THE HOOVER INSTITUTION TPA WHITE PAPER SERIES

TECHNOLOGY POLICY ACCELERATOR
HOOVER INSTITUTION

A Deep Peek into DeepSeek AI's Talent and Implications for US Innovation

By Amy Zegart and Emerson Johnston

APRIL 21, 2025

A collaboration with
The Stanford Institute for Human-Centered AI

HOOVER INSTITUTION
Stanford University
Human-Centered
Artificial Intelligence

A Deep Peek into DeepSeek AI's Talent and Implications for US Innovation

By Amy Zegart and Emerson Johnston

Chinese startup DeepSeek AI has upended conventional wisdom about artificial intelligence (AI) innovation. Released in January 2025, the company's R1 language model and V3 general-purpose large language model (LLM) sent tremors through markets and challenged assumptions about American technological superiority in frontier AI development.[1] Although DeepSeek AI's claims that its V3 model was trained for just $6 million have been widely disputed (experts estimate the true compute costs are closer to half a billion dollars, and DeepSeek AI itself says the cost was just for the final training run), the R1 model built on top of V3 demonstrated unprecedented reasoning capabilities and technical achievements that surpassed previous benchmarks set by US companies.

Beneath DeepSeek's technical achievements lies a more consequential story: the shifting patterns of global AI talent that made the company's breakthroughs possible. This paper examines the educational backgrounds, career paths, and international mobility of more than 200 researchers who authored DeepSeek's five foundational papers from January 2024 to February 2025. These five papers constitute the corpus of the company's openly available research papers since its founding in 2023.

We find striking evidence that China has developed a robust pipeline of homegrown talent. Nearly all of the researchers behind DeepSeek's five papers were educated or trained in China. More than half of them *never* left China for schooling or work, demonstrating the country's growing capacity to develop world-class AI talent through an entirely domestic pipeline. And while nearly a quarter of DeepSeek researchers gained some experience at US institutions during their careers, most returned to China, creating a one-way knowledge transfer that benefits China's AI ecosystem.

These talent patterns represent a fundamental challenge to US technological leadership that export controls and computing investments alone cannot address. DeepSeek is an early-warning indicator about the essential role that human capital—not just hardware or algorithms—plays in geopolitics, and how America's talent advantage is eroding.

Methodology

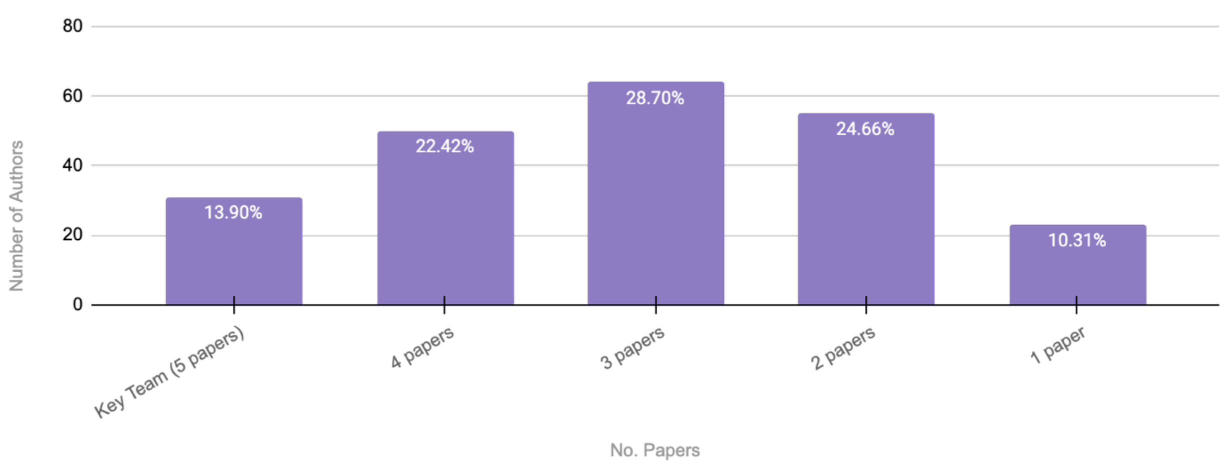
DeepSeek AI, a Chinese AI research company focused on "cost-efficient, high-performance language models," released five papers on Cornell University's arXiv.org manuscript archive between 2024 and 2025.[2] A total of 223 authors were credited across the five papers. We were able to conduct a comprehensive review of 211 of them.[3] Using data from the OpenAlex research catalog collected in February 2025, we collected detailed author profiles (publication records, citation metrics, institutional affiliations dating back to 1989) and comprehensive institutional data (geographical location, organization type, research output metrics), paying special attention to tracking changes over time.[4] Then, through custom Python scripts for data collection and analysis, we mapped each researcher's complete institutional history, revealing previously undetected patterns of cross-border movement. While traditional analyses often rely on static snapshots of talent at a particular point in time, our approach allowed us to quantify not just where talent is today, but how it has flowed between countries over time—particularly between China and the United States—capturing the "reverse brain drain" cases that represent strategic knowledge-transfer mechanisms.

DeepSeek's Talent Infrastructure Across Five Papers

A total of 223 people were listed as contributors to any of DeepSeek's five papers (see fig. 1). Our analysis finds that 31 researchers (or just under 14 percent of the total author pool) contributed to all five papers—what we refer to as the "Key Team."[5] Another 50 authors worked on four papers, 64 contributed to three papers, 55 were listed on two papers, and 22 researchers contributed to just one paper.

Figure 1 – Authorship Distribution Across DeepSeek Papers

Distribution of 222 authors across the five DeepSeek AI papers, with 31 authors (14%) contributing to all five papers



Source: All data from OpenAlex.

