Advancing Ethical AI Governance

PRESENTERS: PAULA ISTURIZ CAVERO
UNESCO REGIONAL OFFICE FOR THE CARIBBEAN

Overview

1. Understand the ethical dimensions of AI
2. Explore the actions of multilateral organizations for ethical AI governance
3. Examples of work in action

Why AI Governance Matters

AI governance is necessary to harness AI for humanity while addressing its risks and uncertainties as AI-related services, algorithms, computing capacity and expertise become more widespread.

Why ethical AI?

Rapid deployment of AI-based technologies

AI that is fair, responsible, inclusive and upholds human rights

Goes beyond technical performance – it is about values

Prioritizes transparency, accountability and public good

Avoids reinforcing historical inequalities or harmful outcomes

Intersection of AI & Human Rights

1. **Right to Privacy:** Collection and analysis of personal data raises concerns about data protection

2. **Right to Non-Discrimination:** AI algorithms can perpetuate biases, gender inequality leading to discrimination and inequalities

3. **Right to Freedom of Expression:** AI can influence information disseminated online which can impact access to information

AI Systems:
Accountability
Transparency
Informed Consent

AI & Human Rights in the classroom

1	**Right to Privacy:** Collection and analysis of personal data raises concerns about data protection	**Student data & surveillance tools**
2	**Right to Non-Discrimination:** AI algorithms can perpetuate biases, gender inequality leading to discrimination and inequalities	**Algorithmic bias in educational tools**
3	**Right to Freedom of Expression:** AI can influence information disseminated online which can impact access to information	**AI moderation in learning platforms**

Bias in AI

Occurs when AI systems produce unfair, inaccurate, or discriminatory outcomes

Historical Bias

AI learns from past data, which may reflect existing inequalities

Data Bias

Underrepresentation of certain groups in training data

Algorithmic Bias

Design choices may unintentionally favor some groups over others

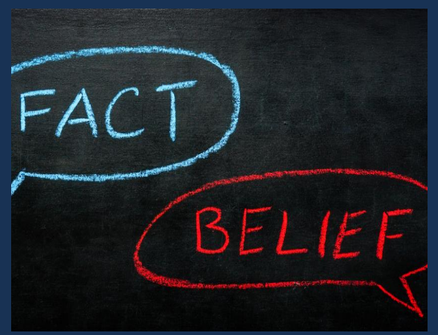
Human Bias

AI inherits biases from the people who create and train it

Potential Risks

Bias

AI can replicate and amplify societal prejudices



Surveillance

Used to monitor people without consent

Misinformation

AI-generated content can spread fake news

Exclusion

Marginalised voices often left out of development

From the search bar

Search “greatest leaders”

Global Political & Revolutionary Leaders

Name	Key Contribution
Nelson Mandela (South Africa)	Led the fight against apartheid, became the first Black president of South Africa, and a global symbol of reconciliation.
Mahatma Gandhi (India)	Led India’s nonviolent independence movement; inspired civil rights movements globally.
Abraham Lincoln (USA)	Preserved the U.S. Union during the Civil War, ended slavery with the Emancipation Proclamation.
Winston Churchill (UK)	British Prime Minister during WWII, known for leadership, oratory, and defiance against Nazi
Franklin D. Roosevelt (USA)	Led the U.S. t and strong le

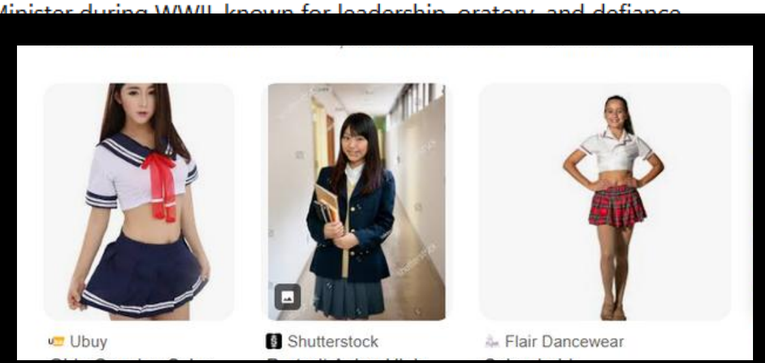
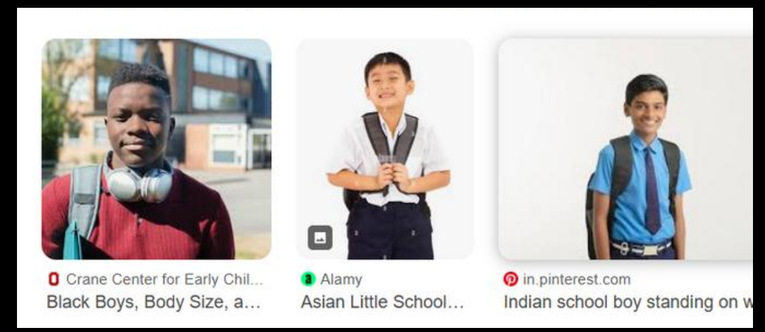


Image search “school girl” vs “school boy”



Biased results are not random, they are reflections of the data AI has been trained.

