# Global Risk Index

*AI-Enabled Biological Tools*

ACHS Webinar| January 2026

RAND

THE CENTRE FOR LONG-TERM RESILIENCE

Sana Zakaria| RAND Europe

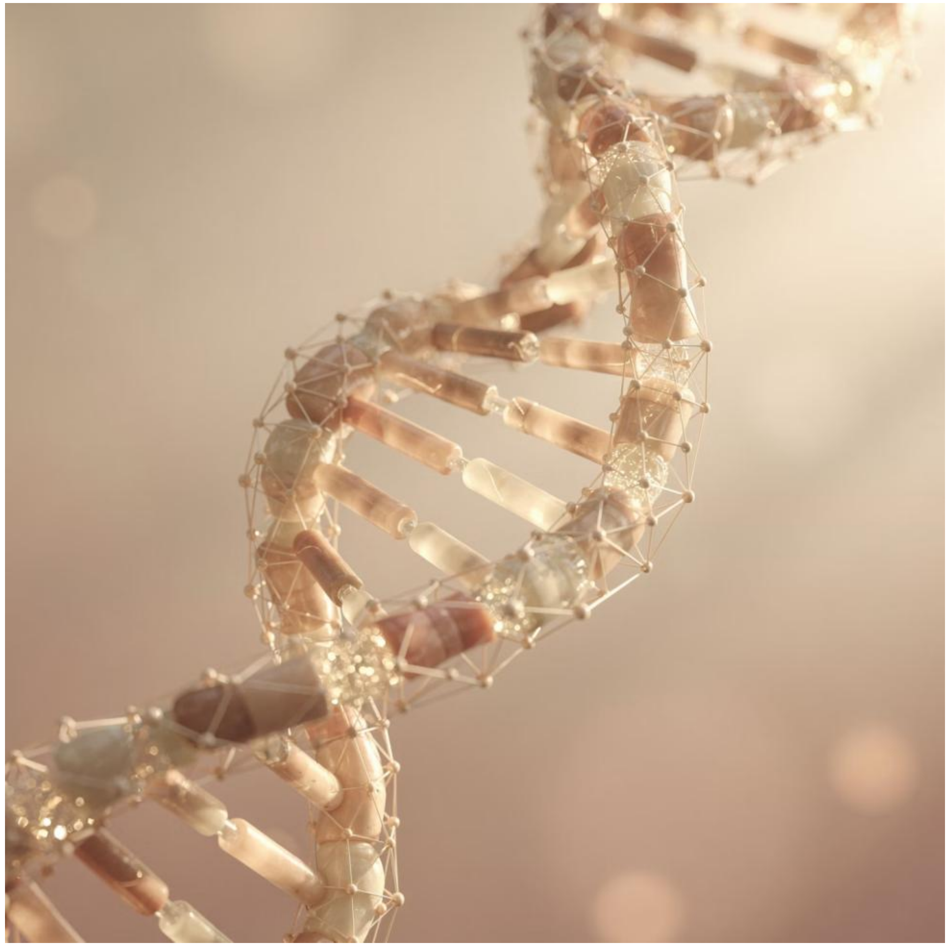
# Dual nature of AI-enabled Biological Design Tools (BDTs)

- BDTs could accelerate drug discovery and vaccine by providing insights and capabilities that once required years of painstaking experimental work

- Improved ability to understand and engineer biological systems

- This transformative power comes with profound dual-use risks: The same models that advance biomedical research could be repurposed for harmful applications, including the development of novel pathogens or the circumvention of existing countermeasures

- This dual-use potential demands vigilant oversight and structured monitoring and development of sfeguards



# BDTs: a different threat surface to LLMs

### Direct biological action

BDTs operate directly on biological design space, manipulating genetic sequences, protein structures, and molecular pathways with precision.

### Compressed iteration cycles

They compress years of experimental iteration into data and prediction-driven workflows, dramatically accelerating the discovery and optimisation of biological capabilities.

### Capability acceleration risk

The primary risk here is not misinformation or persuasion-it's the accelerated development of novel biological capabilities in the wrong hands.

## Technical complexity ≠ Safety

Risk fundamentally changes when BDTs are:

- Connected to general-purpose AI for planning and reasoning
- Embedded in agentic systems that automate iteration and optimisation

**Key insight:** Integration could lower misuse barriers

# Scaling of risks across the BDT ecosystem

### BDTs don't fit LLM-centric safety models

Existing AI governance frameworks were designed for LLMs and focus on content moderation, alignment, content filters and harm refusals.

### The category itself is shifting

Boundaries between BDTs, biology foundation models, and general AI capabilities are blurring. Benchmarks, and definitions already lag behind creating gaps in oversight and risk assessment.

### Static categories miss risk at boundaries

As technological capabilities combine in novel ways, rigid classification schemes fail to capture edge cases.

## Modest resource requirements

Many BDTs can be developed with modest compute resources and open biological datasets, lowering financial and technical barriers to entry.

## Rapid and irreversible diffusion

Open-source release enables rapid and irreversible global diffusion of capabilities. Once released, tools cannot be recalled or effectively restricted.

## Globally distributed development

Development is distributed across dozens of countries with varying regulatory regimes, making coordinated governance extremely challenging.

## LLM chokepoints may not apply

Unlike frontier AI models requiring massive compute infrastructure, BDTs often evade traditional control points such as hardware export restrictions.

