



Puzzle: Talent, Data, and R&D

Anirudh Suri

FEBRUARY 2025

The Missing Pieces in India's Al Puzzle: Talent, Data, and R&D

Anirudh Suri

© 2025 Carnegie Endowment for International Peace. All rights reserved.

Carnegie does not take institutional positions on public policy issues; the views represented herein are those of the author(s) and do not necessarily reflect the views of Carnegie, its staff, or its trustees.

No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from Carnegie India or the Carnegie Endowment for International Peace. Please direct inquiries to:

Carnegie Endowment for International Peace Publications Department 1779 Massachusetts Avenue NW Washington, D.C. 20036 P: + 1 202 483 7600 F: + 1 202 483 1840 CarnegieEndowment.org

Carnegie India Unit C-4, 5, 6, Edenpark, Shaheed Jeet Singh Marg New Delhi – 110016, India P: +011 4008687 CarnegieIndia.org

This publication can be downloaded at no cost at CarnegieIndia.org

Contents

Introduction	1
The Missing Pieces in India's AI Puzzle	3
Talent	4
Data	10
R&D	15
Conclusion: Balancing India's Competitiveness in AI With "AI for AII"	21
About the Author	23
Notes	25
Carnegie India	31



Introduction

The world is at a critical moment in the race for artificial intelligence (AI) leadership. As the global competition for leadership in AI heats up, the current trend is toward the concentration of data, capital, talent, and cutting-edge research in the hands of a few firms and even fewer countries.

The United States and China, the world's two "AI superpowers," are locked in what is being called an "AI arms race" for the faster development and adoption of AI.¹ Firms in these countries are building newer applications—commercial as well as military—for global adoption. The January 2025 release of DeepSeek-R1, an open-source model developed by a Chinese AI start-up, sparked panic in the United States' AI sector, serving as yet another example of the AI race heating up.²

At the same time, other countries—notably, India, Japan, France, Germany, the United Kingdom, Singapore, and the United Arab Emirates (UAE), among others—want to prevent such concentration and are charting their own AI strategies to compete in this arena. These countries are attempting to find ways to avoid being relegated to observer status in the global AI race.

This paper explores the question of whether India specifically will be able to compete and lead in AI or whether it will remain relegated to a minor role in this global competition. The paper argues that if India is to meet its larger stated ambition of becoming a global leader in AI, it will need to fill significant gaps in at least three areas urgently: talent, data, and research. Putting these three missing pieces in place can help position India extremely well to compete in the global AI race.

India's national AI mission (NAIM), also known as the IndiaAI Mission, was launched in 2024 and rightly notes that success in the AI race requires multiple pieces of the AI puzzle to be in place.3 Accordingly, it has laid out a plan across seven elements of the "AI stack": computing/AI infrastructure, data, talent, research and development (R&D), capital, algorithms, and applications.4

However, the focus thus far has practically been on only two elements: ensuring the availability of AI-focused hardware/compute and, to some extent, building Indic language models. India has not paid enough attention to, acted toward, and put significant resources behind three other key enabling elements of AI competitiveness, namely data, talent, and R&D.

Without plugging in these missing pieces, India is likely to fall short of its stated ambitions. Only by building key strengths in data, talent, and research will India be able to compete in building AI models and applications, which can in turn help Indian entrepreneurs build valuable AI companies. In each of the subsequent sections, this paper breaks down the constraints or problems concerning these three elements in detail and also lays out a clear set of recommendations for India to elevate its AI strategy across those three dimensions.

First, India must double down on boosting its AI talent and build up an optimal mix of top-, middle-, and low-tier AI talent. While Indian information technology (IT) services firms and global AI majors will naturally do their bit to create an AI-enabled workforce by upskilling India's existing IT services talent, doing so will not be enough to meet the country's ambitions. India will need to attract, nurture, and retain cutting-edge, top-tier AI research talent to ensure that AI innovations for the world emerge from India. This paper suggests ways for India to achieve this goal.

Second, India must immediately build up "digital public data" to provide the "oil" for Indiaspecific AI models and research. Despite being one of the largest smartphone, internet, and digital transactions markets in the world, American and Chinese firms have an advantage in this aspect vis-à-vis India. The vast majority of the digital data footprint of Indians is locked within platforms owned by global tech firms. This paper shows that to accelerate its unique data advantage, India needs to identify ways to proliferate multilingual data as well as other India-specific datasets. This can provide a differentiating factor for Indian large language models (LLMs) and small language models (SLMs) vis-à-vis their global counterparts.

Third, this paper argues that India must aim to become a leader in both cutting-edge AI research and the development of India-specific applications of AI. By enabling sufficient AI infrastructure through various public and private initiatives; attracting, nurturing, and retaining top-tier AI research talent; and accelerating the availability of large volumes of India-specific datasets, India can truly become a leader in cutting-edge AI research. This, in turn, will cause a proliferation of applications built on top of these enabling layers of infrastructure, top-tier research talent, India-specific data, and cutting-edge AI research.

Ultimately, India's AI strategy has to be crafted in the context of a challenging global AI landscape. Therefore, its "AI for All" approach—as outlined by India's government think tank, NITI Aayog⁵—must be complemented with a "Competitiveness in AI" strategy. This paper lays out the case and the way forward for this fundamental change in approach.

The Missing Pieces in India's AI Strategy: Filling the Gaps

With the launch of the national AI mission in 2024, India has laid out the framework for its approach to AI competitiveness. Early in March 2024, relying on many of the recommendations of the expert working group report of October 2023, the Indian cabinet approved a "comprehensive, national-level IndiaAI mission with a budget outlay of Rs. 10,371.92 crore" (approximately \$1.3 billion) over a period of five years.⁶

To begin with, the Indian government has attempted to establish the basic foundations of the AI ecosystem. In the past year, it has therefore focused on filling gaps in the AI infrastructure. India's Ministry of Electronics and Information Technology (MeitY) has facilitated procurement of AI chips and compute capacity, specifically, 10,000 graphics processing units (GPUs), to support India's start-ups, researchers, and academics.⁷

Geopolitically, access to AI chips and compute was at the top of the agenda for many countries in 2024. This was mainly due to the heavy concentration of advanced chip capacity in Taiwan and the growing risk of a Chinese invasion of Taiwan.8 India, recognizing the geopolitical and geoeconomic risk, rightly placed emphasis on the AI compute dimension first.

But, as with many other countries, the emphasis on chips and compute has come at the expense of some of the other crucial elements of the AI stack. This has meant that the holistic strategy needed to build out the AI ecosystem has not been fully acted upon. India's approach to AI competitiveness and its key pillars can currently best be characterized as a work in progress.

The following three sections argue why the three other elements of the AI stack—namely, talent, data, and research—need significantly greater focus from the various actors in India's AI strategy. Each section delves deep into the specific element to identify the constraints or gaps India currently faces and suggests ways to overcome those gaps.

Talent

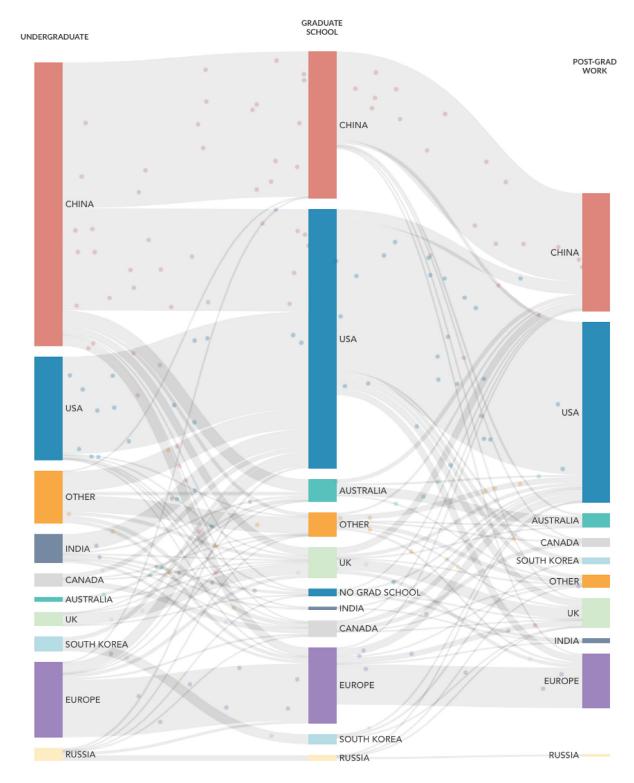
Talent in AI is touted as India's key strength and should be one of the layers of the AI stack that India specializes in. Today, India has one of the largest pools of science, technology, engineering, and mathematics (STEM) talent. However, the talent dimension for AI and India's talent gap in this area requires a more nuanced examination.

Breaking Down the Talent Gap

There are four problems that India needs to address on the talent front.

