Emerging Trends in Artificial Intelligence and Machine Learning

Jason Augustyn, Ph.D.
Chief Technology Officer
BluePath Labs
jaugustyn@bluepathlabs.com

What Is AI/ML?

Brought to you by Microsoft

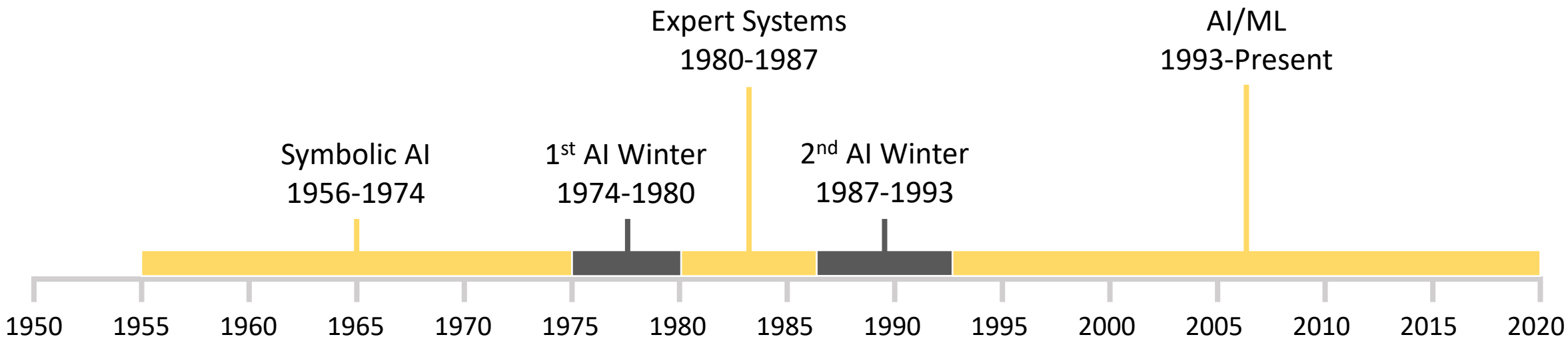
Artificial Intelligence

The capability of a computer system to **mimic human cognitive functions** such as learning and problem solving. Through AI, a computer system uses math and logic to **simulate the reasoning that people use** to learn from new information and make decisions.

Machine Learning

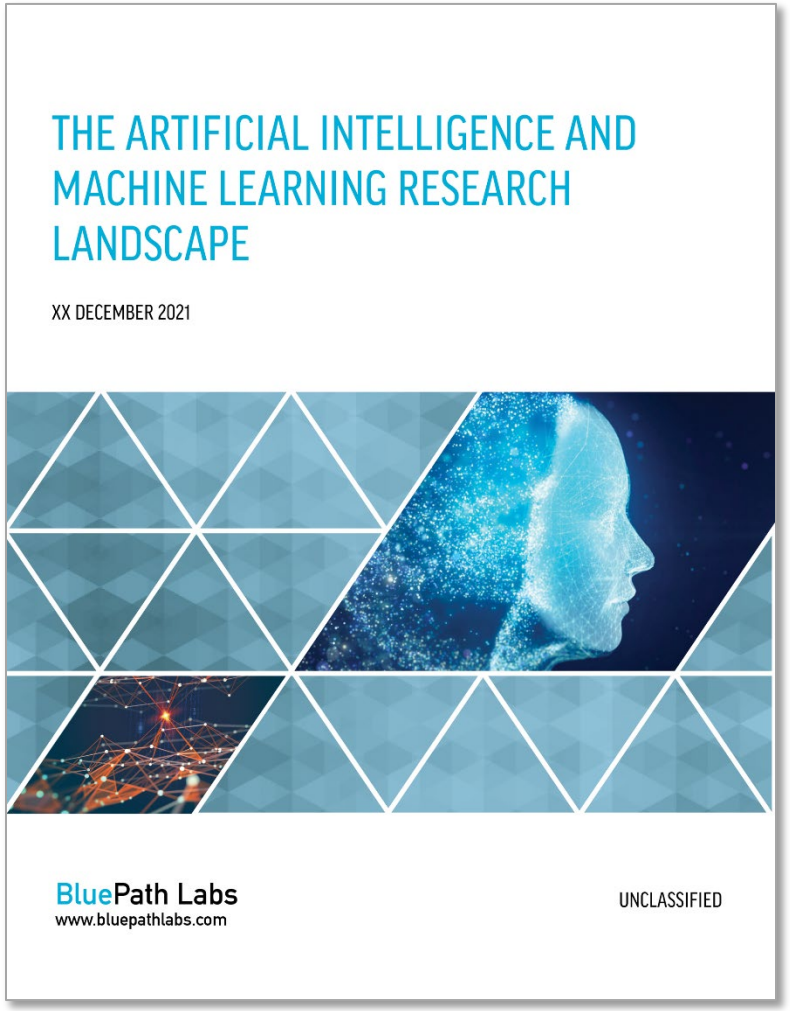
An application of AI. It's the process of **using mathematical models of data to help a computer learn without direct instruction**. This enables a computer system to continue learning and improving on its own, based on experience.

A Brief History of AI/ML



The AI/ML Research Landscape

- What are the major research topics driving AI/ML today?
- Who are the national and institutional leaders in AI/ML?
- What trends are likely to shape AI/ML over the next 3-5 years?