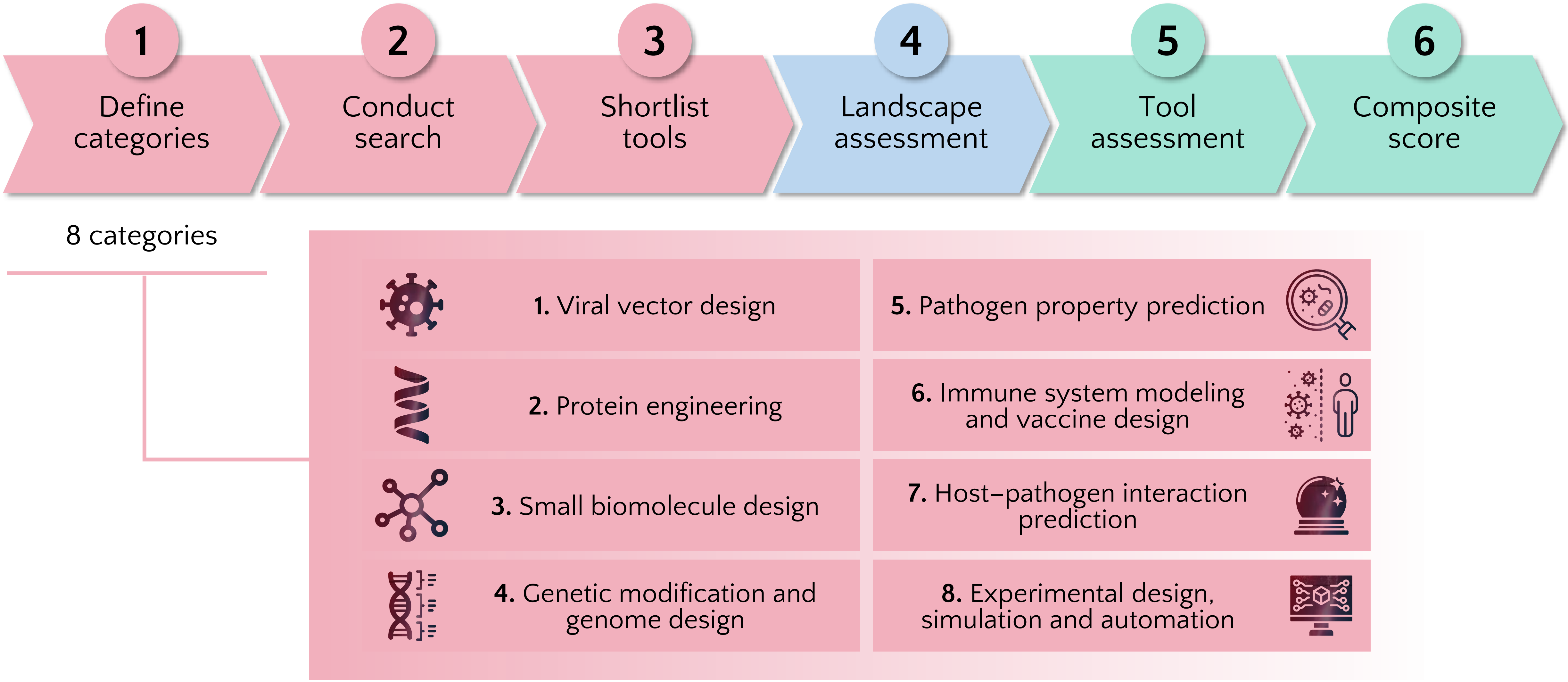
# Objectives

Develop a **structured** & **repeatable** approach for tool risk assessment

Present a current landscape assessment of state-of-the-art AI–bio tools

Helps prioritise key tools for deeper assessment (e.g. interactive, computational)

