LESSON 03

AI Ethics, Bias, Privacy, Environment & Sovereignty

The deeper responsibilities of AI and what libraries must understand, question, and advocate for

⏱ 60–90 min | All library professionals

Why These Questions Cannot Wait

AI adoption is accelerating faster than ethical reflection — libraries must be ahead of this curve

Ethical failures happen now
Facial recognition systems have wrongly identified innocent people. Hiring algorithms have systematically discriminated against women. These are not hypothetical futures.

Bias is structural, not accidental
AI bias flows from training data that reflects historical injustice. It does not fix itself without deliberate intervention.

Patron data is at risk
Every AI system that touches patron queries, browsing, or reading history raises serious confidentiality questions that libraries are legally and ethically obligated to address.

The environmental cost is real
Training a single large AI model can emit as much CO_2 as five cars over their lifetimes. Libraries committed to sustainability must account for this.

Who controls AI controls knowledge
A small number of corporations control the AI systems now shaping how information is found, filtered, and presented globally. This is a sovereignty question.

AI Ethics

What is right? What is harmful? Who decides?

Ethics are not a constraint on good AI; they are a prerequisite for it.

Major AI Ethics Frameworks

Global institutions have produced frameworks — librarians should know and apply them

UNESCO (2021) *Human dignity & inclusion*

The first global normative instrument on AI ethics. Emphasises human rights, gender equality, environmental sustainability, and peaceful societies. Explicitly covers AI in education and information.

EU AI Act (2024) *Risk-based regulation*

Classifies AI systems by risk level (unacceptable → high → limited → minimal). Libraries using AI for patron profiling or employment decisions may fall under high-risk categories.

IFLA (2020) *Access, privacy, trust*

International Federation of Library Associations statement emphasising open access, privacy, freedom from surveillance, and the right to access information without algorithmic mediation.

OECD (2019) *Inclusive growth & accountability*

Five principles: inclusive growth; human-centred values; transparency; robustness; accountability. Adopted by G20 nations and referenced in most national AI strategies.

How AI Causes Harm: A Taxonomy

Understanding the different routes through which AI systems cause real-world harm

Harm Type	Real-World Example	Library Relevance
Allocative Harm	AI denies someone a loan, job, or housing benefit based on biased prediction.	AI that denies patron services or flags accounts based on demographic patterns
Representational Harm	AI associates certain groups with negative traits ('criminal' images biased by race).	Catalogue discovery AI that underrepresents literature from the Global South.
Quality-of-Service Harm	Voice recognition works worse for women and non-native speakers.	AI reference tools that perform better for English speakers than others.
Interpersonal Harm	Deepfakes used to harass, defame, or manipulate real people.	Patrons targeted by AI-generated disinformation based on library-linked data.
Epistemic Harm	AI confidently provides false information, eroding trust in facts.	Patrons citing generated hallucinations in academic or civic research.

Algorithmic Bias

Bias is not a bug. It is often a feature of the data.

If the world is unjust, and AI learns from the world, AI learns injustice.

Where Bias Enters AI Systems

Bias accumulates at every stage of the AI development pipeline — not just in the data

Data Collection	Under-representation of minority groups; historical discrimination encoded in records; scraped internet data skewing toward wealthy, English-speaking populations.
Data Labelling	Human annotators bring cultural assumptions. Labels reflect the worldview of those paid to annotate usually not the communities being studied.
Model Design	Optimisation metrics (e.g. accuracy on majority groups) may mask poor performance on minorities. Architecture choices embed assumptions.
Deployment Context	A tool trained in one context deployed in another inherits mismatches. A medical AI trained on US patient data fails in Sub-Saharan Africa.
Feedback Loops	If biased outputs influence future data (e.g. biased hiring AI → biased workforce data → feeds next AI), bias compounds over time.

Algorithmic Bias

« A group of golfers »

Algorithmic Bias

Algorithmic Bias

Algorithmic Bias

Garbage in, garbage out (GIGO)

