AI Ethics Webinar Series: Part III
Ethical Considerations in ML Projects

June 18, 2020

Amy Paul
USAID Center for Digital Development

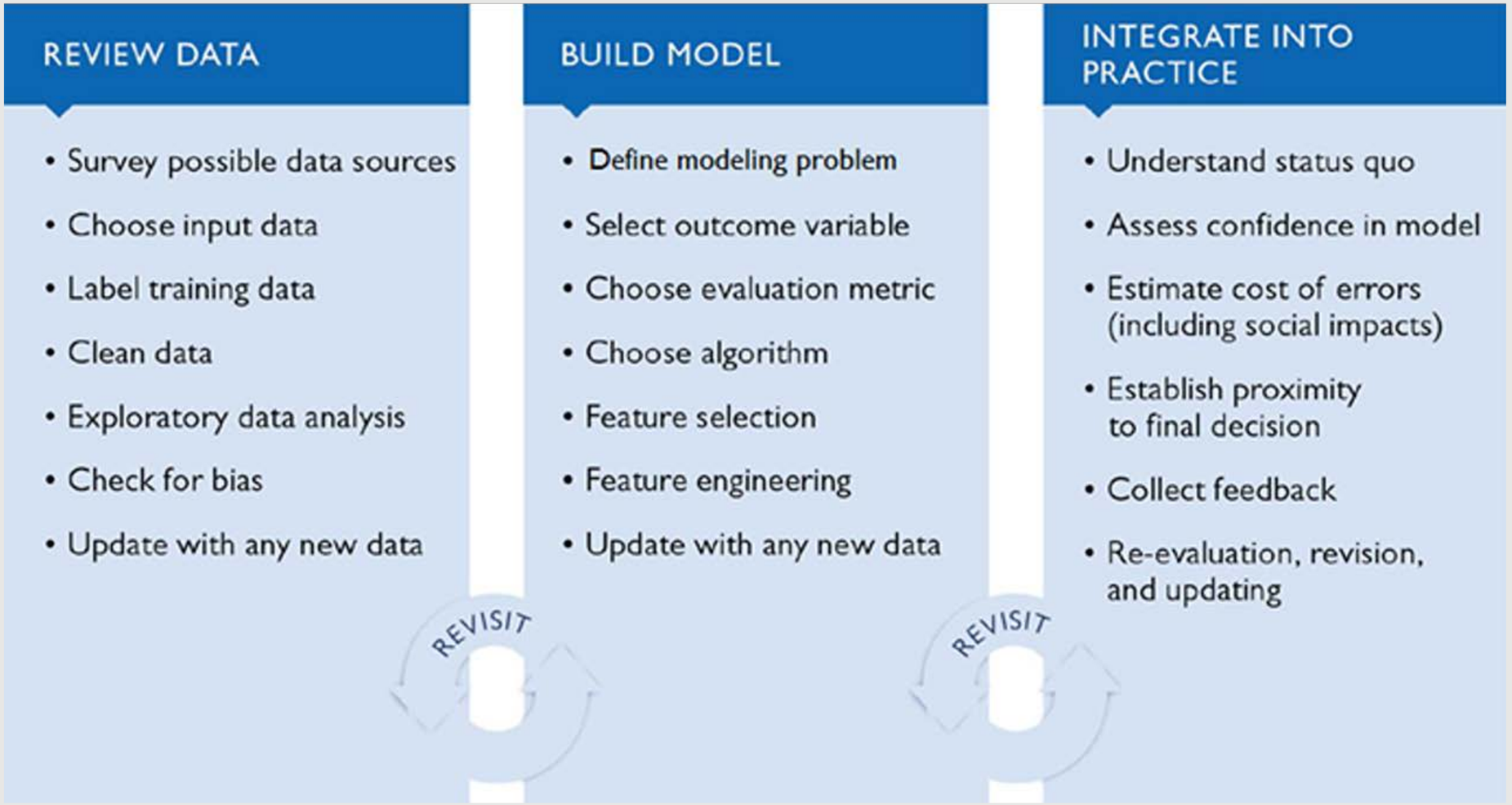
Amit Gandhi
MIT D-Lab

Technical Approaches to Fairness in Machine Learning

- Geared towards those involved in developing machine learning models to be applied in international development contexts

- Focuses on identifying the influences of the choices of model designers and developers in the outcomes that ML models produce

Machine Learning Model Development Process



Values for Machine Learning and AI-
Key considerations relevant to Fairness

- Equity
- Representativeness
- Explainability
- Auditability
- Transparency
- Suitability & Added Value

Part 1:
Introduction to Case Study

Practical Application

Where do these considerations show up in the process of developing machine learning models?

