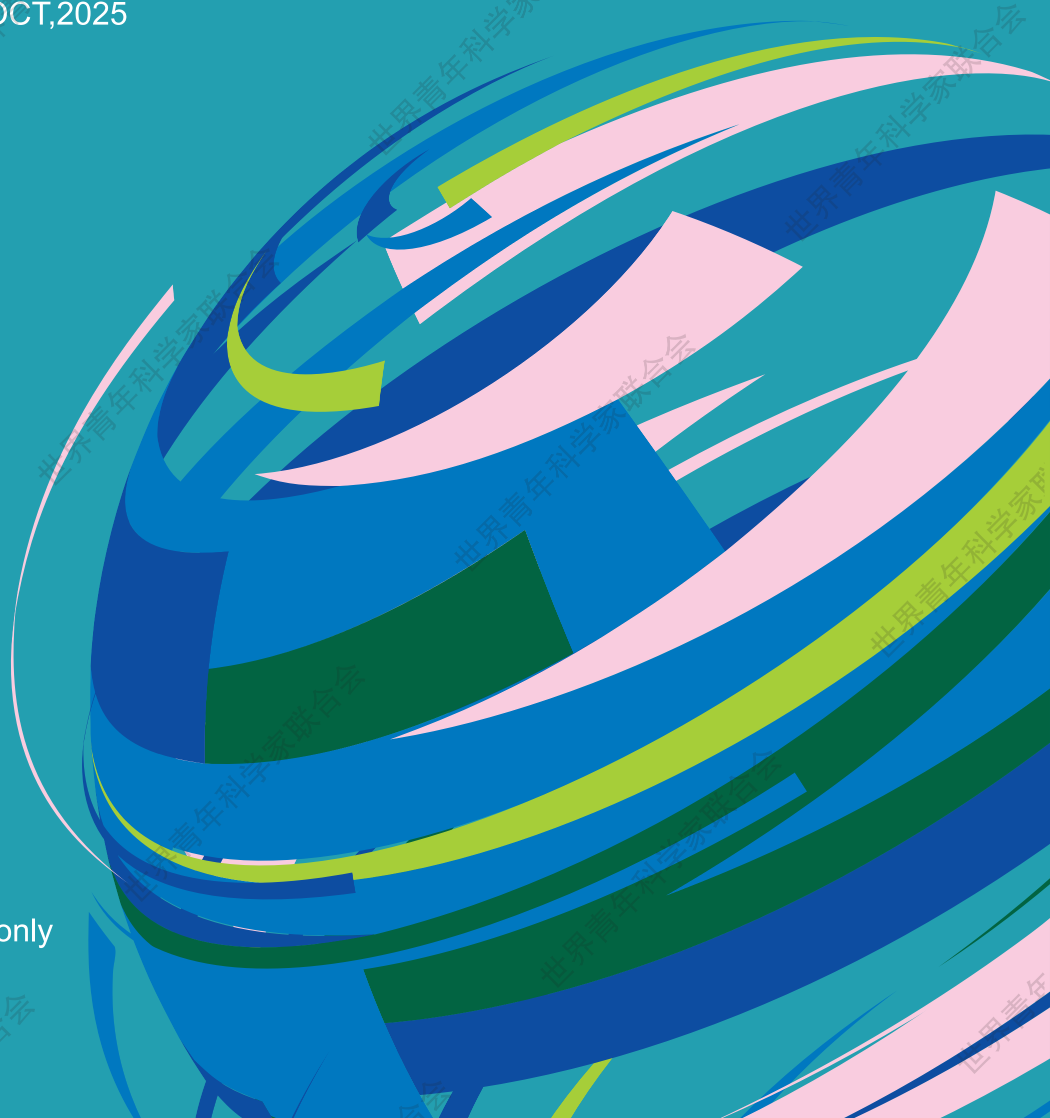
Report on the Distribution and Mobility of Global Young AI Scientists 2025

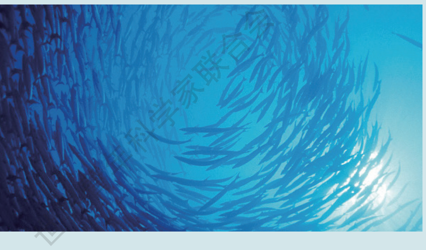
WYSS
2025

世界青年科学家联合会
2025 年 10 月

World Association of Young Scientists
OCT,2025



Internal report for reference only



Editorial Team

Report Planning: Yang Daiqing, Gao Jiping, Wang Yunhong

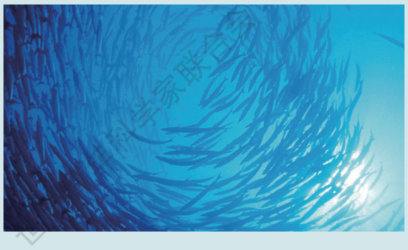
Authors: Li Bohui, Wang Yunhong, Ye Xianting

Reviewers: Wang Yunhong, Zheng Chuhua, Zhang Aixia

Technical Team: Tian Ruiqiang, Ma Chen, Yang Xiao, Xu Xiaoyang

Expert Team: Liu Xiaojuan, Wuyun Qiqige, Zhong Zurong, Zuo Xiaoli

Compiling Organization: Institute of Scientific and Technical Information of China, Scientometrics & Evaluation Research Center



Introduction

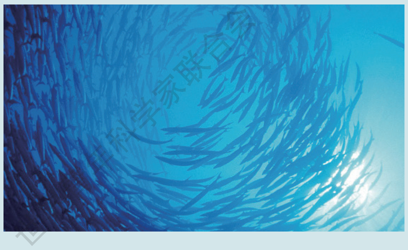
With groundbreaking advances in cutting-edge fields like deep learning and large-scale models, artificial intelligence (AI) is reshaping research paradigms and industrial landscapes worldwide. Young scientists are currently at a prime stage for academic innovation and career growth. They not only form the backbone of knowledge creation and frontier exploration, but also serve as key drivers and communicators of the rapid evolution and transformative applications of AI technologies.

Young scientists are uniquely positioned to transcend geographical and cultural boundaries, building trust and fostering collaboration across borders. This capability is essential for the global community to build consensus on critical issues such as data security and ethical standards, tackle shared challenges, and promote a more equitable and inclusive global governance framework. A deep understanding of the scale, structure, distribution, collaboration, mobility, and growth of young AI researchers offers valuable insights into the talent strategies of different countries and serves as a foundation for assessing the global development of AI research.

Through in-depth analysis of the geographical distribution, mobility paths, and international collaboration characteristics of young scientists worldwide, this report aims to help policymakers understand regional talent development within a global context and offer insights for developing clearer and more effective strategies—including how to identify comparative advantages and strengthen international cooperation.

Ultimately, this report aspires to contribute to the global cultivation and sustainable development of AI talent.

陆朝阳



Key Findings

Young Scientists Are Highly Concentrated in a Few Countries: Top 10 Countries Account for Over 80% of the Total

The global distribution of young AI scientists is highly concentrated, with the top 10 countries accounting for over 80% of the total talent pool. China leads with more than 50,000 researchers, followed by the United States. India, England, Germany, and other nations also feature prominently. Collectively, China and the United States represent more than half of all young AI scientists worldwide.

Within China, young scientists are primarily affiliated with national research institutes and leading universities. The top 10 host institutions are: Chinese Academy of Sciences, Tsinghua University, Zhejiang University, Shanghai Jiao Tong University, Peking University, University of Science and Technology of China, Tianjin University, University of Electronic Science and Technology of China, Xi’an Jiaotong University, and Northwestern Polytechnical University

Over 60% of Scientists with Multiple Moves Ultimately Return Home Countries

Overall, young AI scientists predominantly build their careers in their home countries, with approximately 20% having experienced cross-border mobility. International mobility is highly concentrated, with the largest and roughly balanced bilateral mobility occurring between China and the United States, far exceeding that of any other country pair. Beyond China–U.S. exchanges, there are also substantial bidirectional mobility links between China and countries such as Australia and Singapore, while the United States maintains significant cross-border exchanges with England, Canada, and India.

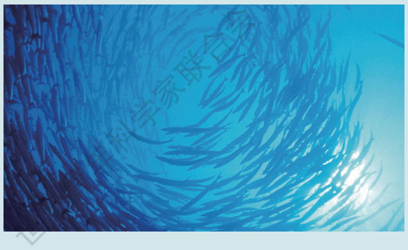
Multiple moves and “circular mobility” are common among young AI scientists. More than 60% of those with multiple international moves ultimately return to their home countries. This suggests that brain drain is not a permanent loss, but rather a manifestation of cyclical resource allocation on a global scale.

Around 30% of Papers Involve International Collaboration

Research activities among young AI scientists are predominantly collaborative, with individually completed studies accounting for only a very small proportion. Most papers exhibit a “small-team” collaboration pattern, with the number of authors typically ranging from 2 to 7, and both the average and median number of authors being four.

In terms of authorship roles, most young scientists participate as “co-authors,” indicating that they tend to play supporting or executing roles within their research teams. At the same time, a significant portion of them serve as both first and corresponding authors, showing that young AI researchers often lead projects while remaining deeply involved in frontline research.

In terms of international collaboration, about 30% of all papers involve cross-border partnerships, with China–U.S. collaboration standing out prominently, far exceeding that of any other country pair.



