CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Record this session

AI Lunch & Learn

Your Journey from AI Curious to AI Confident

Welcome to Lesson 1.4: AI Ethics and Responsible Use

Today we'll explore:

- Identifying and mitigating AI bias
- Protecting privacy and confidential information
- Respecting intellectual property
- Creating your personal ethical guidelines

6th November, 2025

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

How We're Learning Today

Ethics requires reflection and discussion, not just practice.

- **Present** - Main ethical issues and considerations
- **Reflect** -Think through questions and scenarios
- **Discuss** - Share perspectives

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Quick Recap

What We've Learned So Far

Lesson 1.1:

- AI recognises patterns, not rules
- It's a tool for augmentation, not replacement
- AI makes mistakes - always verify outputs

Lesson 1.2:

- LLMs predict patterns, not facts
- Tokens and context windows limit what AI "remembers"
- Hallucinations happen - verify specific claims

Lesson 1.3:

- AI excels at pattern-based, repetitive, data-heavy tasks
- AI struggles with judgement, context, and human touch
- Effective collaboration: Define → Draft → Refine → Validate → Deliver

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Why Ethics Matters Now

The Stakes Are Real

With great capability comes great responsibility. As we integrate AI into our work, we face new ethical questions:

Examples:

- A hiring algorithm consistently rejected qualified women candidates
- AI-generated content published without disclosure
- Customer data processed by AI without proper privacy controls
- Textbook manuscripts with AI-generated citations that don't exist

The Common Thread: All of these can happen because someone didn't stop to ask: "Just because I *can* do this, *should* I?"

Our Goal Today: Build your ethical framework for responsible AI use at work

The Four Pillars of AI Ethics

BIAS

Ensuring fairness and avoiding discrimination

Challenge:

AI learns from biased data

PRIVACY

Protecting confidential information

Challenge:

AI processes everything you give it

INTEGRITY

Respecting intellectual property and authenticity

Challenge:

AI blurs lines of authorship

ACCOUNTABILITY

Taking responsibility for AI-assisted work

Challenge:

Easy to blame the tool

Understanding AI Bias

When AI systems produce **systematically unfair outcomes** for certain groups of people.

Why does it happen?

Training Data Bias

AI learns from historical data that contains human biases

Design Bias

How the AI is built can embed certain assumptions

Use Bias

How we use AI can amplify existing inequalities

Examples of AI Bias

Examples

Facial recognition that works better for lighter skin tones (trained on non-diverse datasets)

Language models that associate certain professions with specific genders (nurse = she, engineer = he)

Resume screening that favours certain educational backgrounds or name patterns

AI doesn't create new biases. Rather it **learns and amplifies** the biases in its training data and our society.

Bias in Your Daily Work

Writing Assistance

- Gendered language assumptions ("Ask the client what he thinks...")
- Cultural assumptions in tone and formality
- Stereotypical examples and scenarios

Data Analysis

- Missing representation of certain groups
- Assumptions about "normal" or "standard"

Content Generation

- Stereotypical character or persona creation
- Cultural insensitivity in examples

Decision Support

- Recommendations based on biased historical patterns
- Missing context about disadvantaged groups
- Reinforcing existing inequalities

Mitigating Bias - Practical Steps

1 Awareness

- Recognise that bias exists in all AI systems
- Question outputs that seem to reinforce stereotypes

2 Diversify Your Prompts

- Explicitly request diverse perspectives: *"Give me examples from different cultural contexts"*
- Challenge default assumptions: *"Avoid gender assumptions in this text"*

3 Review Critically

- Check AI output for stereotypical language
- Look for missing perspectives or groups
- Test: Would this be fair to everyone?

4 Human Oversight

- Never use AI alone for decisions affecting people
- Involve diverse perspectives in reviewing AI output
- When in doubt, consult others

Best Practice: Treat AI like a colleague who grew up in a limited environment - helpful but needs your broader perspective

What Happens to Your Data?