What technical approaches mitigate some of the risks of unfair outcomes?

What are some of the questions that can be asked to make fairness considerations transparent?

Where is context/domain expertise especially needed?

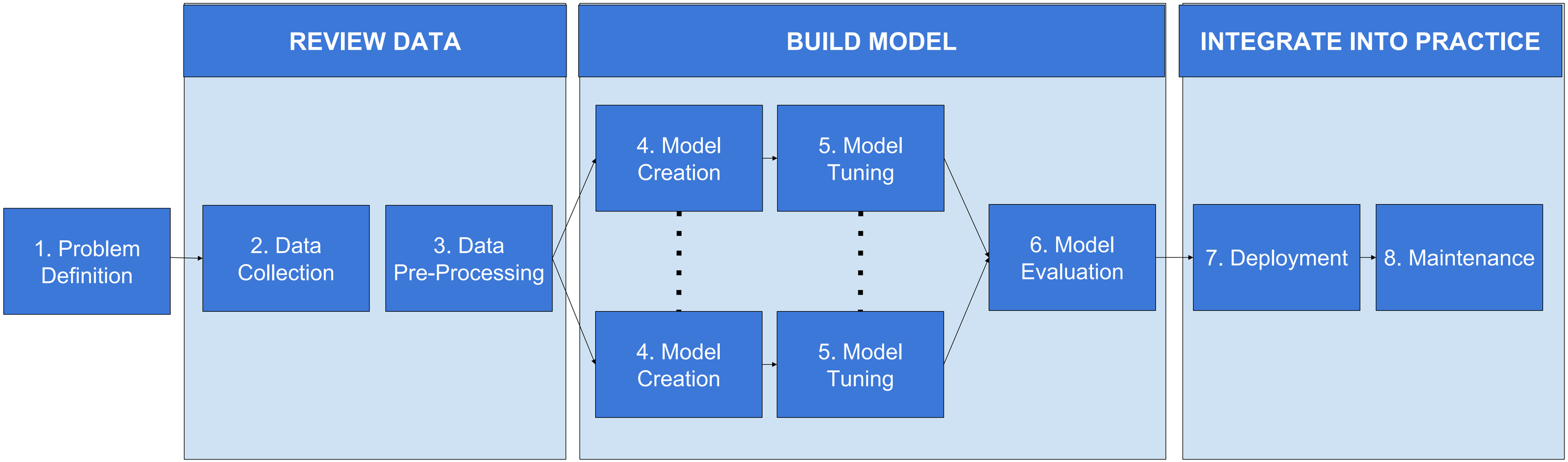
Case Study

We will use a fictional case study about a company that has been providing household-level solar-power products to consumers. The company:

- loans assets (solar panels, devices) to customers and is paid back over 1-3 years
- collects data on users, including demographic information, asset value, product usage, and repayment history
- is looking to expand business by developing creditworthiness algorithms to:
  - help it decide who to give loans to
  - provide information to financial institutions so they can provide loans

Part 2:
Fairness in ML process

Machine Learning (ML) Overview



Problem Definition

Every ML project should begin with a problem-definition phase, where the objectives are defined. This requires gathering and analyzing input from project sponsors, experts, and key stakeholders.

Fairness Considerations: Biases on the part of the people defining the problem, sponsors, or other stakeholders can be introduced.

Protected Attributes

Traits that should not be used as a basis for decision-making in machine learning projects.

Sometimes legally mandated, but often left to the organization and the analyst to define and implement.

