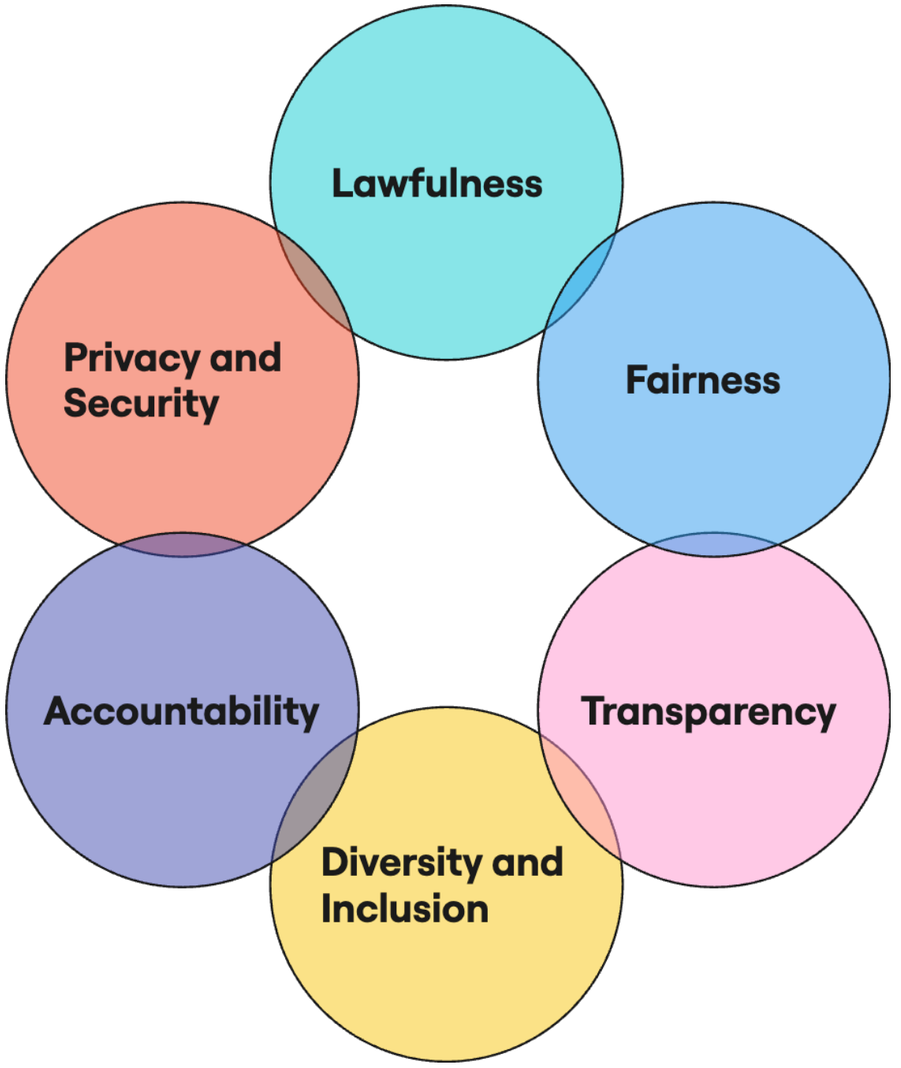
Responsible data dimensions

RESPONSIBLE AI DATA MANAGEMENT

Maria Prokofieva
Lead ML Engineer

Responsible data management

- Ethical data management
- Evaluate models with technical metrics
- Responsible AI



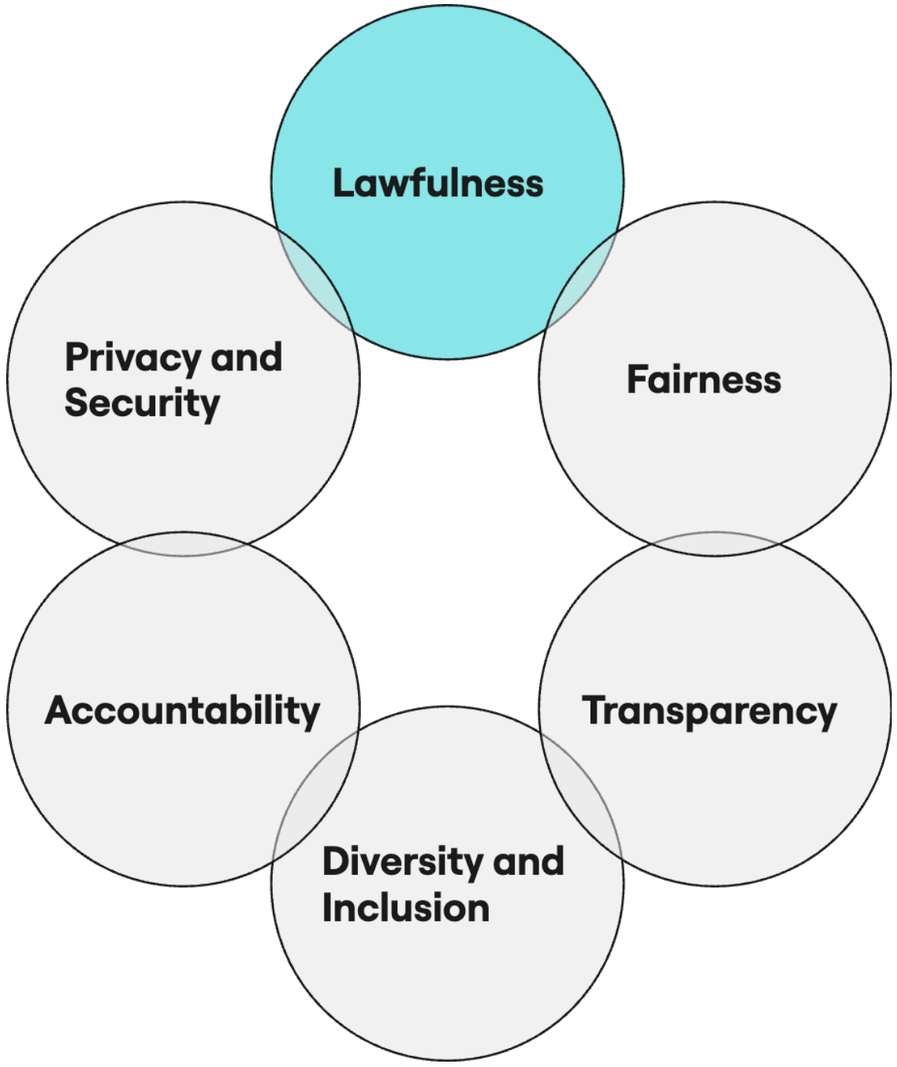
In this course

- Introduce responsible AI dimensions and metrics
- Apply concepts to the real world
- Overview of regulation and licensing
- Data governance and acquisition
- Validation and bias mitigation