Contents

1 Background and Objectives

1.1 Background and Significance

Artificial intelligence (AI) is a strategic technology spearheading a new wave of scientific and industrial revolution, positioning itself as a key frontier and a strategic priority for major countries globally. Enabled by parallel advances in algorithms, computing power, and data availability, AI has achieved leapfrog progress and is profoundly reshaping research paradigms, industrial structures, and societal production and lifestyles.

Young scientists have become central drivers of AI innovation. They are not only active in academia but also deeply engaged in industrial R&D and real-world applications. Understanding their geographic distribution, mobility paths, and collaboration characteristics is therefore of considerable relevance and practical significance.

AI talent development is now a major concern for domestic and international scholars, international organizations, government agencies, and academic institutions have also released various reports on AI talent, providing in-depth analyses of its overall scale, geographic distribution, training mechanisms, mobility trends, and competitive landscape. These studies offer important references for understanding the global development of AI talent. However, compared with the broader research population, young scientists exhibit distinct characteristics in terms of research output, collaboration patterns, international mobility, and academic career paths. Most existing studies focus on the overall AI workforce or top-tier talent, while systematic analyses specifically targeting young AI scientists remain noticeably insufficient.

This report leverages publication data from top-tier AI journals and conferences, combined with information on authors' educational backgrounds, research output, and career histories. Using scientometric analysis and machine learning methods, it systematically identifies and filters a cohort of young scientists aged approximately 45 and below. Building on this foundation, the report analyzes their geographical distribution, mobility pathways, and collaboration patterns, aiming to provide data-driven insights and references for a deeper understanding of the global landscape of young AI scientists.

1.2 Research Objectives

This report leverages extensive bibliometric and career profile data of young AI scientists to establish an analytical framework for assessing the development landscape of young AI scientists globally. It aims to deliver multidimensional insights covering geographic distribution, talent mobility, and international collaboration. The specific objectives include the following three aspects:

First is understanding the global distribution pattern. By analyzing the institutional affiliations of paper authors, this report systematically delineates the distribution of young AI scientists across countries and regions.

Second is analyzing talent mobility patterns. Through longitudinal tracking of authors' institutional affiliations, this report visualizes cross-border mobility paths of young AI scientists.

Third is identifying research collaboration patterns. Utilizing co-authorship data, this report analyzes collaborative practices among young AI scientists, with focus on collaboration scale, author roles, and international partnership patterns.

2 Literature Review

2.1 Review of Research on AI Talent

In recent years, AI talent has emerged as a critical global focus. Numerous authoritative institutions have published reports continuously tracking and analyzing the distribution, development, and mobility of AI talent across countries, regions, and professional tiers. Tsinghua University's Aminer platform, for instance, ranks leading AI researchers based on author data from top-tier journals and conferences across 20 core AI fields over the past decade, selecting 2,000 individuals for its annual list of top AI talents.[1] In 2024, MacroPolo (under the Paulson Institute) released the Global AI Talent Tracker, which samples authors from NeurIPS as a proxy for elite AI talent and incorporates CV data to examine their career trajectories and geographic distribution.[2] Draup's View on Global AI Talent Landscape (late 2024) analyzes worldwide talent distribution, career paths, compensation, team structures, and mobility using a combination of career profiles and recruitment data.[3] The UK-based data company Zeki Data released two reports —AI Moneyball[4] and The State of AI Talent 2025[5] — which analyze the characteristics and career trajectories of 2,000 top AI professionals, as well as the overall competitive landscape of global AI talent. The International Finance Forum (IFF) released Part III of IFF Global Artificial Intelligence Competitiveness Index Report in 2025, offering a comprehensive analysis of AI talent distribution and mobility across 20 countries.[6] Stanford HAI's 2025 AI Index Report (8th edition) includes an in-depth look at AI talent training[7], while the U.S. Council of Economic Advisers issued its own AI Talent Report in February 2025[8]. Taken together, these reports provide a multi-dimensional and authoritative overview of the global AI talent landscape.

In terms of geographical distribution, different studies have reached varying conclusions due to differences in data sources and statistical criteria. Industry reports generally indicate that the United States leads in both the quantity and quality of AI talent, while studies based on publication data highlight China's strengths in research output and its growing pool of young scientists. Overall, global AI talent is highly

---

1 https://www.aminer.cn/ai2000/about/introduction

2 https://www.paulsoninstitute.org/press_release/study-finds-us-remains-a-magnet-for-worlds-best-and-brightest-ai-talent-but-more-global-talent-are-staying-home-instead-of-going-abroad/

3 https://draups3assets.s3.us-east-2.amazonaws.com/wp-content/uploads/2025/01/15055404/3.0-Draup_Global-AI-Report_compressed-1.pdf

4 https://zekidata.com/report/ai-moneyball/

5 https://zekidata.com/report/the-state-of-ai-talent-2025

6 https://ifforum-r2.static.ifforum.org/dist/2025/07/AIEP3/%E5%9B%BD%E9%99%85%E9%87%91%E8%9E%8D%E8%AE%BA%E5%9D%9B%EF%BC%88IFF%EF%BC%89%E5%85%A8%E7%90%83%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD%E7%AB%9E%E4%BA%89%E5%8A%9B%E6%8C%87%E6%95%B0%E6%8A%A5%E5%91%8A%20-%20%E7%AC%AC%E4%B8%89%E7%AF%87.pdf

7 https://hai.stanford.edu/ai-index/2025-ai-index-report

8 https://bidenwhitehouse.archives.gov/cea/written-materials/2025/01/14/ai-talent-report/

concentrated in a handful of countries, with the United States and China standing as the two major centers.

From an industry-data perspective, the United States leads in both talent quantity and quality. The global distribution of AI talent shows a strong concentration trend, with the United States and China standing as the two major centers. According to IFF Global Artificial Intelligence Competitiveness Index Report released by the International Finance Forum (IFF), there are approximately 3 million AI professionals worldwide, with 33% based in the United States and 22.4% in China. Analyses focusing on top-tier talent further indicate that the United States continues to hold a qualitative edge. The MacroPolo report finds that 57% of the world's top 2% of AI experts are located in the United States, compared with 12% in China. Similarly, data from Zeki Data suggest that more than 70% of frontier AI talent is concentrated in the United States.

Studies based on publication data emphasize China's strengths in research output and its reserve of young talent. According to a report by Science, China has approximately 30,000 AI researchers across different age groups, compared with around 10,000 in the United States — and China's AI talent pool is overall younger[1]. This finding aligns with several other studies based on academic publications, which likewise indicate that China holds a relative advantage in research output and the scale of its young talent.

In terms of growth trends, the proportion of top-tier AI talent originating from China is rising rapidly. According to MacroPolo data, China accounted for 47% of newly trained world-class AI researchers in 2022, a significant increase from 29% in 2019. Meanwhile, emerging markets such as the United Arab Emirates, Saudi Arabia, and Brazil are steadily expanding their AI talent pools, while traditional technology powerhouses—including the United Kingdom, South Korea, Canada, and Australia—continue to play key roles in the global AI talent landscape and competition.

In terms of talent mobility, the United States remains the primary hub for high-level AI talent, but its long-standing "one-way attraction" pattern is weakening, giving way to an increasingly multipolar mobility. China, meanwhile, is gradually transforming from a major "talent exporter" into a "two-way hub," demonstrating greater output efficiency in China–U.S. research collaboration.

China, meanwhile, is gradually transforming from a major "talent exporter" into a "two-way hub," demonstrating greater output efficiency in China–U.S. research collaboration. According to research by MacroPolo, Chinese nationals account for as much as 38% of the talent in top U.S. AI research institutions—roughly on par with the proportion of American-born researchers (37%). This indicates that the United States remains the world's leading destination for top-tier AI professionals. At the same time, MacroPolo notes that the inflow and outflow of AI PhDs in the United States have nearly balanced out, suggesting a decline in the share of researchers choosing to stay long-term and signaling the end of America's once-dominant "brain-drain magnet" effect.

Concurrently, China's attractiveness is growing rapidly. A study published in Nature observed a rising inflow of top AI scholars into China, reflecting its transition from a pure exporter of talent to a dual-flow hub

1 Science. China tops world in artificial intelligence publications: Database analysis reveals [EB/OL] . Science, (2025-07-11)[2025-09-26] . https://www.science.org/content/article/china-tops-world-artificial-intelligence-publications-database-analysis-reveals

that both exports and attracts talent. [1] China is not alone in this shift—India shows a similar pattern: while nearly all Indian AI researchers worked abroad in 2019, by 2022 one-fifth had chosen to remain and work in India. Together, these developments underscore the emergence of a more multipolar pattern in global AI talent mobility.

In terms of institutional distribution, there are clear differences across countries. Both the Aminer AI2000 ranking and MacroPolo reports show that China's top AI talent is largely concentrated in leading universities and research institutions such as Tsinghua University, Zhejiang University, and the Chinese Academy of Sciences.In contrast, the United States' AI professionals are primarily clustered in major technology companies such as Google, Microsoft, and Meta, reflecting its traditional strength in industry-driven innovation.

This contrast aligns with the global trend of AI research increasingly shifting toward the industrial sector. Stanford University's AI Index Report 2024 points out that industry is now taking the lead in frontier AI research. According to data from Georgetown University's database, 3.56% of all computer science papers published globally in the past five years included at least one industry-affiliated author. Among these, Chinese scholars contributed 23.58%, indicating that the research quality and impact of Chinese enterprises are improving rapidly.