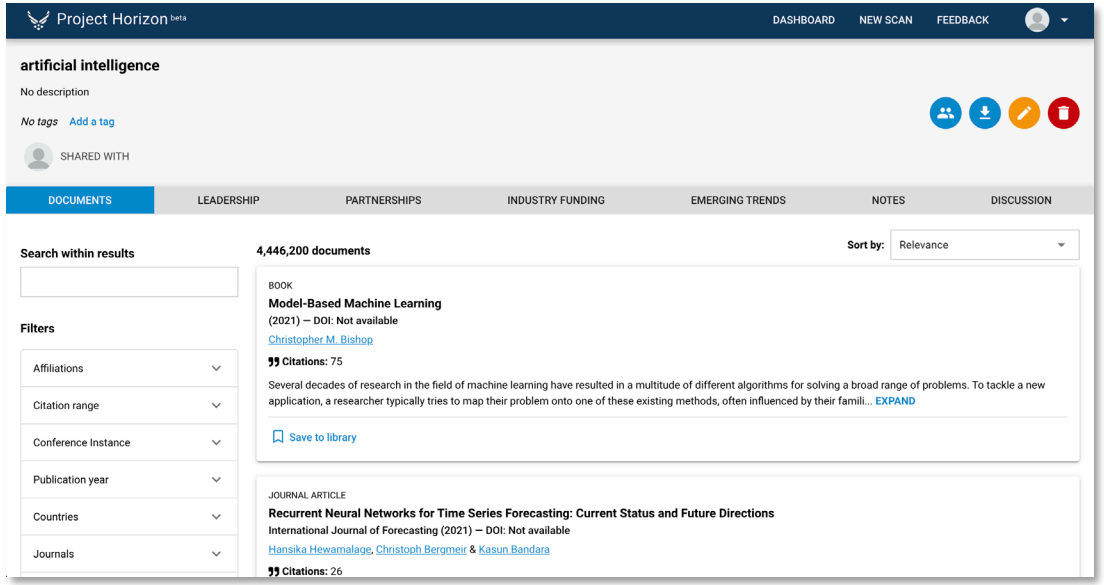


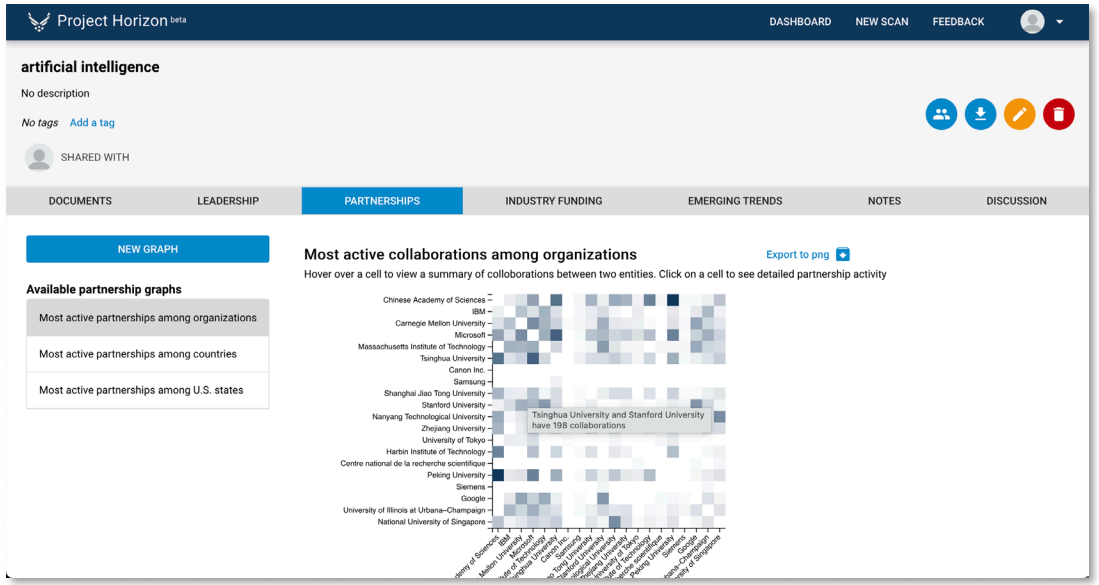
Landscape Analysis Methodology

BluePath Labs
www.bluepathlabs.com

Methodology

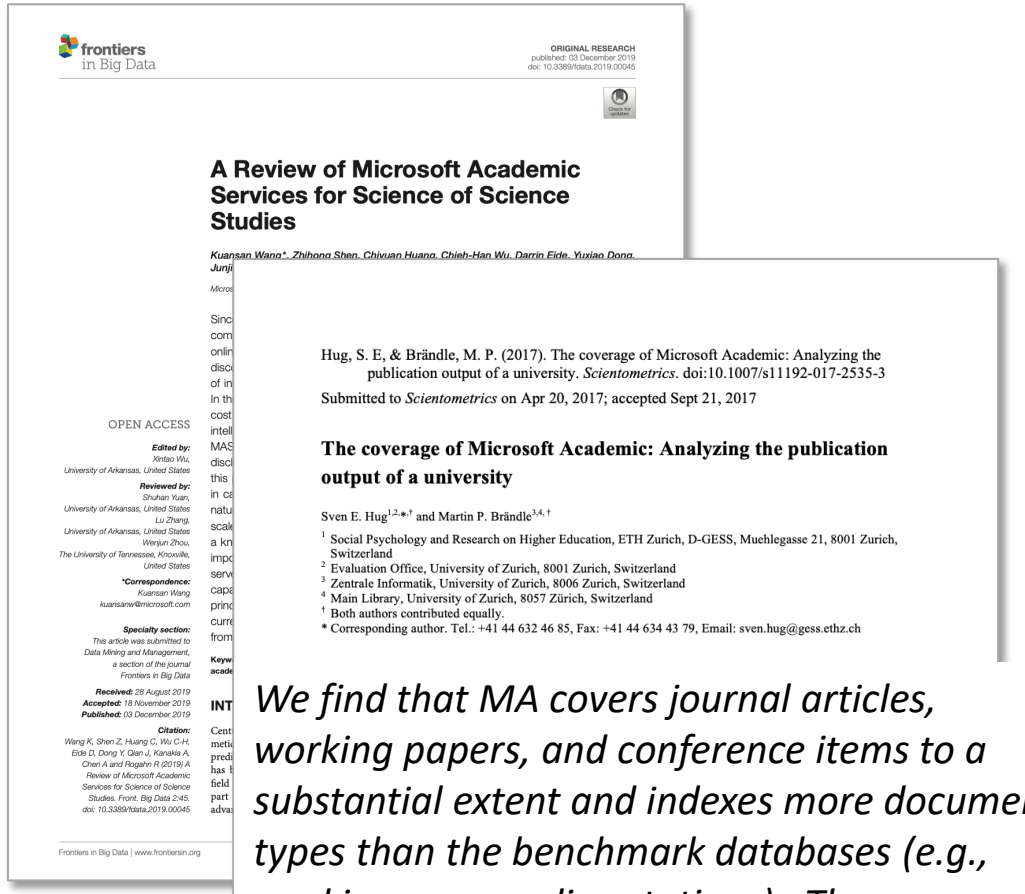
- Analysis-employed tools developed for Project Horizon.
- Conducted broad search for academic journal articles and conference papers on AI/ML published in the past 5 years.
- Filtered documents to identify high-quality, relevant research.
- Applied natural language processing (NLP)-based clustering techniques to identify technical trends.





Data

- Primary data source was Microsoft Academic Graph (MAG), an open bibliographic database covering over 240 million research documents, books, patents, theses, and data repositories.
- Independent analyses have found MAG comparable to Scopus, Web of Science, and other bibliographic databases.
- MAG powers a variety of open research tools, such as Semantic Scholar.



We find that MA covers journal articles, working papers, and conference items to a substantial extent and indexes more document types than the benchmark databases (e.g., working papers, dissertations)…The coverage of MA is favorable for evaluative bibliometrics in most research fields, including economics/business, computer/information sciences, and mathematics.

- Hug & Brändle (2017), *Scientometrics* 113, 1551–1571

Garbage In, Garbage Out

- Previous analyses have taught us the critical importance of filtering out low-quality research prior to conducting any landscape analytics.
- Traditional approaches using citation-based quality metrics are problematic.
  - Citations are easily gamed by journals and individual researchers.
  - Citation patterns differ across technical fields.
  - Method of calculation is often opaque.
- Alternative metrics exist - but have their own problems.

Our Approach to Eliminating Low-Quality Research

COLLECT INITIAL DATA SET

363,979 documents from 10,558 journals and 2,272 conferences.