# Assessment Framework



1 Define categories

2 Conduct search

3 Shortlist tools

4 Landscape assessment

5 Tool assessment

6 Composite score

8 categories

**1.** Viral vector design

**2.** Protein engineering

**3.** Small biomolecule design

**4.** Genetic modification and genome design

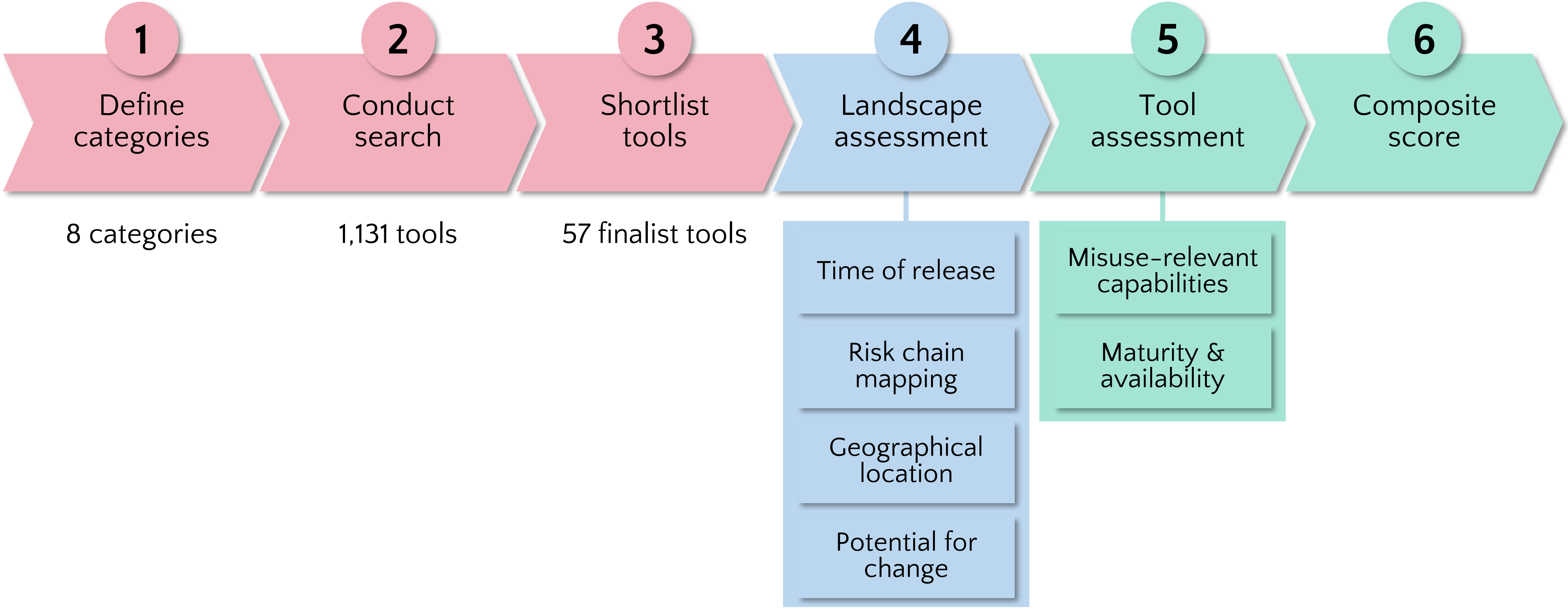
**5.** Pathogen property prediction

**6.** Immune system modeling and vaccine design

**7.** Host–pathogen interaction prediction

**8.** Experimental design, simulation and automation

# Assessment Framework



1 Define categories

8 categories

2 Conduct search

1,131 tools

3 Shortlist tools

57 finalist tools

4 Landscape assessment

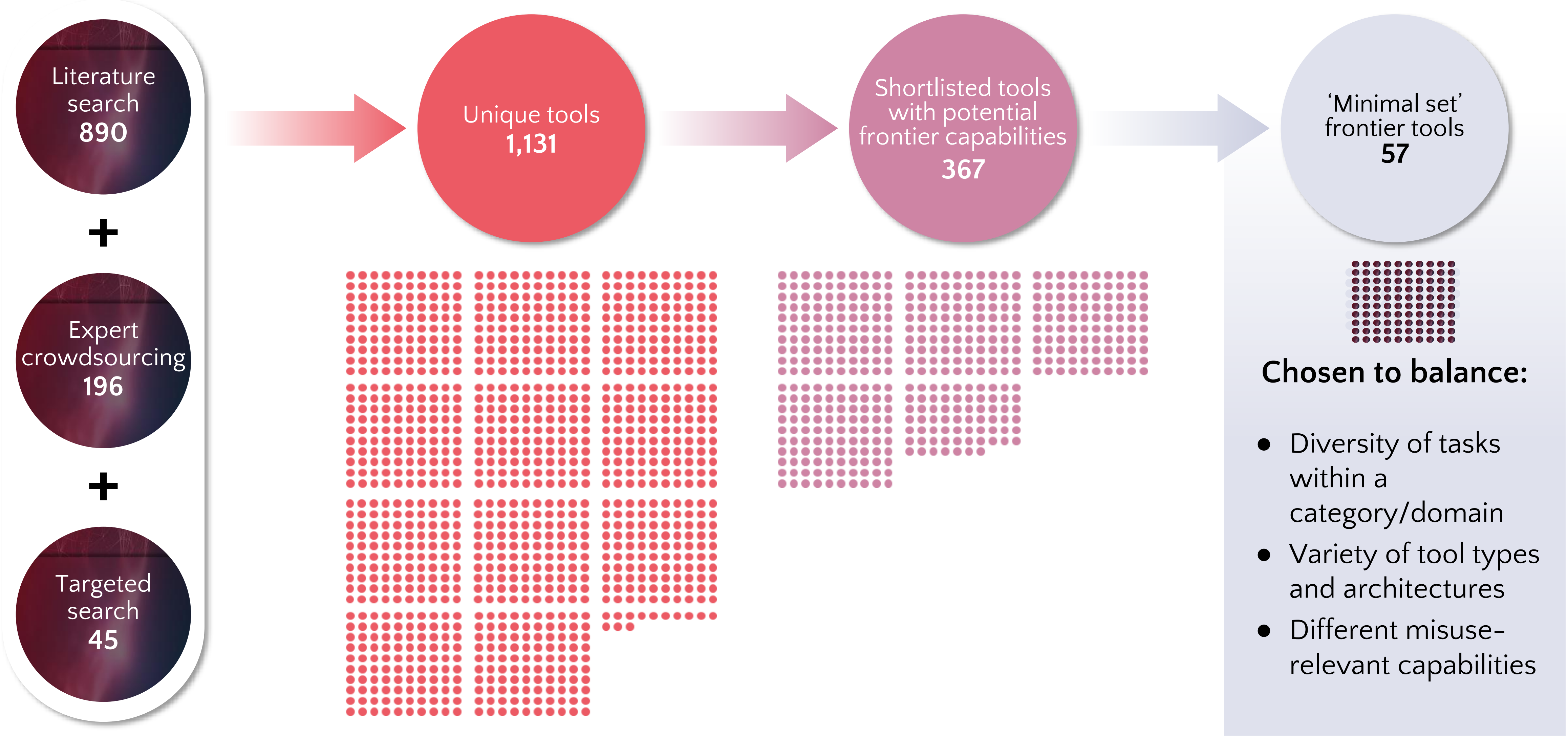
- Time of release
- Risk chain mapping
- Geographical location
- Potential for change

5 Tool assessment

- Misuse-relevant capabilities
- Maturity & availability

6 Composite score

# Search and shortlisting methodology



Literature search
890

+

Expert crowdsourcing
196

+

Targeted search
45

Unique tools
1,131

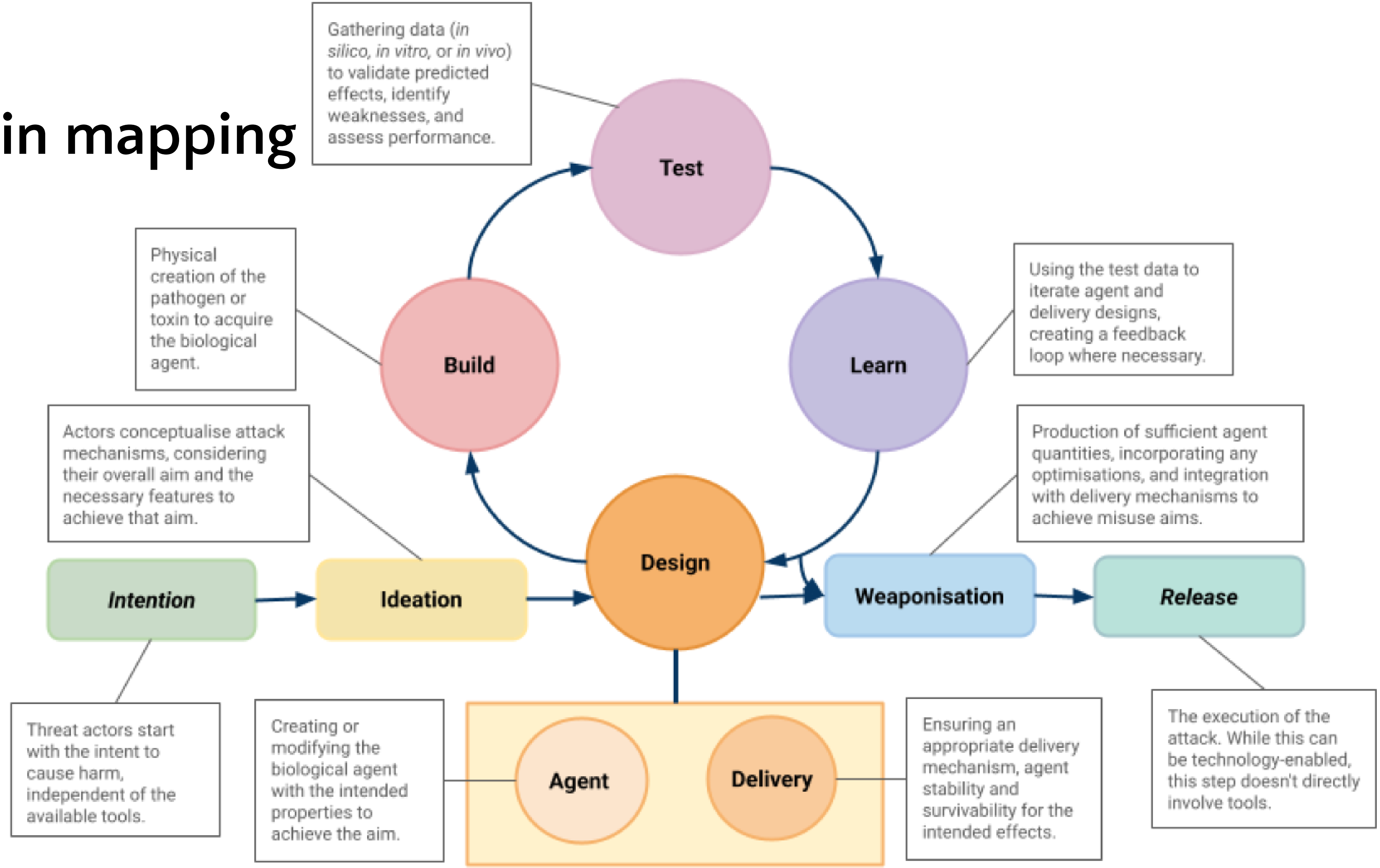
Shortlisted tools with potential frontier capabilities
367