- 1. Overall shortage: By most estimates, countries around the world, including the United States, China, and Europe, as well as India, are facing a severe shortage of AI talent. This problem will only further exacerbate as demand booms in the coming years, but the training and education lags.9 The shortage is even more critical for India, given that AI presents a huge opportunity for the country. Without sufficient talent, India will not be able to capitalize on the opportunity fully. As AI and macrotrends enable globalization to shift from goods to services, India will also face a shortage of talent, despite being "a digital talent nation." ¹⁰
- 2. Talent Migration: The second issue for India to tackle is that of the migration of some of its best talent. On the talent aspect, the Indian government's strategy documents simply recommend increasing the number and type of AI courses at different levels. 11 But they fail to address, for example, the reasons why India's top-tier AI talent often migrates outside India. Neither do the MeitY working group's recommendations offer ways to reverse that flow.
 - As **Figure 1** shows, top-tier AI talent (that is, talent doing cutting-edge AI research) being trained in India until the undergraduate level, presumably at institutions such as the Indian Institutes of Technology (IITs), ends up working in the United States or Europe after completing their graduate (or postgraduate) work. Losing some of its high-potential AI research talent is a problem that India needs to rectify immediately.
- 3. The Quality of Talent and the Need for Upskilling: While India undoubtedly has a large pool of STEM talent, employers always complain about the job readiness of a large portion of Indian engineering graduates. 12 Fixing the quality element will also be critical; else, many of these engineering graduates will likely be unable to compete with AI for basic coding tasks.¹³ At the same time, the need to develop India's existing IT talent into an AI-enabled workforce through upskilling will be a critical challenge for India's tech and IT services industry.¹⁴

Figure 1. Global AI Talent Flow



Source: "The Global AI Talent Tracker 2.0," MacroPolo, accessed January 12, 2025, https://macropolo.org/digital-projects/the-global-aitalent-tracker/.

4. A Suboptimal Mix of Talent: AI talent can be sorted into three categories: top-tier (those conducting cutting-edge research, data scientists, and AI researchers), mid-tier (domain experts and application developers), and low-tier (project managers and implementers).¹⁵ For any country to become an AI superpower, it will, depending on its overall AI strategy, inevitably require an optimal mix of talent across all three tiers.

While some reports suggest that India is short on talent in all tiers, as per most estimates, India currently has predominantly low-tier AI talent. 16 As Figure 2 suggests, India has a growing comparative advantage in medium- and low-expertise talent, given its fast-growing developer base in the AI/machine learning (ML) space, especially compared to the rest of the world. India is one of the top contributors to GitHub AI projects, a fairly reliable metric for the number of developers undertaking AI coding projects.¹⁷ The Indian developer community on GitHub is now the second-largest and fastest-growing one, and it is expected to overtake that of the United States by 2028.18

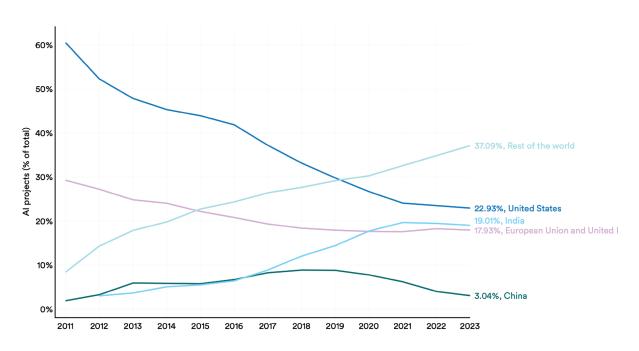
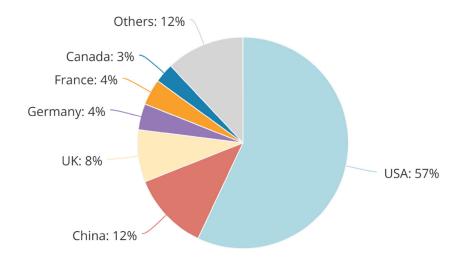


Figure 2. GitHub Developer Projects (Nonacademic) by Country

Source: Nestor Maslej et al., "Chapter 1: Research and Development," The Al Index 2024 Annual Report, Al Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024, accessed January 13, 2025, https://aiindex.stanford.edu/ wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter1.pdf.

Figure 3. Concentration of Top-Tier AI Talent





Source: "The Global Al Talent Tracker 2.0," MacroPolo, accessed January 30, 2025, https://macropolo.org/digital-projects/the-global-ai-talent-tracker/.

> But clearly, as Figure 3 suggests, the top-tier talent is currently concentrated in the United States, China, and European nations. Given the deep research and innovation ecosystems that exist in these countries, without a major push, it will be an uphill task for India to build and retain top-tier AI talent. And without an optimal mix across all three tiers, India will struggle to become a truly cutting-edge global AI power. To meet this ambition, India must therefore pursue various approaches to nurture, attract, and retain top-tier or high-level AI talent to supplement its midand low-tier talent base.19

Key Recommendations

For India to fully leverage its "talent nation" tag in the great AI game, it will need to evolve strategies to address the four problems highlighted above. Most importantly, it will need to grow the size and quality of its AI talent pool further, develop an optimal mix of top-, middle-, and bottom-tier talent, and continue exporting talent but also retaining and building up its domestic talent pool. Some recommendations in this regard are as follows:

- 1. Augment India's Overall Al Talent Pool: India should leverage the role global tech companies can play in this regard. For example, Microsoft CEO Satya Nadella has committed to training and equipping 10 million people with essential AI skills over the next five years.²⁰ India needs to bring at least twenty-five to thirty such tech companies on board to take on similar tasks and ensure that a variety of companies are training different types of AI talent within India.
- 2. Build an Optimal Mix of Talent: India will need a plan to build the various types of talent needed.²¹ India's talent strategy must have three layers: one, AI-related R&D scientists and researchers (the top-tier talent layer); two, AI developers, builders, and architects who are able to create AI technologies, systems, and applications in various sectors; and three, AI integrators with sufficient understanding of AI tools and solutions to be able to integrate them into their company's workflows.²²
- 3. Fix India's Top-Tier Al Research Talent Gap: High-level AI talent usually finds its way to high-tech sectors or emerging tech sectors and tends to concentrate in regions that have the most conducive and dynamic ecosystems (that is, other talented professionals, academics, researchers, companies, funders, and engineers, among others). As Yann LeCun, Meta's chief AI scientist, urged during a visit to India in October 2024, Indian PhDs and scientists also need to focus on AI research, not just engineering and development.²³

The high-performing talent problem for India can only be solved by evolving the overall AI ecosystem, which includes research, talent, industry, academia, and policy.²⁴ While this is a difficult task, it is not impossible. India needs to identify at least twenty-five to thirty universities that will train the next generation of top-tier AI talent.

In addition, India will need to focus on investing deeply in four to six centers of excellence in AI research through a combination of public and private sector sources. The Anusandhan National Research Foundation (ANRF) should also include a substantial AI pillar under its ambit. 25 India must also seek to bring some of its top-tier talent (currently working in the United States or Europe) to these centers to complement India's local AI research talent. Funded through a combination of sources, these centers should seek to offer globally competitive salaries, world-class infrastructure, and appropriate incentive structures for the AI researchers. These labs, like the Facebook AI Research lab set up in Paris a decade ago, would also help retain AI research talent in India.26

The MacroPolo Global AI Talent Tracker's 2023 update suggests that India's ability to retain its top-tier AI talent has grown since 2019. In 2019, while nearly all Indian AI researchers ended up pursuing opportunities outside of India, by 2022, almost 20 percent had chosen to stay in India.²⁷ This demonstrates that retention of top-tier AI talent, though difficult, is achievable.

- 4. Growing the Medium- and Low-Level Al Talent: This will be a relatively easier task for India, as its IT services firms, global capability centers (GCCs), start-ups, and large corporations attract and train talent for the adoption and implementation of existing AI technologies, algorithms, and models. Indian IT services firms should set aggressive targets for upskilling their existing IT workers rapidly to ensure their continuing relevance for clients.
- 5. Integrate Al Into the Education Curriculum at All Levels: India's current strategy does emphasize increasing the number of AI courses at the K-12, graduate, and postgraduate levels. As MeitY's IndiaAI 2023 expert group report rightly suggests, given the immediate short-term demand, industry and academia will need to collaborate to make sure that industry gets the specific types of trained talent it needs.²⁸ Planning of academic programs cannot be done in isolation. Academic institutions need to coordinate closely with industry so that the former can design the necessary programs to impart the requisite skills anticipated by the latter over the next three to five years.²⁹ For example, facing such a shortage, semiconductor companies in India are forging partnerships with academia to bridge this gap.³⁰
- 6. Develop a Broad Set of Al Skills Across Disciplines: The MeitY expert group has suggested career path mapping to supplement the AI curriculum so that India can produce not just AI engineers, but also AI entrepreneurs, product managers, designers, researchers, and ethicists.³¹ For example, management schools can train their students to become AI product and program managers. Science and research institutes should focus on ensuring promising AI research careers for their students and produce AI researchers and data scientists. Engineering schools would produce AI/ML engineers and DevOps professionals. Schools and online edtech platforms can teach fundamental AI skills at scale. Social science schools should encourage the study of sociopolitical, philosophical, and economic implications of AI for India.
- 7. Attract the Best STEM Talent From the Global South: India should develop a visa policy that can attract the best science and engineering talent from its neighboring regions, including AI talent from regions such as South Asia, Southeast Asia, Central Asia, Eastern Europe, Africa, and the Middle East. India currently does not attract many high-skilled immigrants, but such a visa regime will also allow India to benefit from the brain drain occurring in these countries.

India's G20 talent visa, announced in December 2024 and expected to come into effect in 2025, aims to attract top research scholars and fellows from G20 nations as a way to boost innovation in India. Such initiatives could help establish India as the "Silicon Valley of the East." 32 This could be similar to the International Entrepreneur Rule (IER), which allows the United States to attract those who "would provide a significant public benefit through their business venture." Many European countries have also adopted a similar rule. Given its market size and

vibrant start-up and venture capital ecosystems, India will be able to attract a decent portion of the AI talent that might otherwise flow to Western Europe and/or the United States.

8. Tackle the Challenge of Al-Induced Structural Unemployment Head-On:

Lastly, India must keep in mind the potential for AI to cause structural unemployment across various sectors and types of jobs. While Western nations face a chronic shortage of human capital and labor, India is in the opposite situation.³⁴ It must ensure that AI is leveraged for enhancing the productivity of its labor, not for replacing said labor and jobs.

