The Data Science Pipeline: Automation and Ethics

Scott Rubey

Introduction

- The data science pipeline appears straightforward: ingest data, process data, create a model
- BUT:
    - Each of these steps has numerous subprocesses
    - Modern data sizes are huge
    - Raw data is non-uniform and can be incomplete or contain errors

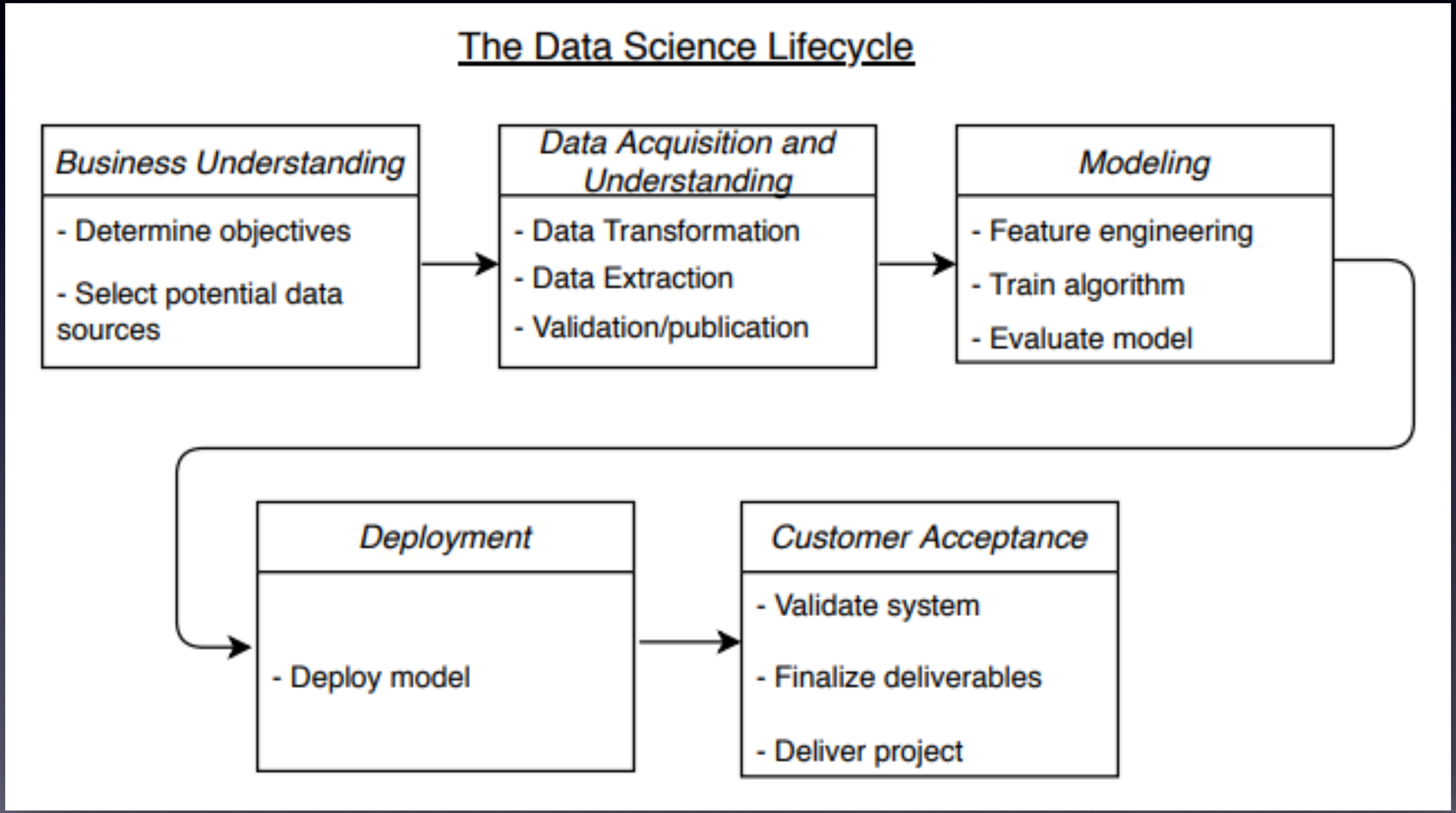
The Solution

- Automation can alleviate manual intervention
- Benefits include:
    - Less time required to create the final model
    - Greater accessibility to those without highly specialized training
    - Cost savings

