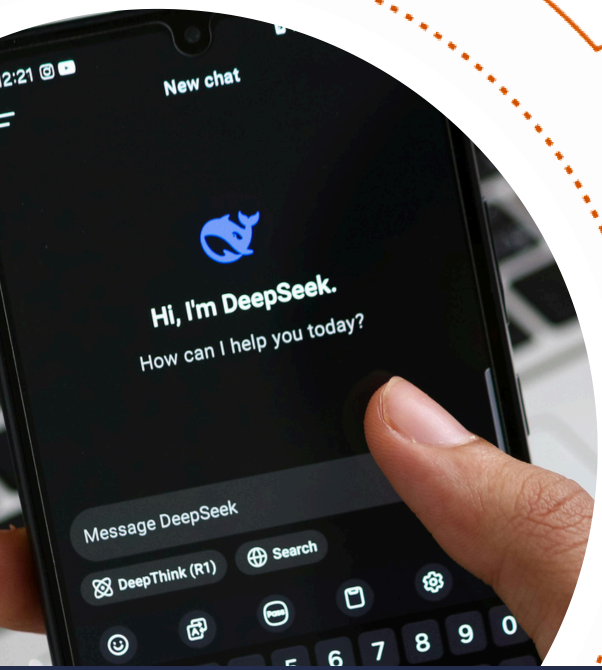
UVA | NATIONAL SECURITY DATA AND POLICY INSTITUTE

PRODUCT NO. 00011

DEEPER QUESTIONS RAISED BY DEEPSEEK

APRIL 18, 2025



NATIONALSECURITY.VIRGINIA.EDU | NSDPI@VIRGINIA.EDU

UVA | NATIONAL SECURITY DATA AND POLICY INSTITUTE

Product No. 00011

Deeper Questions Raised by DeepSeek

April 18, 2025

Did DeepSeek Evade Export Controls?

We have no technical evidence to indicate that DeepSeek circumvented export controls. We cannot independently confirm the company's claim that they used 2,048 NVIDIA H800s (designed to comply with export controls), but we calculate that their training methods would enable the reported efficiencies using that hardware. *Furthermore, proving that export-controlled hardware was used to achieve a claimed outcome through reverse engineering is challenging.*

- It is presently challenging to work backward from a claimed outcome to certainty regarding the hardware employed, resulting in a hurdle to export control enforcement. This represents a gap in U.S. capability. For example, we could reproduce DeepSeek's accelerations by purchasing the H800 chips they claim access to, reproducing their CUDA kernels, and observing the claimed ~45x training speedup, but this would be time consuming and expensive. Simulation can provide insight, but not determinative conclusions.
- Rather than engaging in expensive reverse-engineering, it may be more productive to look for secondary environmental clues that better indicate whether a company is using export-controlled chips. Power consumption, materials, and staffing provide useful secondary indicators of evasion.
- The experience with DeepSeek indicates that "high fence, small yard" export control strategies can drive adversaries to innovate in directions that erode US advantage. The current export control regime, designed to limit the proliferation of weapons of mass destruction to weaker nations such as Iran and North Korea, struggles to contain advances in the world's second largest economy.

Did DeepSeek Use Distillation?

Given the incentives and widely available methods, it is likely that DeepSeek used distillation from a strong, preference-tuned model, such as those developed by OpenAI or Anthropic. Training high-quality LLMs without distillation (the process of training one LLM by using responses from an existing, powerful LLM) requires expensive preference tuning with realistic prompts and human annotators.

- It is possible to test for this hypothesis by distilling different models into new pre-trained models to see if they indirectly acquire either a strong pre-trained model or strong performance on general-purpose tasks. L ooking for unique traces of the distillation data in linguistic patterns from the subsequent models can enable analysts to define similarity metrics between original and distilled models.
- Detering distillation poses a technical challenge, but one that can be overcome. The more serious difficulty is that the acceptability and consequences of distillation are ambiguous. The United States could develop clear policies that weigh the costs of preventing distillation against the benefits of protecting US property.

Did DeepSeek Steal Intellectual Property?

We have no way to determine whether model weights were stolen, but the lack of secondary indicators makes it unlikely. Weight files are massive, and theft is likely to be noticed only by the victim.

- It is most likely impossible to know with certainty whether a company is using outputs from a proprietary LLM because of their inherently probabilistic nature. It is possible, however, to develop models that could provide flags and indicators. The US should invest in these capabilities.
- We have no straightforward way to test whether a company is using stolen LLM development stages. Moreover, many proprietary AI systems build on ideas derived from easily accessible academic research. The United States should consider revised policies toward open approaches to critical capabilities. Protecting US IP requires researching new methods to identify thefts, establishing clear costs for these thefts, and incentivizing industry to institute safeguards and communicate with the US government.