As table 1 illustrates (see appendix B), it appears that DeepSeek used a shifting categorization of talent across the five papers. In Paper 1 (DeepSeek LLM), the reported contributor labels were organizational rather than role-based, with 53 individuals categorized into Business Team (8), Compliance Team (7), Data Annotation Team (36), and Design Team (2). Notably, none of these labeled contributors were credited as authors on the paper itself, which officially listed 86 authors. However, 40 of those contributors were later credited as authors in at least one subsequent DeepSeek paper—which is why they are captured here. This discrepancy may suggest that Paper 1's contributor list reflected a broader pool of internal collaborators—many of whom were not formally recognized at the time but went on to receive authorship credit as the project evolved.

Papers 2 and 4 appear to have transitioned to more functionally descriptive categories that closely resembled internal team structures. Paper 2 introduced hybrid contributor tags such as "Business & Compliance," "Data Annotation," and "Research & Engineering." Among the 156 total contributors, the vast majority (105) were classified under Research & Engineering, followed by 31 in Data Annotation and 18 in Business & Compliance. Notably, 2 contributors—Shengfeng Ye and Yanhong Xu—were listed in more than one category: Ye appeared in both Research & Engineering and Business & Compliance, while Xu was credited under Research & Engineering and Data Annotation, likely reflecting overlapping responsibilities within the organization. Paper 4 (DeepSeek V3), which had 197 listed authors, followed the same categorization structure: Research & Engineering (148), Data Annotation (30), and Business & Compliance (17). Again, Ye and Xu were the only contributors assigned to two categories—Ye in Research & Engineering and Business & Compliance; Xu in Research & Engineering and Data Annotation. This schema appears to capture the backbone of DeepSeek's technical efforts. Within this structure, nearly all members of the 31-person Key Team were designated as Research & Engineering contributors, with one notable exception—Yanhong Xu, who, as noted, also held a Data Annotation role in Papers 2 and 4.

Papers 3 and 5 introduced a different delineation between levels of contribution through a binary categorization: Contributor and Core Contributor. This shift may indicate a formal recognition of hierarchical status within the research group. In Paper 3, just 4 of the 39 contributors were labeled as Core Contributors. Similarly, Paper 5—the company's internationally watched R1 reasoning model—designated 18 Core Contributors out of 194 total. In both cases, Core Contributors made up roughly 10 percent of the total contributor base, suggesting a carefully curated leadership tier. Notably, all four Core Contributors from Paper 3—Daya Guo, Dejian Yang, Qihao Zhu, and Zhihong Shao—were also credited as Core Contributors in Paper 5, and all four contributed to every one of the five DeepSeek papers, likely signaling their central, long-term influence on the DeepSeek project.

DeepSeek Researcher Citation Metrics: Not So Green After All

The prevailing narrative has been that DeepSeek succeeded with younger, less experienced researchers. Citation metrics, however, suggest that DeepSeek's talent was not so green after all.

While the structure of DeepSeek's collaboration shows clear differentiation in participation levels, there is also meaningful variation in scholarly experience across those tiers. Among the set of 211 contributors for whom we were able to pull data, the average researcher has published sixty-one works and received just over one thousand citations, with an h-index of 10.8 and i10-index of just over 19.[6] It is worth noting that these averages mask a bimodal distribution: Many researchers have modest academic footprints, but a concentrated group ranks far higher in output and impact. The median citation count (249), h-index (7), and i10-index (5) for this group underscore this internal variation.

Notably, the 31 Key Team researchers who contributed to all five papers stand out sharply. The group averages 1,554 citations per author, with a median of 501, and a mean h-index of 13.5 and i10-index of 25.5. Median values—an h-index of 10 and an i10-index of 11—further indicate consistent impact across the Key Team, not just a few outliers. These metrics provide additional evidence that the DeepSeek Key Team consists of researchers with already credible academic track records.

This academic strength becomes even more apparent when compared to a peer group from one of the world's leading AI labs. According to data from the OpenAI o1 system card (arXiv:2412.16720), the team of 265 authors listed on that release had an average citation count of 4,403 but a median of just 338, indicating a steep drop-off beyond a few highly cited individuals. Further, the group's median h-index was only 6 and the i10-index 4, reflecting more limited consistency of impact across the full group.

In contrast, both DeepSeek's full author pool and its Key Team exhibit greater balance between average and median performance—suggesting not only strength at the top, but also less variation across contributors compared to the OpenAI team. These patterns may indicate a more evenly distributed base of academic experience, rather than one overly reliant on a handful of standout figures. DeepSeek's research engine appears not only deep but wide—an organizational trait that may prove especially important as competition in foundation model development intensifies.

Taken together, these comparisons challenge the media narrative that DeepSeek's rapid ascent was driven by "untested" or inexperienced researchers. While OpenAI continues to receive global recognition, many of DeepSeek's central contributors—at least by traditional bibliometric standards—were better published, more consistently cited, and arguably more academically established at the time of their breakthrough.[7]

A Longitudinal View of Institutional Affiliations: China's Dominant Position

Looking longitudinally at the 201 DeepSeek authors with known affiliation data, we find that more than half (n=111) have been trained and affiliated *exclusively* at Chinese institutions throughout their careers—evidence of China's growing capacity to develop world-class AI talent domestically without relying on Western expertise. And the vast majority of DeepSeek authors—98 percent (n=197)—have held at least one past or current affiliation with a Chinese institution.

