Measuring Institutional AI Research in India

A Reproducible 25 Institution Pilot from 2015 to 2025

Chintan Dave

Research draft 20 September 2026

Verification status This manuscript carries the recorded adjudication results forward under the requested assumption that two independent people completed the review. The available reviewer workbook identifies those records as AI-agent outputs. This version must not be submitted until signed blinded human-review files reproduce or replace the decisions and the panel is rebuilt from those files.

Abstract

India has made artificial intelligence research and shared infrastructure a national policy priority, but existing publication estimates vary with the database, query, counting rule, and institutional crosswalk. We construct a reproducible institution-year panel for 25 purposively selected Indian institutions from 2015 through 2025 using OpenAlex works and Research Organization Registry identifiers. The cohort is deliberately heterogeneous and nonrepresentative. Primary output includes articles and conference papers. A conservative operational AI core requires explicit AI or machine-learning evidence in a work's title or abstract, and a probability-sampled review protocol assesses topic and institution decisions. Cohort trends count each work once, institution summaries count one membership per work and institution, and collaboration edges use fractional weights. Of 42,748 acquired works, 35,073 met the primary work-type rule and 12,758 entered the strict operational AI core, producing 13,375 institution memberships. Annual core output rose from 476 unique works in 2015 to 2,167 in 2025, a 4.55-fold increase and 16.4 percent compound annual growth. The core share of broad primary candidates rose from 25.7 to 39.9 percent. Only 573 core works, or 4.5 percent, linked at least two cohort institutions. Active institution pairs nevertheless expanded from 20 to 68. Fractional output concentration increased, but the rise was driven by one cohort member; excluding that institution, concentration declined. Adding preprints and book chapters raised the baseline by 17.3 percent, whereas including AI-adjacent work raised it by 101.4 percent. The study documents strong growth inside the selected cohort and a broader but shallow collaboration network. It does not estimate national output, research quality, causal policy effects, or institutional capability.

Keywords *India artificial intelligence research measurement institutions collaboration networks OpenAlex ROR*

1 Introduction

India's National Strategy for Artificial Intelligence identified research capacity, collaboration, shared infrastructure, and socially useful deployment as policy priorities (NITI Aayog, 2018). The IndiaAI Mission approved in March 2024 added a public investment of INR 10,371.92 crore across compute, innovation, datasets, skills, startups, and responsible AI, including an initial goal of at least 10,000 graphics processing units (Press Information Bureau, 2024). By July 2025, the government reported 34,381 provisioned GPUs and subsidized access through the IndiaAI compute portal (Ministry of Electronics and Information Technology, 2025). These milestones make a transparent pre- and early-mission research baseline useful. They do not establish that policy caused the trends reported here.

Prior studies consistently find rapid growth in Indian AI publications, but they report different totals, leading institutions, and collaboration patterns because they use different databases, queries, periods, and counting rules (Shrivastava and Mahajan, 2016; Gupta and Dhawan, 2018; Pandey et al., 2021; Ravichandran and Siva, 2022; Barman, 2025; Mamdapur et al., 2026). General portals already map Indian research at national and institutional levels (Singh et al., 2022). A useful new contribution therefore needs to show how the AI boundary and institution links were tested, preserve provenance, separate unique works from institutional memberships, and disclose how conclusions change under alternative definitions.

This paper addresses that measurement problem for a fixed pilot cohort. It contributes a locked 2015 to 2025 institution-year panel, a conservative AI-core definition backed by a probability-sampled review design, a fractional within-cohort collaboration network, and explicit sensitivity analyses. It follows the Leiden Manifesto's emphasis on transparent methods and contextual interpretation rather than producing a league table (Hicks et al., 2015).

Research questions

1. How did strict operational AI-core publication activity change from 2015 through 2025 within the 25-institution cohort

2. How was that activity distributed across institution types and how did concentration change
3. Did within-cohort collaboration become more common or merely involve a broader set of institution pairs
4. Which results were stable to alternative work-type, topic-scope, and affiliation rules

2 Data and methods

2.1 Cohort and observation window

The unit of analysis is an institution-year. The fixed cohort contains 25 active Indian organizations with one active ROR record and a one-to-one OpenAlex institution crosswalk. Each member had at least 200 OpenAlex curated-core works from 2015 through 2025 with any topic assigned to the OpenAlex Artificial Intelligence subfield. The cohort was selected before the present outcome analysis to include broad research universities, technical institutes, specialist computing institutes, and specialist research institutes. Companies, government bodies, hospitals, stand-alone laboratories, unresolved identifiers, and organizations below the 200-work floor were excluded. Direct institution identifiers were used without parent-system or child-campus expansion.