Lawfulness

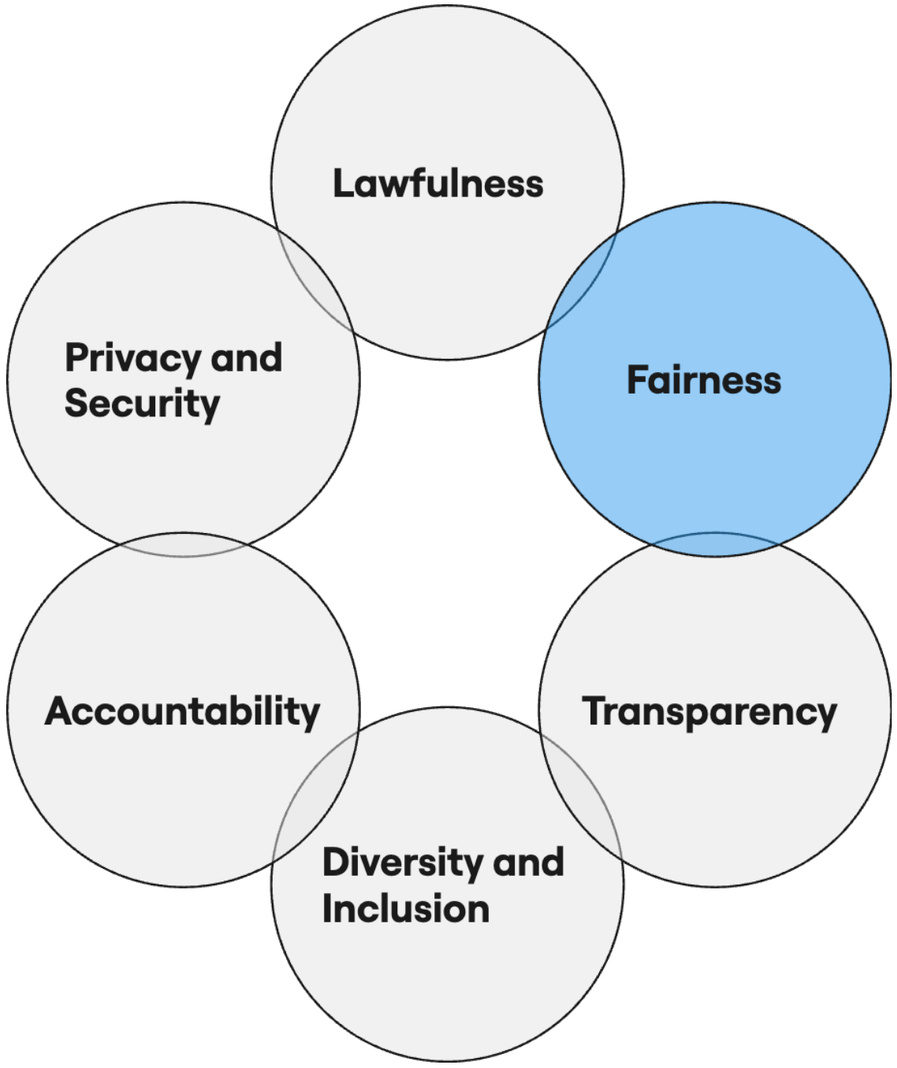
- Compliance with laws and regulations
- Ensures data is collected, processed, and used correctly
- Some laws and regulations include:
  - Data protection laws
  - Human rights laws
  - Ethical regulations towards stakeholders
  - Can differ depending on the governing body or country

Always confirm what applies!



