Emerging Trends in Artificial Intelligence and Machine Learning

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

BluePath Labs www.bluepathlabs.com

What Is AI/ML? Brought to you by Microsoft

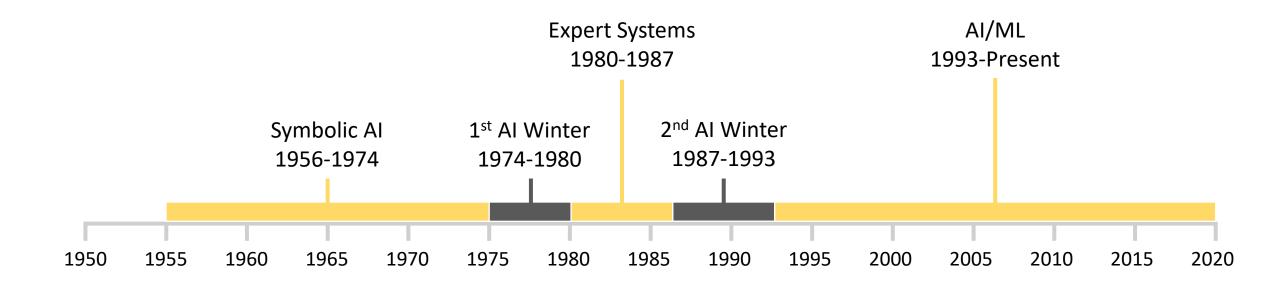
Artificial Intelligence

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?



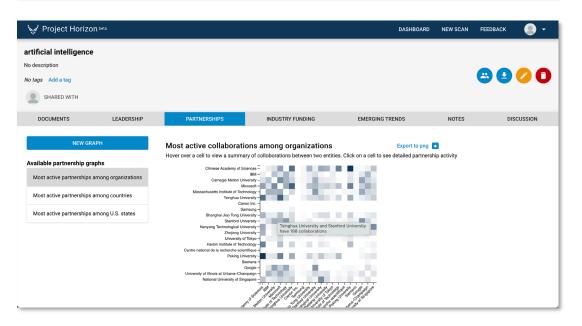
Landscape Analysis Methodology

BluePath Labs www.bluepathlabs.com

Methodology

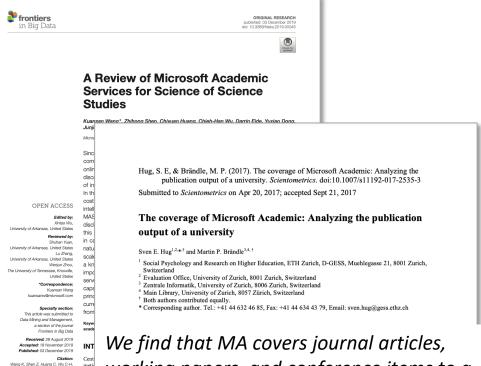
- Analysis-employed tools developed for Project Horizon.
- Conducted broad search for academic journal articles and conference papers on AI/ML published in the past 5 years.
- Filtered documents to identify highquality, relevant research.
- Applied natural language processing (NLP)-based clustering techniques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

🦙 Project Horizon 🔤	а			DASHBOARD	NEW SCAN	FEEDBACK	•
artificial intelligence No description No tags Add a tag SHARED WITH						8 1	00
DOCUMENTS	LEADERSHIP	PARTNERSHIPS	INDUSTRY FUNDING	EMERGING TRENDS	NOTES	DI	SCUSSION
Search within results	4	4,446,200 documents			Sort by: Rele	evance	•
ilters		BOOK Model-Based Machine Learning (2021) – DOI: Not available <u>Christopher M. Bishop</u>					
Affiliations	~	99 Citations: 75 Several decades of research in the field of n	anakina laavaina kawa waxultad in a ay	this do of different algorithms for onlying	- h	f nachlama. Ta taal	
Citation range	~	application, a researcher typically tries to m					NIC A HEW
Conference Instance	~	Save to library					
Publication year	~	JOURNAL ARTICLE					
Countries	~	Recurrent Neural Networks for Time S International Journal of Forecasting (2021)		s and Future Directions			
Journals	~	Hansika Hewamalage, Christoph Bergmeir &					



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.

