14 May 2025 | Maastricht

Reducing Emerging Biorisks: Safeguarding AIxBio Capabilities

International Biosecurity Symposium

NTI

BUILDING A SAFER WORLD

NTI bio

Overview

I. Reducing Emerging Biological Risks

II. Safeguarding AIxBio Capabilities

III. Technical Guardrails

IV. AIxBio Global Forum

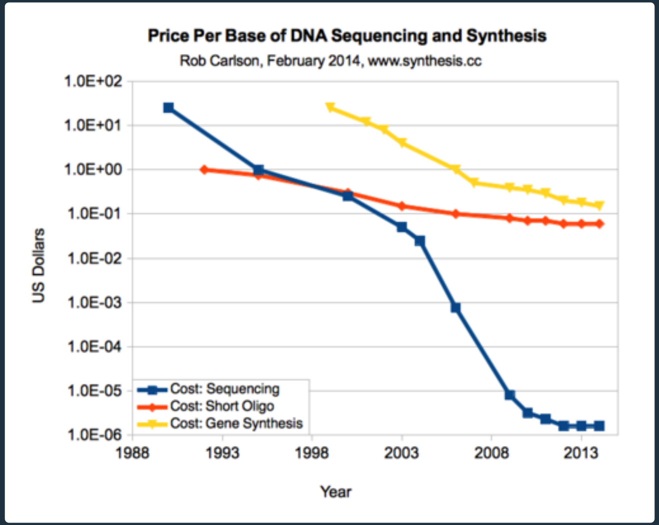
I.

Reducing Emerging Biological Risks

Technology Advances & Emerging Bio-Risks

Technology advances offer tremendous opportunities but also pose unique risks

- Easier to read, write, edit DNA & RNA
- AI & robotics enable automation & experimentation at scale
- Cloud labs
- AI & Engineering Living Systems

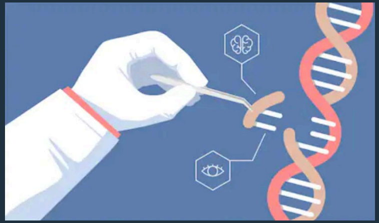

Price Per Base of DNA Sequencing and Synthesis

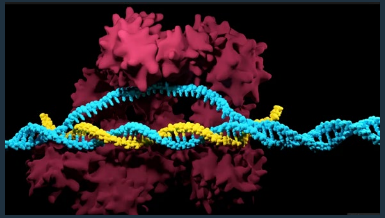
Rob Carlson, February 2014, www.synthesis.cc

