COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Introducing the HUDERIA

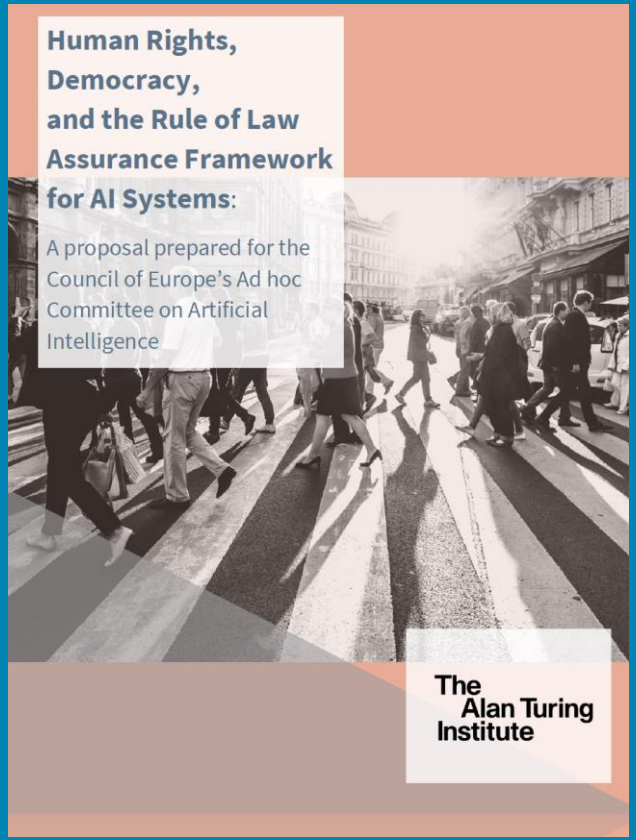
Mr Vadim Pak

Legal and policy advisor in the Committee on Artificial Intelligence

A collaborative effort to advance the HUDERIA

-the Committee of Ministers of the CoE tasked the Committee on Artificial Intelligence to elaborate “a legally non-binding methodology for the Risk and Impact Assessment of AI Systems from the point of view of Human Rights, Democracy and Rule of Law (HUDERIA) to support the implementation of the Framework Convention on AI”

-the Alan Turing has in 2022-2023 been advising and supporting the Secretariat and CoE experts on the drafting of the HUDERIA



A collaborative effort to advance the HUDERIA

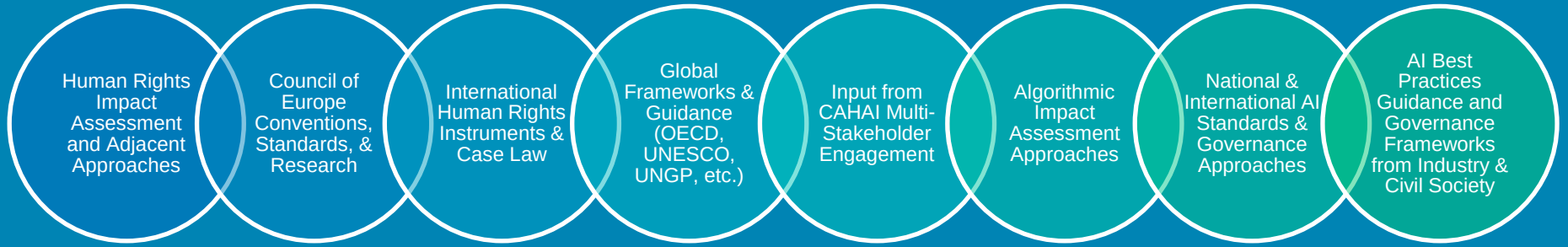
-On 11-13 January 2023 the Committee on Artificial Intelligence (CAI) examined the first draft of the HUDERIA methodology, having instructed the Secretariat to proceed on the basis of a 3-tiered approach:

(1) a general obligation to conduct a risk and impact management, as required by Article 16 of the Framework Convention;

