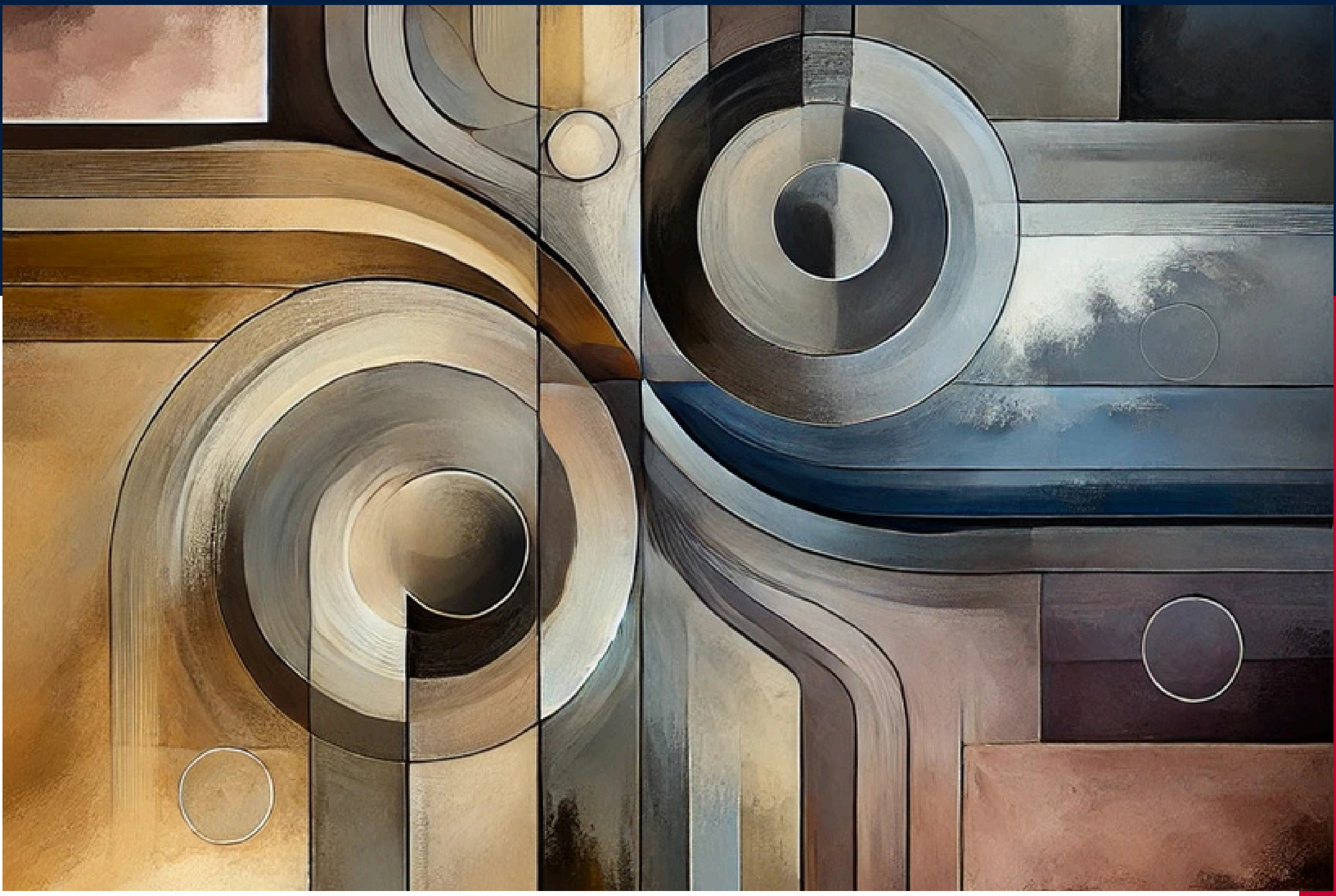
POLICY BRIEF
2025 MARCH

Looking Ahead: Synergies Between the EU AI Office and UK AISI



Authors: Lara Thurnherr, Risto Uuk, Tekla Emborg, Marta Ziosi, Isabella Wilkinson, Morgan Simpson, Renan Araujo, Charles Martinet

Looking ahead: Synergies between the EU AI Office & UK AISI

Lara Thurnherr*, Risto Uuk*, Tekla Emborg*, Marta Ziosi*[1], Isabella Wilkinson, Morgan Simpson, Renan Araujo, Charles Martinet.