UNCLASSIFIED

Ede D, Dong Y, Qian J, Kanakia A, Chen A and Rogahn R (2019) A

Review of Microsoft Academic Services for Science of Science Studies. Front. Big Data 2:45.

doi: 10.3389/fdata 2019.0004

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.

in Psychology				paleet 20 Agat 1919 ex: 10.338/stype 2010.0147 Execution
	Th	e Impact Factor Fa	alla	су
	Friede	r M. Paulus1*, Nicole Cruz2.3* and Sören Kra	ich1*	
	Departm	next of Developing and Developing	I ab. I fai sa	mba el Alash- I Alash- Demons
	² Depa École			
	The			
	man			
	Wea	A brief hi		e Journal Impact Factor: ritique, and discussion of adverse effects
	and To a			
	illust		Vincent I	Larivière ^{1,2} & Cassidy R. Sugimoto ³
	the .			s sciences de l'information, Université de Montréal, Canada.
	or in	Observatoire des science la science et la ten	s et des ti hnologia	echnologies (OST), Centre interuniversitaire de recherche sur
OPEN ACCESS	Keyw	3 School of Infe		🟁 © 1998 Nature America Inc. • http://neurosci.nature.com
Edited by:	Conn.			edito
Uhich Hoffrage, Université de Lausanne, Switzerland	INT	Table of Contents		
Reviewed by: Zaida Chinchila-Rodriguez,	The j			
Zada Chinchilla-Hodriguez, Consejo Superior de Investigaciones Científicas (CSIC), Spain	(Gari In d€	Abstract		nature
Andy Wei Kan Yeung, University of Hong Kong, Hong Kong	publi then,	2. Calculation and repro-		nature <i>neuroscience</i>
*Correspondence:	well	3. Critiques		110111105C1011C0
Frieder M. Paulus paulus@snt.uni-luebeok.de	to m	3.1 The numerator / d		<i>iwmoscmm</i>
Nicole Cruz ncruzd01@mail.bbk.ac.uk	assoc and 1	3.2 Journal self-citatic		
Sören Krach krach@snl.uni-kebeck.de	aim 1 visibi	3.3 Length of citation		Citation data: the wrong impact?
Specialty section:	Gorr maki	3.4 Skewness of citati		Citation data: the wrong impact?
This article was submitted to Cognition,	bonu	3.5 Disciplinary comp 3.6 Journal Impact Fa		
a section of the journal Frontiers in Psychology	whos (Segl	4. Systemic Effects		Every September, a ripple of excitement passes through the sci- entific community as the Institute of Scientific Information (ISI) other. There are fewer ecology papers published, so the
Received: 02 December 2017 Accepted: 27 July 2018	Casa and I	4.1 Journal Impact Fa	E C	publishes its latest set of impact factors, in which some six thou- sand journals are ranked according to the number of citations. Here is somewhere in the dle, but it seems likely that within the field, the most high
Published: 20 August 2018	¹ A fur	4.2 Role of evaluation	ture.	they received in the previous year. The release of these results papers tend to be on molecular and cellular rather than triggers elation or gloom in editorial offices around the world.
Citation: Paulus FM, Cruz N and Krach S (2018) The Impact Factor Falacy.	article the ar	4.3 Application at the	sci.n	but for many scientists it is no more than light entertainment, the scientific equivalent of tabloid gossip. For others, however, it least when comparing across fields) that impact factors w
2016) The Impact Pactor Falacy. Front. Psychol. 9:1487. doi: 10.3389/bsvo.2018.01487	subsec if auth	4.4 Knock-off indicate	neuro	represents something more serious, because their career prospects are increasingly affected by the impact factors of the journals in senting pioneering work in new areas will receive fewer c