In terms of career development paths, top AI talents show remarkable consistency in both their educational and professional backgrounds. Research by Ziki Data reveals several shared characteristics: First, they possess strong academic foundations—80% graduated from the world's top 50 universities, with a preference for disciplines such as computer science, mathematics, and physics. Second, they tend to have high-quality internship experiences—the likelihood of having interned at one of the seven major tech giants (including Google and Meta) is three times higher than that of average AI professionals. Third, they exhibit a stronger willingness for international mobility, with 60% expressing interest in working abroad.

The International Finance Forum (IFF) Report further confirms that AI professionals worldwide are generally highly educated, with over 88% holding a master's degree or above. Since the launch of ChatGPT in 2022, the AI industry has attracted a large influx of young talent, who have become a key driving force in the sector's rapid development.

2.2 Concept and Definition Standards of Young Researchers

To clarify the definition of "young researchers," this study systematically reviewed policy documents issued by governments, research funding agencies, and academic organizations in various countries, while

1 AlShebli B, Memon S A, Evans J A, et al. China and the U.S. produce more impactful AI research when collaborating together [J]. Scientific Reports, 2024, 14: 28576. DOI:10.1038/s41598-024-79863-5.

also drawing on the classification practices of existing studies. The findings show that there is no universally accepted international standard for defining "young researchers," but several common approaches have emerged.

The first approach is the age limit, which is often specified in youth funding or talent programs. For example, China's International Young Talents Program sets the upper age limit at 45, serving as a key threshold for the recruitment of high-level young professionals.[1] In Italy, the Ministry of Education and Research (MUR) restricts applicants for its Young Researchers Program to those aged 40 or below, extendable to 45 if the applicant obtained a Ph.D. within the past seven years[2]. Similarly, Norway's FRIPRO Young Researcher Project requires applicants to be no older than 40[3].

The second approach focuses on years of professional experience, usually calculated as the number of years since earning a Ph.D. For example, the European Research Council (ERC) Starting Grant requires applicants to have obtained their doctoral degree within 2 to 7 years prior to application (to be extended to 10 years starting in 2027)[4]. Similarly, the U.S. National Institutes of Health (NIH) Early Stage Investigator program limits eligibility to researchers within 10 years of earning their highest research degree or completing postgraduate training, provided they have not previously received substantial independent research funding[5].

The third approach concerns employment status, specifically whether the researcher has obtained a permanent position. For instance, the European Science Foundation defines "early career stage" researchers as those who have held a permanent position for no more than five years. In the United States, the Early Career Research Act defines early-career researchers as those holding tenure-track assistant professorships[6].

The fourth approach is based on research career duration, often measured from key academic milestones such as the time of first publication, first first-author paper, or first corresponding-author paper. For instance, the year in which a researcher publishes their first paper is sometimes defined as the starting point of their academic career. In practice[7], this approach is often combined with certain quantitative or qualitative criteria—such as the number or impact of publications—to determine eligibility.

Overall, these four approaches each have distinct characteristics: age and years since Ph.D. are straightforward to implement, while employment status and research output more accurately reflect substantive academic achievement. In practice, a multi-dimensional combination of these indicators is often adopted to strike a balance between research feasibility and academic soundness.

---

1 https://tysp.cstec.org.cn/Webpages/viewdetail.aspx?id=4063868720&

2 https://www.unipd.it/en/mur-funding-young-researchers-2024

3 https://www.forskningsradet.no/en/call-for-proposals/2023/researcher-project-young-talents-fripro/

4 https://erc.europa.eu/apply-grant/starting-grant

5 NIH.New and Early Stage Investigator Policies[EB/OL].[2022-01-13].https://grants.nih.gov/policy/early-stage/index.htm

6 Niu Ping, Meng Xiangli, Su Fen, et al. "Standards and Implications of 'Research Age' and 'Early Career Stage' in Youth Talent Funding[J]."China Science Foundation, 2013, 27(1): 4. DOI: CNKI:SUN:ZKJJ.0.2013-01-006.

7 Zhang Lihua, Yao Changqing. "The Impact of Retraction Events in Early Research Careers on Scientific Performance[J]." Chinese Journal of Scientific and Technical Periodicals, 2024, 35(11): 1510-1522. DOI: 10.11946/cjstp.202407310838.

2.3 Definition Principles and Methods for Research Subjects in This Report

The research subjects of this report are defined as "young scientists." After systematically reviewing relevant international and domestic practices and considering data availability and cross-country comparability, this report adopts a three-step approach to define "young scientists," establishing a complementary mechanism between age limits and publication characteristics.

Step 1: The term "young" is uniformly defined as 45 years old or younger, a baseline consistent with widely accepted international practices, disciplinary training cycles, and the need for cross-country comparison.

Step 2: The age definition does not rely on a single dimension; instead, it integrates Ph.D. graduation year and publication characteristics to align closely with the 45-year threshold.For individuals with ORCID-verified Ph.D. graduation years, the analysis follows the U.S. National Science Foundation's Survey of Earned Doctorates (SED) [1], which indicates an average Ph.D. graduation age of around 30. Accordingly, graduation year is converted into an age indicator. For large groups lacking biographical data, a classification model is trained using 20 publication behavior features—including publication time series, author roles, and output intensity—to estimate the probability of being "≤45 years old." The model's threshold is fine-tuned through cross-validation to ensure consistency with the unified age criterion across large samples.

Step 3: After identifying age, a minimum academic activity level and basic quality threshold are applied to ensure that the selected population is both scientifically active and academically contributive.

This definition framework establishes 45 years old as the unified age standard, uses Ph.D. graduation year as the baseline for determining research samples, and leverages behavioral-feature modeling for large-scale data expansion—balancing international comparability with the breadth and robustness of classification results. Unless otherwise specified, all subsequent analyses and findings in this report are based on the above definition of "young scientists."

1 https://ncses.nsf.gov/surveys/earned-doctorates/2023#data

3 Data and Methodology

3.1 Data Sources

The data sources for the sample of this report are as follows:

(1) Web of Science Core Collection (WOS Core) publication data. The retrieval window covers the period from 2000 to 2025. Based on WOS discipline classifications, AI-related keywords, and high-level AI journals and conference lists recommended by the China Computer Federation (CCF) and the Chinese Association for Artificial Intelligence (CAAI), the search was restricted to document types of article, review, or proceedings paper.As of June 24, 2025, a total of 1,609,509 AI-related papers published since 2000 were retrieved. This dataset includes information on paper titles, abstracts, keywords, author affiliations, and institutional details, providing a solid foundation for identifying young AI scientists and for subsequent analyses of their geographical distribution, research domains, collaboration networks, and mobility patterns.

(2) ORCID Profile Data. Educational background information was retrieved from the ORCID website based on the authors' ORCID identifiers, yielding a dataset of approximately 20,000 researchers with recorded Ph.D. graduation years. This dataset includes details such as the start and end dates of education, institution names, and degree information. It is primarily used to record Ph.D. graduation years and serves as a key reference for the training and validation of the research-age inference model in subsequent analyses.

3.2 Data Preparation

3.2.1 Author Name Disambiguation

First, author identity disambiguation was performed to eliminate the effects of name duplication, variant spellings, and institutional mobility. Based on a corpus of 1.6 million AI-related papers, a comprehensive author disambiguation model was developed using multiple features such as author names, institutional affiliations, email addresses, collaborator networks, and ORCID identifiers.

The model determines whether author records across different papers belong to the same individual through a multilayered process that includes name normalization, institution and address standardization, email and ORCID verification, and similarity calculations based on co-author and research-topic networks. In practice, both deterministic rules (e.g., direct merging of records with identical ORCID or email) and machine-learning models were employed to estimate the probability that candidate pairs represent the same person.For high-risk cases, iterative optimization and manual sampling checks were introduced to ensure accuracy and stability.

After multiple rounds of model tuning and quality control, approximately 2 million distinct researchers were identified from the 1.6 million AI papers. To ensure that the research subjects demonstrate recent scientific activity and a high level of academic quality, a further selection was made: researchers who had published in CCF-recommended high-level AI journals and conferences within the past five years were retained, yielding a final sample of about 420,000 scholars. This sample guarantees both a minimum quality threshold of academic contribution and a representative coverage of the current core AI research community, providing a robust data foundation for the subsequent identification and analysis of the global young scientist cohort.

3.2.2 Design of Author Feature Indicators

Based on the publication data of 420,000 researchers, a feature indicator system was constructed to support subsequent research-age inference and identification of the young scientist cohort. From these 420,000 researchers' publication records, a set of multidimensional indicators related to the onset of research careers, research output, and role evolution was extracted, as detailed below.

The temporal features cover aspects such as the year of first publication, year of most recent publication, research career span, year of first first-author paper, year of first corresponding-author paper, and year of first co-author paper. These indicators reveal when a researcher entered the academic system, the length of their research career, and the trajectory of their role development.

The quantitative features include total publication count, number of first-author papers, number of corresponding-author papers, number of other-author papers, maximum annual publication count, number of active years, and the number of publications within the first three years of a career. These measures reflect both the overall research output and the intensity of activity during specific career stages.

