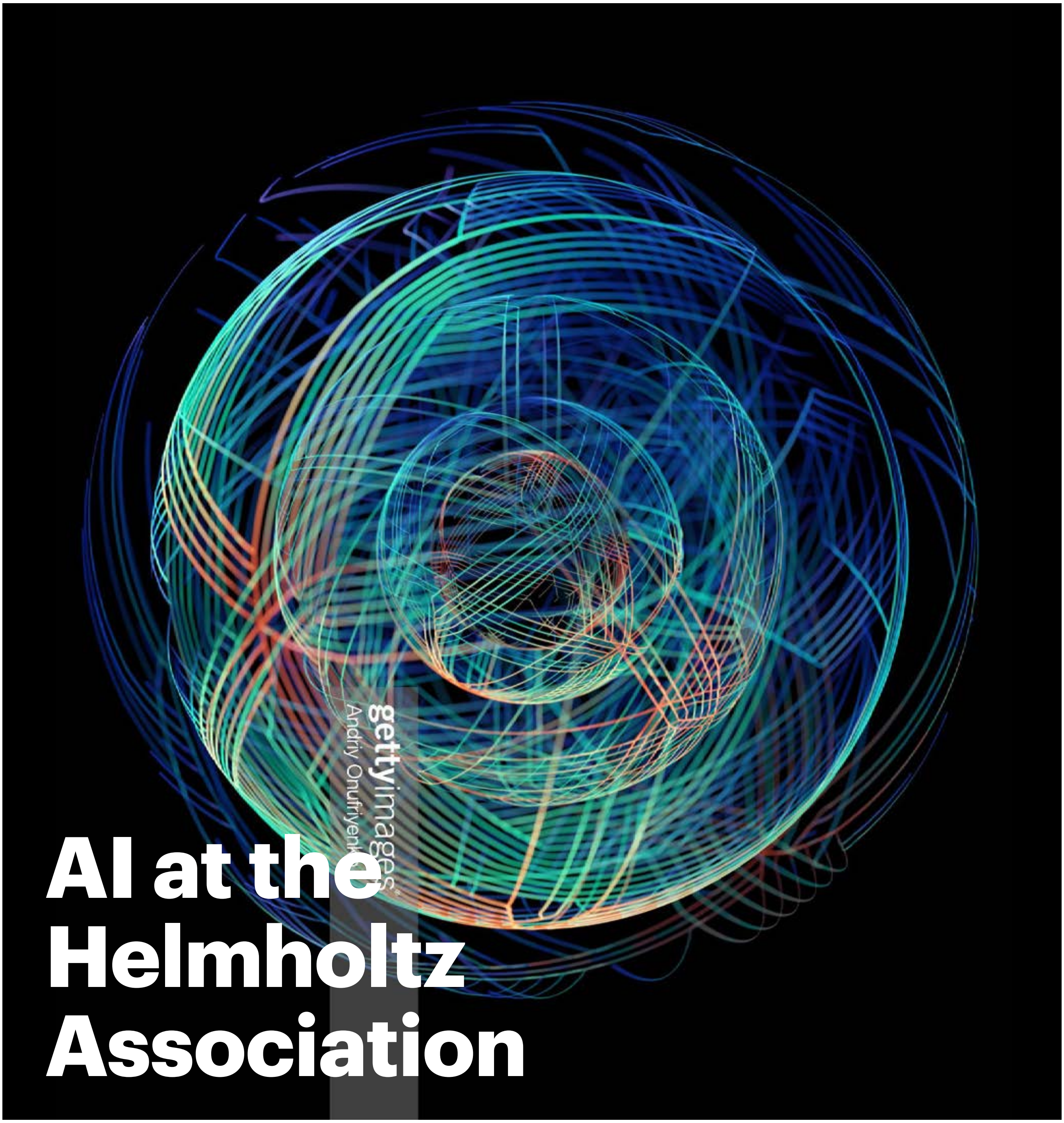
nature
strategy reports



AI at the Helmholtz Association

In partnership with

HELMHOLTZ

nature
research intelligence

January 2023

Contents

ABOUT THIS REPORT

This Nature Strategy Report was produced for Helmholtz AI by Nature Research Intelligence & Nature Research Custom Media, part of Springer Nature.

About Nature Research Intelligence

Nature Research Intelligence summarizes comprehensive research trends to give you relevant and trustworthy insights. You can quickly measure research success, stimulate discovery, uncover hidden connections, and set a research strategy to improve your organization's performance. Find out more at: www.nature.com/research-intelligence

Background
Artificial intelligence (AI) is one of the fastest-growing areas in scientific research. The amount of AI research published in academic journals and conference papers has grown about 10% each year over the past five years, including in top (first quartile) publications. AI is now used in most research disciplines. Publishing patterns indicate that global AI hubs, such as the Pan-Canadian AI Strategy, are starting to develop around certain geographic and institutional centres.

Project summary
Helmholtz AI, an association-wide, artificial-intelligence network for the Helmholtz Association, commissioned Nature Research Intelligence to produce a bibliometric report of the global AI landscape as well as a competitive analysis of the Helmholtz Association's performance among comparable research alliances and an in-depth study of their collaboration patterns.

Results
Global trends
- AI-related publications worldwide grew by almost 60% between 2017 and 2021.
- The largest amount of AI grant funding worldwide was in the medical and health sciences, while the largest number of grants was in core AI.
- Most AI patents globally were awarded in core AI and engineering.
- China is the country that has published the most publications in AI research, with Beijing being the largest global hotspot.

Publishing growth at the Helmholtz Association
- The association published 15,067 publications on AI between 2017 and mid-2022.
- The Helmholtz Association's AI-related papers increased at a compound annual growth rate (CAGR) of 12.6% between 2017 and 2021, which tracks closely with the global rate of 12.4%.
- More than a third of the Helmholtz Association's AI output has been published in first-quartile journals.
- About two-thirds of the Helmholtz Association's AI-related output has been published as open access.
- The Helmholtz Association's AI-related output has been well-received by the academic community, with an average of 15.8 citations per publication — second highest among the competitors considered in this report.
- About two thirds of Helmholtz Association's citations are from international sources.

Collaboration patterns
- AI papers written by Helmholtz Association researchers with international collaborators tended to have higher average citation counts and Altmetric scores than those with no collaborators or domestic collaborators only.
- The top-five international collaborators on AI papers for the Helmholtz Association are:
  1. Harvard University (USA)
  2. University College London (UK)
  3. ETH Zurich (Switzerland)
  4. University of Oxford (UK)
  5. University of Manchester (UK)
- The top-five domestic collaborators on AI papers for the Helmholtz Association are:
  1. Technical University of Munich
  2. RWTH Aachen University
  3. Heidelberg University
  4. Technical University of Dresden
  5. Ludwig-Maximilians-Universität München

Recommendations
- To stay aligned with the global push towards open access, Helmholtz Association should consider increasing the proportion of publications they publish in open-access journals.
- Helmholtz Association could improve their Altmetric performance with increased science communication activities aimed at the public and the popular media.
- Helmholtz Association researchers could make better use of the resources and expertise within the association by increasing intra-association collaboration.

Keeping an eye on AI

An analysis of the global artificial intelligence (AI) landscape using bibliometric, funding and patent data to identify trends in AI research including where the most effort and money is expended and on what areas.

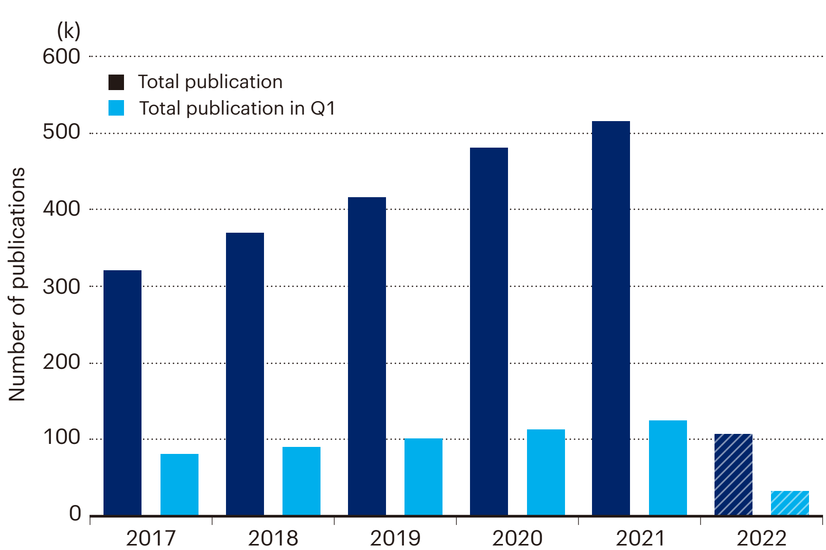
Image credit text/Getty

Key highlights

- The total number of AI-related publications worldwide grew by almost 60% between 2017 and 2021.
- The largest amount of AI grant funding worldwide was in medical and health sciences, while the largest number of grants was in core AI.
- Most AI patents globally were awarded in core AI and engineering.
- China is the country that has published the most publications in AI research, with Beijing being the largest global hotspot.

Figure 1.1

Global yearly publications. Cross-hatching has been used to indicate the data from 2022 is not from a complete year.



1.1 Growth of AI publications

Keywords and fields of research searches were used to extract AI-related publications from the Dimensions database from Digital Science for journal and conference papers published between January 2017 and June 2022.

Figure 1.1 shows that the total number of AI-related articles recorded globally grew by 59.8% over this five-year period (from 322,333 in 2017 to 515,182 in 2021), with a CAGR of 12.4%.

A range of metrics based on publication output is commonly used to measure and evaluate the impact of scientific activity on the international research community. These metrics are used to rank the quality, influence and relevance of each publication. Each subject category of journals is divided into four quartiles: Q1 to Q4, with the first quartile (Q1) consisting of the top 25% of journals in the Scimago Journal Ranking (SJR).

The five-year range 2017–2021 saw a 52.6% growth for AI research in the highest-ranked Q1 publications (from 81,772 in 2017 to 124,812 in 2021), with a CAGR of 11.2%.

This growth pattern has not been consistent, however. Between 2017 and 2021, the greatest year-on-year rise in AI-related publications occurred between 2019 and 2020, when total AI publications jumped by 15.7%, and Q1 AI publications rose by 14.2%. The lowest year-on-year rise was from 2020 to 2021, with 7.3% growth in total publications. This was not mirrored by the Q1 journals, whose lowest year-on-year growth was between 2018 and 2019.

1.2 Publications by research field

Looking at preset Field of Research (FOR) codes from Dimensions, in Figure 1.2 we can see that the most significant research output for AI has occurred in

Figure 1.2

AI publications in a broad selection of research fields.

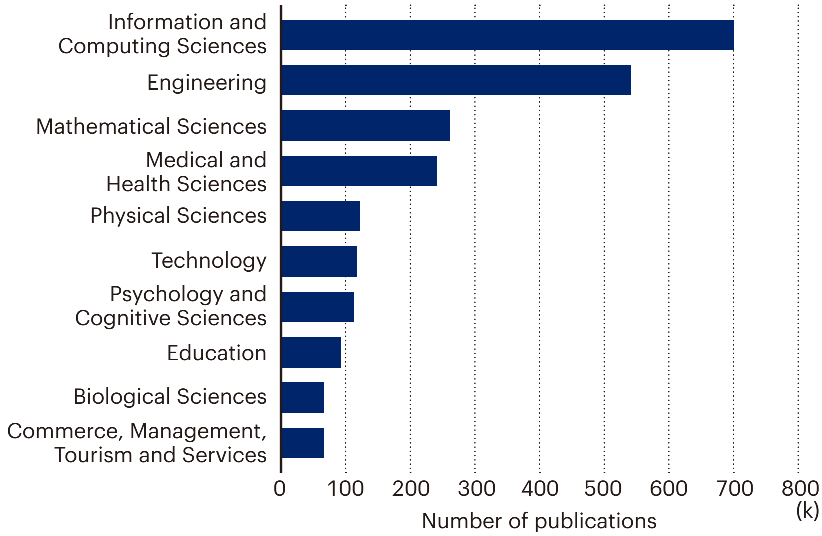
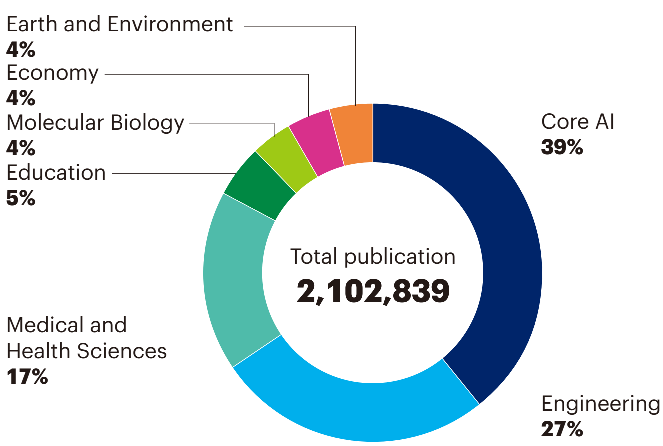
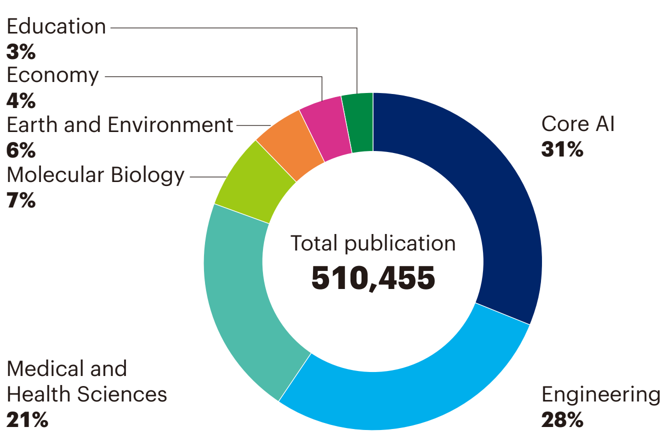


Figure 1.3

Total publications in AI research 2017–2021.



Total Q1 publications in AI research 2017–2021



the Information and Computing Sciences FOR, which accounts for more than 692,563 (37%) of total publications.

The most significant research output for AI was in the field of information and computing sciences, which accounted for more than 692,563 (37%) of total publications. Engineering is the next most-active research field, accounting for around 506,000 (27%) of publications globally, followed by mathematical sciences (11.6%) and medical and health sciences (9.4%).

These fields all reflect the most-important advances in AI listed in 'Gathering Strength, Gathering Storms', the 2021 report by the One Hundred Year Study on Artificial Intelligence (AI100) panel. This included a multitude of real-world applications on AI particularly computational (speech recognition, recommendation models), engineering (robotics and autonomous vehicles), and medical (identifying and diagnosing cancers and medical disorders and supporting clinical diagnoses).

While AI is still an emerging technology in the remaining seven fields (education; psychology and cognitive sciences; biological sciences; commerce, management, tourism and services; Earth sciences; chemical sciences; and economics), the publication count is already in the tens of thousands in each of these fields. As AI expertise within each domain develops, these publication rates are likely to continue to accelerate.