dd: 10.3368/psgg.2016.01467	compe	 What are the alternative The future of journal i 	http://	which they publish. Although bibliometric data undoubtedly have the potential to reveal significant insights into the quality Although these limitations are (or should be) well
Frontiers in Psychology www.frontiersin.s	org	o. The future of journal i	Inc. •]	of scientific work, they are also susceptible to abuse. It is therefore worth examining in some detail how they are derived and how
				they are now being applied to potential subscribers. Nature Neuroscience is of course
	-		a America	ISI is a commercial company, based in Philadelphia, which publishes Science Clation Index and Current Contents in addition Iso Journal Clation Roperts, where impact factors are reported. attention to numbers that they believe reflect well on their attention to numbers that they believe reflect well on their
			1998 Nature	The impact factor for a given year—say, 1997—is calculated as follows: ISI counts the number of citations made in 1997 to papers
			1998	published in the previous two years, 1995 and 1996, and divides by the number of articles published in that two-year period divides by editors to inflate them by artificial means: David Pendlel
			8	The number thus derived is biased in several ways that are not always fully appreciated ¹ . Most obviously, by the time the impact
		Forthcoming in Glänze Science and Techn		factors appear, the papers to which they refer are already two to can manipulate it to their journal's advantage. Needless
				three years old, so any recent changes in a journal's editorial poli- cies will not be reflected in its impact factor. (This is partly avoid- id by looking at the 'immediacy index', which is the average are some strategies: publish more reviews, which is the average
	- 5			
				number is no more than a snapshot, and papers appears but this number is no more than a snapshot, and papers appear ar ing early in the year will be cited more than those appearing later.) According to ISI, the great majority of citations are almost
				invariable to a small faction of the total articles, and so the impact factor, which is the mean citation rate, is a poor measure the total number of citations to any item in the journal,
				of the typical paper in that journal; this is true of high- and low- impact journals alike. In fact, most papers are cited at much lower
				rates than the journal's impact factor would suggest. Giving a disproportionate weight to the most highly cited papers is not nake do various and betters are not normally connected.
				necessarily a disadvantage if the aim is to measure the usefulness is some correlation between quality and impact factor. Me
				of a journal to its field—assuming that the more highly cited papers are likely to be the more significant ones—but it does
				papers are unkey to be the more againctant order—out it does mean that little can be inferred doubte likely citation of an one, casuathered by Eugene Gardiel, due founder out it individual paper from simply knowing the impact factor of the journal it which is appeared. Most importantly, however, different fields have different the set of the set observed with impact factor?
				journal in which it appeared. Most importantly, however, different fields have different ple have become so obsessed with impact factors?
				reflects the tonics it covers. Molecular biology, for instance, tends ly used for a purpose for which they were never intended
				nature neuroscience • volume 1 no 8 • december 1998

