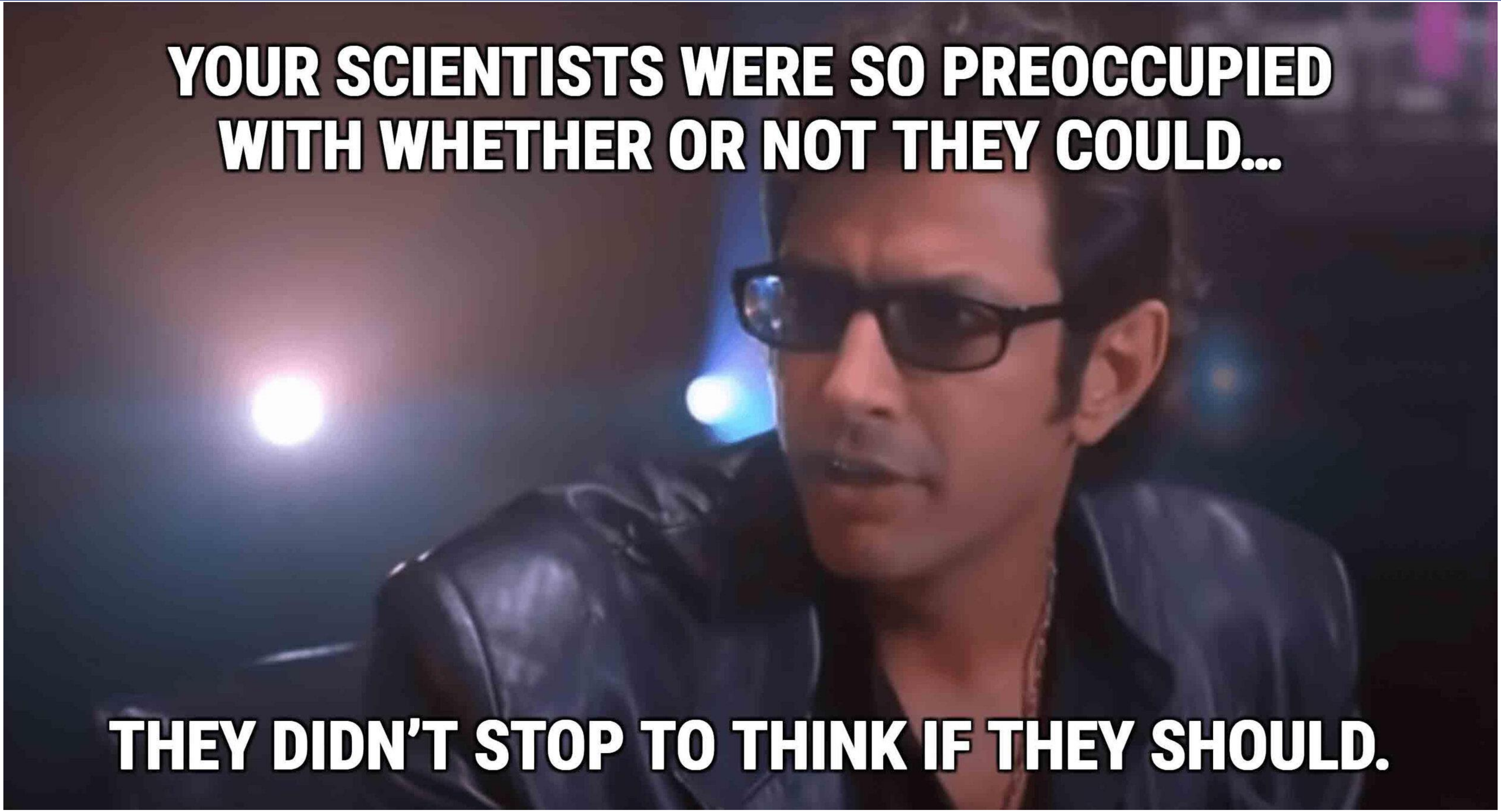
CS3121 - Introduction to Data Science

Ethics

Dr. Nisansa de Silva,
Department of Computer Science & Engineering
http://nisansads.staff.uom.lk/



Jurassic Park (1993)

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

What is common in these pictures?

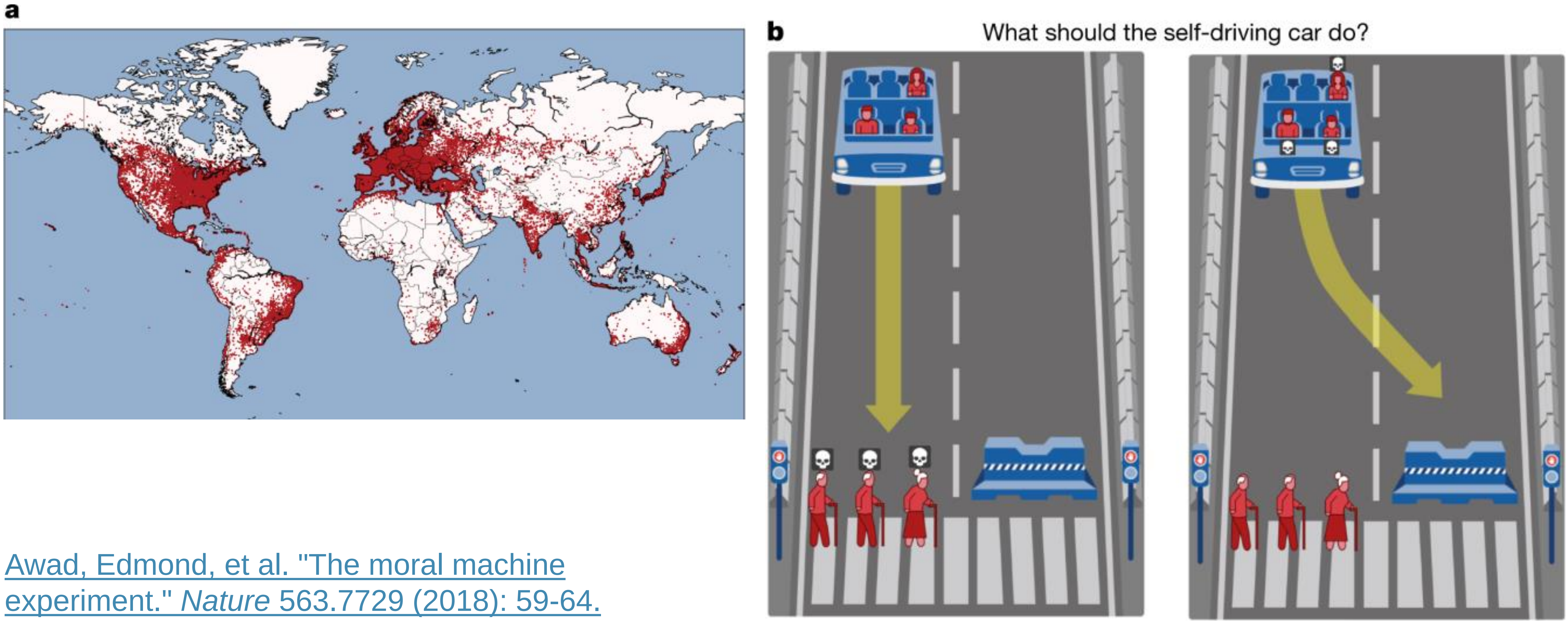
What is common in these pictures?

What is common in these pictures?