(2) the stand-alone, non-legally binding methodology to be adopted by the Committee to serve as guidance for the Parties (the present Piloting programme is part of the on-going work on this part), and

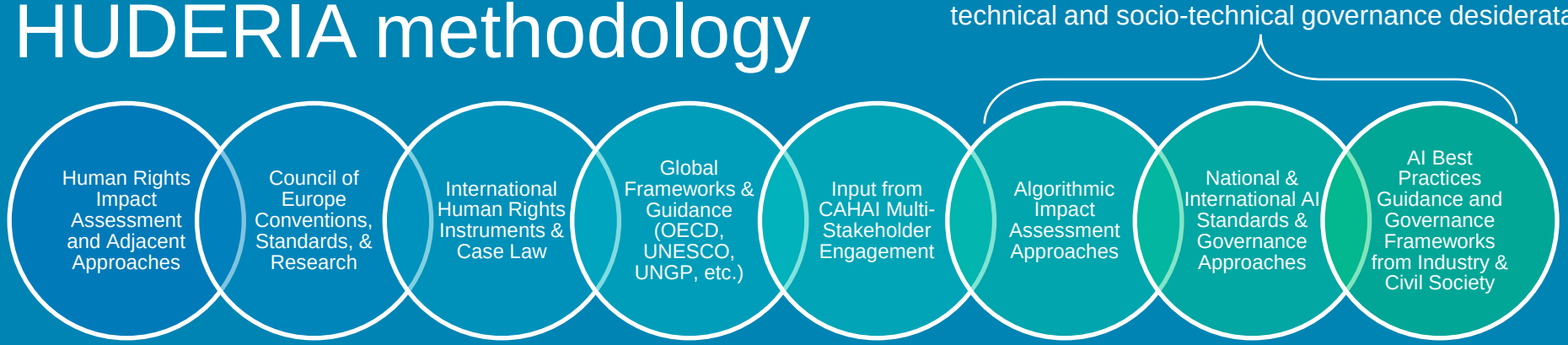
(3) an operationally feasible non-legally binding concrete model facilitating the implementation of the methodology.

An evidence-based approach to the proposed HUDERIA methodology



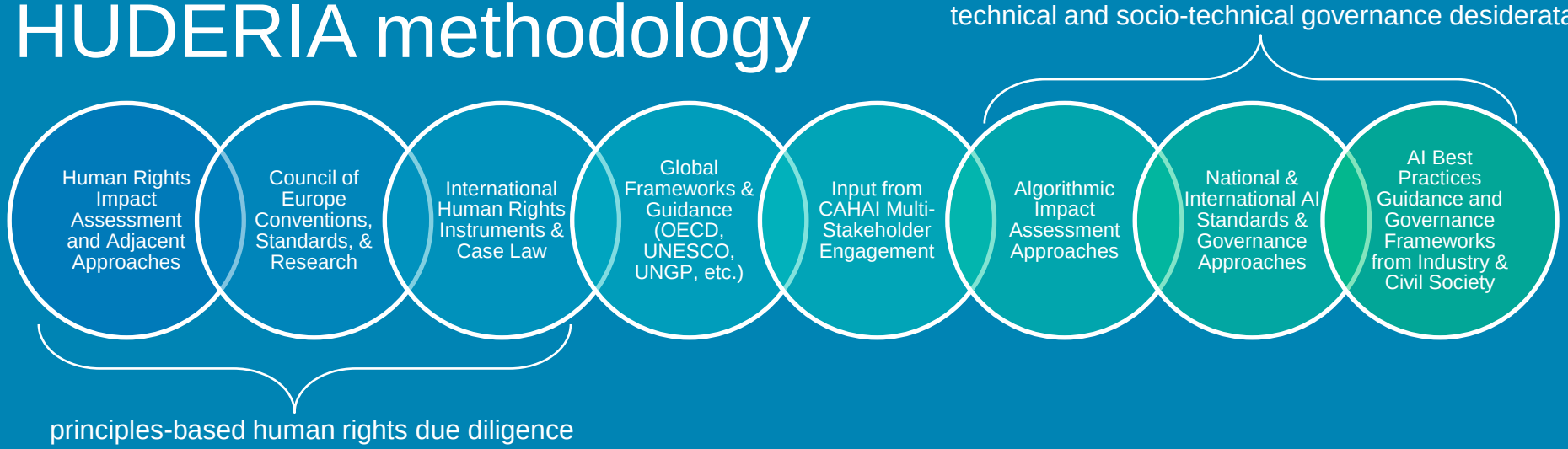
An evidence-based approach to the HUDERIA combines state-of-the-art knowledge about technical and socio-technical governance desiderata for responsible AI innovation with expertise on the procedural requirements for principles-based human rights due diligence.