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

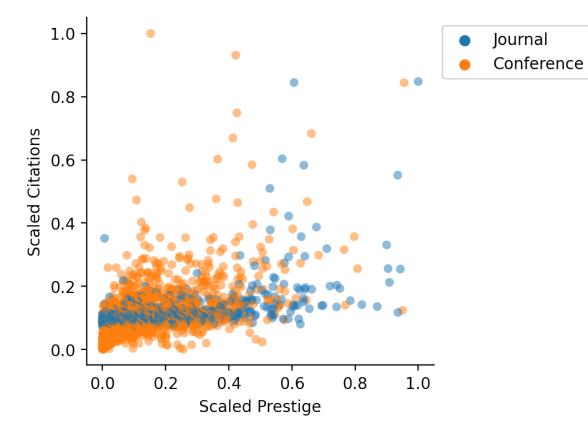
Model quality as a function of the prestige of organizations publishing in a given journal/conference and normalized citation rates.

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

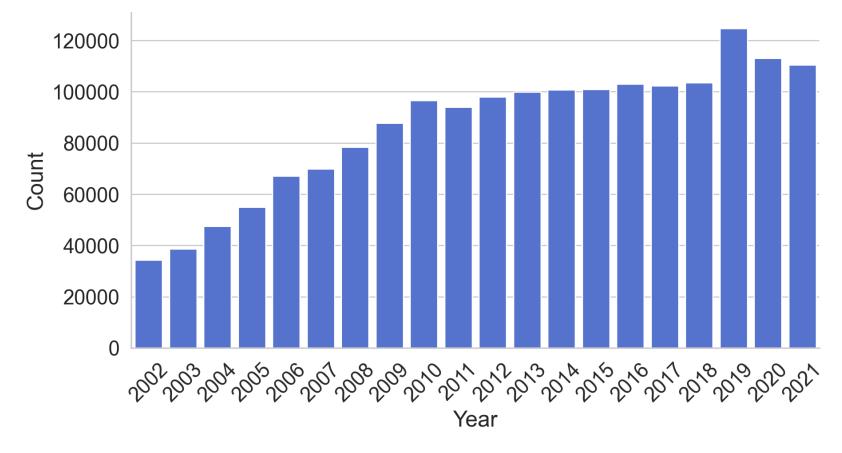
UNCLASSIFIED

The AI/ML Landscape

BluePath Labs www.bluepathlabs.com

AI/ML is Growing

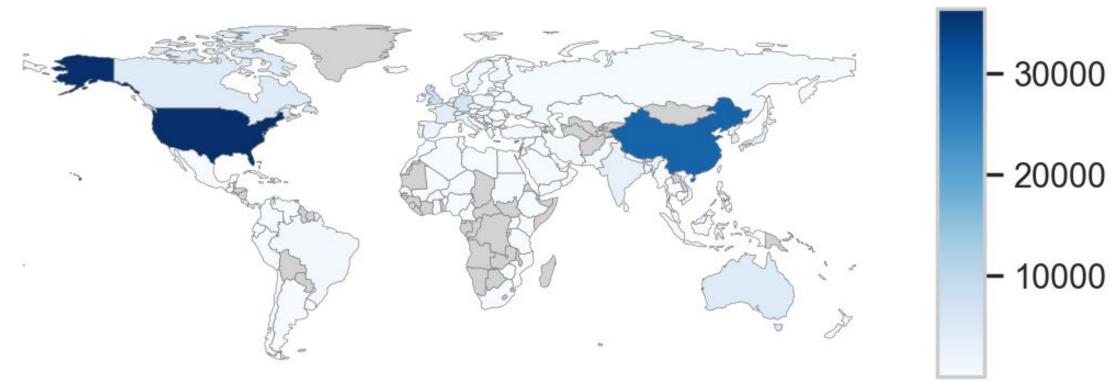
Total number of AI/ML publications per year over the last 20 years



