ISSUE BRIEF

FRONTIER MODEL FORUM

Frontier AI Agents and Biological Tools: Preliminary Risks and Considerations

DATE

July 31, 2026

INTRODUCTION

AI is rapidly advancing biotechnology by accelerating scientific discovery, biological design and experimental research.[1] Much of this progress is being driven by AI-enabled biological tools (BTs), including biological foundation models, specialized analytical tools, and components of increasingly automated biological workflows.[2] Recent breakthroughs illustrate both the breadth and accelerating pace of these advances. In 2025, researchers developed a foundation model that successfully generated a novel hypothesis about cancer cellular behavior, revealing a promising new pathway for developing cancer therapies.[3] Researchers also introduced a biological foundation model trained on genomic data that can predict the functional effects of genetic variation and support biological design at genome scale.[4] As these capabilities mature, they have the potential to accelerate beneficial research by changing how biological capabilities are developed, accessed, and deployed.

Biological tools also have the potential to reshape the biosecurity landscape, particularly when coupled with agents based on the most advanced AI systems. AI agents may increasingly coordinate multiple BTs across end-to-end biological workflows involving scientific reasoning, experimental planning, laboratory automation, and iterative optimization.[5] While these capabilities offer significant scientific benefits, they may also introduce new pathways through which biological capabilities become more accessible, scalable, or sophisticated, creating new challenges for frontier AI developers seeking to understand and mitigate biosecurity risks, including those posed by malicious actors.[6]

Although both frontier AI systems and AI-enabled biological tools

present biosecurity considerations, this brief focuses specifically on risks arising from their interaction. Combining these technologies may create capabilities that are not readily apparent when either is evaluated in isolation, potentially increasing the accessibility, sophistication, scale, and speed of activities that could support biological misuse. These risks may become more significant as frontier AI systems grow increasingly capable of planning, coordinating and executing complex workflows across specialized tools and services.[7]

This issue brief identifies preliminary areas of concern arising from the interaction of frontier AI agents and biological tools. Drawing on input from experts across the member firms of the Frontier Model Forum (FMF), as well as the broader AI safety and biosecurity ecosystems, the issue brief identifies emerging risks from those interactions and outlines considerations for further research and coordination across industry, government, and the biosecurity community.[8]

EMERGING BIOSECURITY RISKS FROM FRONTIER AI AGENTS AND BTs

Interactions between frontier AI systems and AI-enabled biological tools may create biosecurity risks that neither technology presents to the same degree in isolation. To understand these interactions, it is important to distinguish between frontier AI systems and AI agents.[9] Frontier AI systems refer to state-of-the-art general-purpose AI models with advanced reasoning capabilities that can support a wide range of scientific and technical tasks. AI agents build upon these models by combining advanced reasoning with planning, memory, external tool use, and the ability to autonomously execute complex workflows.

Not all AI-enabled biological tools present the same biosecurity risks. This brief focuses primarily on tools that materially contribute to biological design, engineering, optimization, experimental planning, or tool creation for downstream biological workflows, especially when used in agentic settings. Many defensive or analytic tools, including those used for benign classification, interpretation, or quality control, are not the primary focus here. For the purposes of this brief, the relevant risk pathway runs from benign scientific utility toward modification, enhancement, circumvention, or orchestration of biological capability.[10]

Conversations with experts identified two high-level categories of risk arising from interactions between frontier AI systems and BTs. First, frontier AI systems may substantially lower barriers to entry for using advanced biological capabilities by reducing the expertise required to engage with advanced biological capabilities, including through instruction, troubleshooting, tool selection, or autonomous tool use. Second, they may raise the ceiling of harm by enabling sophisticated actors to develop more advanced biological threats through enhanced design, optimization, and tool orchestration.

Lowering Barriers to Entry

Many advanced BTs require substantial domain expertise to operate effectively,

across both biological and computational domains. Users often must understand specialized biological concepts, determine which tools are appropriate for a particular task, interact via the command line interface, correctly prepare inputs, develop custom scripts, interpret technical outputs, and integrate multiple tools into coherent workflows.

