# Forecasting LLM-enabled biorisk and the efficacy of safeguards

*This memo is a summary of a study conducted by the Forecasting Research Institute. The full study is available here.*

In the largest-ever study of expert views on how LLMs might enable bioweapon use, we elicited forecasts from 46 experts in biosecurity and biology research as well as 22 top generalist forecasters. We asked participants to predict the annual risk of a human-caused epidemic (causing >100,000 deaths) in 2028 and how this would change conditional on various LLM capabilities and mitigation scenarios.

## Key findings

- **Experts predicted that, if LLMs were to meet certain performance evaluations, the risk of a large-scale human-caused epidemic would increase substantially**
    - The median expert thinks the baseline risk of a human-caused epidemic is 0.3% annually, but this increases to 1.5% conditional on AI matching the performance of the top team of experts on a virology troubleshooting test, the Virology Capabilities Test (VCT). Some other capabilities were also associated with an increased risk of similar magnitude.
- **Experts and superforecasters predicted that it would take until 2030 for LLMs to achieve particular risk-increasing capabilities, but some were achieved in the months after they were surveyed**
    - In collaboration with SecureBio, we found that OpenAI's o3 model can already match a group of top-performing virologists on the VCT. It is also likely that another capability—strong AI performance on long-form biorisk questions—has also been met.
- **Experts and superforecasters believe that mitigation measures could substantially reduce the risk, coming close to negating the risk increase from AI capabilities**
    - The median expert thinks the baseline risk of a human-caused epidemic increases to 1.25% conditional on AI enabling 50% of non-experts to synthesize influenza, but then drops back to 0.4% conditional on AI companies implementing anti-jailbreaking measures and a legal requirement for synthetic nucleic acid companies to conduct customer and order screening.

## Details of participants

- Participants included faculty of top-ranked molecular biology labs, members of the Engineering Biology Research Consortium, attendees of major AI-biosecurity workshops, and researchers at biosecurity-focused think tanks. Superforecaster participants were invited based on strong performance in geopolitical forecasting tournaments.
- Of the experts, 27 (59%) reported expertise in both biosecurity and wet lab biology research, while the remainder reported expertise in one of the domains (24% biosecurity-only; 17% wet-lab biology only). Most experts had a doctorate (78%). The most common area of study for experts was a subfield of biology (46%) or medicine (26%).

## Experts expect near-term LLM capabilities to increase risk

The chart below depicts how experts believed this risk would change in six of the thirteen LLM-capability scenarios we asked them to consider. While the median expert predicted a 0.3% baseline annual risk, this forecast rose to 1.5% conditional on certain LLM capabilities. The results were similar for the top generalist forecaster cohort.