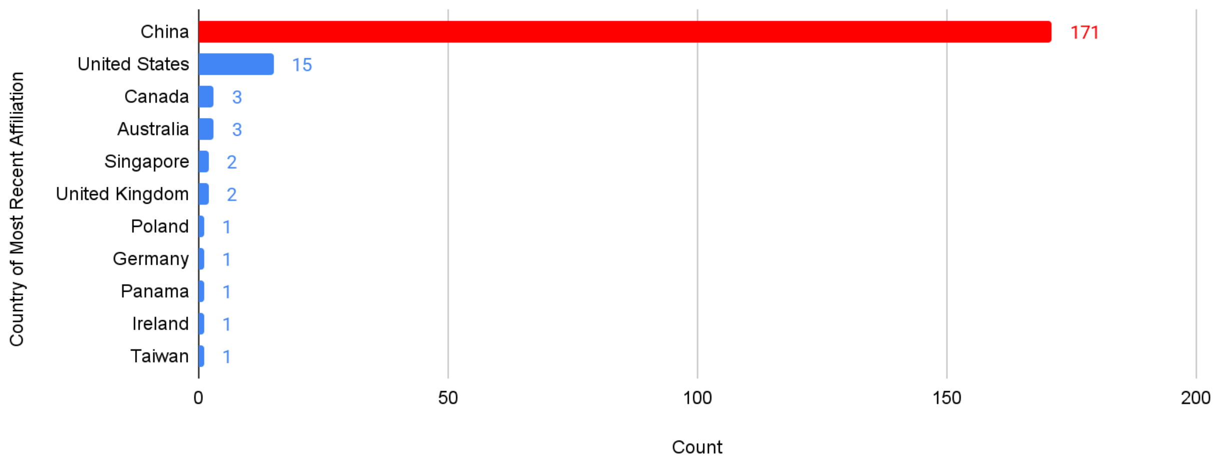
Four authors appear to have not studied, trained, or worked in China at all. Their academic and professional roots spanned a range of global institutions: Erhang Li was trained in the United Kingdom and the United States and is affiliated with Intel UK; Y. Q. Wang studied in Germany at Johannes Gutenberg University Mainz; Yuduan Wang received education in Singapore at the National University of Singapore; and Panpan Huang studied in the United States at Purdue University. While these individuals represent exceptions within the broader DeepSeek ecosystem, they highlight the international reach of the global AI research community. Still, their small number underscores how uncommon this path is among DeepSeek contributors—further reinforcing the observation that China's domestic pipeline is now capable of producing world-class AI researchers largely on its own.

We found that only a quarter of DeepSeek researchers (24.3 percent, n=49) have ever held an academic or professional affiliation with a US institution—further illustrating the limited role American institutions have played in shaping this cohort.

China Dominates the 2025 Snapshot of Institutional Affiliations, Too

As figure 2 shows, 171 of the 201 DeepSeek authors with known affiliation data were affiliated with Chinese institutions in 2025 (the most current year available).[8]

Figure 2 – Geographic Distribution of Current Institutional Affiliations
Current geographic distribution of 201 DeekSeek AI researchers with known affiliations



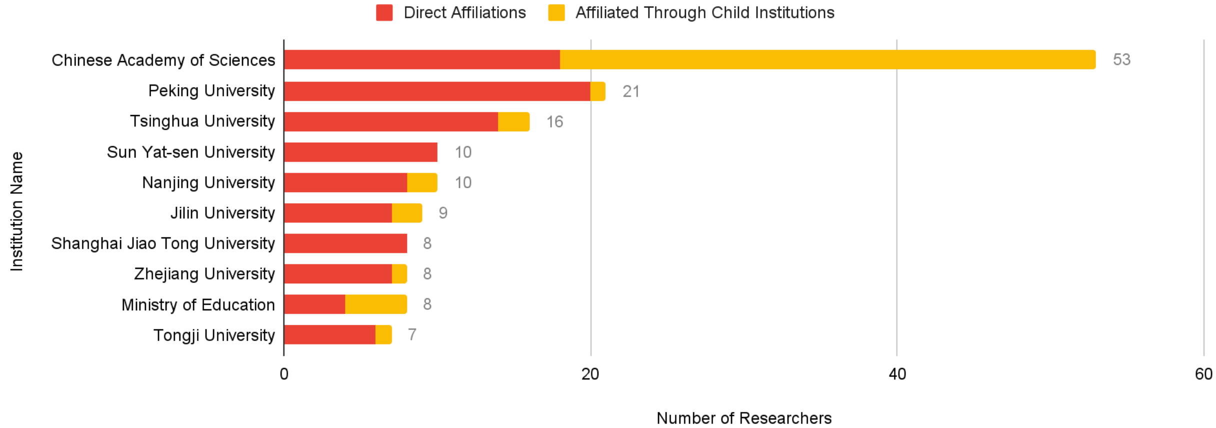
Source: All data from OpenAlex.

Just 7 percent (n=15) of researchers currently hold US-based affiliations. These include positions at prominent research universities (such as Stony Brook University, University of North Texas, and the University of California, San Francisco), medical institutions (such as Boston Children's Hospital), and tech or biotech companies including Google, Otsuka, and Health First. The remaining researchers are spread across a small set of other countries, including Australia, Canada, the United Kingdom, and Singapore, with single cases in Germany, Ireland, Panama, Poland, and Taiwan. This geographic consolidation around China further reinforces the central role of its domestic institutions—not just as training grounds, but as long-term professional destinations for AI talent.

The Central Role of the Chinese Academy of Sciences

The broader institutional landscape supporting DeepSeek's development reflects the full career trajectories of its researchers, encompassing all known affiliations across time. In total, the 211 analyzed authors were connected to 499 unique institutions globally, with Chinese institutions accounting for 368 (74 percent) of them (see fig. 3). The network is predominantly anchored in academia, with universities and research institutions forming the backbone, but it also features some training from private companies (n=17), government institutions (n=12), and nonprofit organizations (n=9).

Figure 3 – Top 10 Institutions by Researcher Count



Source: All data from OpenAlex.

Within this institutional landscape, the Chinese Academy of Sciences (CAS) emerges as the strategic center of gravity. While directly hosting only 18 authors, CAS encompasses a total of 53 researchers when accounting for its network of 153 affiliated institutions. This extensive institutional reach—where a "child institution" refers to an organization with a subsidiary relationship (as defined by OpenAlex) to CAS as its parent organization, including research institutes, laboratories, and specialized centers—combined with remarkable research impact metrics (over 840,000 works and 23.7 million citations), positions CAS as the dominant player in this ecosystem.[9] Peking University comes in second with 21 total affiliations, but it leads in direct affiliations with 20 researchers. Tsinghua University follows with 16 authors, then Sun Yat-sen University and Nanjing University with 10 authors each. This distribution reveals how China has leveraged its institutional infrastructure to support AI development, with a network centered around CAS but distributed across multiple prestigious universities. The concentration of talent within this network of Chinese institutions has created a fertile environment for AI innovation that challenges the US advantage in institutional resources.