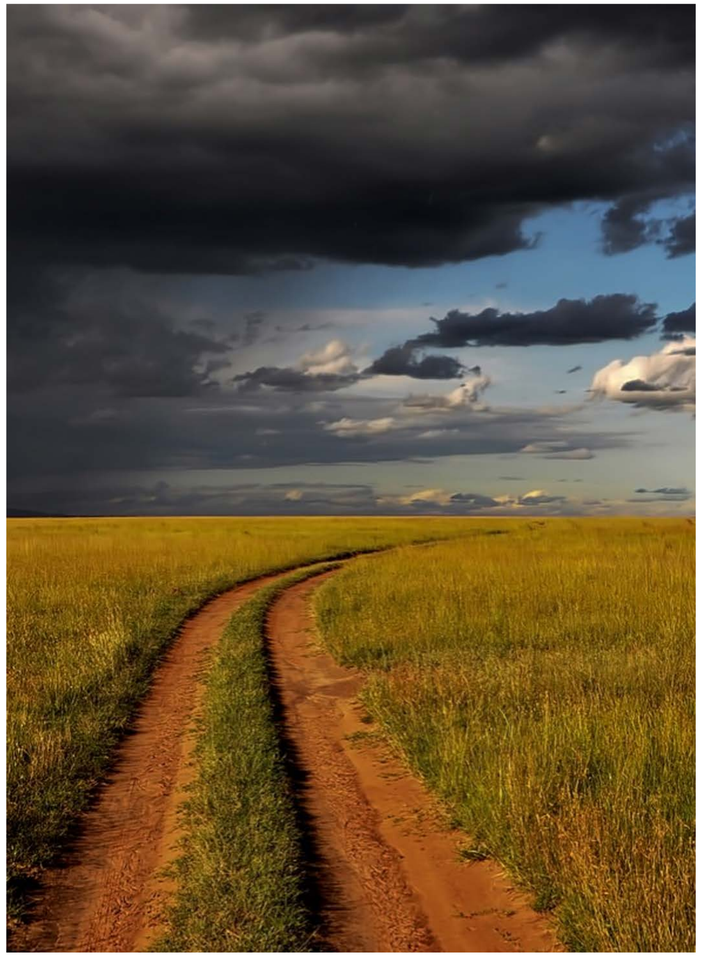
Typically include:

- race
- age
- gender
- religion
- socio-economic status

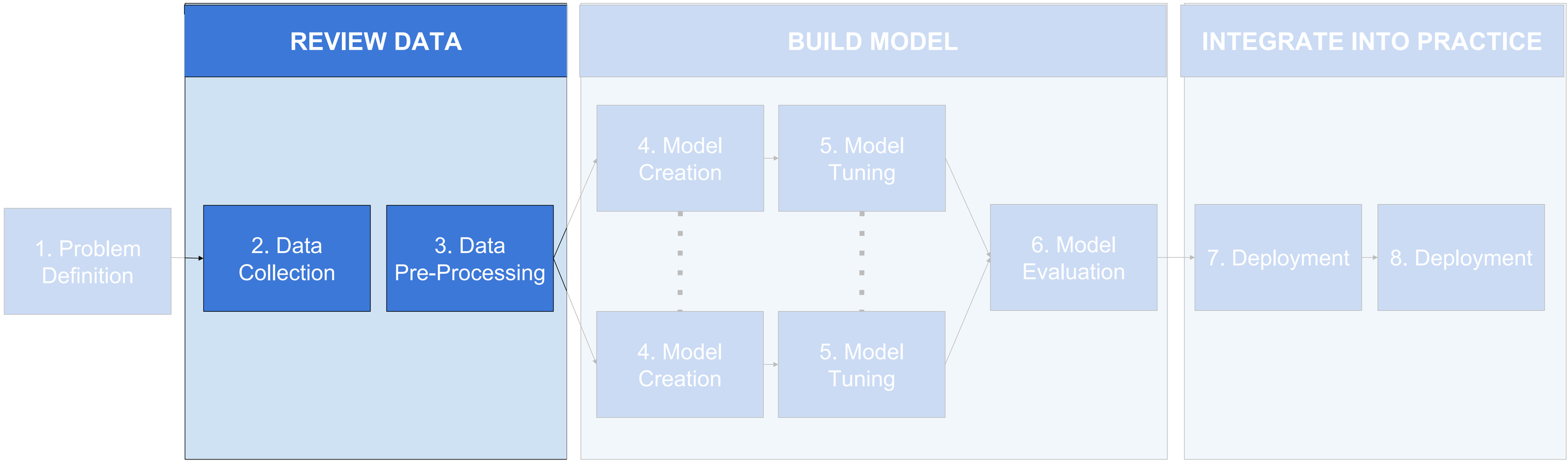
Case Study: Asking the Right Question

Make sure you are asking the right question:

- “Who is likely to repay loans?” rather than “Who have we given loans to before?”
    - Should off-grid areas be preferentially targeted vs. those in urban areas who are supplementing grid power with solar power?
- Is repayment of loans from solar products actually an accurate predictor of creditworthiness?



Review Data:
Data Collection and Data Pre-Processing



Data Collection

This involves aggregating data collected by the organization or from external sources. May include conducting a study to collect field data, purchasing data sets, or collecting data from published sources.

Fairness Considerations:

- many systematic biases can be introduced at this phase:
  - choosing which type of data to collect introduces bias
  - the way data is collected introduces bias
  - external data that is being used may have its own biases
- protected attribute data must be collected

Data Pre-Processing

This phase is primarily cleaning and labeling of data, including extraction and transfer of data to a form suitable for ML. Cleaning refers to identification and correcting (or removal) of erroneous data. Labeling refers to to assigning tags to data to indicate the quantity the user is trying to predict.