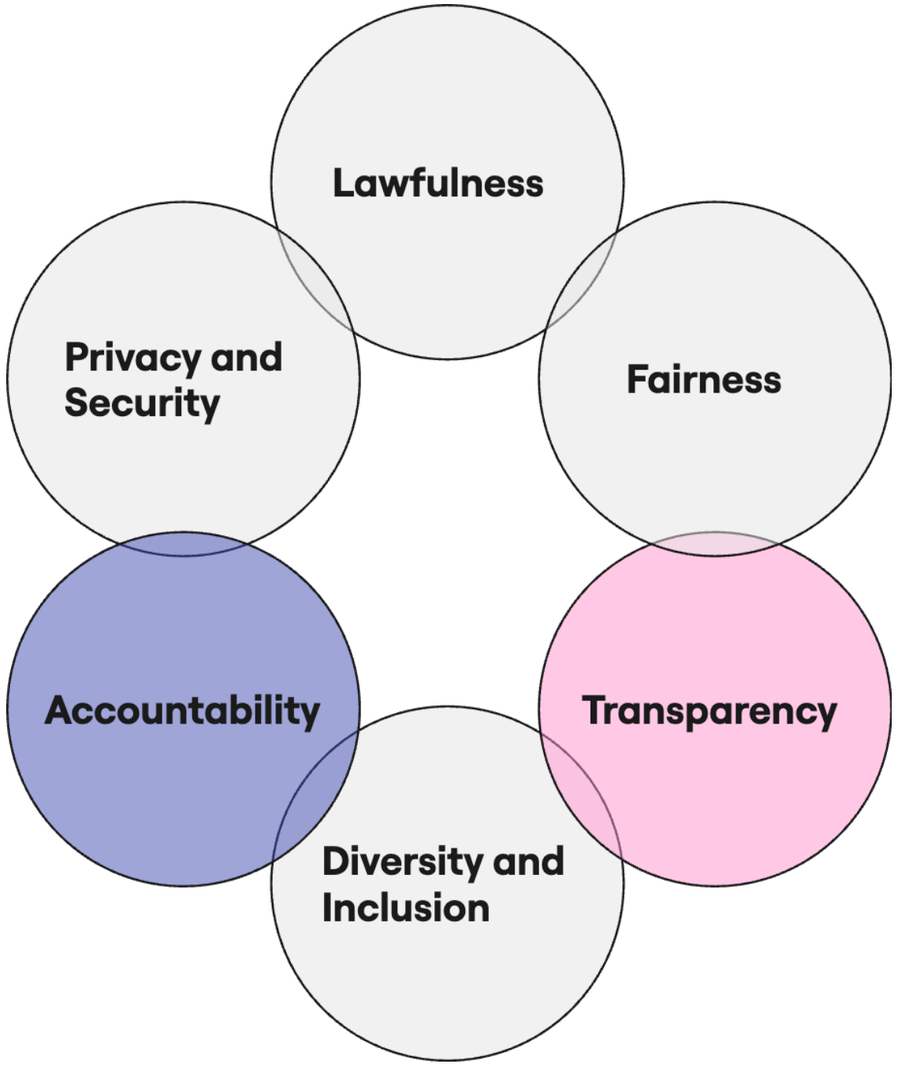
Fairness

- Algorithms and data practices do not create inequalities
- Treat everyone fairly, without discrimination



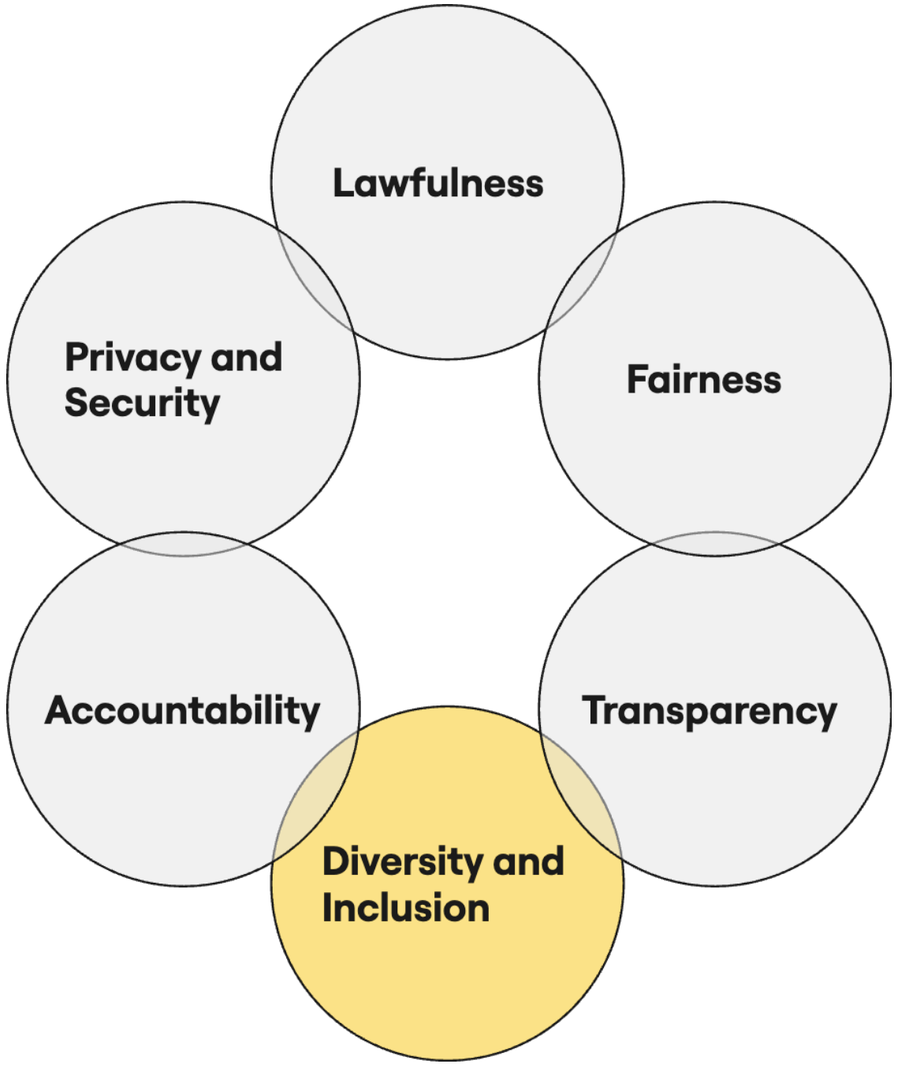
Transparency and accountability

- How the data is used
- How the model is developed
- How decisions are made
- Explain the AI
- Build stakeholders trust