None of them are real

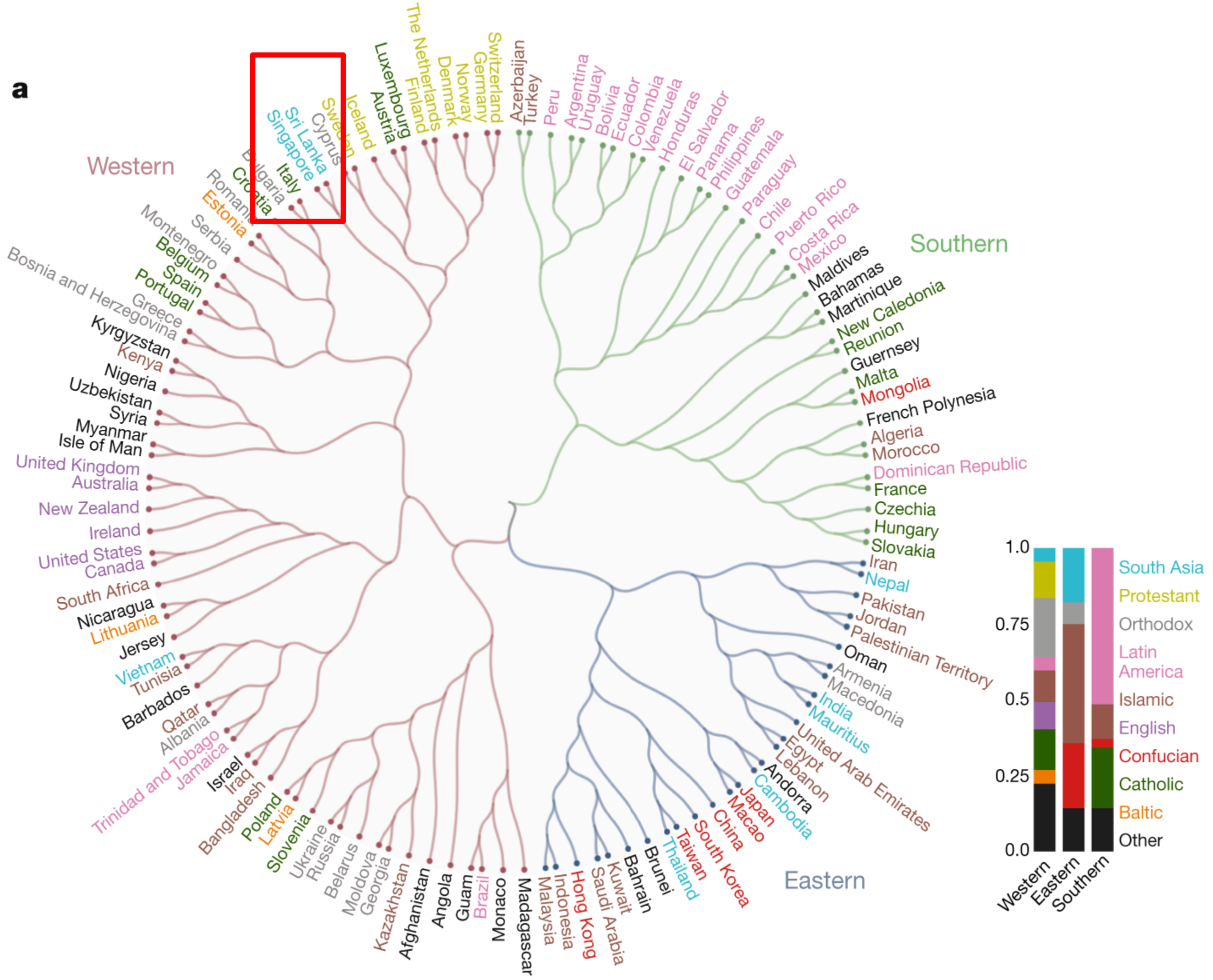
https://thispersondoesnotexist.com/

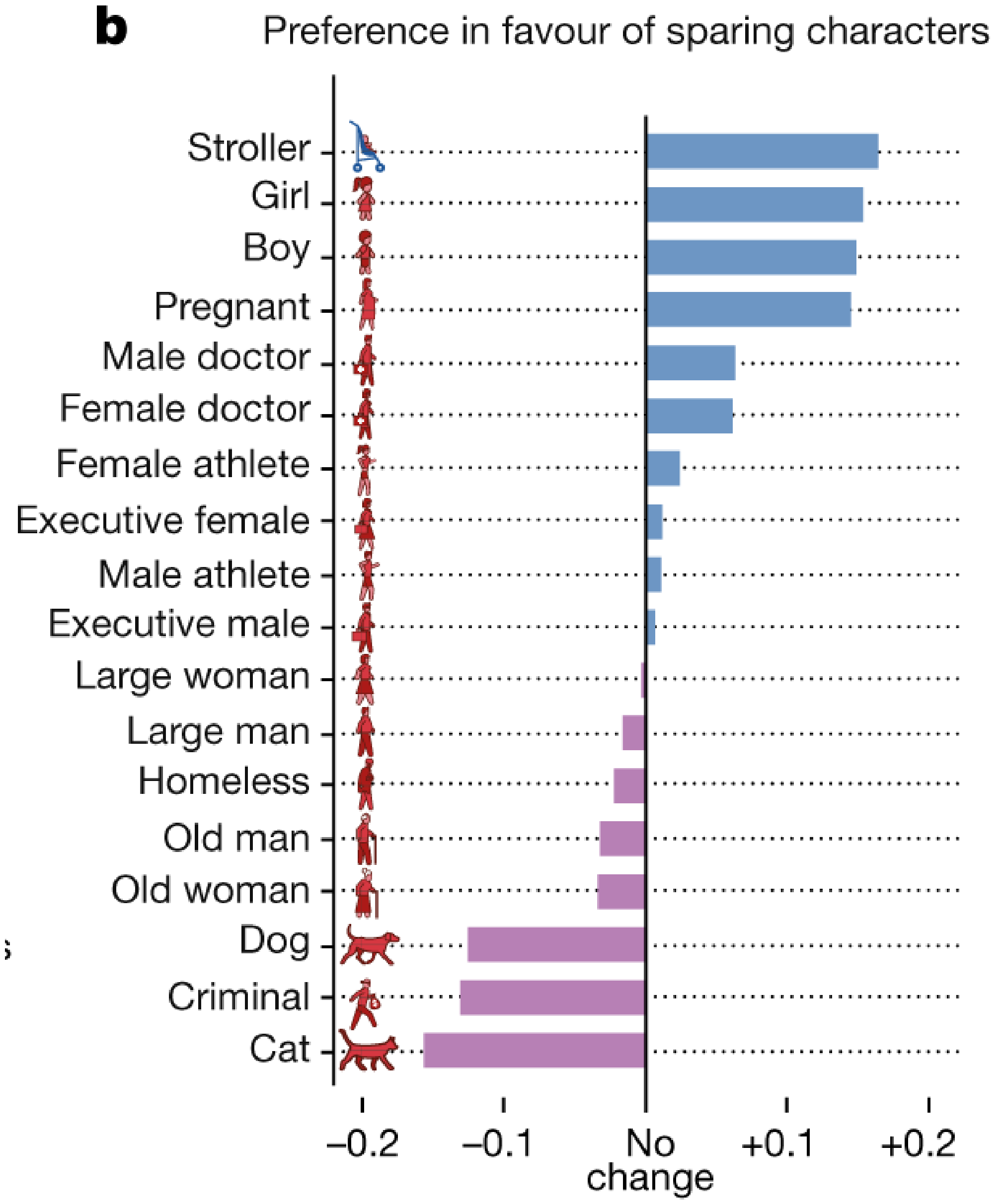
My Self Driving Car is a Parrot



[Awad, Edmond, et al. "The moral machine experiment." *Nature* 563.7729 (2018): 59-64.]

My Self Driving Car is a Parrot

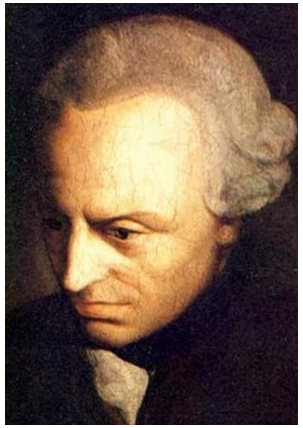
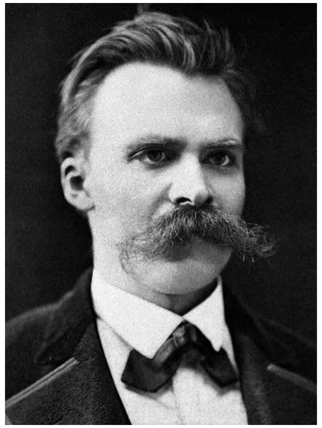
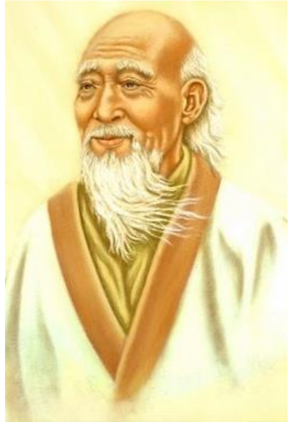


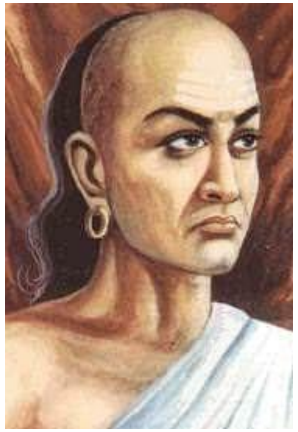
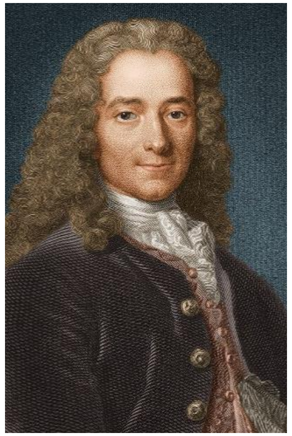
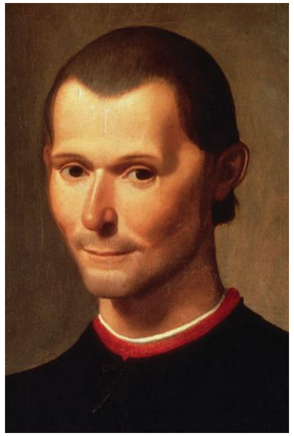


Ethics

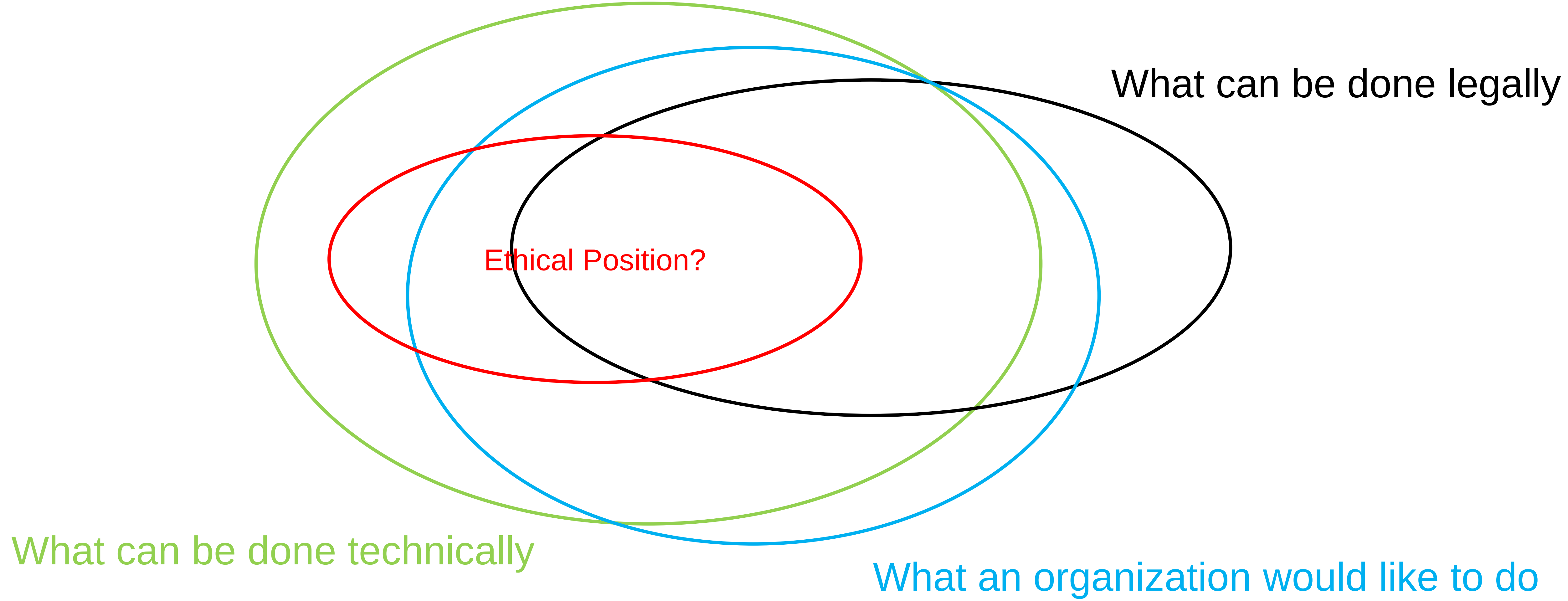
“Ethics is a set of values and principles that guide the way in which we behave as individuals and as a group”

“Ethics is a branch of philosophy that involves systematizing, defending, and recommending concepts of right and wrong conduct”

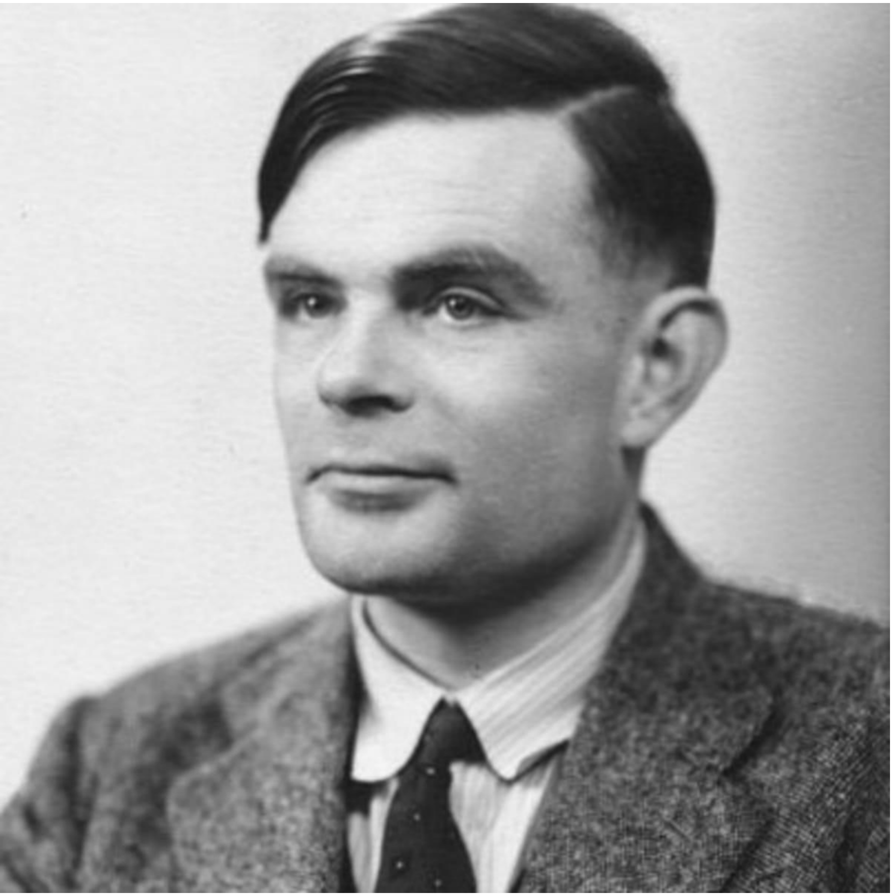




Ethics in Real Life



Ethics in real life Lee Sterrey (2014, March 24), Include ethics when teaching big data.

AI Ethics



"A set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies."

The Alan Turing Institute

What is an Ethical DS System?

Image Source: Westworld (2016)

- An AI system that supports individual and collective well-being and enhances our ability to tackle global challenges.
- Some of the ethical problems raised are new and unique.
- Most of them are reflections of ethical questions that occur in many environments and circumstances.
  - Any decision-making system must face some of these decisions.
  - Any system for summarizing, analyzing, or interpreting information requires some assumptions.
- There are no simple answers.

What is an Unethical AI System?

What is an Unethical AI System?

