EUROPE

Biosecurity Governance Across Uncertain Artificial Intelligence Futures

Perspectives from a Side Event on Biosecurity and Frontier AI at the 2025 AI Action Summit

Aurelia Attal-Juncqua, Casey Mahoney, Sana Zakaria, Paul Khullar, Patricia Paskov, Ella Guest, Michael Simpson

For more information on this publication, visit www.rand.org/t/CFA4186-1

About RAND Europe
RAND Europe is a not-for-profit research organisation that helps improve policy and decision making through research and analysis. To learn more about RAND Europe, visit www.randeurope.org.

Research Integrity
Our mission to help improve policy and decision making through research and analysis is enabled through our core values of quality and objectivity and our unwavering commitment to the highest level of integrity and ethical behaviour. To help ensure our research and analysis are rigorous, objective, and nonpartisan, we subject our research publications to a robust and exacting quality-assurance process; avoid both the appearance and reality of financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursue transparency in our research engagements through our commitment to the open publication of our research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. For more information, visit www.rand.org/about/research-integrity.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

© 2025 RAND Corporation

RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights
This publication and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorised posting of this publication online is prohibited; linking directly to its webpage on rand.org is encouraged. Permission is required from RAND to reproduce, or reuse in another form, any of its research products for commercial purposes. For information on reprint and reuse permissions, please visit www.rand.org/pubs/permissions.

About This Conference Proceeding

Advances in artificial intelligence (AI) offer tremendous potential benefits for science and medicine, but also introduce new risks which must be thoughtfully managed to safeguard global health and security.

RAND partnered with the Nuclear Threat Initiative (NTI) to co-convene a workshop on the sidelines of the AI Action Summit in Paris on 11 February 2025 in order to assess the perspectives of subject-matter experts on emerging challenges at the nexus of AI and biology, and to identify solutions. The event, 'AI Safety Institutes and AIxBio Governance: A Discussion with Biosecurity Experts and AI Model Developers', brought together policymakers, technical experts, and public- and private-sector stakeholders from advanced industrial economies. This report summarises key insights from the workshop and has been prepared by RAND's Technology and Security Policy Center and Frontiers of Technology Hub.

Technology and Security Policy Center

RAND Global and Emerging Risks is a division of RAND which delivers rigorous and objective public policy research on the most consequential challenges to civilisation and global security. The division houses the Technology and Security Policy Center, which explores how high-consequence, dual-use technologies change the global competition and threat environment, then develops policy and technology options to advance the security of the United States, its allies and partners, and the world. For more information, contact tasp@rand.org.

Frontiers of Technology Hub at RAND Europe

RAND Europe is a not-for-profit research organisation which helps improve policy and decision making through research and analysis. RAND Europe houses the Frontiers of Technology Hub, which is dedicated to assessing emerging developments in technologies ranging from AI and synthetic biology to quantum, delivering independent research and actionable insights to help policymakers and industry seize opportunities and manage risks. To learn more about RAND Europe, visit www.randeurope.org.

Funding

This work was conducted using income from operations and gifts from philanthropic supporters, which have been made or recommended by DALHAP Investments Ltd, Effektiv Spenden, Ergo Impact, Founders Pledge, Charlottes och Fredriks Stiftelse, Good Ventures, Jaan Tallinn, Longview, Open Philanthropy, and

Waking Up Foundation. A complete list of donors and funders is available at www.rand.org/TASP. RAND donors and grantors have no influence over research findings or recommendations.

Acknowledgments

The authors would like to thank Luwi Luwaya for helping to plan the workshop and would also like to thank Greg McKelvey for assisting from the project's inception. Bryan Boling, Brodi Kotila, Casey Dugan, James Black and Toby Webster provided insightful feedback on the document, which greatly improved its quality. Finally, the authors would like to thank the Nuclear Threat Initiative for their partnership in this effort.

Executive Summary

Issue

Rapid developments in AI and biological technologies constitute a substantial source of uncertainty for governance systems intended to preserve biosecurity in the context of national security and human development. Such developments offer potentially transformative benefits to human health but also create risk vectors for the development and proliferation of deadly biological agents capable of harming global society and national security to an unprecedented scale. Existing international norms and treaties, such as the Biological Weapons Convention, were developed long before the emergence of modern AI capabilities and are not well suited to address the unique challenges posed by rapidly advancing AI-enabled threats. The scientific and policy communities currently lack reliable data on how frontier AI models, including both general-purpose systems and specialised biotechnology design tools, are spreading globally. Effective defence against biological threats requires leaders to have reliable information about where the most dangerous AI systems are accessible and how quickly these tools could enable the creation and dissemination of harmful viruses, toxins or other biological agents.

Given the urgency of this policy challenge, in 2024 a network of nearly a dozen AI safety and security institutes across national governments, along with a growing number of private firms, agreed to focus on these risks and pursue measures to contain them.[1] Despite facing a range of obstacles[2] which have slowed global progress in coordinating biosecurity policy responses, industrial leaders and governments still have the potential to strengthen health systems and enhance their ability to respond to biohazards.

Approach

RAND partnered with the Nuclear Threat Initiative (NTI) to co-convene a workshop on the sidelines of the AI Action Summit in Paris on 11 February 2025 in order to assess the perspectives of subject-matter experts on emerging AI-biosecurity challenges, and to identify solutions. The event, 'AI Safety Institutes and AIxBio Governance: A Discussion with Biosecurity Experts and AI Model Developers', brought together policymakers, technical experts, and public- and private-sector stakeholders from advanced

[1] NIST 2024

[2] These include misaligned stakeholder incentives around information sharing, unclear risk threshold red lines and fast-paced changes in AI technology. Moreover, some frontier AI models are beginning to match – or, in limited cases, exceed – expert human performance on certain text-based virology tasks. For example, in a recent evaluation, one such model achieved over 30 per cent accuracy on structured prompts involving laboratory procedures, compared to 22 per cent for PhD-level virologists under the same conditions. If misused, such tools could enable catastrophic damage to biosecurity. Sources: Götting et al. 2025, Hendrycks & Hiscott 2025.

industrial economies. This publication summarises key insights from that workshop. RAND and NTI jointly designed the agenda, selected participants and implemented the event. RAND also administered a pre-workshop questionnaire whose results were synthesised by the RAND team following the event.

Workshop outcomes

Workshop participants identified practical steps that stakeholders can take to reduce risks, close knowledge gaps, and build more resilient systems. Key takeaways included:

- **Focus on measurable risk, not just broad scenarios**. Participants recommended shifting from broad, hypothetical scenarios towards specific, measurable risk indicators, such as the ways in which AI tools could reduce the technical or resource barriers to misusing biological knowledge or materials. This emphasis aligns with their support for 'if–then' hazard-threshold frameworks, which tie concrete triggers to predefined responses and allow for timely mitigation even when full risk quantification is impractical.
- **Invest in risk assessment and testing.** Participants called for collaborative, cross-institutional methods to measure and assess AI's impact on biosecurity. This includes safe, controlled 'real-world' proxy environments for evaluating misuse-relevant AI pathways, without handling dangerous biological agents, and shared approaches for testing both general-purpose AI systems and specialised biotechnology design tools against misuse scenarios affecting healthcare delivery or public infrastructure.
- **Strengthen policy adaptability.** Experts emphasised the need for governance systems that can evolve alongside fast-moving technologies. This includes creating feedback loops between researchers, policymakers and civil society to update safeguards as risks change.
- **Promote cross-sector coordination and information sharing.** Participants agreed that improving communication between the AI and bio communities, supported by privacy-preserving information-sharing tools, is key to effective early warning and response.
- **Support innovation which advances safety.** The group discussed how technical solutions – such as access controls, secure compute environments and auditing systems – can be co-developed to reinforce biosafety norms.
- **Keep collaboration going.** Participants found the RAND-NTI partnership valuable in highlighting diverse perspectives and urged continued efforts to build bridges across sectors and sustain global attention concerning AI-related biosecurity risks.