Importantly, India's policies for job creation must also focus on training, reskilling, and upskilling Indian professionals and college graduates for the field of AI. There will be a considerable global shortage of AI professionals globally, much like what happened in the field of software development as well as cybersecurity.³⁵ If India's training, education, and skills institutes can reorient themselves toward this emerging technology, the country can capture a bigger portion of the jobs that will be created globally.36

Data

Data ultimately lies at the core of AI and is the actual oil for AI algorithms and models. It is a necessary but not sufficient condition for the development of useful AI-based products, models, and services. Therefore, a clear, comprehensive strategy for having continued access to data is necessary for success. American companies such as OpenAI and Google already have access to vast amounts of data, both public and proprietary, that have been leveraged for training their AI models.³⁷ Similarly, the Chinese have access to a huge amount of data, which they consider as a key advantage for them in the global AI race.³⁸

On the data element of the AI stack, good-quality, India-specific data in the volumes needed to train LLMs has thus far not been made readily available to start-ups, researchers, and innovators. Even well-funded Indian AI start-ups such as Sarvam have pointed to this fundamental problem, due to which they have had to rely on synthetic data for training their models.39

Breaking Down the Data Gap

Indian companies, start-ups, and researchers find themselves disadvantaged against their global peers on the data element of the AI stack due to various reasons. For India, there are at least a few dimensions to the data problem.

- 1. Lack of Access to Large Volumes of Existing Data: To start with, Indian startups and researchers do not have access to the massive volumes of data that Google, Meta, Microsoft, and others have access to by virtue of being Big Tech firms with consumer- and business-facing global platforms. It is either impossible for others to access this data (since these firms might not share this data) or it is very expensive (if one tries to purchase this data through data brokers).
- 2. Sparse Unique or Proprietary Data Within India: Indian firms currently do not have access to unique or proprietary data that could give them an edge over these global platforms. This unique or proprietary data could refer to data specific to Indian consumers or businesses, such as data from Unified Payments Interface (UPI) transactions. It could also refer to data in Indian languages that is not on the internet yet. (The Indian language data on the internet can be scraped by Big Tech firms anyway.)
- 3. Siloed, Unstructured, and Poor Data: Though large amounts of data are being generated in India, given its growing digital public infrastructure (DPI) and its large digital user base, the data is either lying in silos or has not been tapped yet. For example, data on trade through Indian ports or data on the usage of Indian toll booths is available yet lying in silos. Moreover, in addition to a dearth of easily accessible data or data-generating platforms, India also lacks well-annotated, regularly updated, feature-rich datasets.⁴⁰
- 4. Overreliance on the Government to Solve the Data Gap: The strategy outlined in the national AI mission seems overly reliant on the government to build and manage a data platform. The India AI Datasets Platform, expected to go live in Jan 2025, will aim to build a platform where developers can access and use datasets sourced from the private and public sectors. 41 The vision, as outlined by India's National e-Governance Division (NeGD), is for India to build a platform similar to Hugging Face, a private, venture-funded global repository of datasets and opensource models.42

However, a single government-managed platform will have a low likelihood of solving the data gap. A cursory analysis of leading, cutting-edge AI research work, or applications or models being built by leading AI firms, will show that their data needs are exhaustive. A government-managed data platform, despite best efforts, is likely to be plagued with issues, such as non-exhaustiveness of data, unstructured or unannotated data sets, or just bad data.⁴³

More thinking, therefore, will need to be done on whether the data element is best solved by the government or by private players or start-ups. The global AI datasets market is growing very rapidly,44 and Indian start-ups and researchers will need to find ways to plug into and access that data in cost-effective ways. Moreover, India

needs a clear, long-term strategy to pool together massive volumes of Indic language data to allow Indian companies to build custom models and applications for Indian consumers and businesses.

Key Recommendations

Data for AI is probably the toughest problem eventually. Some have argued that "those who solve the data dilemma will win the AI revolution."45 The solution lies in developing a longterm, creative strategy toward overcoming the "data disadvantage" that currently plagues Indian researchers, start-ups, and companies. This holistic strategy must incorporate a plan for data generation, access to globally publicly available and licensed data, improving data quality, multilingual data, and multimodal (voice-, text-, image-, and video-based) data.

A long-term strategic approach, rather than a short-term, tactical approach, will be needed to solve this constraint for India. India's strength in DPI and public data commons should be leveraged and incorporated for data across all sectors (public, private, and academic).

- 1. Leverage Indian Consumer/Transaction Data: India has to figure out ways to access, unlock, and leverage the vast amounts of data its large internet user base is creating. India's technology firms, including prominently its telecom, e-commerce, logistics, and fintech firms, generate immense amounts of multimodal data that is currently not easily accessible or leveraged for AI research or innovation. Firms such as Jio, Airtel, Flipkart, Zomato, Blinkit, Swiggy, Delhivery, MakeMyTrip, PhonePe, and many others are home to substantial repositories of data. The right regulatory frameworks and market-based approaches need to be developed to unlock this data for driving another wave of innovation, the way UPI has allowed for innovation in the fintech space.
- 2. Develop Multiple Data Marketplaces: India could develop platforms and protocols for sharing non-personally identifiable data—under the ambit of the necessary privacy, anonymity, and rules-based access norms—with AI entrepreneurs, researchers, and innovators. The India Datasets Platform could serve as a repository of data, but the government could also encourage the evolution of a broader set of marketplaces to solve this problem. In the United States, for example, several companies, such as SAP, Amazon Web Services (AWS), Databricks, Snowflake, and many others, operate such data marketplaces. 46 This model could be more scalable for India's needs, especially given such diverse data across the country.
- 3. Unlock Government Department Datasets: Similarly, data that is currently locked up within government departments (such as agriculture, health, finance, education, railways, civil aviation, and others) or other sources within the public sector should be opened up. Rural agricultural surveys, consumption surveys, flight and train data, toll booth data, UPI transaction data, health and medical data, trade

(import and export) data, and many other datasets reside in government departments. The focus should be on taking unstructured datasets and making them available for use in LLMs and other AI-based applications, especially in sectors such as education, finance, healthcare, travel and logistics, and agriculture, among others.

An interesting example that has emerged recently in India, for example, is the Integrated Geospatial Data-Sharing Interface (GDI) set up under the National Geospatial Policy of 2022 by the Department of Science and Technology, Government of India, through the Geospatial Data Promotion and Development Committee. 47 The GDI has compiled and made easily accessible datasets from various public and private partners pertaining to the sectors of agriculture, livelihoods, transportation, and logistics. Such sector-focused data platforms could serve as an interesting model for unlocking government data.

4. Scale Up Current Efforts: The government has already moved a few steps in this direction. In addition to the India AI Datasets Platform, which helps provide easy access to public sector datasets, India's Open Government Data Platform also hosts and provides application programming interface access to various datasets. 48 However, even though a decent number of datasets have been curated to facilitate research and innovation, these initiatives are still in very early stages of execution. Moreover, these data platforms currently suffer from various issues. For example, they are not updated periodically, nor do they provide standardized data in easily readable forms.49

Much like China's National Data Bureau (NDB), India had also proposed setting up the India Data Management Office to serve as India's data regulator, as part of its draft National Data Governance Framework Policy.⁵⁰ In addition, the MeitY IndiaAI expert working group report of 2023 also focused on operationalizing the India Datasets Platform and envisioned the establishment of data management units within each ministry/department.51

These efforts, while commendable, will need to be executed and scaled up significantly to yield results that are truly impactful. The commitment to providing past as well as real-time, updated, and complete data from various arms of the government must be genuine and driven from the top down. A half-hearted approach of putting selective data up on the platform will not serve any real purpose.

Appointing a chief data officer for India would help identify useful datasets across the government, establish ways to ensure good quality data streams, and streamline the efforts across ministries. Moreover, standard operating protocols need to be adopted and adhered to for ensuring data quality. An interesting benchmark that has been suggested for adoption by India is the European Union's Metadata Quality Dashboard, which can help assess the quality of uploaded data in terms of metrics such as accessibility, interoperability, and usability.⁵²

5. Build Up Repositories of Multilingual Data: The government is building opensource datasets in various Indian languages through Bhashini, an AI-powered translation system, with the goal of enabling the development of AI applications using these datasets.⁵³ Bhashini seems to be progressing well and reported having clocked over 100 million inferences in September 2024 across its multiple applications and use cases.⁵⁴ However, despite efforts to crowdsource multilingual data, training data for multilingual AI models is still scarce for mostly all Indian languages. India therefore needs to think of ways to collect, generate, and access multilingual data at scale.

A significant amount of multilingual data today is being generated on India's telecommunications platforms (through voice, messaging, and the creation of digital entertainment content in local Indian languages).⁵⁵ If India can set up processes to leverage this vast and continuously growing consumer-generated multilingual data for AI with the necessary regulatory guardrails, it could provide a big boost to India's aim to build multilingual AI models.⁵⁶

Of course, past regional radio, television, and newspaper records, once digitized and transcribed (potentially using AI itself), could also fill a big gap. This would be a more scalable way to have a deep pool of multilingual data than the on-ground crowdsourcing strategy suggested by some or by leveraging synthetic data as some other Indic-language models have done.⁵⁷

6. Develop a DPI-Like Approach and Guardrails for Data Commons: Of course, the availability and sharing of data raise questions of privacy, consent, and security. Previous efforts at commercializing government data sets have been criticized for these lacunae.⁵⁸ In August 2023, after several years of deliberations, India put in place a "modest and pragmatic" Digital Personal Data Protection (DPDP) Act, 2023, to enable data usage for lawful purposes by data fiduciaries while providing necessary protections to individuals or data principals over their data.⁵⁹ These are commendable steps to improve India's data capabilities within an effective data governance framework.