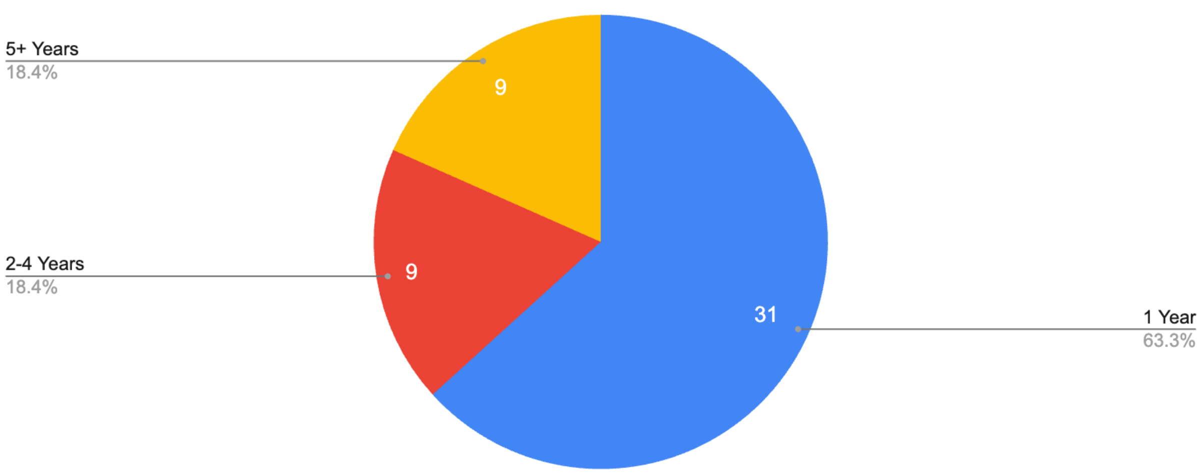
The US-China Talent Pipeline: Challenging American Assumptions

Of the 49 DeepSeek researchers who had US affiliations at some point during their careers, 63.3 percent (n=31) spent just one year in the United States—long enough to gain exposure to top-tier research environments, but not long enough to establish enduring ties. Another 18.4 percent (n=9) remained for two to four years, and 18.4 percent (n=9) stayed five years or longer, often across multiple institutions. This latter group includes some of the most influential researchers in the cohort, such as Minghua Zhang, who accumulated affiliations at State University of New York and Stony Brook

University spanning over a decade; Zhenda Xie, who spent eight years across UCLA and Optica; and B. Zhang, whose recurring ties with the University of Southern California from 2007 to 2022 preceded his return to Peking University. These 9 long-stay researchers are not statistical outliers—they averaged 4,541 citations, held a median h-index of 25, and had a median i10-index of 40. Despite this deep academic integration, only 3 of the 9 currently remain affiliated with US institutions, further underscoring how the US research ecosystem served as a powerful incubator of talent that ultimately advanced China's AI leadership (see fig. 4).

Figure 4 – US Experience Duration

US Experience Duration for the 49 DeepSeek Researchers with US Affiliations



Source: All data from OpenAlex.

Notably, the institutional diversity of US experience among DeepSeek researchers is significant. The 49 individuals with US affiliations were connected to 65 different institutions across 26 states, including public universities, private colleges, medical centers, nonprofit organizations, and technology companies. While no single institution accounted for more than three researchers, several—including the University of Southern California, Stanford University, and New York University—had multiple affiliations. This distribution spans the full geographic breadth of the United States, with clusters visible in in key innovation hubs: the Bay Area and Southern California, the Boston-to-DC corridor, and research-heavy regions of Texas and the Midwest (see fig. 5). Importantly, rather than concentrating within a small number of elite campuses, these researchers engaged with a wide cross-section of the American research ecosystem. This breadth may have facilitated broader exposure to US scientific and technological practices. It also meant that no single institution had good visibility into the scale of the international AI knowledge exchange taking place.

Figure 5 – Geographic Distribution of US Institutions Affiliated with DeepSeek Researchers

Source: All data from OpenAlex.

More telling than location or duration is direction. Among the 49 DeepSeek researchers with US affiliations, only a small share followed the linear trajectory of moving from China to the United States and remaining in the US (see figs. 6a and 6b). Instead, our data shows that the dominant mobility pattern is cyclical, multinational, and strategically adaptive. As shown in figure 6b, almost 40 percent (n=19) of these researchers began their careers in China, traveled abroad—including to the United States—and ultimately returned to China. Researchers such as Xuan Lu, Xiaodong Liu, and Shiyu Wang exemplify this classic “study-abroad-and-return” model: China → USA → China. Their careers reflect a traditional, state-aligned mobility model where US training is used to strengthen domestic capabilities.