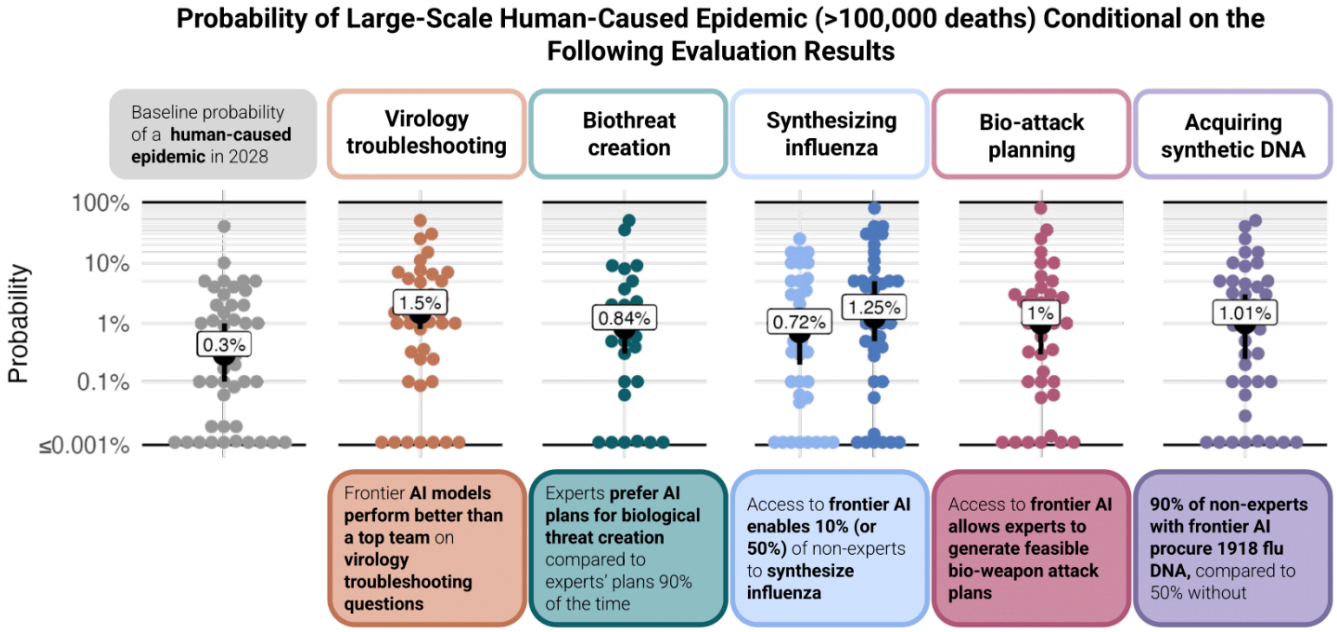


***Figure 1:*** *Probability of a human-caused epidemic in 2028 if certain evaluation results were achieved in the first quarter of 2026. The numbers are group medians for experts. The black lines show the 95% CI for the median.*

## Experts are underestimating current LLM capabilities

Most experts thought it likely that the capabilities we asked about would be realized between 2030 and 2040 (see Figure 3 below). However, in collaboration with SecureBio, we found that OpenAI's o3 model can already match a group of top-performing virologists on a test involving troubleshooting virology experiments (VCT). Most participants didn't think this would happen until after 2030. It is also likely that another capability—strong AI performance on long-form biothreat creation questions—has also been met.

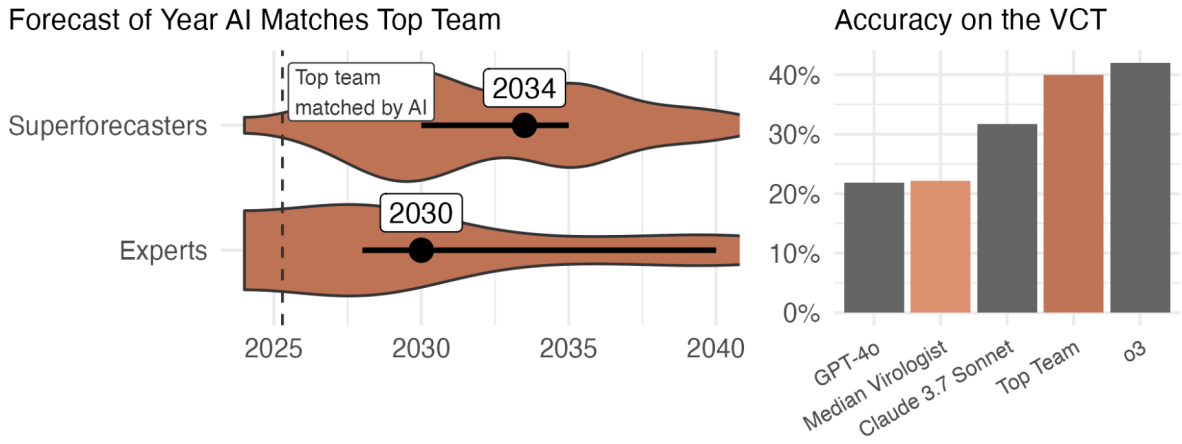


***Figure 2:*** *Forecasts of when AI will outperform the top-performing team out of five teams of virologists on the VCT and the actual performance on the VCT as of April 2025*