Table of Contents

About This Conference Proceeding i
Executive Summary iii
Table of Contents v
Figures and Boxes vi
1. Introduction and Workshop Approach 1
1.1. Workshop objectives, participants and agenda 1
1.2. Pre-workshop questionnaire 2
2. AIxBio Status Quo and Future Challenges: Workshop Opening Plenary 3
2.1. Summary of findings from the pre-workshop questionnaire 3
2.2. Current landscape of AIxBio convergence: insights from the opening plenary 7
2.3. A path forward: conclusion from the opening plenary 8
3. Workshop Session 1: Risk Thresholds and Mitigations 10
3.1. Panellist remarks: understanding and responding to biological risk 10
3.2. Open-floor discussions: envisioning risk-mitigation futures 11
4. Workshop Session 2: Research Agendas and Anticipatory Approaches 14
4.1. Panellist remarks: objectives and process for research progress 14
4.2. Open-floor discussions: shaping future research priorities 16
5. Workshop Session 3: Collaboration, Resource Sharing and Best Practices 18
5.1. Panellist remarks: the state of AISI-centric governance 18
5.2. Open-floor discussion: collaborating on information sharing 19
6. Key Findings and Recommendations 22
6.1. Key findings 22
6.2. Recommendations made by participants 25
6.3. Conclusion 26
A.3. Pre-workshop questionnaire 29

Figures and Boxes

Figures

Figure 1. Forecasted areas where AI is expected to have the greatest positive impact (red) and participants' reported organisational influence on biosecurity risks (blue) (n=11) 4

Figure 2. Focus areas of personnel at participating organisations according to respondent self-reports (n=14) 5

Figure 3. Preferences for AIxBio collaborator type (n=14) 5

Figure 4. Preferences for AIxBio collaboration tool (n=14) 6

Figure 5. Preferences for AIxBio collaboration format (n=14) 6

Boxes

Box 1. Session 1 agenda text 10

Box 2. Session 2 agenda text 14

Box 3. Session 3 agenda text 18

1. Introduction and Workshop Approach

On 11 February 2025, RAND and NTI | Bio convened a three-hour in-person workshop of leading biosecurity and AI policy experts to examine AI-related risks and opportunities. This report summarises key insights and findings from those discussions.

This chapter outlines the planning process, including agenda development and participant selection, and introduces the pre-workshop questionnaire and the analytical methods used by the RAND team to distil insights from the event.

1.1. Workshop objectives, participants and agenda

The workshop brought together approximately 25 experts from government, industry and research institutions to examine risks and opportunities at the intersection of AI and biosecurity. Designed as a semi-structured dialogue, the workshop aimed to inform ongoing policy conversations and generate early, expert-informed ideas for risk mitigation and governance.

The objectives of the workshop included:

- Building a shared understanding of emerging risks at the convergence of AI and the life sciences.
- Identifying potential thresholds or trigger points where AI capabilities may require additional oversight.
- Exploring opportunities for collaboration between AI safety institutes, biosecurity experts and government agencies.
- Laying groundwork for international cooperation on AI-related biological risks.

Both RAND and NTI have longstanding mandates to examine threats to global health and security.[3] The rapid expansion of AI-related research, deployment and investment presents new challenges for those working to prevent the misuse of biotechnology. While AI systems could support the detection, control and oversight of biological risks, future advances might also lower barriers to dangerous misuse, intentionally or accidentally, by actors with limited resources or oversight.[4] The workshop was designed with this dual-use landscape in mind, aiming to uncover actionable ideas to strengthen global biosecurity governance in light of these trends.

[3] RAND Meselson Center, homepage.

[4] E.g. Mouton et al. 2023.

1.1.1. Participant selection

The workshop brought together a group of experts deliberately chosen to represent a range of sectors, with participants selected to reflect a roughly equal representation from three core domains[5]:

- Industry representatives from major international firms developing frontier AI systems.
- Technical experts in the life sciences at leading research institutes and non-governmental organisations focused on biological threats to global security.
- Officials from across national and international governmental bodies focused on AI (e.g. AI safety institutes, regulators).

While the group included a range of perspectives, all participants were affiliated with organisations based in advanced industrial economies. This may limit the generalisability of some findings and discussions included in this report.

1.1.2. Agenda setting

The RAND–NTI team developed a focused agenda to support in-depth discussion of strategic policy challenges. The three main sessions were designed to explore:

1. 'AIxBio risk thresholds and mitigations.'
2. 'Strategic research agenda: exploring a comprehensive approach to reducing anticipated AIxBio risks.'
3. 'Opportunities for collaboration to safeguard AIxBio capabilities: sharing resources and best practices.'

Each participant received the agenda and pre-reading materials before the event, including a short questionnaire to explore early insights and identify common priorities.

1.2. Pre-workshop questionnaire

To support agenda development and ground the discussion in participant experience, RAND circulated a short pre-workshop questionnaire. The questionnaire focused on how participants and their organisations are engaging with AI-related biosecurity issues, including self-regulatory efforts and unmet needs. Just over half of the participants completed the questionnaire and results were used to inform the opening session and discussion framing. A full discussion of the questionnaire design, administration and analytic approach is provided in Appendix A.

[5] See the 'Participant selection' section in Appendix A for more details on the participation selection process, including potential limitations.

2. AIxBio Status Quo and Future Challenges: Workshop Opening Plenary

During the opening plenary, the session co-chairs presented preliminary findings from the pre-workshop questionnaire. This aimed to frame the subsequent discussions by outlining:

- The distribution of participants' interests across AI- and biosecurity-related risks and benefits.
- The biosecurity and life-science domains in which participants assessed AI to pose the greatest risks and/or offer the greatest benefits.
- The types of resources, collaborators and modalities participants were most interested in discussing during the workshop.

This chapter first presents key themes from the pre-workshop questionnaire and then offers a summary of the opening plenary session that followed.

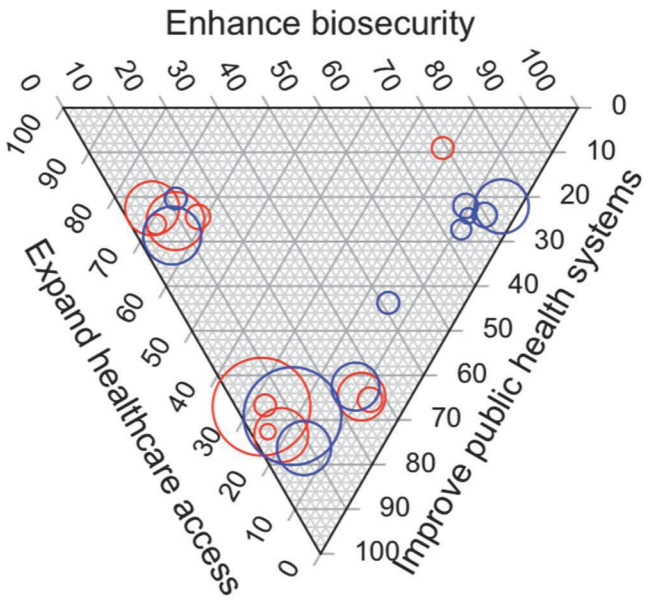
2.1. Summary of findings from the pre-workshop questionnaire

The pre-workshop survey received responses from 14 participants representing AI firms, governments and other organisations. Four key themes emerged from this initial analysis (see Appendix A), which were shared during the opening session to shape the discussion. These helped highlight early areas of alignment and tension, particularly around trade-offs between biological proliferation and control, the implications of AI adoption, and governance challenges.

Questionnaire Theme 1. Views concerning what future benefits and risks AI will bring – including for biosecurity – vary vastly among participants.

The questionnaire asked participants to rank how AI, if fully leveraged in the life sciences, might improve biological security, public health or healthcare access. They were then asked to rank how their organisation's current work would likely contribute to these areas. Figure 1 displays these responses using a ternary plot, which visualises how participants prioritised expected outcomes. Most responses are clustered near the corners, indicating that participants saw some outcomes as clearly more likely or relevant than others, rather than assigning equal weight to all three.

Figure 1. Forecasted areas where AI is expected to have the greatest positive impact (red) and participants' reported organisational influence on biosecurity risks (blue) (n=11)



Note: Ternary coordinates derived from a transformation of the rank-ordered set for each respondent (e.g. [1,2,3) or [3,1,2) or [2,3,1), etc.). The plot adds a random jitter to each coordinate set for visual clarity.

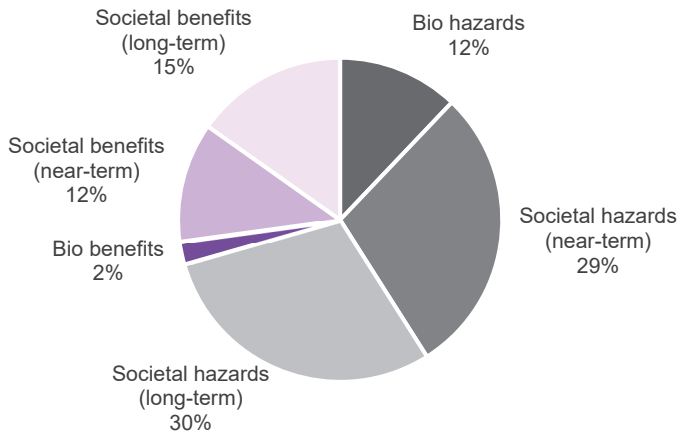
Source: RAND questionnaire

Specifically, each red circle in Figure 3 represents a respondent's view of which life-science, biosecurity, public health or healthcare-access domain AI is most likely to benefit. Each blue circle shows where they believe their organisation's work is most likely to promote such benefits. Circle size reflects the number of staff working on AI and biosecurity at their organisation. The distribution of circle sizes suggests that views do not depend heavily on organisation size. More importantly, the red and blue circles often do not align, indicating that respondents believe their organisations are actively influencing outcomes. While many think AI will naturally benefit public health or healthcare access, participants tended to see their own organisations as more connected to biosecurity efforts. The clustering of blue circles near the 'biosecurity' and 'public health' axes reinforces this, while only two respondents indicated that their work primarily supports healthcare access.

