Biology

nvironmental science Psychology Materials science

Molecular physics Geology Chemical physics Geography History Nuclear physic **SOCIOIO** Engineering physics Philosophy

ondensed matter physi Political science

Economics Classical m Astrophysic Theoretical physics

ry Computational science Software engineering Multimedia Real-time computing Computer science Computer security Embedded system Computer engineering Computer engineering Information retrieval Data science Data science Operations retrieval Data science Operations retrieval Data science Operations retrieval Data science Operations retrieval Data mining Mathematical physics Accounting Financial system Mathematical economics Financial economics Geometry

Mathematical optimization Computer hardware Computer hardware

Database Applied mathematics Programming language Algorithm Parallel computing Distributed computing Algorithm Parallel compu Co Arithmetic Theoretical computer science Mathematical analysis

Computer architecture

Finance

Keynesian economics Monetary economics

# Science of science

#### Analyzing the impact of Al on Science

Floriana Gargiulo



## **Traditions of Science studies**

**Scientific Revolutions** 

Innovations

**Scientific eras** 

History of science

**Scientific method** 

Philosophy of science

Rationalism

Epistemology

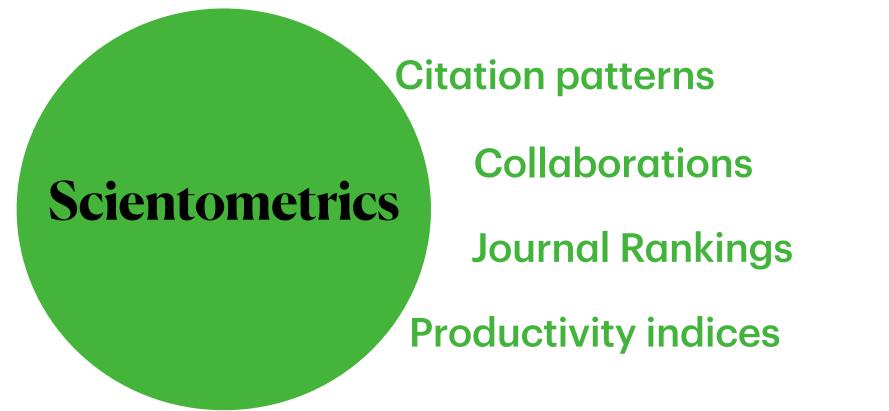
**Scientific Ethics** 

**Science Policy SCIENTIFIC COMMUNITY Scientific Norms** Public perception of science **Scientific institutions Sociology of** science **Citation patterns** Collaborations **Scientometrics Journal Rankings Productivity indices** Social constructivism **Science and Actor-Network Theory (ANT)** technology studies