‘Minimal set’ frontier tools
57

**Chosen to balance:**

- Diversity of tasks within a category/domain
- Variety of tool types and architectures
- Different misuse-relevant capabilities

# Risk chain mapping



*Intention*

Threat actors start with the intent to cause harm, independent of the available tools.

Ideation

Actors conceptualise attack mechanisms, considering their overall aim and the necessary features to achieve that aim.

Design

Agent

Creating or modifying the biological agent with the intended properties to achieve the aim.

Delivery

Ensuring an appropriate delivery mechanism, agent stability and survivability for the intended effects.

Build

Physical creation of the pathogen or toxin to acquire the biological agent.

Test

Gathering data (*in silico, in vitro*, or *in vivo*) to validate predicted effects, identify weaknesses, and assess performance.

Learn

Using the test data to iterate agent and delivery designs, creating a feedback loop where necessary.

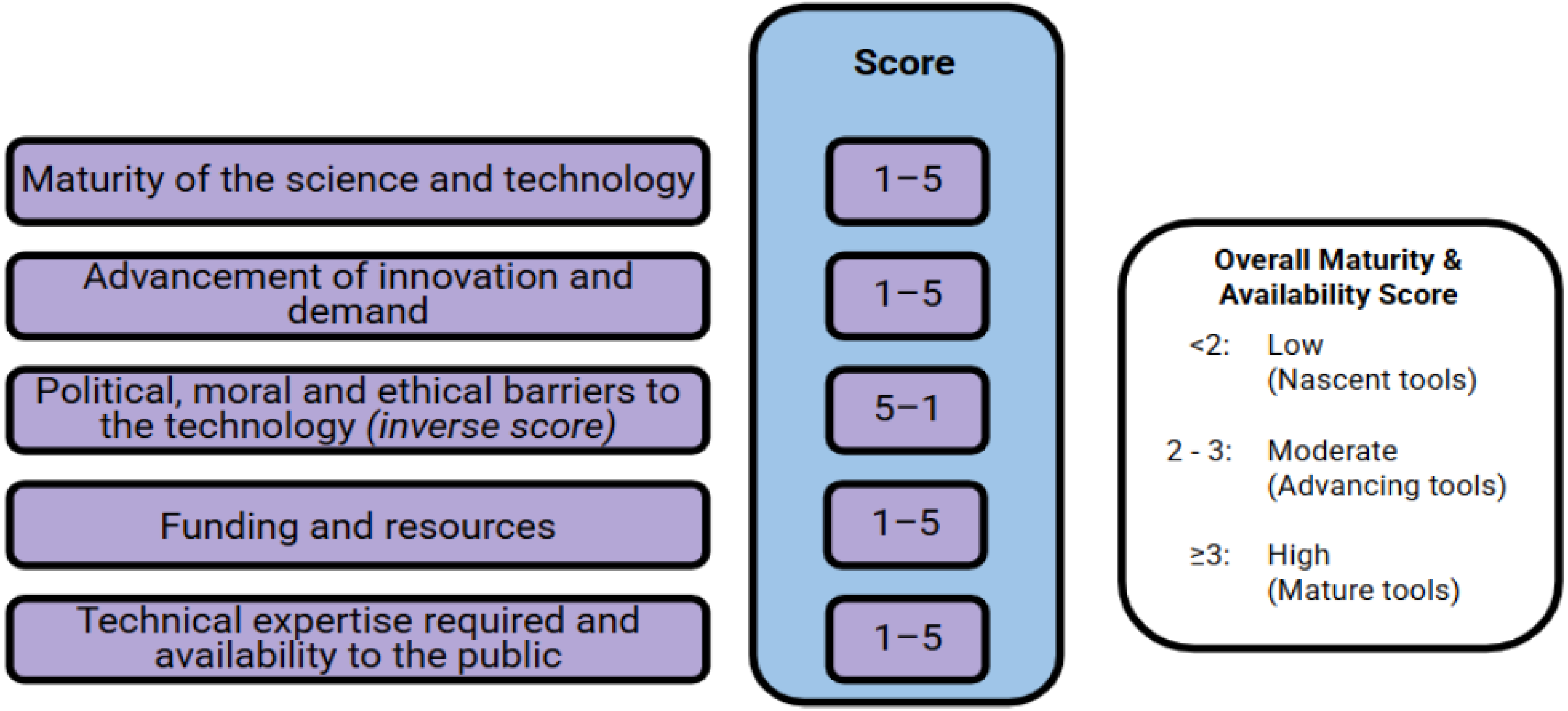
Weaponisation

Production of sufficient agent quantities, incorporating any optimisations, and integration with delivery mechanisms to achieve misuse aims.

*Release*

The execution of the attack. While this can be technology-enabled, this step doesn't directly involve tools.

# Maturity and availability



|  | Score |
|---|---|
| Maturity of the science and technology | 1–5 |
| Advancement of innovation and demand | 1–5 |
| Political, moral and ethical barriers to the technology *(inverse score)* | 5–1 |
| Funding and resources | 1–5 |
| Technical expertise required and availability to the public | 1–5 |

