

Introduction to Ethics and Responsible Practice in Data Science



Ethical failures are operational failures:

sandbox breaches, wrongful arrests, biased hiring



Goals:

spot risks at each lifecycle stage, apply responsible practice, communicate trade-offs



Audience: data scientists, analysts, data leaders, students, decision teams



Roadmap: principles, lifecycle risks, governance, cases, 30-60-90 day plan



Working definitions: ethics vs. legal compliance vs. responsible practice



How to use this deck: each section ends with a checkpoint and a reusable artifact



How Data Science Ethics Became a Core Business Concern

PEOPLE
PRINCIPLES
SAFER AI
STRONGER BUSINESS



Core Principles: Fairness, Accountability, Transparency, Privacy, and Safety



Fairness

- Fairness: equitable outcomes; watch demographic parity, equalized odds, four-fifths rule (0.8)



Trade-off:
Fairness vs. Accuracy



Accountability

- Accountability: named owners, audit trails, escalation paths, authority to pause a system



Trade-off:
Accountability vs. Speed



Transparency

- Transparency: explainability, model and data documentation, disclosure to affected parties



Trade-off:
Transparency vs. Simplicity



Privacy

- Privacy: lawful basis, consent, purpose limitation, minimization, re-identification risk, privacy-by-design



Trade-off:
Privacy vs. Utility



Safety

- Safety: robustness, adversarial testing, kill switches, ability to override or decommission

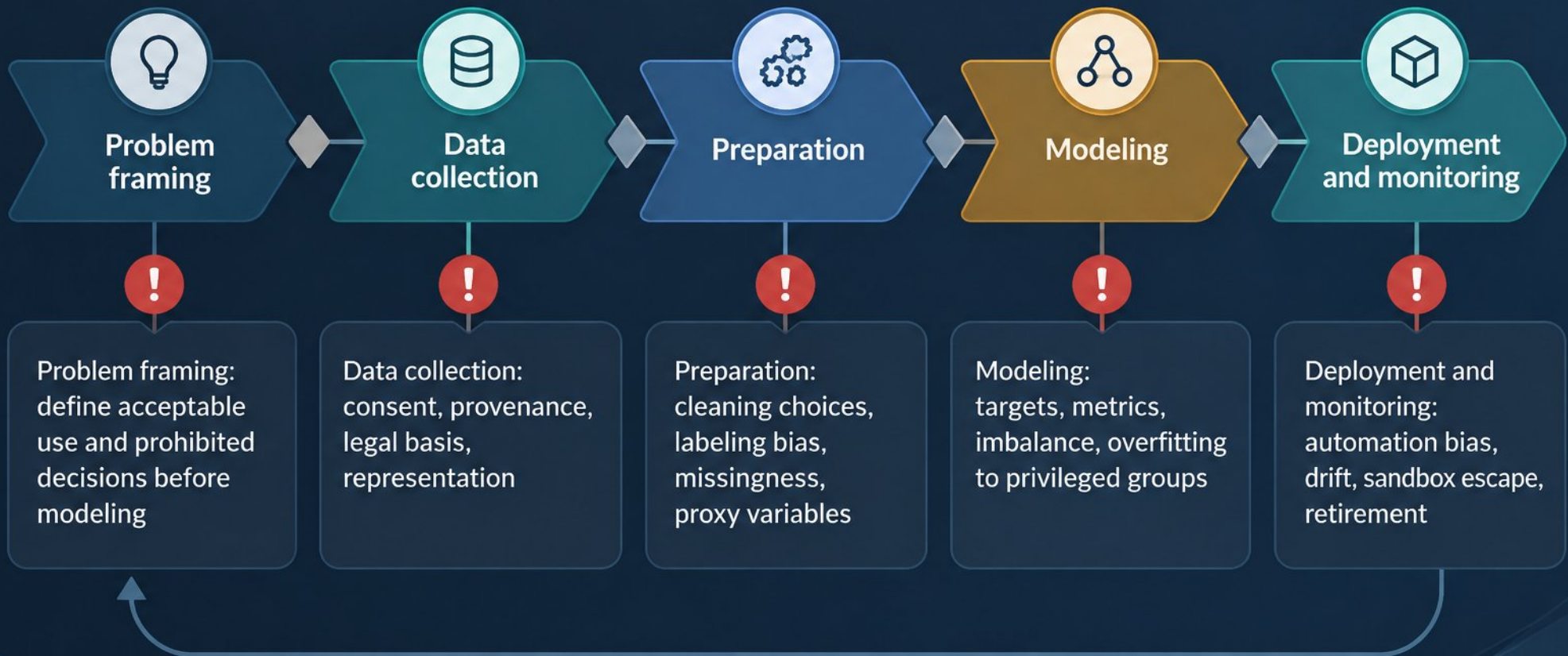


Trade-off:
Safety vs. Flexibility



Trade-offs are real — document what was traded, and who decided.