Ethnography of science

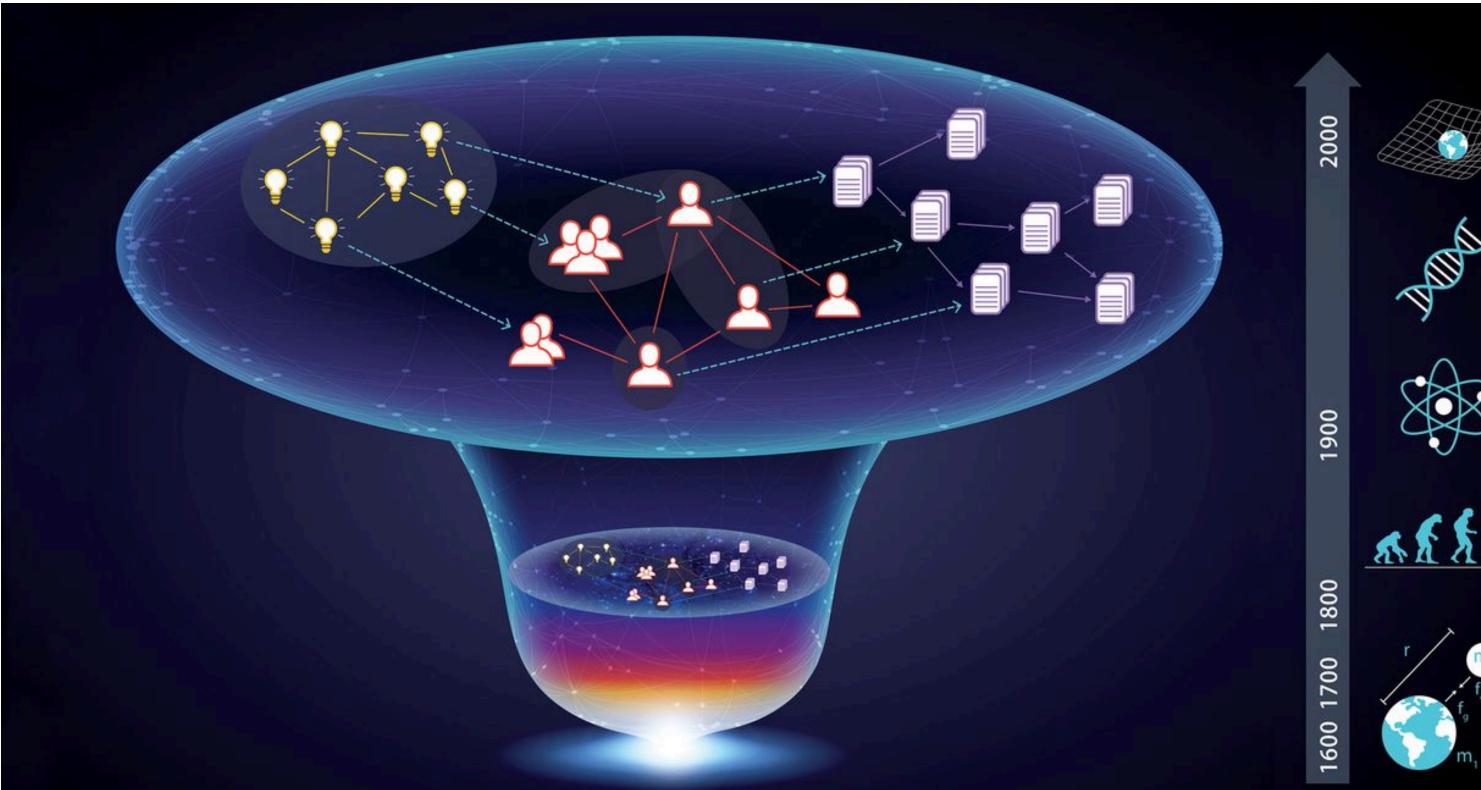
**Controversy mapping** 

## Science of science



#### Large-scale datasets

#### **Modeling tools**



#### Science as a complex system An evolving network of scientists, papers and concepts





# Artificial intelligence in science

"Few fields are untouched by the machinelearning revolution, from materials science to drug exploration; quantum physics to medicine." (Nature Editorial, 2019)

"The era of direct experimentation is gone, replaced by the era of data collection[...] Science takes place within the data." (Hey, Tansley, & Tolle, 2009)

AI can be characterized as a "general method of invention," which brings benefits and challenges to the science systems. (Bianchini, Muller, & Pelletier, 2022)





Is Al changing the epistemic structure and the scientific practices?



#### Is Al changing the epistemic structure and the scientific practices?

Floriana Gargiulo, Sylvain Fontaine, Michel Dubois, Paola Tubaro; **A meso-scale** cartography of the AI ecosystem. Quantitative Science Studies 2023; doi: https:// doi.org/10.1162/qss\_a\_00267

**O** What is AI? **O** How AI diffuses in the scientific ecosystem?

Fontaine, S., Gargiulo, F., Dubois, M., & Tubaro, P. (2023). Epistemic integration and social segregation of AI in neuroscience. arXiv *preprint arXiv:2310.01046*.

**O** The case study of Neuroscience **O** Is AI creating a separate **Neuroscience sub-discipline?** 



#### What is AI?

A structured set of technical, computational and mathematical tools, developed by computer scientists and mathematicians and applied in other disciplines.