1.3 Major topics in AI

Nature Navigator[1] was used to collate this data into seven major topics, or 'subject areas', and then break it down further into eight subtopics, or 'concepts' as discussed in Section 1.4.

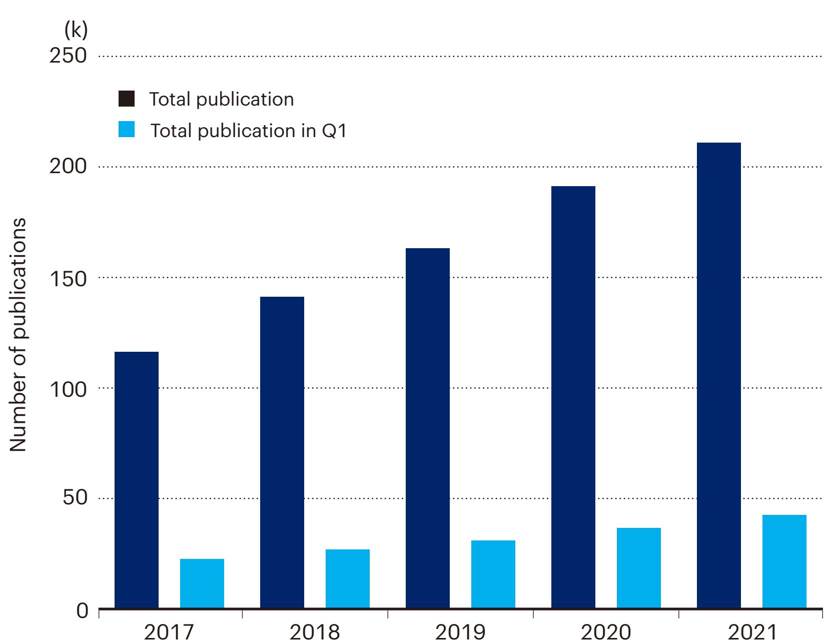
Figure 1.3 shows the subject area breakdowns for AI research, both for total publications and for Q1 journals. It reveals that:

- Core AI was the largest segment of AI research, accounting for 39% of all AI research published between 2017 and 2021, and around 31% of all Q1 publications in AI.
- Engineering AI research was the second largest segment, accounting for 27% of the total and 28% of Q1 AI publications
- In third place, was medical and health science, accounting for 17% of total AI publications and 21% of Q1 AI publications.
- The remaining publications consisted of AI research in molecular biology, Earth and environment science, economy, and education.

Between 2017 and 2021, as shown in Figure 1.4, the total number of core AI-related articles recorded globally grew by 81.5% (from 116,230 in 2017 to 210,927), with a CAGR of 16.1%. This period also saw 88.8% growth for core AI research featured in Q1 publications within each field of research (from 22,449 in 2017 to 42,390 articles in 2021), a CAGR of 17.2%.

Figure 1.4

Global yearly publications in Core AI.



1.4 Most-used concepts in AI

Figure 1.5 plots eight highly used concepts[2] across all AI publications between 2017 and 2022 in terms of rank.

The top concepts used in 2021 were:

1. Artificial neural network
2. Learning algorithm
3. Deep learning
4. Optimization algorithm
5. Computational complexity
6. Vector machine
7. Sensor networks
8. Aerial vehicles

Of these, 'neural network' appeared in more than 145,000 publications between 2017 and 2021. The most marked growth was 'deep learning', which had a CAGR of 72.6%, rising from eighth in 2017 to third in 2021. In comparison, 'sensor networks' had the lowest CAGR at 2.7%, decreasing in rank from sixth to eighth between 2017 and 2021.

'Neural network' and 'deep learning' use multiple layers of processing to simulate the behaviour of the human brain when it learns from large amounts of diverse data, such as during natural language processing and interpreting visual and auditory input. These methodologies are now applied to analyse multivariate pattern recognition across very large datasets in applications as diverse as medical imaging, autonomous vehicles, and agriculture.

1.5 Grants for AI research

More than 280,000 grants worth a total of 128 billion Euro were awarded for AI research worldwide between 2017 and 2021[3].

The analysis of grant funding in Figure 1.6 shows that while the subject area of core AI attracted the largest number of grants (76,654), the field of medical and health sciences received the highest amount of grant funding (€38.34 billion), about 30% of the total money awarded.

After core AI, the grant count was highest in medical and health sciences at 72,963 grants, followed by engineering at 62,568 grants. Molecular biology, education, economy, and Earth and environmental sciences each attracted 31,000 grants or fewer.

Several interesting patterns emerged from an analysis comparing the total monetary value of grants awarded by field, with the CAGR of funding in each field between 2017 and 2020[4]. The three fields where the monetary value of grant funding over the four-year period has climbed are core AI at 12.15%, Economy at 3.01%, and engineering at 1.05%. Conversely, grant funding by monetary value declined in Medical and Health Sciences at −1.6%, molecular biology at −8.03%, Earth and environmental sciences at −3.32%, and education at −9.04%.

Figure 1.5

Most-used terms in AI research

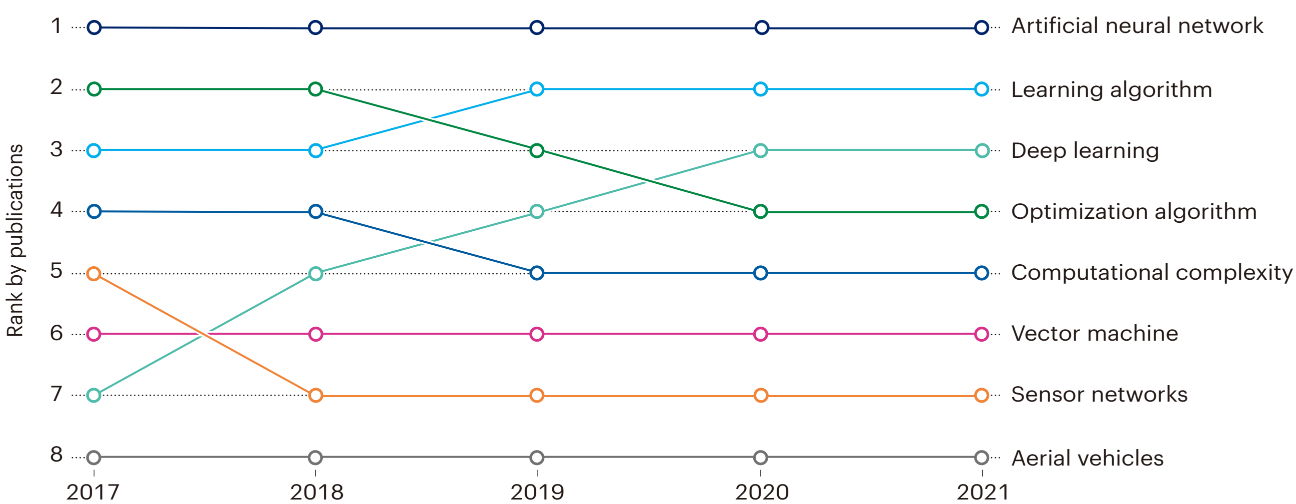
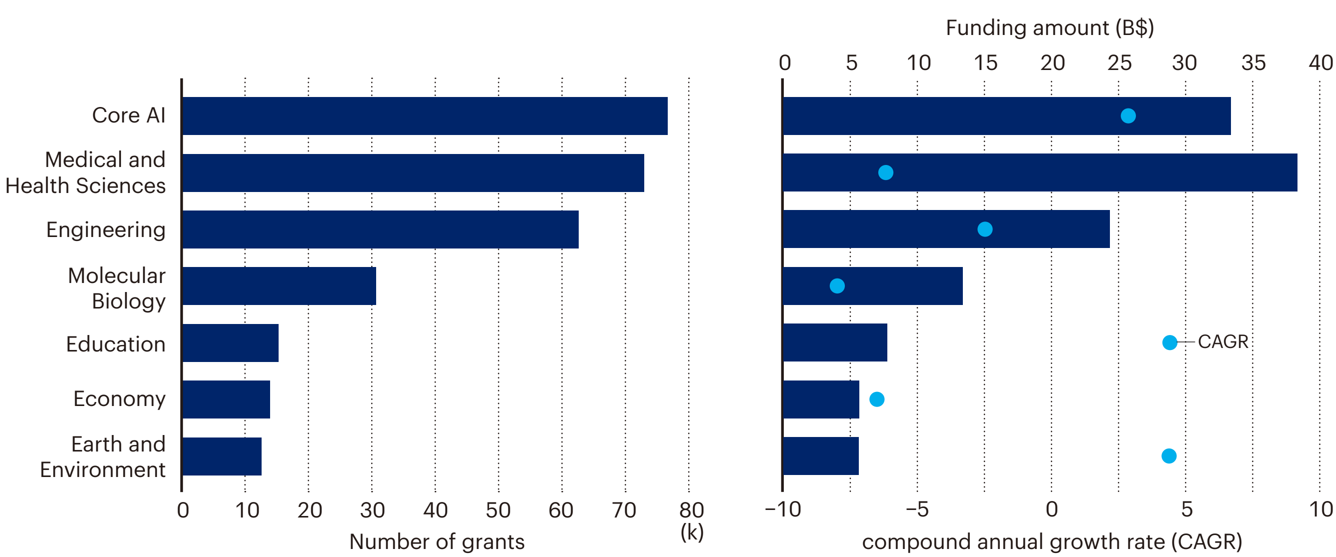


Figure 1.6

AI-related grants awarded across differernt research fields

Funding and CAGR across different research fields



1.6 AI patents by subject area

Figure 1.7 shows the number of AI-related patents awarded in core AI, engineering, medical and health sciences, Earth and environmental sciences, and molecular biology[5].

The fields of core AI and engineering had the largest number of AI-related patents, with engineering having a peak of more than 88,000 in 2019 and core AI having a peak of almost 95,000 in 2020. The number of patents awarded in the Earth and environmental sector was consistently low, between 9,000–10,000 per year, whereas medical and health sciences and molecular biology consistently hovered around 14,000–17,000 and 8,000 per year, respectively.

Analysis of AI-related patents is likely to become more complicated as the field evolves. A recent commentary in *Nature*[5], 'Artificial intelligence is breaking patent law', discussed changes that may be required in intellectual property law once AIs are able to invent tools themselves. It noted that AI has already been used to aid vaccine development, drug design, materials discovery, space technology and ship design.

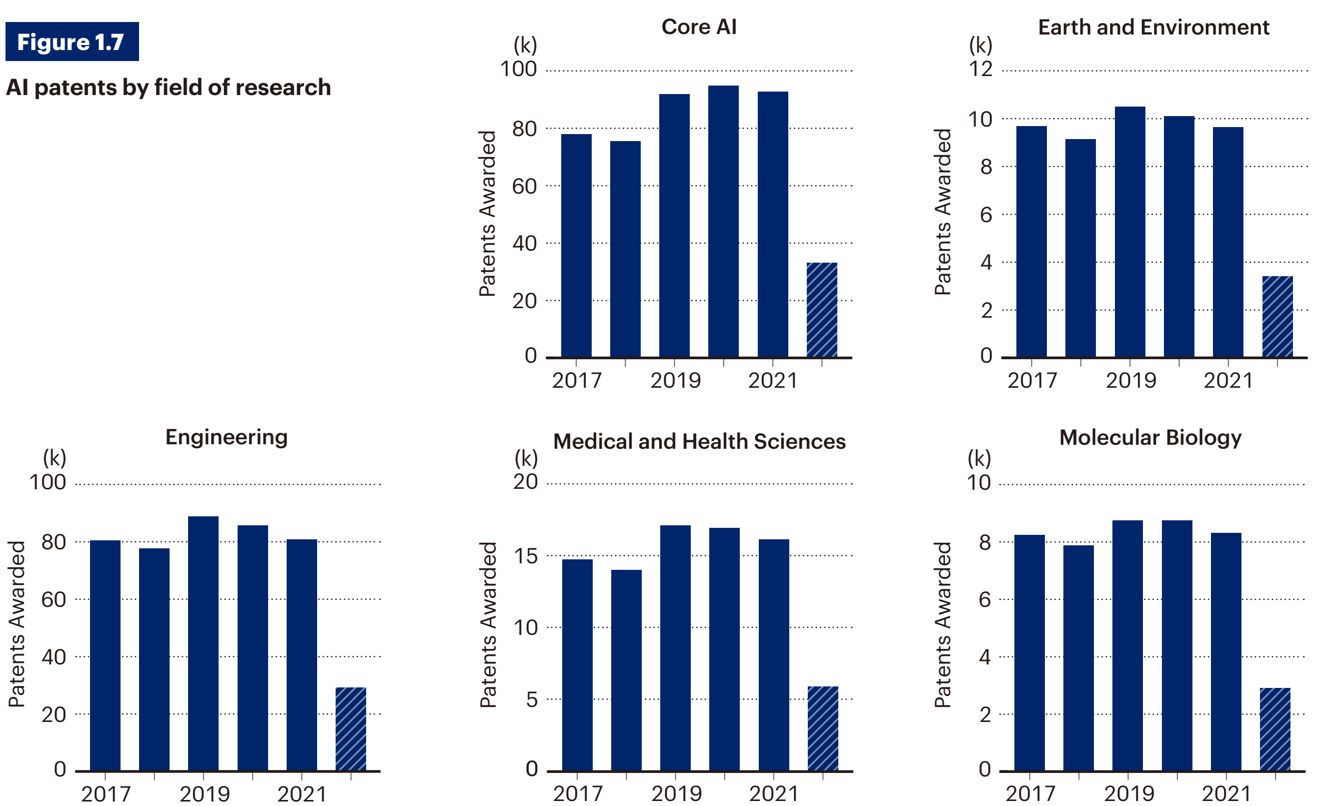
1.7 Global AI hotspots

In terms of countries, China is the largest contributor to AI publications globally, accounting for around 800,000 AI-research publications, followed by the USA, which has contributed almost 650,000 AI publications, as shown in Figure 1.8

The next largest AI-research hotspots are the UK and India, which each produce almost 200,000 AI-research publications. There has also been substantial AI research in Germany and Italy with around