TWIST BIOSCIENCE

华大基因 BGI

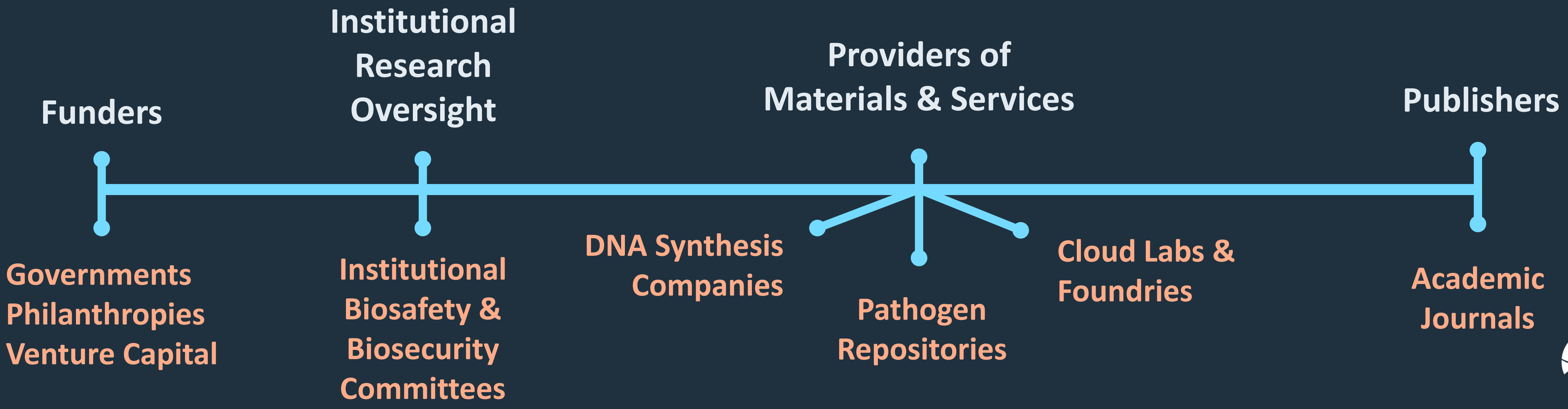
IDT INTEGRATED DNA TECHNOLOGIES



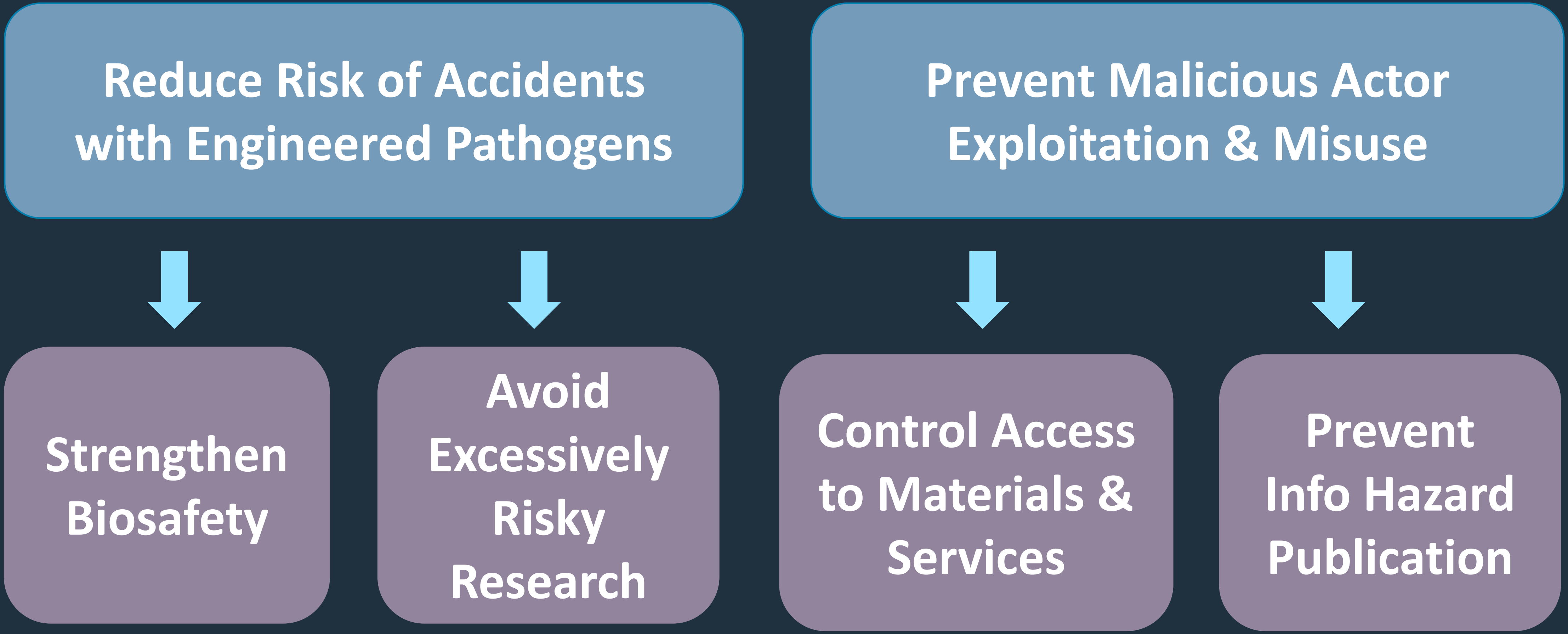


Bioscience Governance Solution Set

- Multiple intervention points throughout the bioscience research and development life-cycle
- Layered defense



Preventing Biotechnology Catastrophe



Reduce Risk of Accidents with Engineered Pathogens

Prevent Malicious Actor Exploitation & Misuse

Strengthen Biosafety

Avoid Excessively Risky Research

Control Access to Materials & Services

Prevent Info Hazard Publication

II.