An evidence-based approach to the proposed HUDERIA methodology



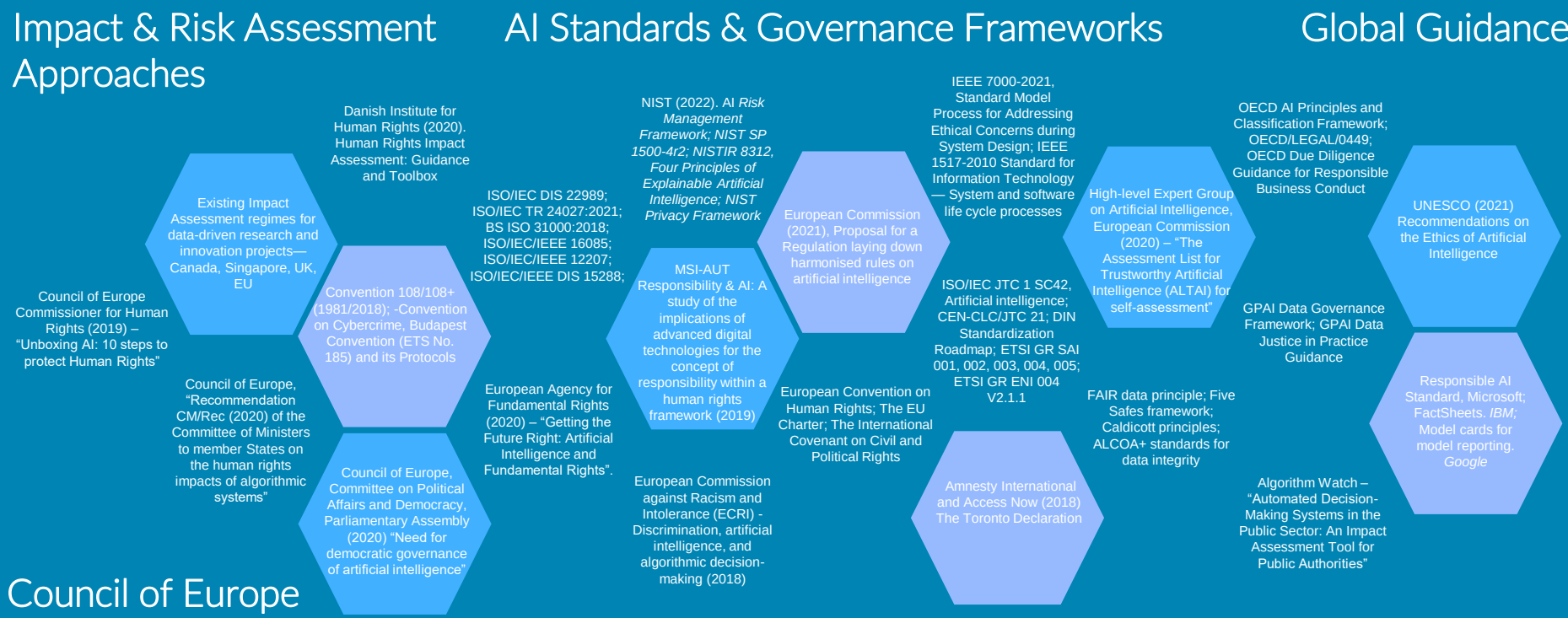
An evidence-based approach to the HUDERIA combines state-of-the-art knowledge about technical and socio-technical governance desiderata for responsible AI innovation with expertise in the procedural requirements for principles-based human rights due diligence.

