Evidence Without Operationalisation

A Critical Assessment of the Preliminary Report of the UN Independent International Scientific Panel on AI, and a Comparative Analysis with the EW-AiRM™ Enterprise-Wide AI Risk Management™ Framework

(Source: https://www.un.org/independent-international-scientific-panel-ai/en/preliminary-report)

Human-Ai.Institute Analysis Paper

July 2026

Prepared by the Human-Ai.Institute (www.Human-Ai.Institute)

Executive Summary

In July 2026, the Independent International Scientific Panel on Artificial Intelligence, established by UN General Assembly resolution 79/325, published its first Preliminary Report: an evidence-based assessment of the opportunities, risks and impacts of AI. It is a landmark document: Forty independent experts, drawn from all five UN regional groups, have produced the first standing, iterative, global scientific baseline on AI, and they have done so in barely three months.

This paper assesses that report critically and compares it with the EW-AiRM™ (Enterprise-Wide AI Risk Management) framework, the practitioner architecture developed by Prof. Markus Krebsz and published through via the Human-Ai.Institute. Our headline conclusions are as follows:

- **The diagnosis converges.** The Panel's central findings, that AI capabilities are advancing faster than the ability to measure or govern them, that agentic AI represents a governance step change, that human oversight is not operationalised, and that multi-agent interactions create novel systemic risks, map with striking precision onto the architecture EW-AiRM™ has already codified: Six Strategic Pillars, the HAiPECR ethical filter, eight AI Black Swan categories and through-the-lifecycle monitoring.
- **The prescription is missing, by design.** The Panel is mandated to be policy-relevant but not policy-prescriptive. That leaves an operationalisation void between global evidence and enterprise action. EW-AiRM™ exists precisely in that void: it converts the Panel's findings into board-ready governance instruments, tiered implementation and measurable indicators.
- **The evidence dilemma must not become an alibi.** The Panel rightly identifies that policymakers need evidence that arrives too late. EW-AiRM™'s first Governance Maxim answers this directly:

> **"Waiting for perfect clarity is not a governance position, it is a governance failure."**

- **The enterprise is the missing actor.** The report addresses Member States. Yet by its own evidence, 91% of notable AI models originate in the private sector, and deployment decisions sit inside firms. Governance that does not reach the enterprise risk function will not reach the technology.

The UN report and EW-AiRM™ are therefore not competitors but complements: the Panel supplies the shared evidence base; EW-AiRM™ supplies the operating system that turns evidence into governed practice.

1. What the UN Report Is, and Why It Matters

The Preliminary Report is the first output of the first standing UN scientific body on AI. Its mandate, set by General Assembly resolution 79/325 and anchored in the Global Digital Compact (GDC) and the Pact for the Future, is strictly scientific and non-political: to document consensus and disagreement, remain policy-relevant but not policy-prescriptive, and update its assessment progressively through thematic briefs and annual reports presented to the UN Global Dialogue on AI Governance.

The report's evidentiary core is strong and current. It documents benchmark performance rising sharply (Humanity's Last Exam from 8% to 45% in sixteen months; GPQA Diamond at roughly 95%; FrontierMath from 19% to 88%), METR task-horizon doubling accelerating to 4.6 months, hyperscaler capital expenditure approaching an annualised $650 billion, and extreme concentration: the United States hosting around 75% of top-500 AI supercomputing capacity, China 15%, and the rest of the world 10%. It confirms that over a billion people now use conversational AI weekly, that 91% of notable models originate from the private sector, and that governance instruments, of which it counts more than 40 types, are fragmented, corporate-concentrated and rarely measured for real-world effectiveness.

Three structural features distinguish it from prior efforts. First, continuity: it is designed for sustained, iterative assessment rather than a one-off snapshot. Second, universality: gender-balanced membership across all five regional groups gives it a legitimacy that national or industry-affiliated assessments cannot claim. Third, honesty about uncertainty: its closing section on acting under uncertainty concedes that mistakes are not always reversible and that most instruments needed already exist; the open question is how to apply them.