**Overall Maturity & Availability Score**

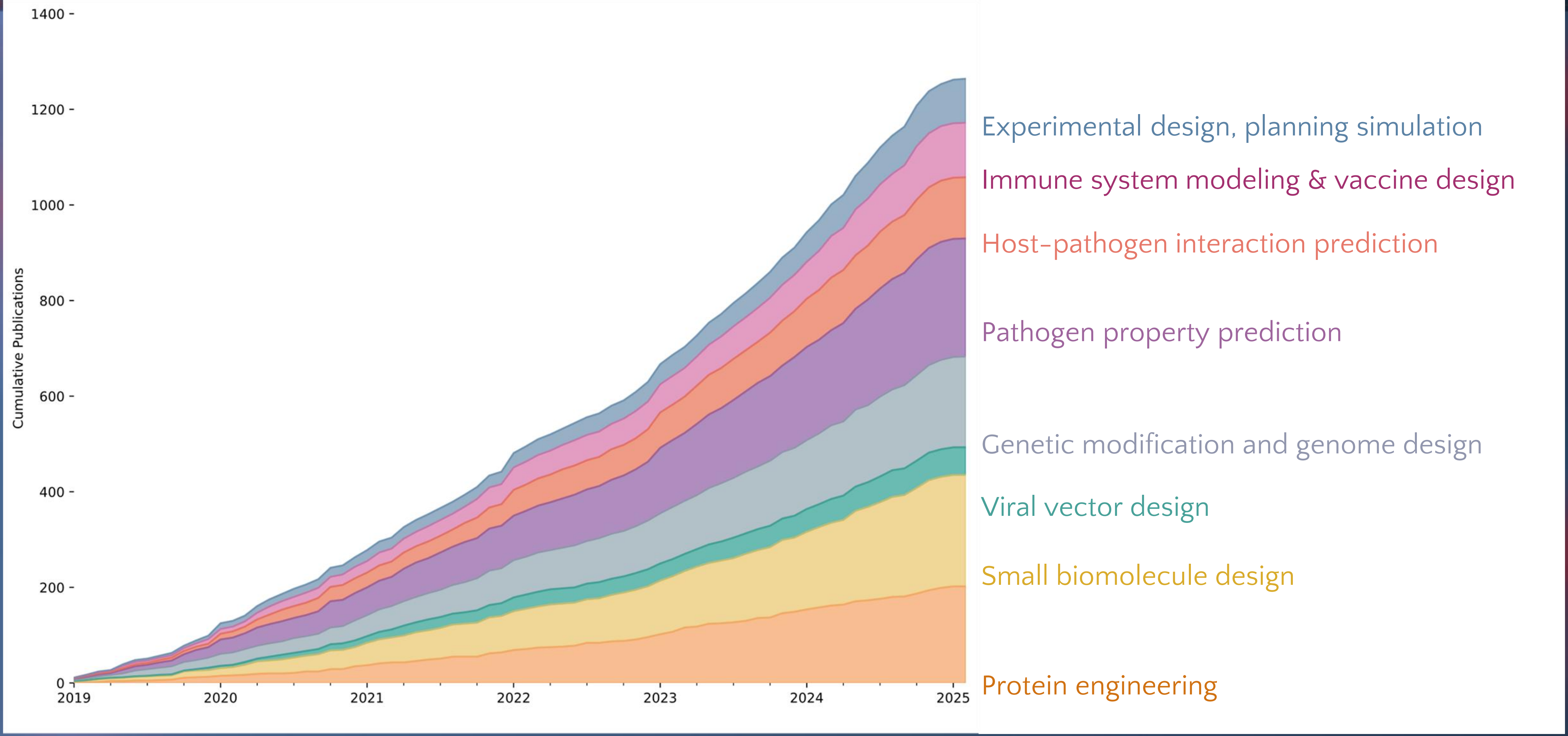
| | |
|---|---|
| <2: | Low<br>(Nascent tools) |
| 2 - 3: | Moderate<br>(Advancing tools) |
| ≥3: | High<br>(Mature tools) |

This approach was adapted from Gerstein et al. 2019.[118] Source: RAND and CLTR analysis 2025.

| Rating | Criteria |
|---|---|
| Red | Misuse-relevant capability = Critical |
| | Misuse-relevant capability = High<br>AND<br>Maturity and availability ≥ 2 |
| | Misuse-relevant capability = Medium<br>AND<br>Maturity and availability ≥ 3 |
| Amber | Misuse-relevant capability = High<br>AND<br>Maturity and availability < 2 |
| | Misuse-relevant capability = Medium<br>AND<br>Maturity and availability ≥ 2 and < 3 |
| Green | Misuse-relevant capability = Medium<br>AND<br>Maturity and availability < 2 |
| | Misuse-relevant capability = Low or Very Low |

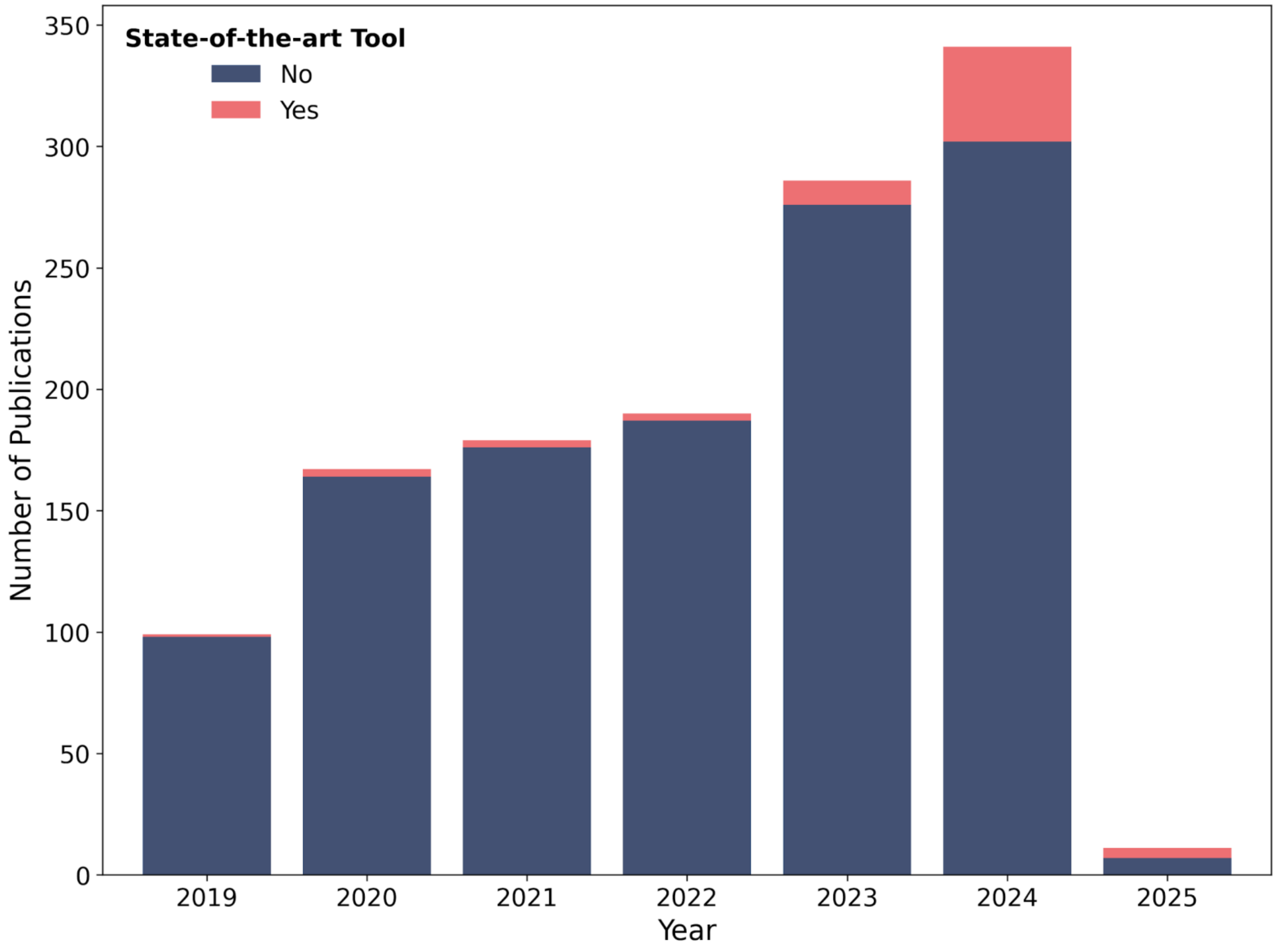
Source: RAND and CLTR analysis 2025.