UNCLASSIFIED

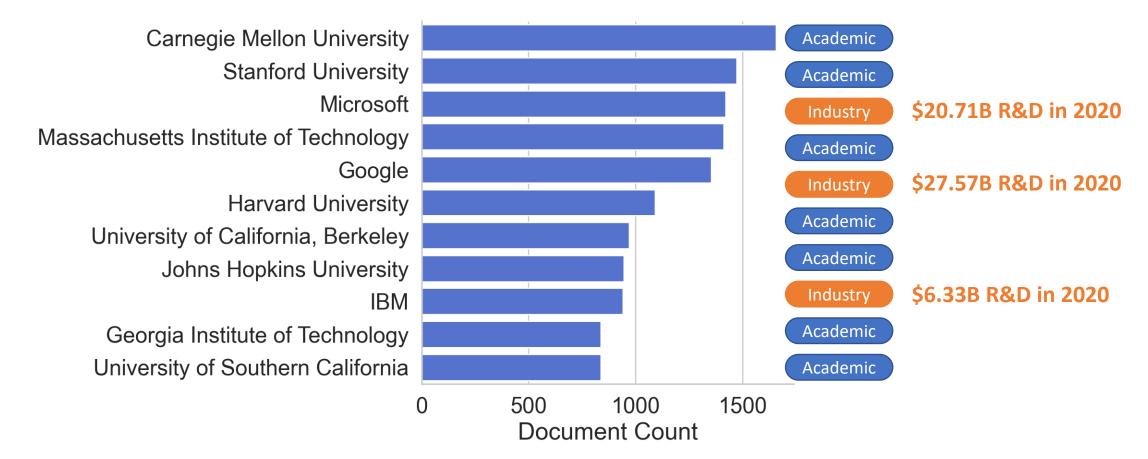
The US and China Lead AI/ML Research

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



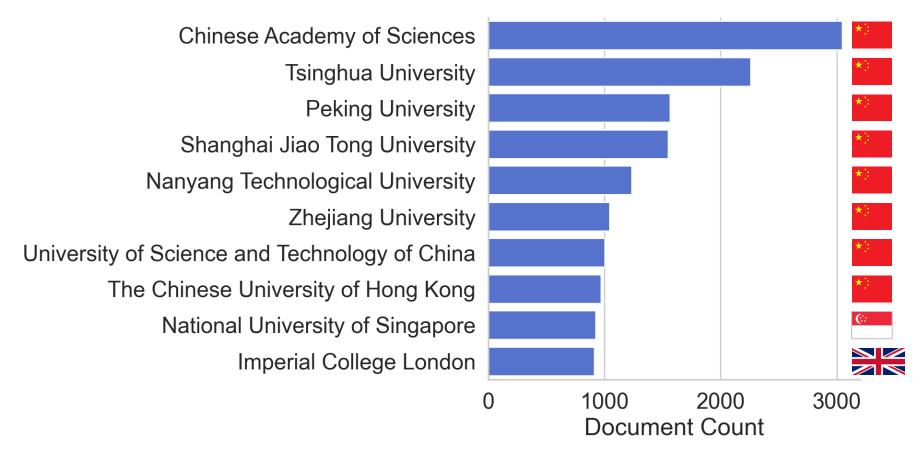
Who are the leading U.S. Research Institutions?