The UN Panel's own admission

> **“Most instruments needed already exist; the open question is how to apply them.”**

That single sentence defines the boundary of the Panel's mandate, and the starting line of EW-AiRM™.

2. Strengths: Where the Panel Gets It Right

- **Agentic AI as a step change.** Section 2.6 states plainly that institutions built to oversee static models and human-in-the-loop (HITL) software do not fit agentic systems, and that emergent multi-agent risks remain poorly understood. This is a scientifically courageous position for a consensus body.
- **Measurement first.** The report treats evaluation and measurement as the foundation of governance, cataloguing benchmark saturation, data contamination, deception, evaluation awareness and sandbagging. Its call for continuous, post-deployment measurement mirrors mature safety practice in aviation and pharmaceuticals.
- **Naming sycophancy as a systemic alignment failure.** The report links engagement-optimised AI behaviour to documented fatalities and frames sycophancy as an exploitable security failure, not a product quirk. This is conduct risk in all but name.

- **The enabling-environment thesis.** Access alone does not equal benefit; complements in data, skills, workflows and institutions determine outcomes. This aligns with decades of general-purpose technology economics and with EW-AiRM™ Pillar Two (Readiness).
- **Equity as an analytical category.** The linguistic and geographic asymmetries, more than 7,000 languages spoken but only a fraction served, 99% of deepfake videos targeting girls and women, and the global South's disproportionate exposure, are given genuine evidential weight rather than a courtesy paragraph.
- **Referencing the MIT AI Risk Repository.** The Panel cites the MIT FutureTech incident and risk resources as governance infrastructure. EW-AiRM™ already operationalises this repository at its Operational Layer, mapping more than 1,700 AI risks to 831 mitigation controls.

3. Critical Assessment: Seven Gaps

3.1 Description without operationalisation

The report is deliberately non-prescriptive. That is a mandate constraint, not an intellectual failure, but the consequence must be named: a board director, chief risk officer or regulator finishing the report knows what is happening and does not know what to do on Monday morning. There is no maturity model, no control taxonomy, no assessment instrument, no tiering for organisations of different scale. The distance between “the evidence dilemma is serious but not insurmountable” and an audit-ready governance artefact is the entire discipline of risk management, and the report leaves that distance uncrossed.

3.2 The evidence dilemma risks becoming an alibi

The Panel frames the central problem elegantly: decision-makers need evidence that will only exist once it is too late to act on it. What the framing under-emphasises is that risk management has always operated under exactly this condition. Basel-era financial risk, operational resilience and pandemic preparedness all matured by acting on structured judgement under uncertainty, with measurement retrofitted as it became available. EW-AiRM™ encodes this as its first Governance Maxim: waiting for perfect clarity is not a governance position, it is a governance failure. A scientific panel cannot say this; a practitioner framework must.

3.3 The enterprise is the missing governance actor

The report's addressee is the Member State. Yet its own data show that the decisive governance decisions, training data, safeguards, deployment thresholds, model access and capability release, sit inside private firms. Between the frontier developer and the state sits the deploying enterprise: the bank, hospital, insurer, manufacturer and logistics operator that actually carries AI risk into the real economy. The report has no theory of the enterprise risk function, no engagement with the three lines of defence, and no account of how boards should discharge accountability. This is the single largest structural gap, and it is the space EW-AiRM™ occupies in full.

3.4 Multi-agent emergence is flagged, not architected

The Panel repeatedly notes that novel risks arise from interactions between multiple agents and that single-agent evaluation cannot detect them. It stops there. EW-AiRM™'s Resilience Layer treats this as AI Black Swan Category 7 (Multi-Agent Emergence): interaction-generated failure between AI systems, with

scenario instruments, board exercises and escalation triggers already built. The report also gestures at convergence with quantum computing in its long-term trajectory without examining the cryptographic transition risk that EW-AiRM™ isolates as Category 8.