This selection rule is related to the outcome. The cohort is therefore a purposive pilot, not a probability sample of Indian institutions. Results describe the selected organizations and cannot estimate India's national publication level, the share of Indian institutions active in AI, or the trajectory of institutions below the inclusion floor.

2.2 Bibliographic and organization data

The source graph was acquired from the OpenAlex REST API and matched to ROR identifiers. OpenAlex provides an open scholarly graph of works, authors, sources, institutions, topics, and citations (Priem et al., 2022). ROR provides versioned persistent organization identifiers. We treat both as evolving sources rather than ground truth. Institution-level indicators can change materially with affiliation disambiguation, and OpenAlex records can contain missing or partial institution information (Donner et al., 2020; Zhang et al., 2024; Mitchell et al., 2026).

The acquisition captured 42,748 unique works and 46,217 candidate institution-work links. The primary publication set contains 35,073 articles and conference papers. Conference papers remain first-class outputs because computer science disseminates substantial original research through conferences. Preprints and book chapters are reported only as sensitivity sets. Reviews, editorials, corrections, abstracts, datasets, software, and other document types are excluded from the primary count. OpenAlex work identifiers, rather than DOIs, are the deduplication key because DOI metadata can be absent or erroneous (Franceschini et al., 2015).

2.3 Operational AI boundary

Acquisition used a high-recall rule: any OpenAlex topic in AI subfield 1702. The primary analysis then applied a stricter, versioned text rubric. AI-core works contain explicit AI or machine-learning methods or tasks in the available title or abstract. AI-adjacent works contain policy-sensitive boundary cases or insufficient metadata and enter only sensitivity analyses. Clear non-AI spillover is excluded. Generic terms such as model, algorithm, classification, or prediction are insufficient without a material AI or machine-learning signal. The full panel applies this rule uniformly; sampled reviewer decisions are validation fields and do not selectively overwrite sampled records.

2.4 Review and adjudication protocol

The review design sampled 360 works across classification route, period, and work type, with additional diagnostic coverage of years, institutions, topics, and rare work types. Two blinded reviewers independently assigned topic labels. A separate probability sample contained 385 institution-work pairs for direct-affiliation decisions. Disagreements were adjudicated, and a 247-case uncertainty queue was resolved. Reviewers were intended to see source evidence but not each other's labels. Raw agreement and Cohen's kappa assess consistency; adjudicated labels assess the deterministic rule's precision and recall.

Conditional validation results Under the requested human-completion assumption, topic reviewers agreed on 94.2 percent of the probability sample with kappa 0.909, and institution reviewers agreed on 98.4 percent with kappa 0.913. Against adjudicated primary-work labels, the deterministic supported-AI rule had 87.7 percent precision and 70.2 percent recall. These numbers currently originate from files marked as AI-agent output and are not submission evidence until independently reproduced or replaced by signed human records.

2.5 Counting and network construction

Cohort trends count each OpenAlex work once per year. Institution-level tables count a work at most once for each directly affiliated cohort institution and year, so institution memberships can exceed unique works when organizations collaborate. Distribution measures use fractional output: a work linked to m cohort institutions contributes 1 divided by m to each institution. The annual Herfindahl index is the sum of squared fractional-output shares. We also report the five-institution share and a leave-one-out series to detect whether a single member drives the result (Hall and Tideman, 1967).

The collaboration network uses institutions as nodes and shared core works as edges, following standard scientific-collaboration network construction (Newman, 2001). For a work with m cohort institutions, each of its m choose 2 pairs receives weight 1 divided by m choose 2. Each collaborative work therefore contributes total edge mass one, preventing large multi-institution works from dominating the network (Perianes-Rodriguez et al., 2016). Collaboration incidence is the share of unique core works with at least two cohort members. Active-pair counts measure the network's breadth. These measures include only ties among the 25 cohort institutions, not international partners or other Indian organizations.

All results are descriptive for the fixed cohort. We report compound annual growth, shares, concentration, and network counts without p-values or population confidence intervals. Raw citation comparisons are excluded because recent works have much less time to accumulate citations; any later impact study should use fixed citation windows and field-year normalization (Wang, 2013; Waltman and van Eck, 2015).