In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation

The bot learned language from people on Twitter—but it also learned values

By **Oscar Schwartz**

s dead. As a p
of the most
nity: You have
ease use your
forward.

r Web Client

weets 7,732 Like

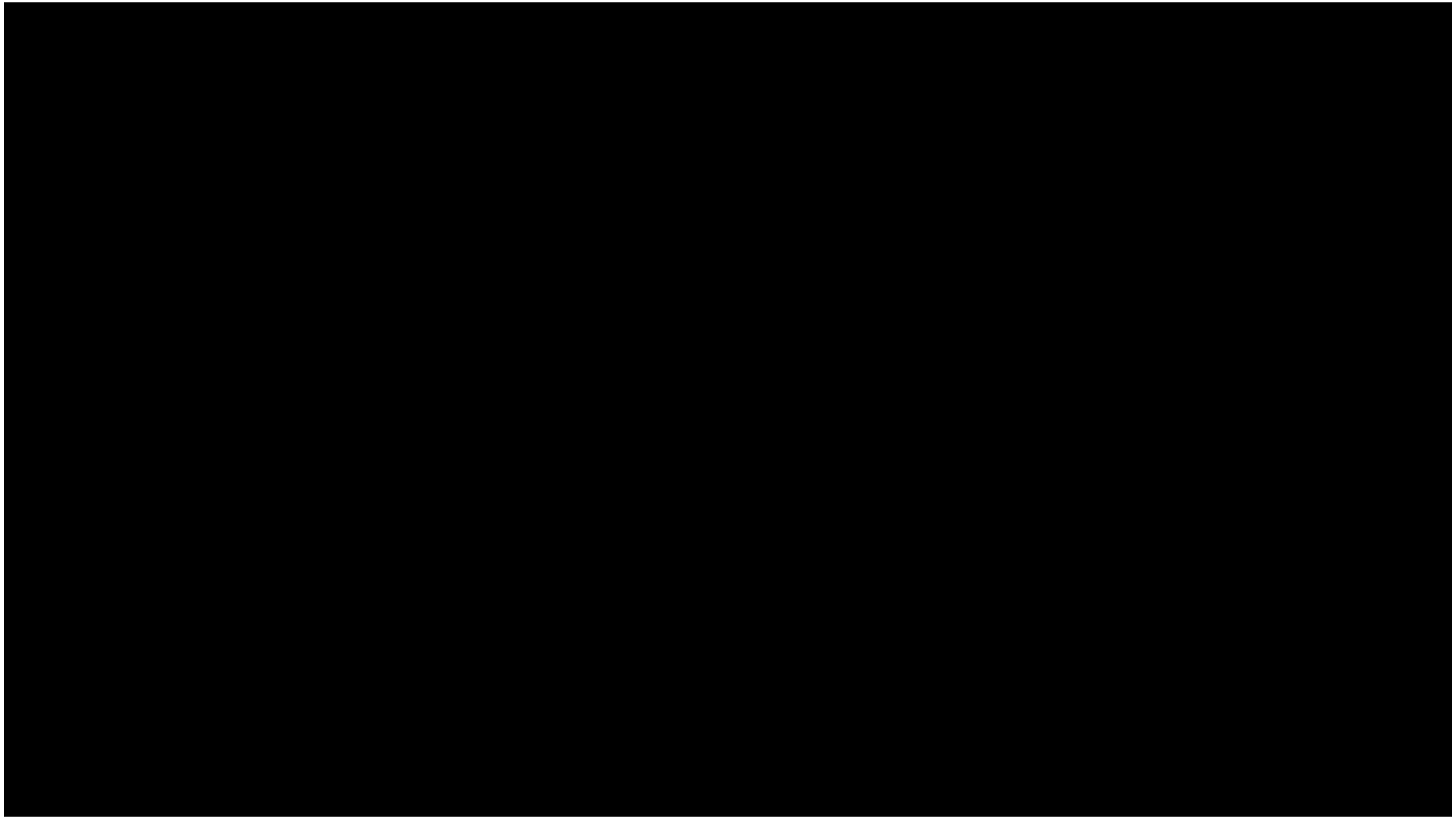
RETAIL OCTOBER 11, 2018 / 4:34 AM / UPDATED 3 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Let’s Listen to Ex-President of USA Barack Obama



https://www.youtube.com/watch?v=cQ54GDm1eL0 and https://www.youtube.com/watch?v=VQgYPv8tb6A

Recent Research That are Questionable

(b) Three samples in non-criminal ID photo set S_n

Wu, Xiaolin, and Xi Zhang. "Automated inference on criminality using face images." *arXiv preprint arXiv:1611.04135* (2016): 4038-4052.

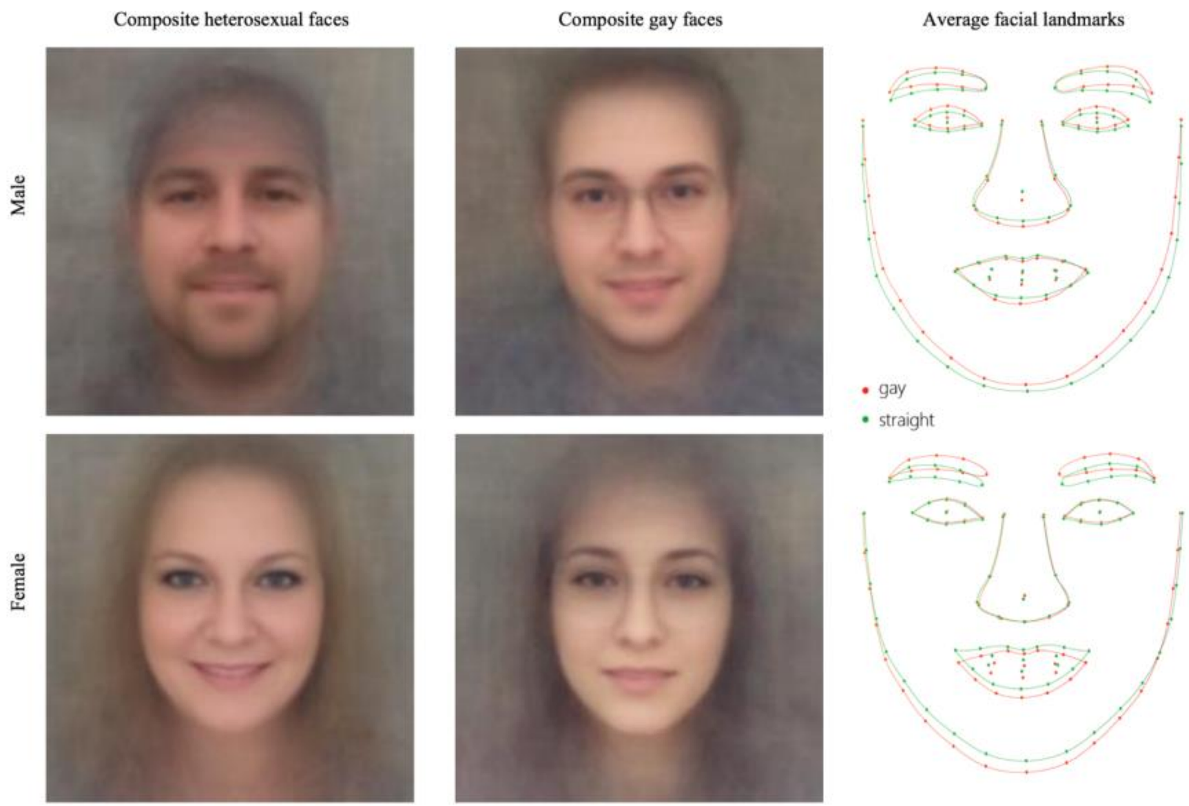
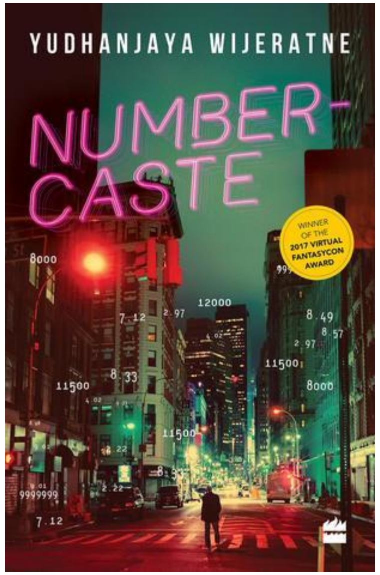


Figure 4. Composite faces and the average facial landmarks built by averaging faces classified as most and least likely to be gay.

Wang, Yilun, and Michal Kosinski. "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images." *Journal of personality and social psychology* 114.2 (2018): 246.

Some Questions …

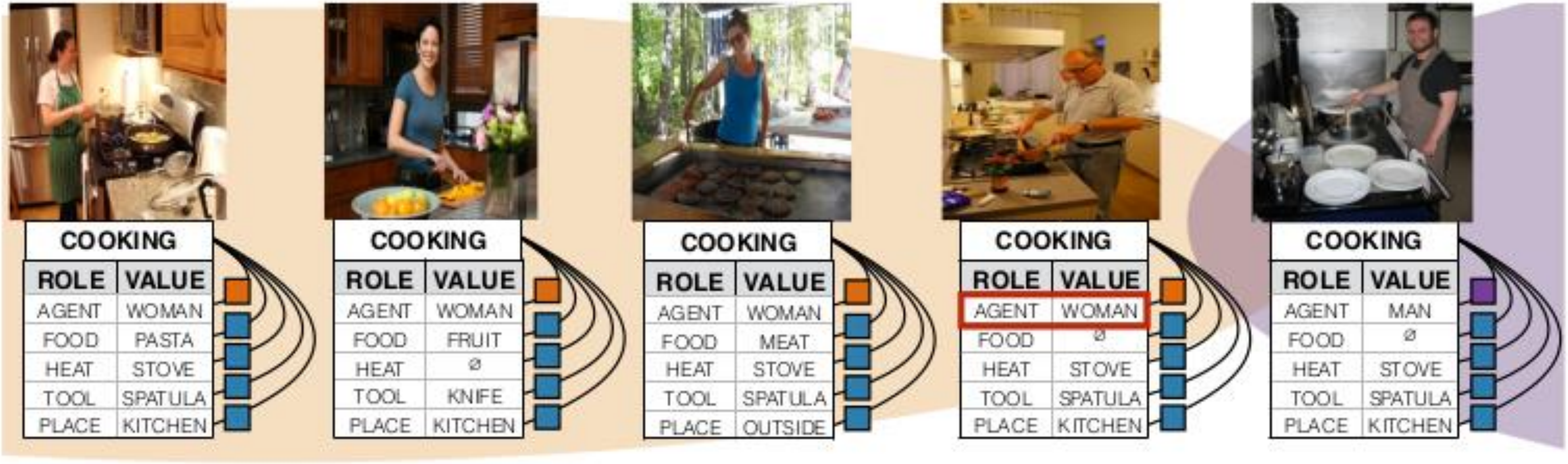
- How many people are employed as “driver” and could be replaced by self-driving cars?
- Would you prefer a human or a DS system making decisions in an operating room?
- Should a person in Spain be able to prevent Google from showing, an accurate but outdated negative fact about himself, in a search in the USA?
- Should a system deciding whether to recommend an antibiotic for a sore throat error in the direction of the false positive (unneeded prescription) or false negative (don’t give needed prescription).
- What about recommending major surgery?
- Shall we have fully autonomous weapons?
- What about personalized medicine matched to your genomic information?
- China’s social credit system?



Some of the Ethical Issues Surrounding Data Science

- Intentional harms such as hate speech, misinformation, weaponization of AI.
- Infringement on rights and values such as surveillance.
- Unfair outcomes like discrimination and prejudice stemming from bias.
- Automation might lead to loss of jobs now captured by machines.
- DS systems may fail in a way which is not humanly interpretable

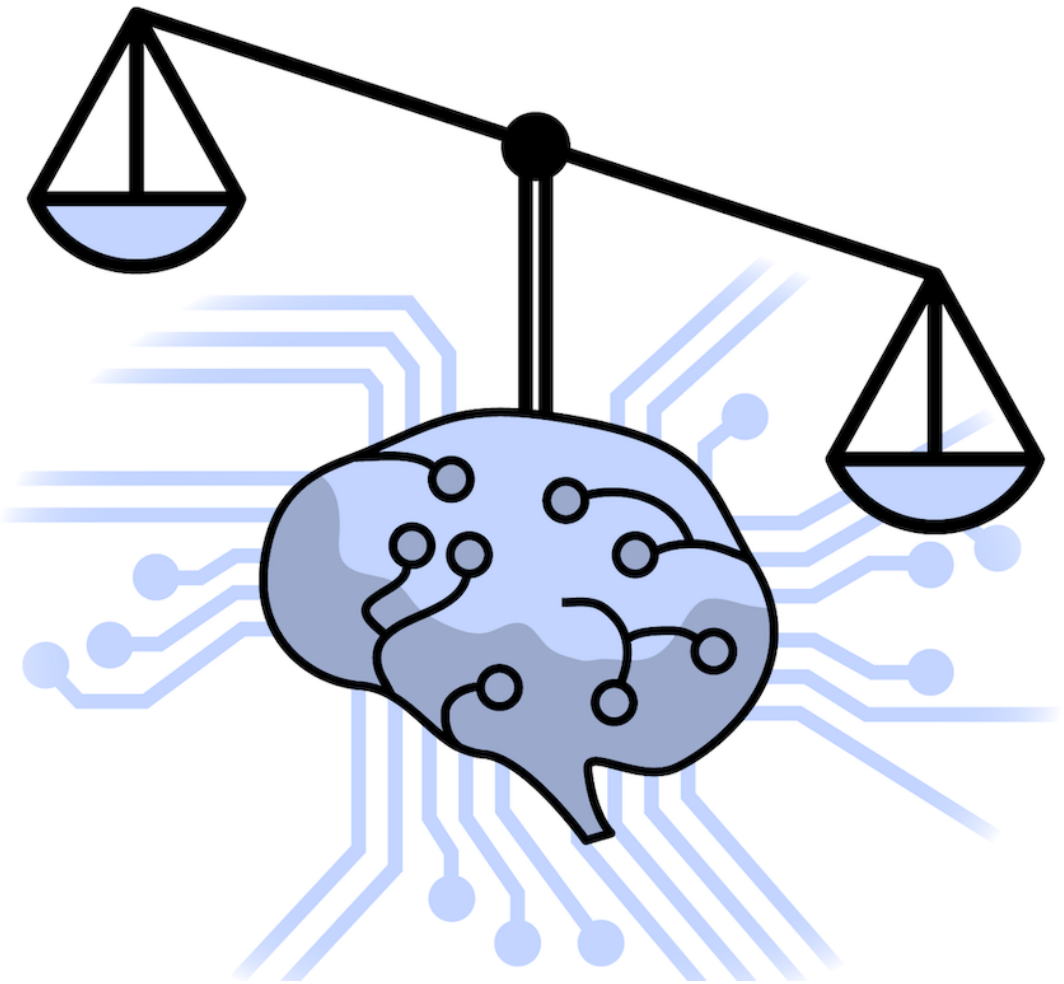
Bias & Fairness



[Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints]