MEASURE VENUE QUALITY

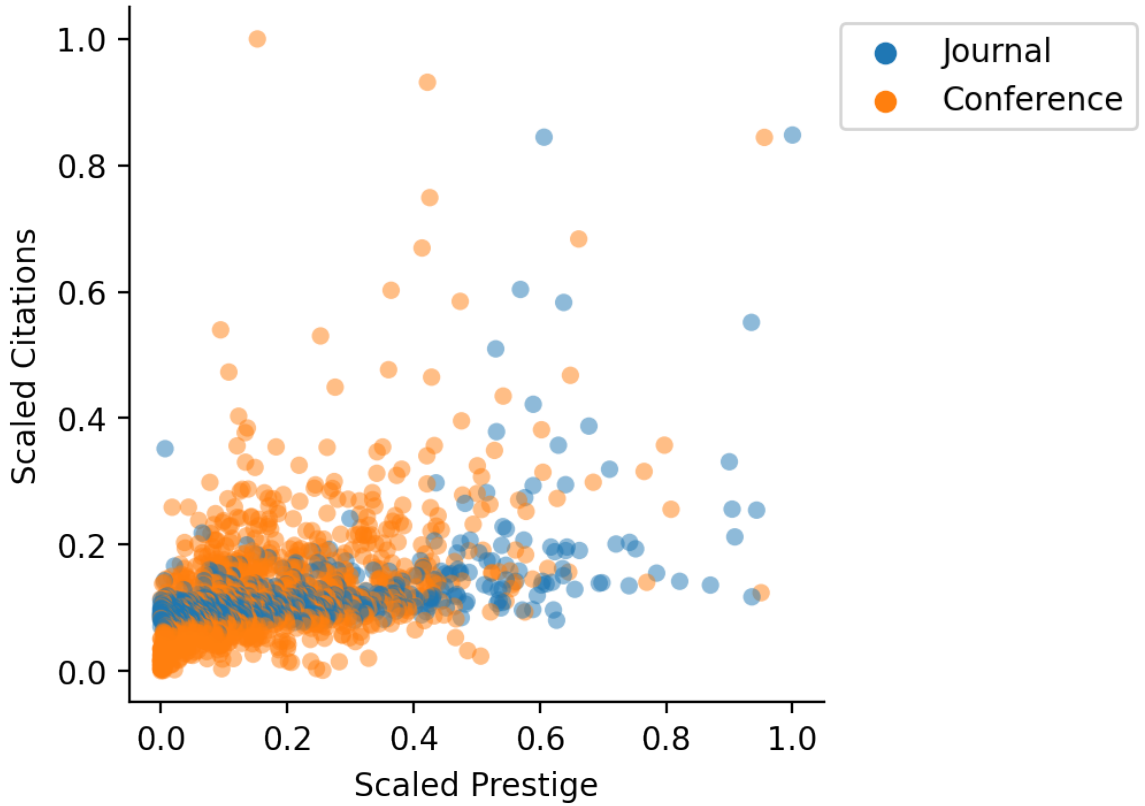
Model quality as a function of the prestige of organizations publishing in a given journal/conference and normalized citation rates.

FILTER ON QUALITY

99,153 documents from 1,061 journals and 856 conferences.

Quality

Prestige- and citation-based quality scores for all publication venues appearing in AI/ML data set



Sample of the top 10 publication venues based on weighted quality metric

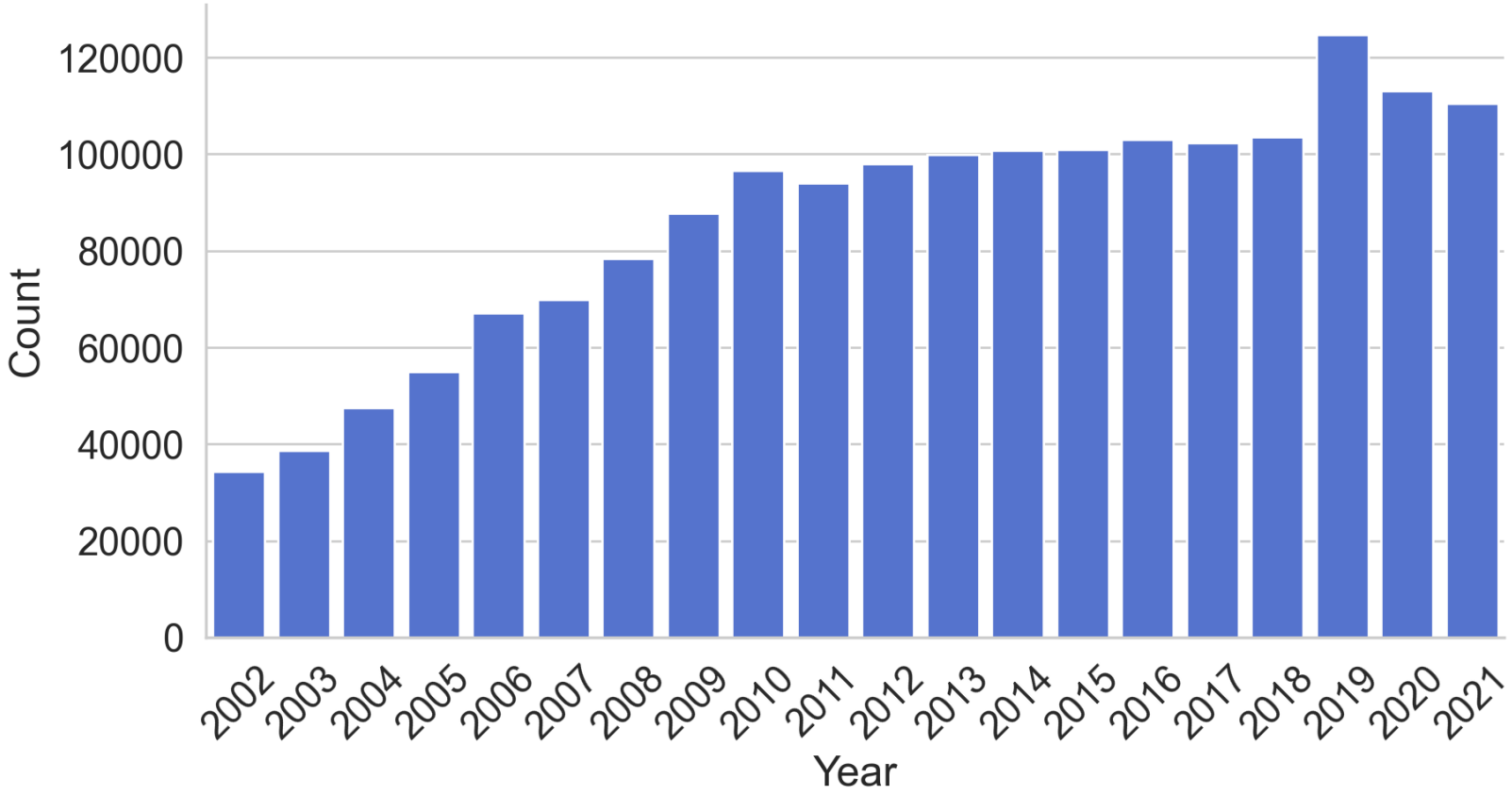
International Conference on Learning Representations (ICLR)	Conference
Transactions of the Association for Computational Linguistics	Journal
Conference on Neural Information Processing Systems (NeurIPS)	Conference
IEEE Conference on Computer Vision and Pattern Recognition	Conference
Nature Medicine	Journal
Nature Neuroscience	Journal
International Symposium on Computer Architecture	Conference
European Conference on Computer Vision	Conference
International Symposium on Microarchitecture	Conference
IEEE Transactions on Pattern Analysis and Machine Intelligence	Journal