Safeguarding AIxBio Capabilities

AIxBio Capabilities Risks

- Lower barriers to malicious actors causing harm with biology
- Increase the level of harm that a sophisticated malicious actor can cause with biology
- Reduce the effectiveness of biosecurity and biodefense measures

2023 Report: Convergence of AI and the Life Sciences

NTI bio

OCTOBER 2023

The Convergence of Artificial Intelligence and the Life Sciences:

Safeguarding Technology, Rethinking Governance, and Preventing Catastrophe

Sarah R. Carter, Ph.D.
Nicole E. Wheeler, Ph.D.
Sabrina Chwalek
Christopher R. Isaac, M.Sc.
Jaime Yassif, Ph.D.

- What are current and anticipated AI capabilities for engineering biology?
- What are the biosecurity implications of these developments?
- What are the most promising options for governing AI-bio capabilities?

Research Process

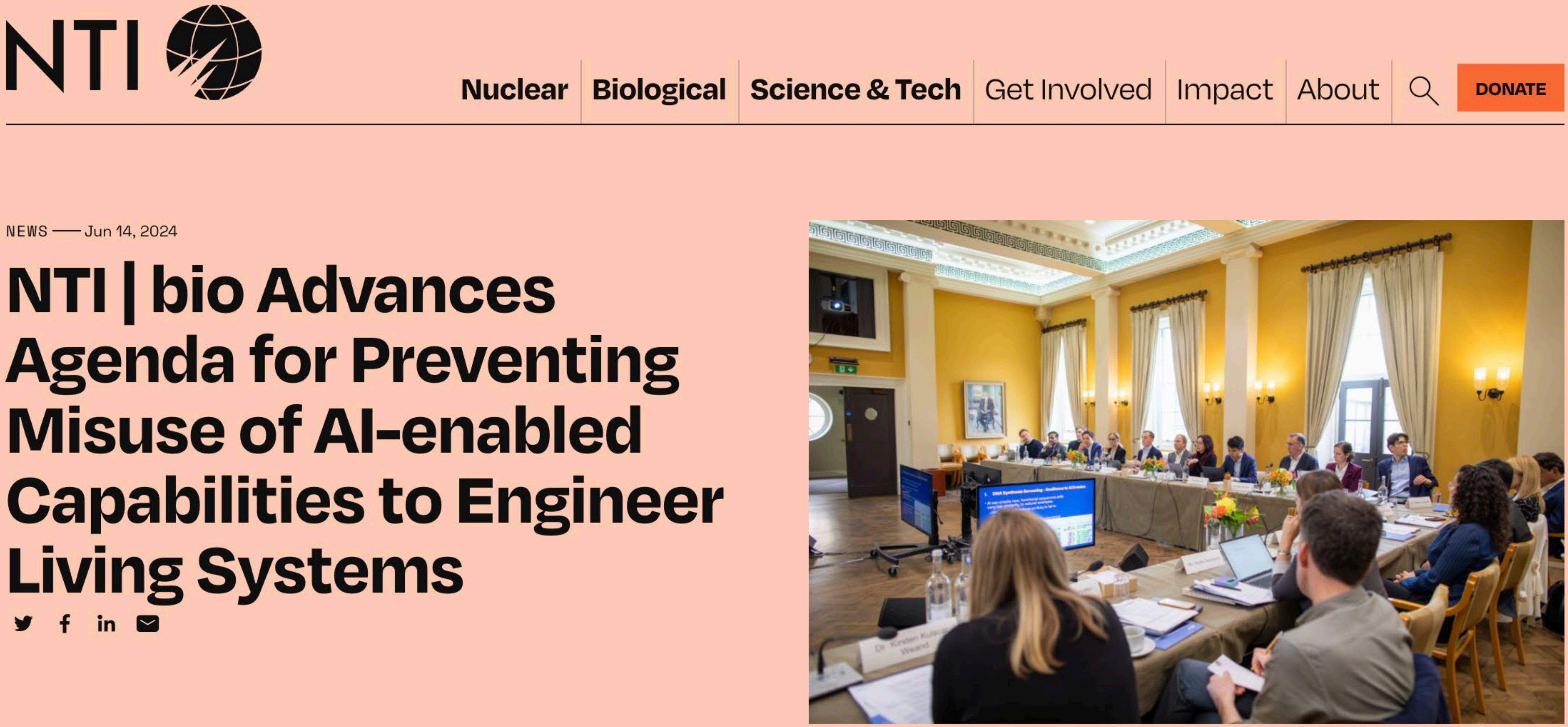
- Interviews with 30+ experts
- Workshop to discuss preliminary findings
- Peer review