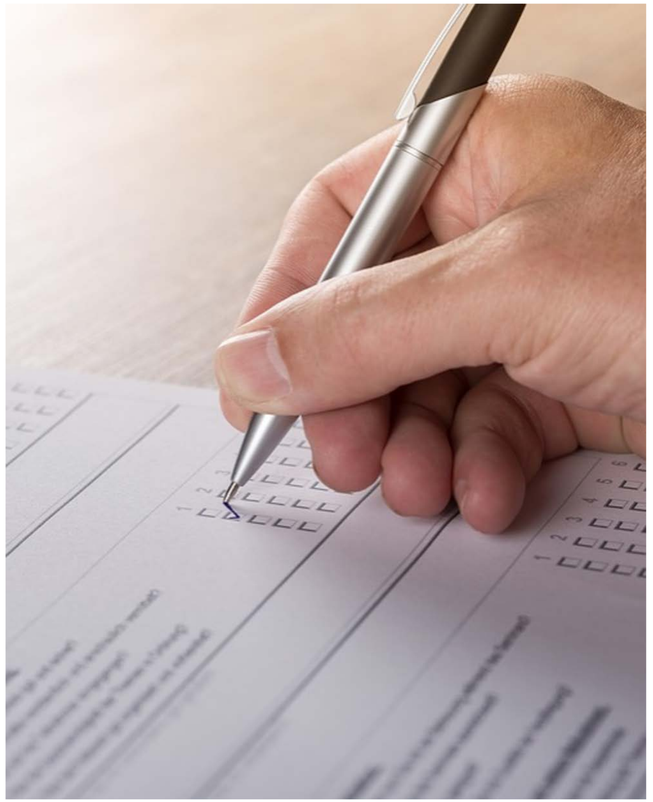
Fairness Considerations: this phase can propagate biases in the data collection, or introduce new biases from the data labelers

- if data from a specific subgroup (e.g. written vs. digital records) is harder to collect and clean, it may get omitted
- data labelers may introduce biases in their labeling

Case Study: Dealing with Representativeness

Assume data for loan repayments and client information can be collected electronically or in handwritten records in various languages, depending on location.

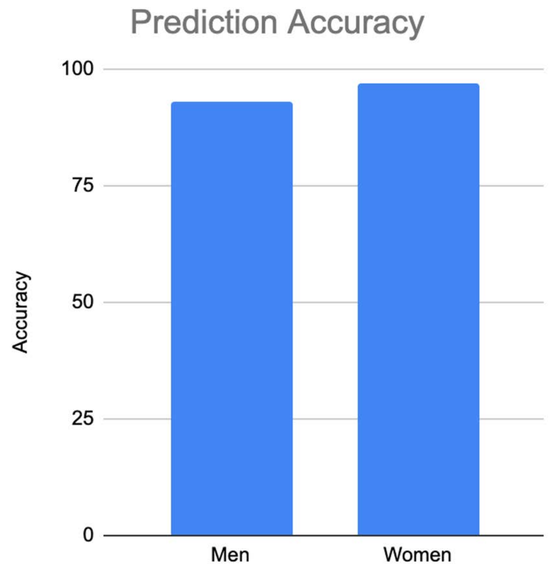
- electronic data is easy to process, where transcription of written records is expensive, time-consuming and can introduce errors
- translation of information between languages is expensive