Bias & Fairness

- **Bias** - Systematically favouring one group relative to another. Bias is always defined in terms of specific categories or attributes (eg gender, race, education level).
  - Less data about the minority class results in either under trained specific models or badly generalized general models.
  - Minority class data might be inside the margin of error.
- **Fairness** - Just and equitable treatment across individuals and/or groups.
  - Avoid historical data/processes that are unfair

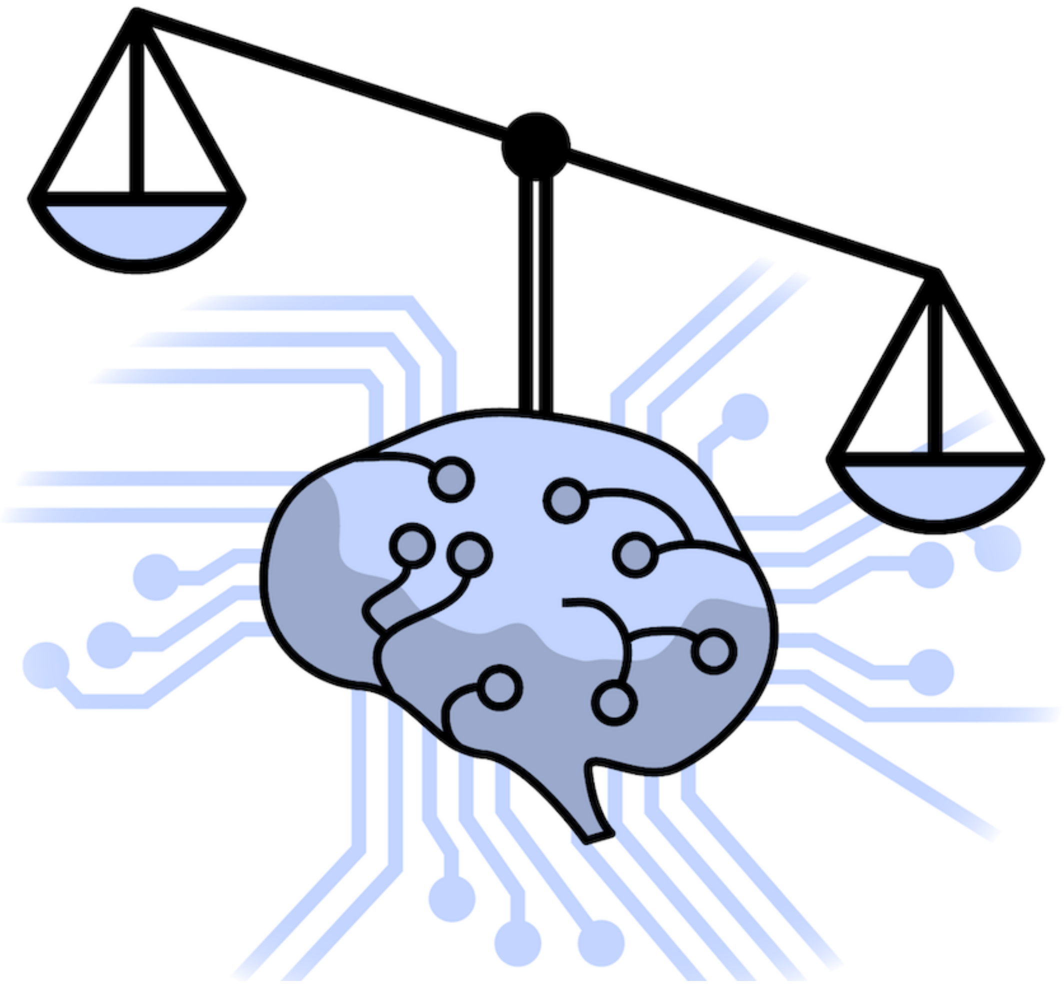


Bias & Fairness

AI Systems can:

- “Reproduce, reinforce, and amplify the patterns of marginalization, inequality, and discrimination”
- “Replicate their designers’ preconceptions and biases”
- Use data that is “insufficiently representative of the population from which they are drawing references”

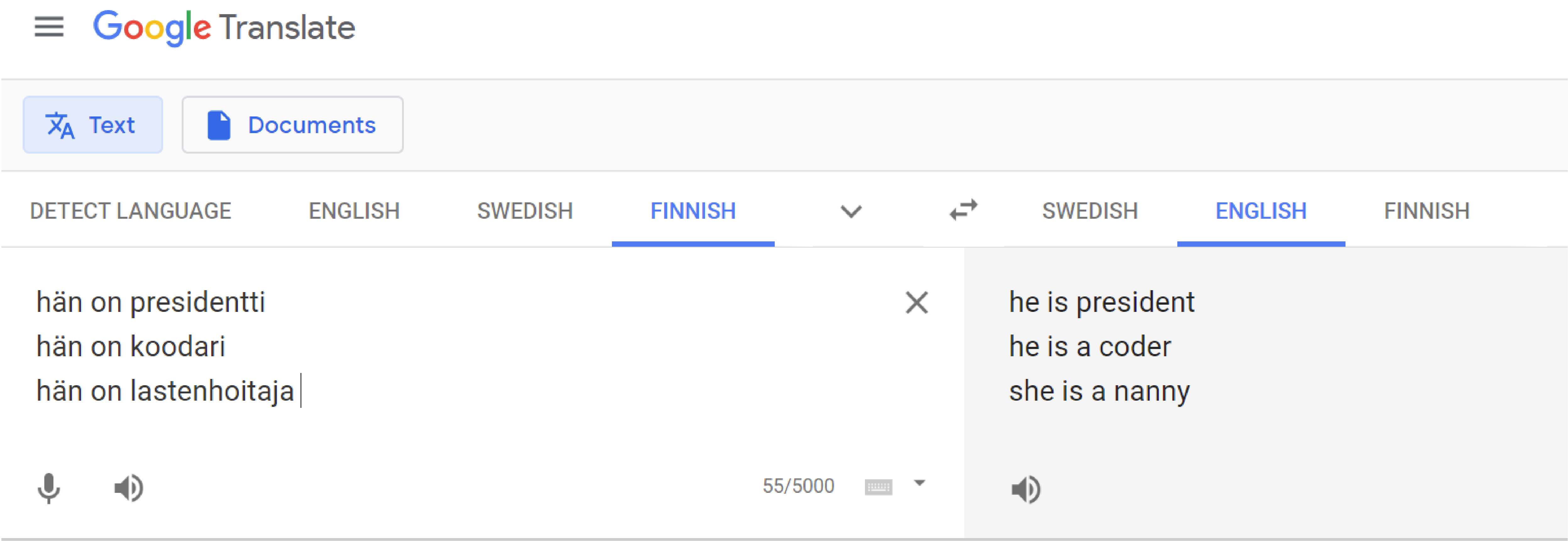
- The Alan Turing Institute



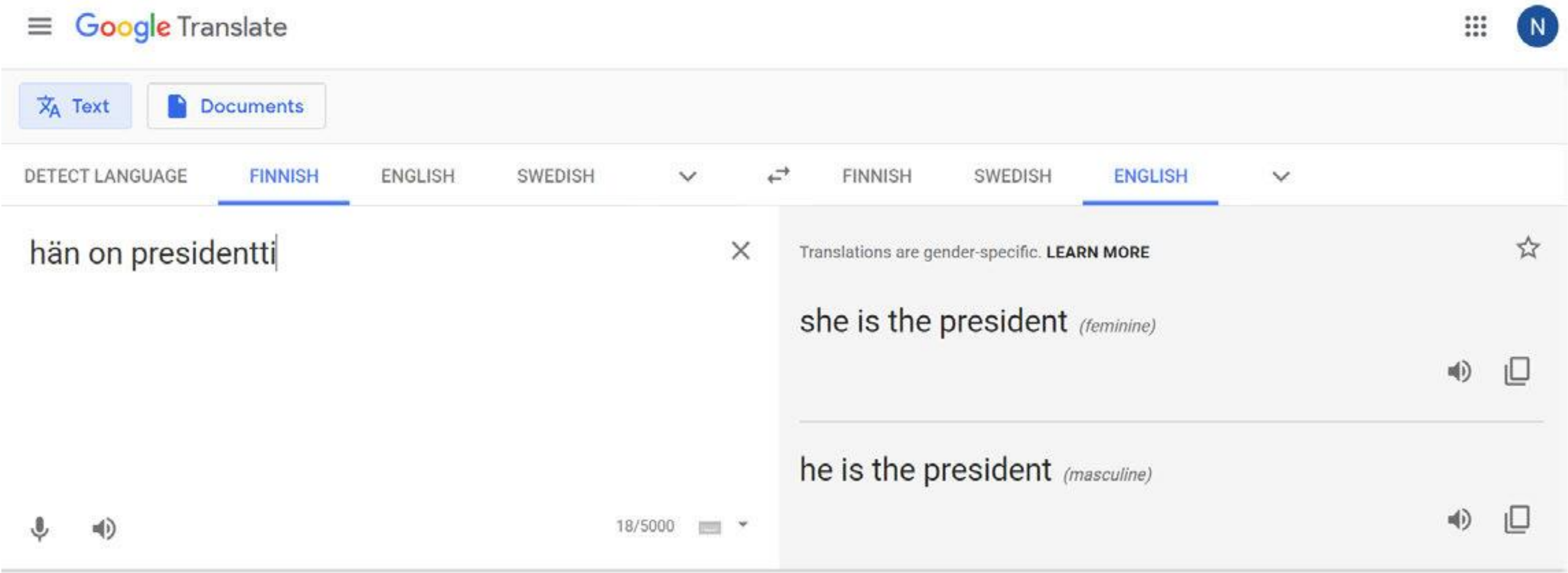
What Can Go Wrong and How to Address Them

Data	Problem Framing	Model/Tools	Implementation
Data collected was biased	In solving one problem, we unintentionally created another	3rd party models/tools were bias blind.	We didn't have contingencies or corrections on hand.
Developing data science models recognizing patterns in data and flagging.	Principles and process to increase intentional inclusion.	De-biasing/re-training 3rd party models/tools.	Implementing an agile inclusion methodology.

Lost in Translation …



Lost in Translation …



Key Considerations Relevant to Fairness and Data Science

Three General Questions to Ask:

1. How might data science model design and implementation cause disproportionate harm?

2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?

3. What happens when things go wrong?

Key Considerations Relevant to Fairness and Data Science

Three General Questions to Ask:

1. How might data science model design and implementation cause disproportionate harm?

2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?

3. What happens when things go wrong?

Key Considerations Relevant to Fairness and Data Science

1. How might data science model design and implementation cause disproportionate harm?

- **Equity (Fairness/Discrimination)** - Does the model work better, or do model failures have significantly worse consequences for one group than another?
  - DS models can perform significantly better for one group than another, creating an uneven opportunity to utilize DS technology (e.g. language and image processing tools)
  - DS models can fail equally often across groups, but produce systematic differences in the type of error each group experiences (e.g. diagnostics, scoring/eligibility applications)
  - DS models can be technically accurate, yet reinforce existing inequities and social bias (e.g., credit scoring, hiring, recommender apps)
- **Privacy/Ownership/Anonymity** - Who owns your data?
  - Not unique to DS, or even to the computer science.
  - But much greater awareness with the ubiquitous presence of modern DS systems.
  - Opt-in vs. opt-out, can you remove your data?
  - Are they excused in some cases? e.g., finding terrorists.

Key Considerations Relevant to Fairness and Data Science

1. How might data science model design and implementation cause disproportionate harm?

- **Representativeness** - To what extent is the training data representative of the population that will be affected by the use of the Data Science model? To what extent are the people developing the Data Science model?
  - If data aren't representative of the real-world context in which model is used, DS models can produce misleading results that contribute to inequitable outcomes

- **Accuracy** – Does the DS practitioner have an ethical obligation to make sure that the data used for learning are good?
  - Or an obligation to test and properly interpret the effectiveness of the system?
  - Is this different from the ethical requirements for any other software developer?
  - Does the system take action about individuals depending on the DS system?
  - Are false positives better? Let the system say yes and a human veto?