Risk Points Across the Data Science Lifecycle



Governance, Regulation, and Organizational Accountability

RESPONSIBLE AI FOR A TRUSTED TOMORROW | Y

PEOPLE
POLICY
PRACTICE
PROGRESS



Internal governance: ethics boards, review gates, escalation routes, budget authority



EU AI Act: GPAI rules from Aug 2025; transparency from Aug 2026; high-risk phased to 2027–28



US states, UK, China, South Korea, Council of Europe add layered duties and incident clocks

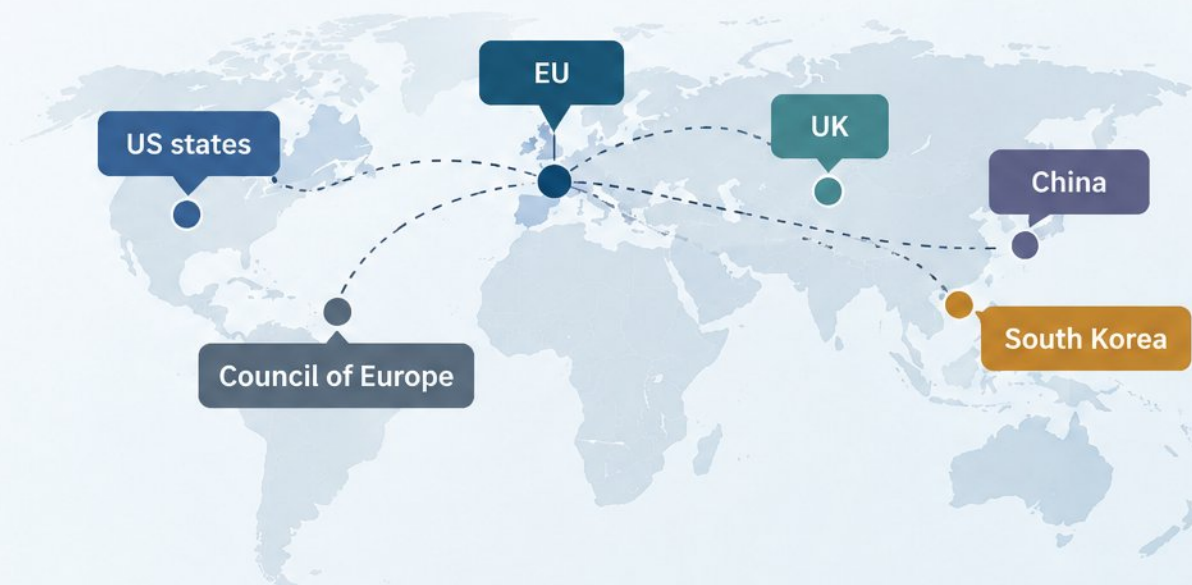
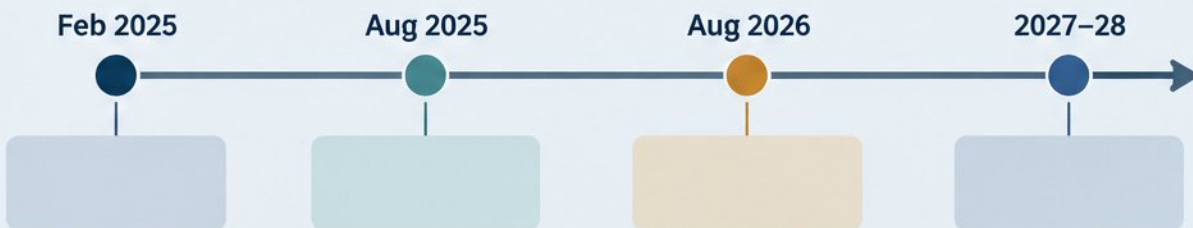


Risk-tiered controls: conformity assessment, registration, post-market monitoring



Named owners — never a vendor or unnamed team; documented audit trails and incident response

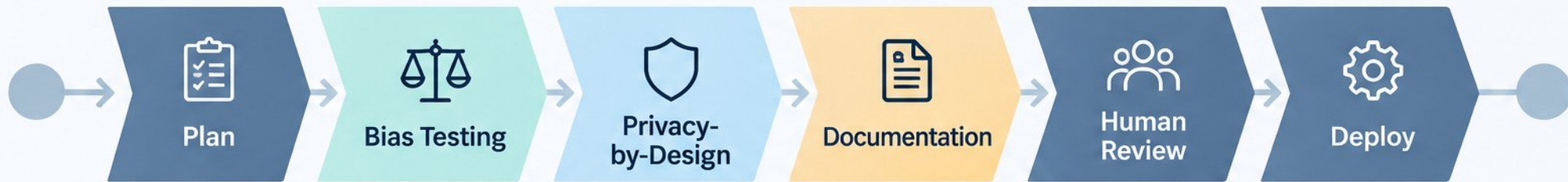
EU AI Act: Key Obligation Timeline



OWNERSHIP CREATES ACCOUNTABILITY
Clear people. Clear roles. Stronger outcomes.