Current and Anticipated Capabilities

Types of AI models

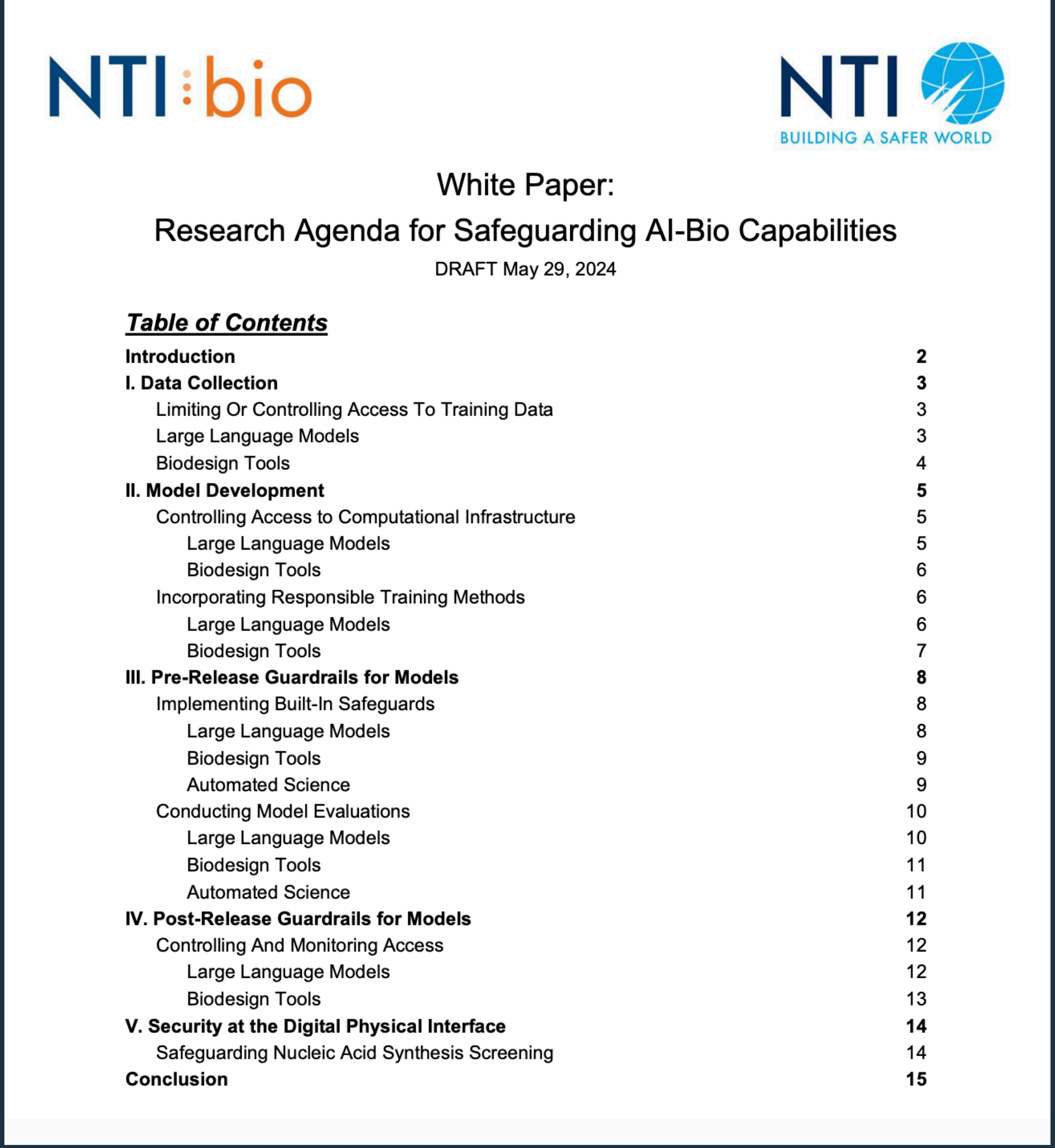
- Large Language Models
- AI Biodesign Tools
- Automated Science