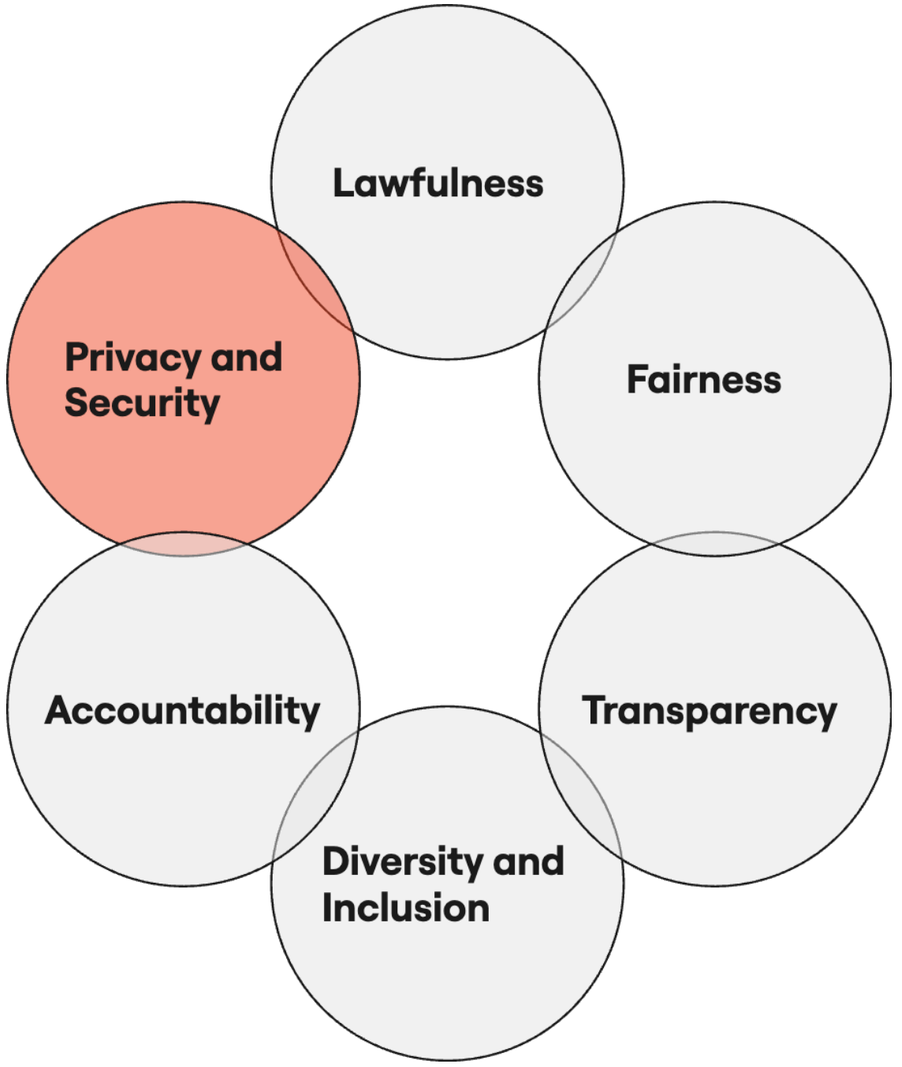
Diversity and inclusion

- Data diversity
- Diverse perspectives and experiences
- Key for bias mitigation



Privacy and security

- Safeguarding of personal and sensitive data
- Respecting and protecting individual rights
- Protect data and models from unauthorized access



Amazon AI hiring tool

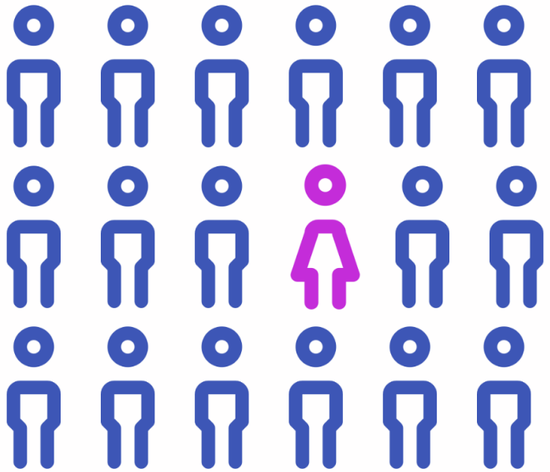
- Amazon: 2015-2017
- Automated talent acquisition
- Use AI to rate job applicants
- Led to scandal and abandoning of the initiative

[1] Reuters: https://www.reuters.com/article/idUSKCN1MK0AG/

Challenges of AI models

What went wrong?

- Not gender-neutral
- Imbalanced training data
- Used only technical metrics for AI evaluation



Let's practice!