The Cost

- Perception: computers can't discriminate
- Reality: bias and discrimination are prevalent in data science
    - This can lead to discriminatory models that can negatively impact communities and individuals

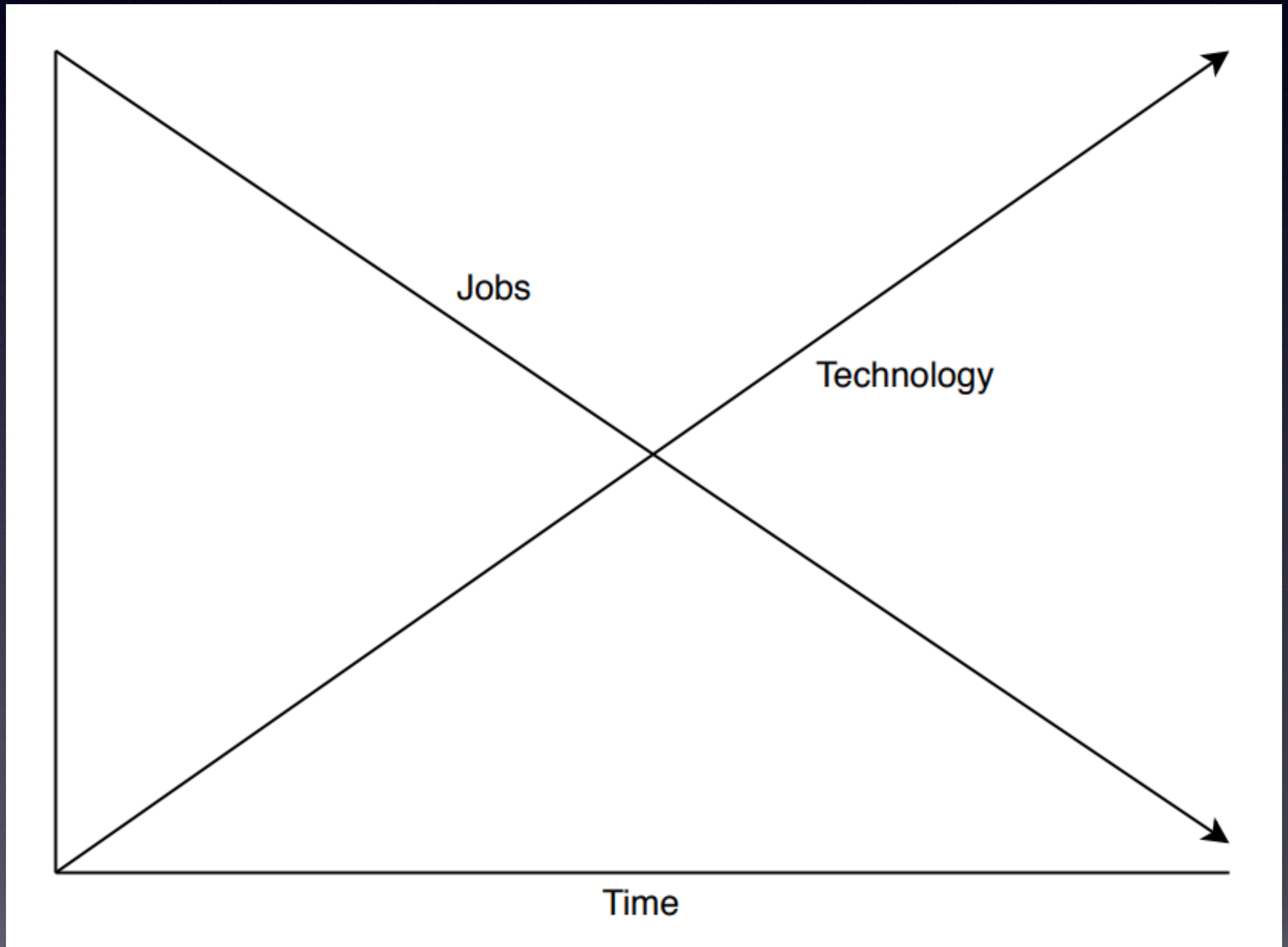
The Data Science Pipeline



Ethical Considerations

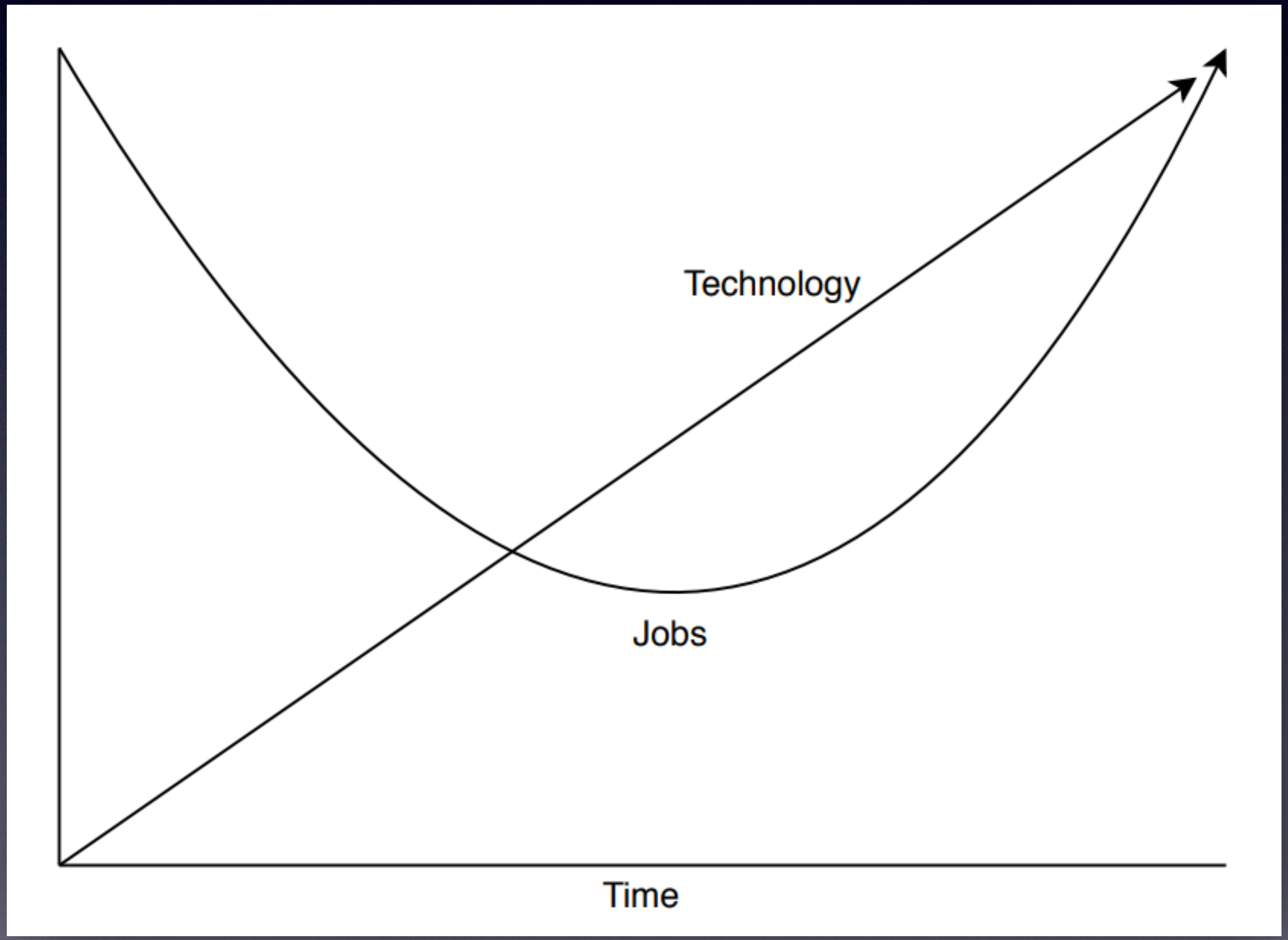
Jobs

The fear...



Jobs

More likely...