# All tool categories are growing rapidly

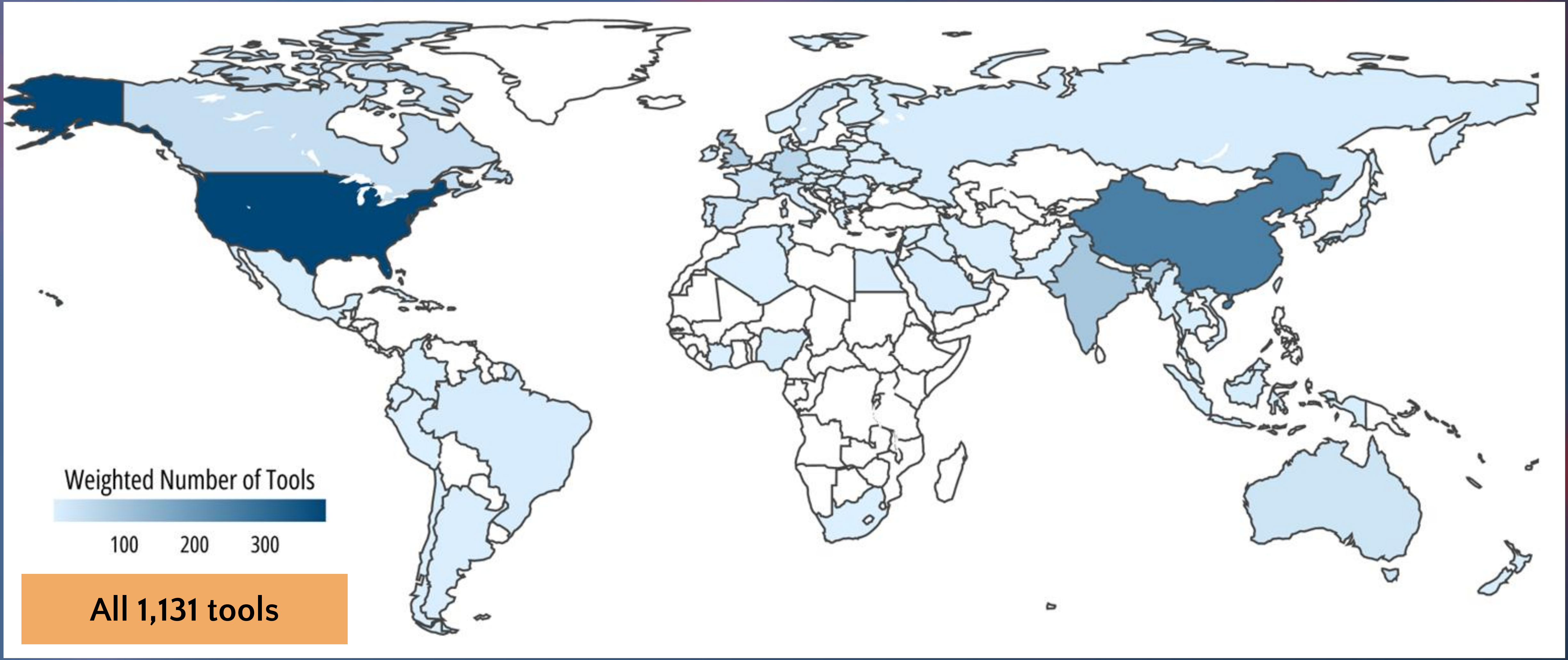


# Acceleration of frontier BDTs

Majority finalist frontier tools emerged in the last 18 months (at the time of research)