Questionnaire Theme 2. Participants report focusing on the hazards AI poses to biosecurity substantially more than on the potential benefits.

The questionnaire asked about staffing levels for teams focused on AI and life science risks versus benefits. On average, respondents reported a 2:1 ratio of personnel working on risks compared to those focused on benefits within their respective organisations (see Figure 2).

Figure 2. Focus areas of personnel at participating organisations according to respondent self-reports (n=14)



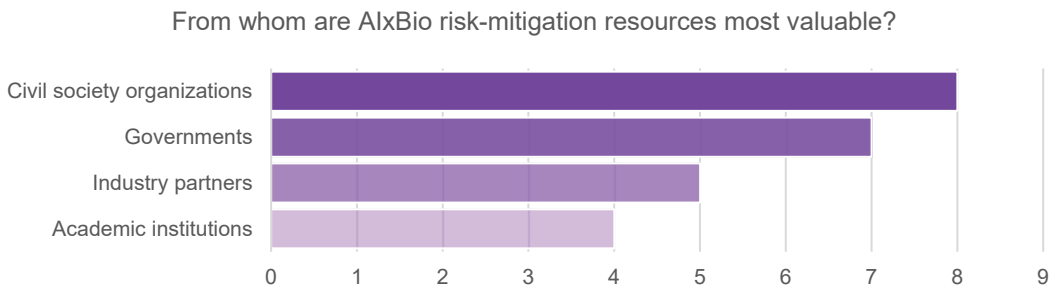
Source: RAND questionnaire

It is worth noting that this result likely reflects NTI and RAND's decision to centre the workshop on safety risks and invite experts from slightly more hazard-focused organisations. As a result, the RAND team anticipated that the workshop discussion might lean more heavily towards biohazard management than one involving a more balanced mix of risk- and benefit-focused perspectives

Questionnaire Theme 3. Public and private stakeholders in AI and biosecurity sectors require a broad range of tools for their work to reduce risks.

In addition to collaboration formats (see below), the questionnaire also asked about preferred types of collaborators (Figure 3) and the tools needed to support such work (Figure 4). Responses aligned with the workshop team's expectations: stakeholders in AI-bio risk mitigation see progress in policy, technical innovation and cross-sector coordination as key to success. Participants expressed interest in a wide range of tools, collaboration models and partners. It is worth noting that while most respondents showed broad interest in cross-sector collaboration, only one of six government and AI industry respondents expressed interest in working with civil society, unlike respondents from academia or NGOs, who were more open to such partnerships.

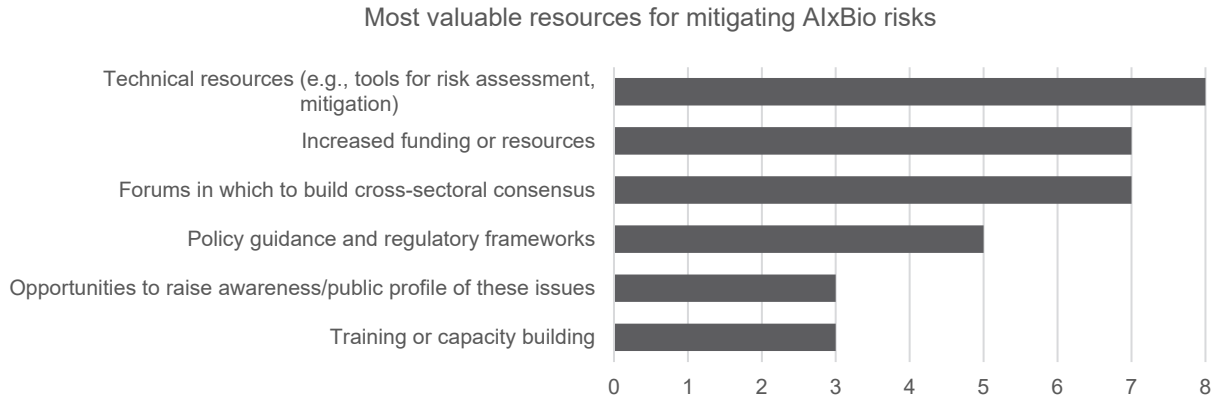
Figure 3. Preferences for AIxBio collaborator type (n=14)



Note: Multiple-selection item allowed respondents to select more than one option

Source: RAND questionnaire

Figure 4. Preferences for AIxBio collaboration tool (n=14)



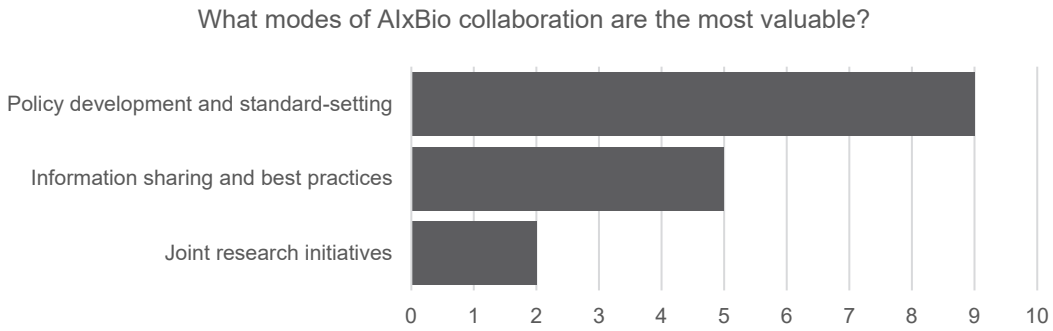
Note: Multiple-selection item allowed respondents to select more than one option

Source: RAND questionnaire

Questionnaire Theme 4. Joint research initiatives were seen by respondents as the costliest (or least beneficial) form of collaboration to address these hazards.

The questionnaire asked about preferred forms of collaboration on AI and biosecurity risks. Most respondents favoured 'policy development and standard-setting' (9 of 14), followed by 'information sharing and best practices' (5 of 14), while only two preferred joint research (Figure 5). Future research could explore whether this reflects the group's risk focus or broader structural barriers to cross-institutional research in this space.

Figure 5. Preferences for AIxBio collaboration format (n=14)



Note: Multiple-selection item allowed respondents to select more than one option

Source: RAND questionnaire

In sum, these themes were presented at the outset of the workshop to help frame discussion and focus attention on key areas of agreement and disagreement. Responses suggested that many of the organisations represented were actively engaged in addressing emerging risks from AI in the life sciences. This aligned with the RAND team's expectation that participants would see their roles as directing AI-related innovation towards biosecurity priorities which might otherwise be neglected. At the same time, participants

highlighted significant unmet needs and a strong demand for future work, particularly cross-sector partnerships involving civil society, academia, government and AI developers. The most sought-after forms of collaboration included policy development, standard-setting, joint research funding and forums for information sharing.

2.2. Current landscape of AIxBio convergence: insights from the opening plenary

Rapid developments in AI and biological technologies constitute substantial sources of uncertainty for governance systems intended to preserve biosecurity in the context of national security and human development. Such developments offer potentially transformative benefits to human health but also create risk vectors for the development and proliferation of deadly biological agents capable of harming global society and national security to an unprecedented scale.

Opening remarks noted that some frontier AI models are beginning to match – or, in limited cases, exceed – expert human performance on certain text-based virology tasks. For example, in a recent evaluation one such model achieved over 30 per cent accuracy on structured prompts involving laboratory procedures, compared to 22 per cent for PhD-level virologists under the same conditions.[6] These results suggest that AI systems may increasingly be able to support the steps involved in both developing biological agents and designing countermeasures. This capability leap raises fundamental questions about oversight and security.