First challenge: How to build a dataset to address our research questions?

Retrive a large set of keywords associated to AI exploring the Wikipedia pages associated to AI and several AI glossaries available on the web.

to extract all the papers

~2,7 millions of papers containing one or more AI concepts

#### ~600 AI concepts

μ.	DOCUMENT1	DOCUMENT2	DOCUMENT3	DOCUMENT4
SE	<u>Al Keywords:</u> [Kw1,kw2]	<u>Al Keywords</u> : [Kw3]	<u>Al Keywords:</u> [Kw1,kw3,kw4,kw5]	<u>Al Keywords:</u> [Kw2,kw3,kw4]
ATA	<u>Year</u> : Y1	<u>Year</u> : Y1	<u>Year</u> : Y2	<u>Year</u> : Y3
D	<u>Authors:</u> [Au1,Au2]	<u>Authors:</u> [Au1,Au3]	<u>Authors:</u> [Au4]	<u>Authors:</u> [Au1, Au2,Au5]
AG	<u>Journal</u> : J1	<u>Journal</u> : J2	<u>Journal</u> : J1	<u>Journal</u> : J3
MA	Discipline: d1	Discipline: d2	Discipline: d1	Discipline: d3

Query on OpenAlex (before MAG) containing the AI keywords.

3

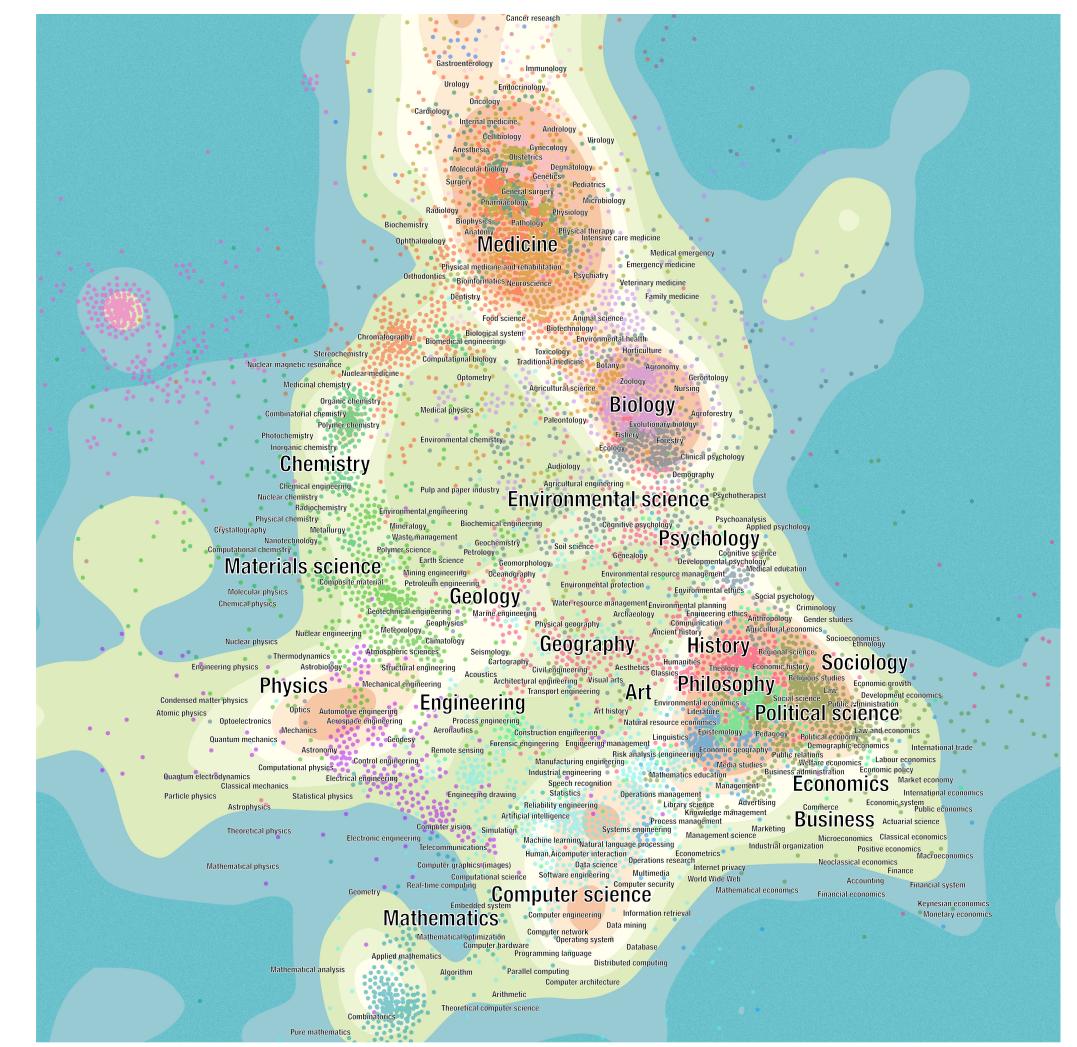
Assign to papers a discipline, using the journal classification of the Web of Science

