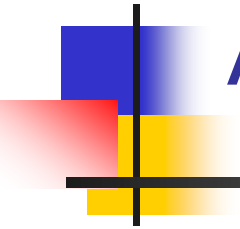
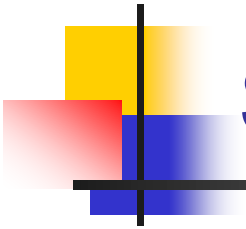


AI, Ethics, and Professionalism

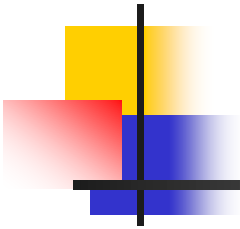
Blay Whitby
blaywhitby.org

blayw@sussex.ac.uk



Structure

- Introductory remarks about ethics
- A case study for discussion
- Some remarks about professionalism
- Take homes

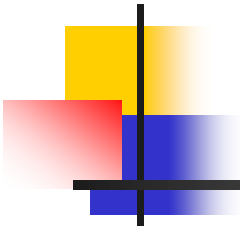


Ethics

- There has been little adoption of any sort of *enforced* ethical standards in AI. (Though there has been plenty of talk.)
- The widespread adoption of AI technology has thrown up plenty of new problems.

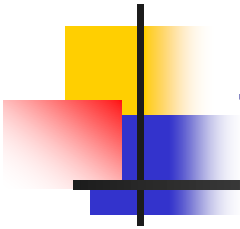
(although they don't seem new to me) ;-)

- BCS, The Chartered Institute for IT has been pushing hard.
- However ethics is not, at present, considered essential for practicing AI and it is often seen as optional in teaching AI.
- Ethical decisions are often disguised as technical or business decisions.



Case Study

- The Amazon recruitment AI system



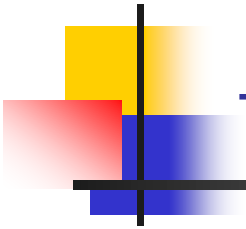
The Amazon recruitment AI system

According to Reuters, Amazon Inc started building a secret AI tool In 2014 that sorted applicants' CVs.

(My best guess is) It was a ANN or multiple ANNs trained on 10 years data of successful Amazon job applicants. It learned about 50k words that showed up on successful applicants' CVs.

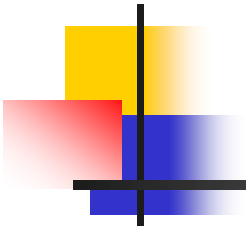
By 2015 it was clear that it was biased against female applicants and becoming more so.

Amazon stopped development in 2017 because they could not fix the gender bias problems.



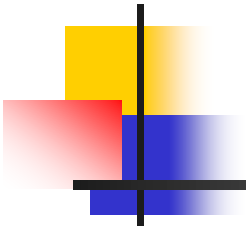
The Amazon recruitment AI system

- There are at least 5 important lessons.



1) You can’t blame the data

- This is ***not*** a case of GIGO.



1) You can’t blame the data

- This is ***not*** a case of GIGO.
- Yes, the training set contained biases (maybe all data does).