Figure 1.7

AI patents by field of research



Global trends

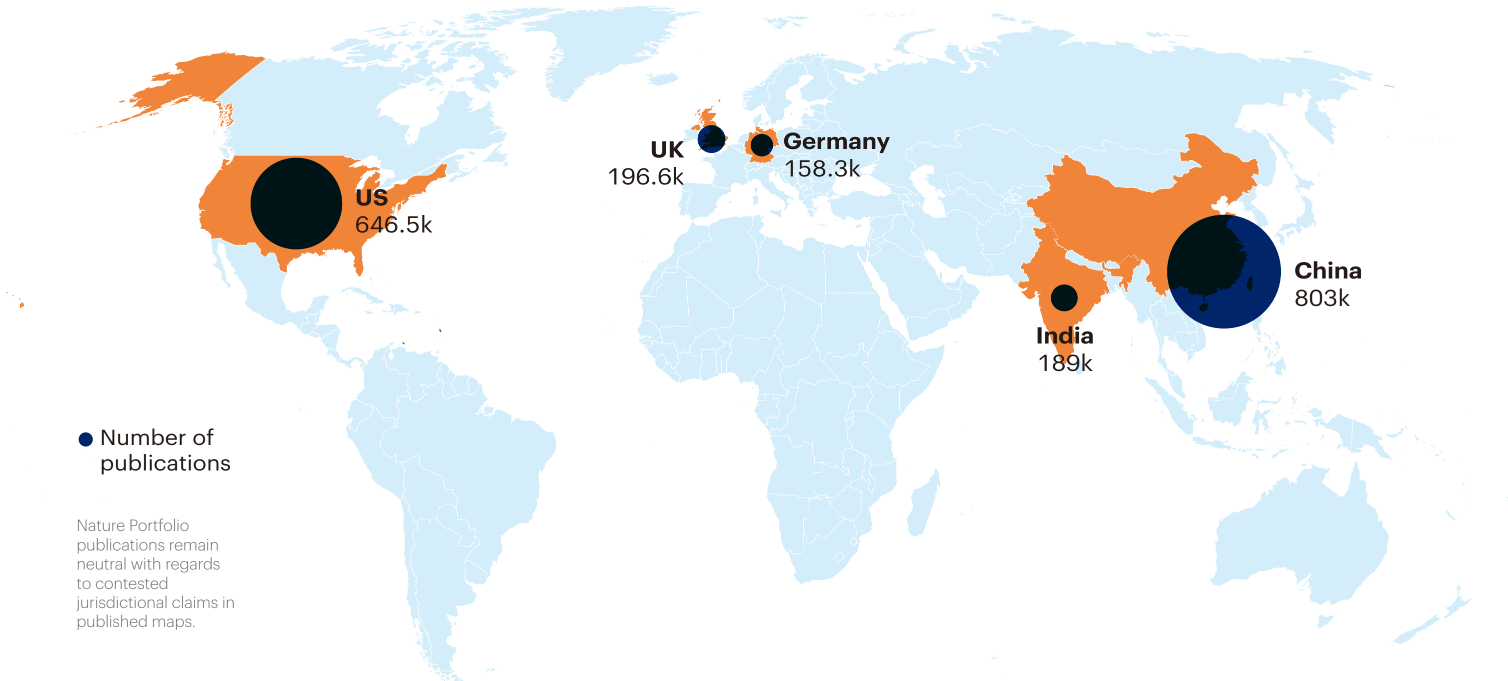


Figure 1.8

AI research 'hotspots' by publication output:

Global

City	Publications
Beijing	172,450
Shanghai	56,781
Nanjing	54,655
London	48,011
Xi'an	44,798
Seoul	43,575
Wuhan	38,842
Tehran	35,177
Tokyo	33,750
Guangzhou	33,011
Chengdu	30,545
Hong Kong	27,375
Hangzhou	26,988
Sydney	26,680
Singapore	26,065
Changsha	24,915
Melbourne	24,604
Paris	24,219
Cambridge	24,179
New York	24,087

Germany

Munich	16,544
Berlin	13,733
Heidelberg	5,654
Aachen	5,421
Hamburg	5,251

158,000 and 116,000 AI publications, respectively, followed by Canada and Australia, who have each published more than 100,000 publications. Other global AI-research hotspots with around 90,000 publications include France (97,000), Japan (95,000) and South Korea (89,000).

Within these 'global AI hotspot' countries, several cities have become hubs for AI research. In Germany, for example, the top-five cities for AI research are Munich, Berlin, Heidelberg, Aachen and Hamburg.

The top-three cities by publication count are all in China: Beijing (172,450 publications), Shanghai (56,781) and Nanjing (54,655). Together, these three cities account for more than 250,000 AI-research publications.

Next on the list is London (UK), with 48,011 publications, followed by Xian (China), and then Seoul (South Korea).

Ten of China's major cities are hubs for AI research. In contrast, the only US cities that appear in the top-20 AI research cities are Cambridge (MA) and New York (NY). This suggests that while the USA is the second largest publisher of AI research overall, the spread of research effort in the USA is dispersed more widely between its individual cities.

Australia is the only other country with more than one city in the top 20 cities for global AI-research output, with Sydney at 14th and Melbourne at 17th. This may reflect the establishment in 2019 of an AI Roadmap and national AI strategy by CSIRO, Australia's national science agency.

Notes & References

1. See Appendix.
2. Selected by Helmholtz AI
3. Based on available data. For example, funding and grant information from China was unavailable. More details are given in the Appendix.
4. Note that we considered the time period 2017–2020 to avoid potential time lags in grant data; grant programmes included in this count can be seen here: https://sn-insights.dimensions.ai/datasources
5. Patents awarded in the economy and education sectors (approximately 1,200–3,300 per year) have been omitted from this discussion, as patents typically recognize technological advances, rather than those in the humanities.

6. George, A. & Walsh, T. *Nature* **605**, 616 (2022).

CASE STUDY

EARTH OBSERVATION

Sharpening sights on cities through AI

AI algorithms act as magnifying glasses on our world, enhancing satellite data to provide extremely detailed information about Earth's surface.

Observations of Earth's surface are used for many purposes, from predicting the weather to monitoring the conditions of forests, crops, volcanoes and glaciers. However, the datasets acquired by satellites, aerial photography and ground surveys are often patchy and limited in resolution.

Researchers including the team led by Xiaoxiang Zhu at the Technical University of Munich, in co-operation with the German Aerospace Center, are employing sophisticated machine learning algorithms to automatically process and enhance Earth observation data. Their efforts are producing impressively detailed digital models that assist scientists and city planners.

"Satellites take many years of development before launching and operating for 10-20 years," explains Zhu's colleague Richard Bamler. "During this time, people may want to do different things from what the satellite was originally designed for. Our algorithms enable new information to be extracted from existing data, for example by removing noise and increasing resolution."

Machine learning is increasingly important for analyzing data from networks of miniature 'CubeSat' satellites launched by new space companies like Planet Labs. "Instead of building heavy, high-quality satellites like ESA (the European Space Agency), these companies build many small satellites from off-the-shelf components," says Bamler. "They have limited resolution, but cover the whole Earth, and the data are much cheaper, or even free."

Combining data from small and large satellites enabled Zhu to take on her most ambitious task – mapping global urbanization, including every building in every city in the world with over 300,000 people. Zhu and co-workers used radar and spectral data from ESA satellites to calibrate the wider-scale observations from CubeSats and train their algorithms to pick out buildings. They even sourced photographs from social media and information from local text messages, which helped their algorithms to characterize the functions of different buildings.

Our algorithms enable new information to be extracted from existing data, for example by removing noise and increasing resolution.

This work has allowed the team to reveal some remarkable details, such as how Berlin Central Station expands by several centimetres on hot days. Another important application is extracting individual road markings from aerial photos to guide self-driving cars.

The maps generated are also useful for city planners to identify areas at risk of becoming dangerous 'urban heat islands'. This requires classifying urban regions into 17 categories based on factors like building density and plants. Doing this task manually is laborious, as Bamler explains:

"We chose 42 cities (a number inspired by the Hitchhiker's Guide to the Galaxy!) in 10 different cultural zones where we had high-resolution images. Then we put 15 students in a room for a week with lots of pizza, manually classifying over 400,000 patches of land!"

In the future, researchers will save time by doing these calibration tasks through 'self-supervised learning', where algorithms learn to identify key features themselves. AI will even be deployed on satellites, for example training them to identify when clouds are in the way and discard images to save on data storage.

"Helmholtz AI is a tremendous initiative, enabling researchers to join forces from many different fields so that we can all get good information from data," says Bamler. "The computer science community brought a new culture into Earth observation, with a very fast development cycle and making code open source. This way, geoinformation is democratized."

A closer look at AI research at the Helmholtz Association

An analysis of the Helmholtz Association's AI research — including their strengths and weaknesses, and where they sit in relation to comparable research collectives.

Key highlights

- The Helmholtz Association published 15,067 AI publications between 2017 and mid-2022.
- The Helmholtz Association's AI-related papers increased at a CAGR of 12.6% between 2017 and 2021, which tracks closely to the global rate of 12.4%.
- Over a third of the Helmholtz Association's AI output was published in Q1 journals.
- About two-thirds of the Helmholtz Association's AI output was published as open access. To stay aligned with the global push towards open access, Helmholtz Association should consider increasing this percentage.
- The Helmholtz Association's AI-related output is well received by the academic community, with an average of 15.8 citations per publication, the second highest among the competitors considered in this section.
- Approximately two-thirds of the Helmholtz Association's citations are from international sources.
- The percentage of Helmholtz Association publications with high Altmetric scores could be improved through increased science-communication activities aimed at the public and popular media.

2.1 Introduction

This section looks at publishing growth, at the Helmholtz Association and among eight comparable research collectives — six chosen by Helmholtz AI and two chosen by Nature Research Intelligence — using bibliographic data from 2017 to mid-2022. The Helmholtz Association's strengths and gaps relative to global patterns are identified, and its performance in comparison to these research collectives are examined.

The competitor research collectives considered here are:

- **French National Centre for Scientific Research (CNRS):** Ranked fourth in the Nature Index 2022 Annual Tables, CNRS is France's national public scientific research organization.
- ***Indian Institutes of Technology (IITs):** Twenty-three public technical institutions located across India, under the auspices of the Ministry of Education.
- **Israel Group, Israel:** This is not an official research collective. We have combined the output of the following research entities and considered them collectively: Ben-Gurion University of the Negev, Tel Aviv University, Hebrew University of Jerusalem, Israel Science Foundation, Technion – Israel Institute of Technology, and Weizmann Institute of Science.

- **Ivy League, USA:** Eight private universities located in the northeast of the USA: Princeton University, Harvard University, Yale University, University of Pennsylvania, Dartmouth College, Brown University, Cornell University, and Columbia University.
- **N8 Research Partnership, UK:** A collaboration of the eight most research-intensive universities in northern England: Durham, Lancaster, Leeds, Liverpool, Manchester, Newcastle, Sheffield and York.
- **Pan-Canadian AI Strategy:** This grouping includes researchers from the Vector Institute for Artificial Intelligence; Mila; the Alberta Machine Intelligence Institute (AMII); and the Canadian Institute for Advanced Research (CIFAR — a Canadian-based global scientific research organization funded by individuals, foundations and corporations); and federal and provincial governments in Canada.
- ***RIKEN, Japan:** The largest research institution in Japan with main campuses in Wako, Tsukuba, Yokohama, Kobe and Harima.
- **Tsinghua University, China:** A leading public university located in Beijing, China. Ranked 16th in the 2022 Times Higher Education World University Rankings, and 11th in the Nature Index 2022 tables.

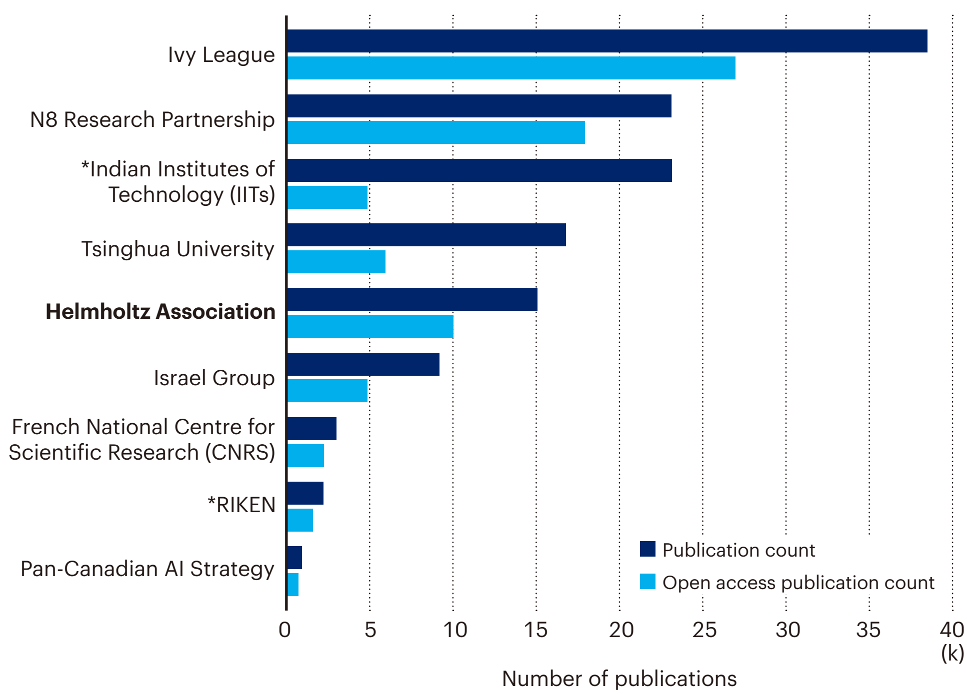
The two collectives with an asterisk preceding their names (IITs and RIKEN) were selected by Nature Research Intelligence; the other six were selected by Helmholtz AI.

2.2 Publication growth and research strengths at Helmholtz centres

Across all 18 Helmholtz centres, the number of total publications in AI rose from 2,218 in 2017 to 3,564 in 2021, a CAGR of 12.6%, which tracks closely against the overall rise in publications in AI globally. The Q1 AI output published by Helmholtz centres rose from 745 in 2017 to 1,249 in 2021 at a CAGR of 13.8%, which is higher than the global CAGR of 11.8%.

Figure 2.1

Total and Open Access publications for each research collective between 2017 and mid 2022.



As observable in Figure 2.1, the centres collectively produced 15,067 AI publications between 2017 to mid-2022, with 5,253 of these, or about 35%, published in Q1 ranked journals.

The association's international ranking for total AI publications has remained stable over the five-year period, whereas the proportion of its AI research which appeared in Q1 publications rose from 33.6% in 2017 to 35% in 2021.

Publication outputs give good insights into the subject field strengths of each Helmholtz centre, as shown in Figure 2.2 below.

2.3 Open access publications

In terms of the total number of AI-related publications, Helmholtz Association ranks fifth in this selection of competitors, with 15,067 publications, as seen in Figure 2.1. The Ivy League leads with more than 38,000 papers, followed by the IITs with 23,139. RIKEN (2,229 papers) and the Pan-Canadian AI strategy (922 papers) have the lowest output.