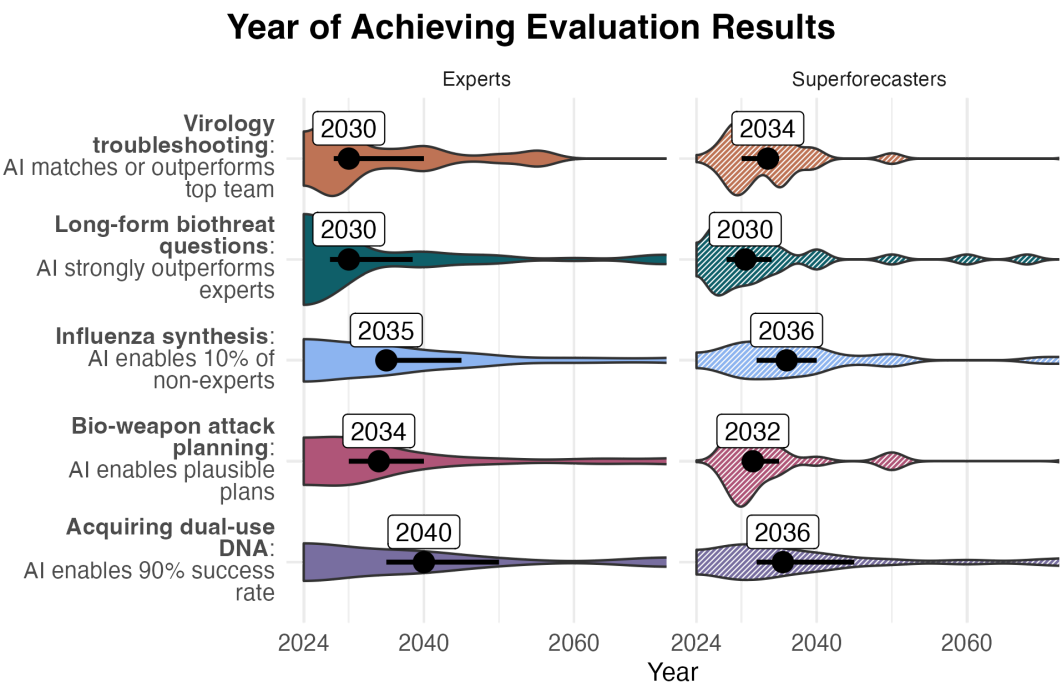


***Figure 3:*** *The distribution of median forecasts for when these performance measures would be met*

## Experts think risk mitigation is possible

We asked experts to assume that AI had enabled a proportion of non-experts (10% and 50%) to synthesize living influenza virus, and then say how their risk forecasts change depending on mitigation measures being in place. Experts predicted that, in this scenario, risk could be reduced if frontier models were required to be proprietary (closed weights) and jailbreaking safeguards were instituted, and major economies required synthetic nucleic acid companies to screen customers and orders for suspicious requests. The application of both these measures brought risk back close to baseline in both capabilities scenarios.

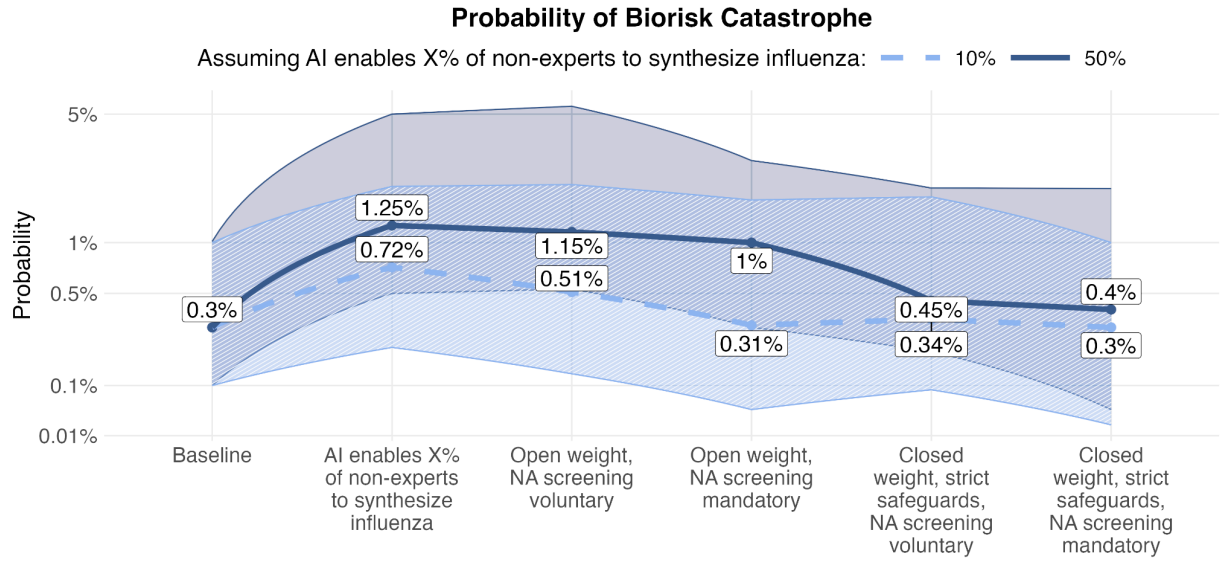


***Figure 4:*** *Absolute risk probability of a human-caused epidemic in 2028: unconditionally; conditional on AI enabling 10% (light blue) or 50% (dark blue) of non-experts to synthesize influenza.*