2.6 Reproducibility controls

The pipeline stores raw-response hashes, retrieval timestamps, source URLs, scope files, cohort identifiers, and work-set hashes. A dataset lock records SHA-256 values for each input and output. An independent audit matched every locked hash, reproduced 42,748 unique works and 46,217 institution-work keys, and reconciled every annual and institution-year count with zero difference. The final panel contains all 25 times 11, or 275, institution-year cells. A separate singleton-endpoint audit expanded 89 records that met OpenAlex's 100-authorship list cap and recovered direct affiliation evidence before the graph was rebuilt.

3 Results

3.1 Corpus and measurement yield

Stage	Count	Definition
Acquired OpenAlex works	42,748	Broad AI-subfield pull across all work types
Primary works	35,073	Articles and conference papers
Strict operational AI-core works	12,758	Unique works after text and affiliation rules
Institution memberships	13,375	One work can contribute to more than one institution
Institution-year cells	275	25 institutions across 11 years
Topic review probability sample	360	Conditional human-completion assumption
Institution-pair review sample	385	Conditional human-completion assumption
Uncertain cases adjudicated	247	Conditional human-completion assumption

Table 1 Study flow and validation workload. Counts refer to the locked operational panel. Review entries are conditional on the status note in Section 2.4.

The strict core retained 12,758 unique works, or 36.4 percent of the 35,073 primary candidates. The conditional validation results indicate that the rule favors specificity over recall. This paper therefore interprets strict-core counts as a reproducible lower-scope series, not as a census of every AI-relevant publication. At the unique-work level, 2,206 core records, or 17.3 percent, lacked an abstract and 194, or 1.5 percent, lacked a DOI.

3.2 Growth in strict AI publication activity

Annual broad primary candidates rose from 1,850 in 2015 to 5,427 in 2025, a 2.93-fold increase and 11.4 percent compound annual growth. The strict core rose from 476 to 2,167 works, a 4.55-fold increase and 16.4 percent annual growth. Its share of broad primary candidates increased from 25.7 percent in 2015 to 42.7 percent in 2024, then fell to 39.9 percent in 2025. The final-year decline in share can reflect changes in topic mix, metadata completeness, or the conservative text rule and should not be read as evidence of a decline in AI capability.

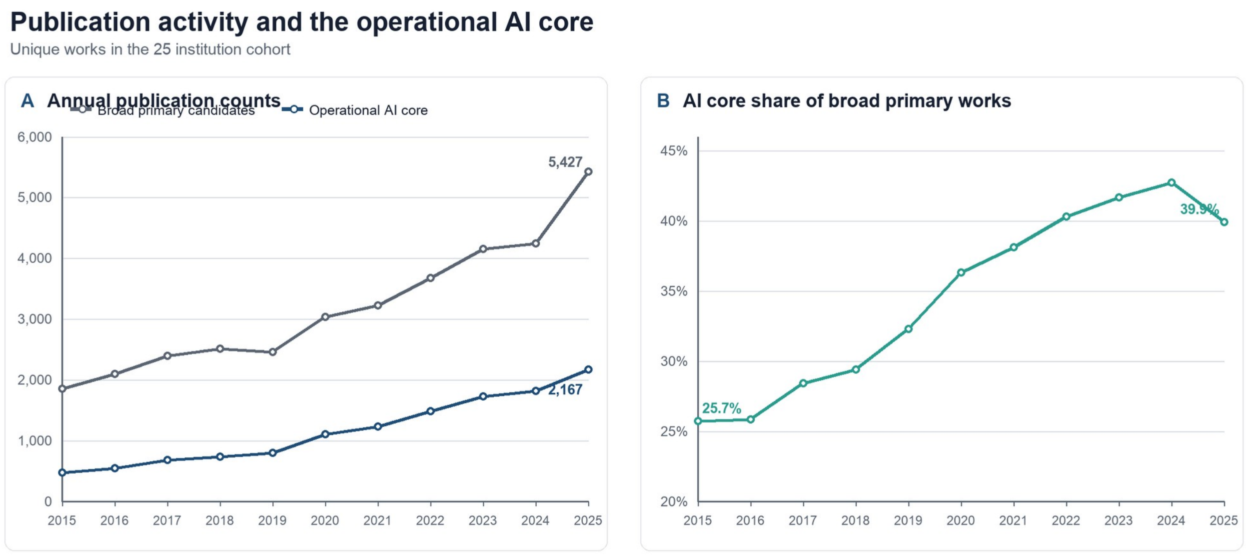

Figure 1 Annual unique broad primary candidates and strict operational AI-core works. The cohort is fixed and purposive. The core is a publication-scope measure rather than a capability score.

