Rensselaer

why not change the world?®

Ethics of Data Science

Ahmed Eleish

Data Science ITWS/CSCI/ERTH-4350/6350

November 4th, 2026

Tetherless World Constellation
Rensselaer Polytechnic Institute



Contents

- Data Privacy
- Bias and Fairness
- Transparency and Interpretability
- Responsible AI Integration
- Regulatory and Policy Frameworks

1. Data Privacy





+ Delete/Purge

https://uwyo.libguides.com/publishing-research-data

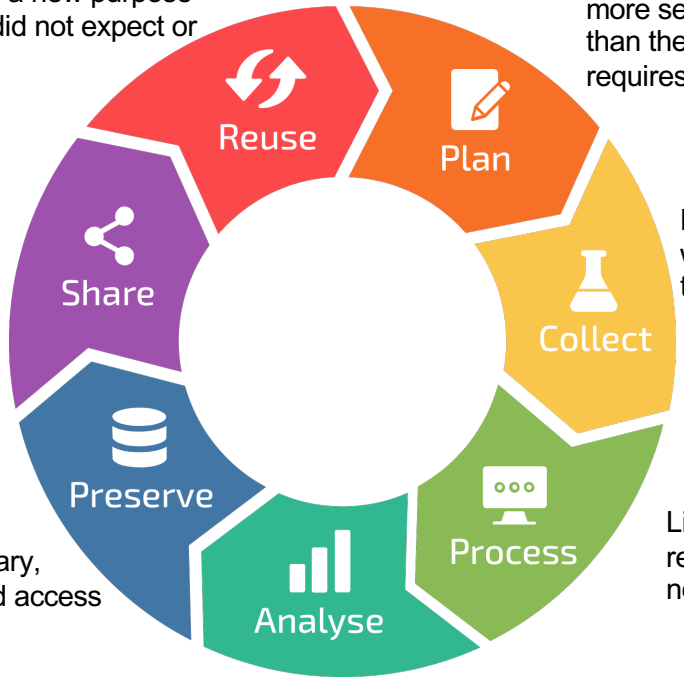
Privacy Across the Data Lifecycle

- Collection: what is gathered, and why?
- Use: analysis, modeling, and secondary purposes..
- Sharing: third parties, public release, data brokers..
- Retention and deletion: how long data persists..

Data are used for a new purpose that participants did not expect or agree to.

The project plans to collect more sensitive information than the research question requires.

Shared data can be re-identified by combining them with other available information.



Participants' data are collected without clear, informed consent for the intended uses.

Data are retained longer than necessary, increasing the chance of unauthorized access or breach.

Linking datasets or transforming records exposes identities that were not obvious in the original data.

Analysis infers sensitive traits about individuals or groups that they did not disclose.

+ Delete/Purge

Copies remain in backups, exports, or collaborator systems after the primary dataset is deleted.

https://uwyo.libguides.com/publishing-research-data

“Available” Does Not Mean “Appropriate”

- Publicly accessible data can still be sensitive.
- Contextual integrity: information flows carry norms..
    - Violating these norms can violate expectation of privacy.
- Combining harmless data can reveal sensitive facts.
- Principles: purpose limitation, data minimization..
    - Limit the use of data to a clearly communicated purpose, only collect and retain data necessary for that purpose.

Example of violation of contextual integrity

1. Health Forum

- Person posts about health concern on forum seeking peer support.

2. Research

- Forum posts are scraped and used to study community needs.

3. Data Broker

- Information from posts is combined with other data to infer a person's health profile.

4. Insurer

- Health profile could be used in a business decision.

Expected vs. Actual Data Use

- Users typically expect the service they signed up for.
- Actual use often includes sharing and model training.
- The gap between the two drives harm and loss of trust.
- Transparency about secondary use narrows the gap.