The comparative features capture the proportions of first-author, corresponding-author, and other-author papers, as well as the time intervals between the first any-author paper and the first first-author or corresponding-author paper. These indicators help illustrate a researcher's transition from collaborator to independent lead investigator.

To validate the effectiveness of these indicators in inferring research age, the subset of 20,000 researchers with known Ph.D. graduation years was used to calculate correlations between each feature and the inclusion criteria derived from ORCID records. The results show that the year of first first-author paper and year of first corresponding-author paper exhibit strong statistical correlations with the inclusion criteria, confirming their value as key proxy variables for research age. These findings provide the data foundation for subsequent machine-learning modeling and age-based filtering. Table 3-1 summarizes the 20 author feature indicators and their correlations with inclusion status.

Table 3-1 Author Feature Indicators

No.	Feature Name	Description	correlation
1	Year of First Publication	The year the author's first paper (any authorship position) was published; reflects the earliest entry into the academic publishing system and is used to estimate age.	0.3668**
2	Year of Most Recent Publication	The publication year of the author's most recent paper (regardless of authorship position); used to calculate the span of the research career and assess whether the author has been engaged in research over the long term.	0.0216**
3	Publication Span (Years)	Difference between the year of the most recent publication and the first publication; a longer research span generally indicates a higher likelihood of older age.	-0.3403**
4	Year of First First-Author Paper	The year the author first published as first author, used as a proxy for estimating the author's age.	0.5173**
5	Year of First Corresponding-Author Paper	The year the author first served as corresponding author, typically after establishing research independence; used as a proxy for age.	0.4232**
6	Year of First Co-Author Paper	The year the author first appeared as a non-lead author, usually during student collaborations; helps identify early career entry and estimate age.	0.3648**
7	Total Publications	Total number of the author's publications, serving as an overall indicator of research productivity.	-0.1923**
8	First-Author Papers	Number of papers published as first author; reflects independent research capability.	0.0929**
9	Corresponding-Author Papers	Number of papers published as corresponding author; reflects project leadership and supervisory capacity.	-0.0307**
10	Co-Author Papers	Number of papers published as other (non-lead) author; reflects collaborative participation.	-0.2635**
11	First-Author Share	Ratio of first-author papers to total publications; a higher ratio suggests an early career stage emphasizing hands-on leadership.	0.3163**
12	Corresponding-Author Share	Ratio of corresponding-author papers to total publications; a higher ratio indicates research maturity and supervisory roles.	0.1568**
13	Co-Author Share	Ratio of co-authored papers (non-lead) to total publications; a higher ratio may indicate a learning or supporting phase.	-0.2562**

continued

No.	Feature Name	Description	correlation
14	Maximum Annual Output	Highest number of papers published in a single year; reflects concentration of output, often coinciding with key career development periods.	-0.0560**
15	Active Publication Years	Total number of years with at least one publication; indicates the span of research activity and supports age estimation.	-0.2650**
16	Publications in First 3 Years	Number of papers published in the first three years of the career; helps identify research activity during the initial stage.	0.0412**
17	Gap: First-Author vs. First Any-Author	Difference between the year of the first first-author paper and the first publication; reflects the time from supporting to leading role.	-0.1340**
18	Gap: First Corresponding vs. First Any-Author	Difference between the year of the first corresponding-author paper and the first publication; reflects the time from initial participation to independent supervision.	-0.1496**
19	Gap: First First-Author vs. First Corresponding	Difference between the year of the first first-author paper and first corresponding-author paper; reflects the transition from independent researcher to project leader.	-0.0618**
20	Average Authors per Paper	Average number of co-authors per paper; indicates the collaboration pattern of the discipline and supports inference of career stage.	0.2253**

3.3 Identification of Research Subjects

The research subjects of this report consist of 126,820 young AI scientists (hereinafter also referred to as “the research sample,” “the sample,” or collectively “global young AI scientists”). The research dataset also includes curriculum and publication data, totaling 448,842 papers produced by the sample group. Unless otherwise specified, the analyses and conclusions presented in Chapters 4 through 8 are all based on this dataset. The sample selection process is described as follows.

Based on the feature and correlation analyses, the year of first publication was widely used as a proxy variable for academic age, showing a strong statistical correlation with the year of Ph.D. graduation. Typically, first-author papers appear during the Ph.D. stage, representing major research contributions, while corresponding-author papers usually emerge during the independent research phase, approximately 2.5 years after Ph.D. graduation.

Using the 20 feature indicators listed in Table 3-1, a machine learning model was trained and evaluated on a sample of 20,000 researchers with known Ph.D. graduation years, employing five-fold cross-validation (5-fold CV). The classification task was framed as a binary problem distinguishing “>45 years old” and “≤45 years old.” A heuristic baseline model based on features such as “first first-author paper,” “first corresponding-author paper,” “first co-author paper,” and “first publication (any authorship)” was used for comparison. The results are shown in Table 3-2.

Table 3-2 Results of Age Verification Experiment

Classification Method	Accuracy	Precision	Recall	F1 Score
Baseline Model	0.728	0.75	0.884	0.812
Random Forest Model	0.841	0.978	0.790	0.874

As shown, the baseline model achieved higher recall—meaning it successfully covered most actual young scientists—but its precision was only 0.75, indicating a higher rate of misclassification, with some researchers over 45 years old included in the list. In contrast, the random forest model achieved a precision of 0.978, indicating almost no misclassification, along with a high F1 score, reflecting good overall balance and an approximately 15% improvement in accuracy over the baseline.

Considering these results, this study adopted the optimized random forest model as the final age-filtering strategy. In the 5-fold cross-validation process, the model achieved a well-balanced performance between precision (97.8%) and F1 score (0.874), providing a robust foundation for subsequent distributional and cross-country analyses. Applying this model to the dataset of 420,000 researchers, 196,000 individuals were initially identified as candidates aged ≤ 45.

To ensure the stability and academic quality of the sample, the study followed the productivity distribution criteria proposed by Larivière et al[1]. and excluded “one-time authors” (those with only one publication). Moreover, all previously selected 420,000 authors were required to have published at least one paper in the past five years (since 2020). Under the combined conditions of a minimum of two publications and at least one publication within the past five years, the 196,000 candidates were further refined to 126,820 researchers, forming the final young scientist sample used for subsequent analyses of geographical distribution, collaboration, and mobility.

1 Larivière, V., & Costas, R. (2016). How many is too many? On the relationship between research productivity and impact. PLoS ONE. https://doi.org/10.1371/journal.pone.0162709

3.4 Research Methods

■ **Literature Review Method.** On one hand, the study focuses on research and authoritative reports related to AI talent, systematically reviewing and comparing white papers, indices, and policy documents on AI talent released by major institutions both in China and abroad. This enables a comprehensive understanding of the global landscape, mobility trends, and competitive dynamics of AI talent. On the other hand, the study systematically examines research findings in the fields of talent studies, education, and research management concerning the concepts of "early-career" and "young talent." By comparing the defining standards across different academic and policy frameworks, the analysis provides a theoretical foundation for identifying and analyzing the global young scientist cohort.

■ **Bibliometric Correlation Analysis.** During the sample selection stage, the study calculated the correlations between researchers' publication characteristics and their known Ph.D. graduation years to verify the representativeness and substitutability of different feature indicators. For example, testing the correlation between the year of first publication and the Ph.D. graduation year helped validate its effectiveness as a proxy variable for academic age and determine its applicability in the research-age inference model. This method was primarily applied in the chapter on sample selection of the study.

■ **Descriptive Statistics and Comparative Analysis.** After processing the sample data, this study applied descriptive statistical methods to calculate the frequencies and proportions of key indicators such as the number of publications per author, team size, collaboration scale, cross-border collaboration volume, and research mobility scope. By conducting horizontal comparisons across different groups, countries, or regions, and combining these with longitudinal annual data, the analysis reveals the characteristics and evolving trends of the research population. This method serves as the primary analytical approach in the chapters on collaboration, mobility, and distribution.

■ **Machine Learning Method.** To accurately identify young scientists aged 45 and below within the large-scale dataset, this study trained and evaluated a random forest classification model based on 20 author features, and used a heuristic baseline model for comparison. Through five-fold cross-validation, the models were compared in terms of accuracy, precision, recall, and F1 score. The random forest model, which demonstrated the best overall performance, was ultimately selected as the age-filtering tool for this study.

■ **Visualization Method.** To enhance the clarity and impact of the findings, this study employed multiple data visualization techniques after completing data processing, including bar charts, line charts, and Sankey diagrams, to present results in an intuitive manner. For instance, distribution patterns were illustrated using bar charts, while collaboration relationships were visualized through Sankey diagrams. These visualization methods were applied throughout the various chapters of the report.

4

Publication Output of Global Young AI Scientists

4.1 Overall Publication Volume Shows Rapid Growth After 2015

From January 2000 to June 2025, a total of 1,609,509 publications were recorded in the field of artificial intelligence. The annual publication trend is shown in Figure 4-1. From 2000 to 2024, the total number of AI publications showed an overall upward trajectory, entering a phase of rapid growth after 2015. By 2024, the number of publications had reached 129,316. Before 2010, publication volumes fluctuated considerably (for example, following a peak in 2001–2002, another rise occurred in 2005–2006). However, since 2015, publication output has increased almost every year, indicating the accelerating growth of AI as a major research hotspot.