RESPONSIBLE AI DATA MANAGEMENT

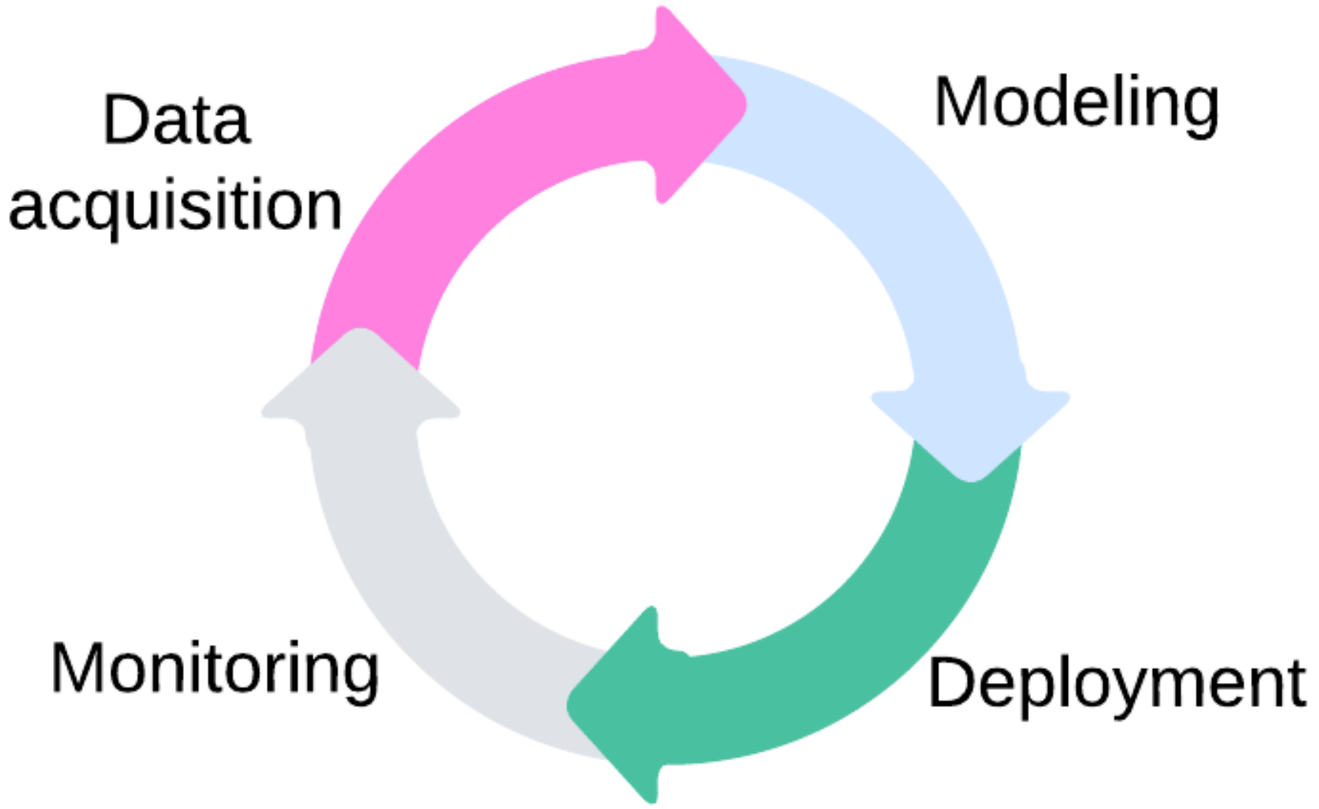
Responsible AI metrics

RESPONSIBLE AI DATA MANAGEMENT

Maria Prokofieva
Lead ML Engineer

datacamp

AI project lifecycle

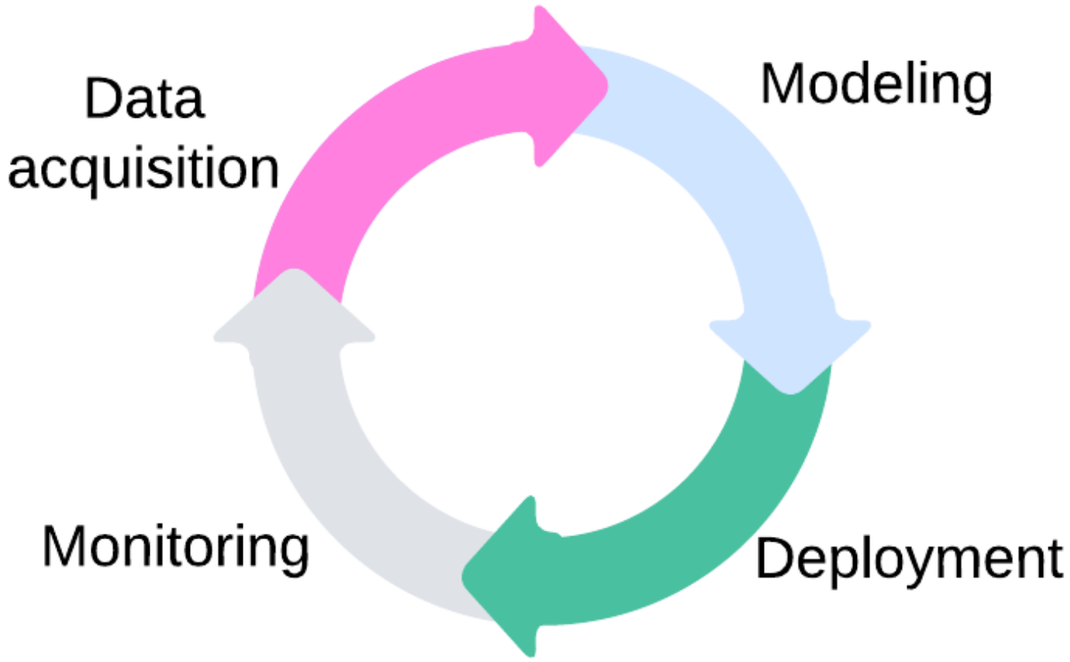


Responsible AI project

- Legally compliant
- Fair and diverse
- Practices are transparent, accountable, and secure

Model fairness:

- Fair, unbiased, and has equitable outcomes for everyone



Protected characteristics

- Groups likely to be treated unfairly and face discrimination
- Defined by protected characteristics:
  - Race
  - Ethnicity
  - Gender
  - Socioeconomic background

Data acquisition

- Equal outcomes
- Demographic disparity
- Laws and regulations

AI in facial recognition

- High accuracy

BUT

- Fail to capture specific ethnicities or genders

WHY...

Lack of:

- Data availability
- Diversity
- Representation

Equal outcomes and demographic disparity

Equal outcomes: benefits are equal across groups

Conditional demographic disparity: differences between groups

- Use descriptive statistics and distributions to assess data diversity
- Corrective measures: weighting and balancing
- Revisit after modeling
- Keep track of tests

Modeling

- Equal performance

For example, medical diagnoses:

- Some more common in protected groups

- Evaluate false negatives, false positives, and accuracy
- Explainability:
  - Local Interpretable Model-agnostic Explanation (LIME)
  - Shapley Additive Explanation (SHAP)