Expected vs. Actual Data Use

What users expect	What can happen to the data
Location is used to give directions to the destination.	Location history is stored over time, and patterns such as home, workplace, and routines can be inferred.
Searches are private, like a one-time lookup.	Searches may be linked to the account and used to personalize results or ads.
Deleting the app removes the data.	Server-side records may remain until the user deletes them or the retention period ends.
Anonymous or aggregated data cannot identify anyone.	Location traces are often unique, so re-identification can still be possible.

Re-identification and Sensitive Inference

- Removing names rarely makes data anonymous.
- Quasi-identifiers (ZIP, birth date, sex) can be linked.
- Models can infer traits that were never collected.
- Harms can fall on groups, not only individuals.

Example of re-identification

- A city publishes a dataset of taxi trips with pickup/dropoff locations/times but without passenger information.

➢ Someone with personal knowledge of a person's pickup location/time could identify the record and learn the destination

OR

➢ Geotagged social media posts from a destination could be matched to a trip revealing the origin or vice versa.

Privacy-Preserving Techniques

- **Differential privacy**: calibrated noise is added to limit what may be revealed by any one person's data.
- **Federated learning**: train without centralizing raw data/ data remains on users' devices.
- **Cryptographic methods**: homomorphic encryption, data remains encrypted while computations are performed on it.
- Pseudonymization and anonymization..

Limits and Trade-offs

- Privacy vs. utility: more noise means lower accuracy.
- Federated learning alone does not guarantee privacy.
- Cryptographic approaches add computational cost.
- Technical tools complement governance; they do not replace it..

- ➢ A complete solution combines policy + technical tools + professional responsibility
- ➢ Continual monitoring ensures compliance

Notice and Consent: Weaknesses and Limits

- Policies are long, and few people read them.
- Consent is often bundled and take-it-or-leave-it.
- Future uses cannot be known at collection time.
- Meaningful consent is specific, informed, and revocable.

End User License Agreement

ADOBE Software License Agreement

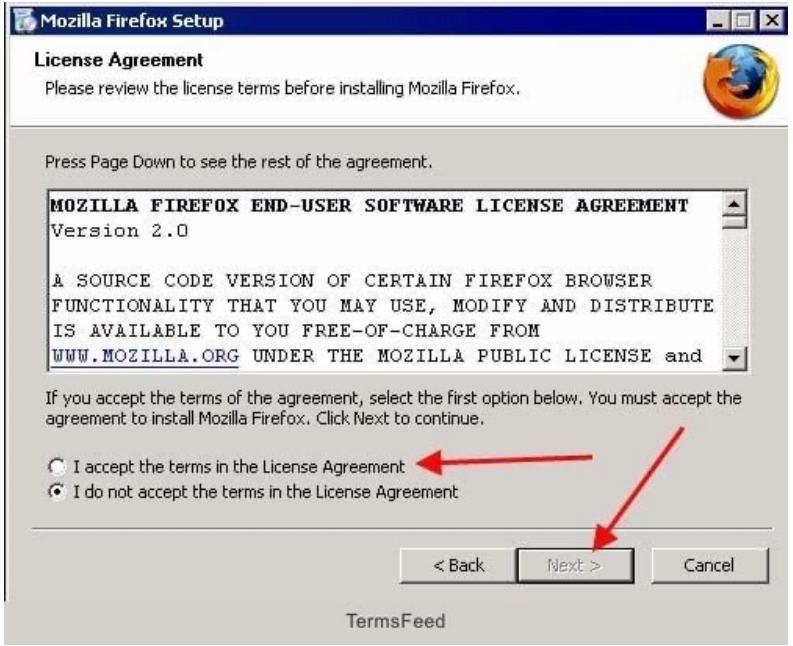
PLEASE READ THIS AGREEMENT CAREFULLY. BY COPYING, INSTALLING, OR USING ALL OR ANY PORTION OF THIS SOFTWARE, YOU (HEREINAFTER "CUSTOMER") ACCEPT ALL THE TERMS AND CONDITIONS OF THIS AGREEMENT, INCLUDING, WITHOUT LIMITATION, THE PROVISIONS ON LICENSE RESTRICTIONS IN SECTION 4, LIMITED WARRANTY IN SECTIONS 6 AND 7, LIMITATION OF LIABILITY IN SECTION 8, AND SPECIFIC PROVISIONS AND EXCEPTIONS IN SECTION 16. CUSTOMER AGREES THAT THIS AGREEMENT IS LIKE ANY WRITTEN NEGOTIATED AGREEMENT SIGNED BY CUSTOMER. THIS AGREEMENT IS ENFORCEABLE AGAINST CUSTOMER. IF CUSTOMER DOES NOT AGREE TO THE TERMS OF THIS AGREEMENT, CUSTOMER MAY NOT USE THE SOFTWARE.