Between January 2000 and June 2025, the research sample in this report produced a total of 448,842 publications. Publication output was at its lowest between 2000 and 2005. It began to rise rapidly from 2006 onward, exceeding 3,000 articles by 2008 and reaching 11,126 in 2014. Thereafter, growth accelerated markedly: between 2015 and 2020, the number of publications increased from 14,704 to 44,098, with an average annual growth rate of over 20%. From 2021 to 2023, the number continued to rise, reaching 62,411 publications, followed by a slight dip to 56,409 in 2024, which may be attributed to delays in publication indexing.

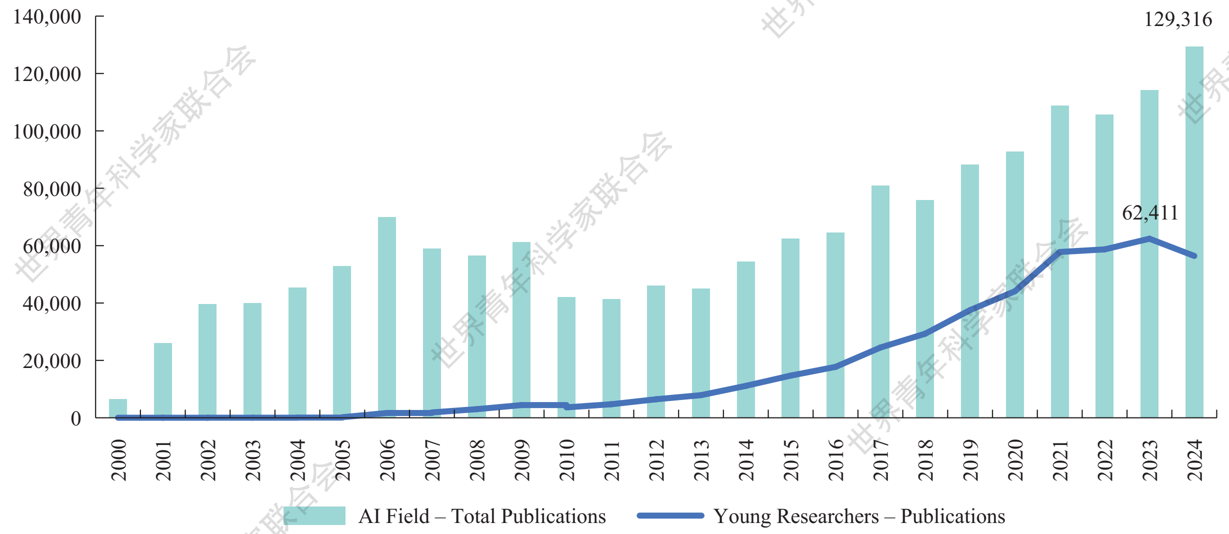


Figure 4-1 Trends in AI Publications (2000–2024)

4.2 Average Publication Output per Researcher is 6, Primarily in Conference Papers

Statistics show that the 126,820 young scientists produced a total of 448,842 publications (795,169 authorship instances), with an average of 6.27 publications per person and a median of 4. Most researchers published between 2 and 7 papers, indicating a moderate concentration of research productivity within this range.

The primary publication type in this field is conference papers, totaling 270,323—accounting for 60.23% of all publications. Journal articles, numbering 178,519, represent a smaller share, suggesting that AI researchers place greater emphasis on academic conference participation and paper presentation.

In the research sample, 289,231 papers (64.44%) received research funding support, highlighting the critical role of funding programs in driving research productivity among young scientists. Young scientists produced a total of 258,479 publications as corresponding authors, yet only 13,868 of these (5.37%) were supported by research grants.This indicates that, during the early stages of their careers, young scientists still face limited access to funding for their independent, corresponding-author research outputs.

5

National and Regional Distribution of Global Young AI Scientists

5.1 Global Distribution Are Highly Concentrated in Top 10 Countries

The institutional affiliations listed in researchers' publications serve as a key indicator of their research activity and academic affiliation. Analyzing the distribution of these institutional addresses provides an objective reflection of the global spatial distribution of researchers. Among the 126,820 samples, 572 lacked institutional address information; therefore, the remaining 126,248 authors were used for the geographical distribution analysis in this chapter.

Based on the latest institutional affiliations of authors, the 126,248 young AI scientists worldwide are distributed across 133 countries and regions, with the majority concentrated in a limited number of research powerhouses. The top ten countries and regions together account for approximately 105,000 individuals, representing 83.74% of the total valid sample (see Table 5-1 for details).

Table 5-1 Top 10 Countries/Regions by Number of Global Young AI Scientists

No.	Country/Region	Count	Share
1	China	53,935	42.72%
2	United States	21,595	17.11%
3	India	6,562	5.20%
4	England	4,608	3.65%
5	Germany	4,547	3.60%
6	Japan	3,038	2.41%
7	South Korea	3,017	2.39%
8	France	2,879	2.28%
9	Canada	2,802	2.22%
10	Australia	2,743	2.17%

Among these top 10 countries, China ranks first with 53,935 researchers (42.27%), followed by the United States with 21,595 researchers (17.11%). Combined, China and the United States account for more than half of the global total. Beyond these two, Asian and European countries also demonstrate strong representation, with India, England, Germany, Japan, South Korea, France, Canada, and Australia ranking third through tenth, respectively.

5.2 Chinese Institutions Are Concentrated in National Research Institutes and Top Universities

A further analysis was conducted on the 53,935 Chinese young scientists to examine their institutional affiliations. The top 20 institutions by number of researchers are presented in Table 5-2.

Table 5-2 Top 20 Chinese Institutions by Number of Young AI Scientists

No.	Institution	Count
1	Chinese Academy of Sciences	1,353
2	Tsinghua University	886
3	Zhejiang University	828
4	Shanghai Jiao Tong University	789
5	Peking University	671
6	University of Science and Technology of China	585
7	Tianjin University	482
8	University of Electronic Science and Technology of China	478
9	Xi’an Jiaotong University	454
10	Northwestern Polytechnical University	450
11	Wuhan University	446
12	Sun Yat-sen University	435
13	Xidian University	419
14	Harbin Institute of Technology	403
15	Huazhong University of Science and Technology	391
16	South China University of Technology	391
17	Shenzhen University	386
18	The Chinese University of Hong Kong	368
19	National University of Defense Technology	343
20	Fudan University	319

Young AI scientists in China are primarily concentrated in national research institutes and top universities.

The Chinese Academy of Sciences ranks first with 1,353 researchers, followed by Tsinghua University (886), Zhejiang University (828), Shanghai Jiao Tong University (789), Peking University (671), and the University of Science and Technology of China (585). These institutions generally possess strong foundations in computer science, information engineering, and related disciplines. They maintain national leadership in disciplinary development, research investment, and talent training within these fields, making them major hubs for young researchers.

Overall, the geographical distribution of young AI scientists in China is characterized by both high concentration and regional diversity. While densely clustered in Beijing, Shanghai, Zhejiang, and Guangdong, they remain active in other key research centers such as Xi'an, Wuhan, Harbin, and Chengdu. This pattern has formed a comprehensive and well-balanced national layout for China's AI research ecosystem.

6

Cross-Border Mobility of Global Young AI Scientists

6.1 Analysis of Cross-Border Mobility Scale

6.1.1 About 80% of Young Scientists Have Not Experienced Cross-Border Mobility

Changes in the institutional affiliations listed in researchers' publications over time often reflect shifts in their workplaces. By analyzing the chronological sequence of institutional addresses, it is possible to accurately trace their cross-border mobility paths. Among the 126,820 samples, 572 lacked institutional address information, leaving 126,248 authors available for the cross-country mobility analysis presented in this chapter.

Of these, 30,123 researchers have experienced cross-border mobility, accounting for 23.86% of the total sample. Together, they generated 58,388 cross-border moves. The remaining 76.14% of young scientists showed no evidence of cross-border mobility. This indicates that while the majority of young scientists pursue their careers primarily within their home countries, there nonetheless exists a significant and active cohort engaged in cross-national mobility. Within the mobile group, 13,982 individuals (46.42%) experienced a single cross-border move, while 16,141 individuals (53.58%) moved two or more times, demonstrating that cross-border mobility among researchers tends to be sustained and recurrent. See Table 6-1 for details.

Table 6-1 Distribution of Cross-Border Mobility Frequency of Global Young AI Scientists

Number of Moves	Count	Share	Total Number of Moves
No Movement	96,125	76.14%	0
1 Move	13,982	11.08%	13982
2 Moves	9,837	7.79%	19674
3 Moves	3,053	2.42%	9159
4 Moves	1,818	1.44%	7272
5 Moves	725	0.57%	3625
6 Moves or More	708	0.56%	4676
Total	126,248	100%	58388

6.1.2 The United States Leads in Net Inflow of Talent, While India Shows Significant Net Outflow

Among the top 30 countries/regions ranked by inflow and outflow count (see Table 6-2), the United States recorded the highest net inflow count, followed by England in second place. Middle Eastern countries such as the UAE and Saudi Arabia also rank among the top net inflow destinations. In contrast, countries including India and Iran show pronounced net outflows, while several traditional research powerhouses—such as Italy, France, South Korea, Japan, and China—also experienced varying degrees of net talent loss. It should be noted that Table 6-2 only includes the top 30 countries/regions by overall mobility scale and does not represent a comprehensive global listing.