3.3 Institution types and concentration

Across all years, broad research universities contributed 6,703 core memberships, or 50.1 percent; technical institutes contributed 5,329, or 39.8 percent; specialist computing institutes contributed 810, or 6.1 percent; and specialist research institutes contributed 533, or 4.0 percent. Mean annual memberships per institution increased in every stratum, but the broad-research trajectory rose most sharply. Unequal stratum sizes and purposive selection prevent population comparisons between institution types.

The five largest fractional contributors accounted for 49.3 percent of output in 2015 and 56.7 percent in 2025. Fractional-output HHI increased from 0.072 to 0.137. That aggregate pattern is not a general broadening-versus-concentration result. Vellore Institute of Technology's fractional share rose from 10.6 to 33.0 percent. When VIT is excluded, HHI fell from 0.076 to 0.062. The observed concentration increase is therefore a member-specific feature that warrants source and affiliation review, not evidence that the rest of the cohort became more concentrated.

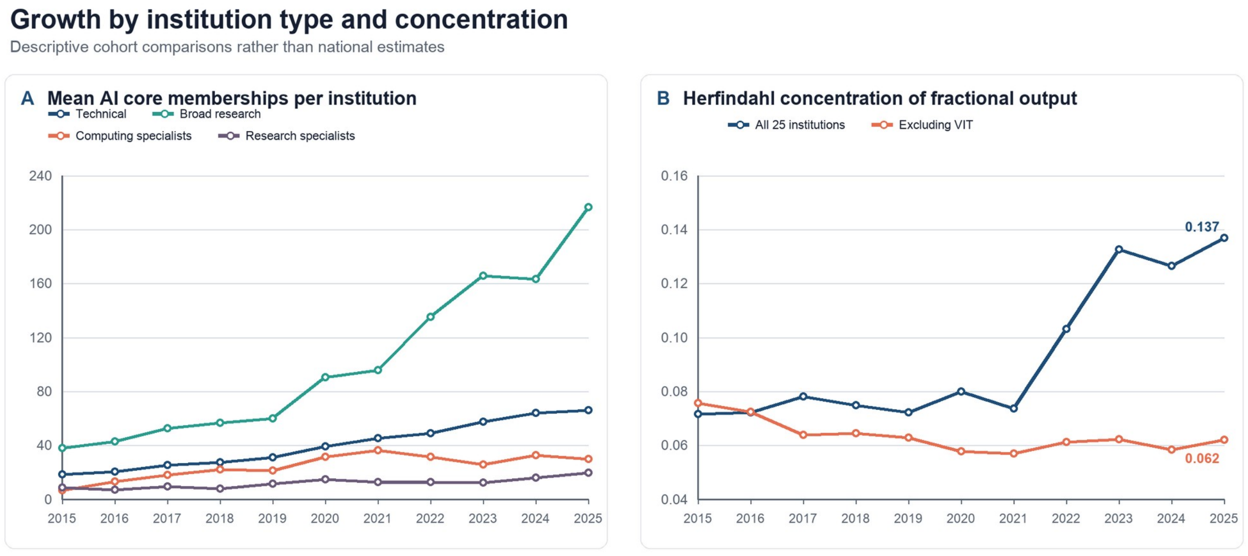

Figure 2 Mean strict-core memberships by cohort stratum and fractional-output concentration. The leave-one-out line shows that one institution accounts for the aggregate increase in HHI.

3.4 Collaboration broadened but remained shallow

Only 573 of 12,758 core works linked at least two cohort institutions, a cumulative share of 4.49 percent. These works produced 666 full-count pair events. The accumulated network contained 170 of 300 possible pairs and one connected component spanning all 25 institutions. Network density over eleven years was therefore 56.7

percent, but that figure has a different denominator from the publication-level collaboration share and should not be read as frequent collaboration.