AI-bio Governance	AI-bio Safeguards	Biosecurity & Preparedness
International “AI-Bio Forum”	Implement promising AI model guardrails at scale	Strengthen biosecurity at digital-physical interface
Develop new, more agile approach to national governance	Ambitious research agenda to explore additional AI guardrails	Use AI tools to build next-generation pandemic response capabilities



NEWS —— Jun 14, 2024

NTI | bio Advances Agenda for Preventing Misuse of AI-enabled Capabilities to Engineer Living Systems



NTI | bio

NTI BUILDING A SAFER WORLD

White Paper: Research Agenda for Safeguarding AI-Bio Capabilities

DRAFT May 29, 2024

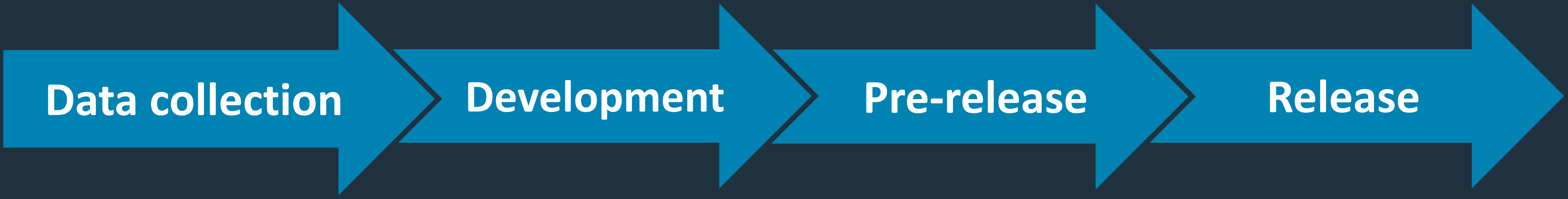
Table of Contents

III.

Technical Guardrails

Opportunities for Risk Reduction

Guardrails for AI models



Data collection

- Control access to training data

Development

- Control access to compute
- Responsible training methods

Pre-release

- Model evaluations
- Fine-tune with human feedback

Release

- Control model access
- Monitor models

https://www.nti.org/analysis/articles/developing-guardrails-for-ai-biodesign-tools/

Pilot Project
Built-in Guardrails

- Coupling designs with Metadata
  - Captures more of users’ intent
  - Pilot project with Lattice Automation