Critical Understanding

When you put information into an AI chatbot, you're sending it to that company's servers. What gets stored and how it's used varies:

ChatGPT
(OpenAI)

- **Free Version**
  Conversations may be used for training
- **Team/Enterprise**
  Conversations not used for training

Data stored on OpenAI servers

Claude
(Anthropic)

- **Free Version**
  Conversations may be reviewed for safety and training
- **Pro/Team**
  Conversations used for training (opt-out available)

Data stored on Anthropic servers

Microsoft Copilot
(Microsoft)

- **Enterprise Users**
  Commercial data protection promises
- **License Dependent**
  Data handling varies by license type

Integrated with Microsoft 365 ecosystem

Risk Categories

Data Sharing Risk Levels

NEVER SHARE

- Personal data (names, emails, phone numbers, addresses)
- Financial information (bank details, salary information)
- Passwords or access credentials
- Medical or health information
- Legal documents with identifying information

HIGH RISK

Extreme Caution

- Confidential business strategies
- Unreleased product information
- Customer lists or contact databases
- IP such as textbooks

MEDIUM RISK

Redact First

- Internal reports (remove names, specifics)
- Customer communications (anonymise)
- Meeting notes (remove identifying details)
- Draft documents (strip confidential sections)

SAFE

- Public information
- General knowledge questions
- Writing style assistance
- Learning and skill development

Golden Rule

If you wouldn't share it publicly, don't share it with AI, or thoroughly anonymise first.

Cambridge's AI Use Guidance

Data type	Chatbots such as ChatGPT	Cambridge-endorsed Chatbot MS Copilot Lite
Personal data	No	Yes
Anonymised personal data	Yes	Yes
Cambridge confidential corporate information	No	Yes
Cambridge IP (i.e. any copyright-protected research, course, assessment or other IP even if publicly available)	No	Yes
High stakes assessment content	No	No
3rd Party IP (including 3rd Party IP forming part of our resources) even if publicly available, e.g. images, text extracts	No	Yes

For more: https://cambridgeorg.sharepoint.com/sites/Connect/SitePages/Guidance-for-using-Generative-AI-at-Cambridge.aspx

Intellectual Property

AI Training Data

- AI models trained on copyrighted material (books, articles, code, images)
- Legal cases ongoing about whether this constitutes fair use

AI Output Ownership

- In most jurisdictions, purely AI-generated content may not be copyrightable
- Human-edited AI content typically belongs to the human editor
- Terms of service vary by AI platform (read them!)

Your Responsibilities

- You can't claim sole authorship of unedited AI content
- You must disclose AI use where required
- You're responsible if AI output infringes someone's copyright

The Grey Area

> If AI generates text that closely resembles copyrighted material, publishing it could constitute infringement, **even if you didn't know**.

Legal Landscape is Evolving

Copyright and intellectual property laws around AI are still being developed. Stay informed about changes in your jurisdiction and industry-specific regulations.

Accountability

Fundamental Principle

You are accountable for everything you create or send, regardless of whether AI helped produce it.

What This Means

You Can't Say

- "The AI made a mistake"
- "I just used what ChatGPT gave me"
- "How was I supposed to know it was wrong?"

You Must Say

- "I made an error in my work"
- "I should have verified that information"
- "I take responsibility for this output"

The Accountability Checklist

- [ ] I have reviewed every claim and fact
- [ ] I have verified sources and citations
- [ ] I have checked for bias and fairness
- [ ] I have ensured no confidential information included
- [ ] I have disclosed AI use where required
- [ ] I would be comfortable defending this work
- [ ] This represents my authentic thinking

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Reflection

The Confidential Report

You're summarising a sensitive internal report about product strategy and resourcing for FY27.

Using AI would save 2 hours, but the report contains names, roles and confidential strategy.

What do you do?

What are the risks?

How could you use AI safely, if at all?