1) You can’t blame the data

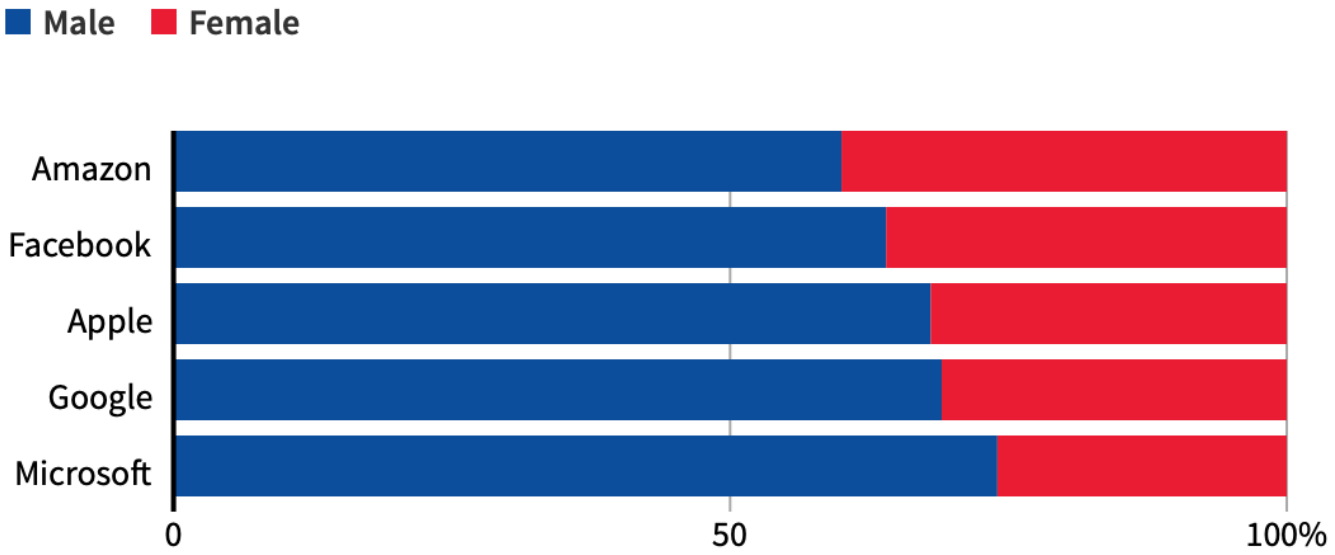
- This is ***not*** a case of GIGO.
- Yes, the training set contained biases (maybe all data does).

- BUT:-

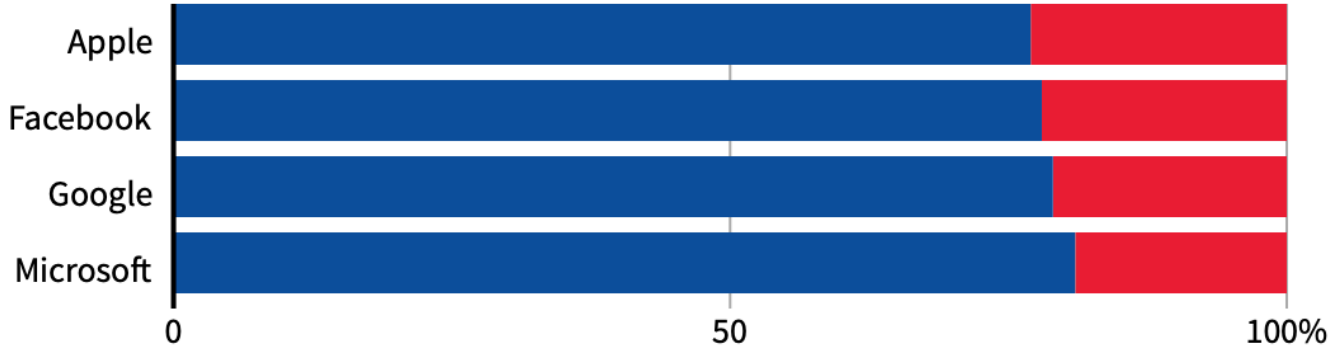
Amazon does not and did never recruit only males.

In terms of overall gender balance, Amazon has a creditable position – at least relative to the big players in the IT industry.

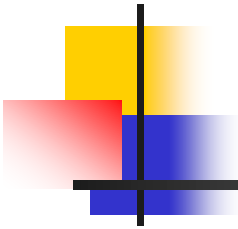
GLOBAL HEADCOUNT



EMPLOYEES IN TECHNICAL ROLES

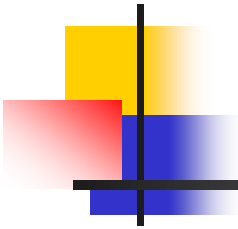


Note: Amazon does not disclose the gender breakdown of its technical workforce.
Source: Latest data available from the companies, since 2017.
By Han Huang | REUTERS GRAPHICS



2) You can’t blame the programmers

- Yes, programmers put their prejudices into their code - often unknowingly.
- Yes, programmers are told to put the prejudices of their organization into the code and they usually do so – almost always unknowingly.

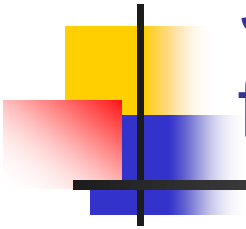


2) You can't blame the programmers)

- Yes, programmers put their prejudices into their code – often unknowingly
- Yes, programmers are told to put the prejudices of their organization into the code and they usually do so – almost always unknowingly.