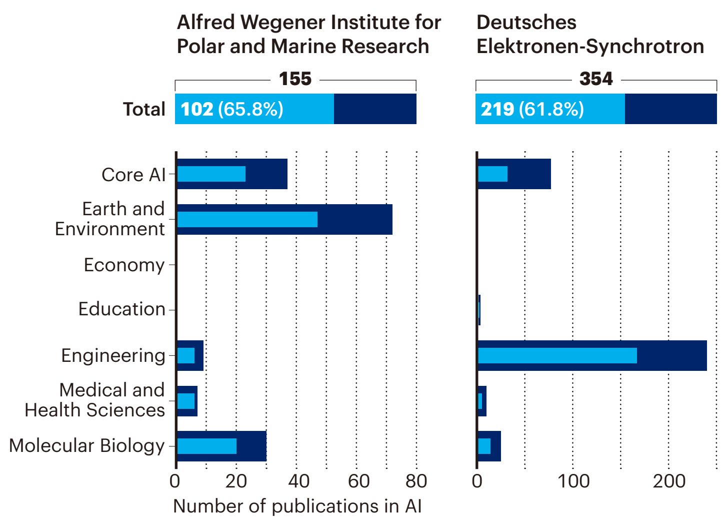
Comparing the total number of publications in AI-related research with the number of open-access (OA) publications reveals that (→ cont. on pg.14)

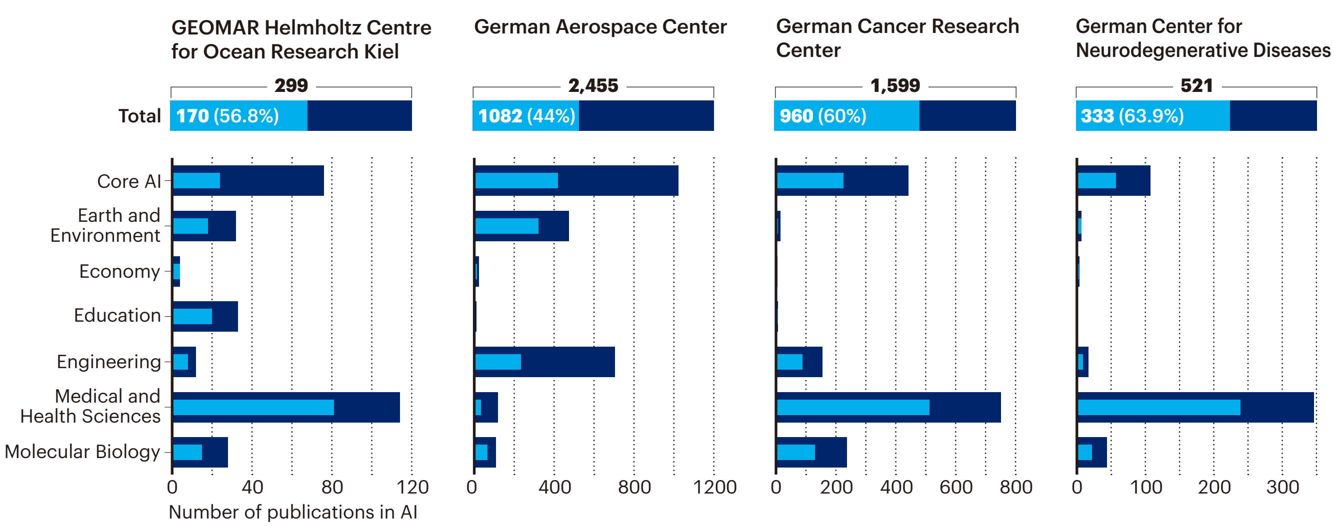
Figure 2.2

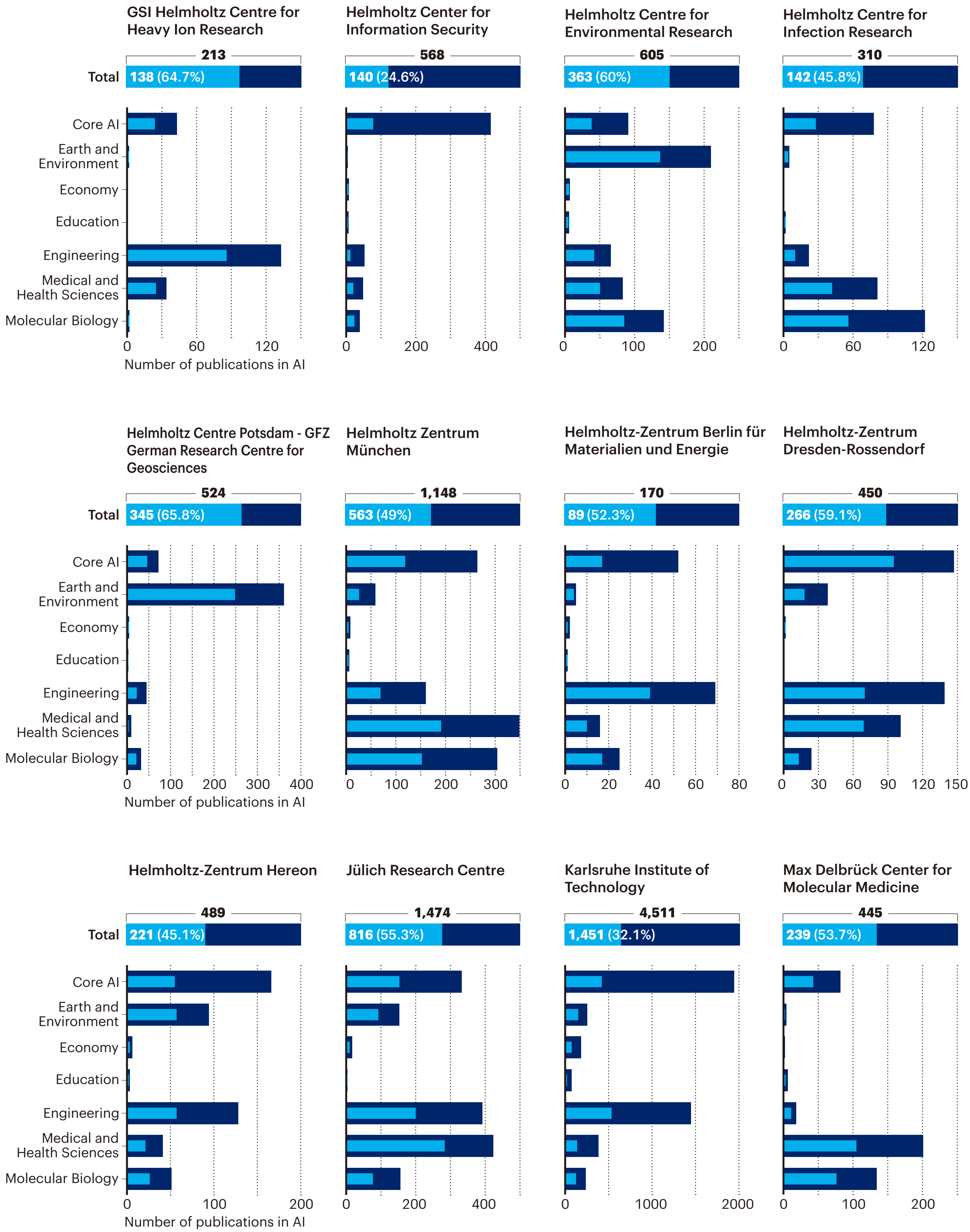
Where do Helmholtz Association researchers publish?

The subject area breakdown for each of the 18 Helmholtz centres' total and Q1 publications. The percentage indicates how much of an institute's output was in Q1 publications.

■ Total publications
■ Q1 publications



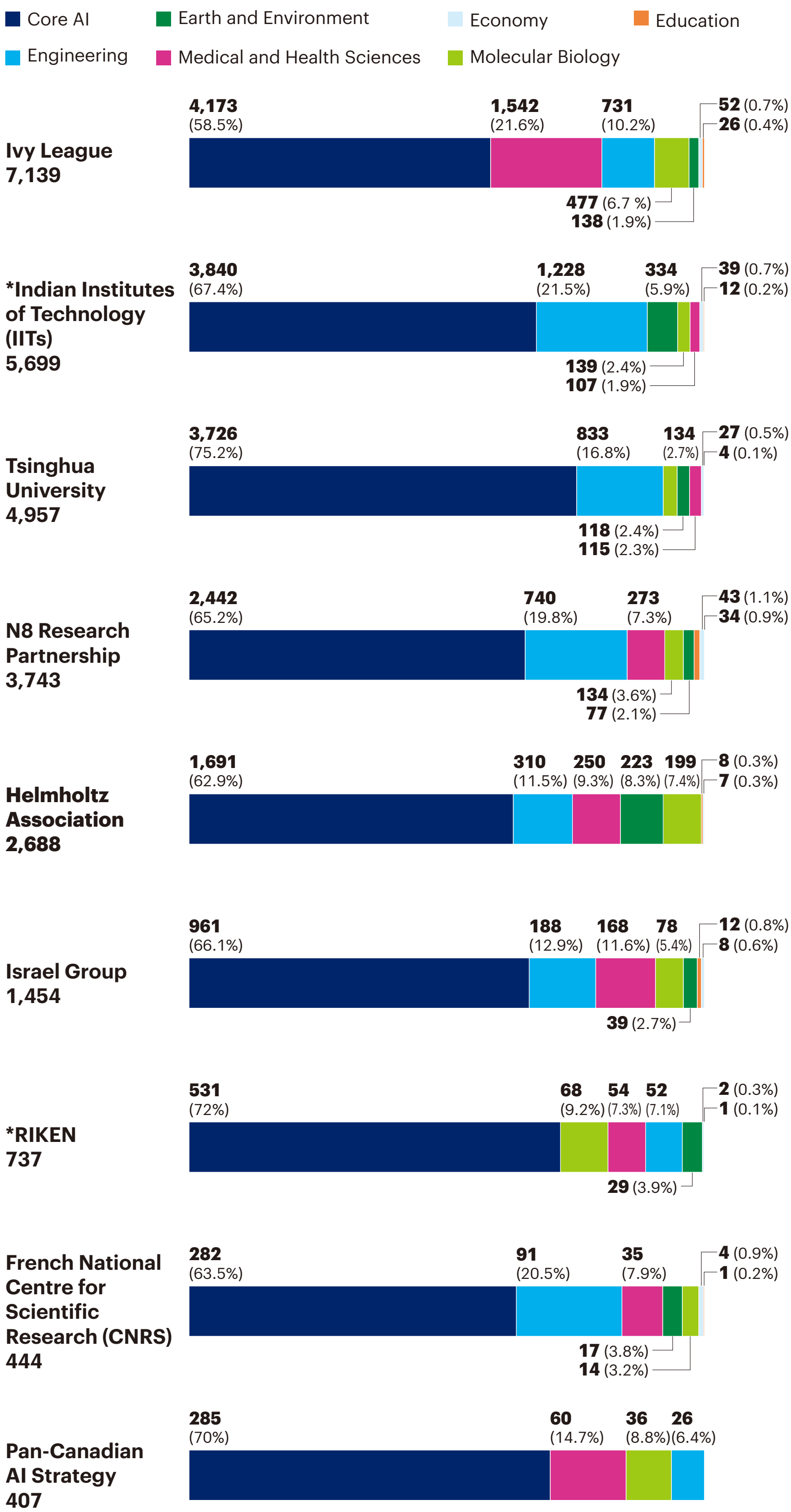



GSI Helmholtz Centre for Heavy Ion Research
213
Total 138 (64.7%)
Helmholtz Center for Information Security
568
140 (24.6%)
Helmholtz Centre for Environmental Research
605
363 (60%)
Helmholtz Centre for Infection Research
310
142 (45.8%)
Core AI
Earth and Environment
Economy
Education
Engineering
Medical and Health Sciences
Molecular Biology
0 60 120
0 200 400
0 100 200
0 60 120
Number of publications in AI
Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences
524
Total 345 (65.8%)
Helmholtz Zentrum München
1,148
563 (49%)
Helmholtz-Zentrum Berlin für Materialien und Energie
170
89 (52.3%)
Helmholtz-Zentrum Dresden-Rossendorf
450
266 (59.1%)
0 100 200 300 400
0 100 200 300
0 20 40 60 80
0 30 60 90 120 150
Number of publications in AI
Helmholtz-Zentrum Hereon
489
Total 221 (45.1%)
Jülich Research Centre
1,474
816 (55.3%)
Karlsruhe Institute of Technology
4,511
1,451 (32.1%)
Max Delbrück Center for Molecular Medicine
445
239 (53.7%)
0 100 200
0 200 400
0 1000 2000
0 100 200
Number of publications in AI

Publishing growth

Figure 2.3

Breakdown of AI publications by subject area for the research collectives.



during 2017–2022, 66% of publications by Helmholtz centres were published on OA platforms (10,014 OA publications out of a total of 15,067). This was the sixth highest proportion of OA publications in the collectives considered.

In comparison, the Pan-Canadian AI strategy published the highest proportion of OA papers (79%), followed by the N8 research partnership (76%), CNRS (75%), RIKEN (71%), Ivy League (70%), Israel group (53%) and Tsinghua University (35%). The IITs had the lowest proportion of OA publications (21%). This could be due to the cost of publishing in OA journals; Siler *et al*. showed that the article-processing fees associated with some OA journals could exclude academics from developing countries[1]. However, this may change with the recent establishment of transformative agreements between Indian institutions and publishers such as Springer Nature in 2020[2].

Making research widely and readily accessible is not only in line with the Helmholtz Association's policy — the Helmholtz Association was among one of the first scientific organizations to sign the Berlin Declaration of Open Access — it is also conforming with the wider AI community, which values free and open access to knowledge.

As OA publications are more widely viewed and cited than those behind paywalls, they provide greater research visibility and enable researchers to conduct collaborative research on a global scale. This analysis indicates that, to remain globally competitive, it would benefit the Helmholtz Association to increase the proportion of articles published in open-access publications.

The top-ten open-access publications that Helmholtz Association publishes papers and preprints in are:

1. bioRxiv (a preprint server)
2. *Remote Sensing*

3. *Scientific Reports*
4. *Sensors*
5. *PLOS ONE*
6. *IEEE Transactions on Geoscience and Remote Sensing*
7. *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*
8. *Journal of Instrumentation*
9. *Fusion Engineering and Design*
10. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*

2.4 AI subject area and highly-used concepts

2.4.1 Subject area strengths

As expected, in Figure 2.3, core AI took the lion's share of AI output in all research collectives considered. Perhaps more interesting is the second slot, which was:

- Medical and health sciences for the Ivy League, comprising 22% of their AI output; 15% of the Pan-Canadian AI Strategy's output, and 7% of RIKEN's AI output.
- Engineering for N8 research partnership, comprising 20% of their AI output; 22% of IIT's AI output; 20% of CNRS' AI output; 17% of Tsinghua's AI output; 13% of Israel Group's AI output.

Helmholtz Association:
Core AI accounted for 63% of Helmholtz Association's AI output, followed by engineering and medical and health Sciences at 12% and 9%, respectively. There is also quite a strong showing in Earth and environmental sciences output (comprising 8% of Helmholtz's AI output), suggesting this might be a promising area to develop further.

Comparing the Helmholtz Association with global trends, we see that core AI share started behind the global volume but has gained over the past five years. Although gains have occurred, citation share shows a negative trend over time compared to global share with publishing in Q1 journals also showing a negative trend comparing to global data. Applied AI could be seen as a strategic driver for research and publishing. Engineering dropped in share compared to global outputs from 2017 to 2021, however, comparing global share of citations has seen an increase. This is a positive result.

2.4.2 Highly-used concepts

The bibliographic analysis has compared AI-related publication output across seven subject areas and eight concepts by all Helmholtz centre and the other research collectives, from 2017 to mid-2022.

The data below is based on research

Table 2.1

Research collectives with the highest publication output and those with the most papers in the top 10% cited for the 8 concepts considered.