The AI/ML Landscape

BluePath Labs
www.bluepathlabs.com

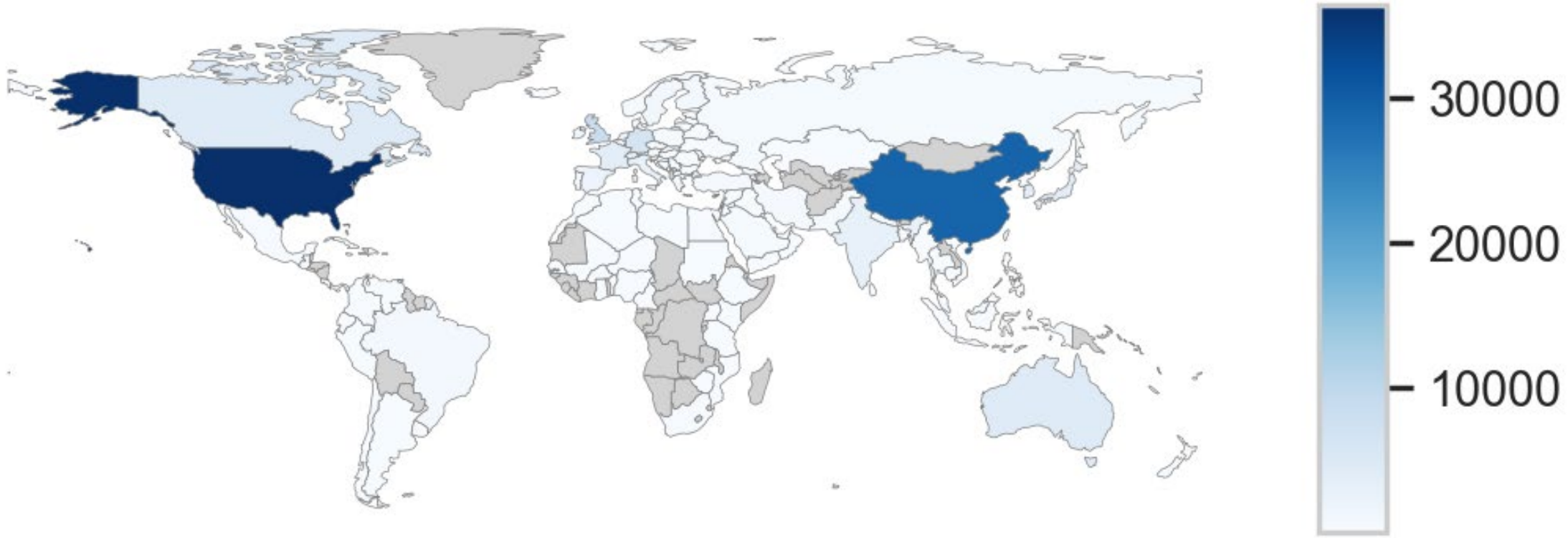
AI/ML is Growing

Total number of AI/ML publications per year over the last 20 years



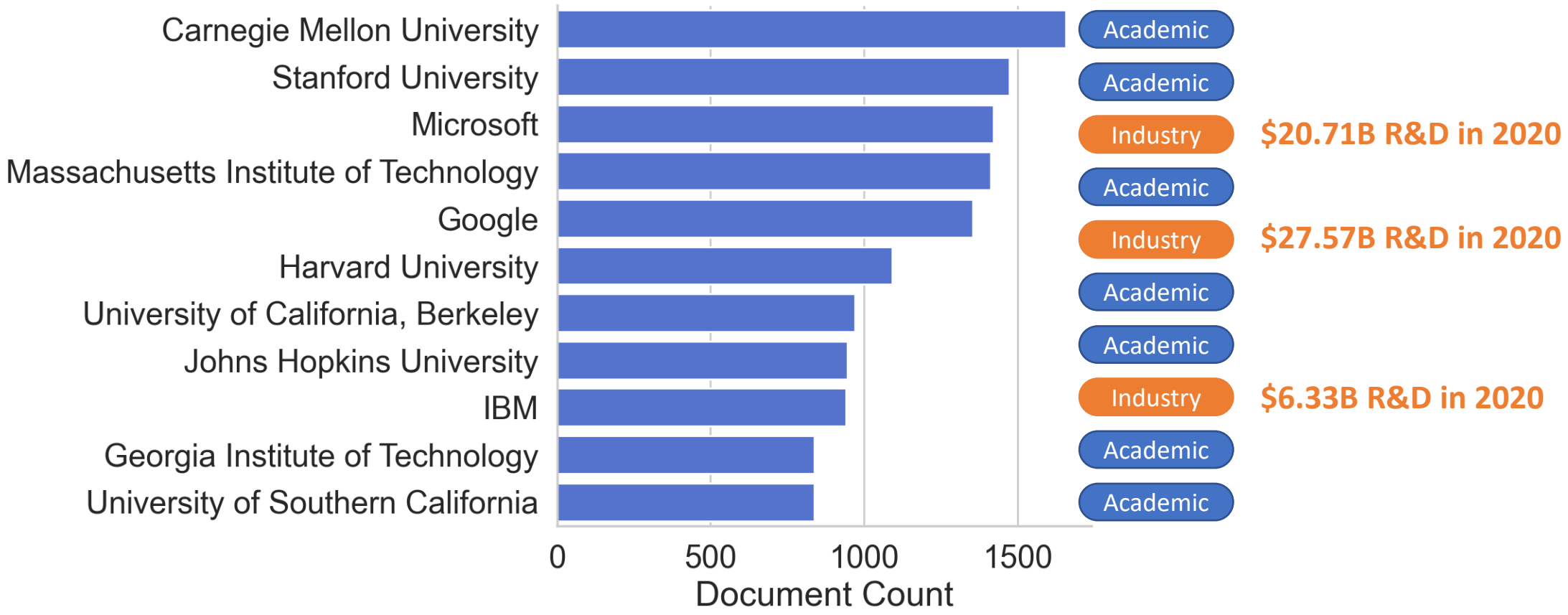
The US and China Lead AI/ML Research

Global leadership in AI/ML research, 2017-2021 (filtered for quality)



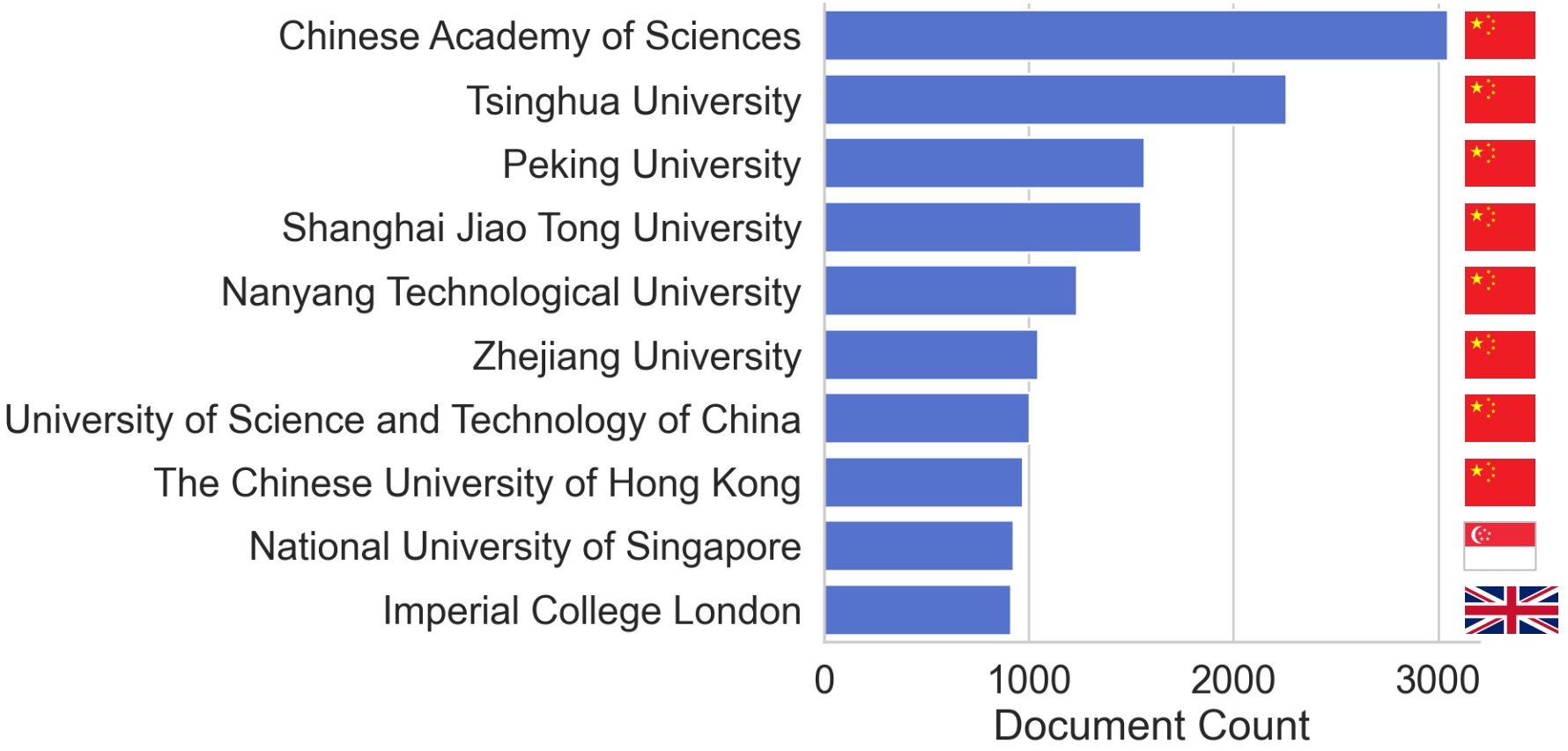
Who are the leading U.S. Research Institutions?

