Data Safety & Ethics

Use AI Responsibly

ITAG Skillnet AI Advantage

90-Minute Lunch & Learn

Protect your data, respect privacy, and use AI ethically in your organization

Today's Agenda

10 min	Why Data Safety Matters in AI
15 min	Understanding AI Data Risks
15 min	The SHIELD Framework
20 min	Hands-On: Ethical AI Assessment
15 min	Compliance & Best Practices
10 min	Building an AI Ethics Culture
5 min	Q&A and Wrap-Up

Learning Objectives

Understand

- Key data safety risks when using AI tools
- Ethical considerations in AI adoption
- Regulatory landscape (GDPR, AI Act)

Apply

- The SHIELD framework for safe AI use
- Data protection best practices
- Ethical decision-making processes

By the end: You'll have practical skills to use AI tools safely while protecting sensitive data and maintaining ethical standards.

The AI Data Dilemma

AI tools are powerful, but they come with risks:

Data Exposure

What you input may be stored, logged, or used for training

Privacy Violations

Personal data can be inadvertently shared or exposed

Ethical Blind Spots

AI can perpetuate bias or make unfair decisions

Key Insight: Every prompt you send to an AI system is data that leaves your control. Think before you type.

Real-World AI Data Incidents

⚠ Samsung Semiconductor Leak

April 2023: Samsung engineers pasted proprietary semiconductor source code and internal meeting notes into ChatGPT within 20 days of company approval.

Result: Company-wide ChatGPT ban

Source: Bloomberg, The Economist (April 2023)

⚠ ChatGPT Payment Data Bug

March 2023: OpenAI bug exposed chat histories and payment info (credit card digits, emails) of ~1.2% of ChatGPT Plus subscribers to other users.

Result: 9-hour service shutdown

Source: OpenAI Blog (March 24, 2023)

⚠ Amazon Internal Data Warning

January 2023: Amazon lawyers warned employees after ChatGPT responses were found to closely resemble confidential Amazon internal data.

Result: Restricted AI use policy

Source: Business Insider (January 2023)

⚠ Italy GDPR ChatGPT Ban

March 2023: Italian DPA temporarily banned ChatGPT citing GDPR violations - lack of age verification, no legal basis for mass data collection.

Result: 1-month service ban in Italy

Source: Garante (Italian DPA) ruling

The Stakes Are High

$4.88M
Average cost of a data breach (2024)

10.7%
Of employee ChatGPT inputs contain sensitive data

4%
Maximum GDPR fine as % of global revenue

Sources: IBM Cost of a Data Breach Report 2024 | Cyberhaven AI Data Security Report 2024 | GDPR Article 83

Beyond Fines: Reputation damage, loss of customer trust, competitive disadvantage, and personal liability for decision-makers.

Types of AI Data Risks

Risk Type	Description	Example
Data Leakage	Sensitive info shared with AI provider	Pasting customer database records
Training Data Use	Your inputs used to train the model	Proprietary strategies in responses
Model Extraction	AI reveals patterns from other users	Memorized PII in outputs
Unauthorized Access	Third parties viewing your interactions	Shared API keys or accounts
Bias Amplification	AI perpetuates existing biases	Discriminatory hiring recommendations

What NOT to Share with AI

Never Share

- Passwords & API keys
- Social Security numbers
- Credit card details
- Medical records
- Personal addresses

Use Caution

- Customer names/emails
- Internal strategies
- Financial projections
- Employee information
- Unpublished research

Generally Safe

- Public information
- Generic questions
- Anonymized examples
- Hypothetical scenarios
- Open-source code

Rule of Thumb: If you wouldn't post it publicly on social media, don't put it in an AI prompt.

The Regulatory Landscape

Key Regulations

GDPR

EU data protection - consent, right to erasure, data minimization

EU AI Act

Risk-based AI regulation - transparency, human oversight

CCPA/CPRA

California privacy - opt-out rights, disclosure requirements

Common Requirements

- **Consent** - Get permission before processing
- **Transparency** - Explain how AI is used
- **Data minimization** - Only collect what's needed
- **Purpose limitation** - Use data only as stated
- **Accountability** - Document your practices

The SHIELD Framework

A systematic approach to safe and ethical AI use

- **S** **Sanitize Data** - Remove or mask sensitive information before sharing
- **H** **Human Oversight** - Keep humans in the loop for important decisions
- **I** **Informed Consent** - Ensure data subjects know about AI processing
- **E** **Evaluate Ethics** - Consider fairness, bias, and impact before use
- **L** **Legal Compliance** - Follow GDPR, AI Act, and industry regulations
- **D** **Document Everything** - Record AI decisions, data flows, and rationale

S - Sanitize Data

Techniques

- **Anonymization** - Remove identifying info entirely
- **Pseudonymization** - Replace with fake identifiers
- **Data masking** - Hide portions (****1234)
- **Aggregation** - Use summaries, not individuals
- **Generalization** - "30s" instead of "age 32"

Before & After

Before:
"Write an email to John Smith (john.smith@company.com) about his $50,000 salary review"

After:
"Write a salary review email template for a mid-level employee"

Quick Check: Before sending any prompt, ask: "Does this contain names, numbers, or details that could identify a real person or company?"