Table 6-2 Top 30 Countries/Regions by Inflow and Outflow Counts of AI Talent

No.	Country	Inflow Count	Outflow Count	Net Inflow Count
1	United States	10,212	9,921	291
2	England	3,864	3,611	253
3	Canada	2,100	1,960	140
4	United Arab Emirates	452	346	106
5	Netherlands	887	788	99
6	Saudi Arabia	752	666	86
7	Singapore	2,032	1,965	67
8	Denmark	379	322	57
9	Scotland	550	497	53
10	Sweden	434	404	30
11	Finland	382	362	20
12	Australia	2,651	2,632	19
13	Pakistan	694	685	9
14	Poland	340	341	-1
15	Germany	2,404	2,414	-10
16	Switzerland	1,167	1,180	-13
17	Belgium	383	397	-14
18	Austria	397	421	-24
19	Brazil	377	408	-31

continued

No.	Country	Inflow Count	Outflow Count	Net Inflow Count
20	China	12,840	12,876	-36
21	Spain	1,066	1,108	-42
22	Israel	444	492	-48
23	Turkey	328	389	-61
24	Japan	1,265	1,360	-95
25	South Korea	950	1,053	-103
26	Malaysia	355	495	-140
27	France	1,790	1,933	-143
28	Italy	1,385	1,539	-154
29	Iran	472	657	-185
30	India	1,641	1,839	-198

6.2 Analysis of Cross-Border Mobility Directions

6.2.1 Cross-border mobility is highly concentrated in specific countries

Mobility directions between all country and region pairs were analyzed, covering every possible "origin country → destination country" combination. Each cross-border move record was counted independently, meaning that multiple cross-border moves by the same individual were included separately in the corresponding country-pair frequency statistics.

The results show that major mobility paths are highly concentrated among a small number of countries or regions. The top 20 mobility directions involve only nine countries: China, the United States, Australia, Singapore, England, Canada, Japan, India, and Germany. The bidirectional mobility between China and the United States is the largest in scale. Beyond China–U.S. exchanges, there are also substantial bidirectional mobility links between China and countries such as Australia and Singapore, while the United States maintains significant cross-border exchanges with England, Canada, and India. See Figure 6-1.

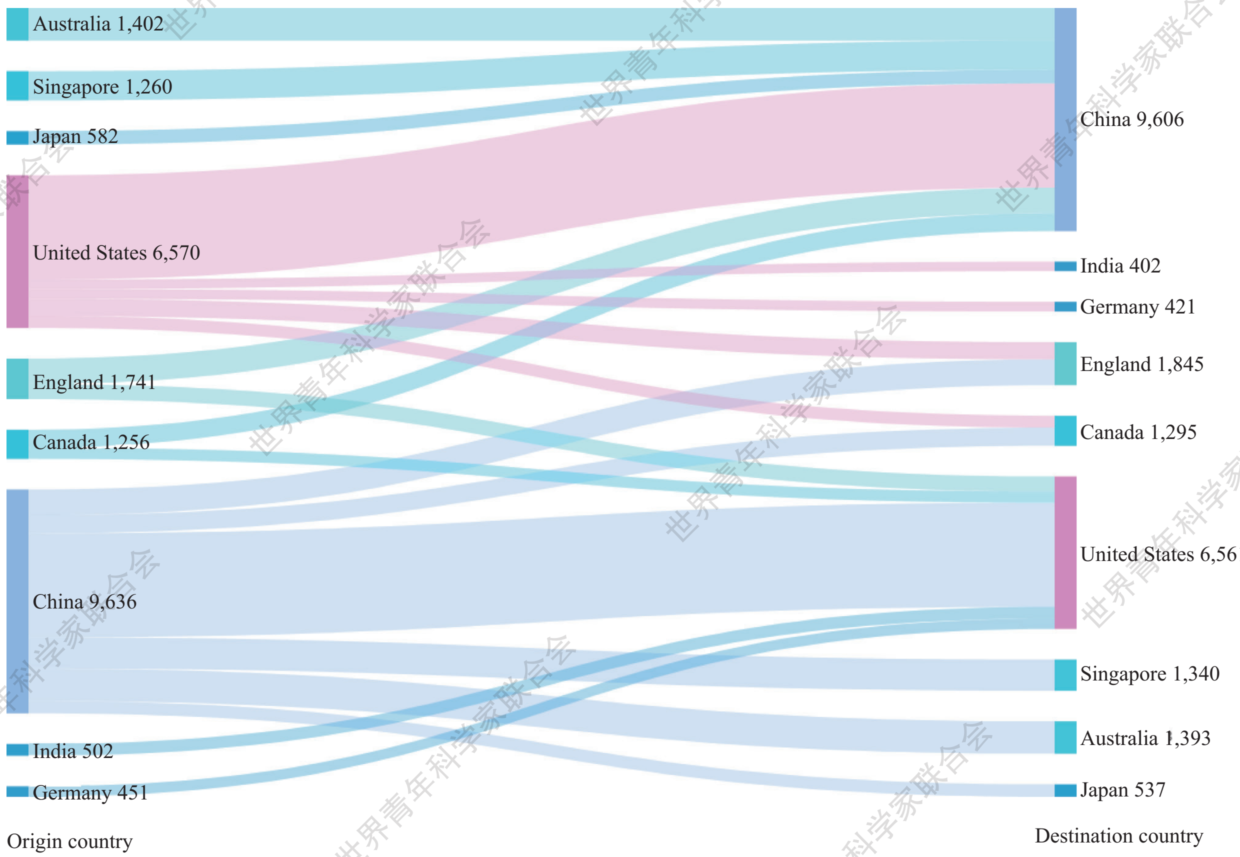


Figure 6-1 Sankey Diagram of Cross-Border Mobility Among Young AI Scientists

The China–U.S. bidirectional mobility represents the largest exchange globally, with 4,493 moves from the United States to China and 4,480 moves from China to the United States, forming a nearly equal bidirectional relationship that far exceeds all other country pairs. The bidirectional moves between China and the United States account for 15.37% of total. See Table 6-3.

Table 6-3 Major Mobility Directions of Global Young AI Scientists

No.	Origin Country	Destination Country	Number of Moves	Share
1	United States	China	4,493	7.70%
2	China	United States	4,480	7.67%
3	Australia	China	1,402	2.40%
4	China	Australia	1,393	2.39%
5	China	Singapore	1,340	2.29%
6	Singapore	China	1,260	2.16%
7	China	England	1,107	1.90%
8	England	China	1,095	1.88%

continued

No.	Origin Country	Destination Country	Number of Moves	Share
9	China	Canada	779	1.33%
10	Canada	China	774	1.33%
11	United States	England	738	1.26%
12	England	United States	646	1.11%
13	Japan	China	582	1.00%
14	China	Japan	537	0.92%
15	United States	Canada	516	0.88%
16	India	United States	502	0.86%
17	Canada	United States	482	0.83%
18	Germany	United States	451	0.77%
19	United States	Germany	421	0.72%
20	United States	India	402	0.69%

6.2.2 Top 10 Mobility Directions Are Concentrated Among Six Countries

Analysis of the top 10 country pairs by annual cross-border moves from 2020 to 2024 reveals that the cross-border moves of young AI scientists remain concentrated within a small group of six countries: China, the United States, Singapore, Australia, England, and Canada. These paths exhibit stable bidirectional structures, indicating that recent cross-border mobility of young AI researchers continues to be channeled through a limited set of key corridors. Furthermore, the bilateral talent mobility between China and the United States has consistently ranked first and second in total move counts. Since 2021, the number of researchers moving from the United States to China has surpassed those moving from China to the United States, indicating that China has become a net inflow country relative to the United States.

China also maintains relatively stable bidirectional mobility links with Australia and England, with these directions consistently appearing in the annual top 10. The United States–England bidirectional mobility appears among the top 10 in total move counts in selected years. See Table 6-4 for details.

Table 6-4 Top 10 Mobility Directions by Annual Person-Moves (2020–2024)

Mobility Direction	2020	2021	2022	2023	2024
United States → China	501 (Rank 2nd)	668 (Rank 1st)	792 (Rank 1st)	726 (Rank 1st)	642 (Rank 1st)
China → United States	554 (Rank 1st)	665 (Rank 2nd)	615 (Rank 2nd)	706 (Rank 2nd)	555 (Rank 2nd)

continued

Mobility Direction	2020	2021	2022	2023	2024
China → Singapore	108 (Rank 6th)	162 (Rank 6th)	195 (Rank 6th)	285 (Rank 3rd)	316 (Rank 4th)
Australia → China	145 (Rank 4th)	188 (Rank 3rd)	240 (Rank 3rd)	244 (Rank 4th)	246 (Rank 5th)
China → Australia	167 (Rank 3rd)	174 (Rank 4th)	215 (Rank 4th)	231 (Rank 6th)	198 (Rank 8th)
Singapore → China	103 (Rank 7th)	128 (Rank 8th)	176 (Rank 7th)	233 (Rank 5th)	338 (Rank 3rd)
England → China	82 (Rank 9th)	130 (Rank 7th)	206 (Rank 5th)	203 (Rank 8th)	213 (Rank 6th)
China →England	112 (Rank 5th)	173 (Rank 5th)	137 (Rank 8th)	192 (Rank 9th)	200 (Rank 7th)
China → Canada	77 (Rank 10th)	108 (Rank 9th)	112 (Rank 10th)	172 (Rank 10th)	Not in Top 10
Canada → China	92 (Rank 8th)	Not in Top 10	121 (Rank 9th)	Not in Top 10	169 (Rank 9th)
United States →England	Not in Top 10	99 (Rank 10th)	Not in Top 10	228 (Rank 7th)	Not in Top 10
England → United States	Not in Top 10	Not in Top 10	Not in Top 10	Not in Top 10	137 (Rank 10th)

6.3 Analysis of Cross-Border Mobility Paths

6.3.1 "Circular Mobility" is Evident in Multiple-Move Paths

A statistical analysis of the individual mobility paths of young AI scientists reveals that some experience multiple cross-border moves. A notable pattern is "circular mobility", in which researchers return to their origin country or region after moving abroad several times. This pattern not only reflects the complexity of researchers' career trajectories, but also highlights the cyclical nature of the global talent mobility network.