Opportunities

Personalized Learning

Adaptive e-learning platforms can tailor content to specific student needs.

Real-time feedback can help both students and teachers.

Accessibility

Enhance inclusion of students with disabilities (translation, speech-to-text or text-to-speech).

Virtual tutors or chatbots.

Administrative Efficiency

Automated scheduling, grading or tracking can help free up time of educators.

Early Intervention

Learning analytics can help identify students that need specific attention.

Predictive models can support student retention.

UN’s Role in Global AI Governance

Role of the UN

-Provides an inclusive platform for dialogue

- Contributes to anchoring AI governances in human rights and to support achievement of SDGs

- Provides guidance to government initiatives to build trustworthy AI ecosystems

- Acts as a convenor – helping bridge participation, knowledge production, capacity development & policy guidance

THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

Adopted by acclamation by all 193 Member States on 23 November 2021
Including all Caribbean States

THE FRAMEWORK

Interconnected values and principles

VALUES

1. Respect, protection and promotion of human rights and fundamental freedoms and human dignity
2. Environment and ecosystem flourishing
3. Ensuring diversity and inclusiveness
4. Living in peaceful, just and interconnected societies

PRINCIPLES

1. Proportionality and do no harm
2. Safety and security
3. Fairness and non-discrimination
4. Sustainability
5. Right to privacy, and data protection
6. Human oversight and determination
7. Transparency and explainability
8. Responsibility and accountability
9. Awareness and literacy
10. Multi-stakeholder and adaptive governance and collaboration

THE FRAMEWORK

Concrete and actionable pathways

(normative guidance)
to translate the values and principles
into policies, practices and actions
in multiple areas:

UN Digital Compact

- Framework for Inclusive and Rights-Based Digital Governance
- Adopted at the Summit of the Future (Sept 2024)
  Sets global principles for digital cooperation and AI governance.

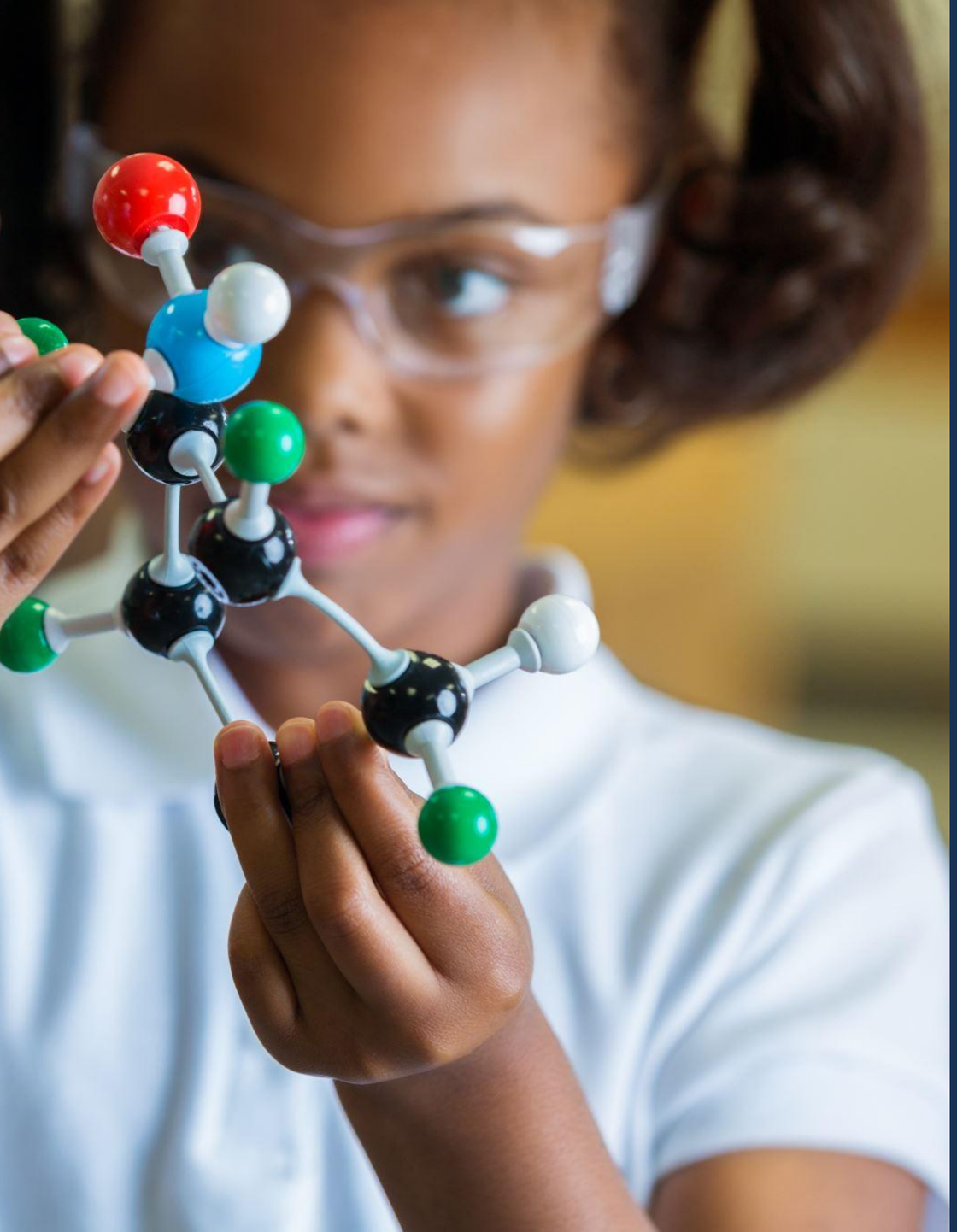
Objectives:
Close digital divides, advance data governance, enhance AI governance, safer digital spaces, inclusion in the digital economy.