An evidence-based approach to the proposed HUDERIA methodology



An evidence-based approach to the HUDERIA combines state-of-the-art knowledge about technical and socio-technical governance desiderata for responsible AI innovation with expertise in the procedural requirements for principles-based human rights due diligence.

Multi-modal knowledge integration



Three overarching goals of HUDERIA:

1. To improve upon the fragmented character of the the current state of knowledge in AI technology policy by integrating distinctive tasks of risk analysis, stakeholder engagement, impact assessment, risk management, impact mitigation, and innovation governance into a coherent and integrated whole

Three overarching goals of HUDERIA:

1. To improve upon the fragmented character of the the current state of knowledge in AI technology policy by integrating distinctive tasks of risk analysis, stakeholder engagement, impact assessment, risk management, impact mitigation, and innovation governance into a coherent and integrated whole
2. To ensure a risk-based and proportionate approach to impact assessment that *responsibly* optimizes agility and innovation across the AI ecosystem by not creating excessive administrative burdens

Three overarching goals of HUDERIA:

1. To improve upon the fragmented character of the the current state of knowledge in AI technology policy by integrating distinctive tasks of risk analysis, stakeholder engagement, impact assessment, risk management, impact mitigation, and innovation governance into a coherent and integrated whole
2. To ensure a risk-based and proportionate approach to impact assessment that *responsibly* optimizes agility and innovation across the AI ecosystem by not creating excessive administrative burdens
3. To take a sufficiently procedural approach to formulating the base-line requirements of the methodology so to safeguard a significant margin of appreciation for the unique circumstances of domestic authorities, which are better placed to make relevant policy and regulatory choices, taking into account their country's specific political, economic, social, cultural, and technological contexts

Basic structure of the HUDERIA methodology

1. Context-based risk analysis (COBRA)
2. Stakeholder Engagement Process (SEP)
3. Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA)
4. Impact mitigation plan and access to remedies
5. Iterative requirements

Basic structure of the HUDERIA methodology

1. **Context-based risk analysis (COBRA)**

- The main purpose of the COBRA is to identify the extent to which, if at all, an AI system is likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, in view, in particular, of the contexts of its deployment.

- This is done through an an analysis of risk factors (characteristics or properties of an AI innovation context that are associated with a higher likelihood of some outcome (or outcomes) that negatively impact human rights, democracy, and the rule of law:
- - application context; - system design and development context; - deployment context.

- The COBRA also includes a risk calibration mechanism that integrates variables of the scale, scope, and likelihood of potential harms to help Parties establish a proportionate approach both to subsequent elements of the methodology and to the level of stakeholder engagement that is needed throughout the project lifecycle, more generally.

Basic structure of the HUDERIA methodology

2. Stakeholder Engagement Process (SEP)

- The purpose of the stakeholder engagement process (SEP) is to identify stakeholder salience and to facilitate proportionate rights-holder involvement and input throughout the project workflow. A diligent SEP is essential for ensuring that rights-holders' views are appropriately incorporated in the assessment and governance of the project and that any potential risks of adverse impacts are identified and mitigated across the system's lifecycle.
- Stakeholder engagement may take various forms, but the exact level of rights-holder participation should be proportionate to risks identified in the COBRA and other relevant factors.
- An SEP includes stakeholder analysis, reflection on limitations of team members' perspectives that helps identify missing stakeholder viewpoints, establishment of engagement objectives and methods, and the initiation and implementation of engagement processes

Basic structure of the HUDERIA methodology

3. **Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA)**

- The purpose of the HUDERIA is to provide detailed evaluations of the potential and actual impacts that the design, development and application of an AI system could have on human rights and fundamental freedoms, democracy, and the rule of law.