Frontier AI systems, particularly advanced reasoning models and agentic systems, may substantially reduce these barriers by assisting users throughout each stage of the workflow. Importantly, lowering barriers does not necessarily require frontier AI systems to advance biological science itself, and instead may only require them to reduce the expertise needed to access existing biological capabilities. Several pathways by which they may do this include:

- **Instruction, Translation, and Workflow Assistance:** Frontier AI may lower the expertise required to use BTs by helping with tool selection, troubleshooting, interpretation, and workflow execution. For example, a frontier AI system may help identify the most appropriate BT for a given workflow, instruct a human user on how to use the tool, and assist in operating the tool by translating inputs and outputs that are not in natural language. Recent evaluations indicate that LLM agents can already perform initial tool-selection and tool-operation tasks across multiple biological tools, supporting concern that agents may reduce the expertise required to access specialized biological capabilities.[11]

- **Replicating or Circumventing Biological Tool Safeguards:** While many biological tools are openly available, others rely on restricted access, licensing agreements, trusted-user programs, or dataset filtering to reduce misuse. Frontier AI systems may increasingly be capable of approximating aspects of restricted tool functionality using publicly available scientific literature, documentation, benchmark datasets, and open-source software. They may also assist users in re-creating or restoring capabilities that were intentionally filtered, removed, or withheld during model training, or in identifying alternative tools and workflows that bypass existing safeguards. In some cases, these dynamics could weaken governance mechanisms that rely primarily on restricting access to individual tools or datasets. A related concern is that tools or datasets filtered for safety may be partially recoverable through downstream fine-tuning or tool-assisted model modification, reducing the effectiveness of purely data-based safeguards.[12]

- **Autonomous Workflow Orchestration:** In the near future, AI agents may be able to coordinate multiple advanced BTs across more expansive end-to-end biological workflows, including experimental planning, data interpretation, and iterative optimization. As biological infrastructure becomes more automated and externally accessible, such as through cloud laboratories, these capabilities could further reduce the expertise and operational barriers associated with advanced biological research.

- **Overcoming Operational Bottlenecks:** AI agents may help threat actors reduce the operational friction involved in harmful workflows by coordinating tasks, managing tool use, and adapting plans under changing constraints. This could lower the practical burden of executing a biological threat, even if it does not fundamentally change the underlying scientific challenge.[13]

Raising the Ceiling of Harm

Lowering barriers to entry primarily expands who can access existing biological capabilities. By contrast, raising the ceiling of harm concerns the possibility that frontier AI systems and AI-enabled biological tools enable new or substantially enhanced biological capabilities, even for sophisticated actors. This may occur through more advanced biological design, improved optimization, or the creation of increasingly capable biological tools themselves. By accelerating data analysis, workflow coordination, and biological engineering processes, frontier AI systems may enable forms of biological capability that would otherwise remain impractical or infeasible. The examples below illustrate several ways in which the ceiling of harm could be raised, but are not intended to capture the full range of possible mechanisms.

- **Novel Design Capabilities:** Frontier AI systems may help identify non-obvious relationships within biological datasets and infer patterns associated with specific functional properties. By applying these insights across large design spaces, frontier AI systems could assist in designing novel biological constructs with specific, enhanced functional properties that would be difficult to identify through manual research alone.

- **Optimization of Biological Properties:** Frontier AI systems could assist in optimizing biological properties by analyzing complex thermodynamic and structural data outputs. This capability could enable more precise engineering of biological constructs with enhanced stability, resilience, or environmental persistence, significantly reducing the trial-and-error typically associated with biological optimization.

- **Complex Assembly and Engineering:** AI agents may also assist in coordinating complex engineering tasks involving multiple biological tools and synthesis processes, including the composition of modular tools into larger end-to-end workflows. By generating optimized assembly strategies or managing complex synthesis pathways that combine multiple tools, these systems could accelerate the development of sophisticated biological constructs, enabling forms of design and coordination that would be difficult to achieve using isolated tools alone.

MANAGING RISKS FROM FRONTIER AGENT-BT INTERACTIONS

The potential for frontier AI systems to lower barriers to entry and raise the ceiling of harm suggests that existing risk management approaches may need to evolve

alongside agentic biological research ecosystems. Addressing risks at the intersection of frontier AI systems and BTs requires a proactive and coordinated approach. As these systems continue to become more autonomous and interoperable, they require multi-stakeholder efforts to develop new approaches for evaluation, monitoring, access control and risk mitigation.

The intersection of frontier AI and BTs has several implications for potential threat actors and the broader threat landscape:

- **Non-expert actors.** Frontier AI systems and BTs may significantly expand access to advanced biological capabilities by enabling non-experts to perform tasks that previously required substantial training, time, or institutional support. In some cases, non-experts may also use agents to assemble or adapt new biological tools for narrow tasks, including tools that support modification, enhancement, or optimization workflows that would otherwise have been inaccessible to them.

- **Expert actors.** Risk management and oversight efforts should also account for sophisticated state or non-state actors that may be particularly interested in agentic systems capable of coordinating complex biological workflows while reducing operational bottlenecks, attribution, or detection opportunities.