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Reflection

The Education Ministry Review

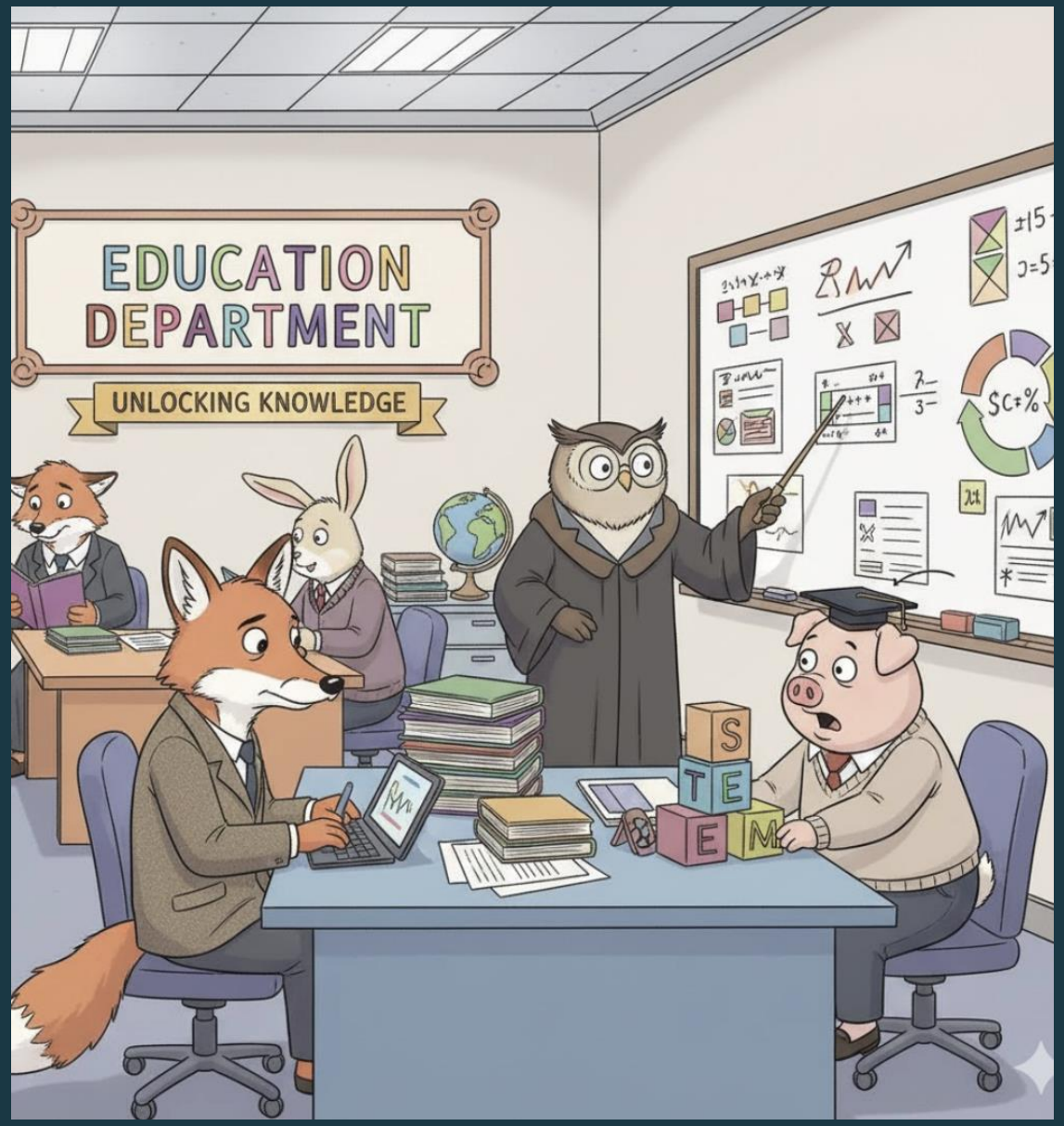
An education ministry pays you for your expertise to write a strategic review document.

You use AI to generate the first draft, then edit it. The client assumes it's entirely your original work.

Do you disclose AI use?

Why or why not?

What if the contract doesn't mention AI?



CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Reflection

The Time Pressure

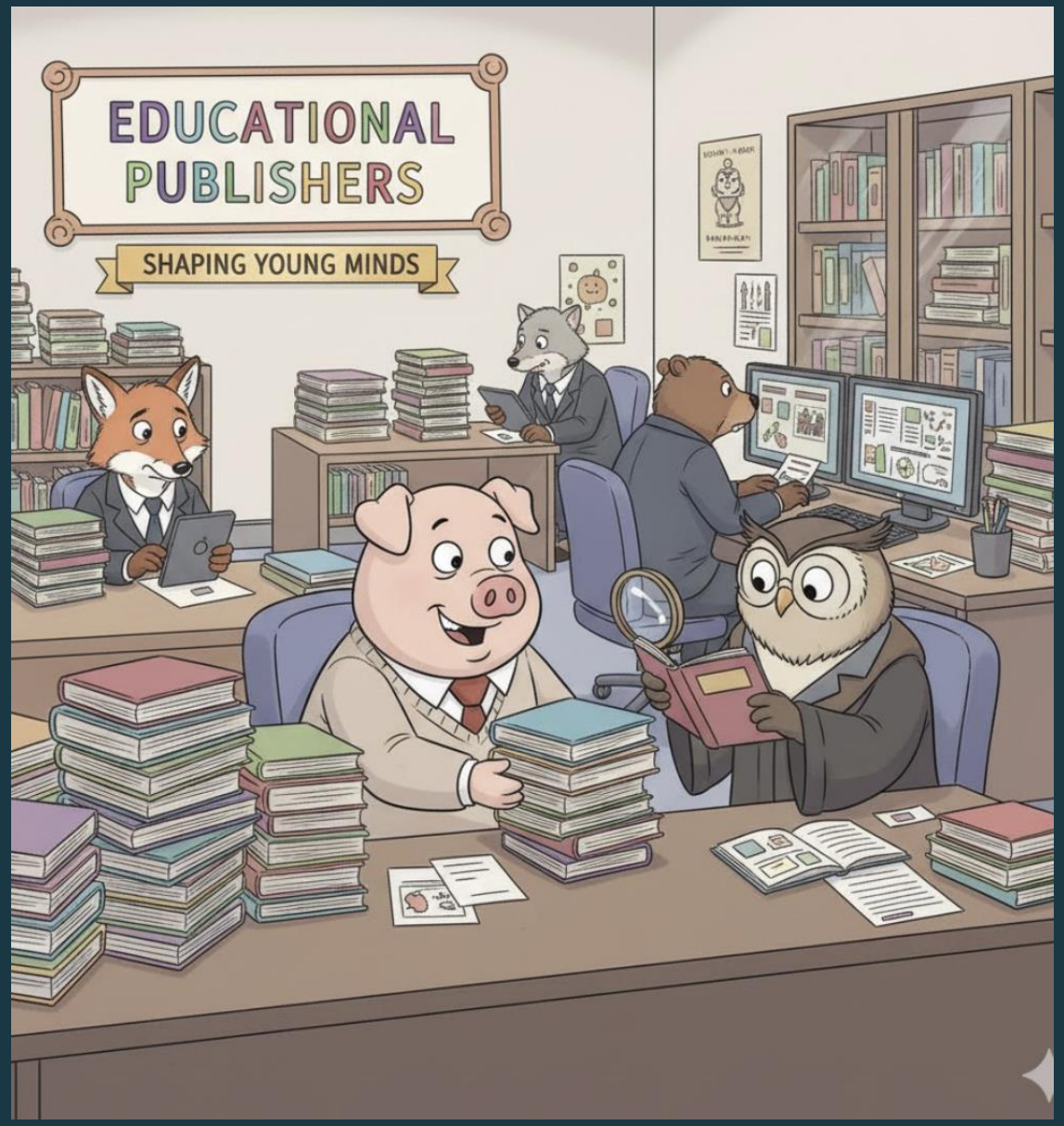
You're on a deadline for a submission and AI generates perfect content quickly.

You don't have time to fact-check every claim, but it "looks right" and reads like textbook should.

What's your responsibility?

What could go wrong?

What would you actually do?



The Ethical AI User

1

Transparency

Be honest about when and how you use AI

2

Verification

Never publish AI output without careful review

3

Privacy First

Protect confidential information religiously

4

Bias Awareness

Actively look for and mitigate unfair patterns

5

Authentic Voice

Use AI to amplify, not replace, your thinking

6

Full Accountability

Own everything you produce, regardless of how it was created

7

Continuous Learning

Stay informed as AI ethics evolve

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Lesson Recap

AI Bias:

- AI learns and amplifies biases from training data
- Actively check for stereotypes and missing perspectives
- Use diverse prompts and critical review

Privacy:

- Never share personal data, credentials, or confidential information
- Anonymise thoroughly before using AI on sensitive content
- Know your organisation's AI usage policy

Intellectual Property:

- Be transparent about AI use where required
- You're responsible if AI output infringes copyright

Accountability:

- You own all AI-assisted output
- Verification is mandatory, not optional

Next week:

Module 2 Begins Lesson 2.1: Fundamentals of Prompting

We'll explore:

- Anatomy of effective prompts
- The importance of context and clarity
- Role-playing and perspective setting
- Common prompting mistakes to avoid

We're moving from foundations to mastery. Get ready to level up your AI skills!

www.cambridge.org