Executive summary

The UK AI Security Institute (AISI) and the European AI Office are the primary bodies covering security and safety of AI systems in their respective jurisdictions. The institutions have overlapping mandates and share various functions. Using a framework of four levels of engagement - collaboration, coordination, communication and separation[2] - this brief provides an overview of potential synergies and strategic alignment, summarized in a table of ideas (Table 1). This framework can further provide a model for other regional strategic arrangements within the broader network of AISIs.

In particular, the institutions could benefit from **collaborating** on jointly developing standards and international engagement, where aligning efforts would amplify global influence and streamline participation in forums such as AI summits. **Coordination** may be ideal for some aspects of evaluations, where avoiding duplication and ensuring interoperability are key, for example by developing consistent evaluation metrics, establishing clear policies on evaluation responsibilities, and holding periodic meetings to interpret results. **Communication** is well-suited to areas like risk monitoring and incident reporting, where differences in institutional mandates and definitions could benefit from consistent information-sharing channels. **Separation** is necessary when confidentiality, sensitivity, or differing priorities demand independent action. For example, exploratory research on future trends might be better handled separately to avoid risks and ensure strategic autonomy. At all levels of engagement, it is important to also keep in mind **institutional differences**, including the UK AISI dealing with proprietary information from companies, the AI Office being a regulator, and particular relationships with their respective defense agencies.

In short, there are a range of avenues for the UK AISI and the EU AI Office to work effectively together on their synergistic tasks while respecting their overlapping but distinct mandates, institutional differences, and contexts. Our key recommendation is for policy practitioners in both jurisdictions, within public bodies or civil society, to concretise these avenues and provide more detailed guidance for their implementation.

[1] This work was initiated prior to, and is entirely unrelated to, the involvement of Marta Ziosi in the EU AI Act Codes of Practice.

[2] Collaboration as a joint working process, where members or teams of both institutions aim to solve a common problem together. We define coordination as a process where both institutions adapt their activities to each other to achieve a higher degree of interoperability and lower degree of unnecessary duplication of efforts. Communication simply means to communicate without an explicit expectation to change the other party's actions. Separation refers to areas where independent action is preferable due to sensitivity, differing priorities, or resource constraints.

Table 1: Areas for potential Collaboration, Coordination, Communication and Separation

Common activities	Collaboration	Coordination	Communication	Separation
Develop Evaluations	Design and develop joint post-deployment evaluations	Develop consistent ways to accredit evaluation developers	Share Evaluation Methods Development Tools	Developing evaluations requiring high levels of sensitive information and or model access
Conduct Evaluations	Executive joint post-deployment evaluations	Establish clear policies on who evaluates which systems	Share evaluation tools	Conducting evaluations with sensitive information and/or model access
Interpret Evaluations	Mutually recognise evaluation results	Periodic Meetings on evaluation interpretation	Share evaluation interpretations	Duplicate evaluation interpretations for insight on uncertain methods
AI Risk estimates	Pool resources to develop risk estimates	Coordinate which risk estimates are currently unresolved and specialise	Share risk estimates and reasoning behind these estimates	Sensitive risk estimates (or reasoning) involving proprietary data.
Risk Monitoring	Joint risk monitoring (UK central-risk function and AIO A2 Unit)	Align risk monitoring system to be interoperable and complementary	Share results from domestic risk monitoring	Avoid sharing sensitive risks found in risk monitoring (or include proprietary information)
Incident Reporting	Develop common AI incident reporting registry	Align "incident" definitions" and create interoperable reporting standards	Share incidents when reported to domestic incident reporting	Separately monitor highly sensitive incidents, e.g. cybersecurity incidents at AI companies.
Risk Thresholds	Jointly research how to measure and set thresholds for risks from an AI model	Align methodology to set risk thresholds and measure risk thresholds for mutual risk measurement recognition	Share methodology to measure risks and related research	Tailor risk thresholds according to domestic priorities and risk tolerance
Risk Response	Joint high-level risk response frameworks on transnational risks.	Align risk response frameworks	Share resources and research helping with the development of risk response frameworks	Adapt the specific risk response framework to domestic institutions and regulations
Research on Verification Methods	Jointly develop verification methods	Coordinate on research agendas to avoid unnecessary duplication.	Share verification research and privacy- preserving tools	Individually conduct verification research on sensitive security issues.
Research on Privacy Preserving Methods	Jointly develop privacy-preserving methods, e.g for audits	Coordinate on research agendas to avoid unnecessary duplication, e.g. for compute usage	Share privacy-preserving tools, e.g. on AI generated content	Individually conduct research on privacy preserving tools on sensitive security issues
Building Research Talent Pool	Jointly fund talent development or facilitate talent exchanges	Coordinate strategies to build research talent pools and identify current talent gaps.	Share insights into specific talent needs.	Uphold competition in attracting talent and individually build talent pools for domestic priorities.