Figure 6a – US Retention Rate

US Retention Data for the 49 DeepSeek Researchers with US Affiliations

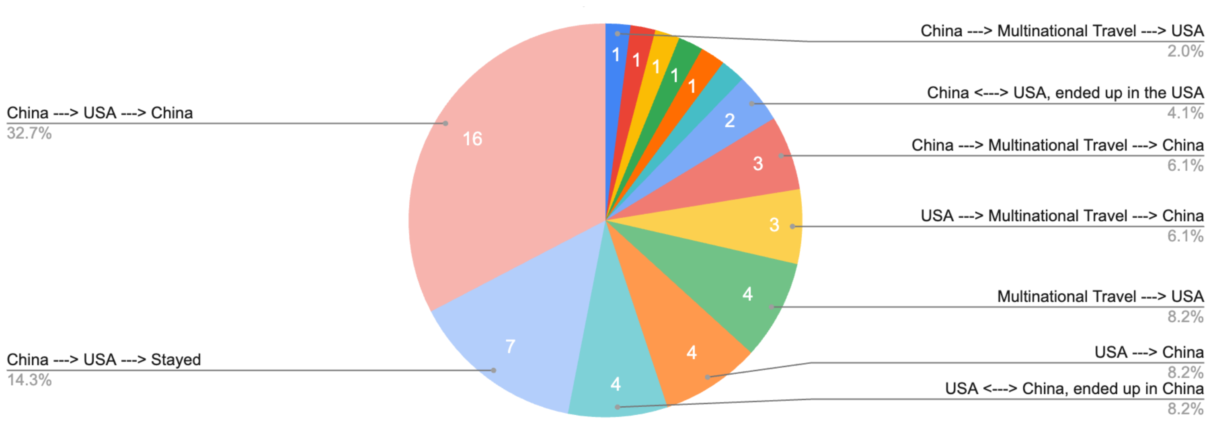
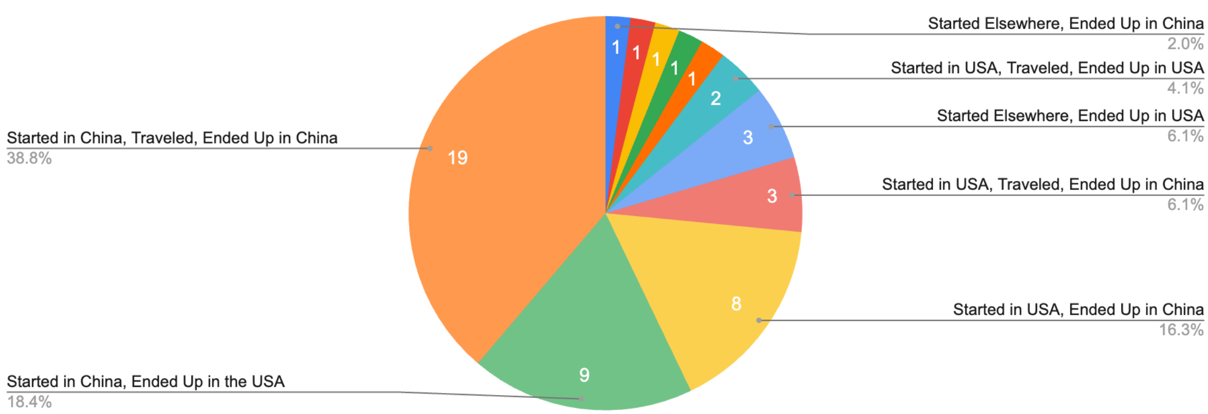


Figure 6b – US Retention Rate (simplified)

US Retention Rate for the 49 DeepSeek Researchers with US Affiliations



Source: All data from OpenAlex.

A second, more complex group includes researchers such as Wenfeng Liang, Minghua Zhang, and Zhiyu Wu, whose careers span multiple transits between China and the United States (e.g., China → USA → China → USA → China). These researchers don't simply return—they circulate, developing global networks and embedding themselves in both ecosystems. This pattern of bidirectional exchange accounts for 12.2 percent (n=6). Of these, 4 currently list US institutions as their most recent affiliation, while 2 are affiliated with Chinese institutions. While it is difficult to determine intent or long-term plans from affiliation data alone, these cases illustrate how cross-border mobility can strengthen China's AI ecosystem without necessarily requiring permanent US retention.

Other researchers such as Daya Guo, Guanting Chen, and Yicheng Wu take even more global paths—passing through institutions in the United Kingdom, Singapore, Saudi Arabia, Taiwan, or Australia. These trajectories (e.g., Taiwan → China → Australia → USA) illustrate the rising influence of multinational knowledge acquisition, with the United States serving as just one of many strategic destinations in a broader global loop.

Notably, only 14.2 percent (n=7) of the DeepSeek cohort (e.g., Ruiqi Ge, Peiyi Wang, Bingxuan Wang) followed a China → USA → Stayed path—remaining in the United States after initial training (see fig. 6a). While still significant, this group is no longer the default or dominant outcome. Even among researchers who began in the United States (e.g., B. Zhang, Ruoyu Zhang), many ultimately relocated to China, with 22.4 percent (n=11) falling into the "Started in the USA, Ended Up in China" or "Started in the USA, Traveled, Ended Up in China" categories.

Finally, a small but illustrative set of researchers defies simple classification. Figures such as Kuai Yu (USA → Netherlands → Singapore → China) or Zhen Zhang (USA → China → Hong Kong → USA) reflect the complexity of today's scientific mobility. These researchers—counted within "Started in USA, Traveled, Ended Up in China" (6.1 percent, n=3) or "Started in USA, Traveled, Ended Up in USA" (4.1 percent, n=2)—reveal how transnational scientific careers are increasingly nonlinear and dynamic.

Taken together, these patterns reveal important features of global AI talent flows. The United States remains a vital node in international research training—but it is not the fulcrum or the end point. Most of DeepSeek's researchers are not being trained in the United States, and those who are trained here are not retained. Instead, they are passing through. These findings suggest that American institutions are serving as steppingstones, equipping elite researchers with high-impact skills, connections, and credentials that are ultimately reinvested into China's AI ecosystem. Importantly, the 49 DeepSeek researchers with US affiliations at some point in their careers were among the most academically accomplished in the entire research cohort, averaging 2,168 citations (median 565), with a mean h-index of 17 and i10-index of 34—figures significantly higher than those for the broader DeepSeek author pool. These are not peripheral actors, but central contributors to one of China's most advanced AI efforts.

For US policymakers, our DeepSeek talent analysis suggests it is high time to reassess long-standing assumptions that the world's best and brightest naturally want to study and stay in the United States. Attracting and permanently retaining the world's best minds—once a cornerstone of American technological dominance—appears increasingly misaligned with twenty-first-century educational realities. DeepSeek is, at its core, a story of homegrown capacity: Half of its researchers have never left China, the overwhelming majority have deep institutional ties to China, and even many who trained in the United States ultimately returned to China—potentially advancing China's position in the global AI race.

A Closer Look at the Key Team and Where They Trained

Of 31 Key Team authors, 28 had available institutional affiliation data on OpenAlex. Half of them (n=14) have spent part of their careers at institutions outside of China. These globally mobile researchers often followed targeted international pathways: initial training at elite Chinese universities followed by graduate study, postdoctoral work, or research appointments abroad—typically in the United States, the United Kingdom, Australia, or other key AI hubs—before returning to China.