Customer may have another written agreement directly with Adobe (e.g., a volume license agreement) that supplements or supersedes all or portions of this agreement. The Software is LICENSED, NOT SOLD, only in accordance with the terms of this agreement. Use of some Adobe and some non-Adobe materials and services included in or accessed through the Software may be subject to additional terms and conditions. Notices about non-Adobe materials are available at http://www.adobe.com/go/thirdparty.

The Software may cause Customer's Computer to AUTOMATICALLY CONNECT TO THE INTERNET. The Software may also require activation or registration. Additional information on activation, Internet connectivity, and privacy is available in Sections 14 and 16.

1. Definitions.

1.1 "Adobe" means Adobe Systems Incorporated, a Delaware corporation, 345 Park Avenue, San Jose, California 95110, if



People Who Never Consented

- Data about others: contacts, family, genetic relatives.
- Inferences made about groups and communities.
- Bystanders captured in images and sensor data.

Walmart surveillance patents

- **Biometric & Emotional Tracking**: Patents describe using **facial recognition**, **in-store cameras**, and **wearable data** (heart rate, glucose) to gauge emotional reactions to products.
- **Vehicle Profiling**: Cameras can measure **vehicle suspension gaps** in parking lots to infer purchase volume and adjust pricing.
- **Routine Mapping**: Technology can track **GPS signals** and **social media** to map daily routines, including work schedules and gym visits.

- Walmart denies using these patents for surveillance pricing

2. Bias and Fairness

Bias Across the Data Pipeline

- Sampling: who is in the data, who is missing?
- Labels and measurement: what is recorded, and how?
    - Instrument calibration, degree of precision, data structures, processing, etc.
- Historical decisions embedded in outcomes..
    - Training models on biased data recreates the bias.
- Deployment context: use differs from design..
    - Model generalizability to unseen data.

Missing and Unrepresentative Data

- Underrepresented groups often see higher error rates.
    - Models have not learned group-specific patters.
- Measurement quality can differ across groups.
    - Access to people, technology, etc.
- Absent data is invisible to aggregate metrics.
    - At the summary level, the nuances of representativeness in the data may be lost.

Historical Bias and Proxy Variables

- Models learn from past decisions and can reproduce them.
- Dropping protected attributes does not remove proxies.
- Label choice matters: a convenient proxy can encode bias.

Example: health-care cost used as a stand-in for health need

  - Problematic because patients with need but no access to health services will be scored as having low need.

Defining Fairness

- Group fairness: demographic parity, equal opportunity, equalized odds.
- Individual fairness: similar people, similar treatment.
- No single metric is correct for every context.

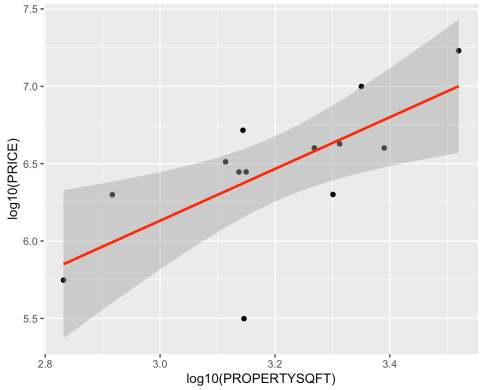
Mitigation Strategies

- Pre-processing: improve data collection, re-weigh samples..
- In-processing: add fairness constraints during training..
- Post-processing: adjust thresholds or outputs..