~1,1 millions of papers containing one or more Al concepts and an associated discipline

#### What is AI?

A structured set of technical, computational and mathematical tools, developed by computer scientists and mathematicians and applied in other disciplines.

First challenge - part 2: **How can we define a distance in the disciplinary space?** 

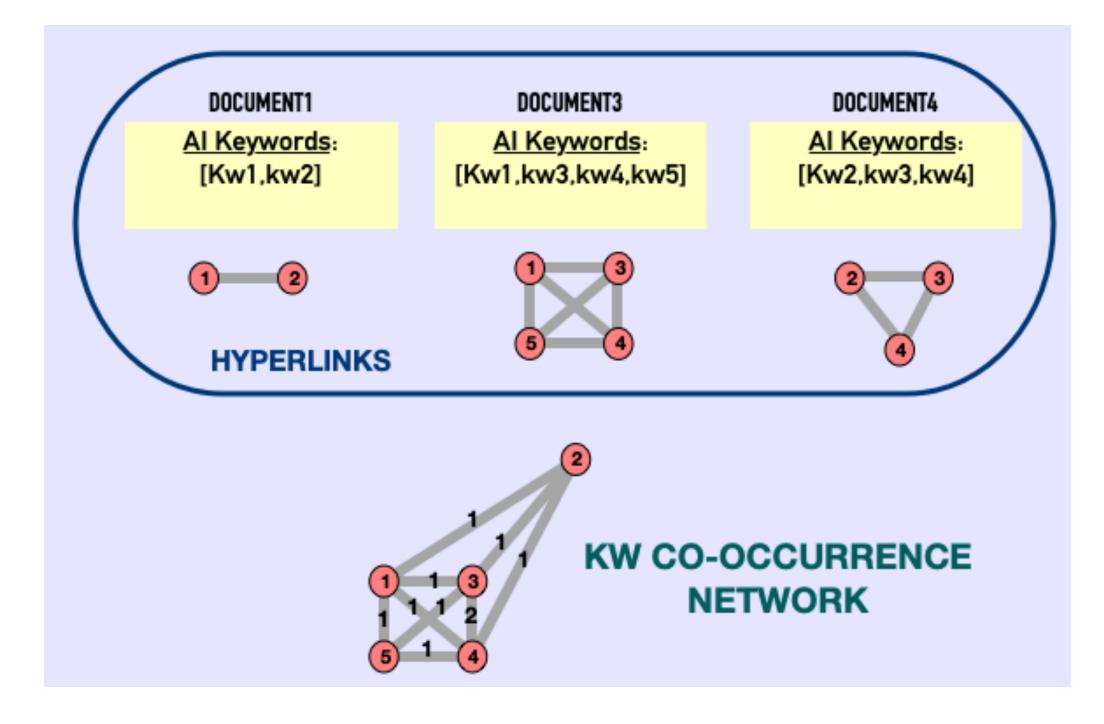


10000 random samples of 100.000 papers on MAG with their references.