Among the 126,248 individuals analyzed, 30,123 had experienced at least one cross-border movement, and 16,141 had moved internationally two or more times, demonstrating the group's high level of activity and mobility within the global research system. Among the 16,141 individuals with multiple moves, 10,258 followed circular paths, accounting for 63.55%. This suggests that for many young AI scientists, cross-border mobility is often phased and round-trip in nature. Most researchers who go abroad for scientific work eventually return to their origin country or region, indicating that international mobility serves primarily as a means of academic collaboration and experience accumulation, rather than as a form of permanent talent loss.

An examination of the top 20 multi-move paths shows that China–U.S. exchanges are the most prominent. A total of 1,266 young AI scientists followed the path "China → United States → China," far exceeding other paths, while 407 followed the reverse path "United States → China → United States." Additionally, several long-chain mobility paths were observed between the two countries. Beyond China–U.S. exchanges, China also exhibits large-scale circular mobility with Australia, Singapore, and England. Similarly, the United States shows significant circular mobility with England, South Korea, Canada, and India. See Table 6-5 for details.

Table 6-5　Top 20 Multi-Move Paths of Global Young AI Scientists

No.	Multi–Move Paths	Count
1	China → United States → China	1,266
2	United States → China → United States	407
3	China → Australia → China	391
4	China → Singapore → China	362
5	China → England → China	288
6	China → Canada → China	269
7	China → Japan → China	140
8	United States → England → United States	120
9	China → United States → China → United States	99
10	United States → China → United States → China	96
11	China → United States → China → United States → China	95
12	South Korea → United States → South Korea	86
13	China → Germany → China	84
14	United States → Canada → United States	82
15	United States → India → United States	82
16	Australia → China → Australia	78
17	Singapore → China → Singapore	70
18	China → South Korea → China	61
19	England → United States → England	60
20	India → United States → India	59

6.3.2 China Exhibits the Largest Circular Mobility Scale, While Italy Shows the Highest Circularity Ratio

In terms of countries and regions where circular mobility occurs, China ranks first, with 4,437 young AI scientists who moved abroad for research and later returned to China. Among the top 10 countries/regions by scale of circular mobility, Italy shows the highest circularity ratio: among 2,650 young scientists currently working in Italy, 12.08% previously left Italy for research in other countries and later returned. See Table 6-6 for details.

Table 6-6 Top 10 Countries/Regions by Scale of Circular Mobility Among Global Young AI Scientists

No.	Start/End Country	Circular Movers	Count	Share
1	China	4,437	53,935	8.23%
2	United States	1,507	21,595	6.98%
3	Germany	356	4,547	7.83%
4	India	338	6,562	5.15%
5	Italy	320	2,650	12.08%
6	England	317	4,608	6.88%
7	France	247	2,879	8.58%
8	Australia	237	2,743	8.64%
9	Spain	204	2,275	8.97%
10	South Korea	196	3,017	6.50%

From the perspective of circular mobility pathways, among the top 20 circular paths by personnel count, 19 paths (see Figure 6-2) follow an "A (origin country) → B (transit country) → A (return to origin country)" pattern, meaning only one country is transited. The sole exception is one long-chain path: "China → United States → China → United States → China".

The "China → United States → China" path ranks first with 1,266 individuals, significantly outpacing all other circular paths and highlighting the most active research exchange between the two nations. In addition to the China–U.S. exchange, frequent mobility links were also observed between China and Australia, Singapore, England, Canada, and Japan, as well as between the United States and England, South Korea, Canada, and India Similarly, the United States shows significant circular mobility with England, South Korea, Canada, and India (see Table 6-7).

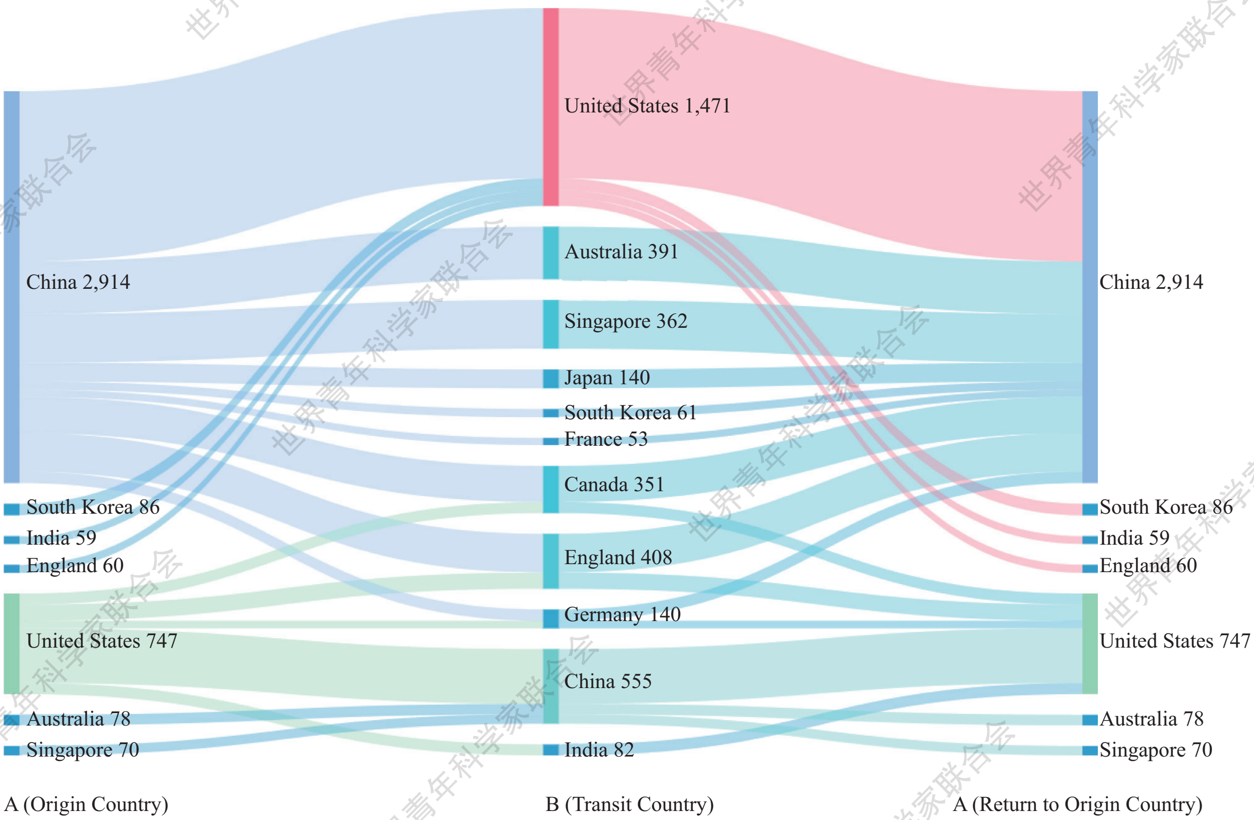


Figure 6-2 Sankey Diagram of Circular Mobility Among Young AI Scientists

Table 6-7 Top 20 Circular Mobility Paths of Global Young AI Scientists

No.	Circular path	Count
1	China → United States → China	1266
2	United States → China → United States	407
3	China → Australia → China	391
4	China → Singapore → China	362
5	China → England → China	288
6	China → Canada → China	269
7	China → Japan → China	140
8	United States → England → United States	120
9	China → United States → China → United States → China	95
10	South Korea → United States → South Korea	86
11	China → Germany → China	84
12	United States → Canada → United States	82
13	United States → India → United States	82

continued

No.	Circular path	Count
14	Australia → China → Australia	78
15	Singapore → China → Singapore	70
16	China → South Korea → China	61
17	England → United States → England	60
18	India → United States → India	59
19	United States → Germany → United States	56
20	China → France → China	53

7

Collaboration Patterns and Authorship Roles of Young Scientists

7.1 Most Collaborations Involve Teams of Fewer Than Ten People

Papers by global young AI scientists are predominantly collaborative. Among the 448,842 papers analyzed, 438,956 were co-authored, accounting for 97.80%, while only 9,886 were single-author papers. while only 9,886 were single-author papers. This indicates that within the AI research community, the scientific output of young scientists is largely the result of team-based collaboration rather than individual work. The distribution of co-authored papers by number of authors is shown in Table 7-1.

