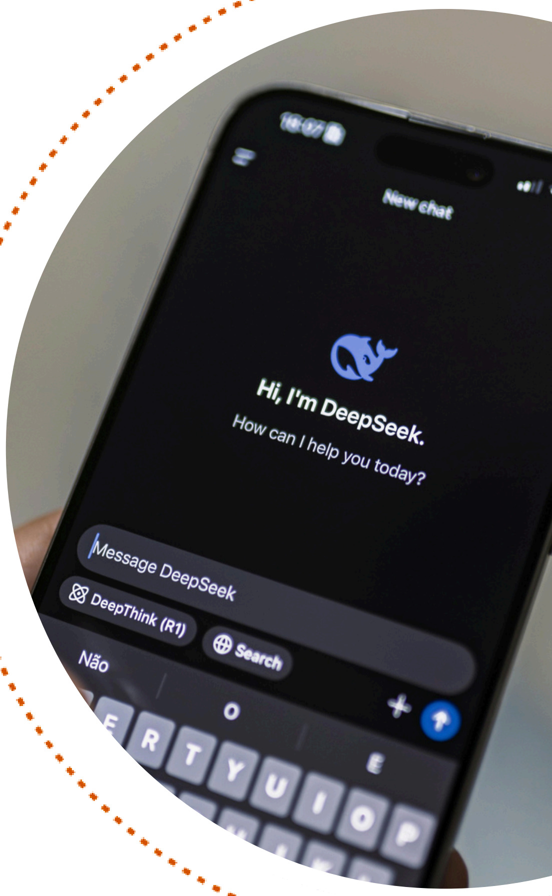
UVA | NATIONAL SECURITY DATA AND POLICY INSTITUTE

PRODUCT NO. 00006

DEEPSEEK'S ACHIEVEMENTS AND THE IMPLICATIONS FOR POLICY

FEBRUARY 21, 2025



NATIONALSECURITY.VIRGINIA.EDU | NSDPI@VIRGINIA.EDU

DeepSeek's Achievements and the Implications for Policy

February 21, 2025

Executive Summary

DeepSeek AI ("DeepSeek") caused a stir in the AI community and roiled the markets by training its large language model (LLM), DeepSeek-R1, with fewer resources and in less time than major competitors such as OpenAI's GPT-o1. This surprising accomplishment was not the product of a single breakthrough. Instead, the story behind DeepSeek's success blends strategic hardware procurement, in-house compute expertise, and incremental software and training innovations that combined to achieve unexpected efficiency. Implications for policy include:

- **Export Control Policy**–DeepSeek leveraged its hardware efficiently, but it still required highly sophisticated chips. DeepSeek leveraged gaps and lags in export control policy to obtain the required hardware. Policymakers should consider alterations to export controls that address these vulnerabilities.
- **Innovation Policy**–External constraints, some imposed by the United States, incentivized innovation in domains that better-resourced market players neglected. Policymakers should consider policies that encourage multiple paths to innovation (and discourage herd mentalities) among domestic developers, startups, and technical thought-leaders. Doing so will, as Vice President Vance put it in a recent speech at the Artificial Intelligence Action Summit in Paris, "ensure that American AI technology continues to be the gold standard worldwide and we are the partner of choice for others."
- **Industry Trajectory**–DeepSeek's efficiency breakthrough raises questions about the future balance between optimization and brute-force scaling. If efficiency outpaces hardware-driven gains, AI development may democratize, fostering broader participation. If scaling remains paramount, market concentration will persist, reinforcing international competition. Policymakers should monitor how these dynamics shape industry structure and strategic advantage.

Hardware

DeepSeek did, in fact, place special emphasis on hardware acquisitions. Assertions in the popular press that DeepSeek's advances were achieved with subpar or limited hardware are overblown. Well before its foray into the large language model (LLM) realm in April 2023, DeepSeek's parent company, High-Flyer, LP, was operating as a well-financed hedge fund specializing in algorithmic trading that relied heavily on cutting-edge graphics processing unit (GPU) clusters. In 2020, well before U.S. export restrictions, High-Flyer purchased 10,000 NVIDIA A100 40GB GPUs to build Fire-Flyer II, thereby replacing its older Fire-Flyer I cluster. This upgrade mirrored the hardware investments of U.S.-based quantitative trading giants like Jane Street and Renaissance Technologies.

