WORKSHOP 1 - REPORT

Ethical Use of Artificial Intelligence in Combating Trafficking in Human Beings

Council of Europe Conference – Malta Presidency

Date: 11 June 2025
Facilitator: Thao Do-Ngoc, PhD Researcher, University of Bath
Workshop Duration: 90 minutes

1. Workshop aim and rationale

As part of the Council of Europe conference *"Empowering Change: Technology and Artificial Intelligence in the Fight Against Human Trafficking"*, a 90-minute workshop titled **"Ethical Use of AI in Combating Trafficking in Human Beings"** was facilitated by Thao Do-Ngoc. The session brought together around 20 participants from diverse sectors, government, civil society, tech companies, and academia, to explore the benefits, risks, and dilemmas of AI deployment in anti-trafficking work, with a particular focus on human rights- based approaches**.**

The session aimed to go beyond technical deployment and interrogate how AI systems in high-stakes contexts such as human trafficking can be designed and governed ethically and inclusively. This framing emerged from ongoing research into AI and human trafficking, and the workshop drew from real-world case studies to make abstract principles tangible. It foregrounded ethical concerns, systemic risks, and power asymmetries often overlooked in technology-driven solutions. The underlying question was not only about how "can AI help combat trafficking?" but **"**how can it be used ethically, responsibly, and without harm to the most vulnerable?"

2. Structure and methodology

The 90-minute session combined critical framing, participatory case study work, and collaborative synthesis, using a structure that enabled participants to reflect on real-world examples through an ethical lens:

- **Framing presentation (15 mins):** Overview of AI technologies used in anti-trafficking, including NLP, facial recognition, and predictive analytics. Emphasis was placed on understanding both the technical logics and their sociopolitical implications. The talk also introduced core principles such as fairness, transparency, accountability, and inclusion, and advocated for embedding human rights and survivor participation into system design.
- **Group case study analysis (30 mins):** Participants joined into groups based on interest in four real-world case studies. Groups engaged with guiding questions designed to unpack assumptions, risks, and rights-based concerns. Case studies include:

    - AI for detecting victims in abuse imagery
    - Trauma support chatbots
    - Sex work post analysis
    - Predictive risk mapping
- **Plenary reflection (15 mins):** Short presentations of key takeaways, followed by synthesis of thematic threads and tensions.
- **Wrap-up (10 mins):** Concluding reflections emphasised the importance of survivor-informed design, safeguards.

3. Key case studies and analytical insights

Case 1: Facial recognition in child abuse image detection

This case explored an AI tool used to scan CSAM (child sexual abuse material) to identify victims. While successful in certain demographics, it showed significant accuracy gaps for racialised individuals due to biased training data.

- **Discussion themes:**
    - **Algorithmic exclusion:** Victims not fitting the training dataset were rendered invisible.
    - **Misclassification harms:** Adults wrongly flagged as minors faced serious reputational and legal consequences.
    - **Opacity and accountability:** Lack of transparency about model performance across demographic groups limited public scrutiny and informed consent.
- **Takeaway:** Even "high-performing" tools can embed structural inequities. "Effectiveness" must be disaggregated and ethically contextualised.

Case 2: Chatbots for trauma support

A chatbot was deployed by an NGO to assist trafficking survivors with PTSD. Though well-intentioned, it raised significant ethical dilemmas.

- **Discussion themes:**
    - **Trauma-informed design** must go beyond soft UX; it requires participatory co-design, clear boundaries, and escalation pathways.
    - **Consent and trust:** Survivors unaware they were talking to AI felt betrayed—highlighting risks of emotional dissonance and secondary trauma.
    - **Data ethics:** Use of anonymised chat data for AI training without survivor control posed risks of exploitation.
- **Takeaway:** Therapeutic interventions via AI cannot replicate human empathy and carry risks of depersonalisation. Survivor-led governance is essential.

Case 3: NLP to detect underage trafficking in sex work ads

This case examined a tool that flags online ads suspected of involving minors. While ostensibly protective, it often misidentifies consensual adult sex workers, particularly racialised and migrant individuals.

- **Discussion themes:**
    - **Language bias:** The model failed to account for the strategic and stylised language used by sex workers.
    - **Digital surveillance as harm:** Flags led to platform bans and policing, resulting in economic loss and further marginalisation.
    - **Consent and due process:** Sex workers lacked appeal mechanisms or notification, raising due process concerns in algorithmic governance.
- **Takeaway:** Protective logic can easily morph into surveillance.

Case 4: Predictive risk mapping

A tool mapped trafficking "hotspots" based on socioeconomic indicators. It was used to inform policing and funding allocations.