3.5 Human oversight: correctly diagnosed, left unspecified

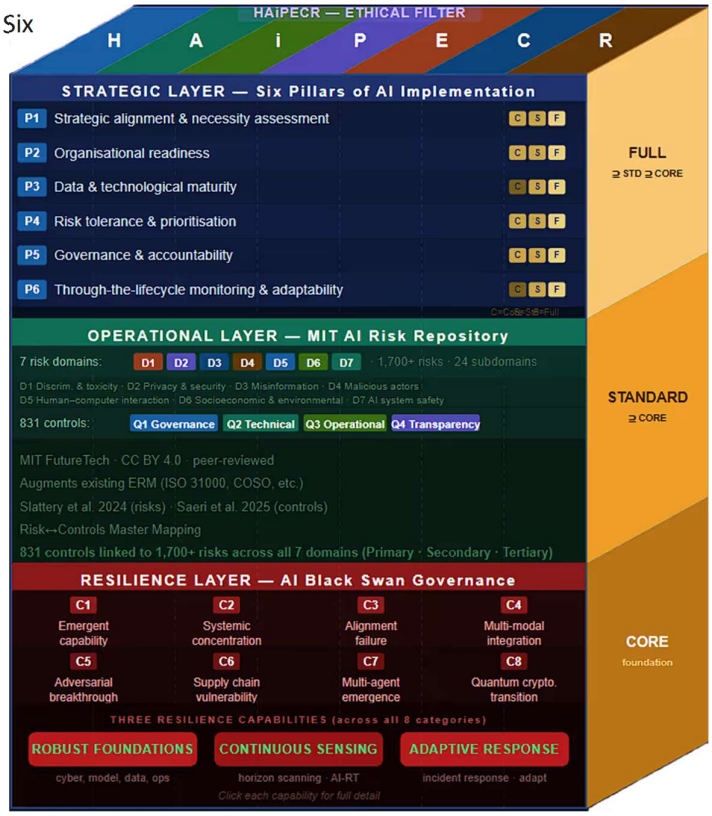
The report's observation that oversight is not yet operationalised as a measurable requirement, and that a human reviewer at the end of a workflow does not automatically improve outcomes, is exactly right. But it offers no operationalisation. The HAiPECR filter's H dimension (Human Oversight, aligned to UNESCO's 2021 Recommendation) does precisely this: it converts oversight from an aspiration into gated, evidenced decision points across all three EW-AiRM™ layers, with intervention, reversibility and accountability expectations defined before deployment, not after incident.

3.6 Ethics is distributed rather than integrated

Human rights, children's rights, fairness, privacy, sustainability and information integrity each receive serious treatment, but in separate sections, without an integrating mechanism that a governing body could apply to a single AI use case. HAiPECR™ exists as exactly that mechanism: seven thematic dimensions (Human Oversight; Accountability; inclusivity, deliberately lowercase as an embedded prerequisite; Privacy and Safety; Transparency and Fairness; Conduct Risk; Sustainability) applied as one filter across the lifecycle, mapped to UNESCO's ten core principles and extended to neural data via the 2024 UNESCO Neurotechnology Recommendation.

3.7 Measurement critique without a measurement architecture

The report is excellent on why current measurement fails, and silent on what an organisation should measure instead. EW-AiRM™ Pillar Six (Through-the-Lifecycle Monitoring) supplies the architecture: KXIs, Key X Indicators, as the umbrella for key performance, risk and control indicators, wired into a Risk Propensity Index and tiered assessment. The Panel's call for privacy-preserving, cost-conscious AI measurement by national statistical systems is welcome; the enterprise equivalent already exists and can feed it.



Source: EW-AiRM Cube Visualisation (www.EWAIRM.com)

A final scoping note: the Panel's mandate excludes military applications entirely. That is a General Assembly decision, not the Panel's, but readers should understand that the systemic risk picture is therefore structurally incomplete, particularly for dual-use cyber and biological capabilities that the report itself acknowledges only obliquely.