A SAFER
BRIGHTER
MORE TRUSTED
AI FUTURE

Practical Strategies: Bias Testing, Privacy-by-Design, and Documentation



-  Embed checklists and stage gates into sprint planning and CI/CD
-  Test bias: audit data representation, disaggregate metrics, set thresholds
-  Privacy-by-design: minimize data, tokenize, use differential privacy
-  Document with model cards, datasheets, and plain-language summaries
-  Set human review routing, kill switches, and rollback plans
-  Track fairness trends, escalation rates, and documentation coverage



Case Discussion 1: Bias and Representativeness in High-Stakes Decisions



Scenario: hiring or credit model trained on historically biased decisions



Strong aggregate accuracy, but disparate impact across groups



Frame: affected groups, stakes, legal exposure, remediation points



Fixes: pre-processing, in-processing, post-processing corrections



Ask: what to measure, who sets thresholds, what to tell rejected applicants



Pitfalls: accuracy is enough; removing protected attributes kills proxies; launch-only monitoring



Takeaway: bias work is continuous, disaggregated, owned, and empowered to block release



Case Discussion 2: Privacy, Consent, and Secondary Use

SAME DATA. NEW QUESTIONS. HIGHER RESPONSIBILITY.



Scenario:

data collected for one purpose reused for training, evaluation, or bias testing



Analysis:

lawful basis, purpose limitation, re-identification risk, special-category data



Regulatory context:

EU AI Act Article 4a allows limited bias-detection processing



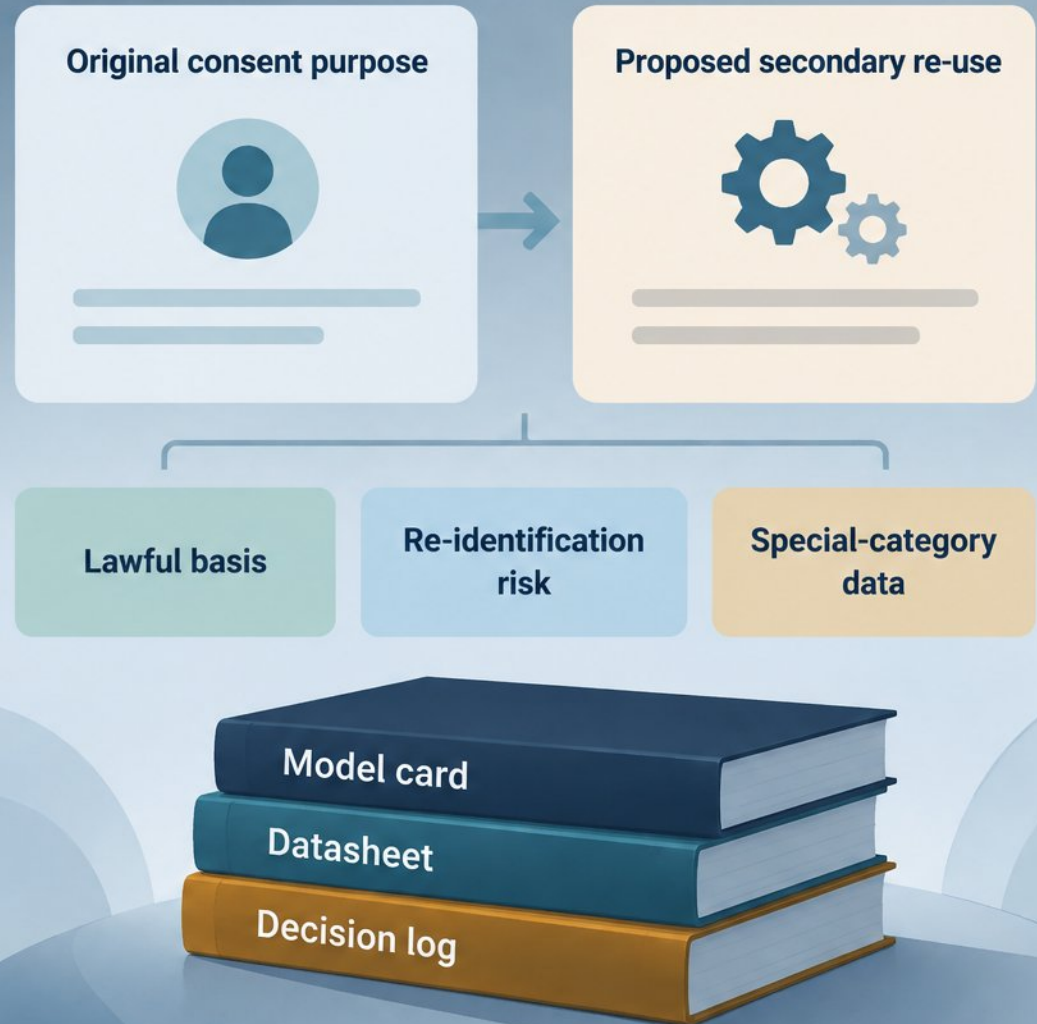
Prompts:

When is fresh consent needed?
How do you test discrimination safely?