H - Human Oversight

When Humans Must Decide

- Hiring and firing decisions
- Credit and loan approvals
- Medical diagnoses
- Legal judgments
- Customer complaints escalation
- Disciplinary actions

Human-in-the-Loop Levels

- **Level 1:** AI suggests, human approves
- **Level 2:** AI acts, human reviews
- **Level 3:** AI acts, human audits
- **Level 4:** Fully automated (low-risk only)

EU AI Act Requirement: High-risk AI systems must include human oversight capabilities. Decisions significantly affecting individuals cannot be fully automated.

I - Informed Consent

What People Need to Know

- That AI is being used
- What data is collected
- How their data will be processed
- Who has access to outputs
- Their rights to opt out
- How to request deletion

Transparency Examples

Good: "This chatbot uses AI to assist you. Your conversation may be reviewed to improve service."

Bad: Hidden AI use with no disclosure

GDPR Article 22: Individuals have the right not to be subject to decisions based solely on automated processing that significantly affects them.

E - Evaluate Ethics

The Ethics Checklist

⚖

Fairness

Does the AI treat all groups equitably?

Transparency

Can we explain how it works?

Safety

What harm could this cause?

Quick Ethics Test

Before deploying AI, ask:

1. Would I be comfortable if this appeared in the news?
2. Would I want this done to me or my family?
3. Does this align with our stated values?

Understanding AI Bias

Sources of Bias

- **Training data** - Historical discrimination baked in
- **Sampling bias** - Unrepresentative datasets
- **Label bias** - Human prejudices in annotations
- **Feedback loops** - AI reinforces existing patterns

Real Examples

- Hiring AI that favored male candidates
- Facial recognition less accurate for darker skin
- Credit scoring discriminating by zip code
- Healthcare AI under-treating minority groups

Key Point: AI doesn't create bias - it reflects and amplifies biases in data and society. The responsibility to check for bias is human.

L - Legal Compliance

GDPR Essentials

- **Lawful basis** - Have a legal reason to process
- **Data minimization** - Collect only what's needed
- **Storage limitation** - Delete when no longer needed
- **Subject rights** - Access, rectification, erasure
- **DPIAs** - Assess high-risk processing

EU AI Act Risk Levels

- **Unacceptable** - Banned (social scoring)
- **High-risk** - Strict requirements (HR, credit)
- **Limited risk** - Transparency obligations
- Minimal risk - No restrictions

Action Item: Check if your AI use case is classified as "high-risk" under the EU AI Act. If so, additional compliance steps are required.

D - Document Everything

What to Document

- AI systems used and their purposes
- Data inputs and sources
- Decision-making processes
- Human review procedures
- Risk assessments conducted
- Incidents and remediation

Why Document?

- **Compliance** - Required by GDPR/AI Act
- **Accountability** - Show you did due diligence
- **Auditability** - Enable reviews and investigations
- **Learning** - Improve processes over time

AI Usage Log Template

Date | AI Tool | Purpose | Data Types | Human Review | Outcome

Scenario: Customer Support AI

Situation: Your company wants to use AI to automatically respond to customer emails about order status.

Risks

- Customer emails contain PII
- Order details are sensitive
- AI might give wrong information
- No human verification

SHIELD Application

- **S:** Use order ID only, not full details
- **H:** Human reviews before send
- **I:** Disclose AI use in signature
- **E:** Test for bias in responses
- **L:** Process under contract basis
- **D:** Log all AI-generated replies

Scenario: AI Resume Screening

Situation: HR wants to use AI to screen 500 job applications and shortlist the top 50.

High-Risk Concerns

- EU AI Act: Employment = High-risk
- Potential for discrimination
- Significant impact on individuals
- Training data biases

Required Safeguards

- Human review of all rejections
- Regular bias audits
- Transparency to candidates
- Right to human review on request
- Detailed documentation (DPIA)

Warning: AI hiring tools are under intense regulatory scrutiny. Several have been banned or penalized for discrimination.

Scenario: Using ChatGPT at Work

Situation: An employee wants to use ChatGPT to help write a client proposal.

What Could Go Wrong

- Client name and project details shared
- Pricing and strategy exposed
- Proprietary methods revealed
- Data retained by OpenAI

Safe Approach

- Use "Client X" instead of real name
- Describe project generically
- Don't include pricing/numbers
- Use enterprise version with data controls
- Or: Use on-premise AI solution

Best Practice: Create a company-approved AI use policy with clear guidelines on what can and cannot be shared.

Enterprise AI vs Consumer AI

Feature	Consumer (Free ChatGPT)	Enterprise (ChatGPT Enterprise)
Data Training	**May use your data**	**Never trains on your data**
Data Retention	30+ days	Configurable / zero retention
Compliance	Basic	SOC 2, GDPR, BAA available
Admin Controls	None	Full visibility and control
Audit Logs	None	Complete audit trail

Recommendation: For business use with any sensitive data, use enterprise-grade AI tools with proper data protection agreements.