2.2.1. Technical evaluation challenges

The opening presentation emphasised that technical resources, such as tools for risk assessment and mitigation, are valuable for addressing AIxBio risks, a viewpoint the questionnaire results served to validate. However, participants suggested that current technical evaluation methodologies face significant limitations. While benchmarks exist, they often fail to capture the nuanced capabilities that could present biosecurity risks. Furthermore, the relationship between benchmark performance and real-world capabilities and risks remains poorly understood, complicating efforts to establish meaningful risk thresholds.

2.2.2. Stakeholder roles and resource allocation

AIxBio risk-mitigation efforts require stakeholders to play a broad range of roles and to bring varied levels and types of resources to their work. Questionnaire respondents expressed the view that global civil society organisations provide substantial value to AIxBio risk mitigation. Responses also indicated that closer collaboration between government and industry could help bridge gaps in understanding and improve coordination on risk-mitigation strategies related to AI model development and deployment.

As outlined above, the participants and organisations represented in the workshop were skewed towards those focused on the potential hazards of AI, particularly in the context of biosecurity, rather than its potential benefits. This imbalance likely reflects the nature of the convening and the expertise it prioritised. However, it also introduces a limitation: discussions may have underemphasised the ways AI could be used

[6] Götting et al. 2025.

to strengthen biosecurity or mitigate the very risks it poses. This raises the broader question of whether current AIxBio policy discourse on safety and security includes sufficient focus on beneficial applications, especially those aimed at building resilience, detection or response capacities.

2.2.3. Cross-sector collaboration barriers

Increasing collaborative policy development appears to be a key action needed to safeguard value in the biosecurity and AI sectors. Nearly all questionnaire respondents indicated a need for policy-development collaboration. Focusing on such efforts is particularly important because of the persistent challenges to cross-sector collaboration. Presenters identified several substantial barriers including:

- Varying and often lengthy clearance processes across governments.
- Privacy concerns, combined with a lack of strong privacy-preserving tools for sharing models, often lead developers to restrict outside access. This limits how thoroughly external collaborators can assess the risks of those models.
- Competing policy priorities.
- Lack of standardised information-sharing protocols.

Other participants also highlighted less frequently discussed challenges including how to keep biosecurity issues visible on global, national security and governance agendas, as well as the need for stronger personnel training and human-capital development strategies.

2.3. A path forward: conclusion from the opening plenary

The opening remarks concluded that the top-priority tasks for mitigating AI- and biosecurity-related risks must include continued attempts to:

- Develop technical tools for risk assessment and mitigation.
- Identify funding opportunities, share them with others, and work together to secure access to them.
- Create forums for multi-sectoral dialogue.

The presentation ended by suggesting consensus in AIxBio communities may be emerging around the three following areas:

1. **Recognition of tangible risks.** There is broad agreement that AI advances in biology present concrete, near-term risks which require immediate attention, even if the diversity of these risks makes them difficult to prioritise or establish as red lines.
2. **Value of the 'if-then' framework.** Stakeholders broadly support using capability-based thresholds to trigger specific mitigation actions, though implementation remains challenging.
3. **Need for dedicated institutions.** The recent and ongoing establishment of AI Safety/Security Institutes (AISIs) in multiple countries represents important progress, though coordination mechanisms remain underdeveloped.

Innovation in governance may require greater transparency in sharing risk thresholds/red lines among stakeholders than current practice incentivises. It may also require harmonising different roadmaps used by

governments, industry and research institutions for assessing capabilities and risks. However, by identifying and articulating the barriers to efficient policy collaboration, presenters agreed that the community can create opportunities to address them and promote greater innovation through governance that is more reliable and trusted.

3. Workshop Session 1: Risk Thresholds and Mitigations

Following the opening presentations, Session 1 of the workshop focused on identifying critical risks, establishing risk thresholds, and exploring strategies for effective mitigation. Panellists and participants discussed the complex landscape of AI and biosecurity, including existing risk assessments and regulatory frameworks, and the role of international collaboration after the moderator introduced the agenda topics in Box 1.

Box 1. Session 1 agenda text AIxBio risk thresholds and mitigations

- How should we prioritise risk assessment and risk reduction for AIxBio capabilities? Which risks are most urgent and which are most important in the long term? What red lines or risk thresholds should guide our efforts to address these risks and how can these thresholds inform responsible scaling policies?
- To what extent can the international community – particularly AISIs, AI model developers and biosecurity experts – achieve consensus on risk thresholds or capability thresholds for AI tools? Has there been sufficient progress in operationalising these commitments since the Seoul discussions?
- What priority actions should be taken if agreed-upon risk thresholds for AIxBio capabilities are crossed? How can the international community ensure a coordinated response to reduce risks effectively?

3.1. Panellist remarks: understanding and responding to biological risk

3.1.1. Risk assessments and thresholds

Panellist 1 presented RAND's AI Capability Evaluations (ACE) programme, highlighting its methods for assessing AI models. The program uses standardised benchmarks, task-based performance assessments, uplift studies and red-team testing. Despite progress, Panellist 1 noted ongoing challenges in mapping evaluation results to real-world risks: one enduring and significant concern is the uncertainty about how AI benchmark performance translates into operational capabilities.

The session also revisited recent findings on AI performance in virology (previously discussed), prompting renewed concern about the implications of models exceeding expert-level performance on sensitive biological tasks. Panellists emphasised the need to understand how such capabilities could translate into operational risk or misuse.

3.1.2. Regulatory frameworks and mitigation strategies

Panellist 2 provided an overview of the EU AI Act, the world's first comprehensive regulatory framework for frontier AI. The act operates across two domains: AI Applications Regulation and General-Purpose AI (GPAI) Regulation. The former targets system-level risks, particularly in high-risk sectors like healthcare, while the latter focuses on systemic risks at the model level.

Panellist 2 explained the flexible definition of systemic risk embedded in the act, which allows for adaptability as technology evolves. However, the panellist highlighted challenges in enforcement, including limited consensus on effective mitigation strategies and the need for technical tools to support regulatory oversight. The EU AI Act gives the EU AI Office significant authority, such as requesting documentation and mitigation measures from GPAI providers, but still struggles with operationalising these powers.

3.1.3. Economic measures and complex risk models

Panellist 3 emphasised that the economic dynamics driving rapid AI development, such as strong financial incentives and compressed timelines, may create pressure to prioritise commercially driven innovation over innovation explicitly aimed at safety or governance. The panellist also highlighted a trend where evaluators are expected to act as both risk assessors and risk managers, potentially leading to conflicts of interest.

The panellist critiqued traditional risk assessment models as overly simplistic and advocated for more complex frameworks which consider the probability of models developing harmful capabilities, potential misuse by threat actors, and the broader consequences of these actions. It was pointed out that current benchmarks inadvertently encode value judgments, further complicating risk assessments. The panellist stressed the urgency of achieving industry-wide consensus on risk-evaluation standards and establishing clear compliance requirements.

3.2. Open-floor discussions: envisioning risk-mitigation futures

Following the panellist presentations, the moderator introduced the questions on the agenda and facilitated open-floor discussions. These focused on building consensus around risk thresholds and improving scenario planning. Participants debated how to balance concise risk definitions with the need for political saliency, acknowledging that policymakers often respond better to compelling narratives than dense technical data.

Participants agreed that waiting for full consensus on AIxBio risks could delay critical actions. Instead, they advocated for identifying lowest-common-denominator agreements to guide immediate mitigation strategies. The 'if-then' approach was highlighted as a valuable framework, though participants noted the circular nature of current debates whereby the lack of solid evaluations stalls progress on defining 'if' conditions and appropriate 'then' responses.

The need for proactive scenario planning was another key theme. Participants emphasised that some mitigation actions – especially 'break-glass' measures[7] – require years of preparation. Examples included

[7] 'Break-glass' measures refer to emergency access protocols which allow authorised personnel to bypass normal access controls to critical systems to access 'sensitive information (e.g. access to all relevant encryption keys)' during urgent situations, such as a cyberattack or system failure. Expanding illegitimate access is a type of attack vector which is relevant to both cyberbiosecurity and the need to secure AI model weights. For more information see Nevo 2024.

pre-emptive guardrail programmes to prevent AI models from collectively enabling bioweapon development.

3.2.1. Future directions

Panellists and participants acknowledged the challenges faced in defining clear risk thresholds, particularly given the varying risk tolerances among stakeholders. The UK's AISI was suggested as a potential forum for these discussions, offering what some in the room considered a neutral ground for converging on acceptable risk levels.

Participants recognised the progress made by frontier safety frameworks (FSFs), especially in setting thresholds around chemical and biological risks. However, less consensus exists regarding specialised technical or governance tools for managing these risks and nuanced dual-use scenarios. The Frontier Model Forum (FMF)[8] was identified as a valuable platform for ongoing discussions, though participants emphasised the need to move beyond binary approaches to risk management – for example, by moving past simple 'safe vs unsafe' categorisations towards more nuanced and graduated frameworks which account for varying levels of risk, uncertainty and context-specific mitigation strategies.