India should explore building out a DPI-driven approach to data as well. Data should be available as part of a digital commons to entrepreneurs to build applications on top of and not remain monopolized by just a few large firms. Data exchanges—with the necessary rules and guardrails—could serve as a key advantage for India in the global AI race. Therefore, the philosophical guiding values adopted in DPI by India, such as a consent-based architecture, and the appropriate data governance frameworks must also be incorporated in the design of these data exchanges.

R&D

Countries seeking to lead in the global AI race cannot ignore the R&D element of the AI stack. The United States, arguably the leader in AI innovation today, has clearly articulated R&D as a top priority for maintaining its global leadership in AI according to its detailed national AI R&D strategic plan, which was issued in 2016 and updated in 2019 and 2023.60

Similarly, China, way back in 2017, laid out a detailed stage-wise plan to become an AI R&D powerhouse by 2030.61 It envisioned starting from R&D in AI technology and applications in the first stage, followed by a research focus on basic AI theories in stage two, and finally a focus on advanced, cutting-edge AI research in the final stage. While it may not have invested as heavily as the United States, estimates still suggested that China had begun investing billions of dollars into AI R&D back then.⁶²

India has not conducted such a long-term-oriented strategic exercise on building its AI R&D capabilities, nor has it invested anywhere near as heavily in R&D the way the United States and China have in the last decade. A country's strength in AI-focused R&D can be measured both in terms of AI articles and journal publications and citations, but also patent applications and patents granted, not to mention the quality and quantity of its research talent pool. Various international studies have detailed out comparisons of R&D outputbased rankings of nations.

Two key takeaways for India emerge from these reports. One, as Figure 4 shows, India is gaining ground vis-à-vis the United States and other countries (but not China) in terms of papers in AI and related fields published between 2014 and 2024. The growth has been exceptionally impressive since 2019.

The second key takeaway is possibly the more important one. On a more quality-based metric, that of patents granted in AI, India does not fare as well (see Figure 5). This reflects the real gap between the global AI leaders, the United States and China, and India.

As with the talent and data dimensions, on the R&D front, a clear identification of the constraints or gaps India faces in building a cutting-edge AI research ecosystem, as well as the ways in which those will be tackled in the short and long terms, is needed.⁶³

120k

Figure 4. Papers in AI (and All Related Fields) by Nation, 2014-24

Source: "Country Activity Tracker (CAT): Artificial Intelligence," Emerging Technology Observatory, Center for Security and Emerging Technology, last updated December 18, 2024, accessed February 26, 2025, https://cat.eto.tech.

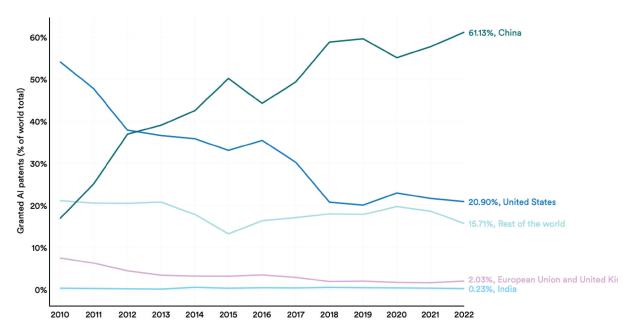


Figure 5. Patents Granted in AI by Region/Country, 2010-22

Source: Nestor Maslej et al., "Chapter 1: Research and Development," The Al Index 2024 Annual Report, Al Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024, accessed January 13, 2025, https://aiindex.stanford.edu/ wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter1.pdf.

2014

Breaking Down the R&D Gap in Al

1. Low R&D Spending on AI (and Innovation in General) by India's Private and Public Sectors: As per the data for 2020, the U.S. federal government now funds roughly 20 percent to industry's 70 percent of total national R&D activity.⁶⁴ In comparison, India's private industry contributes to only about 36.4 percent of gross expenditure on R&D.65 Globally, commercial industries are demonstrably better than government departments at converting the same R&D dollars into functional products. We cannot expect that the reality would be drastically different in India. But unfortunately, the private sector's share of R&D spending in AI is negligible today.

Similarly, the Indian public sector currently has negligible spending on AI R&D compared to its global counterparts. As a share of GDP, India's R&D spending overall is approximately 0.6 percent, compared to other innovation-focused countries that typically average 3 to 4 percent of GDP.66 While exact figures are not available, India's R&D spending on AI is likely even lower as a percentage of GDP compared to the United States, China, and other leaders in AI.⁶⁷

The national AI mission has also largely allocated the majority of its approved outlay toward AI infrastructure (approximately Rs. 4,500 crore or approximately \$ 515 million) and financing start-ups (approximately Rs. 2,000 crore or approximately \$230 million). The funds allocated toward establishing centers of excellence (CoEs), and hence R&D, are approximately Rs. 990 crore (approximately \$110 million).⁶⁸ This needs to be corrected, and R&D spending on AI needs to be substantially increased by both the public and private sectors.

- 2. Dearth of Institutions Focused on AI R&D: The current problem on this front is that India is doing meager AI research compared to others. There are only a handful of well-endowed institutional platforms for research. 69 This is leading researchers to migrate to the United States and Europe. None of India's educational or research establishments make it to the top AI research institutions. By comparison, the Stanford AI Index Report, 2023, lists nine Chinese universities and research institutions in the top ten when ranked by the number of AI publications in all fields during 2010-21.70
- 3. Al Patents Falling Behind Al Research: India also suffers from low quality of research (using globally established metrics of quality and citations). As the publications-to-patents ratio shows, India's AI patents are not keeping up with the quantity of its AI publications. India's share of AI-relevant research publications has grown substantially in the last ten to fifteen years—on this metric, it ranked fourth globally for 2010-2019. Yet, its share of global patents in AI has not increased proportionately, and it ranked eighth globally on this metric for 2002–2019.⁷¹

In terms of citations (again, a metric of research quality), India's rank drops down to fifteen. Even though international collaboration generally benefits research quality and impact, only 16 percent of India's AI research papers published during 2010-2019 had non-Indian co-authors. This was the lowest level of international collaboration among the top ten AI research-producing countries.⁷²

While other contextual factors (such as high costs involved in patenting, insufficient IP protections, and protracted patent litigations) might also contribute to this dichotomy, it is clear that the quality of AI research and AI patent activity in India is not in line with its global peers.

4. Lack of Cutting-Edge Al Infrastructure: India has established a national-level AI Research Analytics and Knowledge Dissemination Platform, which is designed to act as "a common cloud platform for Big Data Analytics with large AI computing infrastructure connecting all COREs, ICTAIs and other academic institutions with National Knowledge Network."73 The National Supercomputing Mission, as per government data, has a total compute capacity of 24.83 petaflops, with a target of reaching 66 petaflops by 2025.74

Further, in March 2024, the Indian government had allocated Rs. 4,564 crore (approximately \$544 million) under the IndiaAI mission to procure 10,000 GPUs through a public-private partnership model.⁷⁵ In February 2025, Indian IT Minister Ashwini Vaishnaw, recognizing the demands for advanced computation capacity for AI research and development, announced that the Indian government had already procured 10,000 GPUs but would enhance the available capacity to over 18,000 GPUs. 76 These GPUs, expected to comprise 12,896 Nvidia H100s and 1,480 Nvidia H200s, would be made available to start-ups as well as researchers at academic institutions.⁷⁷

The constraints placed on advanced GPU imports by certain countries by the US administration in January 2025 might, however, delay future acquisitions by India.⁷⁸

Despite these acquisitions, Indian start-ups and the newly established CoEs will be lagging behind their global counterparts in the United States and China severely, especially since governments and tech firms elsewhere are also prioritizing such compute infrastructure and GPU access, including for research and development.⁷⁹

5. Inadequate Resourcing of CoEs: So far, India has established three CoEs focused on AI. As per the recent budget for FY 2025-2026, it has also announced the establishment of a fourth CoE specifically for AI education. 80 Appointing private sector-led committees to monitor these CoEs should allow them to remain deeply integrated with industry requirements. But these centers will also need to be resourced adequately. For the three initially announced CoEs, Rs. 990 crore (approximately \$110 million) had been allocated over a period of five years, which comes to roughly \$7 million a year per center. 81 Although an additional Rs. 500

crore (approximately \$57 million) has been set aside for the fourth CoE, inadequate resourcing will prevent these institutions from attracting the best AI talent, affording sufficient compute capacity, and buying the necessary datasets to conduct cutting-edge research. These fundamental issues will need to be solved in partnership with the private sector.

Key Recommendations

- 1. Public-Private Push for Greater AI R&D Spending: The government must incentivize the private sector firms (including IT services majors, conglomerates, and others) to invest in AI R&D, either through building in-house AI R&D centers or through the AI research parks at premier universities in India as well as abroad.82 Indian IT services giants, as well as the large conglomerates, with their healthy balance sheets, are well-positioned both to do the R&D and also benefit from it.
- 2. Industry-Academia Partnerships: India should also encourage global tech firms to create industry-academia partnerships in India. For example, Nokia, through its new 6G Lab in Bengaluru, recently partnered with the Indian Institute of Science (IISc) to jointly conduct research in 6G radio, architecture, and AI/ML technologies that have a particular relevance to the Indian and global markets. 83 Such a template of industry-academia partnership should be replicated across at least fifteen to twenty technology companies and research institutions, possibly mapped to India's priority sectors and technological strengths. The government, on its part, should focus on defining functional needs, interface productively with industry, and create processes and policies to support research and development for innovation as well as fast adoption.
- 3. Identify Clear Focus Areas: Given comparatively low R&D spending, developing and charting a clear research strategy along with focus areas will help efficient use of existing R&D expenditure. As per the commerce ministry's AI Taskforce report and the NITI Aayog Strategy Paper, the specific five to ten domains where AI might find the most applicability and social benefit have already been identified.84 Working together with academia, researchers, and industry, the government must play a critical role in kick-starting the AI research ecosystem.