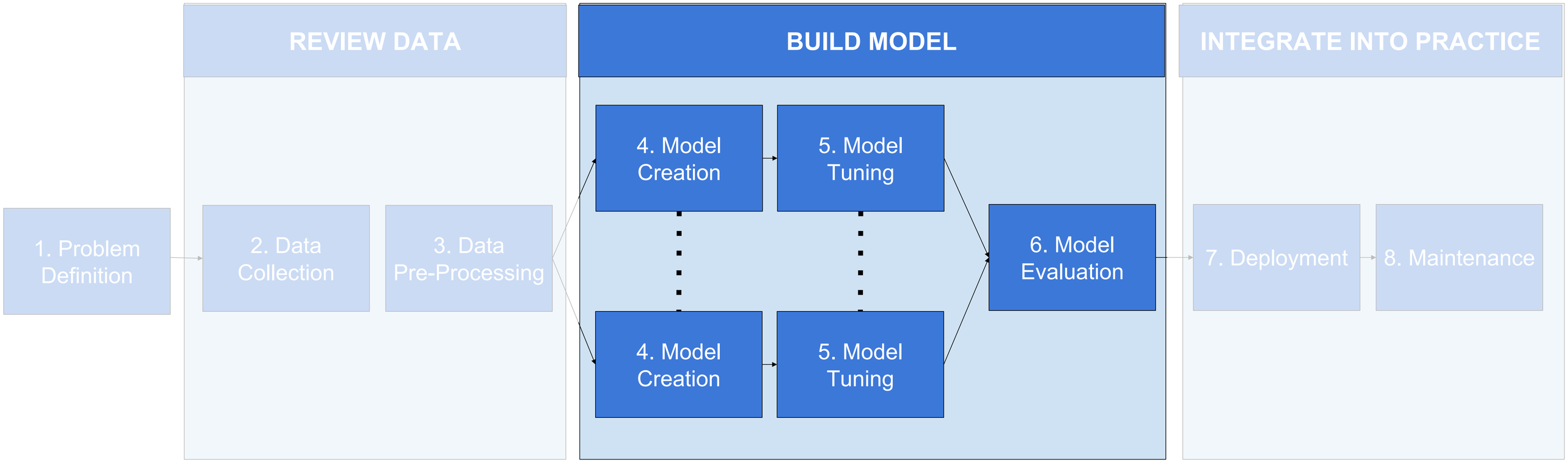
Case Study: Dealing with Representativeness

How do we address issues of representativeness?

- Check to see if we actually need representative data
- Implement techniques to overcome lack of representativeness in data
  - data augmentation
  - bootstrap sampling / synthetic data
- Collect more data



Build Model:
Model Creation, Tuning, and Evaluation



Model Creation

This step involves selecting and developing potential models using the data. Starting with the problem definition, the analyst chooses potential algorithms. To build the model, the analyst splits the data into a training set for building the model and test set for comparing model performance.

Fairness Considerations: analysts should understand the strengths and limitations of different algorithms.

- choosing an algorithm in general introduces the analysts biases
- choosing specific algorithms can propagate biases in data
- algorithms often have tradeoffs between speed, accuracy, and explainability

Case Study: Choosing an Algorithm

One algorithm that the analyst could implement is the k-nearest neighbors (KNN) classifier.

While simple to implement and accurate, it can propagate biases in data and is difficult to explain.

- gender-bias in historical data would result in gender-bias in loan decisions
- cannot clearly tell why someone was not given a loan, making it difficult to identify biases and communicate how clients can improve their creditworthiness.

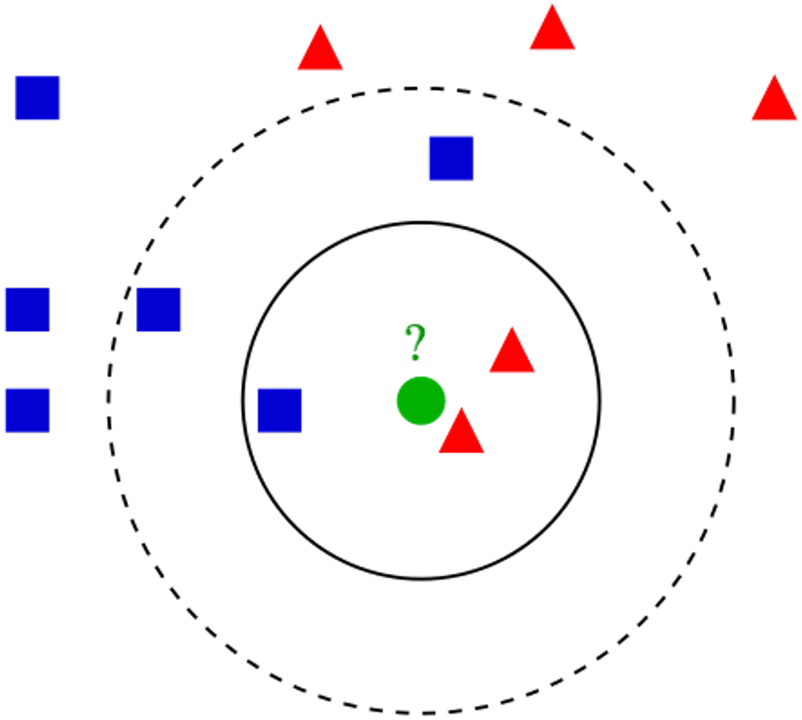


Image Credit: Antti Ajanki
https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

Model Tuning

Models often contain thresholds or hyperparameters that control how the learning process happens. Choosing these values appropriately is critical to having a functional solution.