Common activities	**Collaboration**	**Coordination**	**Communication**	**Separation**
Standards	Harmonization of AI safety standards	Coordinate support in the application of AI safety standards	Share insights on respective standardisation activities	Explanatory guidance re. respective AI safety standards and policies
International engagement	Find consensus on common int. policy goals and jointly prepare for AI summits	Joint meetings to coordinate shared international goals and discuss approaches for how to reach them	Regular discussions on the international AI Governance Landscape	Tailored independent agreements with other countries (E.g. M.o.U of UK AISI with US AISI and Singapore AISI)
Institutional Development		Coordinate institutional development and high-level strategies	Regularly discuss institutional learnings and high-level strategies	
Affirmative Assurance Approaches	Develop shared standards for Assurance Approaches or write a "Crosswalk"	Coordinate mutually beneficial research agendas on Affirmative Assurance Approaches	Share best practices on safety cases (UK) or safety and security reports (EU)	Tailor specific requirements of Assurance Approaches to domestic priorities

1. Introduction

The European AI Office (AI Office) and the UK AI Security Institute (UK AISI) are separate institutions in different jurisdictions. The AI Office focuses on monitoring and supporting implementation of the AI Act, particularly for general-purpose AI systems with systemic risks, while the UK AISI aims to equip governments with an empirical understanding of AI safety. However, they have overlapping mandates, including monitoring the emergence of risks from general-purpose AI and developing tools for evaluations (see Table 1). Further, they face similar challenges navigating these mandates in an extremely complex and rapidly changing international AI policy environment enmeshed with national security and economic interests. The recent AI Action Summit in Paris is a point at hand: several new AISIs were established, including in China, India, France and Chile, while talks of safety were silenced; US-EU tensions over tech regulation heightened; EU announced investment of 200 billions in AI in extension the recently announced competitiveness compass; and the UK and the US abstained from signing the summit statement.

National frameworks alone are insufficient to tackle the shared security and safety risks that advanced AI presents. Given the state of these risks there is an urgent need for optimizing the effectiveness of the two institutions. Both institutions can achieve more of their shared goals and overlapping mandates by conducting their activities at an appropriate engagement level - collaboration, coordination, communication and separation. In this context, it is important to remember that the institutions are hardly one year old, so they remain institutionally flexible (as demonstrated by the recent re-naming of the AISI). This flexibility creates an opportunity for building-in **strategic linkages with reliable counterparts** to strengthen the shared capacity of the institutions. Such efforts would build on precedents of fruitful EU-UK collaboration from other areas, such as digital technical standards.

The relationship between the two institutions is increasingly a topic of interest in policy and think tank conversations in both Brussels and London, but to date is comparatively understudied. Previous work, like Mökander et al., has explored the potential benefits of EU-UK general strategic collaboration on AI. Dennis et al. proposed four conditions under which internationalisation in AI governance is most appropriate, providing a number of relevant considerations to this discussion. But so far, an overview of potential synergies - which could inspire more specific, detailed proposals - is missing.

