LDSSA

SLU17 - Ethics and Fairness

December 3rd, 2023

1. Introduction

Motivation

- Doing nothing is doing something

“Let the data set you free” - this is doing nothing

Overview

Objective: create awareness for ethical and fairness topics in data science

We will cover:

- Machine learning social loop
    - Components of a learning system and how it interacts with the world
- Personal data and sensitive information
- Types of bias in data collection and annotation

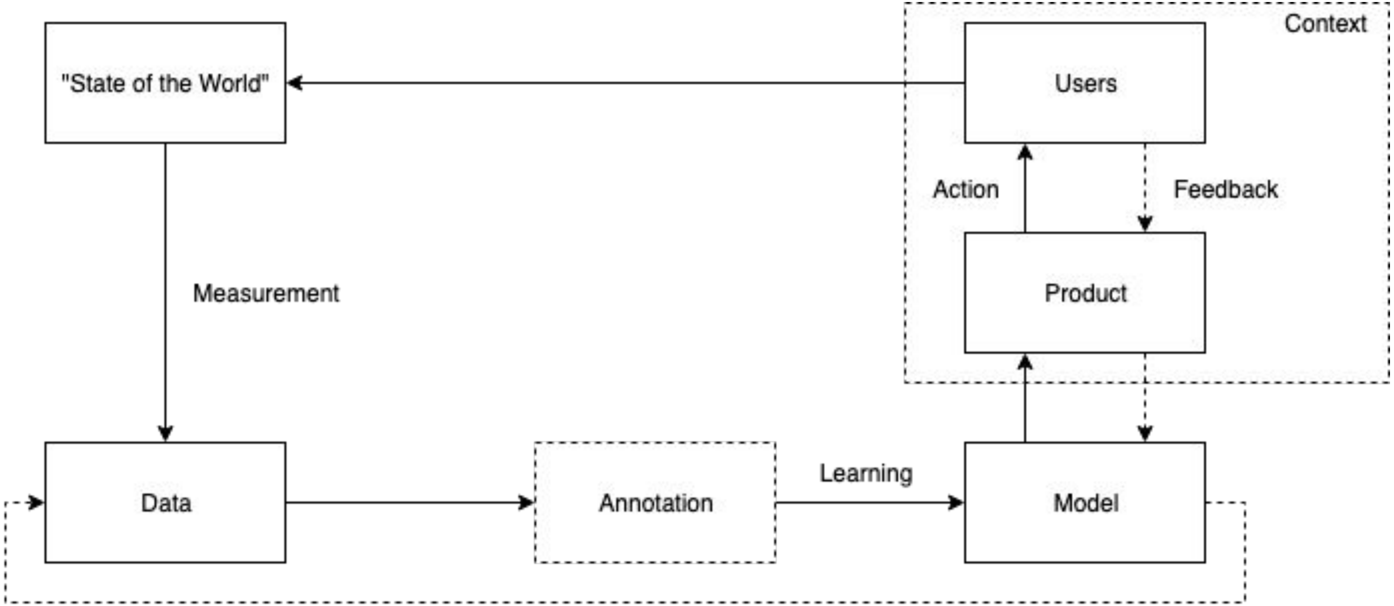
2. Topic Explanation