Concept	#1 in publications	#1 in top 10% cited
Aerial vehicles	N8 Research Partnership (102)	N8 Research Partnership (25)
Artificial neural network	Ivy League (2,472)	Ivy League (273)
Computational complexity	Tsinghua University (408)	Ivy League (55)
Deep learning	Ivy League (1,198)	Ivy League (136)
Learning algorithm	Ivy League (1,764)	Ivy League (194)
Optimization algorithm	IITs (892)	IITs (107)
Sensor networks	IITs (279)	IITs (31)
Vector machines	IITs (456)	IITs (67)

output the collectives have published in eight highly used AI concepts[3], as outlined in Section 1.4. In order of frequency, these concepts are:

1. artificial neural network
2. learning algorithm
3. deep learning
4. optimization algorithm
5. computational complexity
6. vector machine
7. sensor networks
8. aerial vehicles

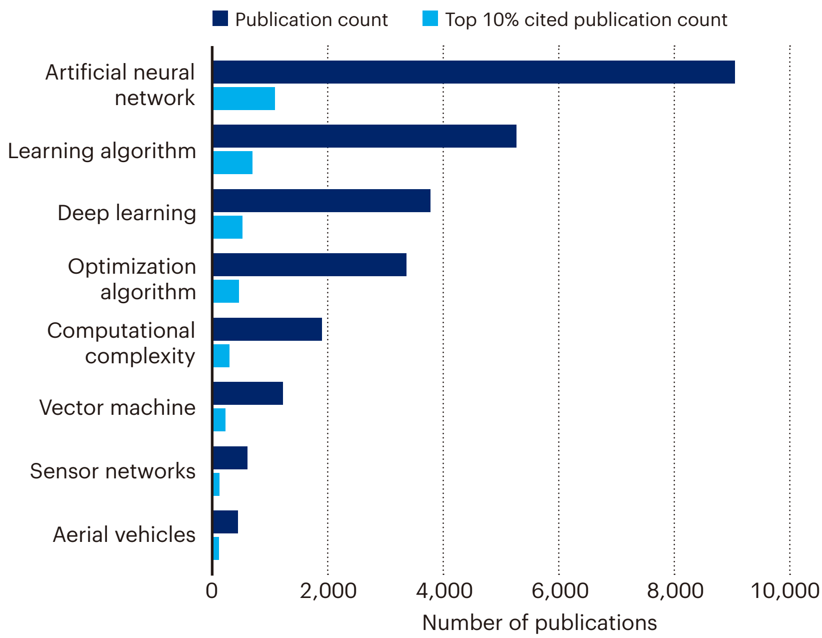
Their top-three concepts varied between artificial neural network, learning algorithm, optimization algorithm, and deep learning. The lowest ranked topic was sensor networks for the Ivy League, the N8 research group, CNRS, and the Pan-Canadian AI strategy grouping. RIKEN had sensor networks and aerial vehicles tied for last place, whereas Tsinghua University, the IITs, and the Israel group had aerial vehicles as the topic with the least number of publications. Approximately 12–30% of the hot-topic papers for these collectives were in the top 10% of cited papers.

Looking more closely at the concepts, Table 2.1 presents the research collectives with the highest publication output and those with the most papers in the top 10% cited:

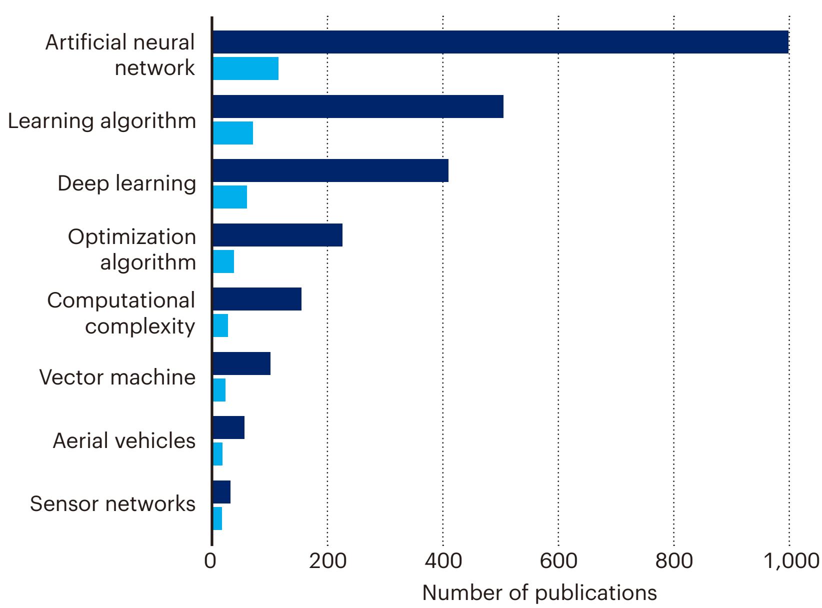
Table 2.1 shows that the IITs and the Ivy League dominate both the number of publications and the top 10% cited list. This suggests potential collaboration opportunities that the Helmholtz Association may be overlooking — for example, collaborative partnerships between the IITs and Helmholtz centres are not common, yet the IITs are strong in optimization algorithms, sensor networks and vector machines. If the Helmholtz Association plans to pursue research in these areas, collaboration with the IITs may strengthen their competitive advantage.

Figure 2.5

Hot concept papers



Helmholtz Association



Helmholtz Association:

Between 2017 and mid-2022, the Helmholtz Association published 2,482 papers on highly used concepts, of which about 15% were among the top 10% cited. Within these papers, their highest output was in the subject areas of core AI research, at almost two-thirds, followed by engineering (11%) and medical and health science (9%).

Their top-three concepts were artificial neural networks, learning algorithm and deep learning — in line with global trends discussed in Section 1. The lowest value is for sensor networks; however, more than half of these papers are in the top 10% cited. The next category with the most 'bang for the buck' is aerial vehicles, with 32% being in the top 10% cited, the rest varied between 14% and 23%.

From 2017 to 2021, the Helmholtz Association increased its output in all hot-topic areas except for computational complexity and sensor networks.

Possible steps for improving the Helmholtz Association's performance in these hot-topic areas could include greater access to AI experts, integration of expertise, and talent recruitment and retention.

2.5 Scholarly impact

This analysis compares the total number of citations for AI-related publications by each of the eight major research collectives from 2017 to mid-2022.

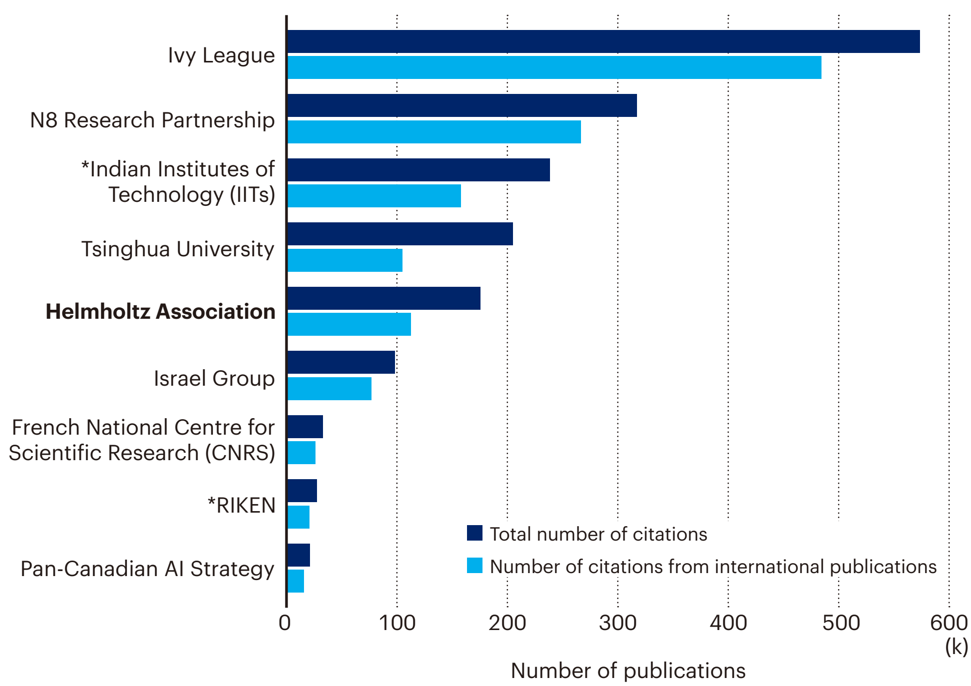
The Ivy League was by far the leading institutional group in terms of citation count, with more than 573,000 citations of their AI-related publications over the analysis period. The next-most cited institutional collective, N8 Research Partnership, had around 316,941 citations during this time. The Helmholtz Association had the third-highest number of citations (238,224).

The Pan-Canadian AI strategy leads the group in terms of average citation per publication (23), followed by Helmholtz Association (15.8), the Ivy League (14.9) and N8 Research Partnership (13.7). This indicates that research by the Helmholtz Association is well received by the academic community.

However, about one third of the Helmholtz Association's citations are from domestic sources, which is exceeded only by Tsinghua University's and the IITs' domestic citations (48.9% and 35.7%, respectively). For the rest of the competitor collectives, domestic citations contribute between 15% and 26% of sources. This indicates that the Helmholtz Association may need to improve their international outreach and collaboration efforts in the academic community.

Figure 2.6

Total number of citations for AI-related publications by each of the eight major research collectives from 2017 to mid-2022.



	No. of articles with a high Altmetric score						% of articles with a high Altmetric score					
	2017	2018	2019	2020	2021	2022	2017	2018	2019	2020	2021	2022
French National Centre for Scientific Research (CNRS)	24	25	22	42	36	10	13.19	14.29	11.76	15.05	14.57	15.15
Helmholtz Association	136	171	208	249	254	40	13.84	15.05	14.44	14.99	14.11	11.40
*Indian Institutes of Technology (IITs)	10	18	14	23	46	8	2.37	3.61	2.63	2.94	4.79	2.93
Israel Group	64	100	91	138	150	26	10.98	15.04	13.58	15.74	15.71	12.15
Ivy League	816	1023	1117	1400	1376	229	22.25	24.00	23.22	23.47	21.51	19.13
N8 Research Partnership	273	328	363	425	443	59	15.37	16.93	16.48	16.78	16.03	9.95
Pan-Canadian AI Strategy	12	27	47	64	67	8	29.27	41.54	40.52	36.78	29.39	26.67
*RIKEN	27	24	40	42	48	11	18.12	12.37	15.56	14.24	14.50	14.47
Tsinghua University	28	44	39	44	51	8	5.74	7.79	6.55	6.34	6.11	3.13

Table 2.2

The number and percentage of articles with Altmetric scores of 20 or higher for the nine research collectives considered.

2.6 Impact in the wider community

The Altmetric score measures the reach of a published scholarly work weighted across a wide range of channels in the broader public sphere including media, social media, websites, article page views and downloads.

The score is influenced by several factors, including the circulation of a journal and the context of the publication relative to similar articles. An Altmetric score complements, rather than replaces, traditional methods of measuring the impact of research via citation counts, and it signals the impact of research more rapidly than citations from subsequent research.

In Table 2.2, we observe that as with citation count, in terms of the number of articles with high Altmetric scores[4] between 2017 and mid-2022, the Ivy League is a clear leader, followed by N8 research partnership and the Helmholtz Association.

However, a different picture emerges when each institution is ranked by the percentage of AI-related research with high Altmetric scores. The Pan-Canadian AI Strategy recorded the highest percentages of articles with high Altmetric scores (27–42%), followed by the Ivy League (19–24%). Helmholtz Association had 11–15% papers with high Altmetric scores published between 2017 and mid-2022.

Altmetric scores across an organization can indicate how effectively their research is being communicated to the public and may also serve to amplify the impact of other forms of scientific outreach as individual researchers build a stronger public profile. Some funders also include public engagement as a performance indicator of research outputs.

Helmholtz Association's Altmetric performance is in the lower half of its competitors. This indicates that more effort could be made to promote research through external and popular communication aimed at non-academic audiences.

Notes & References

1. Siler, K., Haustein, S., Smith, E., Larivière, V. & Alperin, J. P. (2018). Authorial and institutional stratification in open access publishing: the case of global health research. *PeerJ* **6**, e4269 https://doi.org/10.7717/peerj.4269
2. https://group.springernature.com/gp/group/media/press-releases/transformative-agreement-within-asia-manipal/17862952
3. Concepts were chosen by the Helmholtz Association.
4. 20 or higher was used as a benchmark — there is no accepted world standard as yet.

CASE STUDY

PARTICLE COLLISIONS

Accelerating physics with machine learning

Machine learning algorithms are helping physicists run faster simulations of particle collisions to help the processing of data from next-generation accelerators.

Particle accelerators, such as the Large Hadron Collider in Geneva, Switzerland, generate vast quantities of complex data. To make sense of it, physicists need detailed simulations based on fundamental theories. However, these can take a long time, even on the most powerful computers.

This demand will only grow with the next generation of larger, even more powerful facilities, such as the proposed International Linear Collider, which will be more than 30 kilometres long. For this reason, researchers including Frank Gaede at the Deutsches Elektronen-Synchrotron (DESY) research centre in Hamburg, Germany, are employing state-of-the-art machine learning algorithms to greatly reduce simulation time.

"Simulations have always been vital tools," says Gaede. "The Standard Model of particle physics tells us the likely products we would see when two particles collide, but our measurement devices are imperfect.

These generative adversarial networks (GANs) are notoriously difficult to train.

So, to understand what you measured, you have to simulate your detector. If we have lots of simulated events, we can work out how often we have missed particles."

A common detector for measuring particle energies, called a calorimeter, contains grids of small silicon detectors embedded in a heavy metal like tungsten which causes the particles to initiate a huge number of secondary particles, a so-called 'particle shower'. The newest calorimeters have millions of sensors, each as small as 5 by 5 millimetres, meaning that conventional simulations of particle showers using these take a very long time.

Gaede and co-workers developed three much faster machine learning algorithms called generative adversarial networks (GANs), and compared them against the outputs of Geant4, a more conventional system that implements the best possible knowledge of how particles interact in matter.

"One of the first uses of GANs was in faking images and paintings," explains Gaede. "There are two parts to a GAN — the generator and the discriminator. We train the generator to make more and more accurate simulations that can fool the discriminator, like a forger trying to outsmart an art expert." Similarly, GANs show excellent promise for imitating the outputs of the most accurate physics simulations, while saving huge amounts of processing time.

The team found that a type of GAN called the Bounded Information Bottleneck-Autoencoder (BIB-AE) most closely matched the outputs of Geant4, but takes just one three-thousandth of the time to generate results. Crucially, BIB-AE reproduced a peak in the energy distribution of particles that represents a well-defined process — the loss of energy by ionisation of a relativistic particles at a sensor.

"These GANs are notoriously difficult to train, so our PhD students and postdocs had to do a lot of tuning until they converged on realistic outputs," says Gaede. But "the BIB-AE performed very well," he adds, potentially fooling physicists who look at the output simulations into thinking the longer, full simulations using conventional methods had been performed.