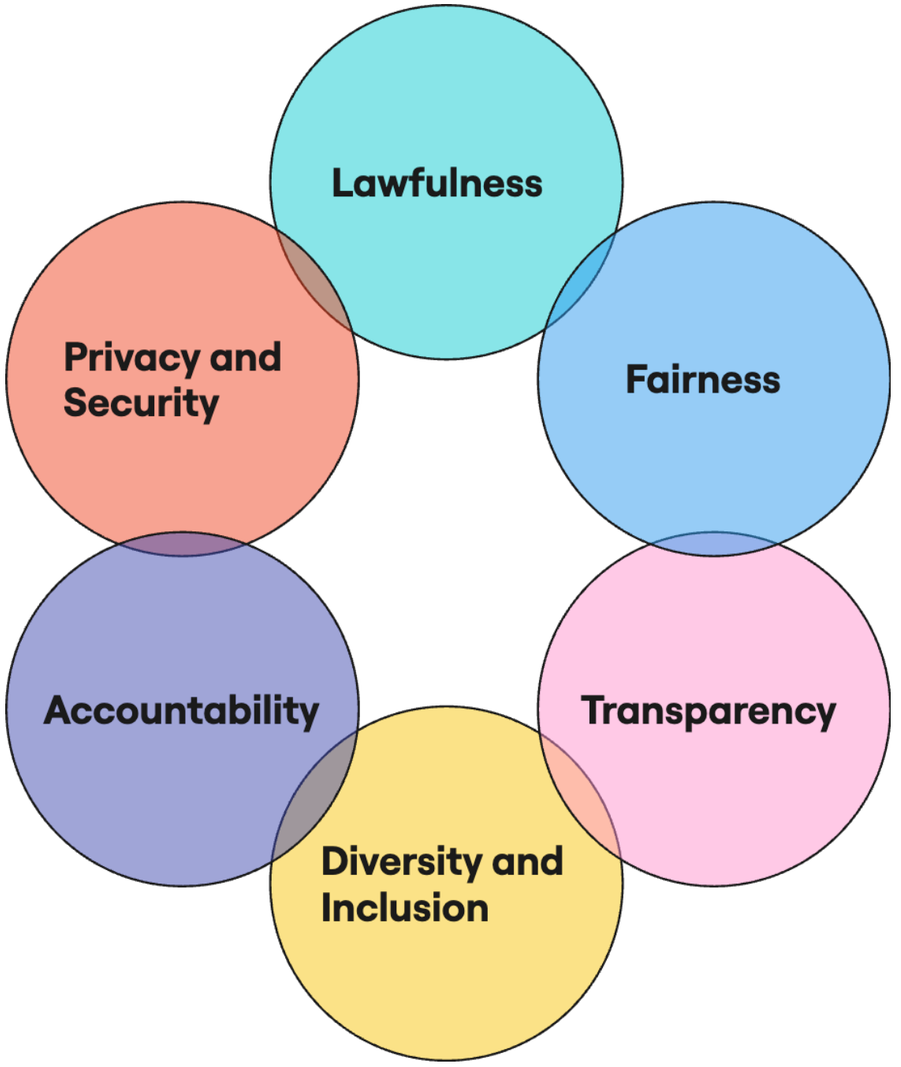
Deployment and monitoring

- Model drift:
  - Changes to model performances over time

- Monitor distributions
- Technical performance metrics
- Adjust model
- Keep track!

Applying metrics

- Understand the protected characteristics
- Many more metrics exist!
- **Always consult appropriate legal and domain experts**
- Conduct privacy and security checks



Let's practice!