This policy brief fills this gap by providing an outline of potential areas for collaboration, coordination, communication and separation between the UK AISI and the EU AI Office. The framework can further provide a model for other regional strategic arrangements within the broader international network of AISIs based on common regional and strategic interests or existing connections. It is informed by the stated purposes and functions of the institution and the external conditions under which different levels of engagement seem most appropriate.

We also take into account less formalised attributes of both institutions. This includes past engagements: For example, the UK AISI has carried out evaluations on 16 models, published a repository of LLM benchmark evaluations, run a bounty programme, and conducted joint pre-deployment evaluation with the U.S. AISI. The EU AI Office has not carried out technical evaluations (it is still in the process of filling the role as Lead Scientific Advisor). However, the office has facilitated a multi-stakeholder process with 1000+ stakeholders led by 13 independent (vice-)chairs from academia producing a draft Code of Practice that guides model providers in complying with the General Purpose AI section of the AI Act. Both institutions have contributed to shared projects like the International AI Safety Report and the AISI network.

Table 1: Comparisons of Functions

	UK AISI	**EU AI Office**
Monitoring and supporting implementation and application of rules	No	Yes
Developing tools, methodologies and benchmarks for evaluating general-purpose AI capabilities	Focus on most advanced current capabilities	Focus on general-purpose AI and general-purpose AI with systemic risks
Evaluating capabilities	Yes (demonstrated)	Yes (expected)
Monitoring the emergence of unforeseen general-purpose AI risks	Yes	Yes
Investigating possible infringements	No	Yes
Engagement with AISI network	Yes	Yes

Sources: Commission Decision of 24 January 2024 establishing hte European Artificial Intelligence Office (C/2024/390), Tasks of the AI Office, AISI about and Understanding the First Wave of AI Safety Institutes.

1. Collaboration

Conditions for collaboration: a collaborative approach is ideal when the issues at hand present minimal sensitivity and no information security concern; when no single institution has the capacity to resolve a shared challenge independently; when there is heightened urgency; when challenges can be resolved quicker or better collaboratively; or when achieving seamless interoperability is a top priority. Collaboration also seems advisable when duplication of work is not critical—particularly if a robust body of evidence already exists to guide a specific effort.

Avenues for Collaboration: The UK AISI and EU AI Office could, for example, collaborate on standards and international engagement to pool resources, avoid duplication, and amplify the influence of standards bodies like CEN-CENELEC JTC 21. Given the current geopolitical environment and the increasing race dynamics between global powers like the United States and China, bundling international influence and collaborative efforts between strategically aligned institutions could become a crucial way to ensure balanced standards consistent with European values. Such efforts could take inspiration from EU-UK collaboration on shared commitments to inclusive, multi-stakeholder approaches to standardisation. Joint efforts could improve contributions to global forums, such as AI Summits or the upcoming AI Standards Hub Global Summit, by co-developing shared goals, issuing joint statements, and preparing unified strategies.

The UK and the EU already both signed the Council of Europe's first ever global treaty on AI; participated in the inaugural meeting of the International Network of AI Safety Institutes in San Francisco; and collectively contributed to the G7 Hiroshima International Code of Conduct and the International AI Safety Report. It now seems appropriate for both bodies to prioritise work on the concrete implications of these discussions, as these international policies could shape outcomes both within and beyond domestic borders. Fragmented approaches bear a higher risk of regulatory arbitrage and national frameworks alone are insufficient to tackle AI safety and security risks with cross-border impacts within the scope of the mandates of the institutions. Furthermore, while internationally applicable standards could take a significant amount of time to develop, they could become important relatively quickly given the speed of AI progress. This increases the importance of efficiency and a reduction of unnecessary duplication.

Joint testing represents another promising area for collaboration once the technical unit in the AI Office is in place. This would build on precedents like the UK joint pre-deployment evaluations with Japan, Singapore and the US. While some evaluation work may be better suited for coordination (see below), collaborative testing of advanced AI systems can leverage combined expertise and resources while establishing shared methodologies.