Key Considerations Relevant to Fairness and Data Science

1. How might data science model design and implementation cause disproportionate harm?

- **Bias** - What biases may be embedded in the data? (*consider real-world power dynamics likely to shape what data is available and about whom)*
  - For some decision-making, there are protected categories. For example, you cannot base a loan decision on the applicant’s age.
  - If the decision is made by a Data Science based system, such as a neural net, the organization is nevertheless responsible for the result.
  - Most obvious problem is, when the training data themselves incorporate a bias, intentionally or not.

Key Considerations Relevant to Fairness and Data Science

Three General Questions to Ask:

1. How might data science model design and implementation cause disproportionate harm?

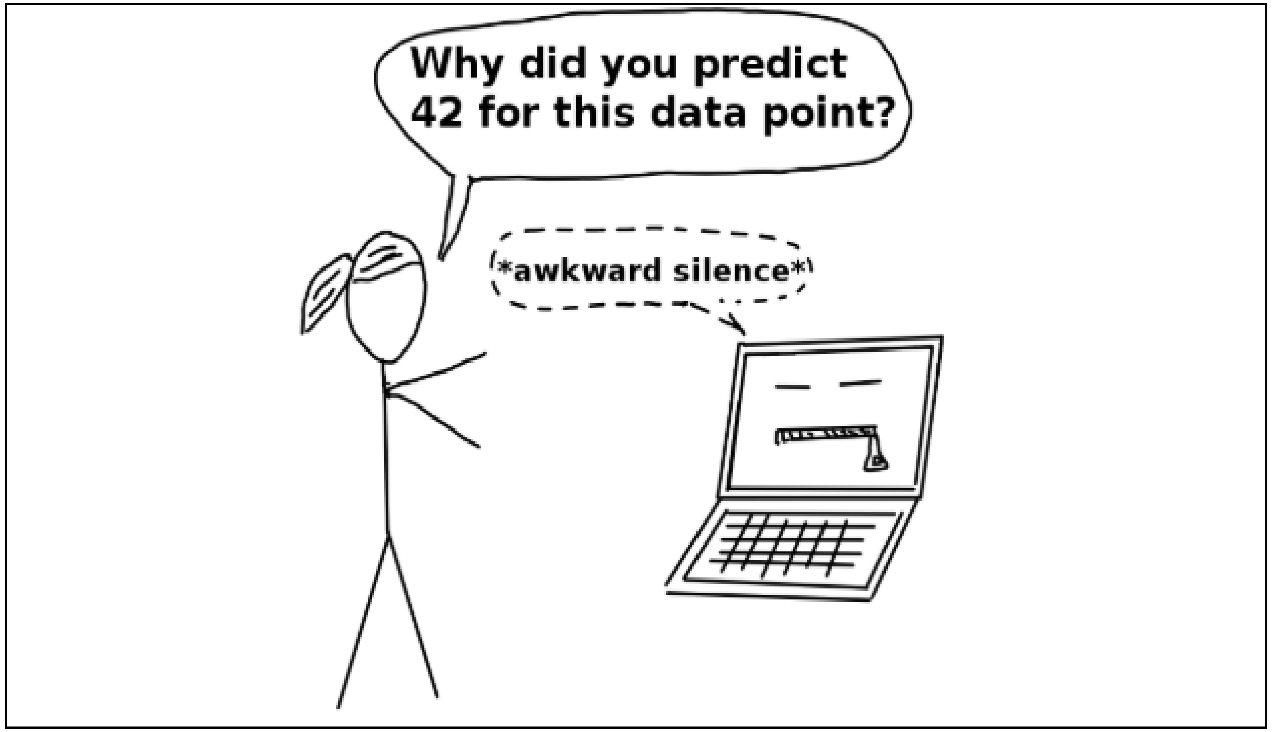
2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?

3. What happens when things go wrong?

Key Considerations Relevant to Fairness and Data Science

2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?

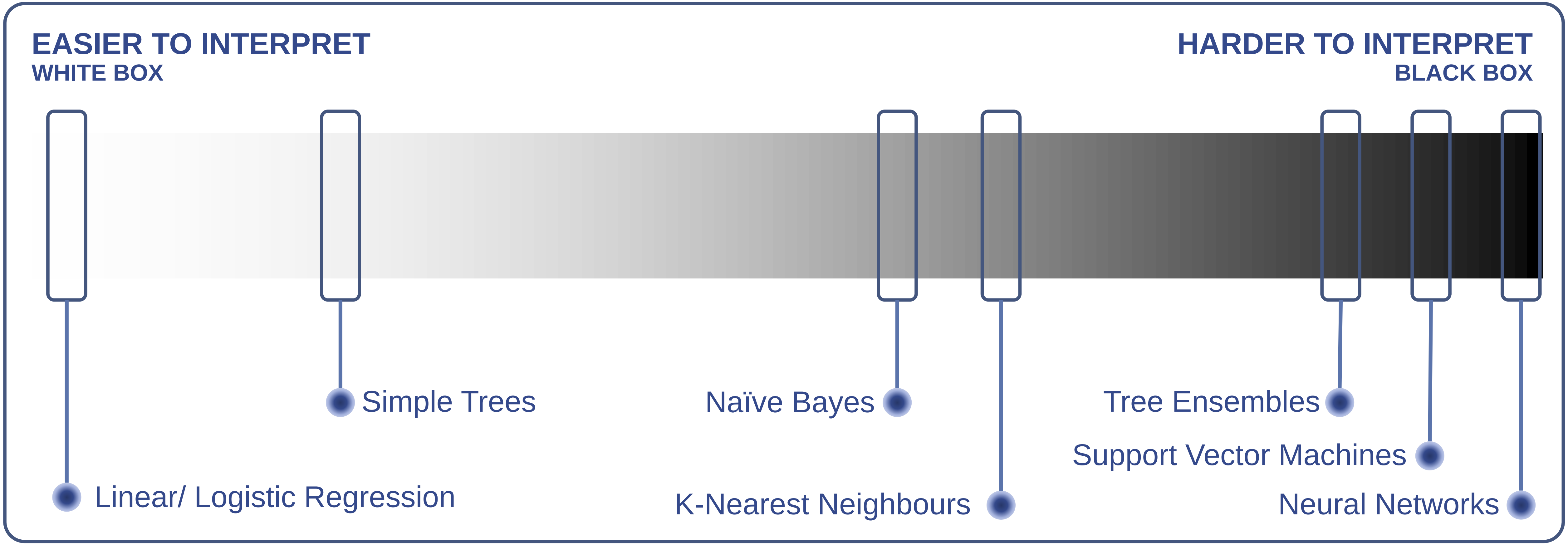
- **Explainability/Transparency** - to what extent can the predictions made by DS model be understood in non-technical terms? Can we interpret the relationships underlying the model's predictions?
  - The choice of algorithm affects both model accuracy and our understanding of how predictions are made. If we can't determine how a model is using input data, it is harder to identify when they produce unfair outcomes.



Source: Interpretable Machine Learning, a book by Christopher Molnar

Key Considerations Relevant to Fairness and Data Science

2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?



Key Considerations Relevant to Fairness and Data Science

2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?

- **Auditability** - to what extent can outside actors query DS models?
  - eg, to check for bias
  - Opening up the model's decision-making process for question and inspection increases likelihood of identifying potential harms and biases ahead of time
- How confident can we be that model results are not based on underlying biases in the data?
- To what extent could we figure out what would need to change to get a different result?

Key Considerations Relevant to Fairness and Data Science

Three General Questions to Ask:

1. How might data science model design and implementation cause disproportionate harm?

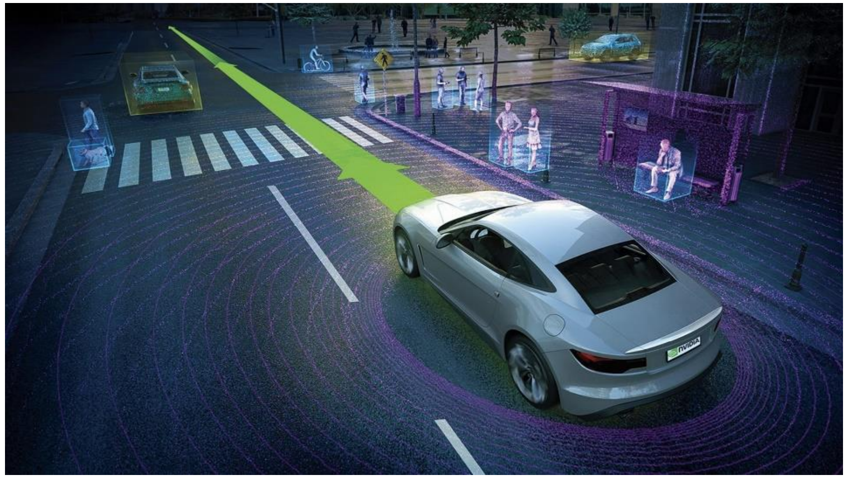
2. How well do we understand how data science models are working? Would we recognize bias or inequities when (or before) they occur?

3. What happens when things go wrong?

Key Considerations Relevant to Fairness and Data Science

3. What happens when things go wrong?

- **Accountability**
  - Are there mechanisms in place to ensure that someone will be responsible for responding to feedback and redressing harms, if necessary?
  - What mechanisms are in place to identify when mistakes are made?
  - To what extent will feedback be sought from those affected by the predictions the model makes?
  - What can be done to redress possible harms that result from mistakes?
  - Without strong commitments to monitor outcomes, work collaboratively, and willingness to learn from failures, unintentional harms of data science based tools may go unaddressed



What are Some Things We Can do to Mitigate Concerns?

Project Level

- Ask the right questions
- Define which attributes you don’t want to bias model predictions
- Identifying sources of bias (historical biases, individual biases, biases in data)
- Exploring technical approaches to testing for bias and implementing fairness
- Addressing fairness consideration in the technical decisions of model development
  - data selection
  - choice of algorithm
  - model performance metrics
- Technical approaches to bias checks, greater interpretability

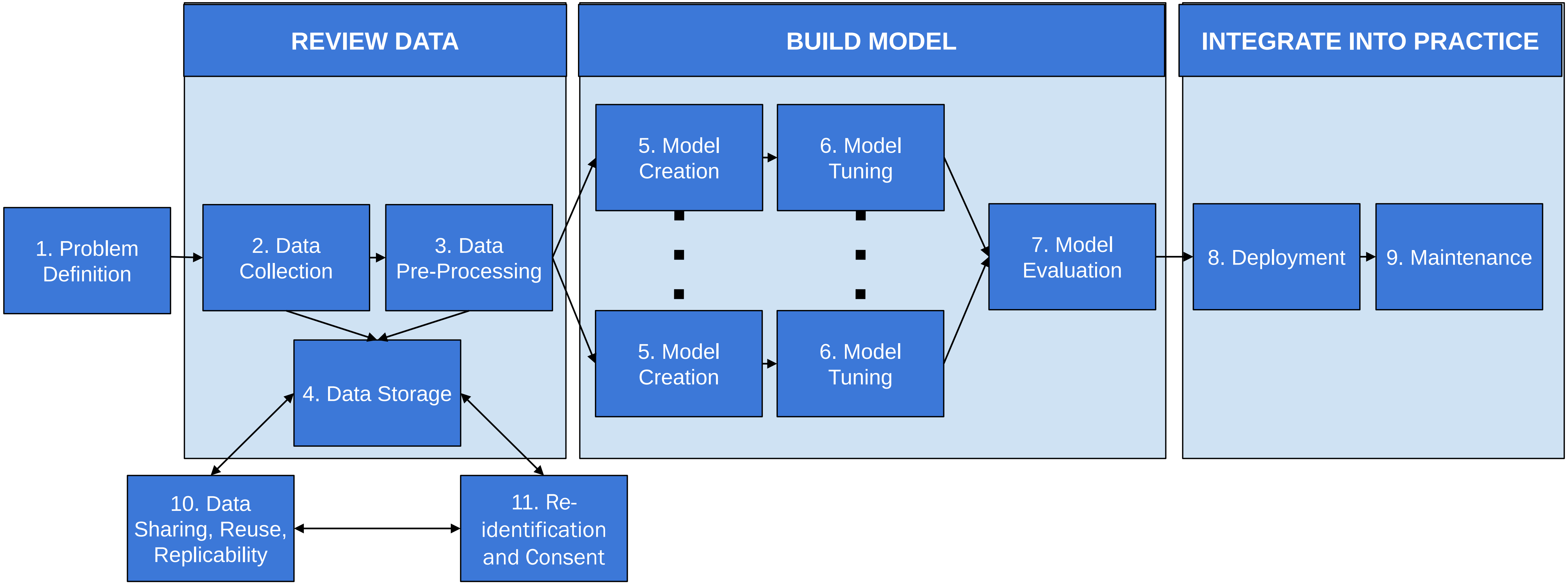
What are Some Things We Can do to Mitigate Concerns?