RESPONSIBLE AI DATA MANAGEMENT

Challenges of responsible AI

RESPONSIBLE AI DATA MANAGEMENT

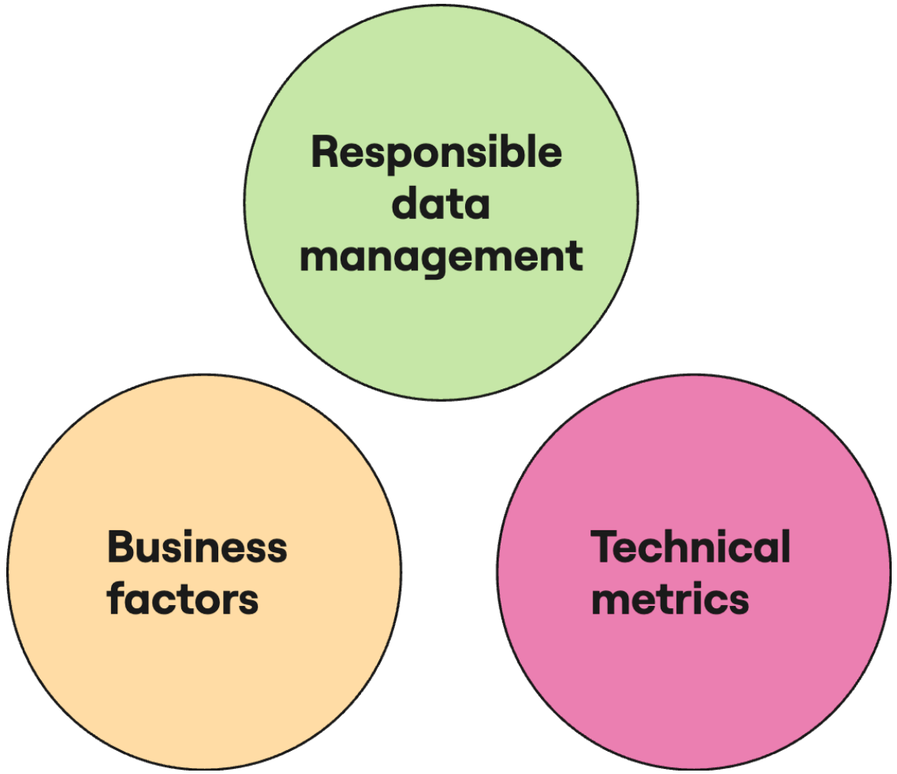
Maria Prokofieva
Lead ML Engineer

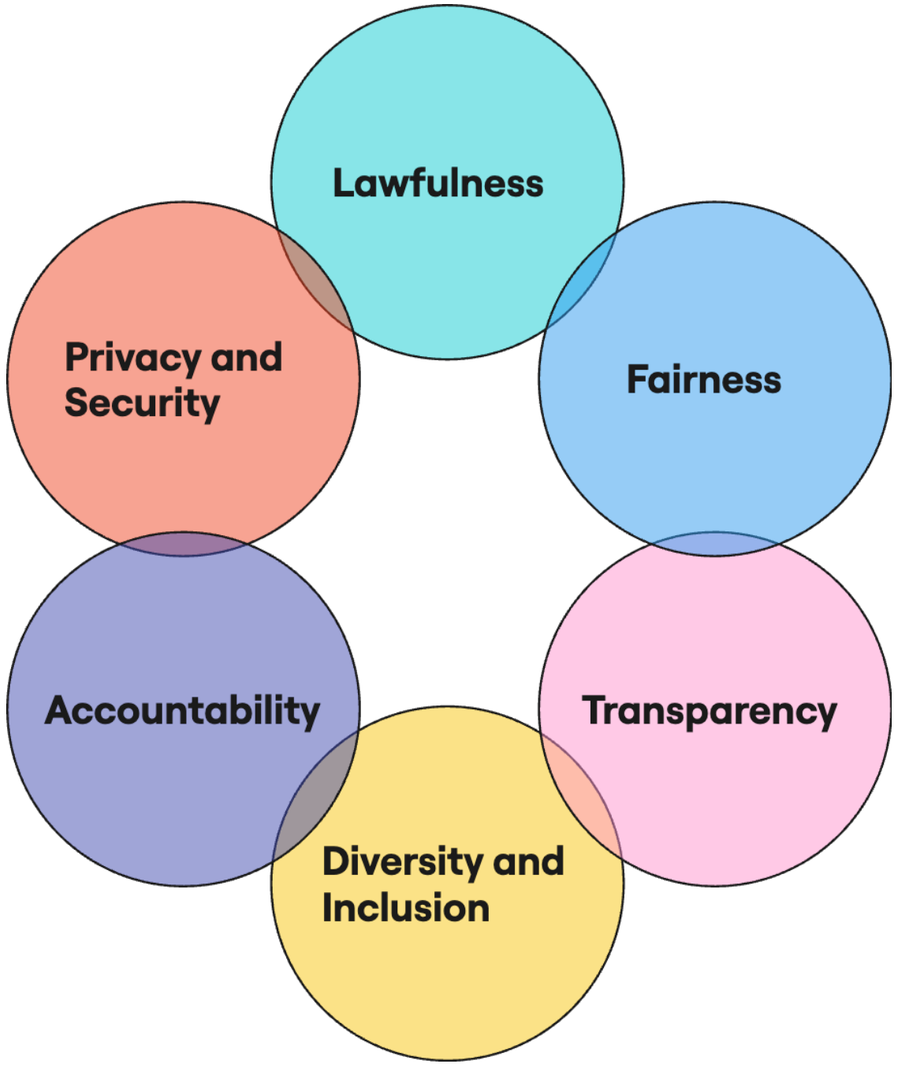
datacamp

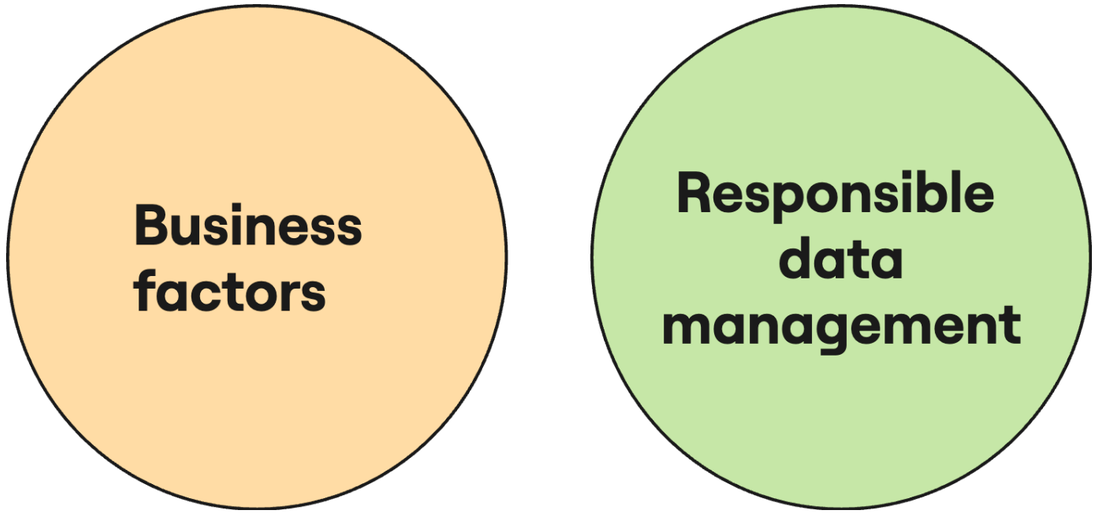
Responsible data management practices in the real world