Learning loop

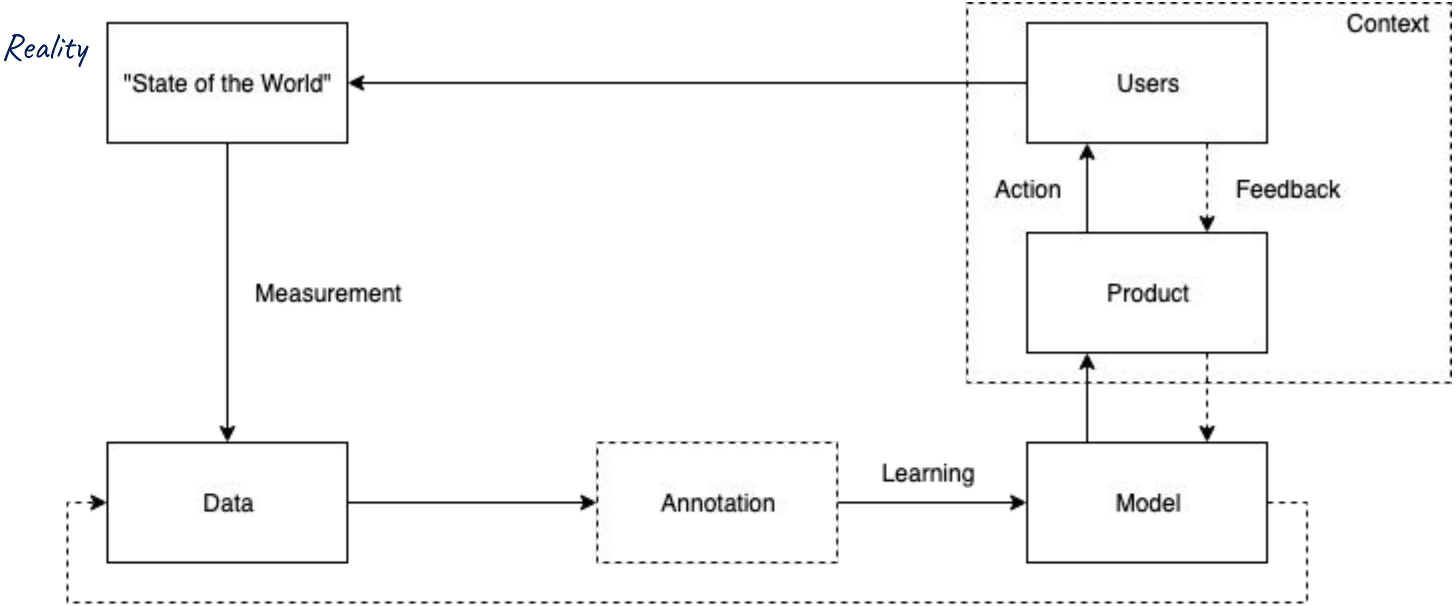
Privacy

Fairness

Meet the data science social loop



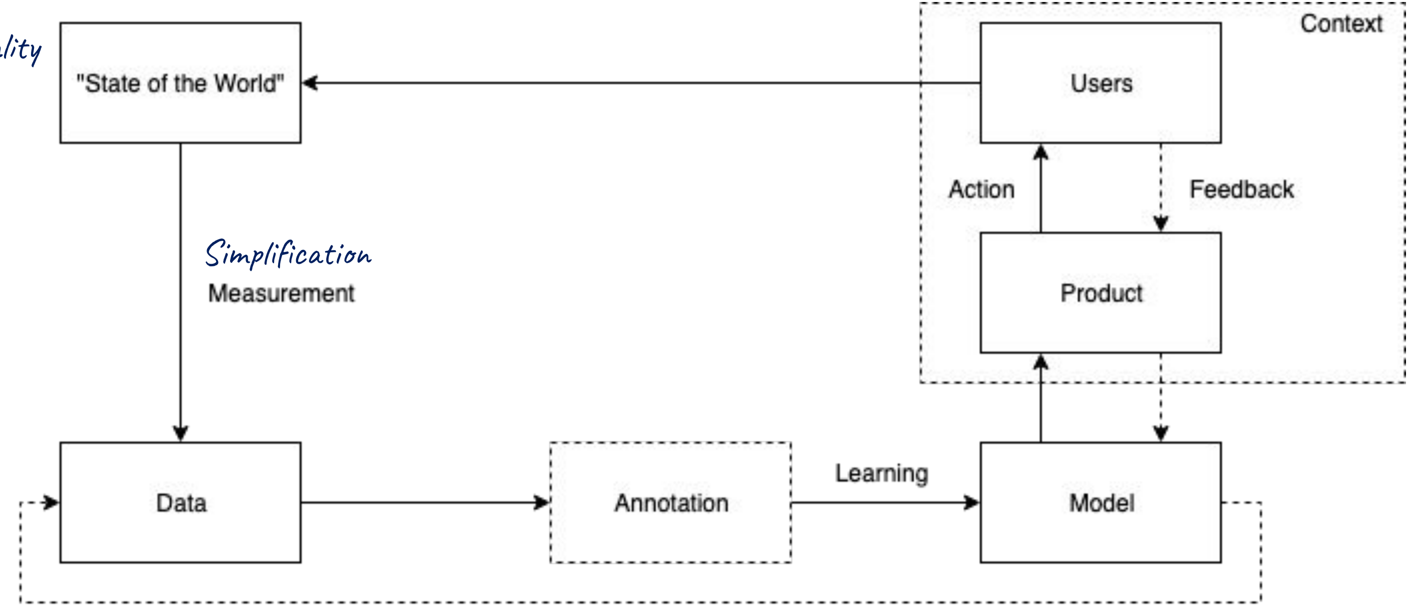
Not a neutral starting point



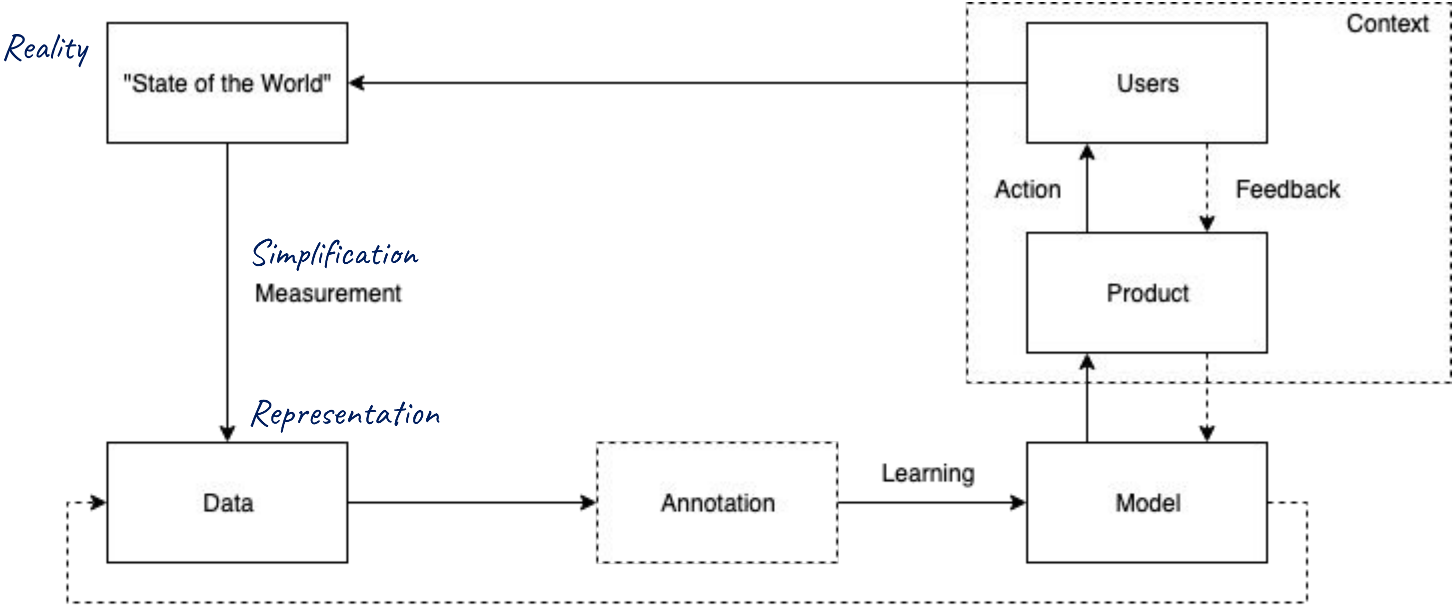
And subjective decisions.

