AI Snake Oil: Chapter 3

Why Predictive AI may not live up the promises made on its behalf

A couple of first impressions:

- In the discussion last week, we concluded that a lot of the critiques of predictive AI in Chapter 2 could be boiled down to human errors linked to ‘bad data science’, including:
  - Using training data-sets that do not fit the context e.g. using national data-sets for very local, area-specific problems or local data-sets for national problems
  - Using proxy input or output variables that do not have a good correspondence with the thing you’re actually interested in e.g. health expenditure rather than health need
  - Failing to understand that any predictions arising from training data will reflect the system as it currently is – not what will happen if you change the basic rules of that system e.g. the example on asthma/pneumonia treatment algorithms in hospitals
  - Humans not being willing or able to question or override decisions coming from computers/algorithms
- In contrast, Chapter 3 seemed to me to be a bit more about statistical/mathematical obstacles to predictive AI
- Also, there seemed to be a little more balance (?) in that the authors didn’t claim that predictive AI never works or can never work. They seemed rather to be saying that there are certain social phenomena where we still don’t know if it can work and where it seems much less likely (from a mathematical/statistical) perspective that it will work well. Human agency seems to be a key factor here.
- In Chapter 3, they also posed a couple of interesting ethical questions:
  - How do we define ‘value’ for a predictive system?
  - Is it better to have a prediction system that is somewhat more accurate but opaque or somewhat less accurate but simple/transparent?

Questions – feel free to add, change, skip or completely ignore:

1. To what extent does the group agree with the following statement:

‘Predictive AI has its uses but it is much less useful for social phenomena where individual human agency and chance events play a significant role’

2. The chapter identified some statistical reasons why predictive AI may not live up to some of the promises made on its behalf – do you agree with these? Are there others?

- Social phenomena having a high degree of inherent randomness/chaos (e.g. virality of certain memes where an initial ‘like’ gets amplified very quickly and which may not have anything to do with the inherent ‘quality’ of the meme itself)
- Social phenomena being influenced by rare but significant chance events that may not be reflected even in very extensive data sets e.g. example of child outperforming educational expectations because of the help of a kindly neighbour
- Social phenomena continually evolving in response to stimuli – including the possibility that people might change their behaviour in response to efforts to exert control through predictive AI

3. Putting human error or ‘bad data science’ to one side, how optimistic is the group with respect to AI’s potential to make accurate predictions about the following social phenomena:

    - Election outcomes
    - Wars
    - Cultural phenomena e.g. hit movies, hit songs, viral memes
    - Individual career or educational outcomes

4. What kind of social phenomena does the group think AI can safely and usefully predict? What makes these phenomena more suited to predictive AI?

5. The chapter mentioned a predictive AI system that used 137 data points to make predictions about the likelihood of individual recidivism. According to the chapter, the predictions were not very much better than a simple model based on one variable (no. of prior offences committed by the individual). This is an extreme example but as a general principle, what is the relative importance of transparency vs. accuracy? Are there situations where we might reasonably value transparency more highly than accuracy in decision-making?

6. The chapter repeated an earlier point, which is that predictive AI should not be tested using its training data and that it ought to have some kind of ‘blind’ comparison test to evaluate its effectiveness. In the real world of commercial AI does this happen? What do industry testing protocols for predictive AI look like? Is there any legislation or regulation in this area?