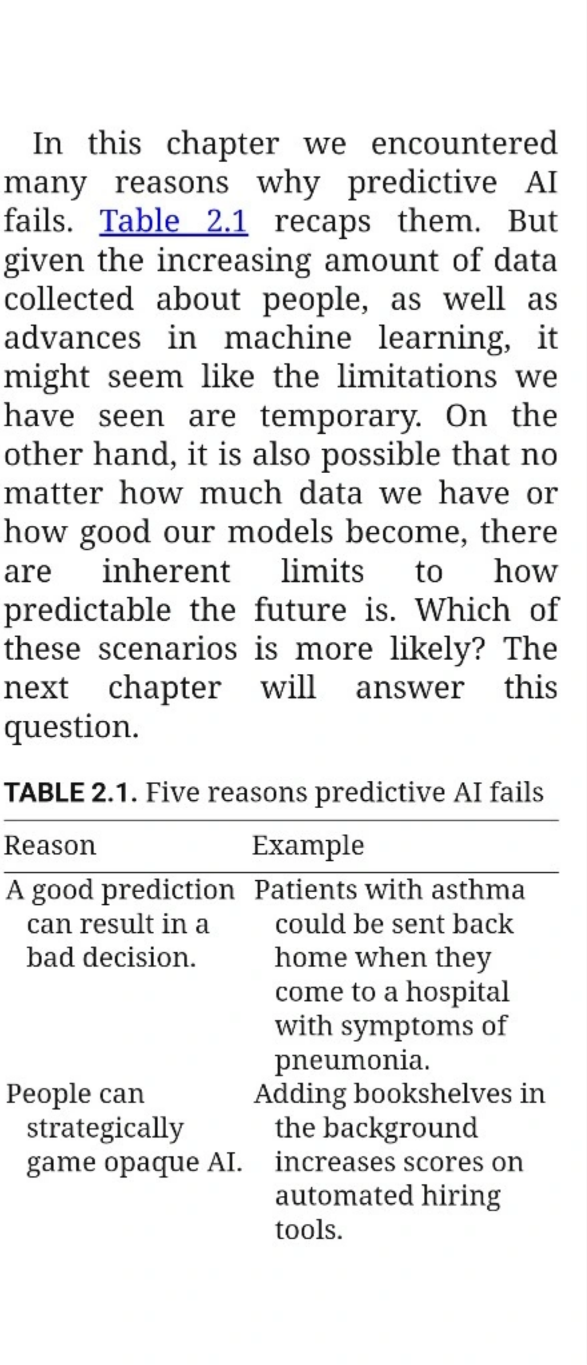
AI Snake Oil – Chapter 2

Why Predictive AI goes wrong

Spoilers:

1. Park any thoughts on generative AI. This is solely about predictive AI

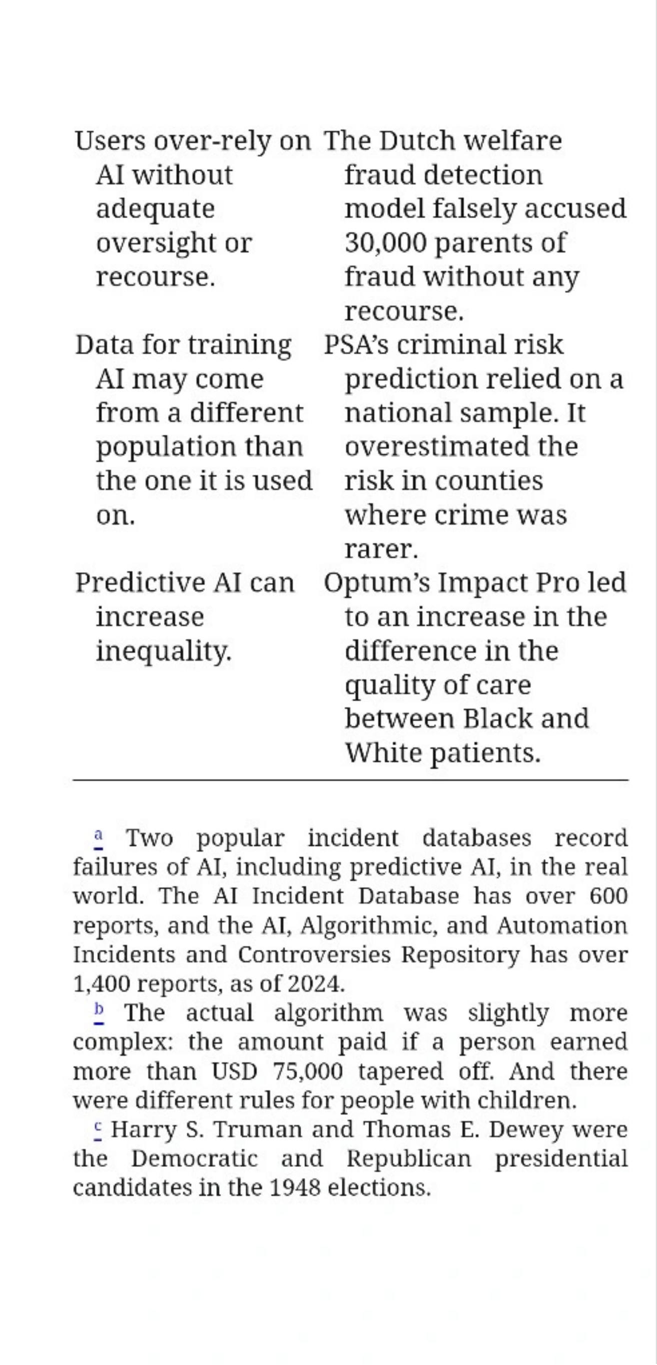
2. Jump to the table at the end of the chapter. Thanks to Andrew for copying