Riddled with technical challenges

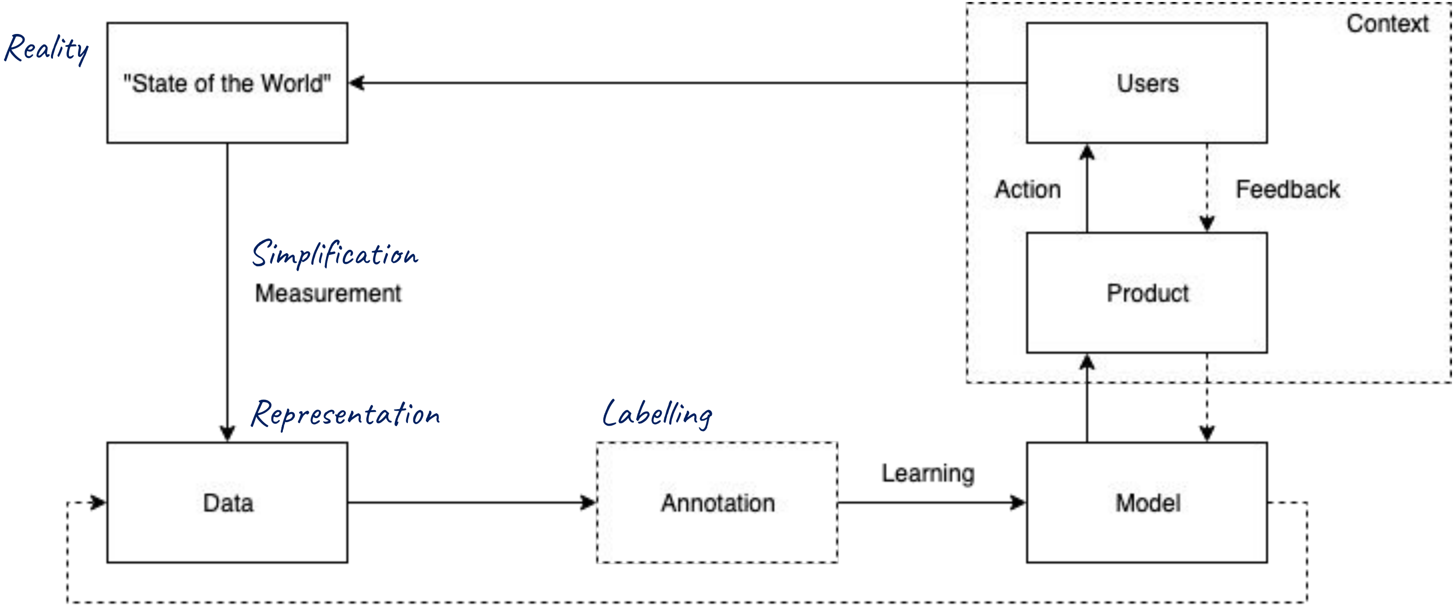
Reality



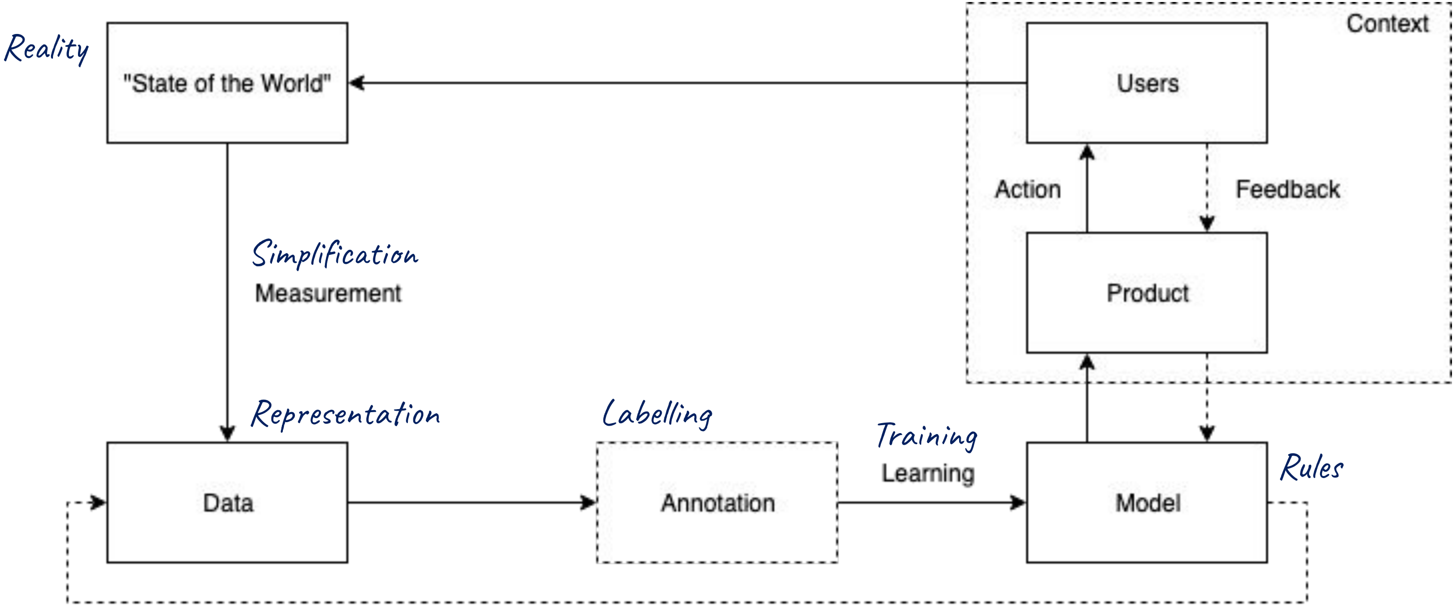
What the model truly “sees”