- **Discussion themes:**
    - **Profiling vs. prevention:** Marginalised communities were labelled as high-risk, reinforcing surveillance and stigma.
    - **Data as power:** Risk scores were treated as objective "evidence," closing space for community contestation.
    - **False neutrality:** By framing predictions as neutral, the tool masked embedded structural bias.
- **Takeaway:** Predictive tools must be seen as interpretive, not determinative. Acting on probabilities without community governance risks profiling and pre-criminalisation.

3. Key themes from discussion

- **Participatory design is not optional:** Across all cases, participants recognised that those most affected by AI systems are rarely included in their design or governance. Survivors, frontline workers, and marginalised communities should not be positioned as beneficiaries or end users but as co-creators of technologies.
- **Ethical principles must be operationalised:** Values such as fairness or accountability are often vague unless embedded in concrete design mechanisms: appeal processes, bias audits, informed consent protocols, data governance standards, etc.
- **Law is not a substitute for ethics, but ethics alone is not sufficient:** Ethical frameworks are critical but non-binding. Legal protections, such as the right to explanation (GDPR) or safeguards under the proposed AI Act, must be integrated into system design, especially in high-risk contexts like trafficking.
- **AI is not neutral:** Participants highlighted how assumptions about who is at risk, who is a victim, and what counts as exploitation are baked into technical systems. These systems mirror social bias, legal inconsistencies, and funding logics.

4. Recommendations for future workshops and policy development

The discussions and case analyses in this workshop reveal that the ethical use of AI in anti-trafficking cannot rely on ad hoc safeguards or isolated design improvements. It requires a comprehensive, multidisciplinary, and rights-based governance architecture. Below are three interrelated recommendations:

a. Establish a structured governance framework for AI in anti-trafficking

AI interventions must be governed by clear, enforceable frameworks that:

- Define high-risk applications (aligned with the EU AI Act and international human rights law)
- Mandate human rights impact assessments (HRIAs) before deployment
- Include procedures for independent oversight, grievance mechanisms, and accountability
- Promote cross-border cooperation for tools used transnationally

This framework should not be developed only by technical experts or institutional actors but must be co-produced with survivor-led groups, legal advocates, social scientists, and ethicists.

b. Develop a structured, multi-level capacity-building programme

AI's application in trafficking prevention, protection, and prosecution is technically complex and ethically challenged. Many frontline actors expressed a desire for deeper understanding and practical guidance.

I recommend:

- Ongoing training programs beyond single-session workshops, with modules tailored for NGOs, law enforcement, policymakers, and survivor-led groups.
- Use of interactive, participatory formats (e.g. simulations, scenario analysis, case clinics) that leverage participants' lived experience and expertise.
- Training should begin with foundational knowledge of AI systems, what they are, how they work, where bias emerges, so stakeholders can meaningfully engage in co-development and accountability processes. The curricula should be interdisciplinary that combine technical understanding with human rights, trauma-informed care, and digital ethics.

Note: After the session, several participants approached me asking for practical tools and expressed interest in applying them, but were unsure where to start. This indicates a clear need for structured, supported capacity building.

c. Institutionalise participatory and survivor-informed design

Ethical AI development in anti-trafficking must centre the perspectives of survivors of trafficking, frontline NGOs and advocates, and marginalised or over-policed communities.

I recommend:

- Creating dedicated platforms or forums that elevate survivor voices, not only for consultation but for co-creation.
- Incorporating survivor narratives, ethical tensions, and design critiques as standard components of technical design processes.
- Institutional actors (e.g. CoE, national agencies) should provide grants, stipends, and resources to support community and survivor participation.

The powerful sharing from a survivor during the session highlighted the gap between system design and lived experience. Designing “with” rather than “for” would also empower the survivors.

d. Anchor ethical AI practice in legal and normative frameworks

While ethical guidelines offer important principles, the **Council of Europe’s Framework Convention on AI, Human Rights, Democracy and the Rule of Law** provides a legally grounded and rights-oriented roadmap.

I recommend:

- Actively incorporating this framework into workshops, trainings, and product development cycles.
- Using interactive discussion tools (e.g., ethical dilemmas, "what would you do" scenarios) to elicit participant reflections on how to apply these norms in their own contexts.
- Supporting translations of the CoE framework into practical compliance tools for different sectors (e.g., NGOs, justice systems, private companies).

These recommendations highlights that ethical AI in anti-trafficking is not only a technical problem but a political and moral challenge that demands a shift in how we design, govern, and evaluate technological interventions.

5. Concluding reflection

The ethical use of AI in combating human trafficking demands more than technical refinement or good intentions. It requires a systemic, participatory, and rights-based transformation of how technologies are designed, deployed, and governed. This session reinforced the need for collective responsibility, participatory design, and rigorous ethical evaluations in deploying AI in high-risk contexts such as human trafficking.