



Monitoring responsible AI an NHRI framework/guidance

This Guidance aims to support countries in navigating these trends



- 1. Why do we monitor?**
Why AI governance matters for NHRIs
- 2. How is AI being monitored?**
Normative and policy reference points
- 3. Why are NHRIs key AI actors?**
Building on the Paris Principles
- 4. What can we monitor?**
- 5. How can we monitor?**
- 6. Who monitors?**
- 7. Institutional readiness and capacity building**

- 1. لماذا نقوم بالرصد؟**
- 2. كيف تتم مراقبة الذكاء الاصطناعي؟**
- 3. لماذا تُعد المؤسسات الوطنية لحقوق الإنسان جهات فاعلة رئيسية في مجال الذكاء الاصطناعي؟**
- 4. ما الذي يمكننا رصده؟**
- 5. كيف يمكننا الرصد؟**
- 6. من يقوم بالرصد؟**
- 7. الجاهزية المؤسسية وبناء القدرات**

The framework aims to support countries in navigating these trends



1. ***Why do we monitor?***
Why AI governance matters for NHRIs
2. ***How is AI being monitored?***
Normative and policy reference points
3. ***Why are NHRIs key AI actors?***
Building on the Paris Principles
4. ***What can we monitor?***
5. ***How can we monitor?***
6. ***Who monitors?***
7. ***Institutional readiness and capacity building***

The framework aims to support countries in navigating these trends



4. *What can we monitor?*

NHRIs, and governments, need to identify where monitoring efforts can be most valuable – and the resources, skills, and expertise required.

- **Specific areas of AI-enabled harm and violation of human rights** (algorithmic discrimination and bias, surveillance and privacy, automated decision-making in social services or criminal justice, labor market and 'gig economy' algorithms; children, young people, and AI; Generative AI and freedom of expression)
- **An emerging set of indicators (risk indicators – below - and positive outcomes):**
 - Discrimination (including measuring demographic parity)
 - Surveillance and privacy (including the extent of false matches from facial recognition)
 - Due process failures (such as gaps in explainability)
 - Labor harms
 - Child safety
 - Freedom of expression and information
 - Automated decision-making in social services and justice

What's missing?

ما الذي يمكننا رصده؟

يتعين على المؤسسات الوطنية لحقوق الإنسان والحكومات تحديد المجالات التي يمكن أن تكون فيها جهود الرصد ذات قيمة قصوى، وكذلك تحديد الموارد والمهارات والخبرات اللازمة لذلك.

مجالات محددة للأضرار وانتهاكات حقوق الإنسان المرتبطة بالذكاء الاصطناعي (التمييز والتحيز الخوارزمي؛ المراقبة والخصوصية؛ اتخاذ القرارات المؤتمتة في الخدمات الاجتماعية أو العدالة الجنائية؛ الخوارزميات في سوق العمل و"اقتصاد المهام المؤقتة"؛ الأطفال والشباب والذكاء الاصطناعي؛ الذكاء الاصطناعي التوليدي وحرية التعبير).

مجموعة ناشئة من المؤشرات (مؤشرات المخاطر - المذكورة أدناه - ومؤشرات النتائج الإيجابية):
التمييز (بما في ذلك قياس التكافؤ الديموغرافي)

المراقبة والخصوصية (بما في ذلك مدى حالات المطابقة الخاطئة في أنظمة التعرف على الوجه)
الإخفاقات في ضمانات الإجراءات القانونية السليمة (مثل الثغرات المتعلقة بقابلية التفسير)

الأضرار المتعلقة بالعمل سلامة الأطفال الحرية التعبير والحصول على المعلومات اتخاذ القرارات المؤتمتة في الخدمات الاجتماعية والعدالة

The framework aims to support countries in navigating these trends



5. *How can we monitor?*

Qualitative, quantitative, and mixed-methods approaches – as well as drawing on broader digital transformation and international development measurement. And identifying whether to monitor **before or after** AI deployments or - ideally – continuously, and in an agile way.

And, leveraging the unique reporting functions of NHRIs.

- **Qualitative approaches:** desk review of AI systems, legal and policy analysis, structured dialogue, community engagement, and public inquiries and hearings.
- **Quantitative approaches:** experimental approaches (including Randomized Controlled Trials, and A/B testing), quasi-experimental designs, and theory-based measurement.
- **And mixed methods.**

What's missing?

كيف يمكننا إجراء الرصد؟ من خلال اعتماد مناهج نوعية وكمية ومختلطة، والاستفادة أيضاً من الممارسات الأوسع نطاقاً في قياس التحول الرقمي والتنمية الدولية؛ وتحديد ما إذا كان الرصد سيتم قبل نشر أنظمة الذكاء الاصطناعي أو بعده، أو - وهو الخيار الأمثل - بشكل مستمر وبأسلوب مرن (رشيق).

وكذلك الاستفادة من وظائف إعداد التقارير الفريدة التي تتمتع بها المؤسسات الوطنية لحقوق الإنسان.

المناهج النوعية: المراجعة المكتبية لأنظمة الذكاء الاصطناعي، وتحليل الجوانب القانونية والسياساتية، والحوار المنظم، والمشاركة المجتمعية، والاستفسارات وجلسات الاستماع العامة.

المناهج الكمية: المناهج التجريبية (بما في ذلك التجارب العشوائية المنضبطة واختبارات والتصاميم شبه التجريبية، والقياس القائم على النظرية. والمناهج المختلطة.

Over to you! Some guiding questions

1. *What measurement and monitoring tools and approaches are you using in your work?*

ما هي أدوات ومناهج القياس والمراقبة (بما في ذلك النوعية والكمية وتلك التي تعتمد على المناهج المختلطة) التي تستخدمونها في عملكم؟

2. *To what extent have you identified indicators of AI governance – and where do you see gaps in your measurement approach?*

إلى أي مدى قمتم بتحديد مؤشرات لحوكمة الذكاء الاصطناعي؟ وأين ترون وجود فجوات في منهجية القياس المتبعة لديكم؟

3. *How is the NHRI mandate intersecting with AI deployment in your context?*

كيف يتقاطع تفويض المؤسسة الوطنية لحقوق الإنسان مع عمليات نشر واستخدام الذكاء الاصطناعي في سياق عملكم؟

4. *What AI and NHRI capacity building gaps exist in your context?*

ما هي الفجوات القائمة في مجال بناء القدرات المتعلقة بالذكاء الاصطناعي وعمل المؤسسات الوطنية لحقوق الإنسان في سياقكم؟

5. *Can we feature your work as a case study in any particular section?*

هل يمكننا استعراض عملكم كدراسة حالة في قسم محدد؟

6. *What is coming in the future? What are you worried, or excited, by?*

ما هي التطلعات المستقبلية؟ وما هي الأمور التي تثير قلقكم أو حماسكم؟



Thank you!