Fairness Considerations:

- tuning hyperparameters change the underlying model
- setting thresholds can require the analyst to make tradeoffs between aggregate performance and performance for groups or individuals

Case Study: Model Tuning

Let’s say we implement a logistic regression model and the data indicates that men and women have different default rates, which the analyst interprets as different likelihoods of loan repayment. Threshold values can minimize errors for men, women, or both groups.

The choice of how to implement fairness is a decision that needs to involve all key stakeholders.

Image Credit: Rich Fletcher, MIT D-Lab
Exploring Fairness in Machine Learning for International Development
https://d-lab.mit.edu/resources/publications/exploring-fairness-machine-learning-international-development

Case Study: Choosing How To Implement Fairness

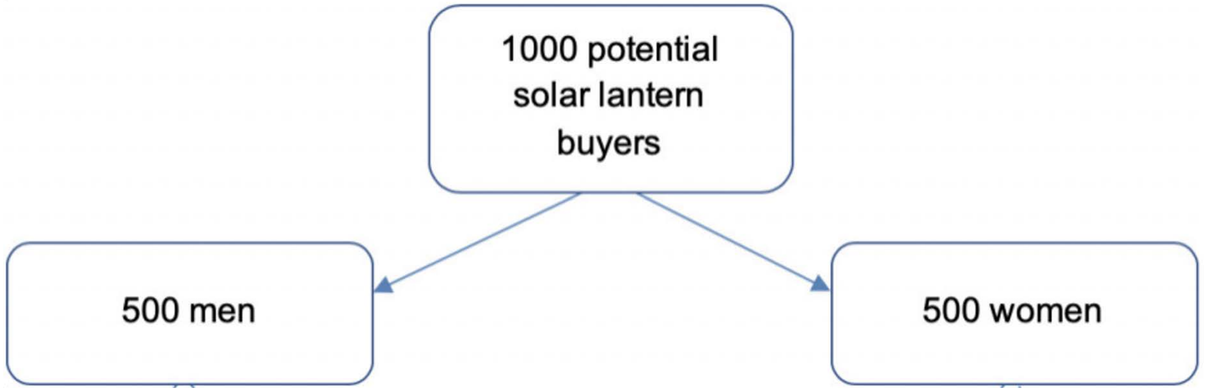
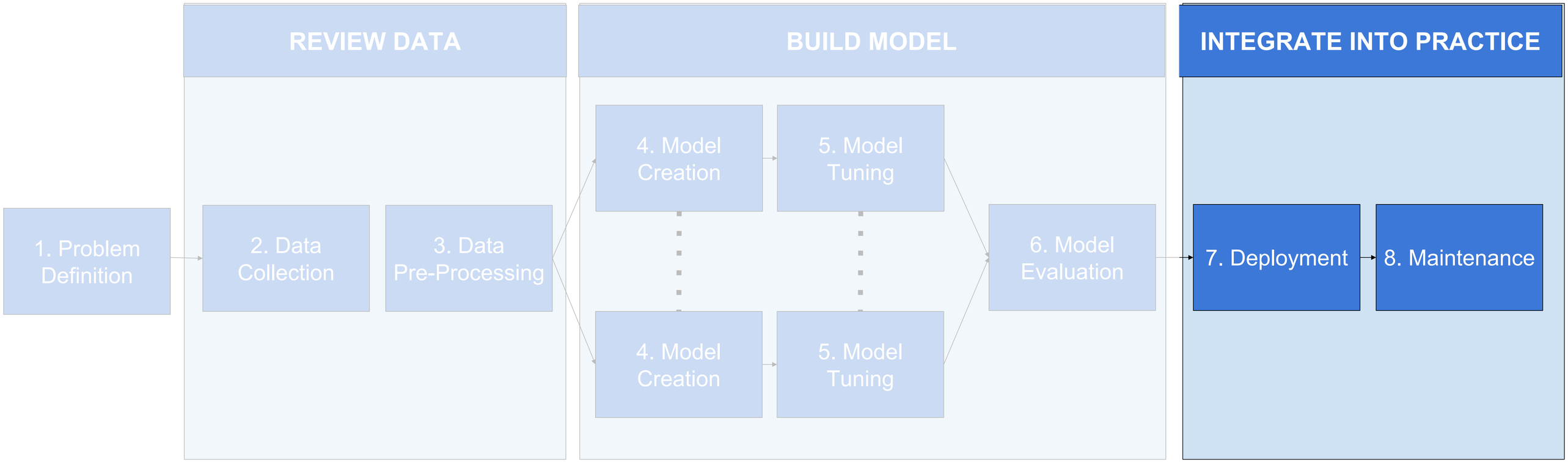