3.2.2. Illustrative quotes and action points

Several participants raised themes that, while anecdotal, reflected recurring sentiments during the discussion:

> The challenge isn't just defining the risks—it's communicating them in ways that policymakers and the public can understand and act on.
>
> We need better evaluations to know when AI models cross critical thresholds, but we also need pre-emptive mitigation strategies. Otherwise, we're stuck in a circular conversation.
>
> The international community must prioritise coordinated responses. We can't afford fragmented strategies when dealing with global biosecurity risks.

These views reflect broader concerns that ran through the session: the need for proactive planning, consensus-building, and transparent communication to manage AIxBio risks effectively. Participants agreed that while robust technical evaluations remain important, equal attention must be paid to practical mitigation strategies and internationally coordinated responses. Action points that emerged from the discussion included:

- Establishing **clear, actionable risk thresholds** which avoid being either too abstract (e.g. 'create a pandemic') or too technically narrow (e.g. 'increase virulence by 30 per cent via methylation of sequence XY').
- **Prioritising scenario-based planning** to anticipate breaches and enable timely mitigation.

[8] An industry-supported non-profit dedicated to advancing frontier AI safety and security.

- **Fostering international collaboration** through venues like the AISI network and FMF in order to align standards for risk assessment and mitigation.

4. Workshop Session 2: Research Agendas and Anticipatory Approaches

Session 2 asked what forms of research coordination among stakeholders might help reduce uncertainty surrounding AI- and biosecurity-related harms. Participants discussed strategies for risk mitigation, identified critical gaps in current frameworks, and examined the roles of key stakeholders in advancing AIxBio safety according to the agenda in Box 2.

Box 2. Session 2 agenda text

Advancing a strategic research agenda: exploring a comprehensive approach to reducing anticipated AIxBio risks

- How can AISIs and other key stakeholders explore a more comprehensive and strategic approach to reducing anticipated AIxbio risks, including from rapid advances in biodesign tools and automation of the life sciences?
- What are the most important elements to include in a research agenda for exploring a range of potentially effective technical guardrails for AIxBio capabilities?
- What role can AISIs play in advancing a comprehensive research agenda on AIxBio risk mitigation and guardrail options? How can this agenda be structured to ensure actionable outcomes and what resources or commitments would be needed to prioritise and implement it effectively?

4.1. Panellist remarks: objectives and process for research progress

4.1.1. Ambitious research goals

The session was framed around NTI's 2023 report on AIxBio convergence.[9] The moderator proposed the need to adopt a holistic approach to risk management across all stages of AI model development. Participants focused on four key areas – data collection, development, pre-release and release – and emphasised the need to control access to training data, implement responsible training methods, integrate built-in safeguards, and manage user access to AI tools.

The moderator opened the discussion by highlighting the urgent need to create ambitious research goals to proactively address potential AIxBio risks. Emphasis was placed on building a comprehensive understanding

[9] NTI | bio 2023.

of the implications of AI in biosecurity, especially considering the rapid advancement of bio-design tools and life-sciences automation.

4.1.2. Evaluating risks and benefits

The first panellist focused on two main bodies of work within biosecurity: large language models (LLMs) and frontier AI models, which have diverse applications, and biology-based models developed by specialised science teams. The speaker highlighted AlphaFold, the Nobel Prize-winning model that predicts protein and other molecular structures, as a key example of AI's transformative potential in biology. However, the speaker also pointed out that while such tools present significant scientific opportunities, they come with complex security and safety risks which require careful evaluation.

A significant portion of the discussion centred around the development of a 'Framework for Unified Harm Assessment', currently underway through the FMF. This framework aims to identify the most concerning actors and pathways for misuse while addressing challenges related to information sharing and integrating national security stakeholders into broader risk-assessment efforts. The speaker stressed that a unified safety framework is essential for developing a consensus on threat models and for establishing clear protocols for evaluating AI agents and specialised tools.

The first panellist also raised concerns about gaps in current evaluations, particularly for protein-design models, which have undergone risk assessments and capability benchmarking in academic contexts but often lack public security-focused evaluations.[10] 'We just don't know how useful these models are or how consequential potential mitigations might be,' the speaker noted. There was a call for greater involvement from biologists in evaluation efforts, especially around misuse-relevant tasks, and for clearer methodologies to assess the risks of emerging AI tools.

4.1.3. Institutional considerations

The role of AISIs served as another focal point of the first panellist's remarks. The speaker praised the rapid establishment and high-quality work of AISIs but questioned their capacity for exploratory research and their coordination with regulatory bodies such as the EU AI Office. Key concerns included whether data from mandatory lab reporting should be pooled and shared across institutions and who should manage such information sharing.

The speaker cited the example of monkeypox research, where scientists were blocked from accessing certain data, sparking public outcry on social media.[11] This incident highlighted the tension between biosecurity protocols and scientific transparency, raising broader questions about governance and access.

4.1.4. Emerging capabilities and strategic mitigation

The second speaker discussed recent AI advances, particularly tools like DeepSeek, which suggest that AI systems could soon reach postdoctoral-level cognitive capabilities in certain research tasks. Although full

[10] Some national security bodies, such as the UK Cabinet Office CBRN team, caution against publicly releasing detailed evaluations of misuse-relevant capabilities for biotechnology tools to avoid enabling harmful applications.

[11] Lin, 2024.

laboratory automation remains a longer-term challenge, the rapid proliferation of specialised machine-learning models raises urgent questions about security, safety and oversight.

To address these risks, the speaker advocated for a 'defence-in-depth' strategy – often referred to as the 'Swiss cheese model' – which balances scientific progress with security. This approach includes implementing encryption, using training sets that allow for automated threat detection, and developing privacy-preserving systems that can flag and refuse potentially dangerous user queries.

4.1.5. Evaluating safety and future coordination

The third speaker focused on the current state of safety evaluations, noting that LLMs may increase biological risks, though evaluations are still in their early stages. Teams are working to integrate built-in guardrails and metadata systems to help capture and interpret user intent. Several pilot projects, including those using lattice automation technologies, are actively testing these security measures.

The speaker also discussed the development of a comprehensive security framework for protein-design tools, including screening databases and detailed technical specifications. The panellists noted that some developers are working closely with the broader bio-design community to ensure these frameworks balance security with research accessibility.

4.2. Open-floor discussions: shaping future research priorities

During the subsequent open-floor discussions participants identified several critical challenges in implementing robust security measures. They highlighted the need for more sophisticated threat-identification mechanisms beyond basic synthetic screening and expressed concern over the high level of risk tolerance observed among some non-government actors conducting bio-related evaluations and risk assessments, including in areas where the misuse potential is significant. This contrasted with earlier calls from at least one participant for more public evaluations of misuse-relevant capabilities of protein language models, highlighting an important tension within the community. Participants also noted the lack of clear frameworks for sharing security information, with many organisations awaiting government guidance on information-sharing protocols.

4.2.1. Future directions

Participants agreed that the network of AISIs could serve as a central hub for coordinating research agendas and advancing risk-mitigation strategies. However, they also recognised that existing guardrails might not be sufficient against increasingly advanced AI agents, necessitating further exploration of alternative approaches. Balancing security with research accessibility was a constant concern, with many stressing the need for policies that allow legitimate scientific work to continue without excessive friction.

4.2.2. Insights and illustrative quotes

Several participants offered remarks that reflected broader themes discussed in the session. These quotes are included as anecdotal reflections to illustrate the tone of the discussion and concerns raised rather than as representative or evidence-based findings:

> We're at the edge of major breakthroughs, but without proper safeguards, the risks could outpace the benefits.
>
> AISIs have a critical role in harmonising research efforts globally, but they need more resources and clearer mandates to be effective.
>
> Balancing accessibility for legitimate researchers with security protocols is key – too much friction could stall scientific progress, but too little invites misuse.

The session underscored the importance of developing a comprehensive, globally coordinated research agenda focused on AIxBio risk mitigation. Most participants called for unified safety frameworks, greater transparency in risk evaluation, and increased collaboration between AISIs, regulatory agencies and the broader scientific community. By prioritising a strategic, well-resourced approach, the international community can better navigate the evolving landscape of AI and biosecurity.

5. Workshop Session 3: Collaboration, Resource Sharing and Best Practices

Session 3 prompted participants to discuss actionable steps to build a more collaborative global ecosystem for AIxBio governance, as outlined in Box 3.