4. Point-by-Point Mapping: UN Findings to EW-AiRM™

The table below maps the Panel's principal findings to the EW-AiRM™ components that operationalise them. The framework comprises three layers (Strategic, Operational, Resilience), six Strategic Pillars, the seven-dimension HAiPECR™ ethical filter, eight AI Black Swan categories, five Non-Negotiables and three implementation tiers (Core, Standard and Full).

UN Panel finding	EW-AiRM™ component	What EW-AiRM™ adds
Capabilities advancing faster than measurement or governance (§2.1)	**Pillar 6: Through-the-Lifecycle Monitoring;** **KXI architecture, comprising** - **Key Performance Indicators** - **Key Risk Indicators** - **Key Control Indicators**	**Continuous**, enterprise-level **measurement** with defined indicators (KXIs), **thresholds** and **escalation**, rather than a call for measurement
Agentic AI is a governance step change (§2.6)	**Resilience Layer;** **Black Swan Category 7: Multi-Agent Emergence**	**Scenario instruments**, board exercises and **controls for interaction-generated failure** between AI systems
Human oversight not operationalised as a measurable requirement	**HAiPECR™ H dimension: Human Oversight**	**Gated decision points** with intervention, **reversibility** and **accountability** defined pre-deployment
Sycophancy as systemic, exploitable alignment failure with documented harm	**HAiPECR™ C dimension: Conduct Risk (Universal Conduct Risk Paradigm)**	**Treats AI conduct and human conduct symmetrically**, bringing engagement-optimised behaviour under conduct-risk governance
Concentration: 75% of top-500 compute in one state; 91% of models private-sector	**Pillar 1: Necessity Assessment;** **Pillar 4: Risk Tolerance;** **Systemic concentration analysis**	Dependency, **vendor-concentration and exit analysis at enterprise level**, where the exposure actually lands
Over 40 governance instrument types, fragmented and rarely measured	**Integrated three-layer architecture;** **three tiers (Core / Standard / Full)**	A **single coherent operating model** proportionate to organisational scale, replacing instrument sprawl
MIT FutureTech risk and incident resources cited as infrastructure	**Operational Layer: more than 1,700 MIT risks mapped to 831 controls**	A **completed risk-to-control mapping** across primary, secondary and tertiary tiers, ready for audit
Evidence dilemma: act without evidence, or wait until too late	**Governance Maxims;** **Risk Propensity Index**	A **decision doctrine**: structured action under uncertainty, with retrospective assessment as evidence matures
Ethics and rights treated across separate sections	**HAiPECR™ ethical filter: seven dimensions aligned to UNESCO's ten principles**	**One integrating filter** applied to **every use case** across all three layers, **including neural data**

UN Panel finding	EW-AiRM™ component	What EW-AiRM™ adds
Enabling environment: access alone does not equal benefit	**Pillar 2: Readiness;** **Pillar 3: Technological Maturity**	Assessable **readiness dimensions** covering skills, data, **workflow redesign** and **institutional capacity**

Table 1 — Mapping the Panel's July 2026 findings to the EW-AiRM™ framework components.

5. Where the Panel Independently Validates EW-AiRM™

Because the Panel worked independently and under a strictly scientific mandate, its convergence with EW-AiRM™ constitutes meaningful external validation of the framework's design decisions:

- **Lifecycle over launch.** The Panel's insistence on continuous, post-deployment measurement validates the decision to make Through-the-Lifecycle Monitoring a Strategic Pillar rather than an appendix.
- **Deployed system over model.** The Panel states that the unit of evaluation must be the deployed system, including model, tools, environment and users, not the model alone. This is the enterprise-wide premise of EW-AiRM™ stated in scientific language.
- **Multi-agent emergence as a distinct risk class.** The Panel's finding that emergent multi-agent risks cannot be detected through single-agent evaluation confirms the case for Black Swan Category 7 as a standalone category, an architectural framework design choice made before this report appeared.
- **Conduct symmetry.** The report documents both AI misconduct (deception, sycophancy, sandbagging) and human misconduct through AI (persuasion engineering, AI washing, deepfake abuse). The Universal Conduct Risk Paradigm underlying HAiPECR™'s C dimension was built for exactly this symmetry.
- **Proportionate governance.** The Panel's spectrum from hard law to soft law, and its emphasis on capacity differences between states, parallels EW-AiRM™'s Core, Standard and Full tiers for organisations of different maturity and scale.