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

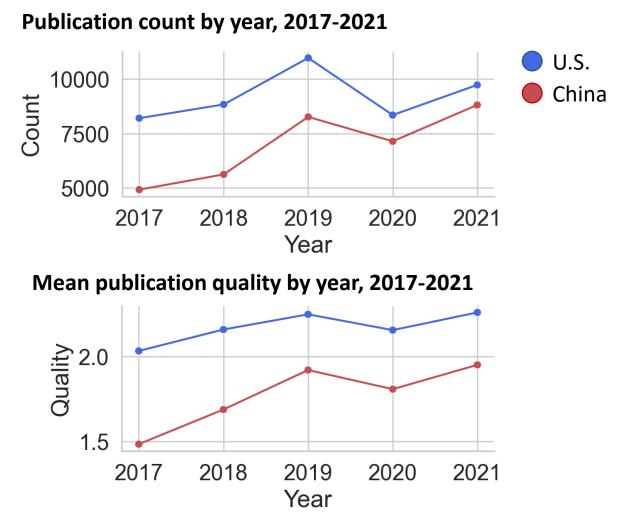


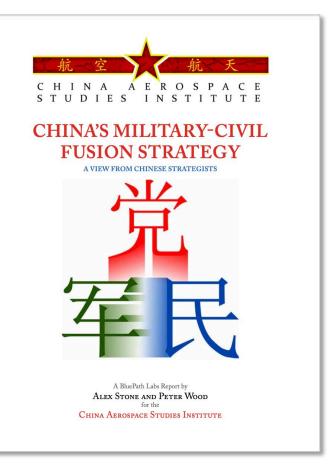
UNCLASSIFIED

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

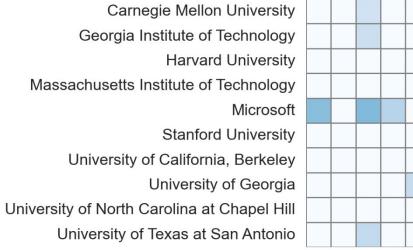


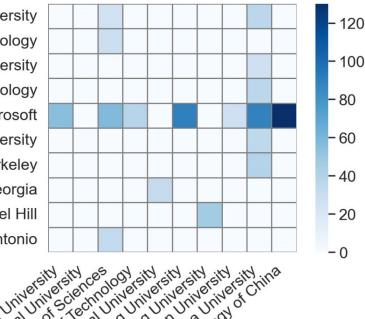


UNCLASSIFIED

China & US: Who's Working With Who?

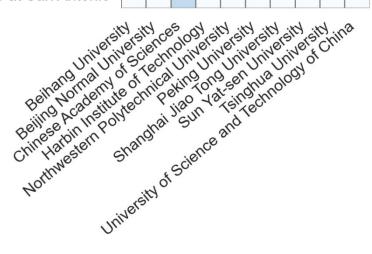
Most active U.S.-China collaborations, 2017-2021





6%

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



Institutions With the Most US-China Collaborations

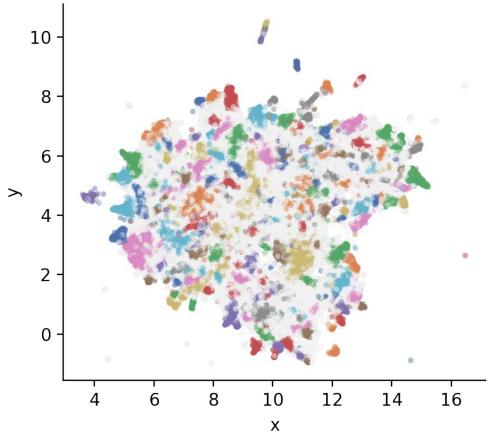
Name	Count
Tsinghua University	684
Chinese Academy of Sciences	646
Peking University	380
Shanghai Jiao Tong University	362
University of S&T of China	312
Zhejiang University	299
Nanyang Technological University	230
Beihang University	222
Tencent	213
University of Electronic S&T of China	187

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

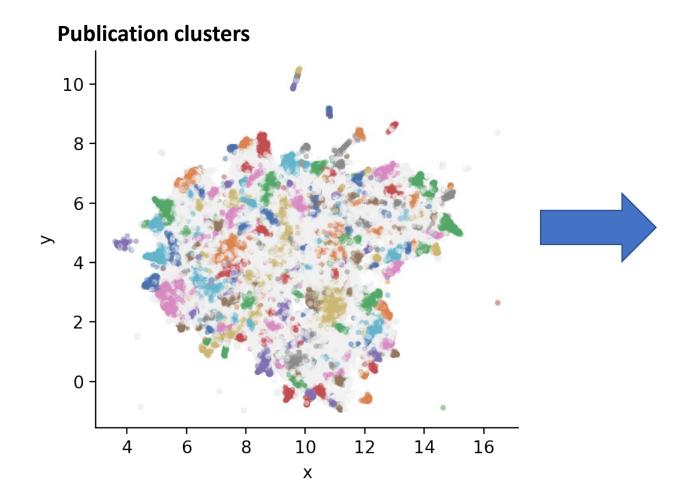
Research Landscape Methodology

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

Publication clusters



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

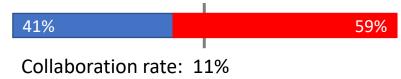
3.81%

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

*1	Chinese Academy of Sciences
*1	Tsinghua University
*1	Chinese University of Hong Kong
	Microsoft
*)	Peking University
	Carnegie Mellon University
	Google
÷	ETH Zurich
*1	Shanghai Jiao Tong University
*1	Tencent