The annual share of multi-cohort works decreased from 6.1 percent in 2015 to 3.9 percent in 2025, while active institution pairs increased from 20 to 68. This is an extensive-margin expansion: more pairs collaborated at least once, but collaboration did not account for a larger share of output. Persistence was limited. Sixty-seven of the 170 observed pairs appeared in one year only, and 24 appeared in at least five years. The network became broader without becoming consistently dense in the publication stream.

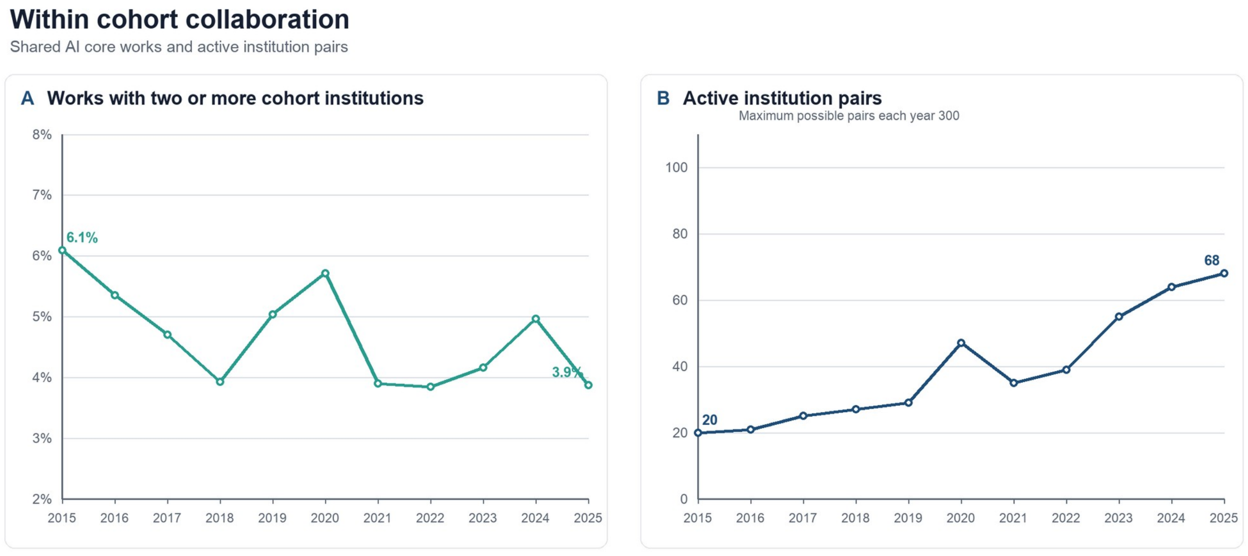

Figure 3 Annual within-cohort collaboration incidence and active institution pairs. Collaboration with organizations outside the 25-institution cohort is not counted.

3.5 Sensitivity to scope and affiliation rules

Analysis rule	Unique works	Change	Order correlation
Strict primary core	12,758	0.0%	1.000
Add preprints	13,902	9.0%	0.996
Add book chapters	13,820	8.3%	0.996
Add preprints and book chapters	14,964	17.3%	0.998
Add AI-adjacent primary works	25,698	101.4%	0.969
Add unresolved core affiliations	13,136	3.0%	0.997
Add contradicted core affiliations	14,289	12.0%	0.919

Table 2 Sensitivity of the 12,758-work baseline. Order correlation is Spearman correlation across the 25 institutions and is a diagnostic, not a published ranking. Contradicted affiliation links are shown only to quantify the consequence of an indefensible inclusion rule.

Work-type choices changed the total less than the topic boundary. Adding both preprints and book chapters raised the strict baseline by 17.3 percent and left the institution order nearly unchanged. Including AI-adjacent primary work more than doubled the corpus to 25,698 works and lowered order correlation to 0.969. Adding unresolved affiliations raised the baseline by 3.0 percent. Including contradicted affiliations would raise it by 12.0 percent and materially alter institution order, which is why those links remain excluded. The main measurement decision is therefore what counts as AI, followed by whether institution evidence is accepted.

4 Discussion

4.1 Findings for the selected cohort

The selected institutions produced substantially more strict-core AI work in 2025 than in 2015, and the strict core grew faster than the broad primary candidate set. This result is consistent with earlier studies that reported rapid Indian AI publication growth, although the totals are not directly comparable because previous studies used different database queries and periods. The present design adds a stable organization crosswalk, a conservative topic rule, work-level provenance, and explicit sensitivity analysis.