A further fruitful avenue is collaboration around building expertise. The AISI and EU AI Office could co-fund researchers to spend time rotating between their offices to contribute to better (and interpersonal) connections in addition to institutionalised knowledge- and expertise-sharing. This could also involve the broader AISI network.

2. Coordination

Conditions for coordination: Institutions should coordinate their efforts when interoperability of processes remains important, and when leveraging one institution's specialised capabilities would significantly advance the shared objective. Coordination is recommended in cases where information sensitivity and cyber- and national security risks are low, and where complementary skill sets can be deployed efficiently without fully harmonising each institution's processes.

Avenues for coordination: One example of an area suitable for coordination between the UK AISI and EU AI Office would be the work on developing, conducting and interpreting model evaluations. This could involve developing consistent ways to measure and report on model and system performance, establishing clear policies on who evaluates which systems, and holding periodic meetings to interpret results.

Interoperability of evaluation results is essential for ensuring findings from different institutions are compatible and actionable across jurisdictions. Standardised methods could enable easier comparisons and foster a shared understanding of the risks of AI systems and capabilities. A starting point here could be a specific Memorandum of Understanding between the two institutions.

Given the breadth of work on evaluations needed across fields and the shared, cross-border risks associated with unsafe models, coordinating efforts could prevent unnecessary duplication and enable specialisation. Dividing responsibilities might allow each institution to focus on its strengths. While both institutions may want to conduct comprehensive evaluations across domains, they could coordinate their research priorities. The UK AISI, with its focus on foundational science and technical safety research, might lead research into novel evaluation methodologies, frontier capabilities and open source evaluation infrastructure. Meanwhile, the EU AI Office could prioritize research into standardizing evaluation frameworks, integrating them into regulatory processes, and supporting an ecosystem of third-party evaluation developers. In particular, the AISI has made great progress on Chemical-Bio-Radiological-Nuclear (CBRN) evaluations which could play an important role for the AI Office and for providers demonstrating compliance with the EU AI Act. If both parties equally benefit from an activity and specialisation would ensure a more efficient outcome, this approach could make the overall evaluation process more efficient while ensuring broader coverage. These conditions aren't always met: If for example results, tools, or expertise are only relevant in the country or context they are developed in or influenced by the unique political priorities of one party, the other is at a significant disadvantage.

Even where certain evaluations, such as those involving sensitive national security issues, might not lend themselves to coordination, aligning on less sensitive areas of those evaluations could still bring substantial benefits.

3. Communication

Conditions for communication: Institutions should communicate their efforts when both parties benefit from establishing a shared body of knowledge or evidence, and when there are minimal information sensitivity or security issues. In this scenario, maintaining open lines of communication suffices without the need for deeper organisational integration.

Avenues for communication: Two areas where communication may be the optimal level of engagement are risk monitoring and incident reporting. The scope of risk monitoring differs for the two institutions. The AI Office is tasked with the implementation of the AI Act, in particular monitoring compliance with the AI Act and investigating potential infringements with regards to general-purpose AI (with systemic risks). In contrast, the UK AISI has the mission to ‘equip governments with empirical understanding of AI safety’. This may lead to different focuses in risk monitoring. Further, the institutions may have different definitions of what qualifies as an incident or have different structures for incident reporting, for example based on frameworks borrowed from other sector-specific agencies within their jurisdiction. This may also result from the difference between the ‘horizontal’ regulatory approach taken by the EU and the ‘vertical’ regulatory approach so far applied by the UK.

In this context, it could be valuable to establish effective and consistent communication channels between the institutions for sharing incidents reported to domestic systems. Cross-border sharing may broaden the awareness of the breadth of risks as well as help identify trends in incidents, including cross-border incidents. This can increase preparedness for both institutions. Such efforts could also be valuable in informing the pending UK AI Bill and updates to the EU AI Act. Further, sharing and comparing incident reports and methods can be essential in identifying divergences and best practices in risk monitoring and incident reporting. Communication could be supplemented with a guide for interpretation to avoid confusion and identify divergences in definitions, scope, and other matters. Such a guide would also be valuable in the context of the wider network of AISIs, where there may be diverging definitions and understanding of concepts including those of ‘risks’ and ‘harm’.