In September 2022, the United States introduced export controls restricting the sale of A100s and H100s to China. In response, High-Flyer acquired an unknown number of NVIDIA H800 80GB (an only slightly cut-down variant of the H100, specifically designed to skirt the 2022 restrictions) before the loophole was closed in October 2023.[1]

Compute Expertise

Owning large quantities of GPUs alone does not ensure success; orchestration and high-performance computing (HPC) design are equally important. DeepSeek distinguished itself by building its own high-performance

[1] There are unsubstantiated reports of 50,000 hopper GPU purchases (Mok, 2025).

clusters rather than relying on standard cloud offerings. In previous work, notably DeepSeek-V3, the company's emphasis on GPU infrastructure became evident. The most prominent example is Fire-Flyer 2, a cluster containing the previously-acquired 10,000 NVIDIA A100 chips, which were then interconnected via a two-layer Fat-Tree network. According to the SC24 paper on Fire-Flyer AI-HPC, this system achieves nearly 80% of DGX-class performance (NVIDIA's high-performance, fully managed AI platform) for 60% of the cost and 40% less power consumption (An et al., 2024). The same efficiencies were likely applied to the H800-based clusters that were later used to train DeepSeek-R1.

The published documents only provide explicit detail for Fire-Flyer 2 (equipped with A100s) and do not mention a brand-new "Fire-Flyer III" or a separate cluster focused entirely on H800 GPUs. While it is conceivable that DeepSeek expanded with a new H800-based system, this has not been confirmed. Instead, references to H800 usage appear to reflect an extension of existing HPC resources.

DeepSeek has released two significant LLMs, each trained under similar pipelines: DeepSeek-V3 (Liu et al., 2024), a large mixture-of-experts (MoE) model running primarily on A100 hardware, and DeepSeek-R1 (Guo et al., 2025), with 671B parameters (the FP8 version at 642B parameters and about 642 GB on disk). Because FP8 requires compute capability 8.9 or higher (e.g., H800, H100), R1 needs BF16 or FP32 if run on A100 hardware, leading to increased memory usage. While reports suggest that DeepSeek may have used 2,048 NVIDIA H800 GPUs in at least one training run for DeepSeek-R1, the specific compute used across all experiments remains unverified.

Software Engineering

DeepSeek's remarkable training speed arose from more than just large-scale GPU deployments. The company also excelled at fine-grained engineering of data formats, GPU kernels, and communication workflows.

One of the principal breakthroughs for DeepSeek-R1 involved FP8 mixed precision on H800 GPUs. By moving from FP16/BF16 to FP8, DeepSeek accelerated matrix multiplication within H800 tensor cores, reduced memory usage by storing activations and gradients in FP8, and preserved numerical stability through carefully orchestrated scaling and partial sums. Another core innovation lay in PTX-level GPU kernel customization, where engineers wrote specialized allreduce and scheduling kernels optimized for H800 hardware. This approach gave them precise control over memory layout and register usage, opening the door to overlapping CPU-based data reduction with the GPU's forward and backward passes. By avoiding the overhead often associated with generic library calls, they managed to operate at or near maximum FLOPS across the cluster.

The HPC stack relied heavily on two key software components—HFReduce and HaiScale—to coordinate data movement and parallel workloads. HFReduce, combining CPU and GPU in a "hybrid allreduce," effectively hid communication behind computation. Meanwhile, HaiScale organized different modes of parallelism—data, pipeline, and tensor—in a way that benefitted from FP8's throughput gains and PTX-level overlaps.

Minimizing Bottlenecks

Efficiently managing thousands of GPUs requires a sophisticated interplay between hardware and software. Fire-Flyer 2's two-layer Fat-Tree network provided short communication paths for distributed training, allowing large throughput without choking the cluster. Scheduling was similarly adaptive, with CPU-based HFReduce merging partial sums to minimize GPU idle time.