Faulty input data leads to faulty outputs

Algorithmic Bias

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

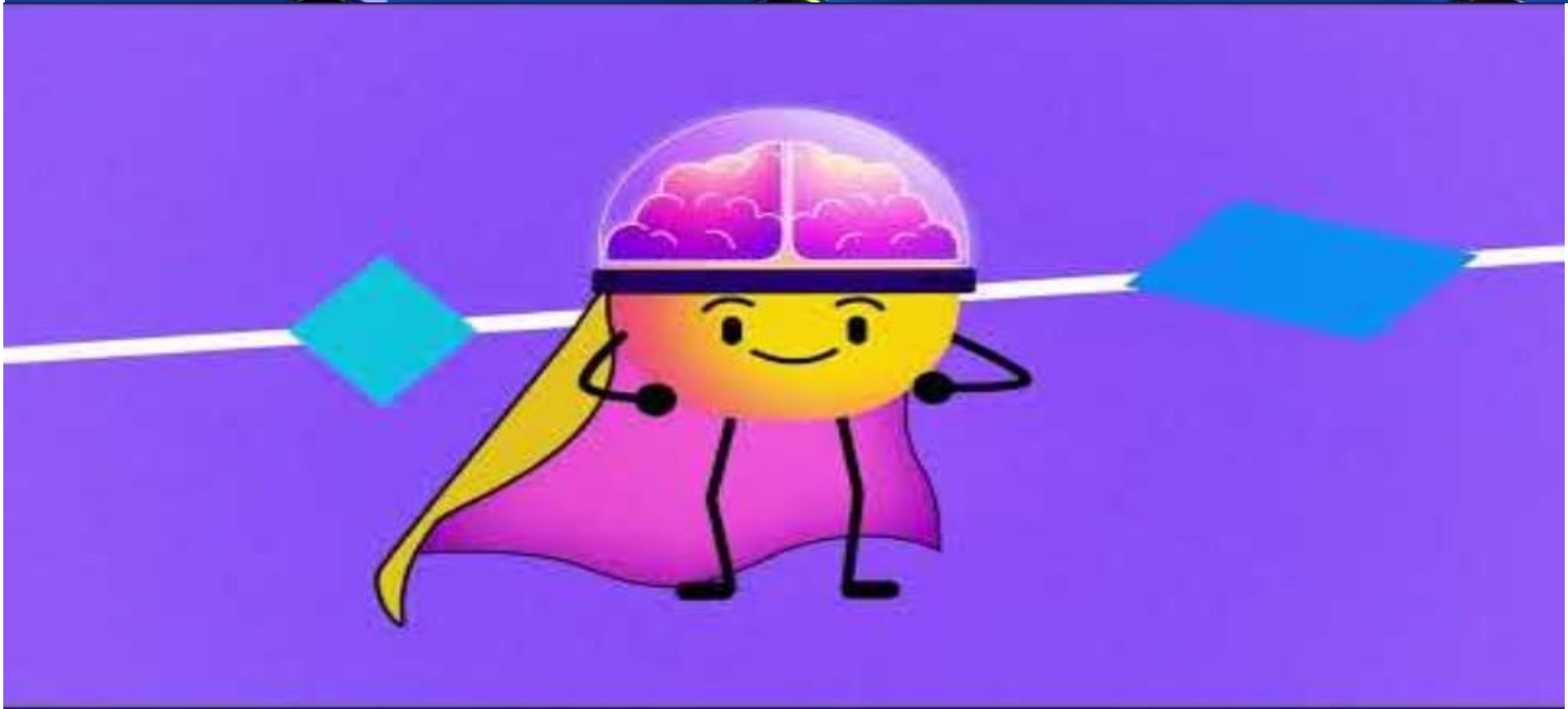
SCORING OF WELFARE BENEFICIARIES: THE INDECENCY OF CAF'S ALGORITHM NOW UNDENIABLE

27 November 2023

Algorithmic Bias

- Amazon's AI recruitment was found to be biased against women: trained on historical tech industry data, which was predominantly male.
- Examples from law enforcement reveal error rates up to 34% higher for darker-skinned faces.
- Medical algorithms have shown bias in predicting treatment needs for Black patients.
- Housing and employment ads can be algorithmically steered based on demographic characteristics.

Algorithmic Bias



Data Privacy & Surveillance

'If you are not paying for the product, you are the product.’

Library patron data is not a resource. It is a right to be protected.

How AI Systems Collect & Use Patron Data

The data flows are often invisible, but their consequences are not

The Data Flow Chain

1. Patron submits a query via AI-powered chat or search
2. Query is processed by a third-party AI API (often in another country)
3. The vendor logs the query for model improvement
4. Usage patterns are aggregated and potentially sold
5. Profiles emerge linking borrowing, searching, and asking behaviour
6. These profiles may be accessible to authorities under certain legal frameworks

Existing Protections

- GDPR (EU) right to access, erasure, data minimisation
- Many national library confidentiality laws extend to digital interactions
- IFLA Privacy Framework for library systems

Key Risks

- Third-country data transfers may lack equivalent privacy protection
- AI tool may grant vendors broad rights to query data
- Patron queries may reveal sensitive information (health, legal, political)

How AI Systems Collect & Use Patron Data

Patron confidentiality is a legal and ethical foundation AI does not waive it

CORE PRINCIPLE: The library's commitment to patron privacy does not end at the human–AI boundary. Every AI system handling patron data must meet the same standard as any other library system.