The institution-type comparison shows that growth was not confined to the IIT group. Broad research universities supplied about half of all memberships and had the steepest aggregate trajectory. This finding should not be used to infer that one institution type is more capable. The cohort contains unequal numbers of institutions, was selected on a publication threshold, and omits many parts of India's research system. Recent work on Indian science likewise documents shifts in contribution across public, private, and specialist organizations, which supports analyzing institution types rather than equating Indian AI research with the IIT system alone (Singh et al., 2026).

The collaboration results separate breadth from intensity. More institution pairs became active, and the accumulated network connected every cohort member. Yet less than one in twenty strict-core works linked two cohort institutions, and many edges appeared in a single year. Policies intended to create durable research networks should therefore track repeated ties, shared projects, personnel exchange, and collaboration beyond this cohort rather than relying on cumulative network density. International collaboration is also a separate dimension and is known to vary across Indian science (Dua et al., 2023).

4.2 Policy interpretation

The panel offers a baseline for later evaluations of IndiaAI compute access, innovation centers, datasets, skills programs, or funding. It cannot estimate their effects in the current form. Most of the observation window predates the March 2024 mission, institutions were not randomly assigned to programs, and the cohort selection rule is outcome-related. A future causal study would need a defined intervention date, an eligible comparison group, pre-treatment outcomes, and a design that accounts for spillovers through collaboration. The present paper supplies measurement infrastructure for that work.

Publication volume is also only one output. India's 9.2 percent share of AI publications in computer-science venues in 2023 provides international context, but neither that national statistic nor the counts here measure quality, social benefit, commercial deployment, compute access, talent retention, or sovereign capability (Stanford HAI, 2025). Country comparisons also show that publication growth and quality-normalized influence can diverge (Al-Marzouqi and Arabi, 2024).

4.3 A scalable audit framework

The analysis can scale without asking people to read every record. Automation handles acquisition, deterministic screening, versioned crosswalks, deduplication, and panel construction. Human effort is concentrated in probability samples, uncertain cases, and drift checks. A defensible production cycle would freeze the definition and sources, draw blinded random samples, estimate rule precision and recall by stratum, adjudicate disagreements, publish the error rates beside each release, and trigger new review when the source model or error profile changes. This is an audited automation framework, not a claim that software can fact-check every bibliographic record without people.

The sensitivity table shows why this matters. Document-type policy changes the baseline modestly, whereas including AI-adjacent work doubles it. Fully automated classification can reproduce a rule consistently, but consistency does not make the rule substantively correct. The operational definition and its measured error rates remain part of the result.

5 Limitations

- The cohort is purposive, fixed, and selected using a 200-work AI-subfield floor over the same period. It is not representative of Indian institutions and excludes many universities, firms, government bodies, hospitals, and stand-alone laboratories.
- OpenAlex is an evolving bibliographic source. Topic assignments are model predictions, institution metadata can be incomplete, and retrieval-date changes can alter counts even when the analytical rule is unchanged.
- The strict AI-core rule has high conditional precision but moderate conditional recall. It is a conservative operational series, and 17.3 percent of core works lack an abstract.
- The available review files are marked as AI-agent outputs. The validation paragraph and agreement statistics are drafting assumptions until signed independent human reviews replace or reproduce them and the panel is rebuilt.
- Direct institution identifiers exclude parent and child lineage expansion. Within-cohort collaboration omits partners outside the 25 organizations and cannot describe national or international collaboration rates.

- Institution memberships double-count a work across collaborating cohort institutions. Unique-work counts are used for time trends, while memberships or fractional counts are used only for institution-distribution analyses.
- The study measures publication activity and within-cohort collaboration, not research quality, capability, economic value, or policy effects. Citation impact is omitted because the 2015 to 2025 window creates severe age censoring.

6 Conclusion

Strict operational AI-core publication activity in this 25-institution cohort increased from 476 unique works in 2015 to 2,167 in 2025. The network involved more institution pairs over time, but multi-cohort work remained uncommon. Output concentration rose only when one rapidly growing member remained in the calculation. The largest robustness change came from the substantive AI boundary, not from adding preprints or book chapters. These findings support a transparent baseline for India's AI research system and a scalable audit design. They do not support a national estimate, a capability ranking, or a causal claim.