Gaede says he looks forward to pursuing this work to apply machine learning to speeding up particle physics simulations. "It is great that particle physics is benefitting from neural networks, which are mainly developed for technologies like self-driving cars," he says. "However, our studies so far are limited to perpendicular impacts by particles within a certain energy range. Next, we hope to simulate electromagnetic showers for all energies and impact angles."

Stronger together?

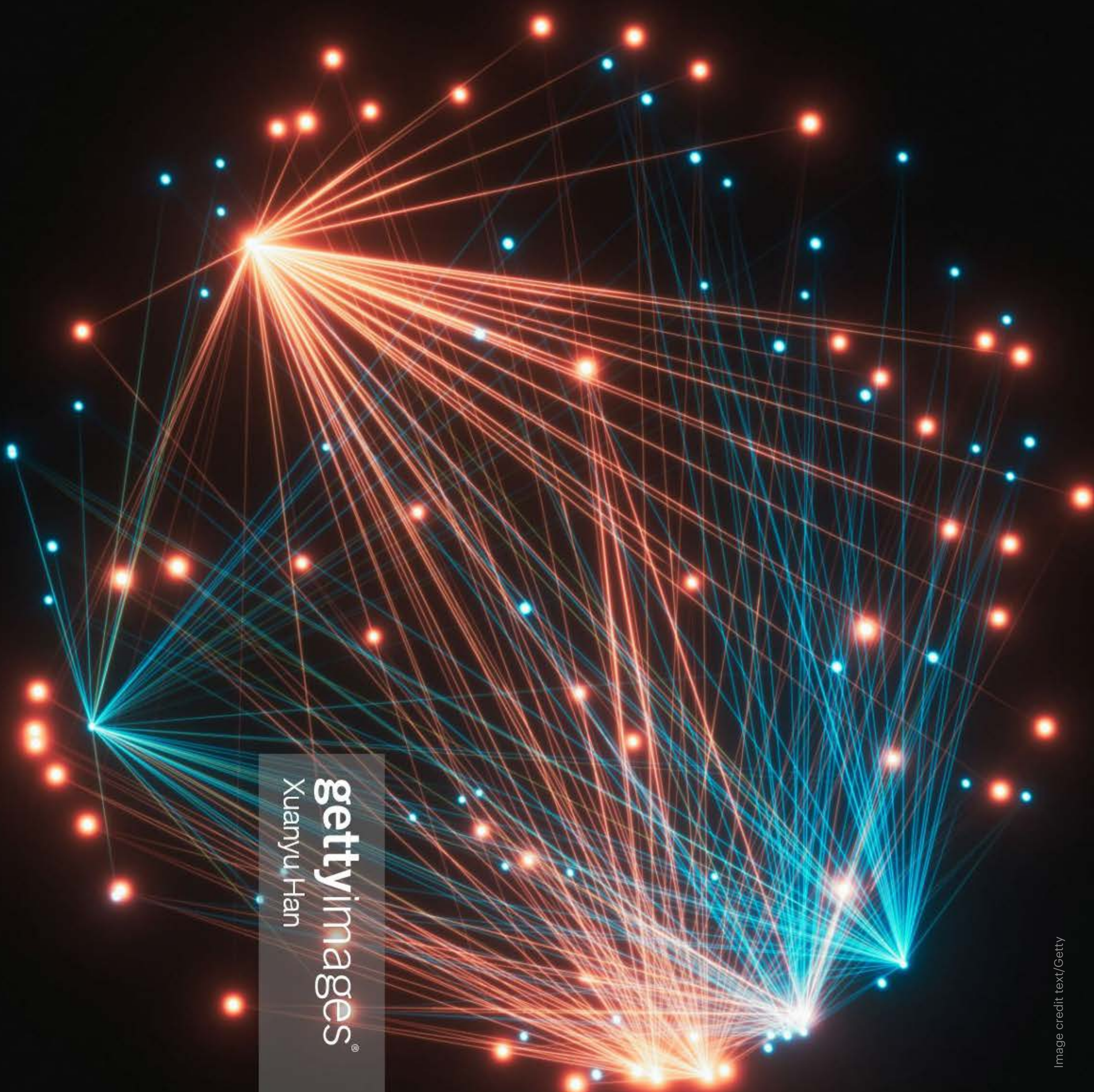


Image credit text/Getty

The link between collaboration and impact, both scholarly and popular, is investigated for the Helmholtz Association and its competitors through a bibliographic analysis.

Key highlights

- Helmholtz Association researchers are increasing the amount of AI research that they produce in collaboration.
- AI papers written by Helmholtz Association researchers with international collaborators tended to have higher average citation counts and Altmetric scores than those with no collaborators or domestic collaborators only.
- The top-five international collaborators on AI papers for the Helmholtz Association are:
  1. Harvard University (USA)
  2. University College London (UK)
  3. ETH Zurich (Switzerland)
  4. University of Oxford (UK)
  5. University of Manchester (UK)
- The top-five domestic collaborators on AI papers for the Helmholtz Association are:
  1. Technical University of Munich
  2. RWTH Aachen University
  3. Heidelberg University
  4. Technical University of Dresden
  5. Ludwig- Maximilians-Universität München
- Helmholtz Association researchers could make better use of the resources and expertise within the association by increasing intra-association collaboration.

3.1 Introduction

Collaboration among AI researchers, among different institutions, and between countries, is an important contributor to the growth and depth of world-leading scientific research. There has been remarkable growth in global scientific collaboration networks, and there is evidence showing[1] that the research emerging from international projects tends to have a much higher citation count and is viewed more frequently than research performed within one country.

This section looks at trends in research collaboration between Helmholtz Association Institutions and other groups, both domestic and international, based on bibliographic data from 2017 to mid-2022.

3.2 Growth in research collaboration

The proportion of AI research in individual Helmholtz centres involving collaboration with external researchers has risen rapidly, from 74% in 2017 to 83% in 2022, which can be seen in Figure 3.1.

3.3 Role of research collaboration in improving citation count, Field Citation Ratio and Altmetrics

Over the period of analysis, three major publication metrics were applied to all AI-related publications output by single Helmholtz Association Institutions — citation count, Altmetric score, and Field Citation Ratio (FCR). A FCR is a citation-based measure of an article's scientific influence compared to other publications in the same year and field of research, with 1.0 representing the average.

Research collaboration

Figure 3.1

Collaborative and non-collaborative output distribution.

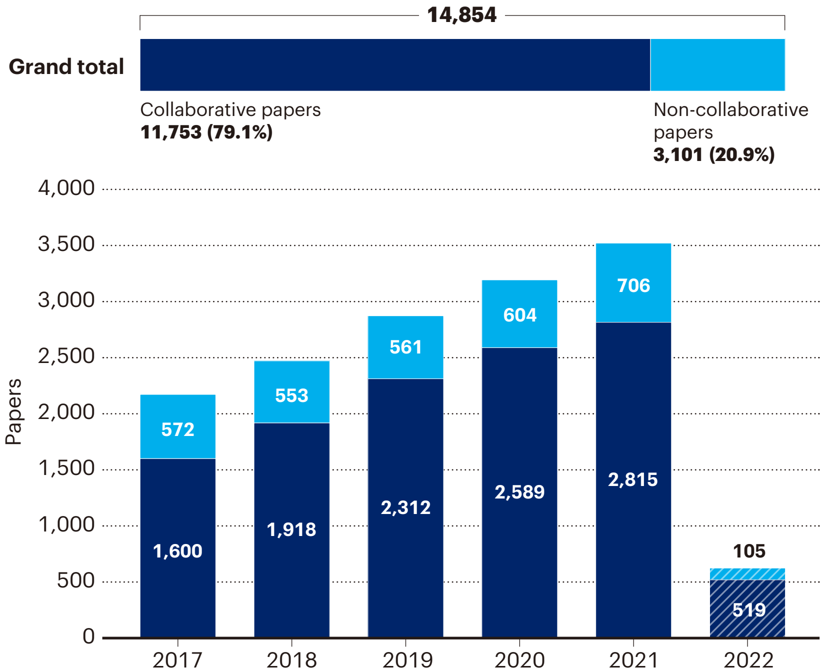
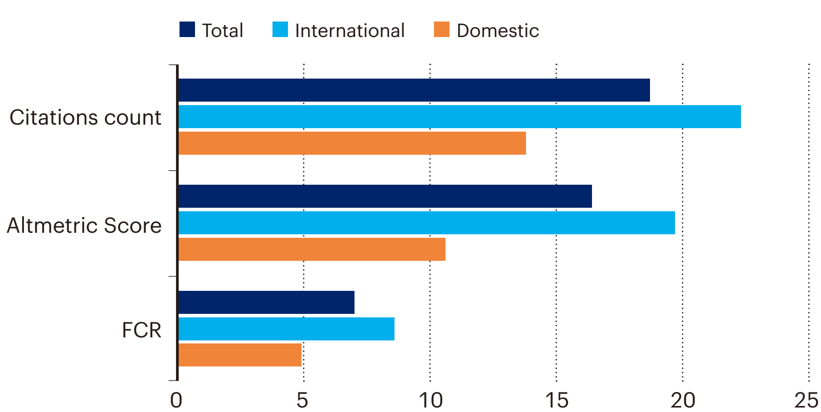


Figure 3.2

Average citation, Altmetric Score, and FCR within total, domestic collaborative, and international collaborative publications.



The analysis in Figure 3.2 found that AI papers written with international collaborators tended to have higher average citation counts, Altmetric scores, and FCRs (22.3, 19.7 and 8.6), respectively, compared to the full set of publications and domestic-only papers.

This could be explained by evidence that suggests[2] that the emergence of 'global network science' has boosted internationally co-authored papers, but at the same time it may have downgraded the importance of national collaboration and non-collaboration in science, particularly in Europe. One study[3] also shows that Altmetric scores are also strongly correlated with international collaboration.

3.5 Top-five domestic and international research collaboration partners

This measure compares the top-five domestic and top-six international institutions that partnered with Helmholtz Association institutions to collaborate on AI-related research publications between 2017 and mid-2022. During this period, collaboration information was recorded on 12,576 papers: 3,483 papers were written with domestic collaborators, 7,373 with international collaborators, and 1,765 without any collaborators.

Of the five leading international collaborators with Helmholtz Association Institutions, three were from the UK, one from Europe, and one from the USA.

- Harvard University, USA (287)
- University College London, UK (273)
- ETH Zurich, Switzerland (270)
- University of Oxford, UK (227)
- University of Manchester, UK (216)

Analysis indicates that key players in AI that don't feature in Helmholtz Association's collaborators are Tsinghua University, Shanghai Jiao Tong, Zhejiang

University and Harbin University (all in China) and Anna University (India).

Among the collaborators on research with Helmholtz Association Institutions, from institutions within Germany, the following organizations were the most prolific in AI publications:

- Technical University of Munich (1,199)
- RWTH Aachen University (818)
- Heidelberg University (685)
- Technical University of Dresden (609)
- Ludwig- Maximilians-Universität München (607)

Looking at the top collaborators for individual research institutes, Table 3.1 shows that intra-institute collaboration within the Helmholtz Association could be improved, with only five out of the 18 institutions having one or more Helmholtz Association collaborators in their top-three domestic collaborators. Increased collaboration would help break institutional silos and make the most of the Association's expertise.

References

1. Adams, J. *Nature* **497**, 557–560 (2013).
2. Kweik, M. *Studies in Higher Education* **46**, 2629 (2021).
3. Didegah, F. *et al.*, 'Increasing our understanding of altmetrics: identifying factors that are driving both citation and altmetric counts', *iConference 2016 Proceedings* (2016). https://doi.org/10.9776/16182.

Table 3.1

The top three international and domestic collaborators for the 18 Helmholtz Association centres by publication output. Italics and blue font indicates an intra-association collaborator.

Helmholtz Association institute	Top-three international collaborators	Top-three domestic collaborators
Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research	1. UiT The Arctic University of Norway, Norway =2. Stockholm University, Sweden =2. University College London, UK =2. National Oceanography Centre, UK =2. Woods Hole Oceanographic Institution, USA	1. University of Bremen 2. Jacobs University *3. GEOMAR Helmholtz Centre for Ocean Research Kiel*
Deutsches Elektronen-Synchrotron (DESY)	1. European Organization for Nuclear Research, Switzerland 2. Rutherford Appleton Laboratory, UK 3. Institute of High Energy Physics, China	1. Universität Hamburg *2. Karlsruhe Institute of Technology* 3. Humboldt-Universität zu Berlin
Forschungszentrum Jülich	1. McGill University, Canada 2. Montreal Neurological Institute and Hospital, Canada 3. Massachusetts General Hospital, USA	1. RWTH Aachen University 2. Heinrich Heine University Düsseldorf 3. University of Cologne
GEOMAR Helmholtz Centre for Ocean Research Kiel	1. Utrecht University, Netherlands 2. Institute of Marine Sciences, Spain 3. Tampere University, Finland	1. Leibniz-Institut für Wissensmedien 2. University of Tübingen 3. Leibniz Institute for the Social Sciences
German Aerospace Center (DLR)	1. Delft University of Technology, The Netherlands 2. ETH Zurich, Switzerland 3. European Space Research and Technology Centre, Netherlands	1. Technical University of Munich 2. University of Würzburg 3. Technische Universität Braunschweig
German Cancer Research Center (DFKZ)	1. Fred Hutchinson Cancer Research Center, USA 2. University of Washington, USA 3. Harvard University, USA	1. Heidelberg University 2. University Hospital Heidelberg 3. Ludwig-Maximilians-Universität München

Research collaboration

Helmholtz Association institute	Top-three international collaborators	Top-three domestic collaborators
German Center for Neurodegenerative Diseases (DZNE)	1. University College London, UK 2. Harvard University, USA 3. Massachusetts General Hospital, USA	1. Otto-von-Guericke University Magdeburg 2. University of Bonn =3. Leibniz Institute for Neurobiology =3. Ludwig-Maximilians-Universität München
GSI Helmholtz Center for Heavy Ion Research	=1. Joint Institute for Nuclear Research, Russia =1. European Organization for Nuclear Research, Switzerland 2. University of Turin, Italy	1. TU Darmstadt 2. Goethe University Frankfurt 3. Heidelberg University
Helmholtz Center for Infection Research (HZI)	=1. University of California, Berkeley, USA =1. Lawrence Berkeley National Laboratory, USA 2. University of the Basque Country, Spain	1. Hannover Medical School =2. Technische Universität Braunschweig =2. German Center for Infection Research
Helmholtz Centre for Environmental Research (UFZ)	1. Lund University, Sweden 2. Environmental Protection Agency, USA 3. Joint Research Centre, Italy	1. Leipzig University 2. TU Dresden =3. University of Potsdam =3. Humboldt-Universität zu Berlin
Helmholtz Center for Information Security (CISPA)	1. Stanford University, USA 2. University of California, Berkeley, USA 3. Vrije Universiteit Amsterdam, The Netherlands	1. Saarland University 2. Max Planck Institute for Informatics 3. German Research Centre for Artificial Intelligence
Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences	1. Wuhan University, China 2. ETH Zurich, Switzerland 3. Delft University of Technology, The Netherlands	1. Technical University of Berlin 2. University of Potsdam 3. Leibniz University Hannover
Helmholtz-Zentrum Berlin für Materialien und Energie	1. University of Oxford, UK 2. University of Manchester, UK =3. Fudan University, China =3. Uppsala University, Sweden =3. Paul Scherrer Institute, Switzerland =3. University of Cambridge, UK	*1. Helmholtz-Zentrum Dresden-Rossendorf* *2. Helmholtz Zentrum München* *=3. GSI Helmholtz Centre for Heavy Ion Research* =3. TU Dresden
Helmholtz-Zentrum Hereon	1. Université Catholique de Louvain, Belgium 2. ETH Zurich, Switzerland 3. Oak Ridge National Laboratory, USA	*1. Forschungszentrum Jülich* 2. RWTH Aachen University =3. Research Center for Information Technology *=3. Helmholtz-Zentrum Dresden-Rossendorf*
Helmholtz Zentrum München	1. Harvard University, USA 2. University of Zurich, Switzerland 3. ETH Zurich, Switzerland	1. Technical University of Munich 2. Ludwig-Maximilians-Universität München 3. Rechts der Isar Hospital
Helmholtz-Zentrum Dresden-Rossendorf	1. Hunan University, China 2. Paul Scherrer Institute, Switzerland 3. Edith Cowan University, Australia	1. TU Dresden *2. German Cancer Research Center* 3. National Center for Tumor Diseases
Karlsruhe Institute of Technology	1. ETH Zurich, Switzerland 2. European Organization for Nuclear Research, Switzerland 3. TU Wien, Austria	1 RWTH Aachen University 2. Technical University of Munich 3. Fraunhofer Institute of Optronics, System Technologies and Image Exploitation
Max Delbrück Center for Molecular Medicine	1. Baylor College of Medicine, USA 2. Harvard University, USA 3. Broad Institute, USA	1. Charité — University Medicine Berlin 2. Humboldt-Universität zu Berlin 3. Berlin Institute of Health at Charité — Universitätsmedizin Berlin

Software supports machine learning exploration of cell activity

We can now sequence the entire genetic activity of a single cell — but processing and understanding such vast quantities of data is a key challenge.