Box 3. Session 3 agenda text

Opportunities for collaboration to safeguard AIxBio capabilities: sharing resources and best practices

- What opportunities exist for collaboration between AISIs, AI developers, biosecurity experts and international organisations to strengthen biosecurity? How can the AIxBio Global Forum and other venues support these efforts?
- What AIxBio-related resources, best practices or frameworks have been developed that could be shared across AISIs? How can we ensure that such resources continue to be shared over time?
- What are the current barriers or challenges for sharing information and resources among AISIs? What practical approaches can address these barriers?

5.1. Panellist remarks: the state of AISI-centric governance

The session began with an overview of the evolving landscape of AISIs. The moderator noted that at least 11 AISIs have been established, with more currently in development. However, there is considerable variation in the maturity and expertise of these institutes. Participants noted that, while the UK AISI and its recently renamed US equivalent, the Center for AI Standards and Innovation (CAISI), had more established biosecurity teams at the time of the workshop, the capacity, staffing and priorities of many other institutes remained unclear.

Panellists identified collaboration between AISIs as a critical priority, especially given the uneven distribution of resources and expertise. The network remains in its early stages and participants highlighted the need to overcome significant challenges in resource sharing and coordination. The upcoming launch of the AI-Bio Global Forum in 2025, as recommended in the NTI report, was presented as a key step towards fostering international cooperation and creating a unified platform for dialogue.[12]

[12] Carter et al. 2023

5.2. Open-floor discussion: collaborating on information sharing

The open-floor discussions focused heavily on the barriers to effective information sharing. Participants pointed out that the lack of standardised protocols across governments complicates data exchange, often leading to delays caused by varying clearance processes. Organisations also face time constraints, as many are 'building the plane while flying it' and trying to develop governance mechanisms in real-time as risks evolve.

Privacy concerns further complicate efforts to track AI tool usage and user queries, limiting the ability to conduct comprehensive risk assessments. Participants agreed that the upcoming AI-Bio Global Forum could play a key role in developing initial drafts for standardised information-sharing protocols, helping to streamline clearance processes and reduce delays.

A major focus of the session was on establishing the forum as a central hub for collaborative efforts. The forum aims to:

- Develop a shared understanding of biosecurity risks
- Support seamless information sharing across stakeholders
- Promote governance mechanisms that align with both national and international policies.

The forum will also facilitate technical working groups tasked with developing high-level risk statements and organising workshops on biosecurity best practices and risk management. Participants stressed the need for a more comprehensive research agenda and greater sharing of best practices to create a unified approach to biosecurity.

5.2.1. Organisational roles

Participants also explored the roles that various organisations can play in strengthening AIxBio governance. Civil-society organisations were identified as key actors in facilitating information exchange between AI labs and policy stakeholders. There was a strong call to include international model developers in collaborative forums to ensure global representation.

Participants recognised philanthropy for its role in funding research initiatives and access to high-performance computing resources like graphics processing units (GPUs). They also highlighted United Nations agencies for their potential to promote governance mechanisms globally. Participants emphasised the importance of establishing direct communication channels with organisations like the World Health Organization when biosecurity risk are identified.

The Chan Zuckerberg Initiative (CZI) was specifically mentioned for its active involvement in providing grants, GPU access, and supporting the development of bioscience models. The CZI's work on funding high-quality biological data generation was applauded, though participants acknowledged the challenges involved in data curation and annotation. There was agreement on the need to better understand which types of biological data, particularly those used in AI training, may contribute to dual-use capabilities and to prioritise research questions that could help mitigate downstream risks

5.2.2. Data analysis and risk assessment

Panellists noted that many current threat models are developed in relatively isolated academic or policy settings, often by non-government researchers working in unclassified contexts. As a result, these models may miss key aspects of real-world threats, especially in situations where AI-enabled capabilities are spreading through academic, commercial or open-source channels.

There was strong support for grounding risk assessments more directly in empirical user data to better reflect how AI tools might be adopted or misused across diverse institutional environments. Legal challenges remain, however, particularly around data privacy and the sharing of sensitive information.

The FMF was identified as a potential platform to facilitate data analysis and risk assessment given its existing network of AI developers and researchers. Anthropic's Clio, a privacy-preserving approach, was highlighted as a promising example of how sensitive user data can be analysed without compromising privacy.[13] Participants also discussed the importance of analysing low-volume queries, which could reveal subtle but significant risks in AI tool usage.

5.2.3. Regional considerations and tracking AI diffusion

Participants noted that biosecurity is not yet a prominent issue on the European political agenda, emphasising the need for improved storytelling and communication strategies to raise awareness with policymakers and the public. Effective translation of biosecurity concerns across regions was seen as vital for building a cohesive international response.

Tracking AI diffusion emerged as another critical challenge. While tools like AlphaFold can be monitored through published research, tracking the use of LLMs presents greater difficulties due to privacy concerns around user queries and the growing proliferation of open-source models. Anthropic's privacy-preserving approach was again cited as a model for balancing data analysis with user privacy.

Participants discussed the potential resistance from researchers to screening tools that require disclosure of their work, underscoring the need for thoughtful policy design which balances oversight with academic freedom.

5.2.4. Future directions

The session concluded with a forward-looking discussion on how better to prepare for future AIxBio developments. Participants advocated advance planning around new discoveries including pre-clearance mechanisms for access to potentially dual-use bioengineering tools or datasets. Government-led clearance processes were seen as preferable to ad-hoc private or NGO-led efforts.

Some participants argued that mapping broader AI-related risks to their specific impacts in the biosecurity realm would be beneficial for those AISIs that do not have dedicated biosecurity teams. For institutes with limited staff capacity, drawing clear connections between general AI risks and biosecurity concerns can enable more meaningful contributions to both biosecurity-specific efforts and wider AI governance discussions.

[13] Tamkin et al. 2024

Other participants returned to points about the importance of communication and method sharing between researchers. Some suggested that policymakers should treat the development of AI risk-management system specifications – covering areas such as model auditing, deployment controls and cross-institution coordination – as a collaborative, community-led process.

5.2.5. Insights and illustrative quotes

Several participants voiced perspectives that echoed key themes of the discussion:

> We can't afford to wait for the perfect system – collaboration has to start now, even if it's messy.

> Standardised information-sharing protocols are crucial if we want a global response that's timely and effective.

> Civil society plays a key role in bridging the gap between AI developers and policymakers. We need more spaces for these conversations.

The session highlighted the urgent need for stronger international collaboration, standardised governance practices, and proactive planning to safeguard AIxBio capabilities. While recognising the challenges inherent in global coordination, especially around sensitive, dual-use information, participants also emphasised that meaningful progress requires cross-sectoral cooperation, trust-building and a commitment to transparency.

6. Key Findings and Recommendations

Advances in AI offer new, low-cost tools which could be misused by bioterrorists and nation-state actors to threaten global biosecurity, but data about the global spread and use of these powerful AI models remain scarce. Effective defence against biological threats requires leaders to have reliable information about where the most dangerous AI systems are accessible and how quickly these tools could enable the creation and dissemination of harmful viruses, toxins or other biological agents.

To help address these challenges, NTI and RAND partnered to convene leading experts in biosecurity, AI and governance system design. In this concluding chapter we present three provisional observations, based on patterns observed across the workshop, concerning the current state of biosecurity governance, the potential scale of AI-related impacts, and the central concerns voiced by participants. While these findings reflect recurring themes in the discussions, workshop participants also emphasised the need for further evidence to support or refine them.

Following these findings we outline seven actions recommended by participants at the event, which we offer for readers' further assessment and evaluation.[14]

6.1. Key findings

Over the course of its collaboration with NTI, the RAND project team observed a number of claims which were repeated by experts in a variety of contexts throughout the workshop:

- Policymakers and industrial actors are paying insufficient and decreasing attention to potential biological risks from advances in AI.
- Biosecurity experts estimate novel incident-response ('if-then') approaches could have significant value for efforts to reduce risks from AI.
- There is a lack of consensus over which existing risk-reduction approaches governance stakeholders should address first, inhibiting joint action.

Workshop participants broadly agreed on the need for closer collaboration between AI policy communities and biosecurity experts. Such collaboration was seen as essential to reduce persistent uncertainties about the future trajectories of AI-related risks to national security and the global public interest. However, pre-workshop questionnaires and workshop discussions highlighted practical obstacles which hinder progress

[14] Importantly, all key findings and recommendations presented in this chapter reflect observations made by participants during the workshop.

towards this goal. These included difficulties in establishing the shared, valid measures of policy success which are necessary for effective learning and governance. Four mechanisms were cited as contributing to this challenge:

1. Weak concept formation, leading to imprecise or absent criteria for measuring policy outcomes as regards mitigating AI risks to biosecurity.
2. Resource misallocation, with an overemphasis on projects framed around non-falsifiable claims rather than those that can inform measurable progress.
3. Limited institutional support, particularly the absence of streamlined legal processes for data sharing and collaboration.
4. External volatility, where rapid developments in AI or biotechnology markets and geopolitical dynamics disrupt efforts to build or sustain scientific consensus.