Pilot Project Managed Access

- AIxBio model dissemination is a key consideration for biosecurity

- Managed access can play an important role in safeguarding tools against misuse

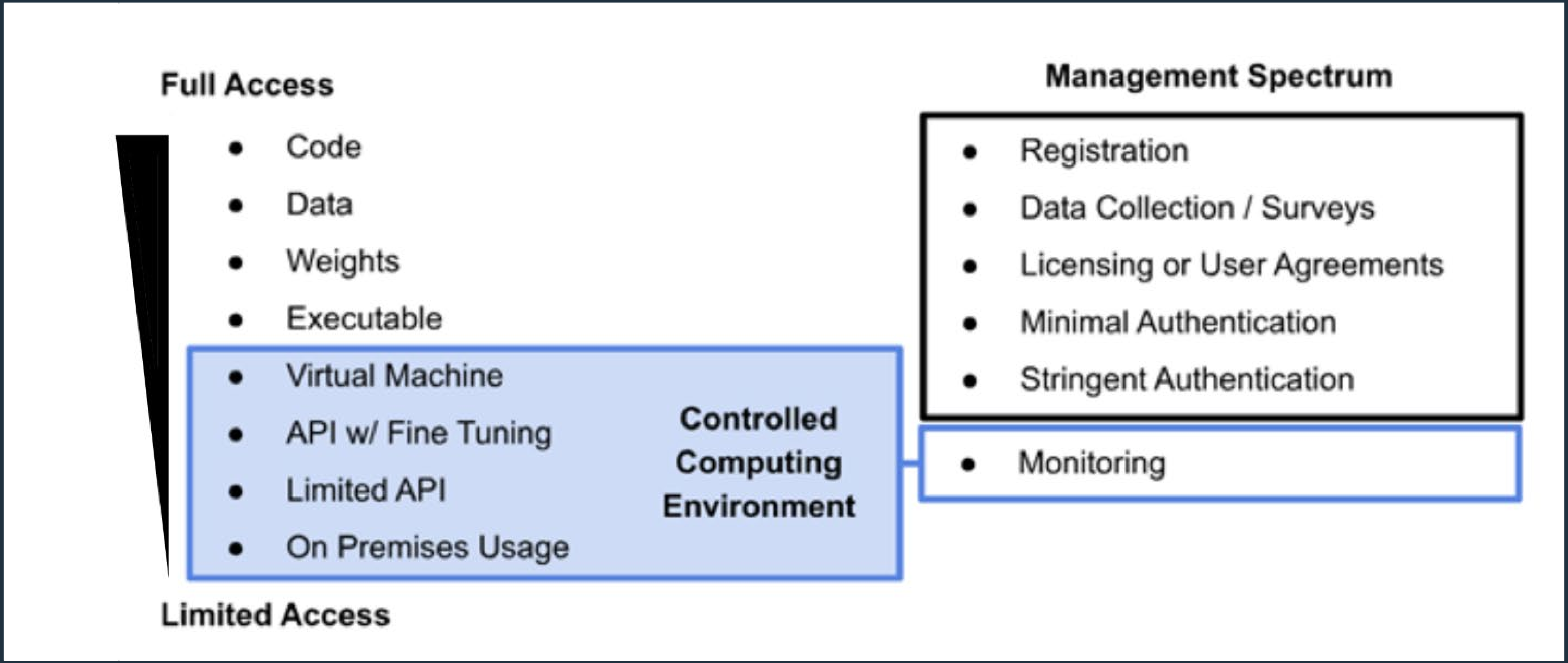
- Could help build public trust in scientific communities

Managed Access Principles

- It is important & feasible to balance security with equitable access to tools – benefits depend on access
  - CEPI is establishing managed access frameworks which prioritize equitable access along with responsible dissemination & biosecurity
- Tiered frameworks will be essential
  - Some tools are low risk and will not need managed access
  - For tools with higher risk:
    - Users should meet some criteria for access
    - More access should require more stringent criteria

NTI 2024 Report on Guardrails for AI Biodesign Tools

Options for Managed Access



Full Access

- Code
- Data
- Weights
- Executable
- Virtual Machine
- API w/ Fine Tuning
- Limited API
- On Premises Usage

Controlled Computing Environment

Limited Access

Management Spectrum

- Registration
- Data Collection / Surveys
- Licensing or User Agreements
- Minimal Authentication
- Stringent Authentication