Leading U.S. research institutions in AI/ML, 2017-2021



Who Are the Leading International Research Institutions (Non-US)?

Leading international research institutions in AI/ML, 2017-2021

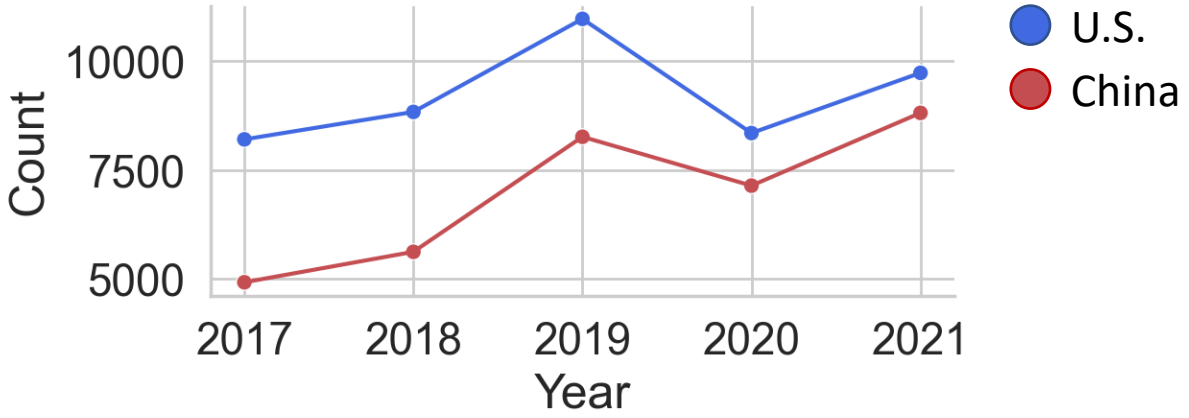


Most Active Chinese Key Laboratories (CAS)

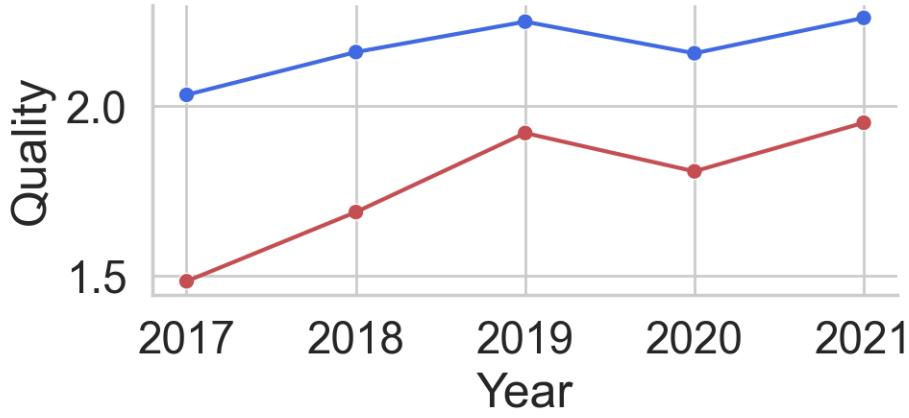
Name	Count
State Key Laboratory of Management and Control for Complex Systems	249
State Key Laboratory of Robotics	122
State Key Laboratory of Transient Optics	104
State Key Laboratory of Computer Architecture	94
State Key Laboratory of Information Security	69
Beijing Key Laboratory of Micro-Nano Energy and Sensors	62
CAS Key Laboratory of Molecular Imaging	43
CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems	36
State Key Laboratory of Information Security	30
State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing	30
State Key Laboratory of Vegetation and Environmental Change	30

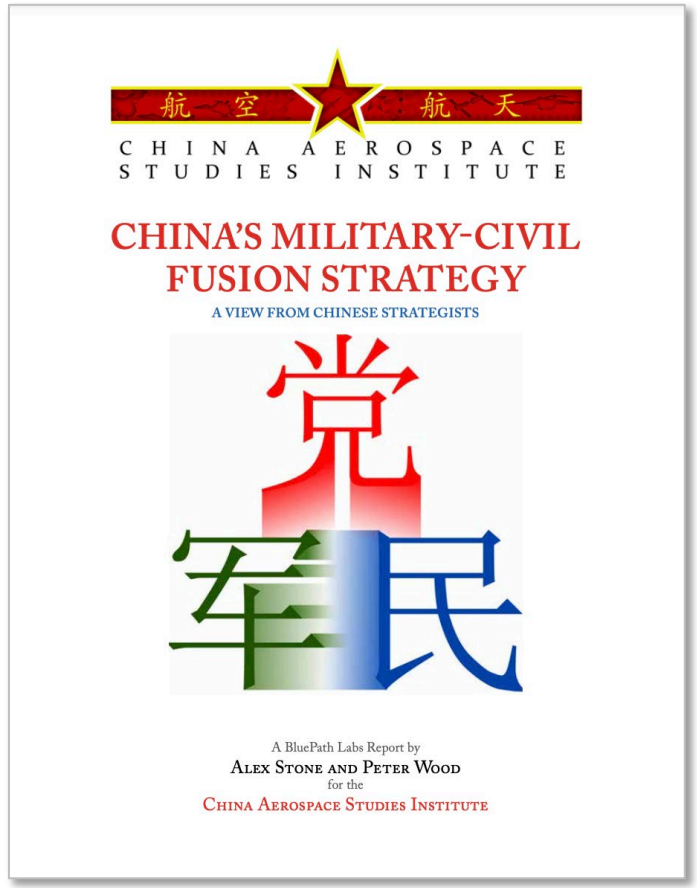
China vs. US: Comparing Output and Quality