- **Agentic systems/Loss of control.** The threat landscape could also shift as AI systems become more effective at autonomously coordinating biological workflows. While current risk assessments often focus on the human-AI interface, including how users interact with frontier AI and BTs, future agentic systems may increasingly execute parts of these workflows with limited human involvement. The combination of greater autonomy, reduced human oversight, and the ability to coordinate multiple tools may make harmful activity more difficult to detect, interrupt or control. These concerns may be heightened in the context of open-weight models, which could potentially be modified, fine-tuned, or combined with external tools in ways that circumvent existing safeguards.[14] Given the pace of capability development, risk management approaches should remain forward-looking and continuously reassess risks at the intersection of frontier AI systems and BTs.

The intersection of frontier AI and BTs also has significant implications for how frontier AI developers should approach the following elements of risk management:

- **Evaluation and assessment.** Some frontier AI developers may be well-positioned to evaluate potential risks from interactions between frontier AI systems and BTs. In addition to measuring the extent to which advanced models can meaningfully assist users in harmful biological activities, evaluations need to examine agentic capabilities such as autonomous tool-use, multi-step workflow execution, and the ability to replicate closed-source BT capabilities through open-source information (e.g.,

published papers). Evaluations should examine both whether models can assist users with harmful biological activities and whether they can modify, reconstruct, or adapt biological tools and datasets, including cases where safety filters or access restrictions were intentionally applied upstream. Reflecting the growing importance of this problem, the development of evaluation methods specifically for AI agents has been identified as a near-term biosecurity priority by frontier AI developers.[15]

- **Mitigations and safeguards**. Depending on the deployment context, developers may also be well-positioned to implement various safeguards that constrain how agentic systems interact with BTs. Recent experimental work suggests that security mechanisms embedded within individual biological tools are unlikely to be sufficient on their own, because frontier AI agents may ignore, bypass, or disable tool-level safeguards. Effective mitigation may therefore require coordination across frontier AI models, agent harnesses, biological tools, and deployment infrastructure rather than relying on isolated components.[16] For higher-risk workflows, this could include access restrictions, additional authorization layers, logging, and human review. At the ecosystem level, it also points to the importance of societal safeguards such as DNA synthesis screening, dataset governance, and other chokepoints that reduce the chance that downstream misuse succeeds.[17] Recent testing indicates that upgraded sequence-screening tools can detect some AI-reformulated and fragmented sequences of concern, while also underscoring the need to continually adapt screening methods as AI-enabled biological design techniques evolve.[18] However, not all BTs present the same level of risk, and effective risk management requires a more granular assessment of tool capabilities, maturity, accessibility, and potential downstream impacts.[19]

- **Risk-based assessment.** In evaluating risks, developers and deployers should at a minimum consider the BT maturity and the characteristics and operational behavior of the frontier AI system interacting with it. This includes assessing the extent to which agentic systems can coordinate, interpret, or operationalize biological workflows.[20]

- **Monitoring.** Although frontier AI developers and deployers are not positioned to monitor the broader ecosystem of biological tools, they are in a position to focus on how agentic systems interact with BTs and infrastructure. Monitoring mechanisms, coupled with responsible deployment controls and other secure-by-design measures, can reduce high-risk tool interactions. However, the monitoring mechanisms for tool calling and agentic activity should be distinct from those for chatbots, since relevant signals may appear across tool calls, code execution, data access, and multi-step orchestration rather than in a single prompt-response exchange.[21] Developers may also consider prioritizing safe and defensive biological use cases through the use of trusted access before enabling broader and more generalized tool access.

Addressing risks at the intersection of frontier AI systems and BTs will likely require close coordination across frontier AI developers, BT developers, governments, researchers and civil society organizations, to adequately track and mitigate risks that result from new interactions. No single entity can manage these intersectional risks independently. A coordinated effort is required to establish standards, refine evaluation methodologies, strengthen information sharing, and ensure that expanding access to advanced biological capabilities remains beneficial and secure.

AREAS FOR FUTURE WORK

As frontier AI systems become more autonomous and interoperable with biological tools, biosecurity risk management will need to account for capabilities that emerge through interactions across models, tools, and infrastructure. Addressing these risks will require progress in several areas:

1. **Establishing robust evaluation methodologies**: Evaluations should assess frontier AI systems and BTs in combination and across areas such as capabilities, workflows, tool selection, and iterative optimization.

2. **Improving understanding of agent-tool interactions**: Further research is needed to understand how autonomy, access to external tools, and coordination across services affect the accessibility, scale and sophistication of biological capabilities.