An analysis of the split of India's AI patents and papers so far would also help identify existing strengths and weaknesses. To begin with, India must focus on research areas where it has a competitive advantage or specific needs, such as personalized and precision medicine, gene therapy, vaccine discovery, drug design, and cancer screening, or optimized crop management.

4. Building the Research Ecosystem: The government's recent efforts, such as the "One Nation One Subscription" scheme, which provides Indian researchers with free access to the world's top journals, and the Partnerships for Accelerated Innovation and Research (PAIR) initiative, which functions as a hub-and-spoke model to pair India's top research institutions with others, are both commendable initiatives to boost the research ecosystem.85

Identification of talent, supporting that talent, and building ecosystems that become pockets or centers of excellence will be key. If needed, India must work on bringing some of the Indian-origin AI scientists working abroad to India and providing them with the necessary budgets and enabling ecosystems such that they can support India's research objectives.

For this, India will also need to evolve better-structured incentives so that researchers can more easily commercialize their research. Like their counterparts elsewhere in the world, India must encourage researchers and faculty members at its premier research institutes to launch start-ups and be stakeholders in the valuable companies that come out of that research.

5. Empower AI Research Parks and CoEs to Become Globally Competitive:

The AI parks that have been set up across the country, including at IIT Bombay and IISc, must be empowered with autonomy, talent, funding, and a strong intellectual property regime to really expedite the creation of dynamic AI research ecosystems. India will also need to pour much greater funding into cutting-edge, next-generation AI research in these institutions to be globally competitive. China, for example, had already set up more than sixty AI tech parks by 2018, which were providing financial incentives to attract AI companies. In addition, Beijing had also announced the setting up of a \$2.1 billion AI tech park, while another province, Tianjin, announced plans to establish a \$16 billion AI fund. 86 This is, of course, in addition to the private investment being poured into AI by China's tech and industrial firms.87

The ANRF set up by India is an excellent step forward and is launching various initiatives to make India's researchers globally competitive. But giving India's research institutions and parks autonomy, along with the ability to raise funds from other sources, easily commercialize their research, and launch start-ups based on that research, will be equally important.

6. Leverage the Existing R&D Centers and GCCs Set Up by Multinational Corporations in India: India has already emerged as a major hub of R&D activity for many global corporations that have set up R&D centers and GCCs in India.88 A key focus area for many of these centers is AI and ML. A recent Zinnov-NASSCOM report suggested that India has over 1,700 GCCs that employ over 1.9 million people and generated over \$64 billion in revenues in FY 2024. Out of this, their revenue for engineering R&D stood at \$36 billion.89

India must systematically work on leveraging the top AI- and ML-focused R&D talent currently housed within these captive R&D centers. They must be incentivized to branch out, raise funding, build innovative start-ups in the AI space, and conduct research in partnership with industry through the AI research parks.

7. International Collaboration: Through collaboration with the United States, Europe, Australia, Japan, and other friendly nations, India must protect against the unethical use of AI, help evolve a global consensus on the guardrails for AI development, as well as prioritize more open-source AI development. At the same time, to enhance its own capacities as well, India should push for international cooperation on AI R&D through joint research with various nations. 90 To bolster joint authorship of cutting-edge AI research, Indian AI researchers must be given greater incentives and funding along with stronger institutional support for conducting collaborative research with non-Indian AI researchers.

Setting up AI R&D exchange programs between Indian research universities and global ones, sponsoring international AI fellowships for emerging AI researchers, and other bilateral and multilateral partnerships to foster the exchange of ideas and expertise in AI R&D can also help bolster the Indian AI research workforce. Some emerging examples of the latter include the declaration of the United States and the United Kingdom on cooperation in AI R&D and the Quad countries' commitment to establish working groups on AI standards development and foundational research.91

Conclusion: Balancing India's Competitiveness in AI With "AI for All"

India's AI approach and competitiveness strategy have to be crafted within a challenging global context, one in which the factors of AI production are concentrated in the hands of a few countries and a few firms. The incumbent tech powers possess substantial advantages in hardware, data, algorithms, software, researchers, and capital. In such a scenario, Indian start-ups and enterprises admittedly lack access to a level playing field to compete in the foundational AI space in the short term.

The first-mover advantage of the United States and China on the one hand and Big Tech firms on the other does appear daunting. India must therefore build on its strong foundations in AI domestically to boost its ability to compete on the global stage. In addition to the existing emphasis on AI compute and hardware, this paper has argued that India must focus on solving the fundamental problems it faces in building out the talent, data, and research ecosystems in AI.

Without plugging these gaps, there could remain a big chasm between India's ambitions and the capabilities of its researchers, entrepreneurs, and businesses to lead in AI. In many ways, China's clear emphasis on building strong talent and research ecosystems in AI has contributed to its recent success with DeepSeek's R1 model challenging the dominance of leading American ones.⁹² Solving these missing pieces in India's AI puzzle, therefore, will be similarly critical for boosting India's AI competitiveness.

Along with these efforts to boost India's AI competitiveness, India must continue its efforts to create a level playing field in AI globally. India has already attempted to use its term as president of the Global Partnership on Artificial Intelligence to bridge the increasingly evident divide between the Global North and South in AI development and adoption.⁹³ It should continue to build a strong voice for the Global South in AI and work hard along with other similarly placed nations to prevent a Second Great Divergence between the AI-haves and AI-have-nots.94

In addition, as it has done with DPI, India must build global coalitions to extend the DPI approach to the global AI landscape. This could be on the AI cloud infrastructure front through initiatives such as the Open Cloud Compute or the data front to build a global data commons.⁹⁵ The extension of India's DPI approach to AI would be a significant contribution toward global AI governance frameworks. More research is needed to suitably craft this extension.

India must also upgrade its capacity to engage in AI standards- and principles-setting processes with organizations globally, including the International Electrotechnical Commission and the International Organization for Standardization, among others. 6 Collaborating with Europe, the Middle East, and other nations on forging a coalition around open-source standards for AI could be an example of a specific AI-focused coalition that India could push. Such multi-stakeholder coalitions of like-minded nations and companies that prioritize open-source development of AI could enable greater collaboration and the development of broad-based AI innovation ecosystems.

Ultimately, balancing its existing "AI for All" approach, both at the domestic level and on the global stage, with a "Competitiveness in AI" approach, as laid out in this paper, will be essential for India to achieve a leadership position in the highly competitive global AI ecosystem.

About the Author

Anirudh Suri is a nonresident scholar with Carnegie India, where his work focuses on the intersection of technology, AI, and geopolitics, and how India is carving its role in the global tech ecosystem.

Suri is also the author of *The Great Tech Game: Shaping Geopolitics and the Destinies of Nations* (HarperCollins, 2022) and host of The Great Tech Game podcast. He serves as the managing partner at India Internet Fund, a technology-focused venture capital fund based in India and the United States. Previously, he has worked with the Government of India in New Delhi, McKinsey and Co. in New York, Goldman Sachs in London, and the China Institute of International Studies in Beijing.

He has also written extensively in publications such as *The Indian Express, Times of India*, *Hindustan Times, Foreign Policy, ThePrint, Hindu, The Week, New Republic, Economic Times, Moneycontrol*, and *Asia Times*.

He completed his MBA from the Wharton School at the University of Pennsylvania, his MPA from the Harvard Kennedy School, and his bachelor's degree from Haverford College, Pennsylvania. He has also studied at the London School of Economics and Beijing Foreign Studies University. Named a Goldman Sachs Global Leader, he has also served on the global board of the Harvard Alumni Association.

Acknowledgments

The author wishes to acknowledge the contributions, support, and feedback of various colleagues at Carnegie India and Carnegie globally, including Rudra Chaudhuri, Anirudh Burman, Matt Sheehan, Milan Vaishnav, and the Carnegie editorial team. The author is also grateful for the valuable inputs of various stakeholders from government, research institutions, start-ups, venture capital firms, think tanks, and industry in India, the United States, Europe, and the Middle East.