Publication count by year, 2017-2021



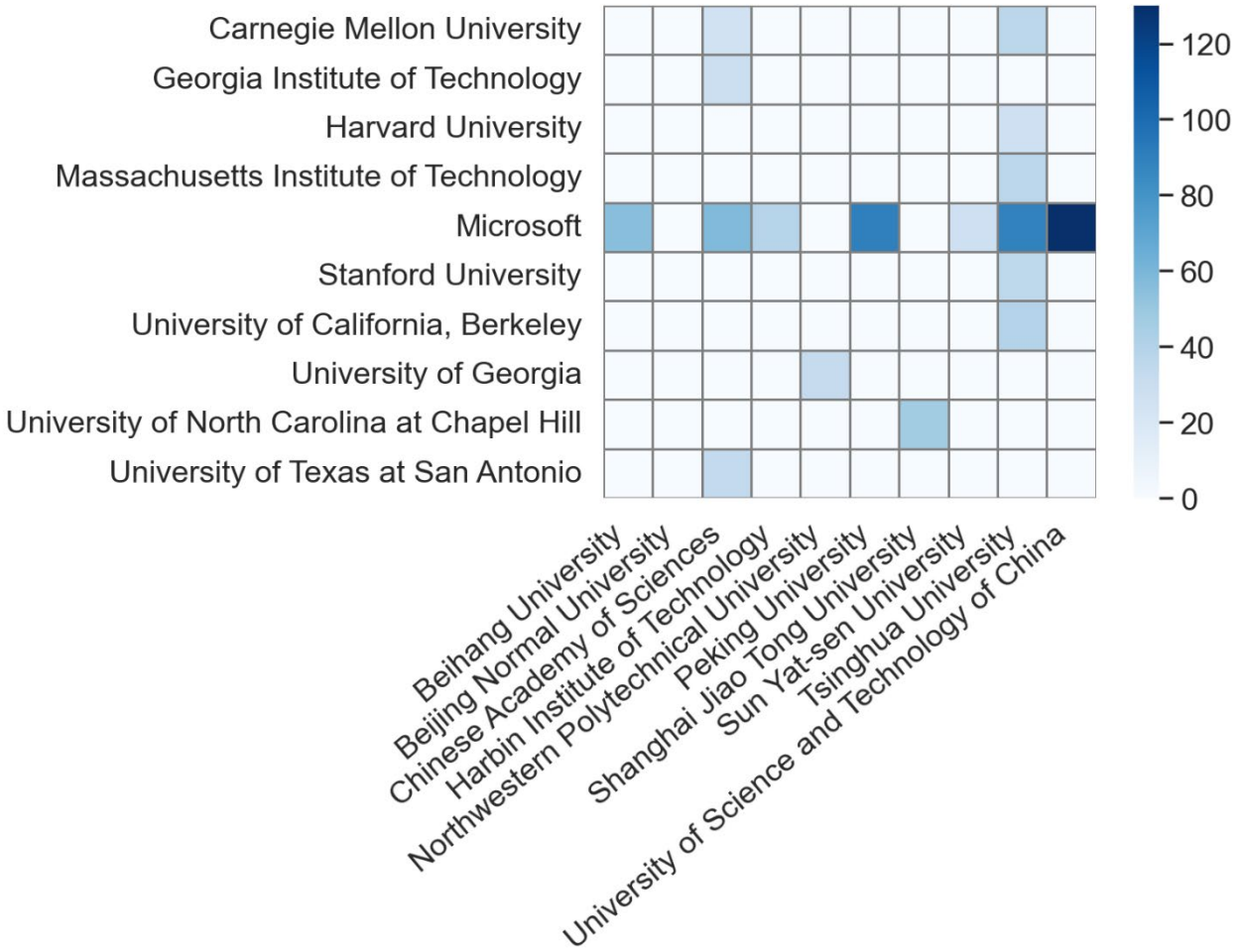
Mean publication quality by year, 2017-2021





China & US: Who's Working With Who?

Most active U.S.-China collaborations, 2017-2021



6%

Overall collaboration rate between the U.S. and China 2017-Present

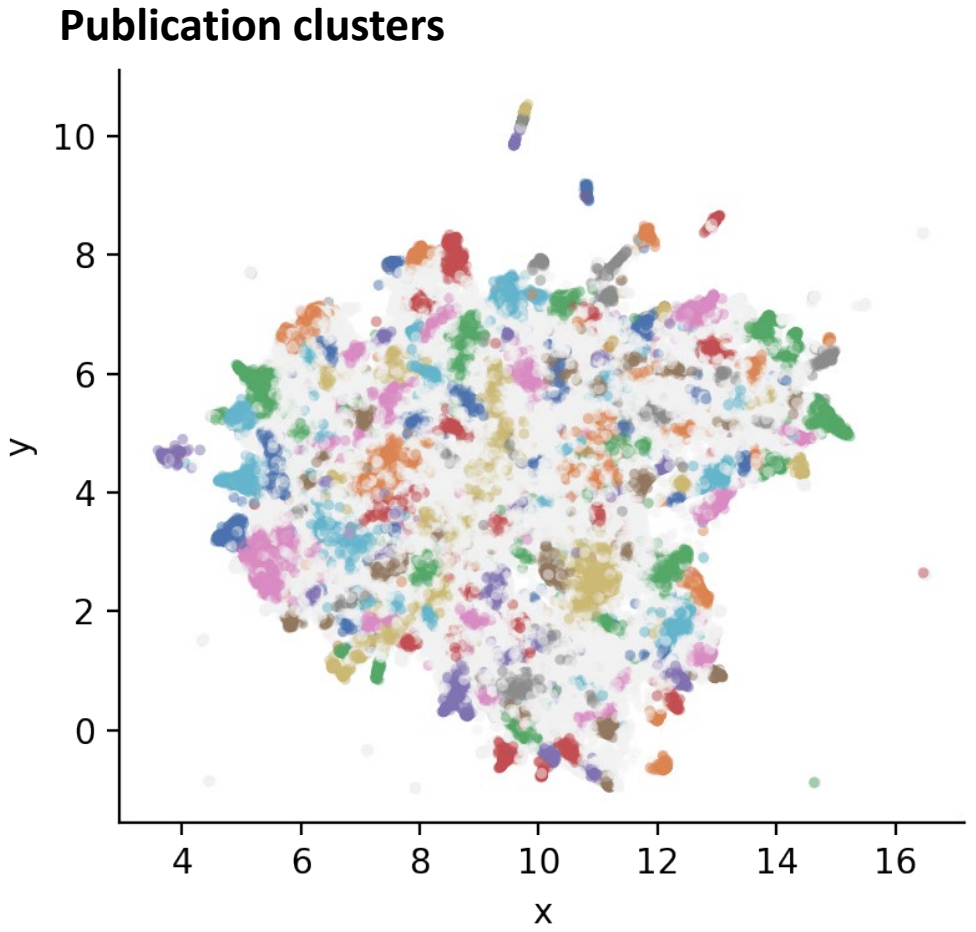
Institutions With the Most US-China Collaborations

Name	Count
Tsinghua University	684
Chinese Academy of Sciences	646
Peking University	380
Shanghai Jiao Tong University	362
University of S&T of China	312
Zhejiang University	299
Nanyang Technological University	230
Beihang University	222
Tencent	213
University of Electronic S&T of China	187