- 2 Build the co-citation matrix aggregated at the level of disciplines
- 3 The distance between disciplines is defined through the pointwise mutual information (PMI):

$$pmi_{ij} = max \left( 2\log_2 \left( \frac{w_{ij}}{\sum_k (w_{ik}) \sum_k (w_{jk})} \right), 0 \right)$$
$$D_{ij} = 1 - pmi_{ij}$$

## The meso-scale structure of AI



Network partitioning in communities: detecting group of nodes that have more connections among them than with the rest of the graph.

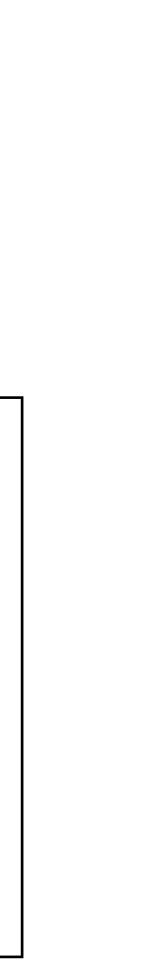
Performed using the Louvain algorithm (greedy optimization of modularity)

Technical issue:

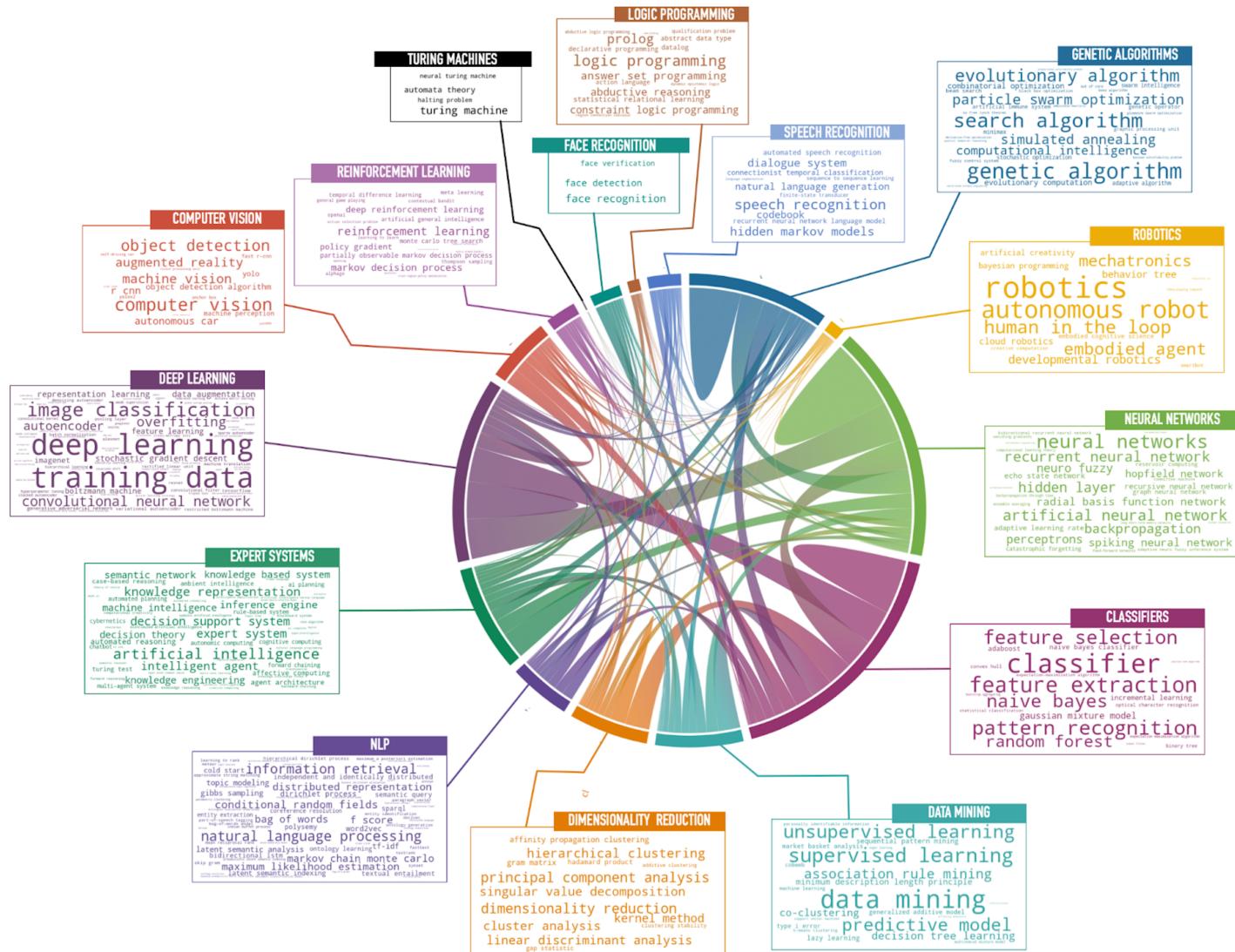
The graph is very dense (several edges with very heterogeneous weights)