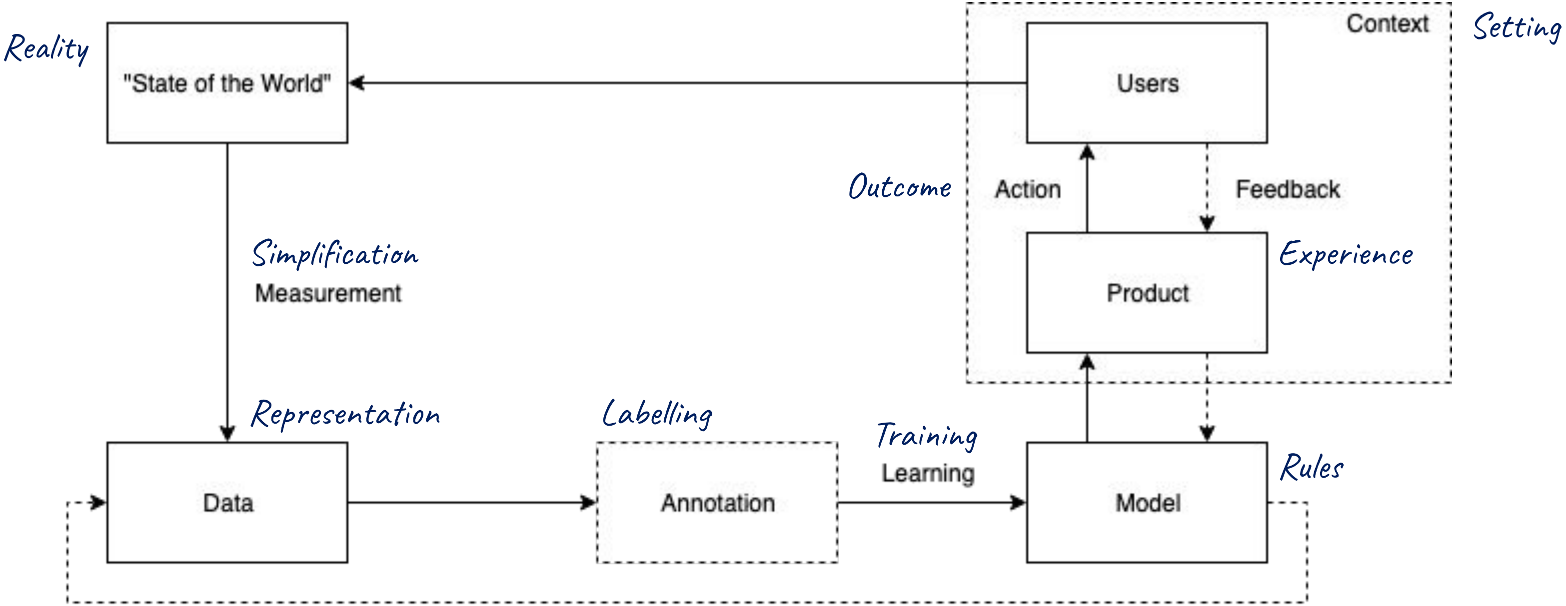
Plus manually imputed knowledge



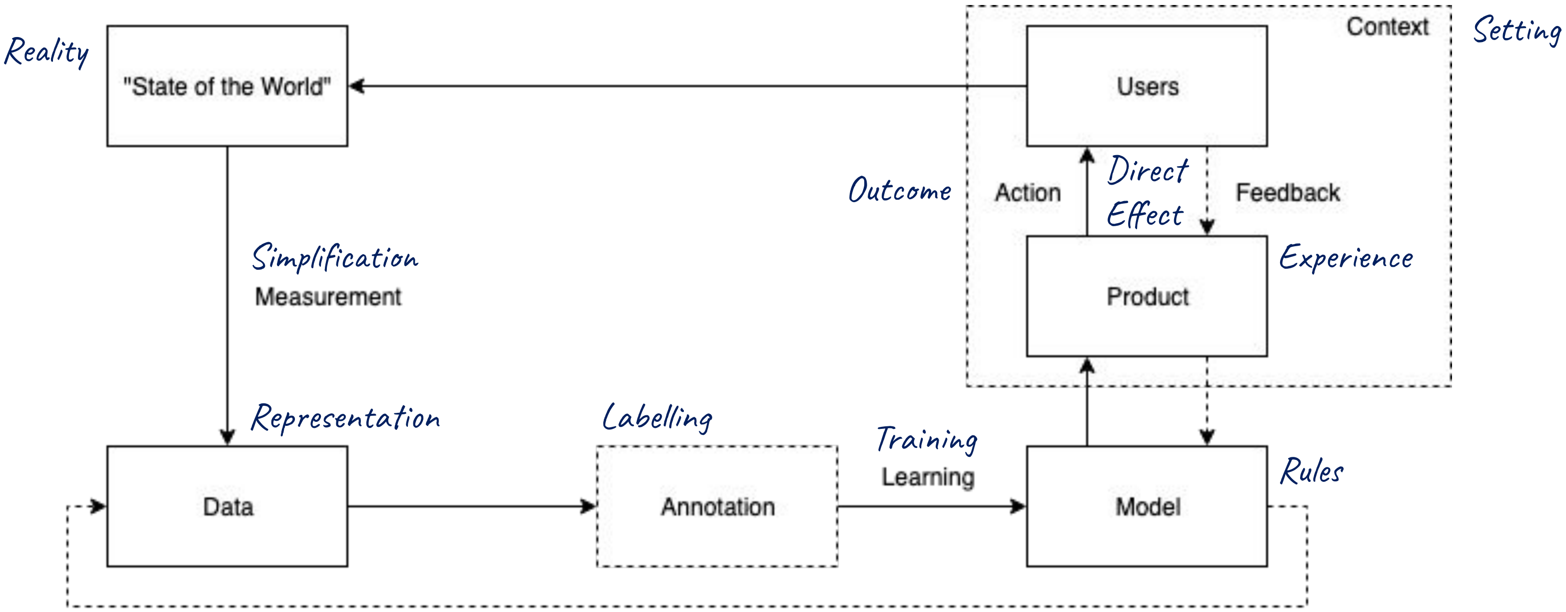
Hopefully it will generalize, though



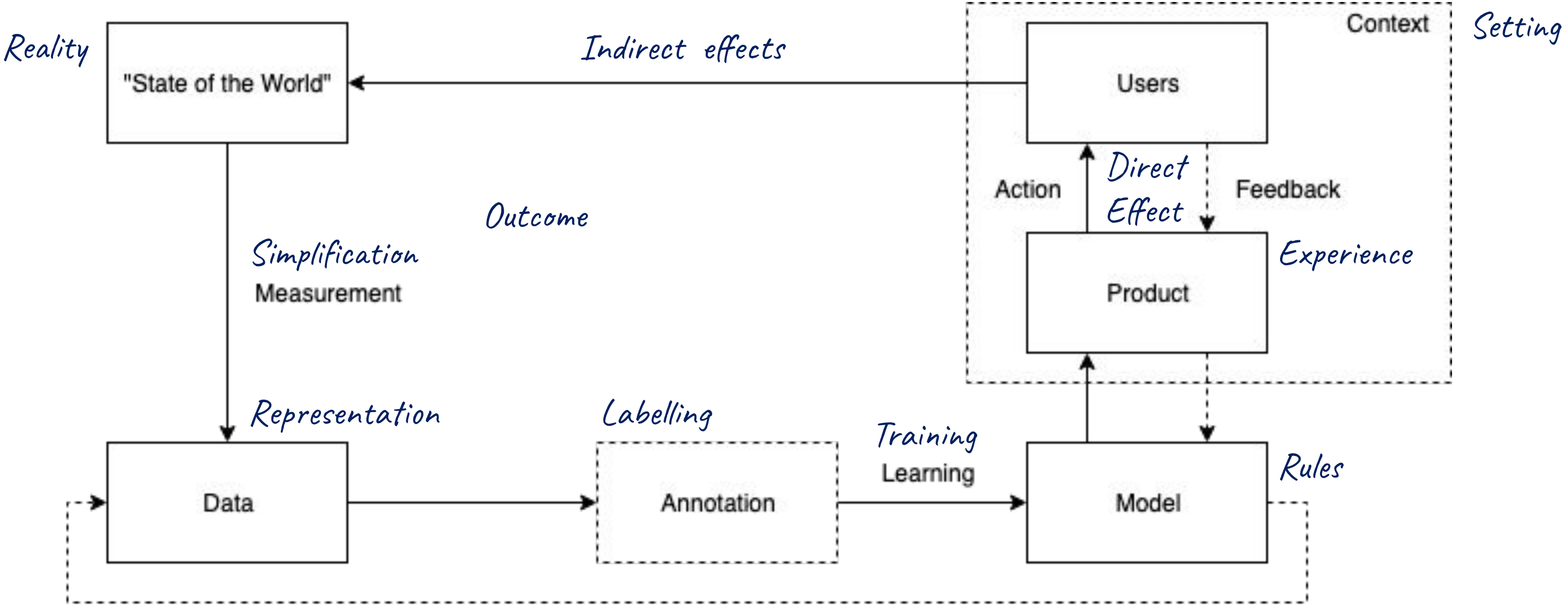