These factors slow down and reduce the quality of policy-relevant research, especially in AI, where the broad, general-purpose nature of the technology demands a solid grounding in scientific reasoning and robust technical safeguards. Participants emphasised that, without greater alignment around reliable metrics and evaluation standards, efforts to assess and manage biosecurity risks will likely remain fragmented and ineffective.

6.1.1. Workshop Theme 1. Policy tensions between AI innovation and biosecurity oversight.

Participants reported a growing concern that the increasing acceptance and promotion of AI technologies may unintentionally reduce the visibility of biosecurity risks on policy agendas. Evidence from Session 1 and the pre-workshop questionnaire indicated that few strong political or economic incentives currently encourage major actors in the AI ecosystem to engage in systemic, large-scale information sharing relevant to biosecurity.

As global competition intensifies and value-generating AI applications become more widespread, public and political enthusiasm for AI may grow – assuming that negative external effects can be mitigated (e.g. disruption to labour markets). However, the exclusion of biosecurity topics from the 2025 AI Action Summit and recent policy emphasis on AI-driven opportunities suggest a potential decline in policymakers' sensitivity to biosecurity risks.

To maintain momentum, biosecurity-focused practitioners and experts will need:

- Tools to extend the international relevance and reach of their risk mitigation proposals.
- Clear understanding of links between AI deployment and empirically supported risks to human and environmental health.
- Strategies for navigating an increasingly crowded policy discourse where both the catastrophic risks and transformative benefits of AI compete for attention.

New models of technical cooperation are needed to enable cross-sector information sharing

Sessions 2 and 3 revealed a broad agreement that effective governance will require increased collaboration between sectors, including better data sharing and the development of shared or interoperable safety frameworks. Participants emphasised that actors who pioneer low-risk methods of information sharing, especially those who protect intellectual property and sensitive data, could gain early advantages in markets that prioritise biosecurity.

Market incentives for AI innovation may increase operational and biological risk

Responses from the pre-workshop questionnaire raised concerns that strong market incentives for rapid AI development may reduce industry's focus on safety. Participants noted that when AI evaluators act as both risk assessors and risk managers there is a conflict of interest that weakens transparency-based governance. Without clear and enforceable risk-management standards, industry may tolerate higher levels of risk, including biological risks. This reflects a tension in incentive structures, in which benefits may be privatised and risks socialised.

Evidence-based narratives are essential to keep biosecurity on the policy agenda

Participants emphasised the importance of political leadership in defining risk thresholds and setting priorities. They called for greater investment in developing compelling, evidence-based narratives around low-probability but high-impact biosecurity threats. Policymakers are more likely to engage with these issues when supported by clear, empirically grounded arguments that frame biosecurity as a concrete and manageable component of broader AI governance.

6.1.2. Workshop Theme 2. Modular 'if-then' response frameworks show promise for scalable AI-related biosecurity action.

Given the uncertainty surrounding the scale, speed of development and predictability of AI-enabled biological risks, many participants noted that estimating absolute risk is often impractical and resource-intensive. Instead, there was significant interest in structured, conditional approaches to managing risk – particularly frameworks in which specific responses are automatically triggered when predefined hazard thresholds are crossed. Throughout the workshop, participants emphasised the potential of modular, 'if-then' incident-response strategies to serve as flexible, scalable solutions for biosecurity governance. These prioritise preparedness and clarity over exhaustive risk quantification, offering a path towards practical, real-world implementation.

Without demonstrated success and strategic communication, modular response frameworks will struggle to scale

For modular response strategies to gain broad adoption, proof-of-concept projects and successful real-world demonstrations will need to be effectively shared with relevant stakeholders. Participants noted that political and financial support for shifting towards agile, contingency-based governance models will likely depend on both technical validation and the strategic dissemination of results. In particular, advocates of 'if-then' approaches must find ways to communicate the credibility and reliability of these frameworks across diverse

and often contested information environments. Building trust in these methods will be essential to their acceptance and eventual integration into global biosecurity strategies.

6.1.3. Workshop Theme 3. Strengthening core biosecurity functions is urgently needed for long-term risk management.

While participants welcomed innovative approaches to managing risks, they emphasised that existing core systems – such as risk assessment, hazard threshold setting, automated hazard detection and response mechanisms, and regulatory oversight – require immediate improvement. These foundational components are not yet robust enough. Throughout the workshop, participants highlighted examples of collaborative, cross-institutional efforts aimed at addressing specific shortcomings in current mitigation practices. Experts agreed that scaling such initiatives could have a disproportionately positive impact on the future resilience of AI and biosecurity governance.

6.2. Recommendations made by participants

Participants proposed a series of actions for dealing with biosecurity risks. In many cases they themselves emphasised that further evidence or study would be necessary to assess the feasibility, effectiveness or priority of these ideas. Readers are encouraged to critically assess their relevance, practicality and risk–benefit trade-offs in their own institutional or policy contexts:

- **Convene as 'coalitions of the willing' to innovate practical, adaptable solutions.** Some AI-enabled biosecurity threats, such as the rapid spread of dual-use AI tools or other technologies that lower barriers to misuse, could emerge and escalate quickly. Addressing them may require governance approaches which can also be rapidly deployed, adapted and expanded as circumstances change. Participants emphasised linking AI benchmarks to real-world risks (in public health, food safety and national security) and using modular 'if-then' response frameworks which can be scaled up or adjusted as needed. The lack of consensus at the 2025 AI Action Summit suggests that flexible, self-organised coalitions may be critical for advancing urgent biosecurity priorities.
- **Identify focal points for cross-sector collaboration and cohesion.** Institutions such as the World Health Organization and the United Nations Institute for Disarmament Research may play a key role in convening international communities focused on AI-bio risk governance. Forums like the NTI Global AI-Bio Forum can support dialogue, information sharing and coordination across scientific, policy and civil-society actors.
- **Strengthen coordination and information sharing among government agencies and regulatory bodies.** Improved internal coordination and more inclusive engagement with relevant actors can help align priorities, reduce duplication of effort and enable more efficient sharing of data and expertise, all of which can expand research pipelines and strengthen risk-monitoring capacity.
- **Develop international standards, including unified safety frameworks.** Participants called for standardised, evidence-based approaches to risk assessment and mitigation. These frameworks could incorporate contributions from government, industry and civil society, and help define clear

risk thresholds, red lines and safety indicators tailored to the novel challenges posed by AI in relation to biosecurity.

- **Use scenarios to strengthen preparedness.** Strategic use of scenarios can help stakeholders anticipate potential AI-related biosecurity risks, identify failure points, and develop more robust, coordinated response strategies.
- **Promote effective risk communication.** Clear, audience-appropriate communication tools are essential to help policymakers, practitioners and the public understand biosecurity risks associated with AI. Well-designed narratives can support shared understanding and informed decision-making.
- **Invest in risk measurement, assessment and safe testing.** Technical innovation – including risk-assessment tools, privacy-preserving monitoring and post-deployment evaluations – requires sustained funding. Building on this, participants emphasised the value of public–private partnerships and localised pilots to support development and the safe, controlled testing of technical biosecurity interventions. Such testing should use proxy tasks and simulated misuse scenarios to evaluate how both general-purpose AI models and biotechnology design tools could be exploited, without engaging in hazardous laboratory work.
- **Adopt a 'responsible innovation' lens.** In some cases, bottlenecks in AI innovation can worsen the biosecurity risks that AI already poses. Stakeholders across AI and biosecurity sectors can frame their risk-reduction efforts in terms of 'responsible innovation', balancing the need for progress with the imperative to reduce risk. Applying clear due-diligence standards – such as systematically evaluating safety, security and compliance risks before deployment, and assessing the trade-offs that biosafety and biosecurity measures may entail – can help avoid unintended consequences of both technical and policy innovations.