Notes

- 1 "The AI Arms Race," FT Series, *Financial Times*, accessed January 11, 2025, https://www.ft.com/content/21eb5996-89a3-11e8-bf9e-8771d5404543.
- 2 See Steve Kopack and Brian Cheung, "Tech Stocks Fall as China's DeepSeek Sparks U.S. Worries About the AI Race," NBC News, January 27, 2025, https://www.nbcnews.com/business/markets/tech-stocks-react-chinas-deepseek-sparks-us-worries-ai-race-rcna189394; Simone McCarthy, "China Celebrates DeepSeek's Breakout AI Success as Tech Race Heats Up," CNN, January 28, 2025, https://edition.cnn.com/2025/01/28/china/china-deepseek-ai-success-tech-intl-hnk/index.html.
- 3 Ministry of Electronics and IT, Government of India, "Cabinet Approves Over Rs. 10,300 Crore for IndiaAI Mission, Will Empower AI Startups and Expand Compute Infrastructure Access," press release, Press Information Bureau, March 7, 2024, https://pib.gov.in/PressReleasePage.aspx?PRID=2012375.
- 4 For the definition of AI stack used in this paper, see *Final Report*, National Security Commission on Artificial Intelligence, 31, accessed January 29, 2025, https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai-full report digital.04d6b124173c.pdf.
- 5 "Responsible AI: #AIForAll, Approach Document for India, Part 1 Principles for Responsible AI," NITI Aayog, February 2021, https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf.
- 6 See IndiaAI 2023, Expert Group, Ministry of Electronics and Information Technology, Government of India, October 2023, https://www.meity.gov.in/writereaddata/files/IndiaAI-Expert-Group-Report-First-Edition.pdf; "Cabinet Approves Ambitious IndiaAI Mission to Strengthen the AI Innovation Ecosystem," Prime Minister of India, March 7, 2024, https://www.pmindia.gov.in/en/news-updates/cabinet-approves-ambitious-indiaai-mission-to-strengthen-the-ai-innovation-ecosystem/.
- 7 "IndiaAI Mission GPU Tender: 19 Companies Submit Bids," *Economic Times*, December 3, 2024, https://economictimes.indiatimes.com/tech/artificial-intelligence/indiaai-mission-gpu-tender-19-companies-submit-bids/articleshow/115940904.cms?from=mdr.
- 8 See Bradley Martin et al., "Supply Chain Interdependence and Geopolitical Vulnerability: The Case of Taiwan and High-End Semiconductors," RAND Corporation, March 13, 2023, https://www.rand.org/content/dam/rand/pubs/research_reports/RRA2300/RRA2354-1/RAND_RRA2354-1.pdf; Cliff Harvey Venzon, "US Chip Supply 'Too Concentrated' Globally, Raimondo Says," *Bloomberg*, March 12, 2024, https://www.bloomberg.com/news/articles/2024-03-12/us-chip-supply-too-concentrated-in-few-nations-raimondo-says.
- 9 "Bridging the AI Talent Gap to Boost India's Tech and Economic Impact," press release, Deloitte Touche Tohmatsu India LLP, August 20, 2024, https://www2.deloitte.com/in/en/pages/deloitte-analytics/articles/bridging-the-ai-talent-gap-to-boost-indias-tech-and-economic-impact-deloitte-nasscom-report.html.

- 10 Rekha M. Menon, "We Are the Undisputed Digital Talent Nation," Times of India, August 17, 2022, https:// timesofindia.indiatimes.com/business/india-business/we-are-the-undisputed-digital-talent-nation-rekha-mmenon/articleshow/93613881.cms; "India as a Digital Talent Nation | NASSCOM Strategic Review Report 2022," NASSCOM Community, NASSCOM, April 20, 2022, https://community.nasscom.in/communities/ it-services/india-digital-talent-nation-nasscom-strategic-review-report-2022.
- 11 Cabinet, "Cabinet Approves Ambitious IndiaAI Mission to Strengthen the AI Innovation Ecosystem," press release, Press Information Bureau, March 7, 2024, https://pib.gov.in/PressReleaseIframePage. aspx?PRID=2012355.
- 12 See Sunainaa Chadha, "Only 10% of India's 1.5 Mn Engineering Graduates to Secure Jobs This Year," Business Standard, September 16, 2024, https://www.business-standard.com/finance/personal-finance/only-10-ofindia-s-1-5-mn-engineering-graduates-set-to-secure-jobs-this-yr-124091600127_1.html; "Over 80% Indian Engineers Are Unemployable, Lack New-Age Technology Skills: Report," India Today, March 21, 2019, https://www.indiatoday.in/education-today/news/story/over-80-indian-engineers-are-unemployable-lack-newage-technology-skills-report-1483222-2019-03-21.
- 13 "Bridging the AI Talent Gap to Boost India's Tech and Economic Impact."
- 15 The Report of the U.S. National Security Commission on Artificial Intelligence, for example, has a useful categorization of talent based on their level of expertise: high, medium, and low. High AI expertise talent (or "researchers") refers to algorithm experts and those involved in cutting-edge research and development of AI technologies, while the medium expertise talent (or "implementers") will refer to those responsible for data cleaning, model training, and other such tasks that require less training and education than high-level AI experts. The third category of AI talent is working at or with the end user to integrate readily available AI solutions into daily business workflows, not unlike enterprise tech teams today that drive the use of currently available software solutions at businesses. See Final Report, National Security Commission on Artificial Intelligence, https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai full report digital.04d6b124173c.pdf. See note 4 for the full citation.
- 16 See Madhav Krishna, "Is India's Talent Pool Ready for India Inc's AI Requirements?," Forbes India, June 21, 2024, https://www.forbesindia.com/blog/technology/is-indias-talent-pool-ready-for-india-incs-airequirements-330398.html; Amit Kapoor, "Mind the Gap, Then Fix It: The Mismatch Between Workforce Skills and Job Market Demands in India," Economic Times, June 11, 2024, https://economictimes.indiatimes. com/opinion/et-commentary/mind-the-gap-then-fix-it-the-mismatch-between-workforce-skills-and-jobmarket-demands-in-india/articleshow/111668404.cms?from=mdr.
- 17 "Data | India Tops AI Projects in GitHub," The Hindu, August 21, 2023, https://www.thehindu.com/data/ data-india-tops-ai-projects-in-github/article67214270.ece.
- 18 "GitHub CEO Thomas Dohmke Praises India for Fastest-Growing Developer Community, Says 'Rise as a Global Tech Titan Is...," Mint, October 30, 2024, https://www.livemint.com/companies/news/github-ceothomas-dohmke-praise-india-global-tech-titan-fastest-growing-developer-community-ai-projects-companynews-11730279405977.html.
- 19 Anirudh Suri, "Winning the AI Race With Research Talent," Hindustan Times, November 3, 2024, https:// www.hindustantimes.com/opinion/winning-the-ai-race-with-research-talent-101730644977334.html.
- 20 "Microsoft Announces US \$3bn Investment Over Two Years in India Cloud and AI Infrastructure to Accelerate Adoption of AI, Skilling, and Innovation," Microsoft, January 7, 2025, https://news.microsoft. com/en-in/microsoft-announces-us-3bn-investment-over-two-years-in-india-cloud-and-ai-infrastructure-toaccelerate-adoption-of-ai-skilling-and-innovation/.
- 21 "How Semiconductor Companies Can Fill the Expanding Talent Gap," McKinsey & Company, February 2, 2024, https://www.mckinsey.com/industries/semiconductors/our-insights/how-semiconductor-companiescan-fill-the-expanding-talent-gap.
- 22 Advancing India's AI Skills: Interventions and Programmes Needed, (NASSCOM, Deloitte Touche Tohmatsu India LLP, August 2024,) https://www2.deloitte.com/content/dam/Deloitte/in/Documents/deloitte-analytics/ in-da-deloitte-nasscom-ai-skilling-in-india-report-noexp.pdf.
- 23 Chandra R. Srikanth and Vikas S. N., "AI Pioneer Yann LeCun: India Must Embrace Open Source, Invest in Research to Become an AI Hub Like France," Moneycontrol, October 23, 2024, https://www.moneycontrol. com/technology/ai-pioneer-yann-lecun-india-must-embrace-open-source-invest-in-research-to-become-an-aihub-like-france-article-12849068.html.
- 24 Anirudh Suri, "Winning the AI Race With Research Talent."
- 25 "Home: Anusandhan National Research Foundation," Science and Engineering Research Board," accessed January 14, 2025, https://serb.gov.in/page.