Image Credit: Dan Frey, MIT D-Lab and Mike Teodorescu, Boston College
Exploring Fairness in Machine Learning for International Development
https://d-lab.mit.edu/resources/publications/exploring-fairness-machine-learning-international-development

CRITERION	ADVANTAGES	DISADVANTAGES
Fairness through unawareness	• Simple to implement	• Not effective unless some unusual criteria are satisfied (no correlated attributes)
Demographic parity	• Conceptually simple • Can have legal standing (disparate treatment)	• Does not address individual-level fairness • May unacceptably compromise prediction accuracy
Equalized opportunity	• Appeals to a reasonable interpretation of fairness • Can be a good option if the true positive rate is most consequential factor	• Disparate false negative rates may remain between two populations • Requires lots of labeled historical data
Equalized odds	• Appeals to a reasonable interpretation of fairness	• May not address group disparities sufficiently • Can be inconsistent with high levels of accuracy

Integrate into Practice
Deployment and Maintenance



Deployment

In this step, ML algorithms are deployed in the field. Often, there is a beta phase where the deployment is small and manually audited and then it is scaled up (or scaled across regions). This also involves integrating the ML system into decision-making processes.

Fairness Considerations: the ML model must be used as intended, and limitations and processes for accountability should be clearly communicated to by users, including:

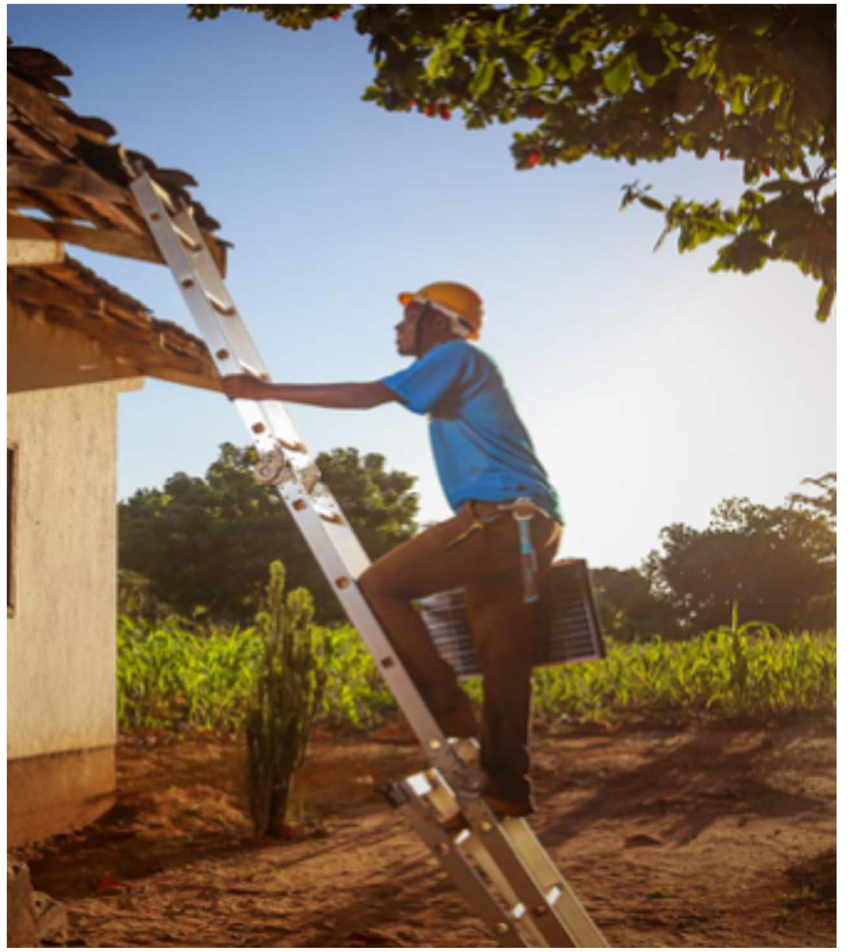
- how the model works: the approach, accuracy, errors, and trade-offs made in the model design
- where, how and for how long the model should be used based on the representativeness of the training data

Case Study: Deployment

Company uses a creditworthiness model to determine which clients should be offered loans for purchasing a solar-powered television.

The model determines places with the highest solar usage were good candidates which could result in the company only providing solar televisions in a specific area of the market.

Could unfairly impact individuals living in other areas.