1	**Data privacy impact assessment** : Before deploying any AI tool that touches patron data, complete a documented DPIA. Assess what data is collected, where it flows, and what risks exist.
2	**Apply Data Minimization** : AI tools should be configured to collect the minimum data necessary. Personally identifiable information should be excluded from AI prompts wherever possible
3	**Review Vendor Privacy Policies** : Read the full privacy policy and terms of service of every AI vendor. Key questions: Does the vendor store queries? For how long? Can data be used to train future
4	**Inform Patrons :**When AI tools are used in patron-facing services, clear notice must be provided. Patrons should be able to opt out of AI-mediated services and access equivalent human alternatives.
5	**Train All Staff on AI Privacy:** Every staff member using AI tools must understand what they may and may not enter. Create an approved data types list and a prohibited data types list.

Activity

Cookie tracking & privacy controls

See Facilitator guide

Environmental Impact

AI has a carbon footprint and libraries committed to sustainability must account for it.

The energy cost of 'intelligence' is not invisible. It is just hidden in data centers.

Environmental Impact

Scale matters the resource demands of modern AI systems are extraordinary

626,000

kg CO_2 equiv.

Estimated emissions to train a single large language model from scratch (MIT study)

10×

more energy

A ChatGPT query uses approximately 10× more energy than a standard Google search

6.4 m

cubic meters water

Estimated water usage by Microsoft's AI data centres in 2022 for cooling systems

2–4%

of global CO_2

ICT sector's current share of global carbon emissions growing rapidly with AI demand

AI Sovereignty & Power

Who builds AI? Who owns it? Whose values does it reflect?

Control over AI infrastructure is control over the future of knowledge itself.

The AI Concentration Problem

Five companies OpenAI (Microsoft), Google DeepMind, Anthropic (Amazon), Meta AI, and xAI currently dominate the most capable AI foundation models. Their decisions about training data, access, pricing, content moderation, and shutdown shape what billions of people can and cannot know.

Compute Concentration	Training large AI models requires vast computing resources owned by a handful of cloud providers (AWS, Azure, GCP)
Data Concentration	The internet data used to train most LLMs was primarily generated in wealthy, English-speaking countries.
Labour Concentration	Content moderation and data labelling for AI is largely performed by low-paid workers in the Global South, often with minimal labour protections.
Value Concentration	Decisions about what AI systems will and will not say, show, or recommend are made by corporate teams in a small number of countries
Regulatory Asymmetry	AI regulation is developing fastest in the EU. Many other regions have limited oversight of AI systems affecting their populations.

AI Sovereignty: What It Means for Libraries

Collection Sovereignty: When AI systems curate, filter, or rank collections on behalf of libraries, decisions about what is visible and what is buried pass from librarians to algorithms designed for commercial, not civic, objectives.

→ Maintain human oversight of AI-influenced collection visibility. Audit rankings.

Search Sovereignty: AI-powered discovery changes what patrons find without changing what exists. A patron who sees only what an AI surfaces has a diminished information environment — shaped by the AI's training, not their actual needs.

→ Offer non-AI discovery pathways. Teach patrons to use controlled vocabularies.

Knowledge Sovereignty: AI systems built primarily on Global North data may present one cultural tradition's knowledge as universal. Libraries serving diverse communities must critically evaluate what AI knows and what it erases.

→ Center underrepresented knowledge traditions. Resist AI that flattens diversity.

Legal Sovereignty: Data processed by AI vendors may be subject to foreign surveillance laws. Libraries must ensure cross-border AI use complies with national privacy regulations.

→ Seek legal opinion on data residency obligations before AI procurement.

Group Activity: The Ethical Audit

Groups of 4–6 | 25 minutes | Choose one AI tool your library uses or is considering

⚖️ Ethics

Does this tool have a published ethical framework?
Who authored it?
What are the documented failure modes and harms?

🔵 Bias

Has the tool been independently tested for demographic bias?
What languages and communities are underrepresented in its training data?

🔒 Privacy

What data is collected, stored, and for how long?
Is the tool GDPR compliant? Is there documentation?

🌿 Environment

Does the vendor publish energy consumption or emissions data?
Are data centers powered by renewable energy?

🏛️ Sovereignty

In which countries is the data processed?
Does the vendor's content policy reflect our community's values and rights?