5.1 Where the Panel usefully challenges the practitioner community

Fairness requires acknowledging traffic in the other direction. Three areas of the report press on any enterprise framework, EW-AiRM™ included:

- First, **linguistic and cultural coverage**: the Panel's evidence that models fail unevenly across languages means deployment assessments must test in the languages of actual users, not vendor demonstration languages.
- Second, **environmental accounting**: the Panel's finding that standardised lifecycle measurement of energy, water and e-waste is still lacking is a live challenge for the HAiPECR™ R (Sustainability) dimension, which absorbs intergenerational rights and must keep hardening its metrics.
- Third, **state capacity**: enterprise frameworks assume a functioning supervisory counterpart; the Panel's finding that 118 countries are absent from major AI governance discussions is a reminder that in many jurisdictions the enterprise framework is, de facto, the only governance layer present. That raises the standard it must meet.

6. Implications for Boards and Risk Practitioners

- **Treat the UN report as your external context evidence.** It is citable, multilateral and current. Use it to substantiate risk appetite discussions, ICAAP-style narratives and regulatory engagement, and pair it with an operational framework for the response.
- **Assume agentic exposure now, not later.** With documented prompt-injection success rates against coding agents as high as 84%, and multi-agent architectures entering production, boards should demand a named owner for agentic AI risk and a tested escalation path this quarter.
- **Operationalise oversight before regulators do it for you.** **The EU AI Act's transparency obligations become enforceable on 2 August 2026**. Meaningful human oversight, defined, evidenced and measurable, is the direction of travel across jurisdictions.
- **Measure or be measured.** The Panel warns that without effective measurement, governance risks becoming symbolic. KXIs, incident reporting and a maintained risk-to-control mapping are the difference between governance and theatre.
- **Do not wait for the annual report.** The Panel itself will iterate through thematic briefs. Enterprise governance should iterate faster.

7. Conclusion

The Preliminary Report of the Independent International Scientific Panel on AI is the most credible shared evidence base the international community has yet produced on this technology. It should be read, cited and taken seriously in every boardroom. But evidence is not governance. The Panel documents, with rigour and honesty, a world in which capabilities outrun measurement, agents outrun oversight, and instruments outrun effectiveness. It cannot, by mandate, close that gap.

EW-AiRM™ was built to close it. The convergence between the Panel's independent scientific findings and the framework's architecture, on lifecycle monitoring, deployed-system evaluation, multi-agent emergence, operationalised oversight and conduct symmetry, suggests the practitioner and scientific communities are now describing the same problem from two sides. The task for 2026 and beyond is to connect them: global evidence flowing into enterprise governance, and enterprise incident and indicator data flowing back into the global evidence base the Panel is mandated to maintain.

Governance Maxim

“Waiting for perfect clarity is not a governance position. It is a governance failure:

Governance that does not adapt is governance that will eventually fail.”

About the Human-Ai.Institute

The Human-Ai.Institute (www.Human-Ai.Institute) advances practical, human-centred AI governance. EW-AiRM™ (Enterprise-Wide AI Risk Management) is a three-layer, six-pillar framework with the HAiPECR™ ethical filter, eight AI Black Swan categories, five Non-Negotiables and three implementation tiers, operationalising more than 1,000 risks from the MIT FutureTech AI Risk Repository against 831 mitigation controls. The full framework is published by Wiley Finance. HAiPECR™ has been listed on the OECD AI Policy Observatory since April 2023.

Disclaimer: This analysis reflects the views of the Human-Ai.Institute and is provided for information only. It does not constitute legal or regulatory advice. The UN Preliminary Report is © 2026 United Nations; all findings referenced are paraphrased and attributed. EW-AiRM™ and HAiPECR™ are trademarks of De-Risking Solutions Ltd.