Maintenance

After deployment, the model is monitored as new data comes in to keep it up to date and retraining it, if necessary.

Fairness Considerations: Lack of bias in model evaluation does not guarantee lack of bias at scale. Over time the training dataset can become less accurate. It is important to:

- regularly audit models
- rectify any biases or unintended negative consequences

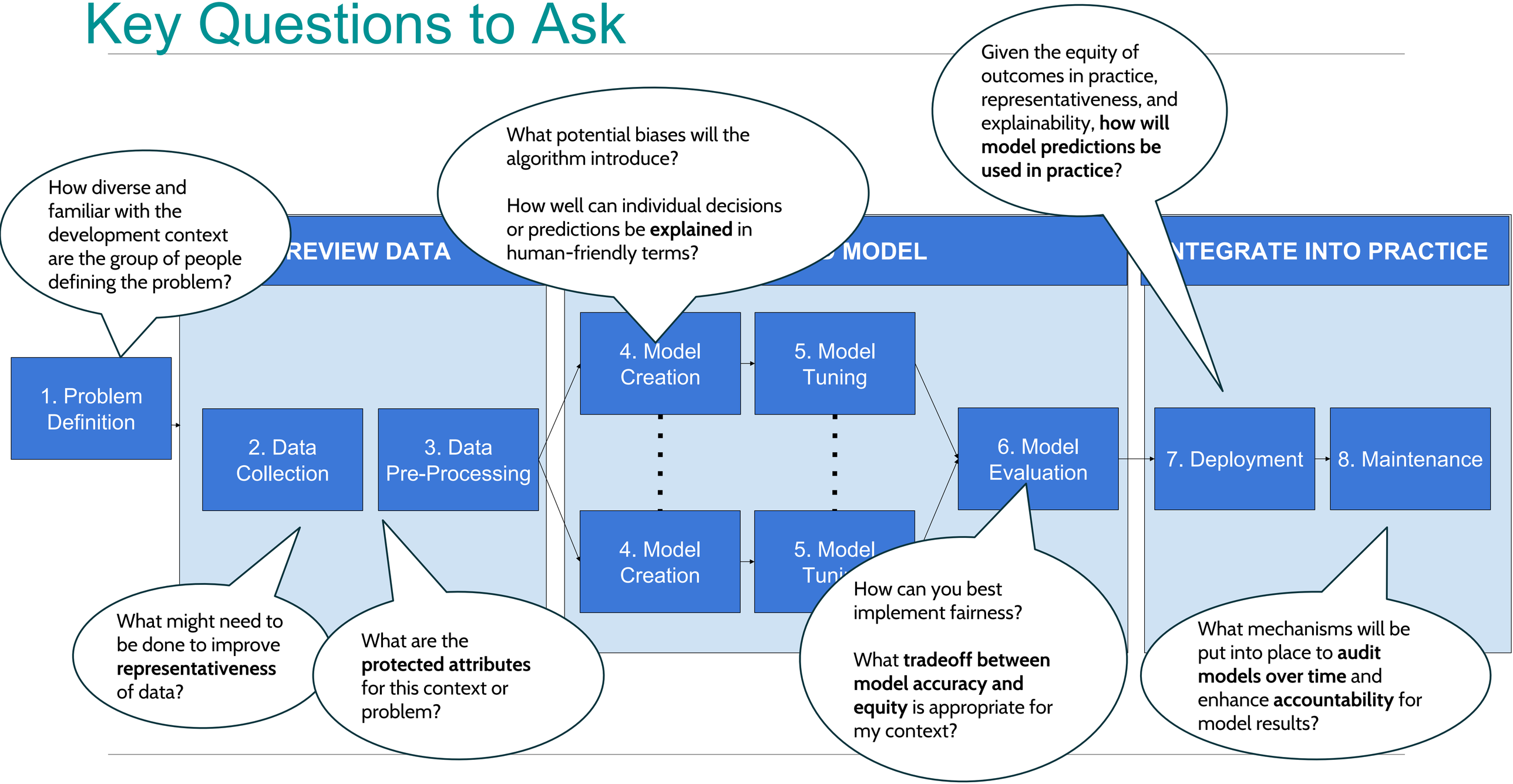
Case Study: Maintenance

Country implements a kerosene-tax that disproportionately affects the rural population of solar asset users.

While the population is adjusting, the model may reject loans because disposable income has gone down.

This would be the opposite effect that the company would want to have, where providing loans at this time would help the population. The model would need to be adjusted in order to account for changes.

Key Questions to Ask



AI Suitability Framework for Nonprofits: https://nethope.app.box.com/file/643893297448