In a controlled environment



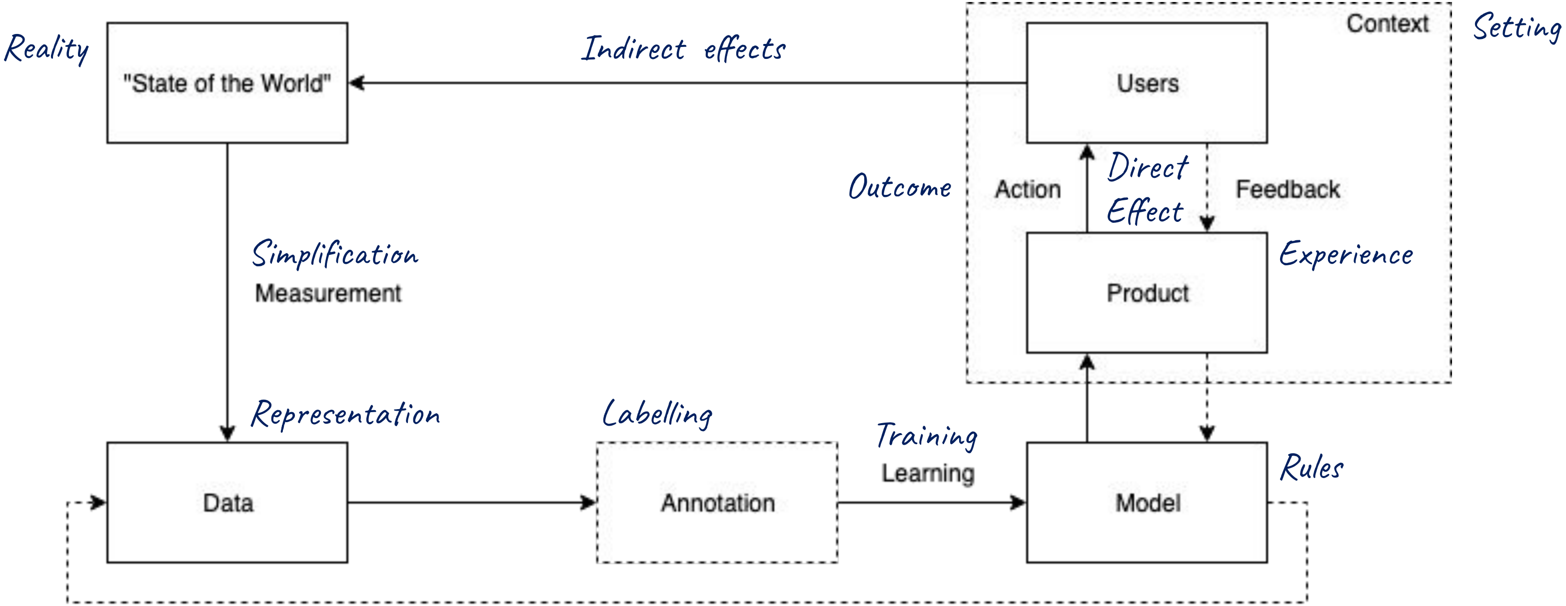
First-order impacts go here



Higher-orders impact go here



Can we have some examples?



Learning loop → Privacy → Fairness

Personal data?

- Any specific information relating to an identifiable person
  - Name
  - Location
  - Physical, physiological, mental information
  - Genetic and biometric data
  - Economic or cultural characteristic

Sensitive data?