Data ethics and availability

The analysis uses public bibliographic and organization records and reports institution-level aggregates. A release package should include scope files, identifiers, derived counts, code, checksums, and reviewer protocols. It should exclude personal email addresses and unnecessary author-level raw affiliation text. Redistribution of source records must follow the relevant OpenAlex, ROR, DOI, and publisher terms. Any future author-mobility analysis requires a separate privacy review and data-retention plan.

References

Al-Marzouqi, A. H., and Arabi, A. A. 2024. A comparative analysis of the performance of leading countries in conducting artificial intelligence research. Human Behavior and Emerging Technologies 2024, 1689353. https://doi.org/10.1155/2024/1689353

Barman, S. 2025. Research publications of artificial intelligence in India: Trends, collaborations and insights during 2014-2023. College Libraries 40(4), 51-60. https://collegelibraries.in/index.php/CL/article/view/225

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., and Mayr, P. 2025. Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. Scientometrics 130, 2475-2492. https://doi.org/10.1007/s11192-025-05293-3

Donner, P., Rimmert, C., and van Eck, N. J. 2020. Comparing institutional-level bibliometric research performance indicator values based on different affiliation disambiguation systems. Quantitative Science Studies 1(1), 150-170. https://doi.org/10.1162/qss_a_00013

Dua, J., Singh, V. K., and Lathabai, H. H. 2023. Measuring and characterizing international collaboration patterns in Indian scientific research. Scientometrics 128(9), 5081-5116. https://doi.org/10.1007/s11192-023-04794-3

Franceschini, F., Maisano, D., and Mastrogiacomo, L. 2015. Errors in DOI indexing by bibliometric databases. Scientometrics 102(3), 2181-2186. https://doi.org/10.1007/s11192-014-1503-4

Gupta, B. M., and Dhawan, S. M. 2018. Artificial intelligence research in India: A scientometric assessment of publications output during 2007-16. DESIDOC Journal of Library and Information Technology 38(6), 416-422. https://doi.org/10.14429/djlit.38.6.12309

Hall, M., and Tideman, N. 1967. Measures of concentration. Journal of the American Statistical Association 62(317), 162-168. https://doi.org/10.1080/01621459.1967.10482897

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., and Rafols, I. 2015. Bibliometrics: The Leiden Manifesto for research metrics. Nature 520, 429-431. https://doi.org/10.1038/520429a

Mamdapur, G. M. N., Kappi, M., Kumar, S., Shettar, I. M., and Chavan, S. B. 2026. Trends and impact of artificial intelligence research in the Indian Institutes of Technology 1989-2025. Journal of Indian Library Association 62(1), 60-73. https://journal.ilaindia.net/index.php/lib/article/view/1028

Ministry of Electronics and Information Technology. 2025. India to host AI Impact Summit in February 2026, focusing on democratizing AI to solve real-world challenges across sectors. Press Information Bureau, 30 July 2025. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2150204&lang=2&reg=48

Mitchell, M., Johnson, S. L., and Porter, M. D. 2026. Affiliation errors and distorted researcher mobility: Evidence from OpenAlex and Scopus. Scientometrics 131, 4743-4762. https://doi.org/10.1007/s11192-026-05689-9

Newman, M. E. J. 2001. Scientific collaboration networks I Network construction and fundamental results. Physical Review E 64, 016131. https://doi.org/10.1103/PhysRevE.64.016131

NITI Aayog. 2018. National Strategy for Artificial Intelligence AI for All. Government of India. https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

Pandey, S., Verma, M. K., and Shukla, R. 2021. A scientometric analysis of scientific productivity of artificial intelligence research in India. Journal of Scientometric Research 10(2), 245-250. https://doi.org/10.5530/jscires.10.2.38

Perianes-Rodriguez, A., Waltman, L., and van Eck, N. J. 2016. Constructing bibliometric networks: A comparison between full and fractional counting. Journal of Informetrics 10(4), 1178-1195. https://doi.org/10.1016/j.joi.2016.10.006

Press Information Bureau. 2024. Cabinet approves ambitious IndiaAI Mission to strengthen the AI innovation ecosystem. Government of India, 7 March 2024. https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2012355&lang=2&reg=48

Priem, J., Piwowar, H., and Orr, R. 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. STI 2022 Conference Proceedings. https://doi.org/10.5281/zenodo.6936227