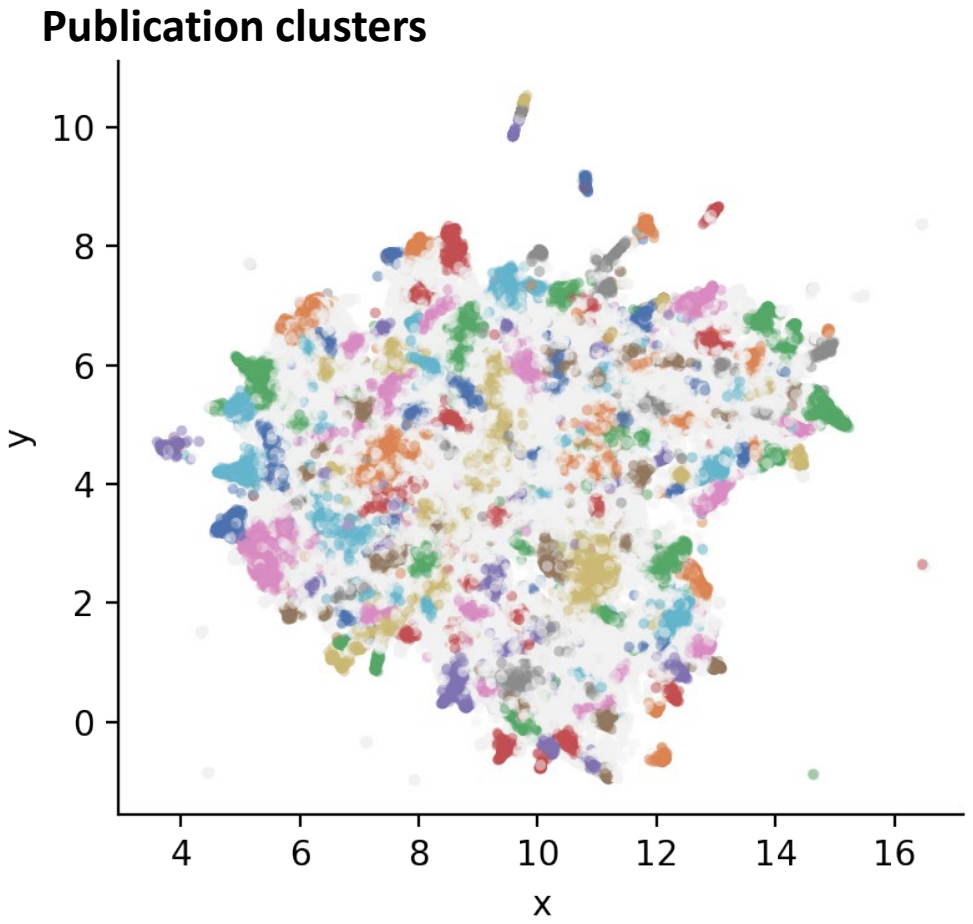
Name	Count
Microsoft	567
Carnegie Mellon University	229
UNC Chapel Hill	187
Georgia Institute of Technology	145
Stanford University	145
U. Illinois at Urbana-Champaign	139
University of Southern California	139
Northeastern University	135
University of California, Berkeley	127
Harvard University	124

Research Landscape Methodology

- Used text embedding model to convert document titles and abstracts into feature vectors
- Reduced vectors to low-dimensional space
- Clustered reduced vectors using density-based clustering
- Resulted in 108 research clusters
- Grouped similar clusters
- Identified fastest-growing trends



Research “Megaclusters”



Megacluster	Subclusters
Computer vision	39
Health	38
Natural language processing	25
Learning algorithms	24
Commercial applications	12
Human-computer interaction (HCI) and robotics	11
Consumer applications	7
Multimodal computer perception	6
Cybersecurity	5
Information retrieval	5

Computer Vision

Publications

12,882

(13.0%)

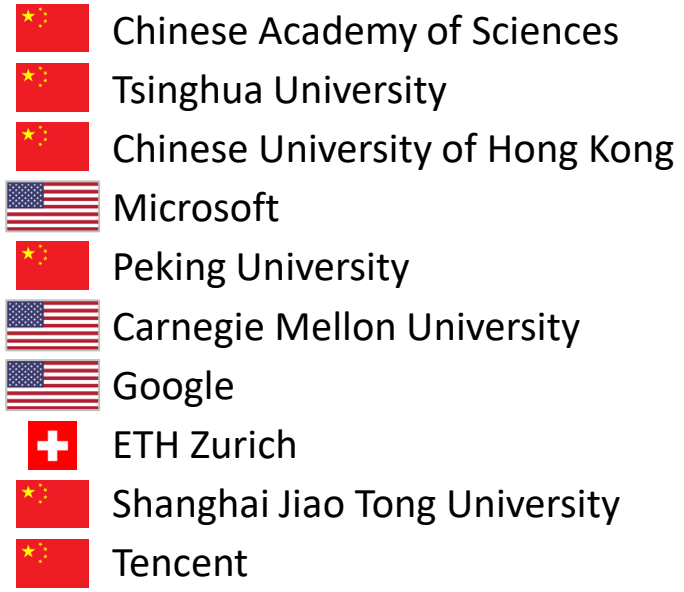
Growth

3.81%

Key Subtrends

- Person reidentification
- Semantic segmentation of scenes
- Thermal and infrared imaging
- Action recognition
- Face recognition image/video synthesis
- Image quality enhancement
- Depth estimation
- Change/anomaly detection
- Salient object detection

Leaders



US vs. China Balance

Collaboration rate: 11%

Health

Publications

11,308

(11.4%)

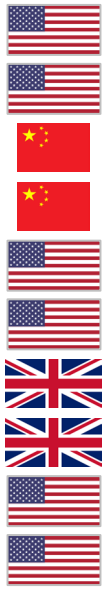
Growth

12.69%

Key Subtrends

- Diabetes
- Heart disease
- Cancer
- Multiple sclerosis
- Parkinson's disease
- Alzheimer's disease
- Autism
- Depression
- Seizure prediction
- Medical image analysis
- Epidemiology (COVID-19)
- Electronic health records
- Pharmaceutical research
- Surgical robots

Leaders



Harvard University
Stanford University
Chinese Academy of Sciences
Shanghai Jiao Tong University
Johns Hopkins University
UNC at Chapel Hill
Imperial College London
University College London
Massachusetts Institute of Technology
University of Pennsylvania

US vs. China Balance

Collaboration rate: 10%

Natural Language Processing

Publications

5,875

(5.9%)

Growth

5.85%

Key Subtrends

- Text classification
- Text embedding
- Topic modeling
- Text summarization
- Dialogue systems
- Machine reading
- Neural machine translation
- Multilingual NLP
- Sentiment analysis
- Neural grammatical error correction
- Language relation extraction
- Named entity recognition

Leaders


Microsoft


Carnegie Mellon University


Tsinghua University


Chinese Academy of Sciences


Google


Peking University


Tencent


IBM


Facebook


University of Washington

US vs. China Balance

Collaboration rate: 10%

Learning Algorithms

Publications

6,035

(6.1%)

Growth

5.17%

Key Subtrends

- Domain adaptation
- Neural architecture search
- Deep metric learning
- Transfer learning
- Graph neural networks
- Multitask/multilabel learning
- Learning on sparse data
- Zero-shot and few-shot learning
- Generative adversarial networks
- Fairness and bias in AI/ML
- Spiking neural networks
- Fundamental algorithms (e.g., clustering)

Leaders


Google


Chinese Academy of Sciences


Carnegie Mellon University


Tsinghua University


University of California, Berkeley


Stanford University


Massachusetts Institute of Technology


Microsoft


University of Oxford


University of Sydney

US vs. China Balance

Collaboration rate: 9%

Commercial Applications

Publications

3,526

(3.6%)

Growth

3.33%

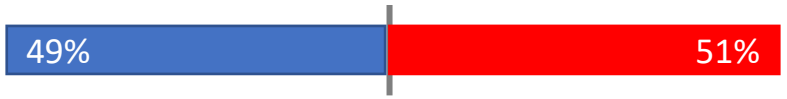
Key Subtrends

- Photovoltaics
- Construction site safety
- Manufacturing
- Structural damage and defect detection
- Machine fault diagnosis
- Agriculture
- Environmental modeling and measurement
- Forestry
- Educational applications
- Internet of things
- Stock market prediction