- Ethnicity
- Gender
- Political opinions
- Religious beliefs
- Higher level of scrutiny to general personal data

Data Collection Checklist

- Informed consent
- Purpose limitation
- Limited to relevant data
- Data accuracy and updated (if not, you probably should discard it)

Data Storage Checklist

- Secure and protect the data against unintended use
  - Internally
  - Externally (including intentional breach and unintentional exposure)
- Empowers users and subjects of interest
  - Access
  - Rectify
  - Erase their personal data (aka right to be forgotten)

Processing & Modeling Checklist

- Personal information should not be used, unless needed
- Honest representation
- Auditability and reproducibility should be ensured
- Data retention plan (periodically discard unnecessary data)
- Evaluate the model (user and social effects, concept drift, unintended use)
- Be ready to roll-back if you need to

Most of these issues are solved with engineering

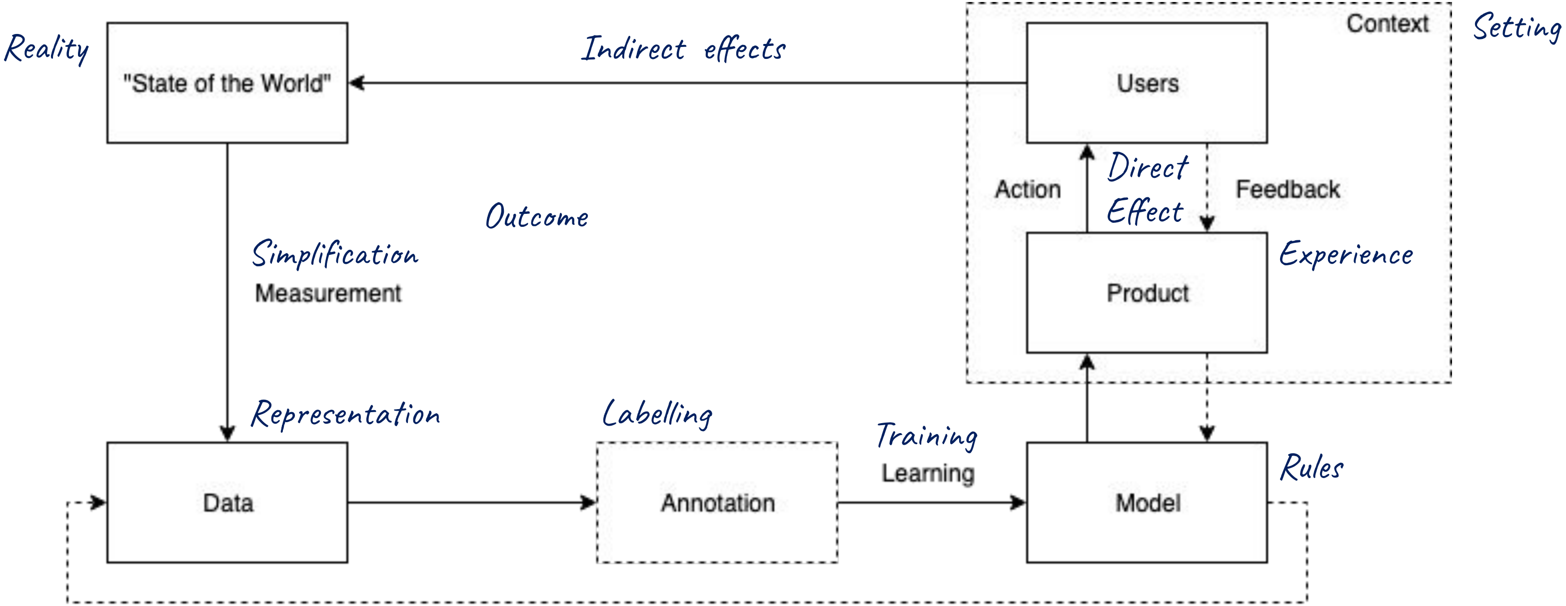
- Document everything
- Create sane APIs
- Have good DevOps
- Engineer your systems so that releases and rollbacks are **business decisions** rather than technical ones
    - i.e. release when you need to release an update, not because it is Tuesday and that's when the release cycle is

Learning loop $\Longrightarrow$ Privacy $\Longrightarrow$ Fairness

Detection of bias

- When evaluating a model we should do more than calculating a loss metric
- Fairness implies fair predictions for different subgroups
  - Audit the training data for data collection and annotation bias
  - Evaluate metrics for subgroups separately (being fairness aware)

Higher-orders impact go here



Be proactive, or else

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

3. Recap

Recap

- Understand the social impact of data science work
- Protect the privacy and security of your users and/or subjects of interest
- Pro-actively audit your data
- Evaluate your predictions for different sub-groups

4. Q&A