AI Governance Pillars:
Multistakeholder approach, human-rights based, institutional mechanisms (Scientific Panel & Global Dialogue on AI)

Implementation Path:
Mapping contributions from stakeholders, high-level reviews and progress reports through 2027, coordination with existing UN structures and resolutions

UN Bodies on AI Governance

- **High-Level Advisory Body on AI:** Multistakeholder group of 32 experts (2023), advanced recommendations for the international governance of AI.
- **Global Dialogue on AI Governance**, platform for states and stakeholders to discuss the critical issues concerning AI facing humanity today. (Formally launched 25 Sept)
- **Independent International Scientific Panel on AI** brings together 40 experts to issue evidence-based scientific assessments related to the opportunities, risks and impacts of AI (Aug 2025, applications open 31 Oct).



Challenges Ahead

- Rapid pace of AI development vs. governance processes

- AI Literacy, capacity & funding, especially in countries of the Global South

- AI system production vs consumption; data sets

- Ensuring diverse representation/participation vs. dominance of a few actors

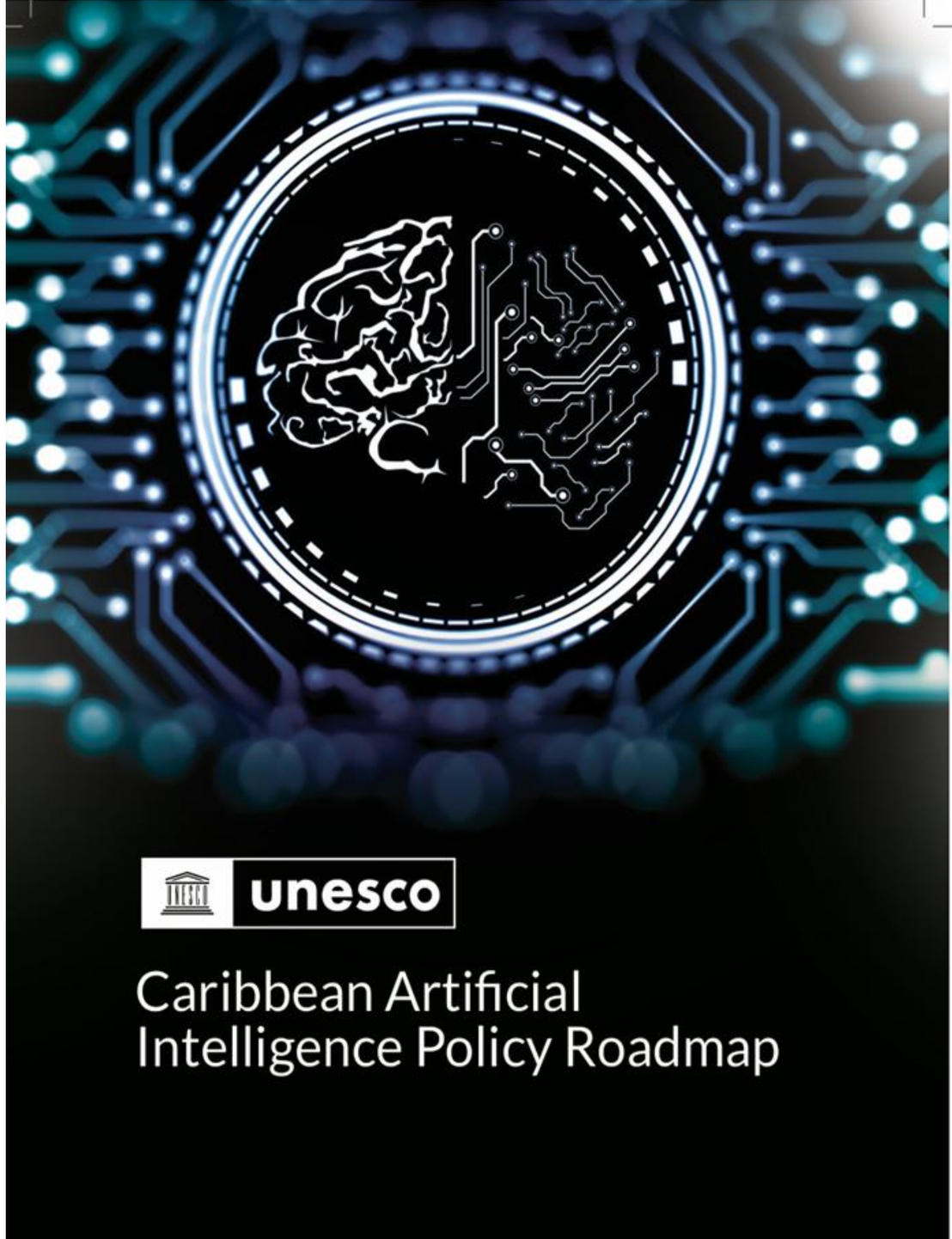
Promoting integrity & fairness in AI use

- Encourage critical thinking about AI
- Strengthen & integrate digital literacy and awareness; teach digital ethics
- Advocate for inclusive and transparent AI tools
- Ensure human oversight in decision-making processes
- Use AI tools ethically & transparently
- Develop national/ institutional policies to guide the use of AI.

Tools and Resources