Bias

- Can algorithms cure racism, sexism, etc?
- If we could just use data that didn't use protected information...
- Then came that infamous ProPublica study

The ProPublica Study

- Algorithm used to predict recidivism in newly-convicted criminals
- Data is used to determine rehabilitation plans
- A sampling of 7000 cases in Broward County, FL showed African-Americans are 2x as likely to commit further crimes than Caucasians

The ProPublica Study

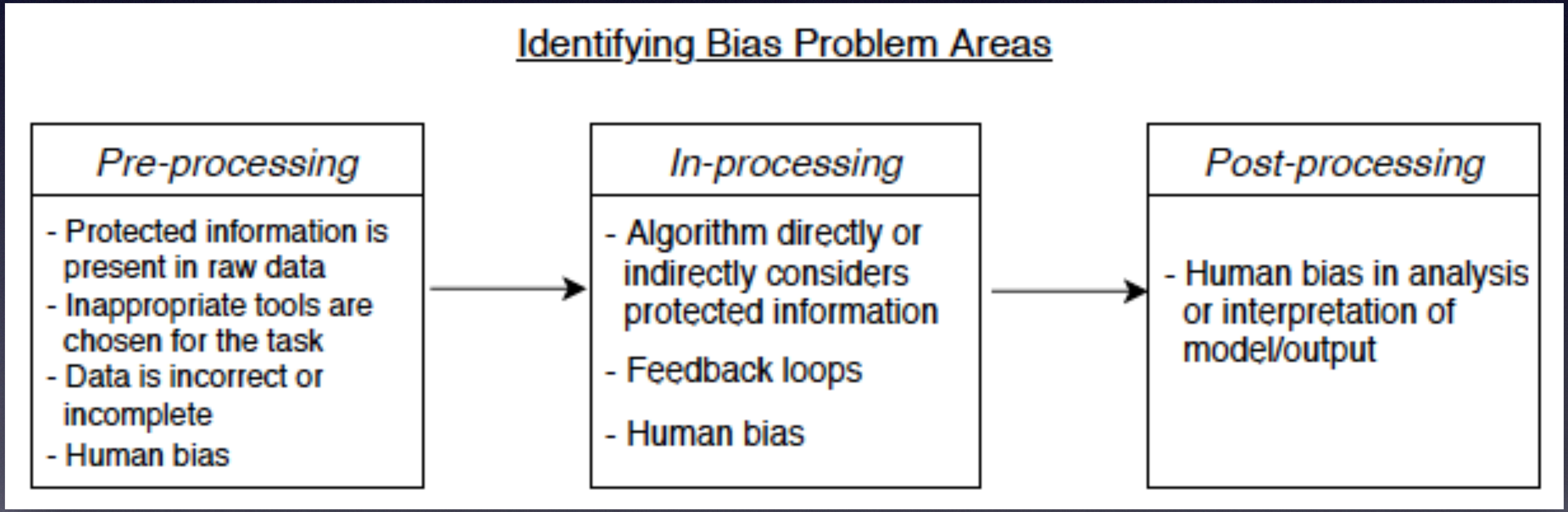
- The algorithm was correct 61% of the time (a little better than a coin flip)
- BUT...
    - Black people 2x as likely to be labeled higher risk but not recidivate
    - Whites much more likely to receive lower risk score but re-offend

The Problem with AI

- Algorithmic bias can occur even...
  - Under human supervision
    - Is the training data biased? Is the human biased?
  - When the raw data is complete/correct
    - Does the data contain protected parameters? Does non-protected data correlate with protected values?
  - When the algorithm does not consider either protected data, nor other data that might correlate
    - Is there a feedback loop?

Solutions

- Identify all potential risk factors for bias



Solutions

- The naive approach: clean the source-data of protected information
- Disparate treatment: discrimination occurs due to explicit consideration of protected class values
- Disparate impact: implicit discrimination occurs due to use of data that correlates to protected class values

Solutions

- Group fairness: an algorithm should yield similar results for all groups
- Individual fairness: an algorithm should yield similar treatment for like individuals regardless of group
- D * X * Y: bias is likely present if some protected class value D can be reasonably predicted by the outcome Y

Conclusion

- Awareness is a good starting point, but that alone won't solve the problem
- Can we eradicate bias from our algorithms wholesale? Probably not, due to human involvement.
- But there's a lot of really great research being devoted to this, so stay tuned...