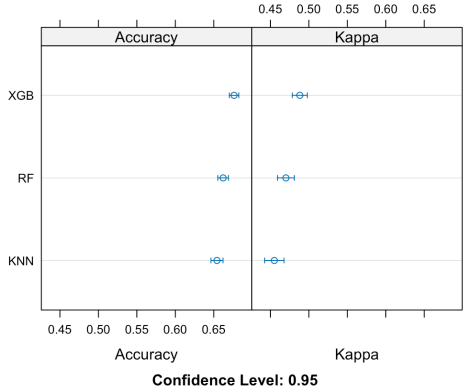


Communicating Uncertainty

- Report estimated probabilities and intervals, not just labels.
- Avoid false precision.
- State known limitations and out-of-scope uses.
- Uncertainty should change how a decision is handled.



Regression line with 95% confidence interval



Prediction accuracy with error bars

3. Responsible AI Integration

Start with the Problem

- Define the decision to be supported and **who it affects**.
- Clarify **goals** and what counts as **success**.
- Ask whether AI is the right tool at all.
- Identify data, constraints, and affected groups first.

Weighing Alternatives

- Compare non-AI and **simpler** approaches.
- List expected **benefits** and foreseeable **harms**.
- Consider scale, reversibility, and who bears the risk.
- “Do not deploy” is a legitimate outcome.

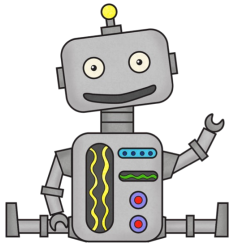
AI vs. Alternatives

Option	Potential benefit	Potential harm	Cost considerations	Oversight needs
AI-assisted review	Helps reviewers process applications and identify patterns at scale.	Biased data or errors may disadvantage some applicants; reviewers may over-trust scores.	Development or vendor fees, integration, testing, and ongoing monitoring.	Validate performance across groups; train reviewers; provide meaningful override and appeal.
Fixed eligibility rules	Consistent, transparent criteria are easier to explain.	Rigid rules may miss individual circumstances or encode unfair criteria.	Cost to establish, maintain, and apply rules.	Review rules for unequal impacts; allow exceptions and appeals.
Human review	Reviewers can consider context and unusual circumstances.	Decisions may be inconsistent or affected by human bias.	Staff time, training, and slower processing at scale.	Use clear criteria, quality checks, bias training, and appeals.

Meaningful Human Oversight

- A sign-off is not oversight.
- Automation bias: people over-trust system outputs..
- Reviewers need authority, information, and time to override.
- Track override rates and outcomes.
- Remedies when an error or harm is confirmed.





Professional Responsibility

- Professional codes (e.g., ACM Code of Ethics) set duties.
- Accountability cannot be delegated to a model.
- Document assumptions and trade-offs.
- Know how and when to escalate concerns.

Shared Responsibility in the AI Supply Chain

- Data providers, model developers, deployers, and users.
- Contracts and documentation should make duties explicit.

Post-Deployment Monitoring

- Track performance overall and by subgroup.
- Detect data drift and changing conditions.
  - Measure real-world impact, not just model metrics.
- Define alert thresholds in advance.
- Provide clear channels for reporting harm.
  - Triage, investigate, and remediate quickly.
  - Disclose significant incidents appropriately.
- Feed lessons back into design and documentation.



Pause, Rollback, Retire

- Set stop criteria before launch.
- Have a tested rollback or fallback process.
- Plan for decommissioning and data disposal.
- Responsibility continues through the end of the system's life.