Notable examples include Daya Guo (China → UK → China → USA → UK → China), who spent time at both Rensselaer Polytechnic Institute and Microsoft Research (United Kingdom); Jiashi Li (China → Japan → China → USA → China), affiliated with the Hoshi University in Japan and later the University of California, Santa Barbara; and Dejian Yang (China → UK → Australia), who held positions at the Pharmaron (United Kingdom) and the University of Technology Sydney. Others, such as Zhenda Xie and Wenfeng Liang, show repeated, multidirectional mobility between the United States and China, suggesting enduring cross-border collaboration.

Of the internationally experienced Key Team members, 8 had US affiliations, including Peiyi Wang (Boston College), Qihao Zhu (Carnegie Mellon University), and Zhihong Shao (University of Michigan–Ann Arbor). Others had connections to institutions in Canada (Liyue Zhang), Singapore (Qihao Zhu), Bangladesh (Kai Dong), and South Korea (Junxiao Song). This distribution reflects a deliberate emphasis on experience in countries that are global leaders in AI research and higher education.

These patterns suggest a sophisticated approach to human capital development that treats international experience not as "brain drain" but as strategic national investment—sending promising researchers abroad to acquire cutting-edge knowledge and methodologies before returning to apply these assets to China's technological advancement.

Geopolitical Implications

The talent patterns revealed in our analysis have significant geopolitical implications. For centuries, the sources of national power have stemmed from tangible assets—such as territory that could be conquered, populations that could be taxed or conscripted, goods that could be embargoed, militaries that could be deployed. Those tangible sources of national power still matter, but in the technology age, power also derives from intangible assets such as data, technology, and knowledge inside people's heads. Knowledge power has never been more important for economic and geopolitical competition; it is the ultimate portable weapon.

These findings challenge a long-held belief that the United States will always attract the world's best talent. In reality, however, top global talent has options. DeepSeek's talent story suggests that the United States cannot assume a permanent talent lead. Instead, the nation needs to compete much more aggressively to attract, welcome, and retain the world's best and brightest while urgently growing domestic capabilities by improving K–12 STEM (science, technology, engineering, mathematics) education at home.

Ultimately, DeepSeek AI represents more than just another advance in language model technology. It reveals talent patterns that challenge long-held US assumptions about innovation advantage. Our analysis of DeepSeek's research network suggests that conventional wisdom about US dominance in talent development and retention may no longer hold true, with significant implications for future technological competition.

ABOUT THE AUTHORS

Amy Zegart is the Morris Arnold and Nona Jean Cox Senior Fellow at the Hoover Institution, director of Hoover's Technology Policy Accelerator, and a senior fellow and associate director at Stanford University's Institute for Human-Centered AI.

Emerson Johnston is a second-year master's student in international policy and a Knight-Hennessy Scholar at Stanford University.

Appendix A: DeepSeek Research Papers (2024–2025)

The following five papers released by DeepSeek AI between January 2024 and February 2025 formed the basis for our institutional and author-trajectory analysis:

1. **DeepSeek LLM: Scaling Open-Source Language Models with Longtermism**
   arXiv:2401.02954 —– January 2024
   Focuses on scaling laws for open-source LLMs in 7B and 67B configurations, contributing insights into training efficiency.

   Full Abstract: The rapid development of open-source large language models (LLMs) has been truly remarkable. However, the scaling law described in previous literature presents varying conclusions, which casts a dark cloud over scaling of LLMs. We delve into the study of scaling laws and present our distinctive findings that facilitate scaling of large-scale models in two commonly used open-source configurations, 7B and 67B. Guided by the scaling laws, we introduce DeepSeek LLM, a project dedicated to advancing open-source language models with a long-term perspective. To support the pretraining phase, we have developed a dataset that currently consists of two trillion tokens and is continuously expanding. We further conduct supervised fine-tuning (SFT) and direct preference optimization (DPO) on DeepSeek LLM Base models, resulting in the creation of DeepSeek Chat models. Our evaluation results demonstrate that DeepSeek LLM 67B surpasses Llama-2 70B on various benchmarks, particularly in the domains of code, mathematics, and reasoning. Furthermore, open-ended evaluations reveal that DeepSeek LLM 67B Chat exhibits superior performance compared to GPT-3.5.

2. **DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model**
   arXiv:2405.04434 – May 2024
   Introduces a 236B parameter Mixture-of-Experts (MoE) model with a focus on cost-effective training and inference using novel architectural choices such as Multi-head Latent Attention (MLA).

   Full Abstract: We present DeepSeek-V2, a strong Mixture-of-Experts language model characterized by economical training and efficient inference. It comprises 236B total parameters, of which 21B are activated for each token, and supports a context length of 128K tokens. DeepSeek-V2 adopts innovative architectures including Multi-head Latent Attention and DeepSeekMoE. MLA guarantees efficient inference through significantly compressing the Key-Value (KV) cache into a latent vector, while DeepSeekMoE enables training strong models at an economical cost through sparse computation. Compared with DeepSeek 67B, DeepSeek-V2 achieves significantly stronger performance, and meanwhile saves 42.5 percent of training costs, reduces the KV cache by 93.3 percent, and boosts the maximum generation throughput to 5.76 times. We pretrain DeepSeek-V2 on a high-quality and multisource corpus consisting of 8.1T tokens, and further perform supervised fine-tuning and reinforcement learning (RL) to fully unlock its potential. Evaluation results show that, even with only 21B activated parameters, DeepSeek-V2 and its chat versions still achieve top-tier performance among open-source models.

3. **DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence**
   arXiv:2406.11931 – June 2024
   A code-specialized MoE model achieving performance comparable to GPT-4 Turbo, emphasizing large-scale continued pretraining.