- 26 Chandra R. Srikanth and Vikas S. N., "AI Pioneer Yann LeCun: India Must Embrace Open Source, Invest in Research to Become an AI Hub Like France."
- 27 "The Global AI Talent Tracker 2.0," MacroPolo, accessed January 14, 2025, https://macropolo.org/digitalprojects/the-global-ai-talent-tracker/.
- 28 IndiaAI 2023, Expert Group, Ministry of Electronics and Information Technology, Government of India.
- 29 V. Ramgopal Rao, "Semiconductor Mission's Great But Academia Can Chip In," Times of India, March 3, 2024, https://timesofindia.indiatimes.com/home/sunday-times/all-that-matters/semiconductor-missionsgreat-but-academia-can-chip-in/articleshow/108168309.cms.
- 30 See Sreeradha Basu and Brinda Sarkar, "In India's Semiconductor Moment, Spotlight on Talent," Economic Times, March 9, 2024, https://economictimes.indiatimes.com/jobs/mid-career/in-indias-semiconductormoment-spotlight-on-talent/articleshow/108336340.cms?; Ayushman Baruah, "Semiconductor Companies Partner With Academia to Bridge Skills Gap," Entrepreneur India, September 16, 2024, https://www. entrepreneur.com/en-in/technology/semiconductor-companies-partner-with-academia-to-bridge/479886.
- 31 IndiaAI 2023, Expert Group, Ministry of Electronics and Information Technology, Government of India, 9.
- 32 "India's New G20 Talent Visa Is All Set to Attract Top Scholars From G20 Nations," India Today, December 19, 2024, https://www.indiatoday.in/education-today/news/story/india-launches-g20-talent-visa-attract-topscholars-from-g20-nations-2652280-2024-12-19.
- 33 "International Entrepreneur Rule | USCIS," U.S. Citizenship and Immigration Services, accessed January 14, 2025, https://www.uscis.gov/working-in-the-united-states/international-entrepreneur-rule.
- 34 See Shefali Anand, "India: What You Need to Know About the World's Largest Workforce," Society for Human Resource Management, August 3, 2023, https://www.shrm.org/topics-tools/news/india-need-toknow-worlds-largest-workforce; "Reaping the Demographic Dividend," EY India, April 11, 2023, https:// www.ey.com/en_in/insights/india-at-100/reaping-the-demographic-dividend.
- 35 Remco Zwetsloot, Roxanne Heston, and Zachary Arnold, "Strengthening the U.S. AI Workforce," Center for Security and Emerging Technology, September 2019, https://cset.georgetown.edu/publication/strengtheningthe-u-s-ai-workforce/; Sean Mitchell, "Global Enterprises Struggling With AI Talent Shortage," IT Brief UK, TechDay, August 1, 2024, https://itbrief.co.uk/story/global-enterprises-struggling-with-ai-talent-shortage.
- 36 Raghav Aggarwal, "Demand for Indian AI Talent to Double by 2027 But Quality a Hurdle: Report," Business Standard, August 20, 2024, https://www.business-standard.com/industry/news/demand-for-indian-ai-talentto-double-by-2027-but-quality-a-hurdle-report-124082000896_1.html.
- 37 "How ChatGPT and Our Foundation Models Are Developed," OpenAI Help Center, accessed January 15, 2025, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed.
- 38 As a 2019 Deloitte report argued, "the data Chinese companies can access are more complex and multidimensional and serve as a solid foundation for the algorithm upgrade of AI technologies and the expansion of application scenarios" (see page 16 in Amit Kumar, "National AI Policy/Strategy of India and China: A Comparative Analysis," Research and Information System for Developing Countries, RIS Discussion Paper Series, Discussion Paper #265, June 2021, https://www.ssc-globalthinkers.org/system/files/2021-06/ DP%20265%20Amit%20Kumar.pdf). Based on the author's correspondence with Matt Sheehan, an expert on China's AI strategy, much of this "purported" Chinese data advantage has been accomplished passively through the presence of various Big Tech firms and not through active design. Nevertheless, the Chinese government has also established a National Data Bureau (NDB) for "coordinating the integration, sharing, development, and utilization of data resources." This is in addition to province-level data management bureaus and various measures to create an "open, safe, fair and vibrant data market, and strengthen the competitiveness of China's data economy globally" (see Jian Xu, "What Does China's Newly Launched National Data Bureau Mean to China and Global Data Governance?," Internet Policy Review, April 25, 2023, https://policyreview.info/articles/news/chinas-national-data-bureau-and-global-data-governance).
- 39 "What Is Sarvam-1, a New AI Model 'Optimized' for 10 Indian Languages?," Indian Express, October 26, 2024, https://indianexpress.com/article/technology/artificial-intelligence/what-is-sarvam-1-a-new-ai-modeloptimised-for-10-indian-languages-9638492/.
- 40 Rudraksh Lakra and Rutuja Pol, "Charting India's AI Future: Overcoming Data Challenges for Responsible Innovation," The Secretariat, July 9, 2024, https://thesecretariat.in/article/charting-india-s-ai-futureovercoming-data-challenges-for-responsible-innovation.
- 41 "IndiaAI Datasets Platform Set to Go Live by January Next Year," Economic Times, October 10, 2024, https://economictimes.indiatimes.com/tech/artificial-intelligence/indiaai-datasets-platform-to-launch-byjanuary-2025/articleshow/114088962.cms.
- 42 "IndiaAI Datasets Platform Set to Go Live by January Next Year," Economic Times.

- 43 Rohan Pai, "India's AI Development Faces Challenges, Calling MeitY's Attention," Deccan Herald, June 17, 2024, https://www.deccanherald.com/opinion/indias-ai-development-faces-challenges-calling-meitysattention-3069360.
- 44 Tajammul Pangarkar, "AI Training Dataset Market to Hit USD 11.7 Billion by 2032," Market.us Scoop, March 18, 2024, https://scoop.market.us/ai-training-dataset-market-news/.
- 45 Jim Stratton, "Those Who Solve the Data Dilemma Will Win the A.I. Revolution," Fortune, August 10, 2023, https://fortune.com/2023/08/10/workday-data-ai-revolution/.
- 46 See "Data Marketplace—AWS Data Exchange—AWS," AWS, accessed January 16, 2025, https://aws.amazon. com/data-exchange/; Lucy Kelly, "50+ Best Data Marketplaces in 2024," Monda, September 27, 2024, https://www.monda.ai/blog/best-data-marketplaces-guide.
- 47 See "GDI | Integrated Geospatial Data-Sharing Interface," GDI, accessed January 16, 2025, https://catalogue. gdi.cdpg.org.in/; Ministry of Science and Technology, Government of India, "Operation Dronagiri Launched Along With GDI Marking a Milestone in the National Geospatial Policy," press release, Press Information Bureau Delhi, November 14, 2024, https://www.pib.gov.in/PressReleasePage.aspx?PRID=2073284.
- 48 See "All Datasets," IndiaAI, accessed January 16, 2025, https://indiaai.gov.in/datasets/all; Ministry of Electronics and Information Technology, Government of India, "Cabinet Approves Over Rs. 10,300 Crore for IndiaAI Mission, Will Empower AI Startups and Expand Compute Infrastructure Access," press release, Press Information Bureau, March 7, 2024, https://pib.gov.in/PressReleasePage.aspx?PRID=2012375#; "APIs | Open Government Data (OGD) Platform India," Open Government Data Platform, accessed January 16, 2025, https://data.gov.in/apis.
- 49 Antara Vats, "A Decade Into India's Open Government Data Journey," Observer Research Foundation, September 28, 2022, https://www.orfonline.org/expert-speak/a-decade-into-indias-open-government-data-
- 50 See Jian Xu, "What Does China's Newly Launched National Data Bureau Mean to China and Global Data Governance?" Also see "National Data Governance Framework Policy (Draft)," Ministry of Electronics and Information Technology, Government of India, May 2022, https://www.meity.gov.in/writereaddata/files/ National-Data-Governance-Framework-Policy.pdf.
- 51 IndiaAI 2023, Expert Group, Ministry of Electronics and Information Technology, Government of India,
- 52 See "Metadata Quality Dashboard," European Union, accessed January 16, 2025, https://data.europa.eu/ mqa/?locale=en; Rohan Pai, "India's AI Development Faces Challenges, Calling MeitY's Attention."
- 53 See "Bhashini Translation System: India Turns to AI to Capture Its 121 Languages," Economic Times, December 4, 2023, https://government.economictimes.indiatimes.com/news/digital-india/bhashinitranslation-system-india-turns-to-ai-to-capture-its-121-languages/105712238.
- 54 Bhashini, "BHASHINI Celebrates Milestone Achievement of 100+ Million Inferences as of September 2024!," LinkedIn post, accessed January 16, 2025, https://www.linkedin.com/posts/digiital-india-bhashinidivision bhashini-celebrates-milestone-achievement-activity-7258692099808624640-WC-6/.
- 55 "Indian Languages—Defining India's Internet," KPMG Assurance and Consulting Services LLP, April 25, 2017, https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf.
- 56 Antara Vats, "A Decade Into India's Open Government Data Journey."
- 57 Vishal Dhupar, "India Enterprises Serve Over a Billion Local Language Speakers Using LLMs Built With Nvidia AI," blog, NVIDIA, October 23, 2024, https://blogs.nvidia.com/blog/llms-indian-languages/.
- 58 Antara Vats, "A Decade Into India's Open Government Data Journey."
- 59 See Anirudh Burman, "Understanding India's New Data Protection Law," Carnegie India, October 3, 2023, https://carnegieendowment.org/research/2023/10/understanding-indias-new-data-protectionlaw?lang=en; The Digital Personal Data Protection Act, 2023 (No. 22 of 2023), Enforced August 11, 2023; Ministry of Electronics and Information Technology, Government of India, "Salient Features of the Digital Personal Data Protection Bill, 2023," Press Information Bureau Delhi, August 9, 2023, https://pib.gov.in/ PressReleaseIframePage.aspx?PRID=1947264.
- 60 See "The National Artificial Intelligence Research and Development Strategic Plan," Networking and Information Technology Research and Development Subcommittee, National Science and Technology Council, Government of the United States of America, October 2016, https://www.nitrd.gov/pubs/national ai rd strategic plan.pdf; "National Artificial Intelligence Research and Development Strategic Plan, 2023 Update," Select Committee on Artificial Intelligence, National Science and Technology Council, Government of the United States of America, May 2023, https://www.whitehouse.gov/wp-content/uploads/2023/05/ National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf.