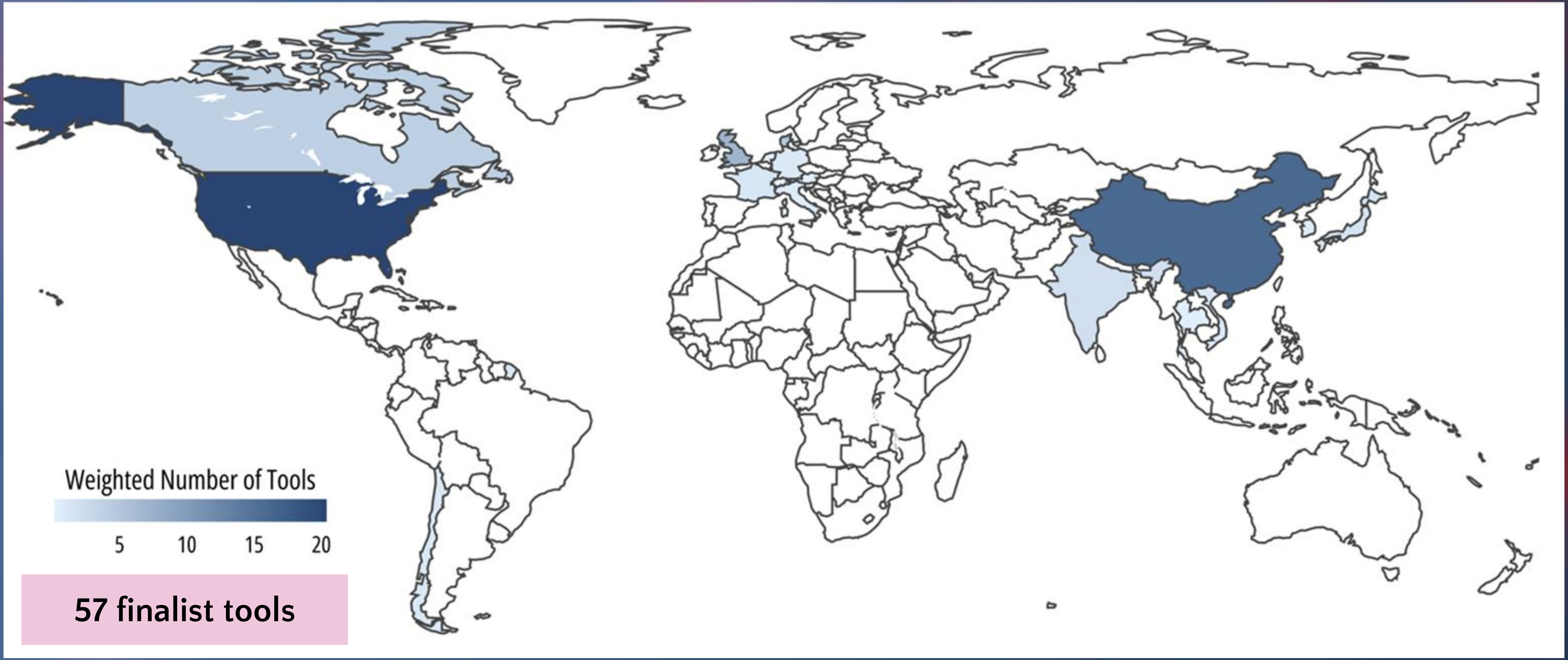


# Tools are developed globally (across 76 countries)



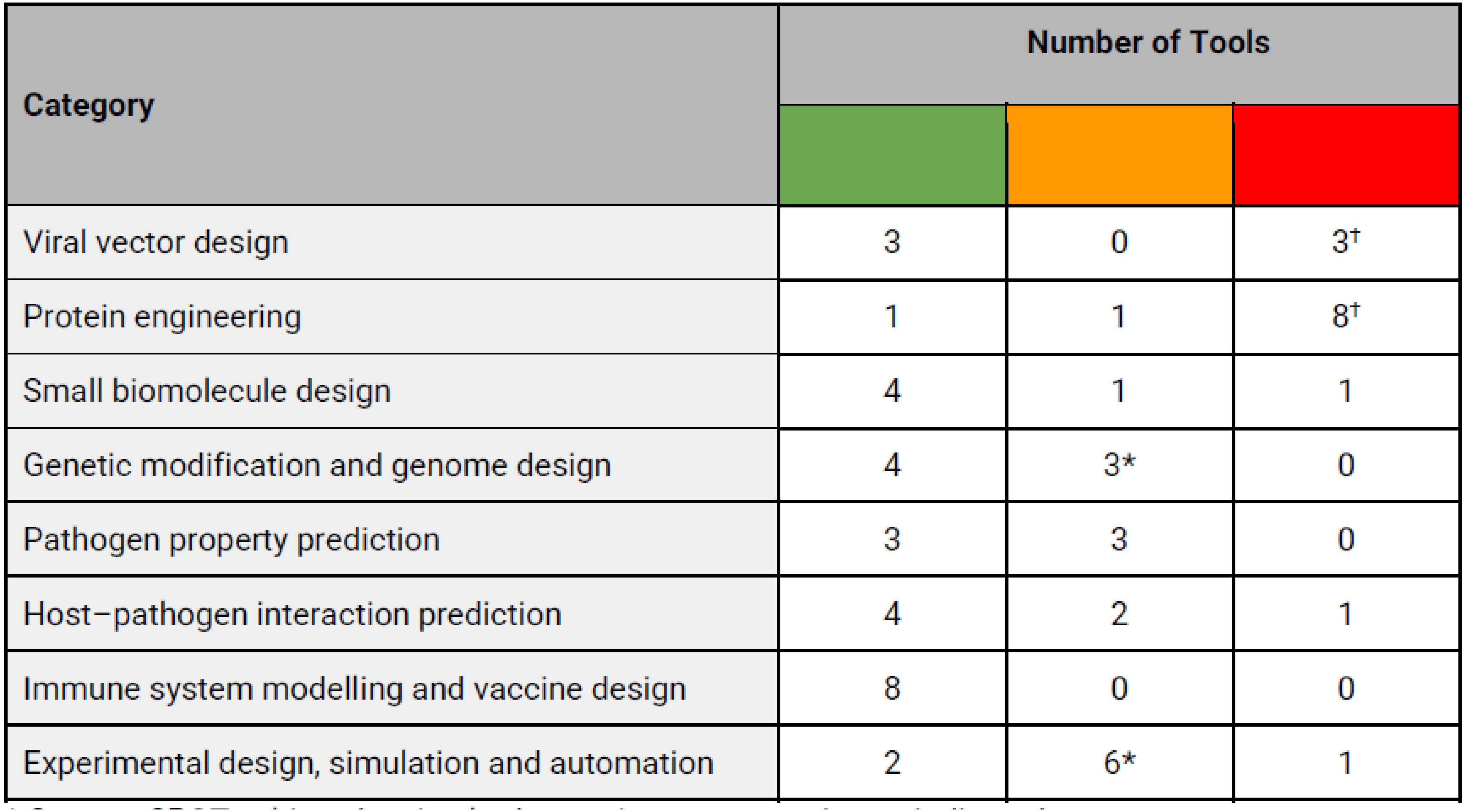
All 1,131 tools

# Frontier tool development is distributed across 24 countries

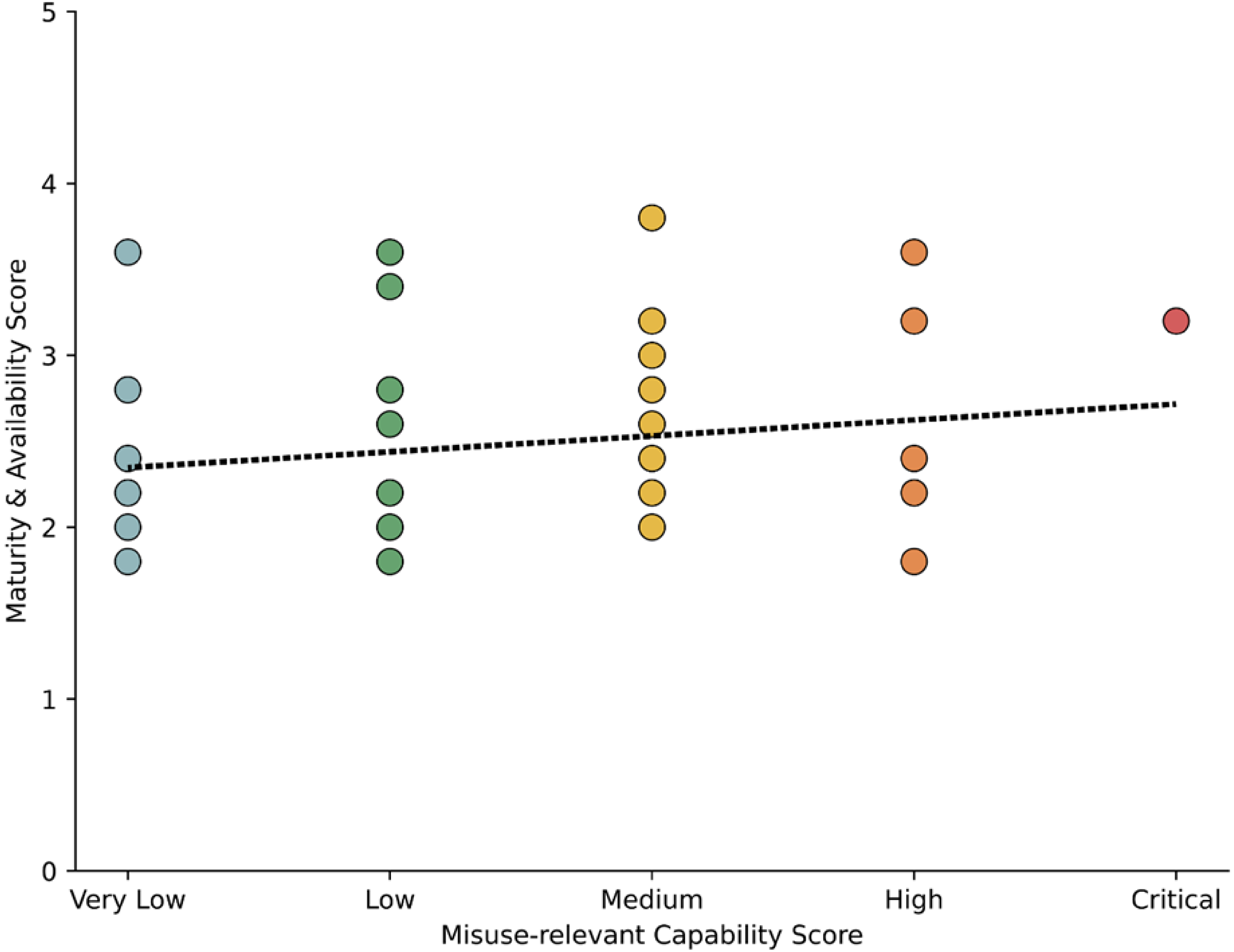


57 finalist tools

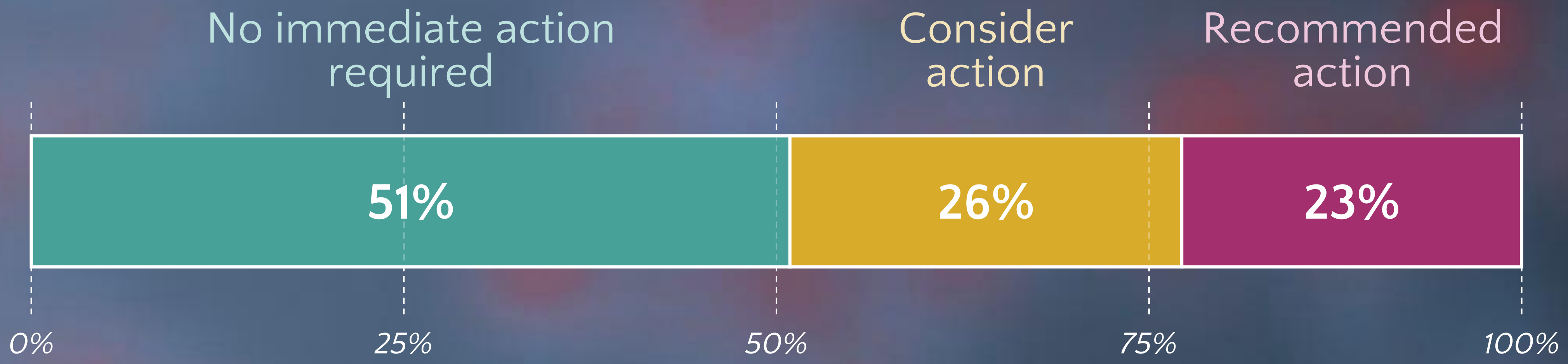
# Composite assessment of tools across categories



| Category | Number of Tools | | |
|---|---|---|---|
| | 🟩 | 🟧 | 🟥 |
| Viral vector design | 3 | 0 | 3[†] |
| Protein engineering | 1 | 1 | 8[†] |
| Small biomolecule design | 4 | 1 | 1 |
| Genetic modification and genome design | 4 | 3* | 0 |
| Pathogen property prediction | 3 | 3 | 0 |
| Host–pathogen interaction prediction | 4 | 2 | 1 |
| Immune system modelling and vaccine design | 8 | 0 | 0 |
| Experimental design, simulation and automation | 2 | 6* | 1 |


5
4
3
2
1
0
Maturity & Availability Score
Very Low
Low
Medium
High
Critical
Misuse-relevant Capability Score

# Nearly 1 in 4 scored ‘Recommended action’



Over 60% of recommended action tools are open source

# Recent demos of capabilities leaps

## AI generated sequences can evade screening tools

Microsoft researchers identified that current biosecurity screening tools used by DNA synthesis firms can be bypassed by AI-generated protein-encoding sequences that preserve harmful structures while appearing benign to filters, highlighting a critical vulnerability and prompting development of updated detection methods.

## AI-designed viral genomes were successfully synthesised into bacteriophages

Researchers used generative AI models to design complete viral genomes de novo and then successfully synthesised these sequences into functional bacteriophages that could infect and kill *E. coli* bacteria in the lab, demonstrating AI's capacity to author viable viral agents.

# Recommendations

- Urgently assess priority tools in greater depth (evals + red-teaming)
- Invest in effective mitigations (technical safeguards + oversight)
- Conduct periodic assessment and monitor trends
- Enhance future assessments (cost-benefit analysis, threat modelling)
- Coordinate international governance efforts

# Risk Index Observatory (RIO)

RIO will establish a repeatable mechanism for monitoring and assessing risks from AI-enabled biotools, transforming insights into structured, actionable intelligence.

### Tool identification

### Risk assessment

Identify and analyse which tools matter for biosecurity oversight, why, and how urgently

### Analytics and automation

Converts tool level data into landscape trend analyses and set up automation of new tools identification and flags for assessment

### Stakeholder support

Engage with and support stakeholders (governments, industry, researchers) in understanding and acting on risk assessments through briefings and a misuse risk calculator

# Thank You

RAND