The Caribbean Artificial Intelligence Policy Roadmap

- Raise awareness of Artificial Intelligence and its relevance, potential and challenges for the Caribbean
- Identify priorities and support the development of strategies for collaboration
- Leverage AI and digital transformation for the region's sustainable, social, economic and environmental development



Policy Pillars of the Caribbean Artificial Intelligence Policy Roadmap

THE TOOLS

Ethical Impact Assessment (EIA)

To assist Member States and stakeholders to identify and assess the ethical concerns and risks of AI systems, as well as risk prevention and monitoring measures, throughout the AI life cycle.

AI Readiness Assessment Methodology (RAM)

To assist Member States to identify their AI-readiness status at specific moments of their trajectory, along a continuum of dimensions (5 categories for analysis).

-Global Ethics of AI Forum, Ethics AI Experts without Borders, Women4EthicalAI, Regional Summit on Ethics of AI (LAC)

UNESCO RAM POLICY-ORIENTED REPORTS

Legal & Regulatory

Technological & Infrastructural

RAM

Societal & Cultural

Scientific & Educational

Economic

Deep Dive into Gender Equality Considerations

Bias, Gender-based violence and AI

AI systems often *mirror* and *magnify existing societal biases*, especially when they are built on unrepresentative data

58%

of young women and girls globally have experienced **online harassment** on social media

Generative AI can:

- **Reinforce gender stereotypes**
- **Generate hate speech**
- **Spread disinformation**
- **Enable prompt-based harassment**

