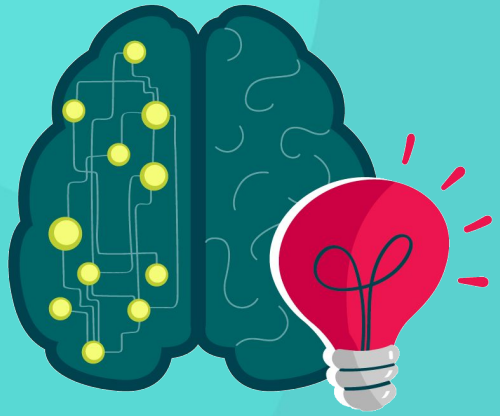
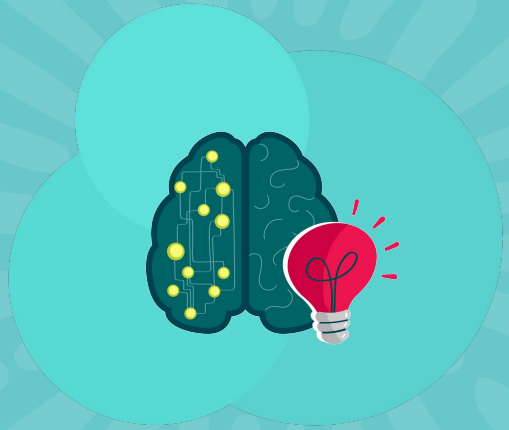


NGO management & AI Ethics in Operation

Session #3:
Ethics and bias in AI decision-making

2025

AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT
Funded by the European Union
forus CONNECT SUPPORT INFLUENCE
INNOVUS



Would you trust a decision made by a machine if it impacts human lives?

AI can help you decide who to give a scholarship or humanitarian aid to, or even prioritize for medical intervention... but ***how do you know that decision is fair?***

NGO management & AI ethics in operation

Introduction to AI for NGOs

Analyze the potential of artificial intelligence in NGO decision-making and operations, evaluating real-world applications and designing an initial mapping of opportunities for use aligned with your organizational context.

1

2

AI Tools for Fundraising and Program Management

Identify and evaluate artificial intelligence tools applied to fundraising and program management, recognizing their advantages and limitations, and designing a preliminary integration strategy aligned with your organizations' operations.

AI Ethics and Bias in Decision Making

Analyze the impact of algorithmic bias on nonprofit work, apply relevant ethical frameworks, and formulate basic principles for responsible and equitable adoption of artificial intelligence in your organizations.

3

Objectives

Ethics and bias in AI decision-making

- Analyze examples of algorithmic bias and its implications for nonprofit work.
- **Apply ethical frameworks** for the responsible use of AI in social impact contexts.
- **Formulate a set of guiding principles** or basic protocols for the ethical adoption of AI in NGOs.

Why does AI ethics matter in NGOs?

AI offers **opportunities** (efficiency, reach) but carries **ethical risks**.

Bias in algorithms can **perpetuate** existing **injustices.**

NGOs must **protect values** (equity, transparency) and the **trust** of their communities.

What is algorithmic bias?

Partial or unfair result produced by an AI system

- **Sources of bias:** Biased training data, poorly designed algorithms, or built-in human biases.

- **Effect:** AI reflects and amplifies existing human biases, affecting decisions.

Examples of algorithmic bias

(real cases)

Health

A diagnostic AI system showed lower accuracy for patients of African descent than for white patients (unrepresentative clinical data).

Recruitment

Amazon scrapped its recruitment algorithm after discovering that it penalized women's resumes, replicating gender biases in technology.

Criminal Justice

Predictive policing algorithms reinforce racial profiling by relying on historical arrest data.

Advertising

A study found that Google Ads showed ads for higher-paying jobs more to men than to women *(among other cases).*

Implications of algorithmic bias in the nonprofit sector

- **Unintentional discrimination:** Biased AI can deny services or resources to vulnerable groups, contradicting the NGO's social mission.
- **Loss of trust and reputation:** Beneficiaries and donors may lose trust if they perceive unfair decisions.
- **Legal and ethical risks:** Possible claims of discrimination (e.g., bias in hiring may violate labor laws) and ethical dilemmas for the organization.
- **Impact on effectiveness: Biased decisions ->** poor program targeting, perpetuation of inequalities we sought to reduce.