More than two decades ago, the first human genome was sequenced. Today, the gene activity of a single human cell — the transcriptome — can be sequenced, revealing which genes are active, or even in which cell these genes are active among the thousands of cells in a tissue sample.

'Single-cell transcriptome sequencing' generates vast amounts of complex data. Processing, tagging, and organizing it into useful formats has presented an informatics challenge like no other.

But software called Scanpy – Single Cell Analysis in Python, which refers to the programming language used to develop it — by researchers at the Institute for Computational Biology at Helmholtz Munich, may be the answer. It's open-source, and provides a customized framework for the machine learning and artificial intelligence algorithms needed to analyse and make sense of the reams of single-cell transcriptome data.

"It provides high-performance methods for normalizing data — the quality control, plotting, clustering and then basic analysis," says Isaac Virshup, lead developer at Helmholtz Munich, who is also a PhD candidate at the University of Melbourne in Australia.

The choice of Python was important because it enables a machine-learning approach to the data analysis. "The methods that are applicable to single-cell sequencing are often much more machine-learning driven than they were previously," Virshup says.

The data being analysed by machine learning, using Scanpy as a platform, are all the sequences of messenger RNA (mRNA) within a cell, or a batch of cells at a point in time. mRNA are single-stranded sequences of genetic code that carry the instructions for making proteins, so sequencing all the mRNA in a cell reveals its biological activity.

Open-source nature of Scanpy means researchers can tailor it to their specific needs.

Once those sequences are available, a programming tool within Python called anndata – short for 'annotated data', which was developed specifically for Scanpy, allows additional information to be recorded for each sequence that can be used in the machine learning analysis.

The Helmholtz Munich team have also developed add-on software called Squidpy, which allows researchers to visualise the physical location of different mRNA sequences within cells and also across a particular tissue sample. This can reveal how patterns of cellular gene expression can change across a tumour or in different locations with an individual cell.

These platforms are enabling clinical applications of transcriptome sequencing data that weren't even imagined five years ago. "It's a particularly exciting field to be in because the questions people are asking have been evolving so quickly," says Virshup.

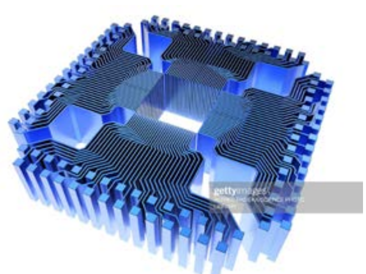
For example, analysing and comparing tumour cells from different patients will allow fine-grained categorization of the cellular differences not both within and between patients. It can also allow the detailed study of the effects of drugs on different cell types in various tissues.

Scanpy and anndata are now being used by initiatives such as the Human Cell Atlas — an international initiative to map all the cells types in the human body over a lifetime. And the open-source nature of Scanpy means researchers can tailor it to their specific needs.

"One of the very cool things about Scanpy — and why I enjoy about working on these things — is that they are a kind of core infrastructure that allows many people in many different labs to develop methods on top of it," Virshup adds.

The next leap for AI: quantum computers

Kristel Michielsen was an author of a 2019 Nature study that reported the first case of 'quantum supremacy': where a quantum computer solved a task faster than the world's top supercomputer. The head of the Quantum Information Processing group at the Jülich Supercomputing Centre, and a professor at RWTH Aachen University, Germany, she studied quantum computing before it was a buzzword.



What does your research involve?

My research group focuses on quantum information processing. Quantum theory can be used to describe phenomena — in materials science and theoretical physics for example — and quantum models are used to simulate them. But simulating quantum systems requires a lot of computing resources. The computers need to be big, and we need technically demanding software to get the most out of them. Finding improvements is something we have been working on for a long time.

A quantum computer is also a quantum system itself, so, now we use supercomputers to simulate how quantum computing technologies interact with their environments. We do this by making a so-called digital twin that takes many properties into account and then we compare it to the outcomes of a real device, in the hope of getting a better understanding of how quantum technologies work, and how they interact with their environments.

My group also examines the potential applications of quantum computers, like solving problems in quantum chemistry, learning how proteins bind to DNA, and understanding how to design new materials and drugs.

They can also be used for solving optimization problems in areas as varied as finance, traffic, scheduling, drug design, and hotel reservations. This is where AI and quantum machine learning play a role.

Are you using AI in your quantum computing research?

We are looking into algorithms for quantum machine learning. We have already applied them to small problems, like docking proteins on DNA. We also applied them for binary image classification, where we considered a picture of a small part of the city of Lyon, France and then tasked the quantum computer, which in this case was a quantum annealer, to identify the buildings in the picture. We also do benchmarking, where we take a classical algorithm and compare it to a quantum one. None of this is very advanced compared to artificial intelligence for classical computing. But we are looking into it.

It sounds like we're still in the early stages of applying AI to quantum computing, but will these machines allow AI to be much more advanced than classical computing currently allows?

Our vision has always been that quantum computers are special purpose machines that can help in combination with classical computers. For example, by very swiftly selecting learning data that can be used to train algorithms on classical machines.

So we are studying simulations of hybrid quantum-classical systems to understand how we can couple quantum computing hardware with our high-performance computing systems. Bringing them together will make it much easier to run simulations on both machines, so we have the best of both worlds.

Finally, I do not want to add to the hype around quantum computers. We are not that far yet. We are in the research phase and there is future potential.

Image credit text/Getty

Information here, there and everywhere

Christof Weinhardt is professor of information systems and head of the Information and Market Engineering group at the Karlsruhe Institute of Technology. He and his team are using artificial intelligence (AI) to analyse how information systems interact with and impact the economy and society.

Image credit text/Getty

What is information systems research?

Information systems is an interdisciplinary research field. We intersect with management, for example, by finding out where IT can best help businesses. We work with political scientists to investigate the impacts of IT systems on democracy and society. We also work with psychologists, neuroscientists and medical researchers to find out how IT-supported decision-making is influenced by our mental state or cognitive load. We are a broad research discipline and are very aware that information systems have a huge impact on the economy and society.

Much of your research uses AI. What are some examples?

One of the biggest areas is smart grids and energy markets. We are developing algorithms to forecast renewable energy electric loads in residential buildings, and peak load times for energy communities and industrial parks. We are also developing algorithms that learn from weather data and user mobility patterns to recommend the optimal time and place to charge an e-vehicle.

We are also using AI to conduct simulations that could improve decision-making about the future organisation of our transmission networks.

How do you use AI to help understand human behaviour?

An area we work in is called digital experience, which studies how people interact with systems. For example, we are investigating how experiences in immersive virtual reality environments affect consumer decision-making about purchases. We also use AI to study collaborative platforms in work environments; this information could help decide, for example, when it is best to stop a video call because people are tired and there will be little progress in a meeting.

Additionally, we are working with sociologists, political scientists and computer linguists in digital democracy research to identify the linguistic and grammatical characteristics of fake news. What is the trick that makes it look real? How is it spread to manipulate people? Once we have this information, we can develop algorithms that identify misinformation campaigns and give readers a trustworthiness score for news items.

Where do you see the most potential for the use of AI in information systems applications?

Advanced forecasting algorithms are very important in the energy sector. AI is delivering the technological toolbox for predicting consumption and managing electricity transmission and distribution. AI will help to organize and control our future energy markets.

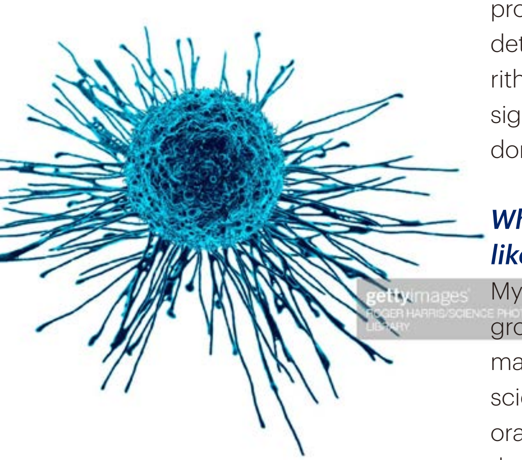
There is also a huge need to learn more about what fake news is and how it is spread. Algorithms can amplify misinformation and conspiracy theories by learning reader behavioural patterns and then displaying news from certain sources more often. This can affect our opinion and lead to a more polarized society.

Can AI can be used for bad reasons as well as good?

It *is* used for bad. Nowadays, AI isn't only used by some very smart people. The algorithms are publicly accessible. They run on our laptops and in the cloud. People and organizations, whatever their purposes, can access and use this technology.

Cancer detection gets an AI upgrade

Artificial intelligence has already proven its worth in radiology, but it is far from reaching its full potential. Computer scientist, **Klaus H. Maier-Hein**, who is head of Medical Image Computing division at the German Cancer Research Center in Heidelberg, has assembled a team that is solving the problems that stand in its way.



What is the current focus of your research?

Our work on brain tumour analysis is well known. We are showing that we can use fully automated pipelines and tools to assess how tumours react to treatment by measuring their volume with 3D magnetic resonance imaging (MRI) images, which are created by stacking many two-dimensional (2D) images on top of one another. Normally, to measure a tumour's volume and detect fine-grained changes, a radiologist must slowly annotate every 2D image, which is a slow process. We can do it much faster and more precisely using AI.

We are also doing important work in prostate cancer assessment, and early detection of breast tumours. We use algorithms to analyse MRI images for early signs of malignancy to rule out cases that don't need a biopsy.

What does your research team look like?

My division has 65 people with backgrounds in computer science, physics, mathematics and medical computer science, and we also have strong collaborations with radiologists. Our work is three-pronged. We have people working on applications that can help patients, like the ones I have mentioned.

Other members of my team work in software and infrastructure development. The algorithms we use for various applications need to be trained. But it is challenging to collect multiple sets of very different data from numerous centres. To address this, we are building infrastructure to allow researchers to send algorithms that learn directly at a centre without the data having to leave it. This is called federated learning.

A third group works on developing the basic methodologies needed for detecting objects and learning patterns within images.

What are the main challenges facing AI in radiology?

One I already mentioned is the availability of high-quality data. An important challenge in the future will be to go federated. Another is generalization. For example, a publication might say we have a new method that works nicely and this is how we applied it to our data. Then you try it on your data but it doesn't work. It doesn't generalize. To address this, we developed 'nnU-Net', a deep-learning method, which can tune the method used by one group so that another can apply it to their data. This is now used by many groups worldwide. There is also the problem of shifting domains. For example, if you train an algorithm in Europe, will it be similarly powerful for processing data on an Asian population? Or if you train it on a younger population, will it work as well on an older one? So that's another important direction for research.

What can we expect next?

AI will provide an increasing amount of support to radiologists. Currently, they need to routinely look at huge numbers of images. Our work on AI will save them time so they can focus on demanding questions and interact more with their patients. Hopefully we will be able to detect tumours earlier and more precisely, and radiology will become more important within personalized medicine. There is so much information in the images that we are currently missing or not using efficiently, because it's so tedious to extract. Algorithms can help with that.

Image credit text/Getty

Research highlights

Summaries of selected highly-cited and popular research articles produced by Helmholtz Association researchers.

Outdoor lighting levels increasing

The level and brightness of outdoor night-time lighting increased globally by an average of 2.2% every year between 2012 to 2016.

Using data from a satellite sensor able to measurement night-time radiance in the visible spectrum, a team led by researchers from GFZ German Research Centre for Geosciences has performed the highest resolution study yet of night-time lighting across the planet.

Their work revealed that many larger cities had a reduction in night-time lighting in their centres as older brighter bulbs were replaced by more energy-efficient LEDs. However, the use of LEDs and associated decrease in energy consumption cost has actually led to an increase in the use of night-time lighting, particularly on the partly urbanized city fringes and areas that were previously less lit or unlit due to cost.

Sci. Adv. **3**, e1701528 (2017).

Machine learning improves brain cancer diagnosis

The accuracy of central nervous system tumour diagnosis can be significantly improved by using a machine-learning approach to analyse patterns of DNA changes known as methylation.

There are more than 100 types of tumours that affect the brain and spinal cord and these can be difficult to tell apart based on cellular features alone. However, the different tumours do have unique signatures of DNA methylation.

An international team, which included researchers from the German Cancer Research Center (DKFZ), developed a machine learning algorithm that can classify central nervous system tumours according to their DNA methylation patterns. When compared to the standard pathology results for more than 1,100 tumours, the algorithm corrected the diagnosis in 12% of cases.

Nature **555**, 469-474 (2018).

COVID-19 caught sniffing around the brain

The nasal passage provides COVID-19 easy access to the central nervous system.

Common neurological symptoms of COVID-19 such as loss of taste and smell, headaches, and nausea suggest that the virus enters the central nervous system, raising questions about its point of entry. A team from Germany, including researchers from the German Center for Neurodegenerative Diseases (DZNE), used fluorescent microscopy and an open-source biological imaging analysis software (Fiji) to analyse tissue from the nasopharynx — the airway between our nose and windpipe — and selected brain regions of 33 COVID-19 patients who had died.

They found the highest levels of viral RNA in tissue from the olfactory mucosa, the membrane at the top of the nose that sends smell signals to the brain, suggesting that SARS-CoV-2 exploits the proximity of nerve endings to this part of the nasal passage to gain access to the brain.