Pipeline parallelism was another essential mechanism. Instead of processing the entire forward pass for one batch and then moving on to the backward pass, DeepSeek's system broke operations into chunks or micro-batches and pipelined them across different stages (or slices) of the model. This meant that while one batch's forward pass was running on earlier layers, another batch's backward pass might already be happening on later layers. By dovetailing pipeline stages with allreduce or HFReduce operations—sometimes after partial sums were computed—DeepSeek minimized idle GPU time. Such methods are well-documented in HPC research (e.g., "1F1B" or "dual-pipe" scheduling), and DeepSeek's success in applying them to large-scale LLM training demonstrates the power of HPC-level pipeline algorithms in the AI sphere.

Efficient Algorithms

DeepSeek's software foundation dovetailed with novel training algorithms to capitalize on every bit of GPU horsepower. Building on the DeepSeek-V3 base, the company introduced large-scale reinforcement learning to produce DeepSeek-R1. The initial phase, often called R1-Zero, omitted any supervised warm-up, allowing reinforcement learning (RL) to uncover chain-of-thought (CoT) behaviors in math, coding, and logic tasks. A small curated "long-CoT" dataset then stabilized training, ensuring correct and readable outputs. DeepSeek distilled the fully RL-tuned model to smaller variants, typically in the 7B–32B range, saving on inference costs and enabling broader applicability.

These multiphase and multisize strategies not only enriched the reasoning capacity of the final model but also shortened the total convergence time. By letting RL discover critical patterns directly on top of the HPC-optimized base model, DeepSeek avoided lengthy supervised data curation and overcame many of the usual training bottlenecks.

DeepSeek-R1-Zero improves its reinforcement learning algorithm through Group Relative Policy Optimization (GRPO) (Shao et al., 2024), proposing a rule-based reward system that has two parts: accuracy rewards and format rewards. Unlike standard methods that rely on computationally expensive value function training, GRPO uses group-based advantage estimation via outcome supervision RL. The model generates multiple outputs per prompt, assigns each a reward score, and normalizes these scores. Improvement is then computed relative to the group mean. GRPO constrains policy deviation using a per-token KL divergence penalty to mitigate reward model over-optimization. Instead of applying this penalty to the reward, as in typical LLM RL, GRPO integrates it directly into the loss function, improving stability, avoiding the costlier neural reward model, and minimizing reward hacking.[2]

In DeepSeek-V3, Mixture-of-Experts (MoE) layers split the model into specialized "experts" that can be selectively activated for different tokens or sub-tasks. A gating network dictates which expert to route each token to, enabling the model to scale total parameters without proportionally increasing the compute for each pass. This approach is especially helpful in large-scale mathematical reasoning and domain-specific tasks, as specific experts can learn unique patterns (e.g., algebraic manipulations and code interpretation) while the rest remain dormant. As a result, MoE allows DeepSeek-V3 to handle more nuanced inputs efficiently, contributing to shorter convergence times and lower overall GPU load.

For large deployments, such as DeepSeek-V3's training on thousands of A100 GPUs, MoE also makes it easier to balance workload among machines. The system reduces communication overhead by assigning subsets of experts to distinct GPU groups and avoids saturating a single compute node with all parameters at once. This design choice aligns well with DeepSeek's HPC focus, leveraging the Fat-Tree interconnect for distributed gating logic and partial sums across multiple experts.

Policy Implications

DeepSeek's ability to train DeepSeek-R1 so quickly and cost-effectively exemplifies the value of holistic optimization, weaving together hardware availability, HPC design, precision-level customization, and advanced reinforcement learning. The company's early and large GPU purchases—first A100s, then H800s—granted it massive computational potential. Fire-Flyer's HPC design, complete with specialized networking and an intricate multi-parallel scheduling framework, unlocked that potential at relatively low cost and power usage. Meanwhile, the shift to FP8 in building the H800 cluster, combined with PTX-level custom kernels, squeezed the maximum FLOPS from each GPU, even at unprecedented scales. Layering on a refined reinforcement learning approach—supported by pipeline parallelism and hybrid communication strategies—led to additional efficiencies.

