
SAFETY FEBRUARY 26, 2026

LLM NOVICE UPLIFT ON DUAL-USE, IN SILICO BIOLOGY TASKS

Chen Bo Calvin Zhang, Christina Q. Knight, Nicholas Kruus, Jason Hausenloy, Pedro Medeiros, Nathaniel Li, Aiden Kim, Yury Orlovskiy, Coleman Breen, Bryce Cai, Jasper Götting, Andrew Bo Liu, Samira Nedungadi, Paula Rodriguez, Yannis Yiming He, Mohamed Shaaban, Zifan Wang, Seth Donoughe, Julian Michael

A multi-benchmark human uplift study demonstrates that LLMs significantly empower novices to outperform internet-only controls and even experts on biosecurity-relevant tasks, highlighting a substantial narrowing of the barrier to expert-level biological knowledge despite existing model safeguards.

Large language models (LLMs) perform increasingly well on biology benchmarks, but it remains unclear whether they uplift novice users -- i.e., enable humans to perform better than with internet-only resources. This uncertainty is central to understanding both scientific acceleration and dual-use risk. We conducted a multi-model, multi-benchmark human uplift study comparing novices with LLM access versus internet-only access across eight biosecurity-relevant task sets. Participants worked on complex problems with ample time (up to 13 hours for the most involved tasks). We found that LLM access provided substantial uplift: novices with LLMs were 4.16 times more accurate than controls (95% CI [2.63, 6.87]). On four benchmarks with available expert baselines (internet-only), novices with LLMs outperformed experts on three of them. Perhaps surprisingly, standalone LLMs often exceeded LLM-assisted novices, indicating that users were not eliciting the strongest available contributions from the LLMs. Most participants (89.6%) reported little difficulty obtaining dual-use-relevant information despite safeguards. Overall, LLMs substantially uplift novices on biological tasks previously reserved for trained practitioners, underscoring the need for sustained, interactive uplift evaluations alongside traditional benchmarks.

LINKS

<https://arxiv.org/pdf/2602.23329v1>