Table 7-1 Distribution of Co-Authored Papers by Number of Authors Among Global Young AI Scientists

Number of Authors	Number of Papers	Share
2	58,024	13.22%
3	101,835	23.20%
4	105,481	24.03%
5	78,079	17.79%
6-10	90,801	20.69%
11-50	4,648	1.06%
51-100	78	0.02%
Over 100 authors	10	0.00%

From the perspective of co-authorship scale, the majority of papers involve ten or fewer authors, representing 98.92% of all co-authored papers. Among them, papers written by four authors are the most common, accounting for 24.03%, followed by those with three or five authors. Papers with six to ten authors also make up a significant proportion. The average number of authors per paper is 4.31, while both the median and mode are four, suggesting that small-team collaboration is a prevailing feature of research activity among this group.

Although most papers are produced by teams of modest size, a few represent large-scale collaborations. For example, the paper with the largest author group included 132 authors and was published in 2024 under the title “The Ninth NTIRE 2024 Efficient Super-Resolution Challenge Report.” This large conference paper, involving multiple international teams, demonstrates that task-driven or competition-driven projects can still lead to extensive, cross-institutional collaboration in the AI field.

7.2 Young Researchers Continue Deep Involvement in Frontline Research While Leading Projects

In collaborative scientific research, the authorship roles undertaken in publications reflect the nature and degree of contribution of each researcher. Within this study's sample of 126,820 young scientists, there were a total of 795,169 authorship instances (the number of authorships exceeds the total of 448,842 papers due to multi-author collaboration among sample members).A statistical analysis of the frequency of different authorship roles reveals significant variation across categories (see Table 7-2).

Table 7-2 Distribution of Authorship Roles in Papers by Global Young AI Scientists

Role Type	Number of Authorships	Share
First Author Only	95,163	11.97%
Corresponding Author Only	83,690	10.52%
Co-Author (Other Author)	441,527	55.53%
First + Corresponding Author	174,789	21.98%

Among all roles, "co-author" is the most frequent, accounting for 441,527 instances, far exceeding other categories. This indicates that in collaborative AI research, most young scientists participate in non-leading roles, contributing primarily as executing or supporting members within their teams.

The core authorship positions in research papers—first author (who formulates key ideas and drafts the paper) and corresponding author (who oversees the project and provides academic supervision)—show more balanced participation. Young scientists served as first authors only in 95,163 instances, slightly higher than the 83,690 instances where they served as corresponding authors only. This suggests that, within the young scientist cohort, those directly responsible for research execution and manuscript writing slightly outnumber those responsible for overall project coordination and academic oversight.

Notably, there were 174,789 instances in which young scientists served simultaneously as both first and corresponding authors, far exceeding the number of cases where they held either role alone. This finding demonstrates that in the field of artificial intelligence, many young researchers not only take on the main research and writing responsibilities but also possess the capacity for independent leadership and coordination. Such dual roles highlight their academic autonomy, integrated leadership potential, and their emerging status as a core driving force within the global AI research community.

7.3 About 30% of Papers Involve International Collaboration

To examine the characteristics of international collaboration from the perspective of young scientists, this section analyzes the cross-country coauthorship patterns in their publication records. Based on the 448,842 papers associated with the sample, the statistics reflect the collaborative modes of research activity among young AI scientists. It is worth noting that the total number of authors across these papers the 126,820 young scientists included in the study sample.

Among the 448,842 papers, after excluding 1,048 records lacking country information, the remaining 447,794 papers show that 320,474 papers did not involve international collaboration, accounting for 71.57% of the total.This indicates that domestic collaboration within a single country remains the dominant mode of research among young AI scientists. The distribution of cross-country collaborations is as follows: 98,013 papers involved researchers from two countries; 21,889 papers involved three countries; 5,102 papers involved four countries; 2,316 papers involved researchers from five or more countries, accounting for only 0.52% of the total. Overall, while international collaboration is clearly present within the young scientist cohort, large-scale multilateral collaboration remains relatively rare. See Table 7-3 for details.

Table 7-3 Distribution of Cross-Border Collaboration Papers by Global Young AI Scientists

Number of Collaborating Countries	Number of Papers	Share
No Cross-Border Collaboration	320,474	71.57%
2	98,013	21.89%
3	21,889	4.89%
4	5,102	1.14%
≥5	2,316	0.52%

The frequency of collaboration between countries and regions is shown in Figure 7-1 and Table 7-4. Overall, China–U.S. collaboration dominates the global landscape, with the number of joint papers between the two countries far exceeding that of any other bilateral combination. Among the top 20 bilateral collaborations worldwide, only “England–Germany” does not involve either China or the United States, underscoring the central roles of both countries in global AI research cooperation.

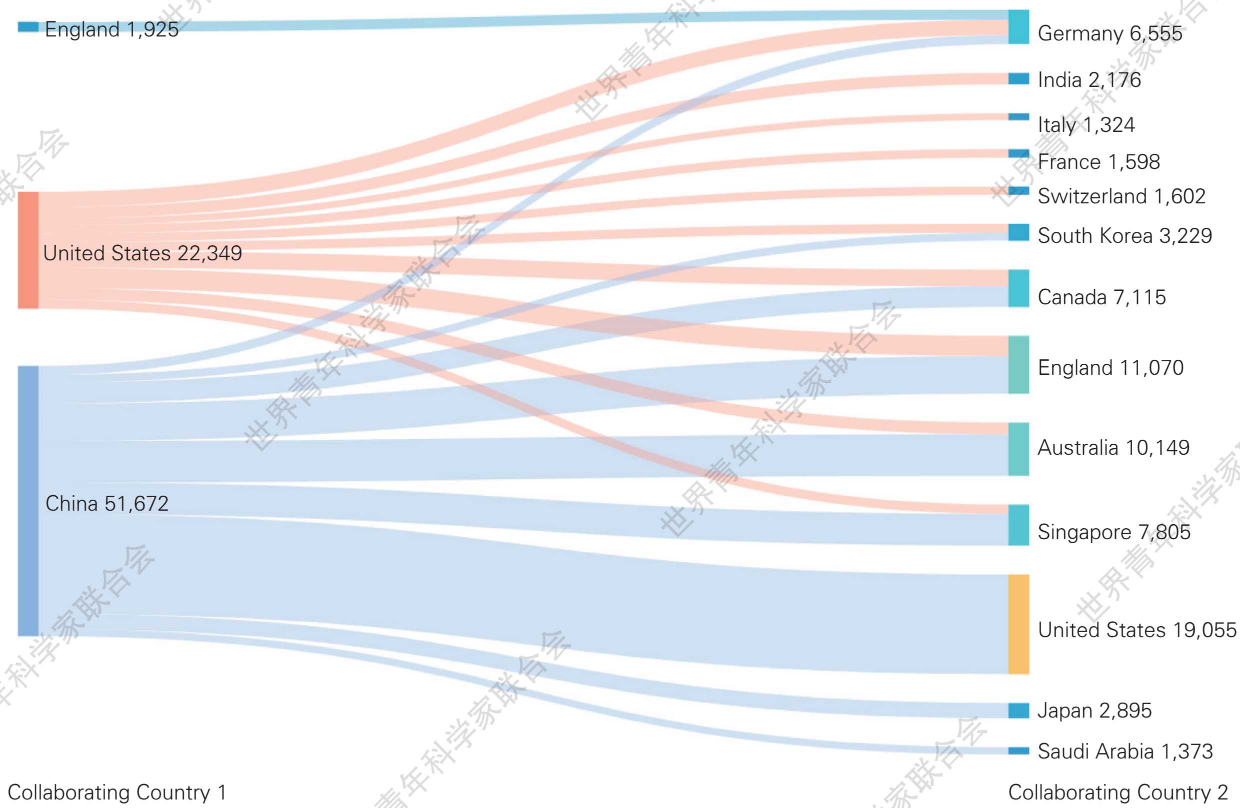


Figure 7-1 Sankey Diagram of Cross-Country Co-Authorship Among Global Young AI Scientists

From a quantitative perspective, there are 19,055 jointly authored papers between Chinese and the United States researchers, significantly higher than any other country pair.

In addition to the United States, China also maintains large-scale collaborations with Australia, England, Singapore, and Canada. The United States, meanwhile, has a broader range of research partners — beyond its collaboration with China, it also engages extensively with England, Canada, Germany, Australia, India, Singapore, South Korea, Switzerland, France, and Italy.

Table 7-4 Top 20 Country/Region Pairs by Frequency of Cross-Country Co-Authorship Among Global Young AI Scientists

No.	Country 1	Country 2	Number of Collaborations
1	China	United States	19,055
2	China	Australia	7,931
3	China	England	7,196
4	China	Singapore	6,067
5	China	Canada	3,950
6	United States	England	3,874
7	United States	Canada	3,165

continued

No.	Country 1	Country 2	Number of Collaborations
8	United States	Germany	2,917
9	China	Japan	2,895
10	United States	Australia	2,218
11	United States	India	2,176
12	England	Germany	1,925
13	United States	Singapore	1,738
14	United States	South Korea	1,737
15	China	Germany	1,713
16	United States	Switzerland	1,602
17	United States	France	1,598
18	China	South Korea	1,492
19	China	Saudi Arabia	1,373
20	United States	Italy	1,324

For detailed reports, please contact World Association of Young Scientists
(email: secretariat@ways.science)