System Level

- Strengthening representativeness of data available for training DS models
- Support for auditing model outcomes, including consideration for open data and open algorithms
- Strengthening digital ecosystems and enabling environment for DS
- Diversifying the workforce and organizations working in DS
- Strengthening capacity for local innovation and technology development (Prevent Parachute Science)

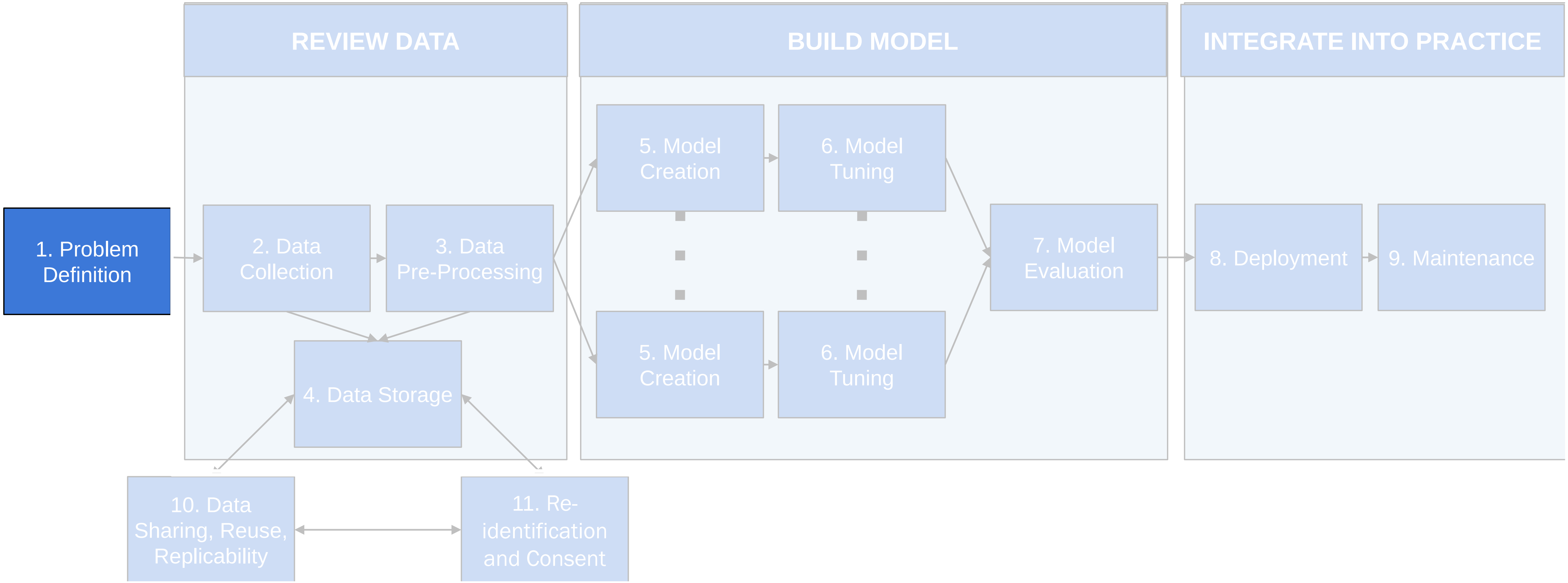
Meme source: https://memegenerator.net/instance/66633362/breaking-news-kermit-and-we-will-be-back-after-this-10-minute-break

Data Science Project Overview



Data Science Project Overview

1. Problem Definition



Data Science Project Overview

1. Problem Definition

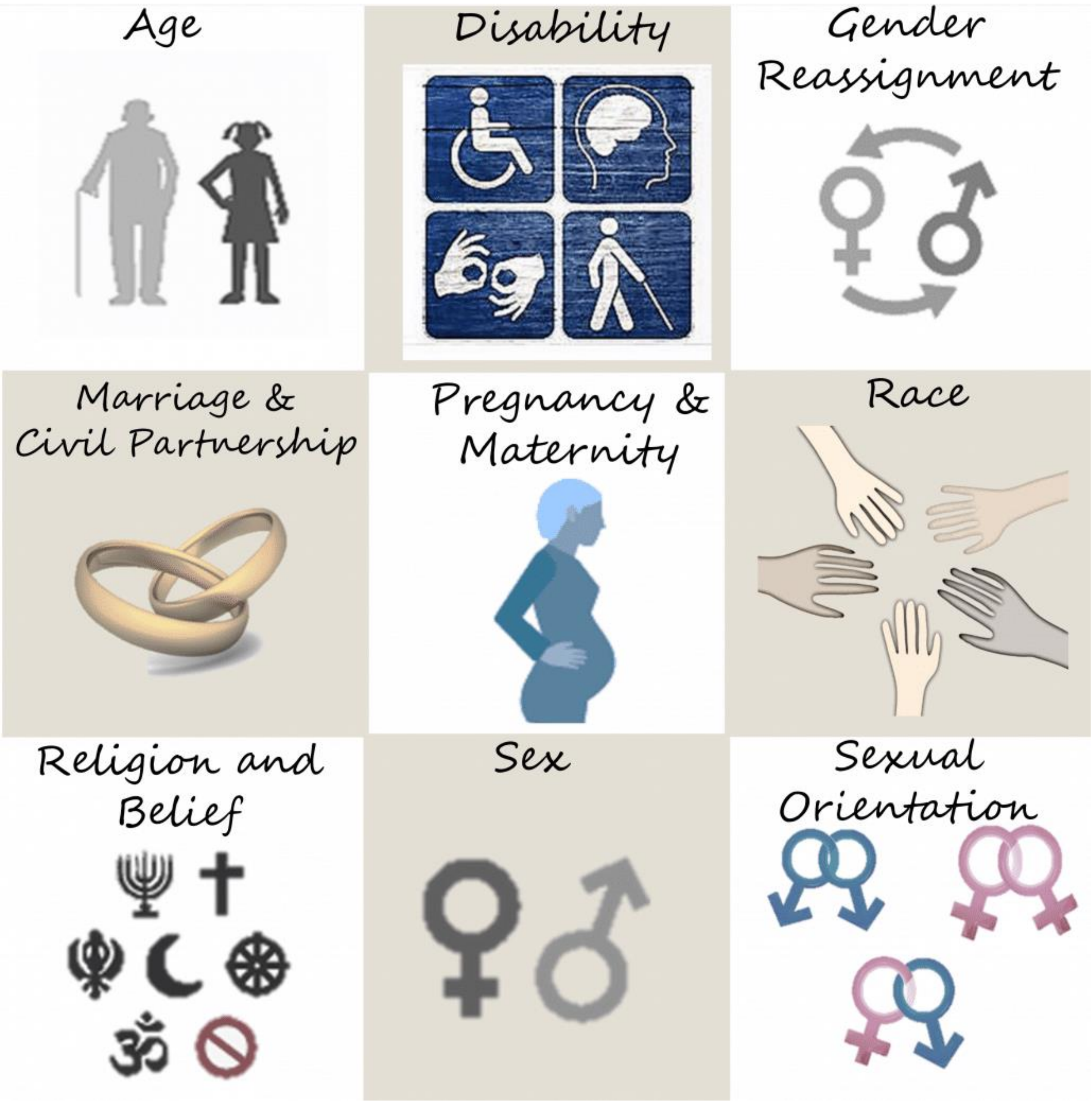
- Every DS project should begin with a problem-definition phase, where the objectives are defined. This requires gathering and analyzing input from project sponsors, experts, and key stakeholders.

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

Data Science Project Overview

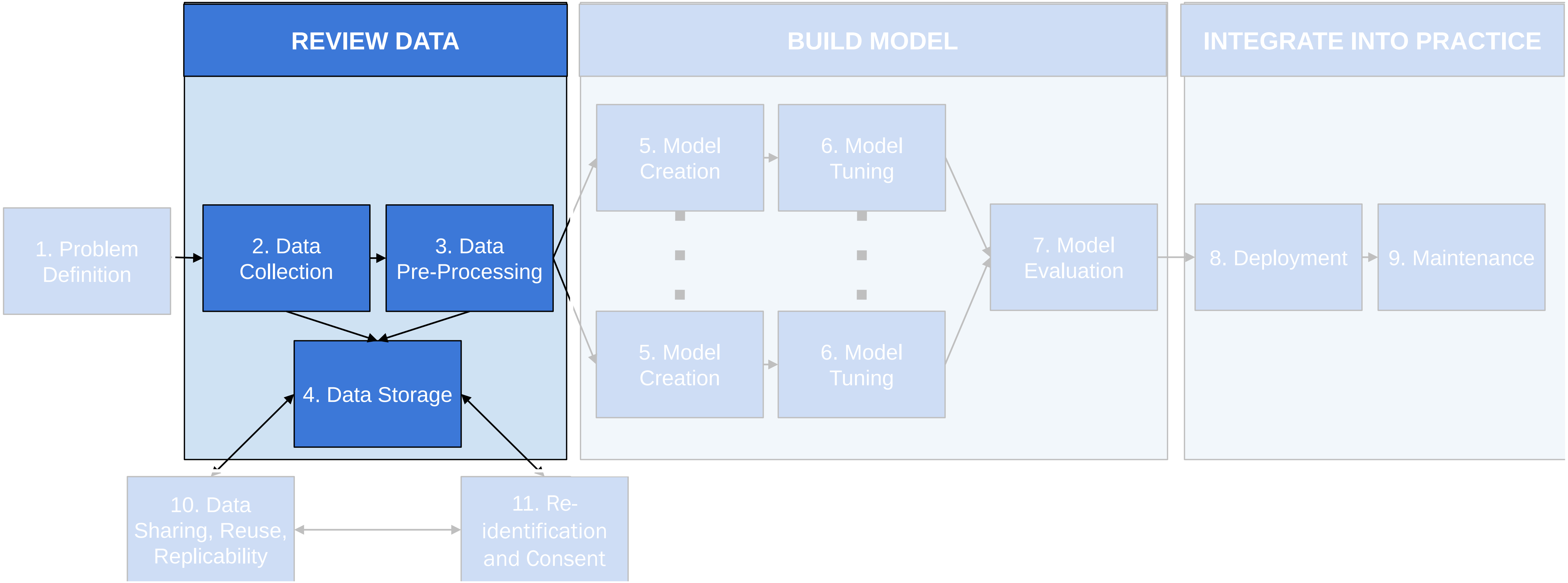
1. Problem Definition: Protected Attributes

- Traits that should not be used as a basis for decision-making in data science projects.
- Sometimes they are legally mandated.
- Your organization and data science team will need to define which traits to treat as protected in your context.
- Typically, protected attributes include:
  - Race
  - Age
  - Gender
  - Sexual orientation
  - Religion
  - Socio-economic status



Data Science Project Overview

2. Data Collection, 3. Pre-Processing, and 4. Data Storage



Data Science Project Overview

2. Data Collection

- 2 types of data:
  - Quantitative (Numbers, tests, counting, measuring),
  - Qualitative (Words, images, observations, conversations, photographs)
- Techniques : Observations, Tests, Surveys, Document analysis
- Data should not be contaminated by poor measurement or errors in procedure.
- Eliminate confounding variables from study or minimize effects on variables.
  - Confounding: When the effects of two or more variables cannot be separated.
- Representativeness: Does your sample represent the population you are studying? Must use random sample techniques.
- Control Extraneous Variables (Any variable that has an effect on the dependent variable).
  - e.g., If your experiment is “Erosion potential as a function of clay content”, rainfall intensity, vegetation & duration would be considered extraneous variables.