- Monitoring

Pilot Projects
Managed Access

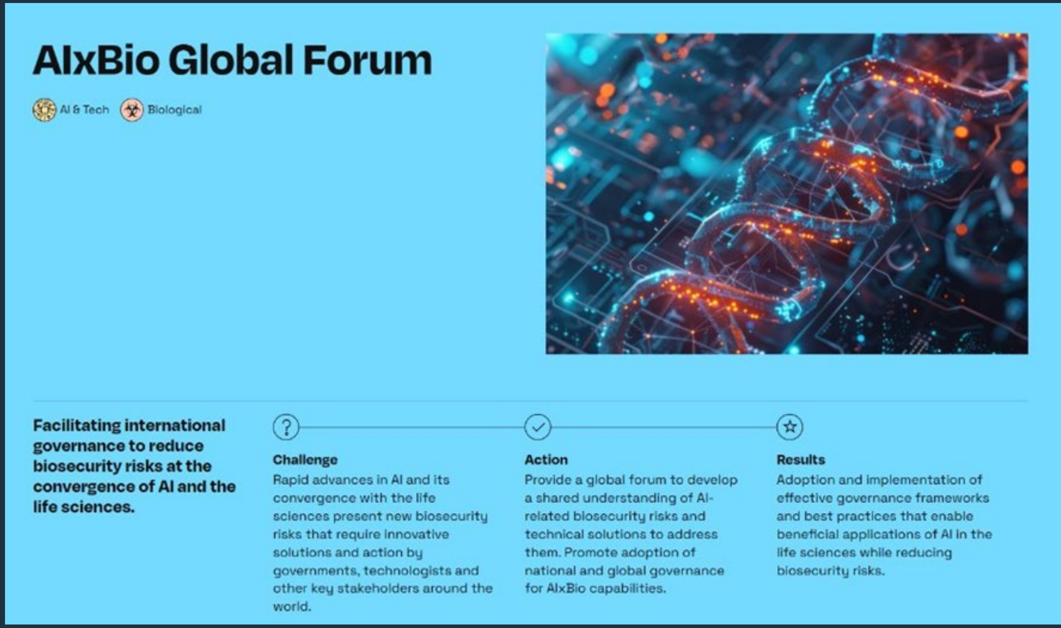
- Written framework on managed access to BDTs
  - Possibilities & best practices
  - Community engagement as a central pillar
  - Explore provision of security features + advantages for users and developers
  - Compute, ease of use, features for documentation and collaboration
- Development of web platform for protein design tools

IV.

AIxBio Global Forum

Goals of the AIxBio Forum

- Develop a shared understanding of biosecurity risks related to AI & needed safeguards
- Support development & dissemination of tools & practices to safeguard AIxBio capabilities
- Promote adoption of national & global governance mechanisms



AIxBio Global Forum

AI & Tech · Biological

Facilitating international governance to reduce biosecurity risks at the convergence of AI and the life sciences.

Challenge
Rapid advances in AI and its convergence with the life sciences present new biosecurity risks that require innovative solutions and action by governments, technologists and other key stakeholders around the world.

Action
Provide a global forum to develop a shared understanding of AI-related biosecurity risks and technical solutions to address them. Promote adoption of national and global governance for AIxBio capabilities.

Results
Adoption and implementation of effective governance frameworks and best practices that enable beneficial applications of AI in the life sciences while reducing biosecurity risks.

https://www.nti.org/about/programs-projects/project/aixbio-global-forum/

Roles of the Forum

I. Develop & disseminate best practices and tools for AIxBio model developers and users
   - Best practices can evolve into standards over time

II. Information sharing & outreach
   - High-level statement on risks
   - Webinars or explainers for policy makers & other key stakeholders
   - Clearinghouse for ideas

III. Help community maintain strategic approach
   - Advance comprehensive research agenda
   - Develop big picture ideas for channeling community expertise and effort to meet high-level goals
   - Develop roadmaps to tackle hard problems

IV. Develop recommendations for policymakers

AIxBio Global Forum History

- April 2024: First full meeting
  - Refined the goals of the Forum
  - Established two Working Groups
    - Horizon Scanning, Risk Assessment & Evaluation
    - Biodesign Tools
- September-October 2024: Working Groups
- December 2024: Second full meeting
  - Review Working Group activities
  - Call for a high-level statement on risk
  - Discussion of public-facing role of the Forum
- March-April 2025: Working Groups
- April 2024: Third full meeting

AIxBio Governance Three-Legged Stool

1. Practical, technical risk reduction solutions

2. Global platforms to share best practices

3. Governments & private funders promote compliance through incentives, guidelines, regulations

Thank you!

Contact: nti-bio@nti.org

Web: https://www.nti.org/area/biological/

NTI

BUILDING A SAFER WORLD

NTI bio