In this chapter we encountered many reasons why predictive AI fails. Table 2.1 recaps them. But given the increasing amount of data collected about people, as well as advances in machine learning, it might seem like the limitations we have seen are temporary. On the other hand, it is also possible that no matter how much data we have or how good our models become, there are inherent limits to how predictable the future is. Which of these scenarios is more likely? The next chapter will answer this question.

TABLE 2.1. Five reasons predictive AI fails

Reason	Example
A good prediction can result in a bad decision.	Patients with asthma could be sent back home when they come to a hospital with symptoms of pneumonia.
People can strategically game opaque AI.	Adding bookshelves in the background increases scores on automated hiring tools.
Users over-rely on AI without adequate oversight or recourse.	The Dutch welfare fraud detection model falsely accused 30,000 parents of fraud without any recourse.
Data for training AI may come from a different population than the one it is used on.	PSA's criminal risk prediction relied on a national sample. It overestimated the risk in counties where crime was rarer.
Predictive AI can increase inequality.	Optum's Impact Pro led to an increase in the difference in the quality of care between Black and White patients.



[a] Two popular incident databases record failures of AI, including predictive AI, in the real world. The AI Incident Database has over 600 reports, and the AI, Algorithmic, and Automation Incidents and Controversies Repository has over 1,400 reports, as of 2024.

[b] The actual algorithm was slightly more complex: the amount paid if a person earned more than USD 75,000 tapered off. And there were different rules for people with children.

[c] Harry S. Truman and Thomas E. Dewey were the Democratic and Republican presidential candidates in the 1948 elections.

Questions / Random thoughts:

How much do we agree with the five reasons? Are there any missing?

> Trying to make predictions with insufficient / poor quality data (St Mary's U / EAB example)
>
> Reuse of data collected for other purposes which contains bias due to its original purpose (pneumonia referrals and hypertension diagnoses, also Allegheny family screening)
>
> Using proxy variables, e.g. arrests / criminality, health spend / need, which have built-in biases.

What makes an 'algorithm' AI? What about expert systems (yes I'm that old...)

COMPAS example: asserts similar input leads to similar output. When is that valid? What are the indicators?

Reuse of datasets in inevitable, due to ubiquity and the cost of gathering new data. What are the pitfalls? What about using synthetic data (not mentioned by the authors).

Would Randomised Control Testing help? Discussed last week. [I believe Control may be very difficult for many AI experiments].

Authors talk about 'gaming' but stop short of adversarial action – maybe in later chapters?

Overautomation – I was expecting Horizon to be covered. Glad to see it's in the exercises! We got past this with alleged ATM fraud in the 90s. What's changed?

Does human oversight compensate against overautomation? What about skills erosion / overseer succession planning?

ORAS / PSA examples show the pitfalls of over- and under-stratification, and of making cross-stratum inferences.

Examples of bias amplification all have proxy variables which display the bias. I'd like to see other examples where bias is introduced by sourcing methods, e.g. the criminality tests.

Why do people and orgs accept (and commission) bad AI?

> Removing 'randomness / chaos'
>
> Offloading responsibility while seemingly retaining accountability
>
> Please add...

yyyy