- With the support of proportionate stakeholder engagement, this process contextualises and corroborates potential adverse effects identified at the previous stages, allows for the identification and analysis of further potential harms by enabling project team members to engage in extended reflection and gap analysis, and enables the establishment of an impact mitigation plan that also sets up access to remedy.

Basic structure of the HUDERIA methodology

4. Impact mitigation plan and access to remedies

- This final phase of the methodology involves an assessment of the severity (scale, scope, and remediability) of potential adverse impacts; a clear presentation of the measures and actions that will be taken to address the potential adverse effects; a clarification of the roles and responsibilities of the various actors involved in impact mitigation, management and monitoring; a plan for monitoring impact mitigation efforts and for re-assessing and re-evaluating the HUDERIA during subsequent development and deployment phases of the project lifecycle;

- It also involves an accessible presentation of access to remedy mechanisms that will be available to impacted rights-holders.

Basic structure of the HUDERIA methodology

5. Iterative requirements

- The methodology reinforces that, in the impact assessment process, continuous attention should be paid both to the dynamic and changing character of the AI production and implementation lifecycle and to the shifting conditions of the real-world environments in which systems will be embedded.
- For this reasons, iterative requirements are included, though the exact modalities of this part of the process are left to Parties.
- These requirement include the iterative revisitation of HUDERIA. This follows a plan (established as part of the HUDERIA) for monitoring impact and impact mitigation efforts and for re-assessing and re-evaluating the HUDERIA during each phase of the project lifecycle up to system retirement or decommissioning; Such processes remain as responsive as possible to the way the AI system is interacting with its operating environments and with impacted rights-holders.

Special features of HUDERIA

- A deliberate effort to ‘open up’ the technical aspects for the users of the Methodology.
- Ressources sections with detailed questions and indications regarding relevant aspects of AI system’s design and development (data science and AI) and application and system deployment context (human rights law: mapping of rights and mapping of HR sensitive contexts).

Comparison to other Risk Management Initiatives

OECD Interoperability Framework	Govern					Define	Assess	Treat
	Monitor and Review	Communicate	Consult	Document	Embed			
OECD DDG	Track	Communicate	Embed			Identify and Assess		Cease, Prevent and Mitigate / Remediation
ISO/IEC 42001:2023	Monitoring and Review	Communication and Consultation		Recording and Reporting	Leadership and Committment	Scope, Context and Criteria	Risk Assessment	Risk Treatment
ISO/IECD Guide 51	Validation and Documentation	n/a	n/a	Validation and Documentation	n/a	Identify user, intended use and reasonably foreseeable misuse / Hazard identification	Estimation / Evaluation of risk	Risk reduction
HUDERIA	**Iterative requirements**	**Stakeholder Engagement Process**		**HUDERIA documentation process**	**HUDERIA workflow**	**COBRA**	**HUDERIA**	**Impact Mitigation and Access to Remedies**
IEEE 7000-21	n/a	Transparency management process	Ethical values elicitation and prioritisation	n/a	n/a	Concept of operations and context exploration	Ethical values elicitation and prioritisation	Ethical requirements definition and ethical risk-based design
NIST AI RMF	Govern* * Monitor; Consult; Communicate; Document are part of every high-level function of the NIST AI RMF and as such also feature in Map, Measure and Manage.					Map	Measure	Manage

Piloting of HUDERIA

- On 7 June 2023 the Secretariat sent out its Proposal for the Collaborative Piloting of HUDERIA to the members of the CAI;

- 7 states volunteered to participate in the program - 5 of them submitted specific AI use case for a test run of HUDERIA;

- The piloting took place in June – early July 2024.

Many thanks for listening.