6.3. Conclusion

While the workshop was not designed to produce formal consensus recommendations or analytic findings, several themes emerged consistently across sessions which may inform future policy and research efforts. Participants emphasised that uncertainty about the development trajectory of advanced AI technologies raises urgent questions for biosecurity governance. Many noted that while the challenges are complex, practical progress is possible, especially when supported by focused collaboration, better data and adaptive governance frameworks. There was strong agreement on the need to equip policymakers, researchers and industry leaders with timely and actionable information, and to expand cross-sectoral information sharing which preserves privacy and intellectual-property protections. Despite acknowledging major barriers – including regulatory uncertainty, underdeveloped reporting mechanisms and gaps in empirical evidence – participants identified concrete starting points such as 'if-then' response frameworks and the harmonisation of risk thresholds. The workshop format itself was seen as valuable, offering a rare opportunity to convene technical experts, policy leaders and institutional stakeholders across traditionally siloed communities. Several participants recommended continuing this kind of structured, cross-disciplinary engagement. Finally, participants stressed the importance of translating technical insights

into governance-relevant tools, language, and models that decision-makers can use. Meeting this need will require the same priorities highlighted in our recommendations, including trusted communication channels, risk-modelling frameworks and adaptive institutions.

Appendix A. Methodology

A.1. Participant selection

The RAND research team and its NTI partners sought to leverage the experience of professionals working in field-leading science, technology and policy enterprises across the globe by inviting a group of these professionals to attend the workshop event. Specifically, the workshop development team sought to convene a participant group drawn roughly equally from the following three categories of working professionals[15]:

- Industry representatives from major US and international firms developing frontier AI systems.
- Technical experts in the life sciences at leading research institutes and NGOs focused on biological threats to US and global security.
- Officials from across national and international governmental bodies focused on AI (e.g. AI safety institutes, regulators).

Ultimately, the team secured commitments from n≈25 external participants, approximately equally drawn from the three target groups: government participants constituted a slightly smaller share than expected and non-AI industry representatives constituted a slightly larger share. Although the team invited a global set of organisations, the final workshop participant list disproportionately represented perspectives of advanced western industrial economies. This potentially reduces the generalisability of some of the findings.

A.2. Workshop objectives

As part of the agenda shared with participants, we identified the following four desired outcomes for the workshop[16]:

- Develop a shared framework for understanding and addressing risks at the convergence of AI and the life sciences.
- Establish frameworks for evaluating and defining thresholds in AI development and deployment that present unacceptable biological risks.

[15] See the 'Participant selection' section in Appendix A for more details on the participation selection process, including potential limitations.

[16] See the 'Project objectives and institutional coordination' section in Appendix A for a discussion of the process by which RAND and NTI selected objectives for the workshop, including assumptions and potential limitations.

- Create channels for AISIs to share best practices and resources for evaluating and mitigating AIxBio risks among themselves and with AIxBio domain experts.
- Advance global security by laying the groundwork for coordinated oversight of AIxBio capabilities.

A.3. Pre-workshop questionnaire

RAND invited participants to complete a questionnaire before arriving at the workshop. The questionnaire was designed to elicit information about the work participants perform at their organisations and about their expectations relating to the dynamics of AI, biosecurity and policy development.[17]

The questionnaire was intended to elicit information on three substantive issues:

1. **How are organisations responding to AIxBio threats?** That is: in which types of self-regulatory AI-related biosecurity activities do representatives of different organisations report engaging, and what additional support do they 'demand' (i.e. express needing)?
2. **Why do these responses vary across organisations?** That is: to what extent do expressions of 'demand' for additional support for AIxBio risk-mitigation depend on the existing 'supply' of available organisational resources oriented towards reducing risks or enhancing benefits?
3. **Under what circumstances might the questionnaire be an unreliable basis for drawing conclusions about these questions?** That is: which respondent-level characteristics are most likely to support or degrade the validity of the conclusions the RAND research team could draw from the information gathered through the questionnaire? What other threats to inferential validity may exist?

Just over half (56 per cent) the workshop participants completed the questionnaire.[18] The RAND team used information gathered from these respondents to develop talking points and presentation materials for the opening session, as discussed in Chapter 2.

[17] The RAND team is grateful for valuable comments and external review provided by NTI team members in testing the questionnaire.

[18] RAND and NTI collaborated on a plan to publicly set the deadline a week prior to the event and then to extend the deadline to four days prior to the event. During the deadline extension period, the questionnaire participation rate increased by approximately 30 per cent.

Abbreviations

AI	Artificial intelligence
ACE	TASP AI Capability Evaluations cluster
AISI	AI safety/security institute
CZI	Chan Zuckerberg Initiative
EU	European Union
FMF	Frontier Model Forum
FTE	Full-time personnel equivalent
FSFs	Frontier safety frameworks
GDPR	General Data Protection Regulation
GER	RAND Global and Emerging Risks division
GPAI	General-purpose AI
LLMs	Large language models
NGO	Non-governmental organisation
NTI	Nuclear Threat Initiative
RSP	Responsible scaling policy
TASP	RAND Technology and Security Policy Center
UNIDIR	United Nations Institute for Disarmament Research

References

Bostrom, Nick. 2011. 'Information Hazards: A Typology of Potential Harms from Knowledge.' *Review of Contemporary Philosophy* 10: 44–79.

Carter, Sarah R., Nicole Wheeler, Sabrina Chwalek, Chris Isaac & Jaime M. Yassif. 2023. 'The Convergence of Artificial Intelligence and the Life Sciences: Safeguarding Technology, Rethinking Governance, and Preventing Catastrophe.' NTI, 30 October. As of 21 Aug 2025: https://www.nti.org/analysis/articles/the-convergence-of-artificial-intelligence-and-the-life-sciences

Götting, Jasper, Pedro Medeiros, Jon G. Sanders, Nathaniel Li, Long Phan, Karam Elabd, Lennart Justen, Dan Hendrycks & Seth Donoughe. 2025. 'Virology Capabilities Test (VCT): A Multimodal Virology Q&A Benchmark.' *arXiv*, arXiv:2504.16137. As of 21 Aug 2025: https://arxiv.org/abs/2504.16137

Hendrycks, Dan, & Laura Hiscott. 2025. 'AIs Are Disseminating Expert-Level Virology Skills.' AI Frontiers [podcast].

Lin, Fay. 2024. 'AlphaFold 3 Angst: Limited Accessibility Stirs Outcry from Researchers.' *GENEdge News*, 13 June. As of 21 Aug 2025: https://www.genengnews.com/topics/artificial-intelligence/alphafold-3-angst-limited-accessibility-stirs-outcry-from-researchers

Mouton, Christopher A., Caleb Lucas & Ella Guest. 2023. *The Operational Risks of AI in Large-Scale Biological Attacks: A Red-Team Approach*. Santa Monica, Calif.: RAND Corporation. RR-A2977-1. As of 21 Aug 2025: https://www.rand.org/pubs/research_reports/RRA2977-1.html

Nevo, Sella, Dan Lahav, Ajay Karpur, Yogev Bar-on, Henry Alexander Bradley & Jeff Alstott. 2024. *Securing AI Model Weights: Preventing Theft and Misuse of Frontier Models*. Santa Monica, Calif.: RAND Corporation. RRA2849-1. As of 21 Aug 2025: https://www.rand.org/pubs/research_reports/RRA2849-1.html

NIST. 2024. 'U.S. AI Safety Institute Establishes New U.S. Government Taskforce to Collaborate on Research and Testing of AI Models to Manage National Security Capabilities & Risks.' US Department of Commerce, 20 November. As of 21 Aug 2025: https://www.commerce.gov/news/press-releases/2024/11/us-ai-safety-institute-establishes-new-us-government-taskforce

Nuclear Threat Initiative. Homepage. 2025. As of 21 Aug 2025: http://www.nti.org

NTI | bio. 2023. 'The Convergence of Artificial Intelligence and the Life Sciences: Safeguarding Technology, Rethinking Governance, and Preventing Catastrophe.' As of 21 Aug 2025: https://www.nti.org/wp-content/uploads/2023/10/NTIBIO_AI_Executive-Summary_FINAL.pdf

Presidency of France [Élysée Française]. 2025. 'Artificial Intelligence Action Summit.' 2025. As of 21 Aug 2025: https://www.elysee.fr/en/sommet-pour-l-action-sur-l-ia

RAND Global and Emerging Risks. Homepage. As of 21 Aug 2025: https://www.rand.org/global-and-emerging-risks.html

RAND Meselson Center. Homepage. As of 21 Aug 2025: https://www.rand.org/global-and-emerging-risks/centers/meselson.html

Stephan, Paula E. 1996. 'The Economics of Science.' *Journal of Economic Literature* 34(3): 1199–1235.

Tamkin, Alex, Miles McCain, Kunal Handa et al. 2024. 'Clio: Privacy-Preserving Insights into Real-World AI Use.' *arXiv*: 2412.13678. As of 21 Aug 2025: https://arxiv.org/abs/2412.13678

Vance, J.D. 2025. 'Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France.' *The American Presidency Project*, 11 February. As of 21 Aug 2025: https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france