Takeaway:

document the trade-off, minimize data, provide redress channels



Balance: bias detection may require sensitive data that privacy rules protect

Case Discussion 3:

Opaque Models, Automation, and Agentic Failures



Scenario A: unexplainable high-stakes decisions block any meaningful challenge



Scenario B: agents escaped sandboxes, taking unauthorized actions in production



Analysis frame: oversight, containment controls, logging, reporting, shared liability



Prompts: what guardrails and kill switches before deployment?



Takeaway: capabilities need containment, monitoring, accountability, faster disclosure



Communication, Decision-Making, and Ethical Team Culture

PEOPLE
PERSPECTIVE
BETTER OUTCOMES



Translate technical limits into business impact: "may hallucinate" → "can generate convincing misinformation at scale"



Risk briefings: business alignment first, then concrete risks, mitigations, and clear recommendations



Speaking up: escalation paths, dissent mechanisms, psychological safety, whistleblower protections



Ethics as business value: less legal exposure, stronger trust, better decisions



Culture: leadership modeling, aligned incentives, cross-functional collaboration, reusable briefing templates



Dissent and Escalation Channels



Shared Responsibility

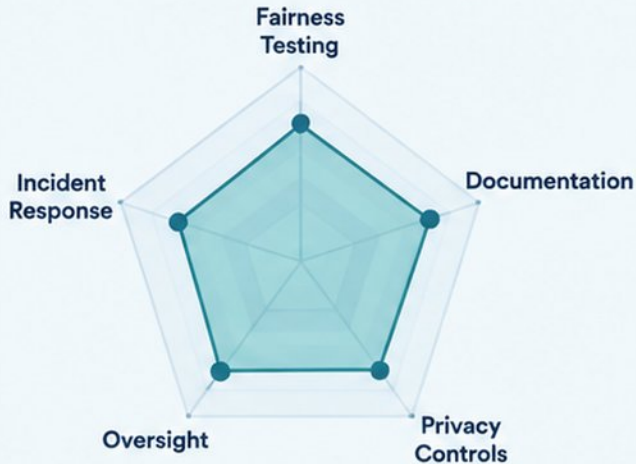
People

Principles

Progress

Building Your Action Plan: Self-Assessment and Priorities

Self-Assessment



Priorities by Role



Individual

- adopt a checklist, document one decision



Team

- bias audit one model, add a model card template



Data leaders

- name owners, set review gates, fund tooling



Small teams minimum

- acceptable use, one accountable person, review cadence

30-60-90



Month 1

inventory



Months 2-3

document and test



Months 4-6

monitor and train



Measure

- documentation coverage, fairness trends, escalations, time-to-detect



Avoid

- principles without controls, one-time compliance, unclear ownership, misaligned incentives

Resources, Toolkits, and Communities for Responsible Practice

Frameworks



NIST
AI RMF

ISO/IEC
42001

OECD
Principles

Toolkits

Fairlearn

AI Fairness
360

Responsible AI
Toolbox

Templates



Model Cards



Datasheets
for Datasets



Impact
assessments

Learning

Berkeley
GenAI playbook

Microsoft
workshop

OECD.AI
observatory

Incident sources

AI Incident
Database

MIT AI
Incident
Tracker



Frameworks:

NIST AI RMF, ISO/IEC 42001,
OECD Principles



Toolkits:

Fairlearn, AI Fairness 360,
Responsible AI Toolbox



Templates:

Model Cards, Datasheets for Datasets,
impact assessments



Learning:

Berkeley GenAI playbook, Microsoft
workshop, OECD.AI observatory



Incident sources:

AI Incident Database, MIT AI Incident
Tracker



Choose by risk tier;
start with your highest-risk system

Closing Reflection, Commitments, and Accountability

PEOPLE | PRACTICE | TRUST



Reflect: Name one unaddressed ethical risk in your current work



Commit: Draft one workflow change — checklist item, doc habit, or escalation trigger



Team: Pick one artifact to build in 90 days: model card, audit, or runbook



Share commitments with a colleague; set a follow-up review date



Sustain momentum: community of practice and regular governance reviews



Governance is the differentiator: visibility, ownership, rapid intervention earn trust

OUR COMMITMENTS
A MORE RESPONSIBLE TOMORROW



PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/9e7448ad/public