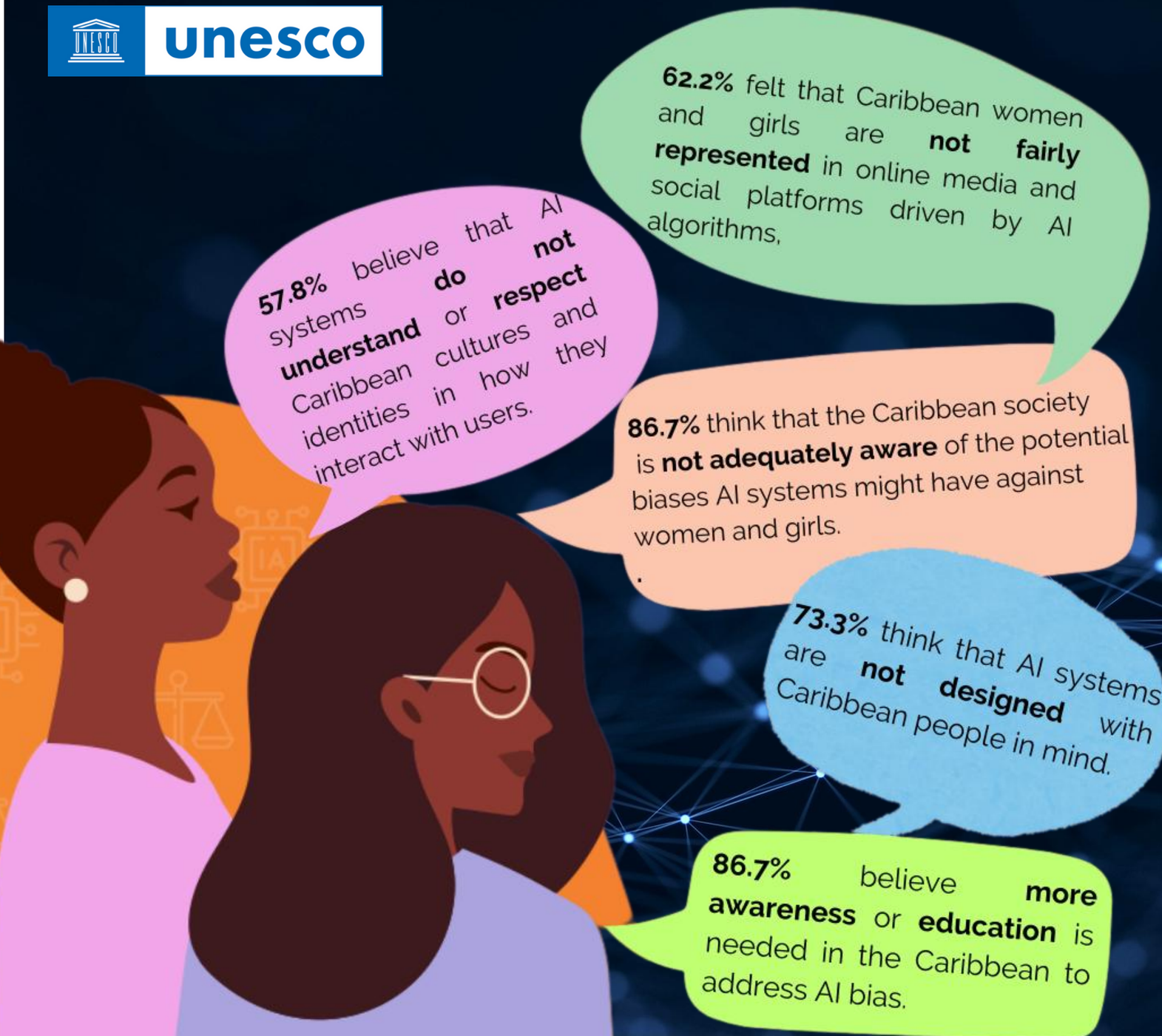
Key Risks Identified:

- **Weak safety guardrails in AI**
- **Norms & culture overlook gender harms**
- **Tools can be manipulated for harassment**

Next Steps:

- **Robust guardrails**
- **Stronger policies & accountability**
- **Enforce rights-based frameworks**
- **Inclusive design & testing**

PERCEPTION AND AWARENESS OF AI-BASED DISCRIMINATION AMONG WOMEN AND GIRLS IN THE CARIBBEAN



AI systems often *mirror* and *magnify existing gender biases*, especially when they are built on unrepresentative data

unesco

- Data on how women and girls in the region experience technology and violence is almost non-existent.

- Regional survey, analysis, consultation & audit of LLMs.

- Fill this regional data gap to ensure that women’s and girls' lived realities inform future policies, protections, and opportunities in AI.

- Complete the survey.

THANK YOU

Paula Isturiz, Programme Specialist for Social and Human Sciences

p.isturiz-cavero@unesco.org

www.unesco.org