- Complex
- Involves trade-offs
- Professional judgement

Common trade-offs



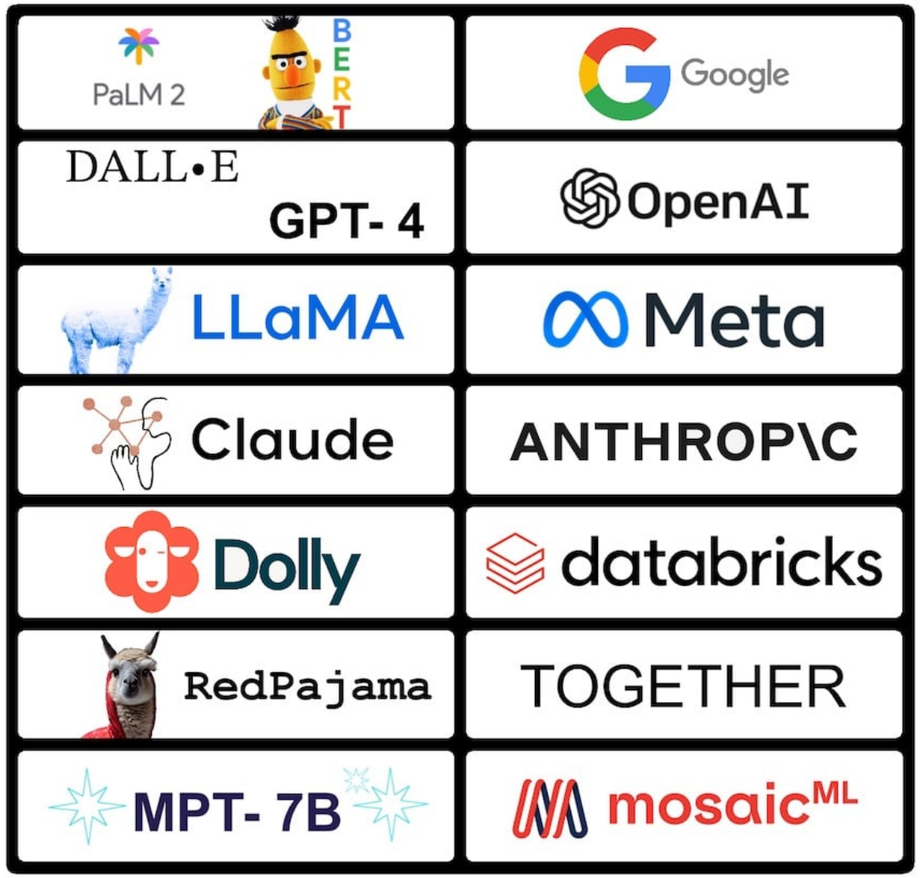




- Profit over fairness and privacy
- Revenue over testing and security

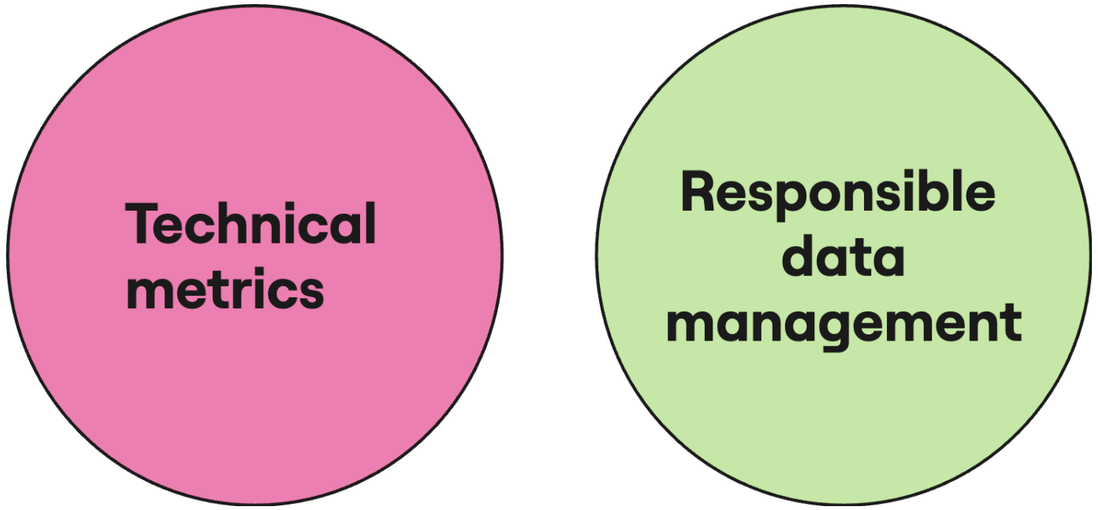
Pre-trained models

- Reduce costs
- Save time and resources
- No need for data collection and training
- Might have biased training data
- Lack transparency



Using pre-trained models

- Due diligence on model source
- Good reputation
- Credibility
- Review model documentation
- Additional tests for fairness and bias

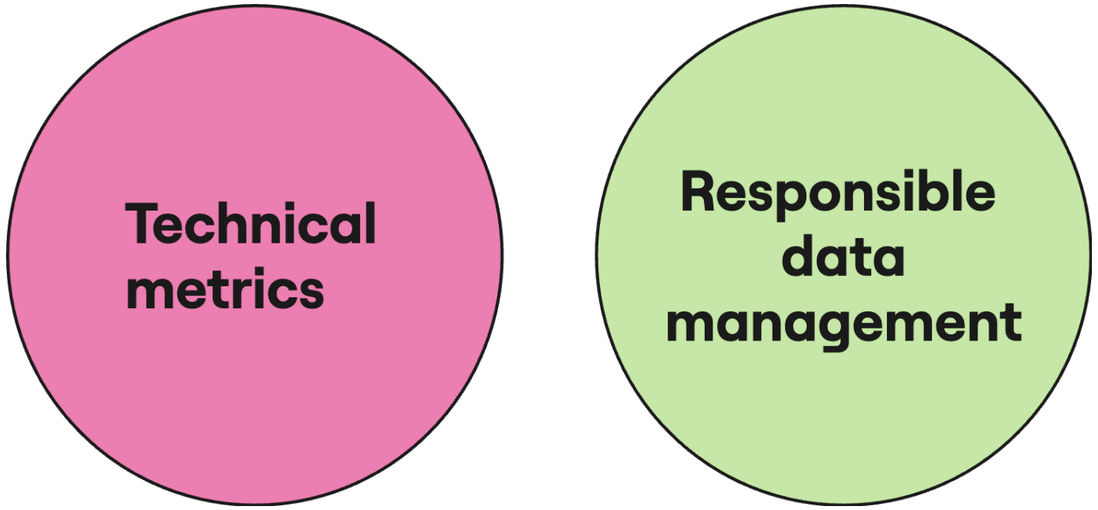


- Accuracy over fairness
- Even in balanced datasets

Accuracy trade-offs

- Lower accuracy for specific groups
- No account for data quality or quantity for underrepresented groups
- Privacy reduces accuracy

Robustness trade-offs



- Robustness versus bias
- Robustness versus fairness

Professional conduct and duties of care

- Code of ethics and conduct
- Guidance varies by country and organization
- Responsibility, non-harm, fairness
- User privacy and confidentiality
- Positive impact on society
- Maintain high standards
- Develop robust and secure systems
- Inclusive and non-discriminatory

[1] https://www.acm.org/code-of-ethics

Let's practice!

RESPONSIBLE AI DATA MANAGEMENT