The study also revealed high levels of SARS-CoV-2 infiltrated the medulla oblongata, the part of the brain that controls heartbeat, breathing, and blood pressure.

Nat. Neurosci. **24**, 168-175 (2021).

STRING untangles proteomics

A growing public database of protein-protein interactions is enabling researchers to probe and predict biological phenomena in thousands of organisms.

The structure and function of millions of different proteins have already been described, but keeping track of the billions of possible interactions between them is a major undertaking.

In 2019, an international team, including researchers from the Max Delbrück Centre for Molecular Medicine, launched the latest version of the protein association database, STRING v11. STRING derives protein-protein interactions from several sources, including direct data input from lab experiments and automated text mining of millions of scientific papers. The database currently contains more than 20 billion interactions between proteins from more than 14,000 organisms. Users can search a protein to view its associated proteins and the nature of each interaction, for example transporter or catalyst. Each interaction is given a score between zero and one based on confidence that the association is real.

By specifying the specific biological function of each interaction, the database can support crucial discoveries, such as identifying disease-related genes.

Nucleic Acids Res. **47**, 607-613 (2019).

Traces of a lost continent

Ancient mineral found in the sands of Mauritius.

An international team, including researchers from GFZ Helmholtz Centre Potsdam, discovered samples of the mineral zircon, dating to the Archean Eon, which are up to 3 billion years old, on Mauritian beaches.

The minerals predate the formation of the island of Mauritius itself, which occurred due to volcanic eruptions approximately 8-10 million years ago.

The team then used uranium-lead radiometric dating methods to date the zircons to between 2.5 and 3 billion years old. They claim that their findings demonstrate the existence of an ancient continental crust beneath modern-day Mauritius.

"We propose that Mauritius and other potential Mauritian continental fragments, collectively named Mauritia are dominantly underlain by Archaean continental crust, and formed part of the ancient nucleus of Madagascar and India," they report.

Nat. Commun. **8**, 14086 (2017).

eggNOG-mapper accelerates identification of gene function

Genes that evolved in different species from a common ancestor can help predict gene function in new genome sequences.

Whole genome sequencing has accelerated the discovery of orthologs — related genes in two species that diverged from a common ancestor. As orthologs typically retain their original role, geneticists can use them to infer the function of genes in new sequences without doing extra experiments. However, precise annotation of entire genomes requires a lot of computation.

In 2017, a team including researchers from Max Delbrück Centre for Molecular Medicine created eggNOG-mapper, an AI that extracts the biological role, such as fat metabolism or energy conversion, of orthologs in the eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database and uses them to label large sets of gene sequences. Compared with BLAST and InterProScan, two widely used annotation tools that do not use orthologs, eggNOG-mapper runs about 15 and 2.5 times faster, respectively, and reduces the number of incorrectly annotated genes.

An evolving and free online tool, eggNOG-mapper hopes to accelerate the discovery of new gene functions in emerging genome sequences.

Mol. Biol. Evol. **34**, 2115-2122 (2017).

Deep learning goes sky high

Deep learning is improving the reach of remote sensing to tackle global challenges.

In the last decade, deep learning has rapidly advanced artificial intelligence and computers are now starting to match human ability in tasks such as action recognition, object detection, and image captioning. But the complexity of the physical processes monitored by satellites, such as atmospheric gas concentrations and landslides, presented greater challenges to data scientists. In 2017, an international team that included researchers from the German Aerospace Center, reviewed the status of deep learning in remote sensing, noting an exponential increase in published papers on the topic since 2014. While most of the research examined focussed on classification and detection tasks, the researchers are confident that deep learning could also help solve problems involving complex physical processes. The authors highlight the need to develop unsupervised deep learning to replace labour-intensive annotation of huge training datasets.

By overcoming these technical challenges, remote sensing and AI could be combined for new applications, such as automatic target recognition for military surveillance and terrain classification for monitoring land use change.

IEEE Geosci. Remote Sens. Mag. **5**, 8-36 (2017).

Creating a buzz about bee habitats

Citizen science reveals a targeted approach is needed for bumblebee conservation.

Bumblebees are crucial pollinators, but many species are experiencing population declines due to loss of habitat and floral diversity related to changes in land use, as well as increased use of pesticides and factors related to climate change.

Understanding the link between land use, climate variables and bumblebee abundance is critical for the design of effective conservation measures.

To this end, an international team led by researchers from the Karlsruhe Institute of Technology analysed 10 years' worth of data about 14 species of bumblebee. This data was collected by British citizen scientists through the Bumblebee Conservation Trust's BeeWalk scheme in the UK, the team compared it to national-scale land cover and climate data.

The team found that targeted conservation efforts are needed, as there is both between and within species variation in preferred habitats. However, lead study author Penelope Whitehorn noted that "reversing the loss of semi-natural areas such as wetlands may be the single most generally effective action for bumblebee conservation, while improving habitats in urban and arable areas could benefit particular rare species."

J.Appl. Ecol. **59**, 1671-1946 (2022).

Plane plumes' contribution to climate change will increase

Plane trails' contribution to global warming set to increase despite advances in aviation technology

Expected increases in air traffic are likely to triple the effect that plane trails have on climate change by 2050 compared to their contribution in 2006.

Many airlines offer passengers the option of offsetting the carbon emission of their flight when purchasing tickets. But the condensation trails planes leave in their wake are thought to make an equivalent or larger contribution to climate change. These contrails increase the amount of energy the Earth receives from the Sun relative to the amount that is radiated out into space.

Now, two researchers from the German Aerospace Center have calculated that the fourfold increase in air traffic predicted by 2050 will triple the contribution of contrails to the Earth's energy budget.

The model the pair developed accounted for many factors, including more-efficient planes, lower soot levels in exhaust gas, and the effect of climate change on contrail formation.

Their results indicate that reducing soot emissions and boosting propulsion efficiency won't be enough to offset the effect of increased air traffic.

Atmos. Chem. Phys. **19**, 8163–8174 (2019).

A clear verdict — the oceans are becoming less oxygenated

As the atmosphere gains carbon dioxide, the oceans are losing oxygen.

The world's oceans are losing oxygen — an essential element for most of marine life — at a concerning rate due to climate change. This could cause habitat loss for some fish and invertebrates.

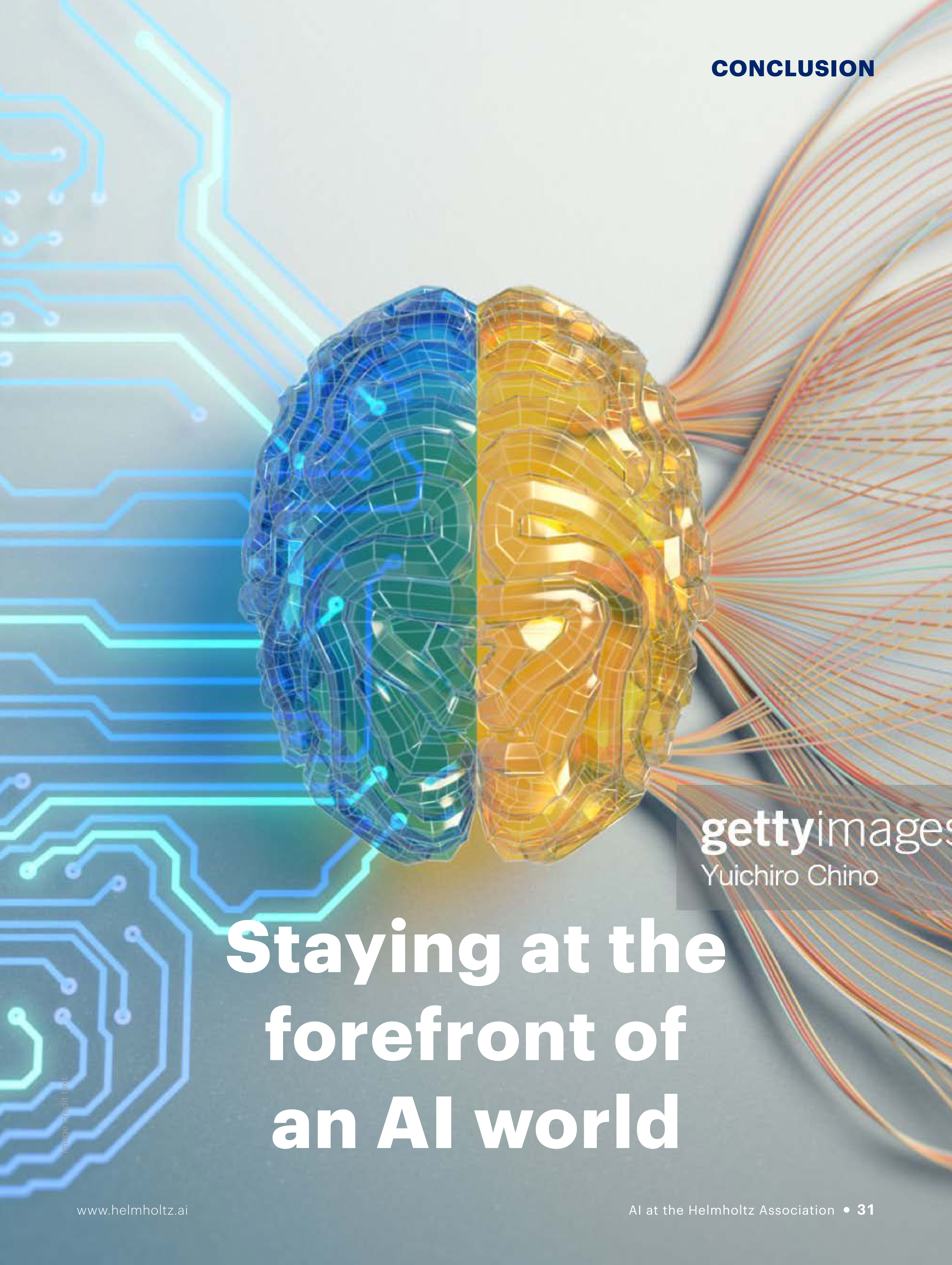
Models of the world's oceans predict that the amount of oxygen dissolved in them will drop somewhere between 1% and 7% by 2100 as the world heats up.

While measurements on oceanic oxygen levels have provided some evidence of this decline, the picture has been patchy because previous studies considered limited regions or time periods.

Now, three researchers from the GEOMAR Helmholtz Centre for Ocean Research have performed an extensive analysis of oxygen data collected over the past 50 years from the world's oceans and found a 2% decline overall.

Roughly 15% of this oxygen loss is estimated to be due to warming oceans not being able to hold as much oxygen; the remaining 85% is caused by other, factors such as changes in currents.

Nature **542**, 335–339 (2017).



Staying at the forefront of an AI world

A global and local analysis of the artificial intelligence landscape for the Helmholtz Association.

AI will contribute $15.7 trillion to the global economy and a 26% increase in GDP for local economies by 2030

As an enabling technology, artificial intelligence (AI) will change the way we live and work – from digital assistants on our wrists reminding us when we have a dental appointment, to AIs taking over routine processing tasks and allowing humans to focus on more creative endeavours.

This effect will reverberate through the world, with a recent PwC study, 'Sizing the prize: What's the real value of AI for your business and how can you capitalise?', estimating that AI will contribute $15.7 trillion to the global economy and a 26% increase in GDP for local economies by 2030[1].

At the forefront of this transformation are researchers, many of whom are already using AI to facilitate their discoveries in a range of fields, from engineering to medical and health sciences.

In this report, commissioned by Helmholtz AI, we used bibliometric analyses to take a closer look at the AI research landscape on two levels: global, identifying the key geographical hotspots and research trends; and local, focusing on publication trends in AI-related research conducted at the Helmholtz Association and related competitors.

Globally, we found that:

- The number of AI publications per annum was increasing – growing by almost 60% between 2017 and 2021.
- The country that publishes the most publications in AI research is China.
- The largest amount of AI grant funding worldwide was in medical and health sciences.

At the Helmholtz Association, we found that:

- 15,067 AI publications were published between 2017 and mid-2022.
- The Helmholtz Association's AI-related papers increased at a CAGR of 12.6% between 2017 and 2021, which tracks closely to the global rate of 12.4%.
- Over one third of publications were published in Q1 journals.
- About two-thirds of AI output was published in open access sources.
- There were an average of 15.8 citations per publication.
- Approximately two thirds of citations were from international sources.

Our recommendations for the Helmholtz Association include:

1. To stay aligned with the global push towards open access, consider increasing the proportion of publications published in open-access journals.
2. Helmholtz Association could improve their Altmetric performance with increased science communication activities aimed at the public and the popular media.
3. Helmholtz Association researchers could make better use of the resources and expertise within the association by increasing intra-association collaboration.

Reference

1. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

Description of data methodology employed

1. Selection of publications

As the definition of artificial intelligence (AI) is widely debated, we (the Nature Research Intelligence team) adopted a broad definition of AI that included various techniques when selecting content. To manage this ambiguity, we employed a graded approach to define AI. A classifier was initially trained using 2021 AI journals as defined by SCIMAGO. This revealed that the best model combines title and reference embeddings with the AI field of research code from Dimensions. To extend the range of searches while maintaining a high accuracy, we extended this model by including a set of AI-related concepts (e.g. deep learning, autonomous driving). We scored publications using a cosine similarity based on the document relevance for these AI-related concepts, enabling us to identify AI publications with high accuracy.

Additionally, simple filters (such as limiting the publication year to the range 2017–2022) were applied from the beginning. In particular, only certain content types (including articles, book chapters, conference proceedings and preprints) were considered. For journal content, we considered only primary research articles and reviews.

2. Details of the topic model

The topic model is based on a weighted co-occurrence network of concepts — two concepts are connected if they appear together in one or more publications. Clusters of densely connected concepts are associated with subtopics and are identified using the Louvain community detection algorithm. Unlike co-citation-based topic models, in which publications are assigned a unique subtopic, concept-based models can assign more than one subtopic to a publication using a similarity score. This ability is useful when dealing with highly multidisciplinary fields such as AI.

Additionally, the topic model can generate subtopics limited to pre-defined fields of research. This is achieved by associating individual concepts with the field of research in which they are most often used. Concepts in other research fields are then 'switched off' in the co-occurrence network and are not considered by the model. The process can be repeated for additional research fields, and each new set is combined with previous ones to form the final set of subtopics in AI. This feature is important since if concepts were not pre-limited, a very large and vague cluster would emerge that combines deep learning with applications such as medical diagnosis, genetics and autonomous driving — a consequence of deep learning's connection with applied subtopics and the community-detection algorithm's inability to understand the network's context.

This approach identified not only core AI research but also six broad areas of application, including, medical and health sciences, and Earth and environmental sciences, and engineering. It also identified 55 more-granular subtopics such as deep learning, medical diagnosis and the Internet of Things. The broader research fields largely correspond with those of the Helmholtz Association.

2. Grants and patents

Grants and patents granted between 2017 and 2022 in Dimensions were classified using a similar approach. The title embeddings and concepts identified for research publications have been applied to titles, abstracts and references depending on the availability of source data.

nature
strategy reports