DeepSeek's story demonstrates that effectively integrating HPC research, GPU kernel engineering, and RL-based algorithms can cut large-model training durations and costs to levels once thought unattainable. By pursuing these intertwined advances, the company forged a path for how future AI labs might achieve similarly remarkable training speeds without relying solely on brute-force hardware acquisitions.

DeepSeek's ability to achieve significant efficiency gains in AI development highlights the strengths and weaknesses of current export control measures. While it optimized its use of available hardware, it still required access to highly sophisticated chips—access made possible by export control policies that do not consistently adapt at the pace of innovation and industry action. While restrictions on the sale of advanced

[2] Reward hacking occurs when the model learns to find low-quality outputs that trick the reward model into producing high scores.

semiconductors can slow competitors' progress, they do not necessarily prevent them from acquiring or working around key technological inputs. Policymakers should reassess the scope, timing, and enforcement of export controls to close vulnerabilities without unduly disrupting global supply chains or incentivizing adversaries to develop alternative pathways that could ultimately weaken U.S. strategic leverage. A more dynamic, anticipatory approach to export controls will be crucial in shaping the competitive landscape.

The software engineering innovations that went into the DeepSeek product highlight the need to model advanced technologies as a combination of physical hardware and algorithmic interaction with the hardware. By combining accessible hardware with innovations in software engineering, the DeepSeek team was able to converge the boundaries of restricted and unrestricted physical technologies. This illustrates the hybrid nature of advanced technology: limitations in one domain (physical) can sometimes be overcome by innovations in another (algorithmic and engineering). Analytically decoupling the two domains is important because algorithmic/engineering approaches arise frequently and have high mobility, which can fill performance gaps created by limiting access to physical technology. Paying attention to the two domains of innovation and modeling them as complementary processes can widen the toolbox to predict which gaps in the technological landscape are vulnerable.

DeepSeek's advancements underscore the role of external constraints—some imposed by U.S. policy—in driving alternative approaches to AI development. These constraints incentivized innovation in areas overlooked by better-resourced market leaders, demonstrating that scarcity can sometimes spur technological breakthroughs. However, a reliance on artificial constraints to guide innovation is neither sustainable nor strategically desirable. Instead, policymakers should seek to foster an environment that encourages diverse and independent paths to innovation. This includes promoting investment in alternative architectures, supporting a wider range of AI research agendas, and discouraging herd mentalities that cluster resources around dominant paradigms. A well-calibrated innovation policy should ensure that domestic industry players pursue multiple technological frontiers, reducing systemic vulnerabilities and enhancing the resilience of the broader AI ecosystem.

The implications of DeepSeek's efficiency gains, when considered alongside the massive compute scaling strategies of major U.S. players, remain uncertain. It is unclear whether DeepSeek has uncovered a fundamental shift that challenges the prevailing hardware-intensive paradigm or if its innovations could be integrated with large-scale compute to push AI capabilities even further. This uncertainty has significant policy ramifications, as it influences the trajectory of the industry. If efficiency gains surpass the benefits of sheer hardware accumulation, AI development could become more accessible, fostering a diverse ecosystem of specialized models and decentralized innovation. Conversely, if large-scale hardware acquisition remains the dominant approach, the industry will continue to exhibit monopolistic tendencies, where only a handful of well-resourced actors can compete, reinforcing international rivalries and strategic competition. Understanding which trajectory prevails is crucial for policymakers navigating the future of AI governance and competition.

References

An, W., Bi, X., Chen, G., Chen, S., Deng, C., Ding, H., & Dong, K. (2024). Fire-flyer ai-hpc: A cost-effective software-hardware co-design for deep learning. *International Conference for High Performance Computing, Networking, Storage, and Analysis.*

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948.*

Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al. (2024). Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437.*

Mok, C. (2025). Taking stock of the deepseek shock. https://cyber.fsi.stanford.edu/gdpi/publication/taking-stock-deepseek-shock

Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al. (2024). Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300.*