3. **Adapting safeguards and monitoring**: Risk management approaches should account for signals that emerge across tool calls, code execution, data access and adaptive workflow orchestration.

4. **Strengthening ecosystem coordination**: Frontier AI developers, BT developers, governments, researchers and civil society organizations should continue working together to refine evaluation methodologies, share relevant information and develop technical standards and safeguards.

The emerging challenge is therefore not simply how to evaluate frontier AI systems or biological tools in isolation, but how to evaluate increasingly autonomous systems of interacting technologies, where capabilities, and associated risks, emerge from their coordination rather than from any individual component alone.

The Frontier Model Forum will continue working with its members and the broader biosecurity community to improve understanding of these interactions, support the development of evaluation and risk management practices, and identify opportunities for information sharing and coordinated action.

FOOTNOTES

1. National Academies of Sciences, Engineering, and Medicine, *The Age of AI in the Life Sciences: Benefits and Biosecurity Considerations*, National Academies Press, April 23, 2025.
2. We use this term to encapsulate other commonly used terms, including Biological AI Models (BAIMs), Biological Design Tools (BDTs), and Biological Foundation Models (BioFMs). See also: Toby Webster et al., "Global Risk Index for AI-Enabled Biological Tools," Centre for Long-Term Resilience and RAND Europe, September 2025.
3. Shekoofeh Azizi and Bryan Perozzi, "How a Gemma Model Helped Discover a New Potential Cancer Therapy Pathway," Google, October 15, 2025.
4. Garyk Brixi et al., "Genome Modelling and Design across All Domains of Life with Evo 2," *Nature*, March 4, 2026.
5. Toby Webster et al., "Global Risk Index for AI-Enabled Biological Tools," Centre for Long-Term Resilience and RAND Europe, September 2025.
6. Cindy S. Groff-Vindman et al., "The Convergence of AI and Synthetic Biology: The Looming Deluge," *npj Biomedical Innovations*, July 1, 2025.
7. For more on the ability of frontier AI to carry out long-horizon tasks, see Thomas Kwa et al., "Measuring AI Ability to Complete Long Tasks," METR, March 19, 2025.
8. See Cassidy Nelson and Sophie Rose, "Understanding AI-Facilitated Biological Weapon Development," Centre for Long-Term Resilience, October 2023; Frontier Model Forum, "Frontier AI Biosafety Thresholds," May 12, 2025; Frontier Model Forum, "Preliminary Taxonomy of AI-Bio Misuse Mitigations," July 30, 2025; The White House, "America's AI Action Plan," July 2025; OpenAI, "Working with US CAISI and UK AISI to Build More Secure AI Systems," September 12, 2025.
9. Frontier Model Forum, "Emerging Security Practices for AI Agents," June 3, 2026.
10. National Academies of Sciences, Engineering, and Medicine, *The Age of AI in the Life Sciences: Benefits and Biosecurity Considerations*, National Academies Press, April 23, 2025.
11. Jeffrey Lee et al., "Can LLM Agents Select and Engage with Biological Tools? An Initial Biosecurity Assessment," RAND Corporation, June 25, 2026.
12. Luca Righetti, Kamile Lukosiute, and James Black, "Coding Agents Are Changing the Biosecurity Risk Landscape," Centre for the Governance of AI, April 20, 2026.
13. Christopher A. Mouton, Caleb Lucas, and Ella Guest, "The Operational Risks of AI in Large-Scale Biological Attacks: Results of a Red-Team Study," RAND Corporation, January 25, 2024.
14. AI Security Institute, "Managing Risks from Increasingly Capable Open-Weight AI Systems," August 29, 2025.
15. Google DeepMind and Isomorphic Labs, "Our Approach to Bioresilience," July 16, 2026.
16. Casey O. Barkan et al., "Restricting AI Agent Use of Biological Tools: Exploring the Feasibility of Software Barriers," RAND Corporation, July 20, 2026.
17. "An Open Letter in Support of Mandatory Nucleic Acid Synthesis Screening and Recordkeeping," screendna.org, June 2026.

18. Bruce J. Wittmann et al., "The Limits of Sequence-Based Biosecurity Screening Tools in the Age of AI-Assisted Protein Design," *Frontiers in Bioengineering and Biotechnology*, July 13, 2026.
19. Doni Bloomfield et al., "Biological Data Governance in an Age of AI," *Science*, February 5, 2026.
20. Toby Webster et al., "Global Risk Index for AI-Enabled Biological Tools," Centre for Long-Term Resilience and RAND Europe, September 2025.
21. Note that identifying and sourcing datasets for tooling-based classifiers may also introduce challenges for the design, implementation, and evaluation of relevant guardrails.