4. Separation

A separation of key policy areas, even in overlapping mandates, would be important under three potential conditions.

Firstly, some activities of both the EU AI Office and the UK AISI involve sensitive or proprietary information. Pre-deployment evaluations on sensitive national security threats, would be an example. Sharing this information with a larger group than necessary could increase its attack surface and thus threaten its confidentiality.

Secondly, resources like time, talent and funding need to be considered. If collaboration, coordination or communication is simply too resource intense on a particular workstream and seems to offer limited value to the other party, it should most likely not be pursued. Agreements that take these considerations into account and state clear expectations of the other party help prevent a waste of resources on unnecessary collaboration, coordination or communication.

Thirdly, tailored separate strategies are appropriate when the long-term priorities or near-term tactics of the two institutions differ. The UK might want to set different risk thresholds for localised risk than the EU. Or they might want to pursue a different approach to a particular international relationship. For example, the UK AISI has formed a close bilateral collaborative relationship with the US AISI and abstained from signing the Statement on Inclusive and Sustainable AI along with the US. In contrast, the EU AI Office has primarily engaged with the US AISI through "technical dialogues" and the multilateral inaugural meeting of the AISI Network, and many EU Member States signed the Statement. These aren't scientific assessments but political decisions which both institutions can tailor to their own priorities and goals.

5. Conclusion

There are ample avenues for the EU AI Office and the UK AISI to effectively address their overlapping tasks while respecting their institutional and jurisdictional differences. We recommend policymakers treat each area on a case by case basis, and that the two institutions collaborate, coordinate, communicate and separate their work accordingly. Our hope is that this overview inspires more detailed proposals and momentum for the shared goals of these important institutions. Lastly, hopefully the conditions and examples outlined in this framework can be useful for identifying synergies in the wider context of the growing network of AISIs.

Further Reading

- Exploring EU-UK Collaboration on AI: A Strategic Agenda
- Understanding the First Wave of AI Safety Institutes: Characteristics, Functions, and Challenges
- Getting the UK's Legislative Strategy for AI Right
- Council of Europe opens first ever global treaty on AI for signature - Portal
- The AI Safety Institute Network: Who, What and How? - ICFG
- Conference on frontier AI safety frameworks | AISI Work
- The AI Safety Institute International Network: Next Steps and Recommendations

About the Authors

Lara Thurnherr* is an independent AI governance researcher and MA student in Cyber strategy and Policy at King's College London, focusing on security-transparency trade-offs and international AI governance.

Risto Uuk* is the Head of EU Policy and Research at the Future of Life Institute and a PhD Researcher at KU Leuven on the assessment and mitigation of systemic risks posed by general-purpose AI. He also runs the biweekly EU AI Act Newsletter with over 40,000 subscribers.

Tekla Emborg* is the EU Policy Research Fellow at the Future of Life Institute, focusing on implementation and enforcement of the EU AI Act, European AI liability, and systemic risks from general-purpose AI.

Marta Ziosi* is a Postdoctoral Researcher at the Oxford Martin AI Governance Initiative (AIGI). Her work focuses on standards for advanced AI systems and on AI policy and governance more broadly.

Charles Martinet is Head of Policy at the French Center for AI Safety (CeSIA), a Research Affiliate at the Oxford Martin AI Governance Initiative, and independent expert in the EU GPAI Code of Practice process. His work, focused on European and international AI governance, has been published by the OECD AI Policy Observatory, Euractiv, and the German Marshall Fund of the US, among others.

Isabella Wilkinson is a Research Fellow in Chatham House's Digital Society Programme.

Morgan Simpson is a research manager at Pivotal Research, and a research management consultant for the cambridge ERA fellowship.

Renan Araujo is a Research Manager at the Institute for AI Policy and Strategy, where he leads the international governance workstream, a fellow at the Oxford China Policy Lab, and a Research Affiliate with the Oxford Martin AI Governance Initiative.