- 61 "Next Generation Artificial Intelligence Development Plan," State Council of the People's Republic of China, September 15, 2017, http://fi.china-embassy.gov.cn/eng/kxjs/201710/P020210628714286134479.pdf.
- 62 Ashwin Acharya and Zachary Arnold, "Chinese Public AI R&D Spending: Provisional Findings," Center for Security and Emerging Technology, December 2019, https://cset.georgetown.edu/publication/chinese-publicai-rd-spending-provisional-findings/.
- 63 Anirudh Suri, "Winning the AI Race With Research Talent."
- 64 "U.S. Research and Development Funding and Performance: Fact Sheet," R44307, Congressional Research Service, updated September 13, 2022, https://sgp.fas.org/crs/misc/R44307.pdf.
- 65 Sarthak Pradhan and Pranay Kotasthane, "Policy Tweaks Can Push Private Sector R&D," Deccan Herald, December 6, 2024, https://www.deccanherald.com/opinion/policy-tweakscan-push-private-sectorrd-3305698.
- 66 See "R&D Expenditure Ecosystem: Current Status & Way Forward," Economic Advisory Council to the Prime Minister, Government of India, July 2019, https://www.indiascienceandtechnology.gov.in/sites/ default/files/file-uploads/roadmaps/1571900991_R%26D%20book%20expenditure%20ecosystem. pdf. Also see Vishwa Mohan, "India Improves Its R&D Expenditure But Lags Behind Many Countries Including China, USA, and Israel," Times of India, November 29, 2024, https://timesofindia.indiatimes.com/ india/india-improves-its-rd-expenditure-but-lags-behind-many-countries-including-china-usa-and-israel/ articleshow/115813953.cms.
- 67 Ahmed H. Al-Marzouqi and Alya A. Arabi, "A Comparative Analysis of the Performance of Leading Countries in Conducting Artificial Intelligence Research," Human Behavior and Emerging Technologies 2024, no. 1, September 12, 2024, https://onlinelibrary.wiley.com/doi/10.1155/2024/1689353.
- 68 See Ashutosh Mishra, "Cabinet Nod for India AI Mission With Rs. 10,372 Crore Outlay for 5 Years," Business Standard, March 8, 2024, https://www.business-standard.com/india-news/cabinet-nod-to-india-aimission-with-rs-10-372-crore-outlay-for-5-years-124030701231 1,html. Also see "Rs. 990 Crore Approved by Government to Establish Three AI Centres of Excellence," India Today, October 19, 2024, https://www. indiatoday.in/education-today/news/story/government-approves-rs-990-crore-for-establishing-three-ai-centresof-excellence-2619679-2024-10-19#.
- 69 Report of Committee—C on Mapping Technological Capabilities, Key Policy Enablers Required Across Sectors, Skilling and Re-Skilling, R&D, (Ministry of Electronics and Information Technology, Government of India, July 2019), https://www.meity.gov.in/writereaddata/files/Committes C-Report-on RnD.pdf.
- 70 Nestor Maslej et al., "Chapter 1: Research and Development," The AI Index 2023 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023, https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_1-1.pdf.
- 71 For the data cited in this paragraph, see Husanjot Chahal, Sara Abdulla, Jonathan Murdick, and Ilya Rahkovsky, "Mapping India's AI Potential," Center for Security and Emerging Technology, March 2021, https://cset.georgetown.edu/wp-content/uploads/CSET-Mapping-Indias-AI-Potential-1.pdf.
- 73 Amit Kumar, "National AI Policy/Strategy of India and China: A Comparative Analysis."
- 74 Amlan Mohanty, "Compute for India: A Measured Approach," Carnegie India, May 17, 2024, https:// carnegieindia.org/posts/2024/05/compute-for-india-a-measured-approach?lang=en¢er=india.
- 75 Amlan Mohanty, "Compute for India: A Measured Approach."
- 76 "India's Own GPU Could Arrive in 3-5 Years, 18,000 AI Servers to Go Live Soon Says Ashwini Vaishnaw," Times of India, February 6, 2025, https://timesofindia.indiatimes.com/technology/tech-news/indias-own-gpucould-arrive-in-3-5-years-18000-ai-servers-to-go-live-soon-says-ashwini-vaishnaw/articleshow/117958923.
- 77 Abhishek Rakhonde, "India's Own AI Platform Already Has 18000+ GPUs: 800% More Than DeepSeek," Trak.in, February 1, 2025, https://trak.in/stories/indias-own-ai-platform-already-has-18000-gpus-800-morethan-deepseek/.
- 78 Subhrojit Mallick and Himanshi Lohchab, "Biden Admin's Cap on GPU Exports May Hit India's AI Ambitions," Economic Times, January 16, 2025, https://economictimes.indiatimes.com/tech/technology/ biden-admins-cap-on-gpu-exports-may-hit-indias-ai-ambitions/articleshow/117245296.cms?from=mdr.
- 79 See Billy Perrigo, "Exclusive: New Research Finds Stark Global Divide in Ownership of Powerful AI Chips," Time, August 28, 2024, https://time.com/7015330/ai-chips-us-china-ownership-research/.
- 80 "Union Budget 2025: Centre of Excellence for AI to Be Set Up With Rs. 500 Crore Outlay," Economic Times, February 1, 2025, https://economictimes.indiatimes.com/tech/technology/union-budget-2025-centre-ofexcellence-for-ai-to-be-set-up-with-rs-500-crore-outlay/articleshow/117819492.cms?from=mdr#.
- 81 Ibid. Also see "Rs. 990 Crore Approved by Government to Establish Three AI Centres of Excellence."

- 82 Sarthak Pradhan and Pranay Kotasthane, "Policy Tweaks Can Push Private Sector R&D."
- 83 "Nokia and Indian Institute of Science Form Strategic Partnership to Research How 6G Can Meet India's Needs," press release, Nokia, Feb 23, 2024, https://www.nokia.com/about-us/news/releases/2024/02/23/ nokia-and-indian-institute-of-science-form-strategic-partnership-to-research-how-6g-can-meet-indias-needs/#.
- 84 See Report of the Artificial Intelligence Task Force, (Ministry of Commerce and Industry, Government of India, March 20, 2018), https://dpiit.gov.in/sites/default/files/Report_of_Task_Force_on_ ArtificialIntelligence 20March2018 2.pdf; and "National Strategy for Artificial Intelligence: #AI for All," (NITI Aayog, Government of India, June 2018), https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf.
- 85 See Vaishnavi Chandrashekhar, "India Takes Out Giant Nationwide Subscription to 13,000 Journals," Science, December 2, 2024, https://www.science.org/content/article/india-takes-out-giant-nationwide-subscription-13-000-journals. Also see Ministry of Science and Technology, Government of India, "Anusandhan National Research Foundation Announces the Launch of the 'Partnerships for Accelerated Innovation and Research' (PAIR) Programme to Transform Research and Innovation in Indian Universities," press release, Press Information Bureau, November 14, 2024, https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2073282.
- 86 See Amit Kumar, "National AI Policy/Strategy of India and China: A Comparative Analysis," 13; Also see "The AI Landscape in China: Segmentation of the Top AI Companies on the Market," Daxue Consulting, July 15, 2019, https://daxueconsulting.com/ai-landscape-china/.
- 87 Amit Kumar, "National AI Policy/Strategy of India and China: A Comparative Analysis."
- 88 Rishikesha T. Krishnan, "India Is an R&D Hub for MNCs. Will Global Protectionism Play Spoilsport?," Founding Fuel, November 2, 2019, https://www.foundingfuel.com/article/india-is-an-rd-hub-for-mncs-willglobal-protectionism-play-spoilsport/.
- 89 Zinnov-NASSCOM India GCC Landscape Report: The 5-Year Journey, (Zinnov, September 11, 2024), https:// zinnov.com/centers-of-excellence/zinnov-nasscom-india-gcc-landscape-report-the-5-year-journey-report/.
- 90 Cameron F. Kerry, Joshua P. Meltzer, and Andrea Renda, "AI Cooperation on the Ground: AI Research and Development on a Global Scale," Brookings Institution, November 4, 2022, https://www.brookings.edu/ articles/ai-cooperation-on-the-ground-ai-research-and-development-on-a-global-scale/.
- 91 See "Declaration of the United States of America and the United Kingdom of Great Britain and Northern Ireland on Cooperation in Artificial Intelligence Research and Development: A Shared Vision for Driving Technological Breakthroughs in Artificial Intelligence," press release, U.S. Department of State, September 25, 2020, https://www.state.gov/declaration-of-the-united-states-of-america-and-the-united-kingdom-of-greatbritain-and-northern-ireland-on-cooperation-in-artificial-intelligence-research-and-development-a-sharedvision-for-driving/; "Fact Sheet: Quad Leaders' Summit," The White House, September 24, 2021, https:// www.whitehouse.gov/briefing-room/statements-releases/2021/09/24/fact-sheet-quad-leaders-summit/.
- 92 See Abhijeet Kumar, "DeepSeek AI to Fighter Jets: How China Is One-Upping Western Dominance," Business Standard, January 28, 2025, https://www.business-standard.com/external-affairs-defence-security/news/chinadeepseek-ai-stealth-jets-global-innovation-leadership-125012801140_1.html; Sylvie Zhuangin and Zhang Tongin, "What Does Rise of AI Firm DeepSeek Mean for US-China Tech War, and The Race for Talent?," South China Morning Post, January 27, 2025, https://www.scmp.com/news/china/diplomacy/article/3296477/ what-does-rise-ai-firm-deepseek-mean-us-china-tech-war-and-race-talent.
- 93 "AI for All—Bridging the Global AI Divide," YouTube video, posted by "Digital India," January 12, 2024, https://youtu.be/XDuUPTpuEGI?si=k-bGctQTzp0IAKmX.
- 94 Upasana Sharma and Shreya Ramann, "AI for All, Beyond the Global North: India's Opportunity?," Carnegie India, November 27, 2023, https://carnegieendowment.org/posts/2023/11/ai-for-all-beyond-the-globalnorth-indias-opportunity?lang=en.
- 95 Swathi Moorthy, "Open Cloud Compute: A UPI Moment for AI?," Economic Times, December 3, 2024, https://economictimes.indiatimes.com/tech/artificial-intelligence/open-cloud-compute-a-upi-moment-for-ai/ articleshow/115910167.cms?from=mdr.
- 96 Cameron F. Kerry et al., "Strengthening International Cooperation on AI," Brookings Institution, October 25, 2021, https://www.brookings.edu/articles/strengthening-international-cooperation-on-ai/.

Carnegie India

Carnegie India, founded in 2016 and based in New Delhi, is part of a robust global network that includes over 200 scholars. The center focuses primarily on three interrelated topics: technology and society, political economy, and security studies. Led by Indian experts with decades of policy experience, Carnegie India engages with governments, policymakers, academics, students, industries, practitioners, and civil society to provide insightful and fresh analysis on India's pressing challenges and its rising global role.

Carnegie Endowment for International Peace

In a complex, changing, and increasingly contested world, the Carnegie Endowment generates strategic ideas, supports diplomacy, and trains the next generation of international scholar-practitioners to help countries and institutions take on the most difficult global problems and advance peace. With a global network of more than 170 scholars across twenty countries, Carnegie is renowned for its independent analysis of major global problems and understanding of regional contexts.