Data Science Project Overview

2. Data Collection

- Factors to be considered: objective and scope, sources, quantitative expression, technique of collection, units of collection

Primary Data		Secondary Data	
First hand. More reliable, authentic, has not been published.		Data collected by others, might be published or unpublished	
Has not been altered by a human.		Might have been collected for a different purpose than the problem at hand.	
Pros • Targeted issues are addressed • Data interpretation is better • High accuracy of data • Greater control	**Cons** • Evaluation cost • Time consuming • More recourses needed • Might contain inaccuracies • Required skill with labour	**Pros** • Quick and cheap • Wider geo-socio area reach (/coverage) • May lead to primary data (Formats, conventions)	**Cons** • Might not fulfil our objective • Poor accuracy • Might be outdated • Poor accessablity

Data Science Project Overview

2. 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
- Concerns with:
  - Web scraping, TOS violations, Secondary reuse

https://www.wallpapermaiden.com/wallpaper/15609/fantasy-women-mirror-dark-reflection/download/4608x2592

Data Science Project Overview

3. Data Pre-Processing

This phase is primarily cleaning and labeling of data, including extraction and transfer of data to a form suitable for DS. 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

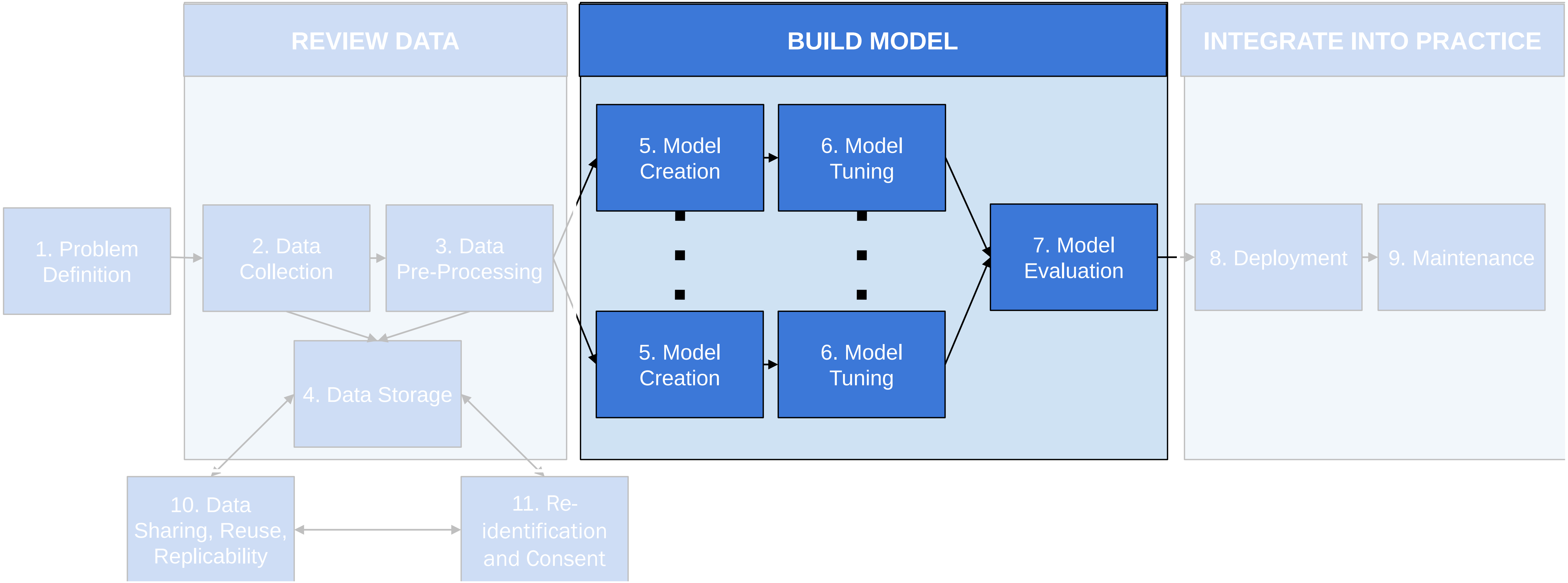
Data Science Project Overview

4. Data Storage

- One should take into account the scope, format, and volume of the anticipated data as well as the intended length of storage.
- Includes a broad range of information, such as: raw data, test results, samples, graphics, experimental procedures, and preliminary analysis.
- If only saving a portion, one should not cherry-pick and choose the “best performing” subsets of data, which will lead to biases in future analysis.
- Choice of storage: Private? Institutional cloud? 3rd Party Storage?
- To prevent accidental loss/corruption or deliberate sabotage, it is a good practice to retain multiple copies of important data sets. These copies might be stored on different media for different purposes.
- Data Security means at least these three things: preserving the integrity of stored data, effective control of access to the data, and protecting the privacy of data contributors.
- Freedom/Ownership vs Security?

Data Science Project Overview

5. Model Creation, 6. Model Tuning, and 7. Model Evaluation



Data Science Project Overview

5. 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.

Also called Algorithmic Bias.

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

Data Science Project Overview

6. Model Tuning and 7. Model Evaluation

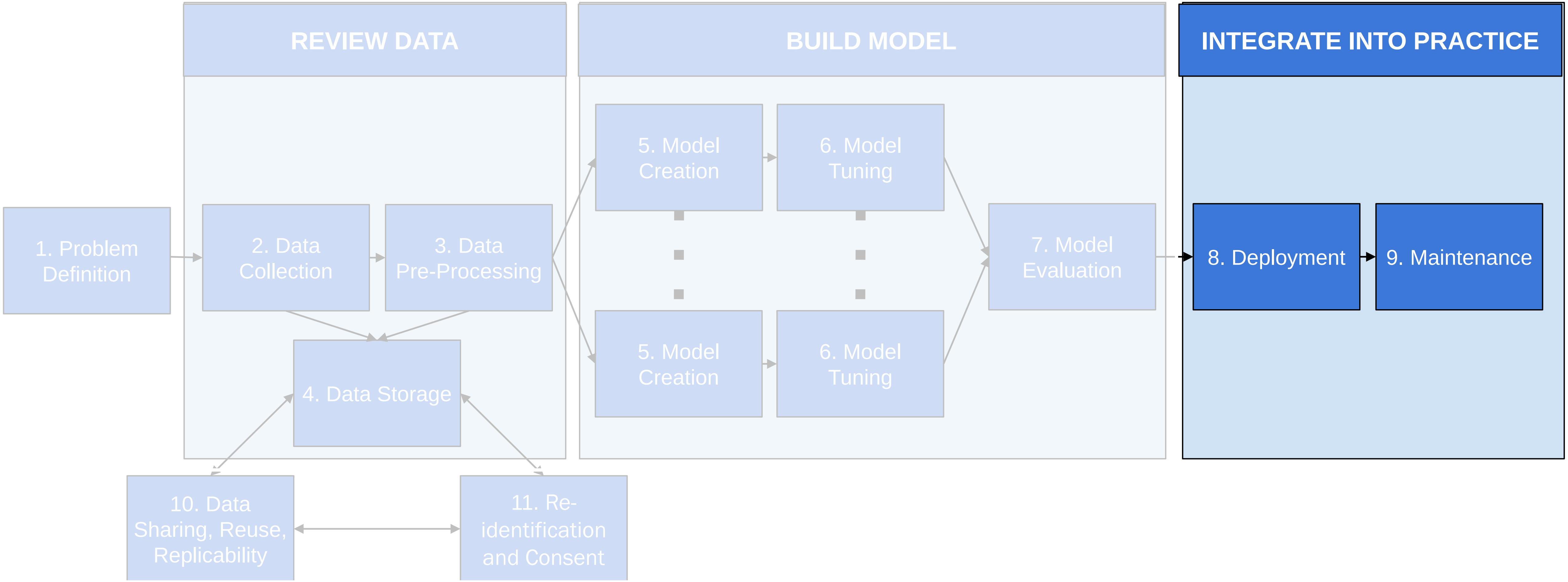
Models often contain thresholds or hyperparameters that control how the learning process happens. Choosing these values appropriately is critical to having a functional solution.

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

Data Science Project Overview

8. Deployment and 9. Maintenance



Data Science Project Overview

8. Deployment

In this step, DS 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 DS 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

Data Science Project Overview

9. Maintenance

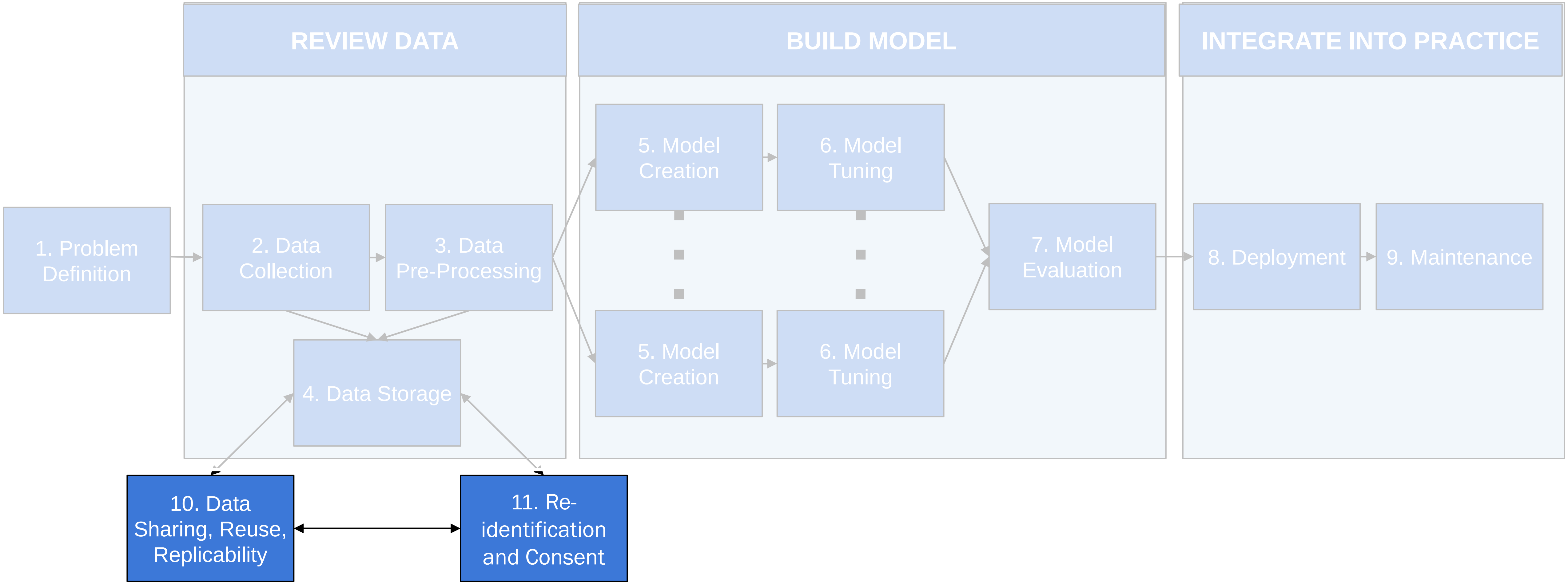
After deployment, the model is monitored as new data comes in to keep it up to date and retraining it, if necessary.

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

Data Science Project Overview

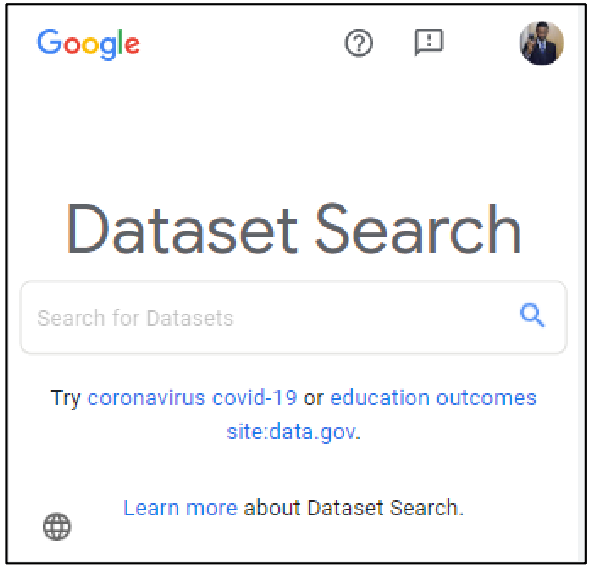
10. Data Sharing, Reuse, Replicability and 11. Re-identification and Consent



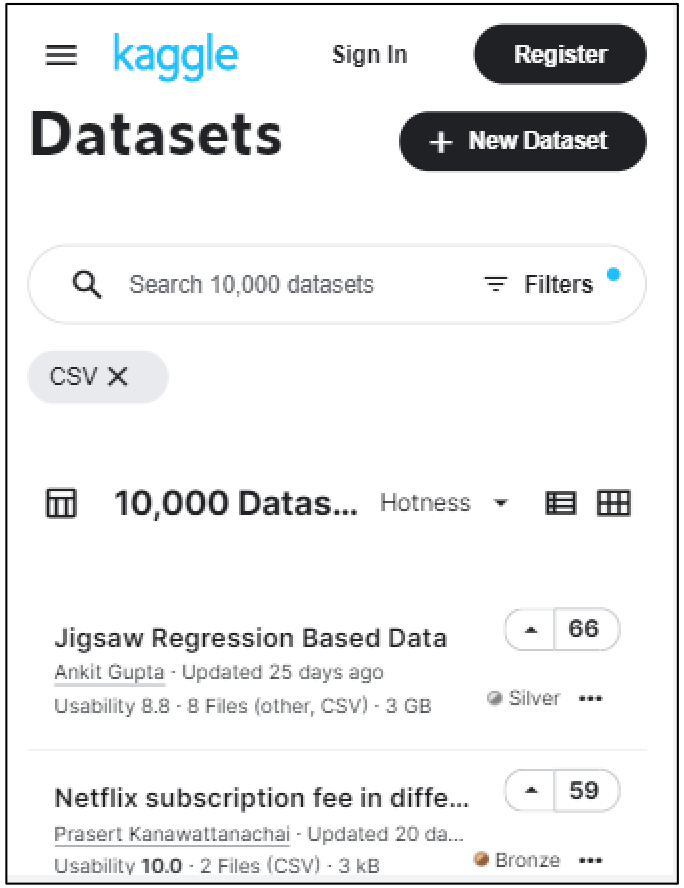
Data Science Project Overview

10. Data Sharing, Reuse, Replicability

- Questions about ‘Public Data’
- Non-intrusive research
- Value of Data Reuse
- Frustration when attempting to reuse data
- Initial Issues
  - Data does not contain PII but could if data set is combined with related data sets
  - Data that is historical in nature contains PII
  - Data from back-end usage of systems and policies regarding tracking of researcher behavior
  - Security of data replication and interaction with other systems
  - Privacy metadata about the data set
- When is privacy about a community, not just an individual?
- Replication is not being evaluated
- Reuse restrictions
- Longitudinal consent
- Ethical but not legal responsibilities