Building an AI Use Policy

Essential Elements

1. Approved AI tools list
2. Data classification rules
3. Prohibited use cases
4. Approval workflows
5. Training requirements
6. Incident reporting

Sample Policy Statement

```
"Employees may use approved AI tools for work tasks.
Confidential data (as defined in our data classification
policy) must never be input into AI systems. All AI-generated
content must be reviewed by a human before external use."
```

Start Simple: A basic policy today is better than a perfect policy never. Iterate and improve over time.

Creating an AI Ethics Culture

Educate

- Regular training sessions
- Share case studies
- Update on regulations

Empower

- Safe reporting channels
- Ethics champions
- No blame culture

Embed

- Include in processes
- Make it part of reviews
- Lead by example

"Ethics isn't a department - it's everyone's responsibility."

5 Questions Before Using AI

1. **What data am I sharing?**
   Could any of it identify a person or reveal secrets?
2. **Who might be affected?**
   Customers, employees, the public?
3. **What could go wrong?**
   Bias, errors, privacy breaches?
4. **Is human review needed?**
   Should a person check this before acting?
5. **Can I justify this decision?**
   Would I be comfortable explaining it publicly?

Safe Prompt Engineering

Risky Prompt

```
"Here's our customer list with names and emails. Write
personalized marketing emails for each:

John Smith - john@email.com - bought Product A
Mary Jones - mary@email.com - bought Product B
..."
```

Safe Prompt

```
"Create a marketing email template for customers who
purchased [PRODUCT_TYPE]. Include:
- Personalization placeholder for [NAME]
- Reference to their purchase
- Upsell for related products

Do not use any real customer data."
```

Template Approach: Create AI-generated templates, then use your own secure systems to fill in real data.

Data Anonymization Techniques

Technique	Original	Anonymized
Masking	john.smith@company.com	j***@***.com
Pseudonymization	John Smith, Dublin	User_A4X7, City_1
Generalization	Age: 34, Salary: $52,340	Age: 30-40, Salary: $50-60K
Synthetic Data	Real customer records	Statistically similar fake data
Aggregation	Individual transactions	"Average order: $45"

Rule: If you can identify an individual from the anonymized data (even combined with other info), it's not truly anonymous.

Interactive Demo

Data Safety Assessment Tool

Try our interactive tool to assess the safety of your AI use cases.

- Input your use case scenario
- Get a SHIELD framework analysis
- Identify risks and mitigations
- Generate a compliance checklist

Access the Demo

Navigate to the **Demo** section in the course materials

Use it to evaluate your own workplace AI scenarios

Take-Home Exercise

AI Safety Audit

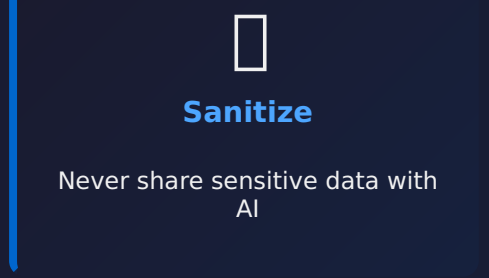
1. **Inventory** - List all AI tools used in your team/department
2. **Classify** - Identify what data each tool processes
3. **Assess** - Apply the SHIELD framework to each
4. **Document** - Note any gaps or concerns
5. **Recommend** - Propose improvements to your manager

Bonus Challenge: Draft a one-page AI use policy for your team based on today's learnings.

Key Takeaways

SHIELD

Apply the framework to every AI use case

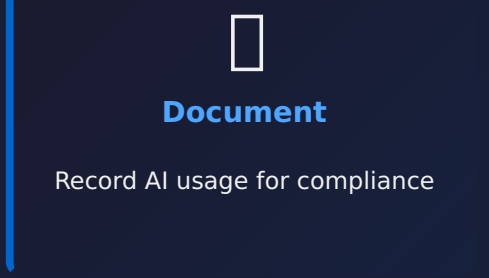


Sanitize

Never share sensitive data with AI

Human Loop

Keep humans in control of decisions



Document

Record AI usage for compliance

"With great AI power comes great responsibility."

Resources

Regulations & Guidance

- EU AI Act Official Text
- GDPR Guidelines (edpb.europa.eu)
- ICO AI Guidance (ico.org.uk)
- NIST AI Risk Framework

Tools & Checklists

- Microsoft Responsible AI Toolkit
- Google AI Principles
- AI Ethics Canvas
- Data Privacy Impact Assessment Templates

Stay Updated: AI regulations are evolving rapidly. Follow your data protection authority for updates.

Questions & Discussion

What data safety challenges have you encountered?

Share

Your AI use cases and concerns

Ask

About specific scenarios

Discuss

Industry-specific challenges

Thank You!

Data Safety & Ethics: Use AI Responsibly

ITAG Skillnet AI Advantage

Remember the SHIELD Framework

Sanitize - **H**uman Oversight - **I**nformed Consent - **E**valuate Ethics - **L**egal Compliance - **D**ocument

Access course materials and the interactive demo through the course portal