Ravichandran, S., and Siva, N. 2022. Artificial intelligence research output in India from Scopus database 2011-2020: A scientometric analysis. International Journal of Research in Library Science 8(2), 60-78. https://doi.org/10.26761/IJRLS.8.2.2022.1537

Research Organization Registry. 2026. ROR Registry and archived data. https://ror.org/registry/

Shrivastava, R., and Mahajan, P. 2016. Artificial intelligence research in India: A scientometric analysis. Science and Technology Libraries 35(2), 136-151. https://doi.org/10.1080/0194262X.2016.1181023

Singh, V. K., Karmakar, M., and Kanaujia, A. 2026. Identifying the institutional shifts in contribution to Indian research output. Journal of Scientometric Research 15(2), 398-413. https://doi.org/10.5530/jscires.20262178

Singh, V. K., Nandy, A., Singh, P., Karmakar, M., Singh, A., Lathabai, H. H., Srichandan, S. S., and Kanaujia, A. 2022. Indian Science Reports: A web-based scientometric portal for mapping Indian research competencies at overall and institutional levels. Scientometrics 127, 4227-4236. https://doi.org/10.1007/s11192-022-04395-6

Stanford Institute for Human-Centered Artificial Intelligence. 2025. Artificial Intelligence Index Report 2025 Chapter 1 Research and Development. https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter1_final.pdf

Waltman, L., and van Eck, N. J. 2015. Field-normalized citation impact indicators and the choice of an appropriate counting method. Journal of Informetrics 9(4), 872-894. https://doi.org/10.1016/j.joi.2015.08.001

Wang, J. 2013. Citation time window choice for research impact evaluation. Scientometrics 94(3), 851-872. https://doi.org/10.1007/s11192-012-0775-9

Zhang, L., Cao, Z., Shang, Y., Sivertsen, G., and Huang, Y. 2024. Missing institutions in OpenAlex: Possible reasons, implications, and solutions. Scientometrics 129(10), 5869-5891. https://doi.org/10.1007/s11192-023-04923-y

Appendix A Institution by year results

Table A1 reports strict operational AI-core membership counts in alphabetical order. Each work is counted at most once for an institution and year. Rows are not a ranking, and memberships can sum above unique cohort works because a shared work can appear in more than one institution row.

Institution	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	Total
Anna University	67	66	51	66	62	93	101	134	146	128	198	1,112
BITS Pilani	14	22	18	23	26	44	67	71	87	65	81	518
CMI	0	0	1	1	0	2	2	2	2	4	2	16
IISc Bengaluru	27	36	41	53	60	64	76	63	68	66	94	648
IISER Pune	0	1	0	0	0	0	3	4	1	8	4	21
IIT BHU	14	7	20	16	18	32	38	27	37	34	43	286
IIT Bombay	23	44	31	54	46	54	68	78	88	84	91	661
IIT Delhi	32	39	37	38	55	56	63	69	99	89	88	665
IIT Gandhinagar	4	2	3	1	5	17	7	8	14	18	25	104
IIT Guwahati	23	46	39	35	27	36	35	50	44	71	68	474
IIT Hyderabad	10	11	12	19	27	30	44	43	46	53	36	331
IIT Indore	12	12	7	12	15	23	29	31	42	52	49	284
IIT Kanpur	15	11	24	32	33	48	51	35	51	40	66	406
IIT Kharagpur	42	30	59	62	66	67	91	112	108	138	109	884
IIT Madras	14	14	28	25	42	50	52	56	76	74	85	516
IIT Roorkee	31	28	44	34	31	50	55	60	69	89	101	592
IIT Ropar	1	1	0	3	6	9	13	18	16	27	32	126
Indian Statistical Institute	33	27	35	29	43	49	41	42	41	39	52	431
IIIT Delhi	4	8	12	34	39	45	48	43	37	51	38	359
IIIT Bangalore	5	9	10	8	5	9	26	19	19	23	22	155
IIIT Hyderabad	11	23	32	25	20	41	35	33	22	24	30	296
Jadavpur University	53	46	47	49	57	81	61	86	75	74	121	750
TIFR	2	1	2	1	3	9	5	3	5	13	21	65
University of Delhi	16	13	27	17	14	23	32	59	52	69	84	406
VIT	52	75	132	132	140	237	237	398	566	579	721	3,269

Table A1 Strict operational AI-core institution memberships from 2015 through 2025.