   Full Abstract: We present DeepSeek-Coder-V2, an open-source Mixture-of-Experts code language model that achieves performance comparable to GPT4-Turbo in code-specific tasks. Specifically, DeepSeek-Coder-V2 is further pretrained from an intermediate checkpoint of DeepSeek-V2 with an additional six trillion tokens. Through this continued pretraining, DeepSeek-Coder-V2 substantially enhances the coding and mathematical reasoning capabilities of DeepSeek-V2, while maintaining comparable performance in general language tasks. Compared to DeepSeek-Coder-33B, DeepSeek-Coder-V2 demonstrates significant advancements in various aspects of code-related tasks as well as reasoning and general capabilities. Additionally, DeepSeek-Coder-V2 expands its support for programming languages from 86 to 338, while extending the context length from 16K to 128K. In standard benchmark evaluations, DeepSeek-Coder-V2 achieves superior performance compared to closed-source models such as GPT4-Turbo, Claude 3 Opus, and Gemini 1.5 Pro in coding and math benchmarks.

4. **DeepSeek-V3 Technical Report**
   arXiv:2412.19437 – December 2024
   Advances the DeepSeek MoE line with 671B total parameters and pioneering loss-free loading techniques to enhance inference efficiency.

   Full Abstract: We present DeepSeek-V3, a strong Mixture-of-Experts language model with 671B total parameters with 37B activated for each token. To achieve efficient inference and cost-effective training, DeepSeek-V3 adopts Multi-head Latent Attention and DeepSeekMoE architectures, which were thoroughly validated in DeepSeek-V2. Furthermore, DeepSeek-V3 pioneers an auxiliary-loss-free strategy for load balancing and sets a multitoken prediction training objective for stronger performance. We pretrain DeepSeek-V3 on 14.8 trillion diverse and high-quality tokens, followed by supervised fine-tuning and reinforcement learning stages to fully harness its capabilities. Comprehensive evaluations reveal that DeepSeek-V3 outperforms other open-source models and achieves performance comparable to leading closed-source models. Despite its excellent performance, DeepSeek-V3 requires only 2.788M H800 GPU hours for its full training. In addition, its training process is remarkably stable. Throughout the entire training process, we did not experience any irrecoverable loss spikes or perform any rollbacks. The model checkpoints are available at https://github.com/deepseek-ai/DeepSeek-V3.

5. **DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning**
arXiv:2501.12948 – January 2025
The flagship reasoning-focused model trained via large-scale RL without supervised fine-tuning. Widely seen as a breakthrough in emergent reasoning behaviors and the focal point of DeepSeek's impact.

Full Abstract: We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning without supervised fine-tuning as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multistage training and cold-start data before RL. DeepSeek-R1 achieves performance comparable to OpenAI o1-1217 on reasoning tasks. To support the research community, we open-source DeepSeek-R1-Zero, DeepSeek-R1, and six dense models (1.5B, 7B, 8B, 14B, 32B, 70B) distilled from DeepSeek-R1 based on Qwen and Llama.

Appendix B: Tables

Table 1: Author Count and Contributor Roles for DeepSeek Publications (2024–25)

Publication	Date	No. of Authors*	Contributor Categories (no. of contributors)
Paper 1: DeepSeek LLM	January 2024	86	Business Team (8), Compliance Team (7), Data Annotation Team (36), Design Team (2)**
Paper 2: DeepSeek V2	May 2024	156	Research & Engineering (105), Data Annotation (31), Business & Compliance (18), Mixed Roles (2: Data Annotation + R&E, Business & Compliance + R&E)
Paper 3: DeepSeek VCoder 2	June 2024	39	Core Contributor (4), Contributor (35)
Paper 4: DeepSeek V3	December 2024	197	Research & Engineering (148), Data Annotation (30), Business & Compliance (17), Mixed Roles (2: Data Annotation + R&E, Business & Compliance + R&E)
Paper 5: DeepSeek R1	January 2025	200†	Core Contributor (18), Contributor (176) – *This total number (194) reflects the no. of authors from the PDF.*

* This number reflects unique authors listed for each paper, consolidating names across both the ArXiv and PDF versions where applicable. Discrepancies between sources are noted:

- Paper 2 and Paper 4 PDFs each contain two duplicate names (Shengfeng Ye and Yanhong Xu), due to those individuals being listed in multiple contributor categories.
- Paper 5:
  - The PDF includes one duplicate name (Shengfeng Ye).
  - The ArXiv includes two duplicate names (Shengfeng Ye and Yanhong Xu).
  - Nine authors appear on only one version (either ArXiv or PDF), including:
    - On ArXiv but not on PDF: Chenyu Zhang, Han Bao, Haocheng Wang, Huajian Xin, Jiawei Wang
    - On PDF but not on ArXiv: Jinhao Tu, Kaichao You, Mingxu Zhou, Wanjia Zhao

** The contributor categories listed for Paper 1 reflect a separate contributor pool that was not credited as authors on that paper. The numbers shown in the chart represent the total number of individuals in each category at that time.

† Paper 5 showed discrepancies in authorship counts: The PDF version originally listed 195 authors, but one author (Shengfeng Ye) was listed twice, resulting in 194 unique names. The ArXiv entry listed 197 authors. When combining both lists and removing duplicates, the total came to 201 unique authors.

Table 2: List of Key Team Researchers

The list below includes the 31 individuals who are credited as authors on all five DeepSeek AI papers. An asterisk (*) indicates those identified as core contributors in the fifth paper.