Health

Publications

11,308 (11.4%)

Growth

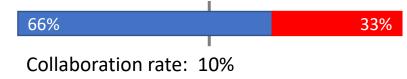
12.69%

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



Natural Language Processing

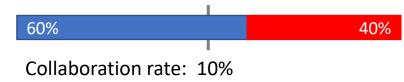
Key Subtrends

- 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



US vs. China Balance



Publications

5,875 (5.9%)

Growth

5.85%

Learning Algorithms

Publications

6,035 (6.1%)

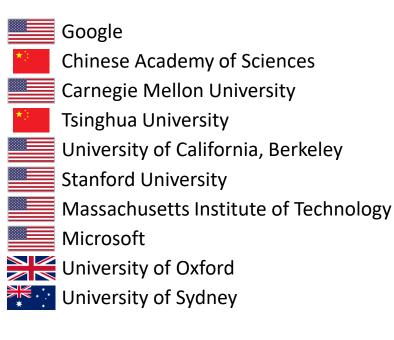
Growth

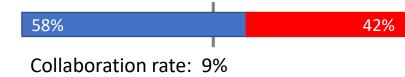
5.17%

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





Commercial Applications

Publications

3,526 (3.6%)

Growth

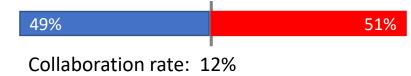
3.33%

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

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



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

	ETH Zurich
*	Georgia Institute of Technology Nanyang Technological University
	Stanford University
*:	Tsinghua University Carnegie Mellon University
	University of California, Berkeley
*>	Massachusetts Institute of Technology Chinese Academy of Sciences

Consumer Applications

Publications

1,090 (1.1%)

Growth

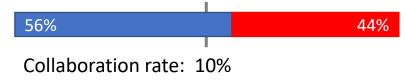
-6.79%

Key Subtrends

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

Leaders





Multimodal Computer Perception

Publications

840 (0.8%)

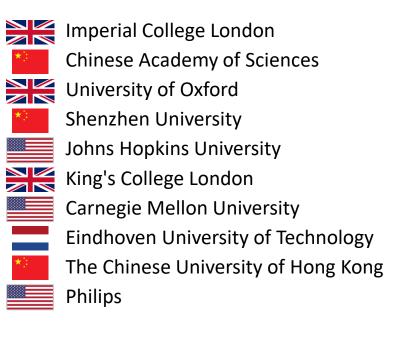
Growth

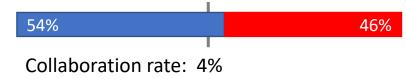
-2.38%

Key Subtrends

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

Leaders





Cybersecurity

Publications

1,268 (1.3%)

Growth

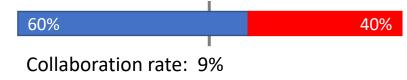
23.08%

Key Subtrends

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

Leaders

*1	Tsinghua University
*1	Shanghai Jiao Tong University
*1	Chinese Academy of Sciences
	Microsoft
	IBM
	Michigan State University
	University of California, Berkeley
*1	Nanyang Technological University
	University of Maryland, College Park
	Norwegian University of S&T



Information Retrieval

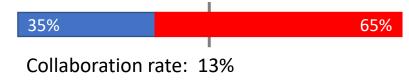
Key Subtrends

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

Leaders



US vs. China Balance



Publications

1,031 (1.0%)

Growth

2.51%

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 THE ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING RESEARCH LANDSCAPE

XX DECEMBER 2021



BluePath Labs www.bluepathlabs.com

UNCLASSIFIED

BluePath Labs www.bluepathlabs.com