Leaders

- Tsinghua University
- Chinese Academy of Sciences
- Nanyang Technological University
- Huazhong University of S&T
- Beijing University
- Zhejiang University
- Hong Kong Polytechnic University
- University of California, Berkeley
- Carnegie Mellon University
- Peking University

US vs. China Balance



Collaboration rate: 12%

HCI and Robotics

Publications

5,566

(5.6%)

Growth

-0.02%

Key Subtrends

- Self-driving car perception
- Traffic flow prediction
- Unmanned aerial vehicles
- Pressure sensors
- Tactile and haptic sensing
- Robot manipulation
- Prosthetics and exoskeletons
- Human-robot and robot-robot interaction
- Group decision-making models
- Motor imagery measurement
- Saccades and gaze

Leaders


Massachusetts Institute of Technology


Chinese Academy of Sciences


University of California, Berkeley


Tsinghua University


Carnegie Mellon University


Stanford University


Georgia Institute of Technology


Nanyang Technological University


ETH Zurich


Shanghai Jiao Tong University

US vs. China Balance

Collaboration rate: 8%

Consumer Applications

Publications

1,090

(1.1%)

Growth

-6.79%

Key Subtrends

- Mobile AI/ML applications
- eCommerce applications
- Food
- Music
- Sports
- Fashion
- Virtual reality

Leaders


Stanford University


Alibaba Group


Tsinghua University


Chinese Academy of Sciences


National University of Singapore


Peking University

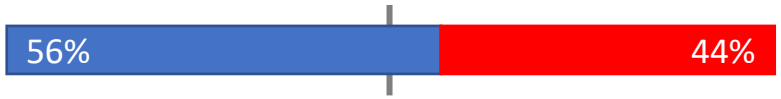

Nanyang Technological University


Zhejiang University


Massachusetts Institute of Technology


KAIST

US vs. China Balance



Collaboration rate: 10%

Multimodal Computer Perception

Publications

840

(0.8%)

Growth

-2.38%

Key Subtrends

- Acoustic scene classification
- Audio source separation
- Underwater imaging
- Texture classification and synthesis
- Radar
- Ultrasound image reconstruction

Leaders


Imperial College London


Chinese Academy of Sciences


University of Oxford


Shenzhen University


Johns Hopkins University


King's College London

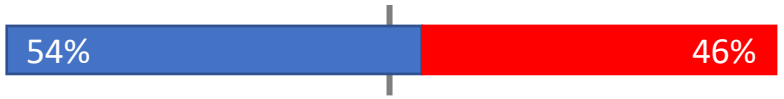

Carnegie Mellon University


Eindhoven University of Technology


The Chinese University of Hong Kong


Philips

US vs. China Balance



Collaboration rate: 4%

Cybersecurity

Publications

1,268

(1.3%)

Growth

23.08%

Key Subtrends

- Fingerprint identification
- Malware detection
- Adversarial examples
- Software defect prediction
- Forgery detection

Leaders


Tsinghua University


Shanghai Jiao Tong University


Chinese Academy of Sciences


Microsoft


IBM


Michigan State University


University of California, Berkeley


Nanyang Technological University


University of Maryland, College Park


Norwegian University of S&T

US vs. China Balance

Collaboration rate: 9%

Information Retrieval

Publications

1,031
(1.0%)

Growth

2.51%

Key Subtrends

- Hashing
- Learning to rank
- Recommendation systems
- Knowledge graphs
- Image retrieval

Leaders


Tsinghua University


Chinese Academy of Sciences


Microsoft


University of Electronic S&T of China


Peking University


National University of Singapore


Tencent


Shanghai Jiao Tong University


Nanyang Technological University


Google

US vs. China Balance

Collaboration rate: 13%

Takeaways

- Dramatic progress in ML over the past 20 years, fueled mostly by the US and China.
- China is producing large amounts of quality AI/ML research.
- U.S. tech companies are prime movers in basic and applied research.
- Cybersecurity is a major growth area for AI/ML research.
- This analysis focused on basic and early applied research – patents would give a different view.

What Are the Future Prospects for AI/ML?

Strengths

- Actual success in valuable use cases
- Expanding computational power
- Open-source tools
- Open data sources
- Cloud computing

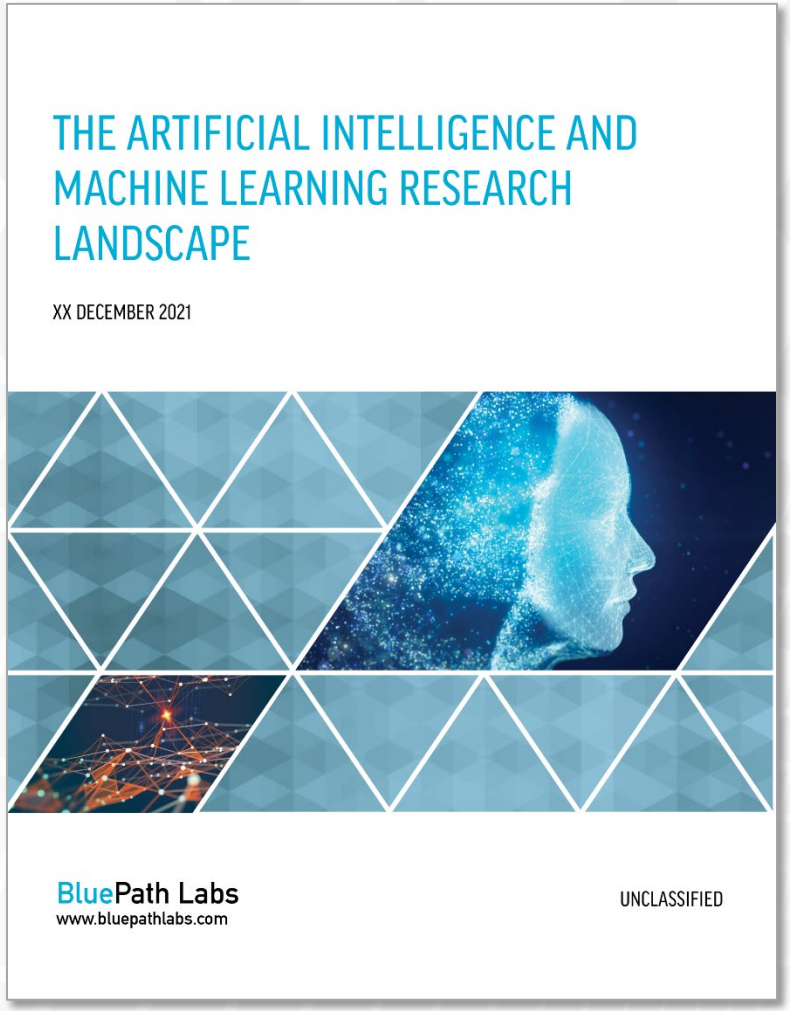
Threats

- Peak of inflated expectations
- Prohibitively high cost of training models could constrain open innovation
- Diminishing returns on increasing model size
- Algorithmic bias
- Public backlash

Machine Learning Is Different Than Artificial General Intelligence (AGI)

Thank you

Jason Augustyn, Ph.D.
Chief Technology Officer
BluePath Labs
jaugustyn@bluepathlabs.com



BluePath Labs
www.bluepathlabs.com