Let's think about it!

Are there any examples close to home where a digital tool has unintentionally discriminated? How would our organization react if something like this happened?

Ethical dilemmas with reactions

What would you do?

An AI recommends excluding people over 60 from a program because "historically, they participate little." *Would you apply this recommendation?*

👍 Yes, I trust the data.

🤔 It depends, I would review it further before deciding.

❌ No, that's discrimination.

What would you do?

Your team wants to use free AI that analyzes emotions based on the faces of participants in videos. You don't know how that model was trained. *Would you use it?*

✅ Yes, if it saves time, let's try it.

⚠️ Only with informed consent.

🛑 No, it's risky without transparency.

What would you do?

Your organization uses AI to prioritize communities with "greater potential for impact." This leaves out small or isolated populations. *Would you continue with this model?*

📈 Yes, maximizing impact is key.

🤝 I would seek balance with human criteria.

🌍 No, that perpetuates exclusion.

What would you do?

An AI system predicts which children will drop out of school. Should this prediction be used to automatically allocate scholarships?

🧠 Yes, it improves efficiency.

💡 Only if accompanied by human intervention.

🚫 No, it can unfairly label people.

Detecting hidden biases in AI decisions

Learning activity *-10 minutes-*

Detecting hidden biases in AI decisions

Case study:

An NGO implements AI to screen job applications. Months later, it notices that almost all of the shortlisted candidates are men, even though the organization promotes gender equality.

Group discussion:

What bias could be at play? How would you identify it? What changes would you make to mitigate it?

Ethical frameworks and international guidelines for responsible AI

UNESCO (2021)

Global Recommendation on AI Ethics – First global standard. Emphasizes human rights, non-discrimination, transparency, human oversight, and “do no harm.”

OECD (2019/2024)

AI Principles – Inclusion and well-being, fairness, respect for human rights, transparency, security, and accountability. Adopted by over 60 countries.

Legislation (GDPR, EU)

Data protection regulations prohibit fully automated decisions without the possibility of human review if they significantly affect individuals. The EU is moving forward with the first comprehensive AI law.

We can base our internal policies on solid references such as UNESCO and the OECD, and we must be aware of the laws in force.

Applying AI ethics in contexts of social impact

- **Impact and risk assessment:** Before implementing AI, analyze possible effects on the community (Who could be harmed?) *– E.g., use ethical checklists or impact assessments (EIA)*
- **Inclusive data and design:** Ensure data **is representative** of the populations served (avoid bias at source) *– Include diverse actors in the design and testing of the tool.*
- **Transparency and explainability:** Communicate how AI works and why it makes certain decisions, in accessible language. *– Allow participants to understand and question results.*

Applying AI ethics in contexts of social impact

- **Continuous human oversight:** Combine AI with human judgment in critical decisions (the machine does not decide alone on matters affecting rights). *– Monitor the algorithm's performance and adjust it in case of deviations.*
- **Training and ethical culture:** Train staff in AI literacy, bias, and privacy. *– Foster a culture where ethical dilemmas are openly reported and discussed.*

Guiding principles for the ethical adoption of AI in NGOs

Proposed principles

1. Equity and inclusion
2. Transparency and explainability
3. Responsibility and accountability
4. Privacy and data security
5. Human oversight
6. Proportionality and "Do no harm"
7. Training and continuous improvement

Conclusions and final thoughts

- Algorithmic biases **are not imaginary:** we have seen that they exist and can cause real harm if left unchecked.
- **Ethics in AI** is indispensable in the social sector to uphold our mission, protect the rights of all people, and foster trust.
- There are **global frameworks and laws** that support us (UNESCO, OECD, GDPR), but there must be an internal commitment from each NGO to turn them into everyday actions.
- When adopting AI, let's think **"morally first**": prioritize people's well-being over mere technical efficiency.

AI used responsibly can amplify our positive impact; ethics and technology must go hand in hand to achieve sustainable social innovation.

Thank you!

www.innovusconsulting.co

Session facilitators: Catalina Valenzuela, Paula Quiñones, and Camilo Forero

Module creator: Carolina Posso