“I'm sorry Dave, I'm afraid I can't do that”

5. Regulatory and Policy Frameworks

Why Jurisdiction and Context Matter

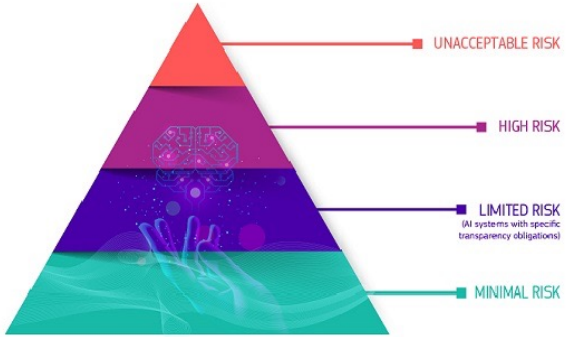
- Duties depend on location, sector, data type, and use.
- Multiple regimes can apply to the same system.
- Law sets a floor; ethics often asks for more (**professional responsibility**).
- Always confirm current requirements with **counsel**.

Data Protection Law

- GDPR (EU): lawful basis, purpose limitation, data minimization, data-subject rights.
- GDPR includes safeguards on solely automated decisions.
    - The largest fine ever issued was **€1.2 billion** against **Meta** in May 2023 for unlawful EU-to-US data transfers.
- U.S.: sector laws (e.g., HIPAA, FERPA) and state privacy laws such as CCPA/CPRA.
- Common thread: transparency, limits on use, individual rights.

AI-Specific Regulation

- EU AI Act: risk-based, with prohibited practices and high-risk obligations:
  - Unacceptable risk is prohibited (e.g. social scoring systems and manipulative AI).
  - Most of the text addresses high-risk AI systems, which are regulated.
  - A smaller section handles limited risk AI systems, subject to lighter transparency obligations: providers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes).
  - Minimal risk is unregulated.



https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

- U.S.: mostly sector-specific rules and emerging state laws.

NIST AI Risk Management Framework

- Voluntary U.S. framework (AI RMF 1.0, 2023)
- Four functions: Govern, Map, Measure, Manage..
- Emphasizes trustworthiness characteristics and context.
- Turns principles into assessment and documentation.

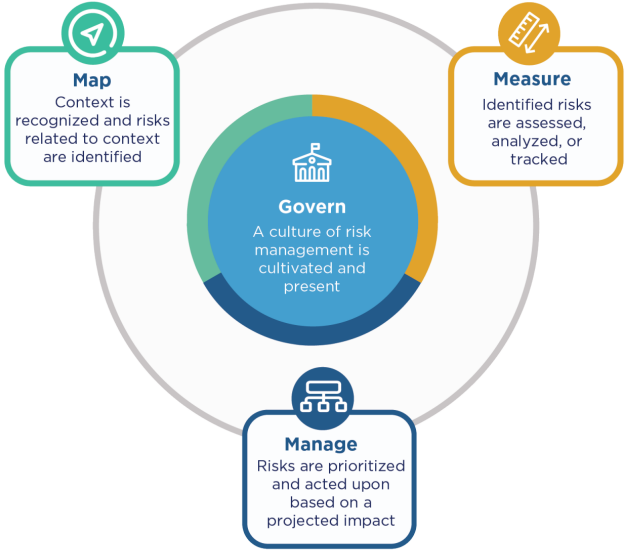


Fig. 5. Functions organize AI risk management activities at their highest level to govern, map, measure, and manage AI risks. Governance is designed to be a cross-cutting function to inform and be infused throughout the other three functions.

https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Other Standards and Principles

- ISO/IEC 42001: AI management system standard.
- OECD AI Principles: widely cited international baseline.
- Professional and institutional codes of ethics.
- Alignment across frameworks reduces duplicated effort.

Summary

- Ethical data science spans the entire lifecycle, not just modeling.
- Fairness, privacy, and transparency involve trade-offs that must be justified.
- Human oversight and accountability must be designed in from inception.
- Regulation sets a floor; professional judgment goes further.

Readings

Select 1 from:

Data in Research

- Ten Simple Rules for Reproducible Computational Research [32]
- Reproducible Research in Computational Science [33]
- An empirical analysis of journal policy effectiveness for computational reproducibility [34]

And select 1 from:

Ethics of Data Science

- EU AI Act [35]
- GDPR [36]
- OECD AI Principles [37]

2 total :-)

Assignment 4

- Group + individual
- Due November 22nd

Thanks!