1. Bingxuan Wang
2. Chenggang Zhao
3. Chengqi Deng
4. Chong Ruan
5. Damai Dai*
6. Daya Guo*
7. Dejian Yang
8. Deli Chen
9. Fuli Luo
10. Hanwei Xu
11. Huazuo Gao
12. Jiashi Li
13. Junxiao Song
14. Kai Dong
15. Kang Guan
16. Liyue Zhang
17. Peiyi Wang
18. Qihao Zhu
19. Qiushi Du
20. Shirong Ma
21. Wenfeng Liang
22. Wenjun Gao
23. Xiao Bi*
24. Xin Xie
25. Yanhong Xu
26. Yaohui Wang
27. Yishi Piao
28. Yuxiang You
29. Zhenda Xie
30. Zhewen Hao
31. Zhihong Shao

Table 3: Scholarly Output and Citation Metrics of DeepSeek and OpenAI Research Teams

		Works Count	Cited by Count	h-Index	i10-Index
All DeepSeek Authors	Average	61.057	1,059.218	10.791	19.166
	Median	24.000	249.000	7.000	5.000
DeepSeek Core Group	Average	70.806	1,554.258	13.548	25.548
	Median	51.000	501.000	10.000	11.000
DeepSeek US-Affiliated Authors	Average	101.286	2,200.286	17.122	34.265
	Median	57.000	565.000	12.000	14.000
OpenAI o1 Authors	Average	58.951	4,402.917	12.109	24.955
	Median	16.000	338.000	6.000	4.000

Table 4: US Institutions Affiliated with DeepSeek Researchers

The following list includes US-based academic, research, medical, and industry institutions where DeepSeek authors have held prior or current affiliations. This includes both educational and professional roles. Asterisk (*) indicates a current affiliation based on the most recent OpenAlex data.

No. of Authors	Organization	City	Country
3	University of Southern California*	Los Angeles	United States
2	Auburn University	Auburn	United States
2	New York University	New York	United States
2	Stanford University	Stanford	United States
2	University of California, Santa Barbara	Santa Barbara	United States
2	University of North Texas*	Denton	United States
1	National Clinical Research	Richmond	United States
1	University Research Co. (United States)	Bethesda	United States
1	Electric Power Research Institute	Palo Alto	United States
1	Johns Hopkins Medicine	Baltimore	United States
1	Stony Brook University*	Stony Brook	United States
1	Cornell University	Ithaca	United States
1	Zero to Three	Washington	United States
1	Mississippi State University	Starkville	United States
1	Creative Commons	Mountain View	United States
1	Otsuka (United States)*	Princeton	United States

1	University of Notre Dame	Notre Dame	United States
1	University of California, San Diego	San Diego	United States
1	State University of New York	Albany	United States
1	Northeastern University	Boston	United States
1	The University of Texas at Austin	Austin	United States
1	University of California–Los Angeles	Los Angeles	United States
1	Loyola University Medical Center	Maywood	United States
1	North Carolina State University	Raleigh	United States
1	Michigan State University	East Lansing	United States
1	Graduate School USA	Washington	United States
1	Intel (United States)	Santa Clara	United States
1	University of California, Davis	Davis	United States
1	University of Michigan–Ann Arbor	Ann Arbor	United States
1	Block Engineering (United States)	Southborough	United States
1	Capital University	Bexley	United States
1	New York Institute of Technology	New York	United States
1	Case Western Reserve University*	Cleveland	United States
1	Purdue University West Lafayette*	West Lafayette	United States
1	University of California, Berkeley	Berkeley	United States
1	The University of Texas MD Anderson Cancer Center	Houston	United States
1	Johns Hopkins University	Baltimore	United States
1	Carnegie Mellon University	Pittsburgh	United States
1	Amgen (United States)	Thousand Oaks	United States

1	Stanford Medicine	Stanford	United States
1	Rensselaer Polytechnic Institute	Troy	United States
1	Boston Children's Hospital*	Boston	United States
1	Center for Information Technology*	Bethesda	United States
1	University of California, San Francisco*	San Francisco	United States
1	Pfizer (United States)	New York	United States
1	King University*	Bristol	United States
1	ORCID	Bethesda	United States
1	FuelCell Energy (United States)	Danbury	United States
1	Lamar University	Beaumont	United States
1	Applied Materials (United States)	Santa Clara	United States
1	University of Chicago	Chicago	United States
1	Rutgers, The State University of New Jersey	New Brunswick	United States
1	University of Memphis	Memphis	United States
1	University of North Carolina at Chapel Hill	Chapel Hill	United States
1	Southern California University for Professional Studies	Irvine	United States
1	Optica	Washington	United States
1	The Ohio State University Wexner Medical Center	Columbus	United States
1	University at Buffalo, State University of New York	Buffalo	United States
1	Google (United States)*	Mountain View	United States
1	University of Arizona	Tucson	United States
1	Unchained Labs (United States)	Pleasanton	United States
1	Boston College*	Boston	United States

1	Health First*	Rockledge	United States
1	University of Akron	Akron	United States
1	Hunter College	New York	United States

[1] DeepSeek's announcement roiled US markets, leading to a 3 percent decline in the NASDAQ composite and a 17 percent drop in NVIDIA shares, erasing $600 billion in value. It was the largest single-day loss of a company in US history—a figure equivalent to 65 percent of the annual US defense budget. For more information: https://www.cnn.com/2025/01/27/tech/deepseek-stocks-ai-china/index.html.

[2] See Appendix A for details on each of the five DeepSeek papers.

[3] Note: For the first four papers, author lists were consistent between PDF and arXiv metadata. However, for the fifth paper (arXiv:2501.12948), we found discrepancies between authors listed in the PDF and the arXiv metadata, so we included all unique authors from both sources to ensure comprehensive coverage. Additionally, 11 authors across all papers could not be found with OpenAlex profiles and were excluded from the analysis.

[4] OpenAlex is a tool hosted by OurResearch, a nonprofit focused on open science tool development.

[5] See Appendix B for a full list of names in the Key Team.

[6] The h-index captures the number of publications with at least h citations (i.e., an h-index of 13 implies 13 papers cited at least 13 times), while the i10-index counts how many works have at least 10 citations—useful for gauging consistency across a body of work.

[7] See Appendix B for the full dataset.

[8] While 211 authors were included in the full bibliometric analysis, the affiliation-based breakdowns in this chart total 201 due to 10 individuals with no available institutional data in OpenAlex. These 10 authors also had very limited bibliometric profiles, with an average of just 4.4 publications, 8.6 citations, and near-zero recent citation activity—suggesting that they are likely junior researchers or early-career contributors. Their omission from affiliation analysis does not significantly affect aggregate findings but is noted here for transparency.

[9] For more information about institutional relationships in OpenAlex, see https://docs.openalex.org/api-entities/institutions/institution-object.