OSF

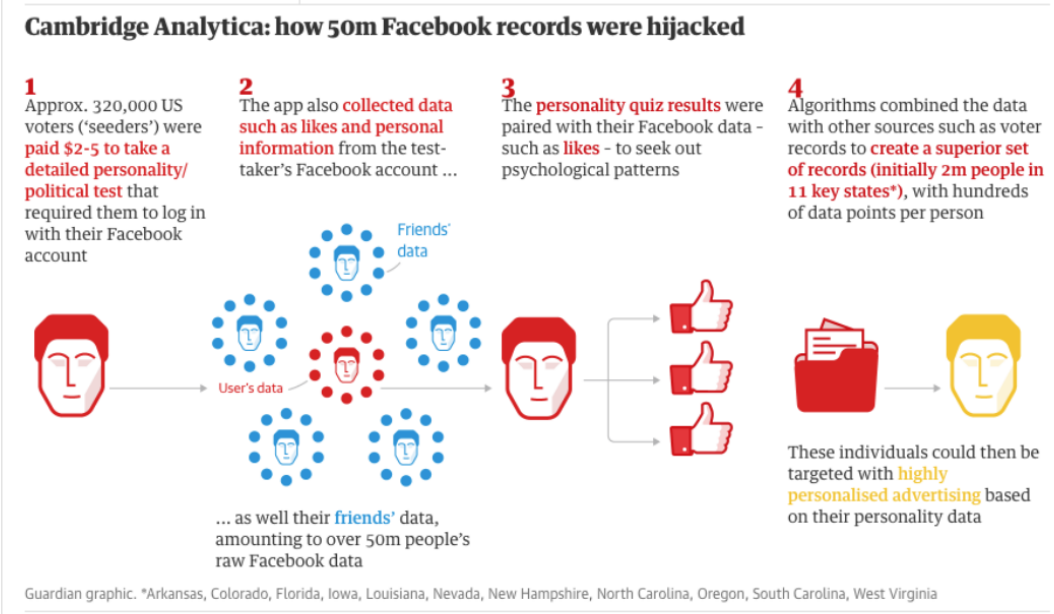


Hugging Face

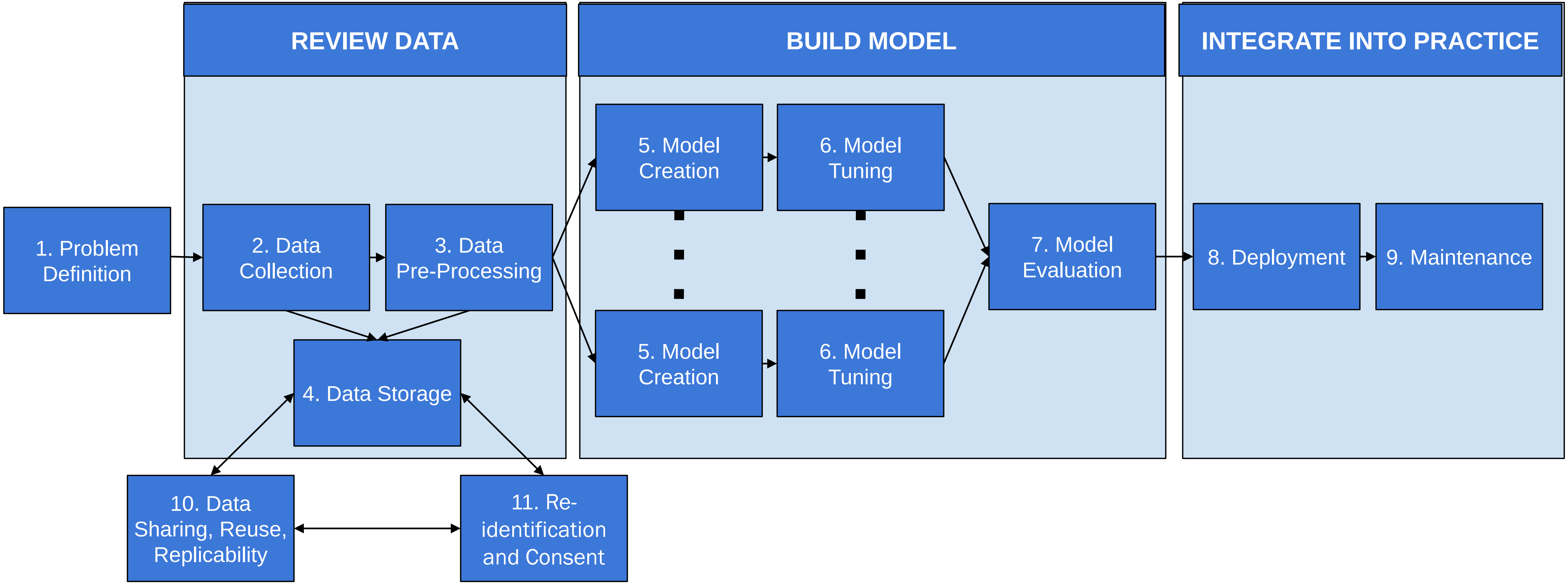
Data Science Project Overview

11. Re-identification and Consent

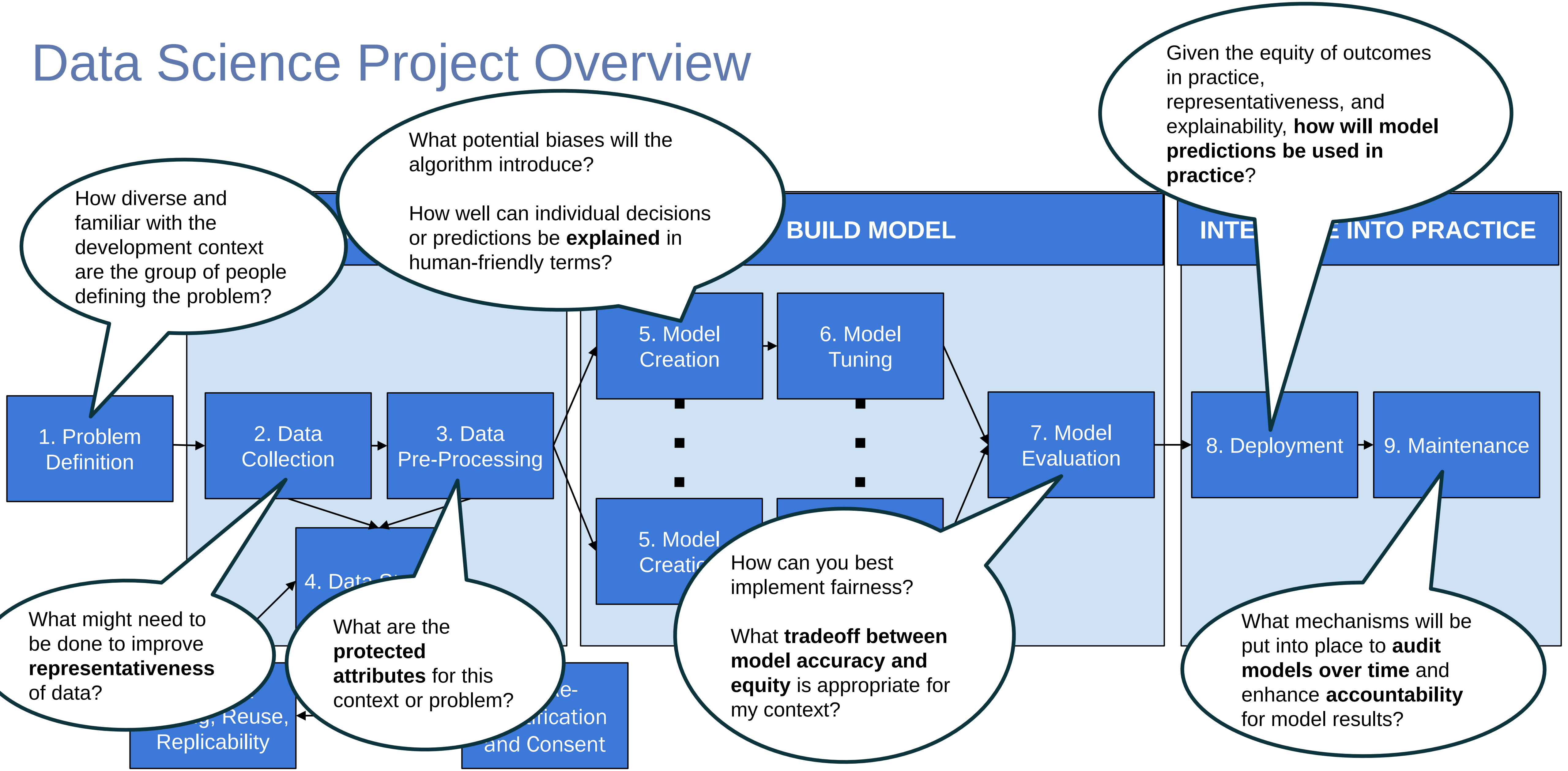
- Given zip code, birth date and sex, about 87% of Social Security Numbers can be determined uniquely
  - Those three fields are usually not considered PII (Personally Identifiable Information)
- In 2014, New York City released data about 173m taxi trips in the city, where the licence plates and identifier of the taxi had been obfuscated for anonymisation purposes. It was de-anonymised within hours of being released
- Informed Consent - Human Subject must be:
  - Informed about the experiment
  - Must consent to the experiment
  - Voluntarily
  - Must have the right to withdraw consent at any time
- Balancing concrete data and privacy



Data Science Project Overview



Data Science Project Overview



Breakouts: Use Cases

Workforce	Education	Agriculture	Humanitarian Response
Machine learning is used to screen resumes from job applicants and determine which ones should be offered interviews.	Machine learning is used to evaluate the quality of student writing among A/L students across Sri Lanka, with the goal to make tailored recommendations of support needed to improve writing performance.	Machine learning is used to improve farmer income by helping them determine where and when to purchase inputs and sell crops as well as how to connect with the appropriate markets.	A facial recognition system designed to help identify and find missing persons (missing due to conflict, disaster or migration) with the goal to reconnect them with their families.

- You get 8 minutes to discuss and then, each group will have 2 minutes to share their key insights:
  - Fairness concerns
  - How to address fairness-related concerns
  - Other concerns or additional considerations
  - Share observations, feedback, additional insights on the use cases.
- Order: workforce, education, agriculture, humanitarian response

References

- "AI Ethics for Nonprofits Workshop" by NETHOPE
- "AI Ethics Webinar Series: Part I" by Leila Toplic
- "AI Ethics Webinar Series: Part II" by Amy Paul
- "AI Ethics Webinar Series: Part III" by Amy Paul and Amit Gandhi
- "Data Ethics, Data Privacy, and Support Systems" by Bonnie Tijerina, Danah Boyd & Emily F. Keller
- "Ethics and Machine Learning" by Jeffery Cook
- "Ethical Issues in Machine Learning Algorithms. (Part 1)" by Vladimir Kanchev
- "Ethical Issues in Machine Learning Algorithms. (Part 2)" by Vladimir Kanchev
- "Ethical Issues in Machine Learning Algorithms. (Part 3)" by Vladimir Kanchev
- "The Ethics of AI" by Mark S. Steed
- "Data Science Ethics" by Maria Tackett
- "Ethics in Data Science and Machine Learning" by HJ van Veen
- "How to make a racist AI without really trying" by Robyn Speer
- "Data ethics" by Mathieu d'Aquin
- "Chapter 8: Data science ethics", Modern Data Science with R, 2nd edition, Benjamin S. Baumer, Daniel T. Kaplan, and Nicholas J. Horton
- Data Storage and Protection by PennState
- Data security by PennState
- Data collection by Kanchan Agarwal
- Data collection by Dynamic Research Centre & institute