BUT: no one, either intentionally or unintentionally, programmed this system to discriminate absolutely against women – it acquired that bias all by itself.

3) How can we detect this sort of thing in future?



- Amazon deserve credit for noticing the problem – though I don't know how long this took. (I was told "instantly".)
- Our existing beliefs that we can easily detect when an automated system is making an error now need revising.

3) How can we detect this sort of thing in future?

- Amazon deserve credit for noticing the problem – though I don't know how long this took.
- Our existing beliefs that we can easily detect when an automated system is making an error now need revising.
- AI does not hang, crash, or output gibberish when it goes wrong.
- Instead, it produces a result that is plausible, nearly right, and may back it up with a convincing natural language justification.

4) What makes ***you*** think that you can use AI for this sort of thing?

- AI recruitment and selection programs are very common.
- We are being sold a story that this is the future.
- If a company with Amazon's resources and experience couldn't get it to work – isn't it a little hubristic to think that *you* can?

5) Building and introducing an AI system is ***not*** a purely technical operation

- We need to take account of *broader issues* than the technical specification.

- In the Amazon case study the selection of which training set to use contained a number of assumptions – some of which are social and political but it was treated as a *purely technical* decision.

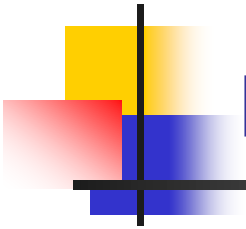
5) Building and introducing an AI system is not a purely technical operation

- We need to take account of *broader issues* than just the technical specification.
- In the Amazon case study the selection of which training set to use contained a number of assumptions – some of which are social and political but it was treated as a *purely technical* decision.
- Training an AI to produce “more of the same” *just is* political!

He should have done my course:

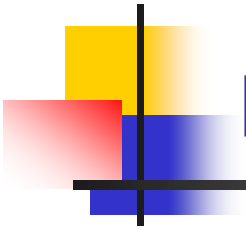
- ZUCKERBERG: We didn't take a broad enough view of our responsibility, and that was a big mistake. And it was my mistake. And I'm sorry. I started Facebook, I run it, and I'm responsible for what happens here.

- So, now, we have to go through our — all of our relationship with people and make sure that we're taking a broad enough view of our responsibility.



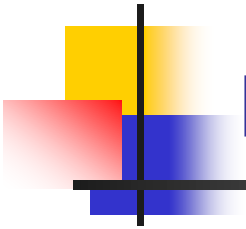
Professionalism – some brief remarks

- Because of at least lesson 5), ethics is not some sort of ‘add-on’ to technical work.
- It needs to be considered at every stage.



Professionalism – some brief remarks

- Because of at least lesson 5), ethics is not some sort of ‘add-on’ to technical work.
- It needs to be considered at every stage.
- There is a significant disparity between IT and comparable areas.



Professionalism – some brief remarks

- Because of at least lesson 5), ethics is not some sort of ‘add-on’ to technical work.
- It needs to be considered at every stage.
- There is a significant disparity between IT and comparable areas.
- Consider CCBT.
- Consider a doctor using an AI diagnosis and treatment system.

My position – you don’t have to agree

- What’s needed in AI is rather more professionalism
- Regulation should be ‘light touch’
- But it has to be international
- The time to do this is now
- There is an existence proof that it can be done.

Thank you for listening

- Whitby, B.R. (2015) Automating Medicine the Ethical Way in Machine Medical Ethics, Volume 74 of Intelligent Systems, Control and Automation: Science and Engineering pp 223-232
- Whitby, B.R. (2012) Do You Want a Robot Lover? in Robot Ethics: The Ethical and Social Implications of Robotics Lin, P., Bekey, G. , and Abney K., MIT Press.
- Whitby, B.R. (2010) Oversold, unregulated, and unethical: Why we need to respond to robot nannies. Interaction Studies, 11:2 pp. 290-294
- Whitby, B.R. (1996) Reflections on Artificial Intelligence: The Social, Legal and Moral Dimensions, Oxford: Intellect Books.
- Whitby, B.R. (1988) Artificial Intelligence: A Handbook of Professionalism, Chichester: Ellis Horwood.