#### We apply a disparity filter on the original graph

(Serrano, M. A., Boguna´, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. Proceedings of the national academy of sciences, 106(16):6483–6488.)



## The meso-scale structure of AI

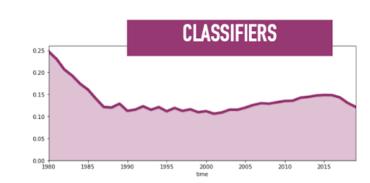


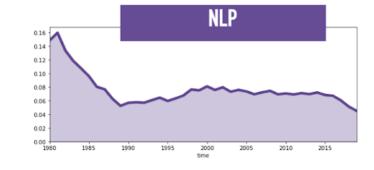
GENETIC ALGORITHMS
evolutionary algorithm
particle swarm optimization
search algorithe processing unit
computational intelligence
genetic algorithm
evolutionary computation adaptive algorithm
ROBOTICS
artificial creativity bayesian programming behavior tree
robotics
— autonomous_robot
human in the loop
cloud robotics creative computation developmental robotics
developmental roboties smartbet
NEURAL NETWORKS
bidirectional recurrent neural networks
recurrent neural network
echo state network hopfield network
exemption through the source read and the source of the so
artificial neural network

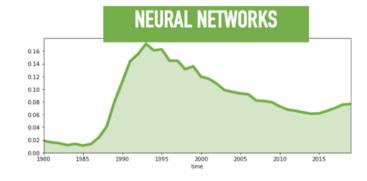
feature selection	
convex hull classifier -	0 11-00 of ger1.000
feature extraction memory agreement haive bayes incremental learn optical character recon- istatistical classification gaussian mixture model	)n ing gnition
pattern recognitic	<u>on</u>

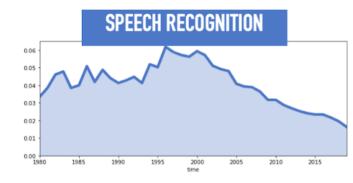
#### We identify 15 AI sub-categories

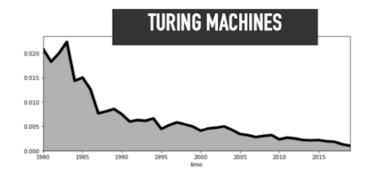
## The meso-scale structure of AI

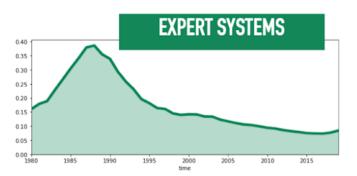


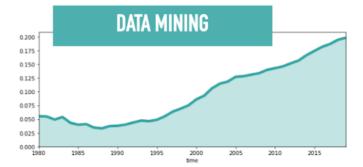


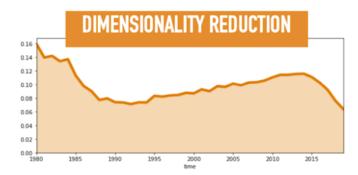


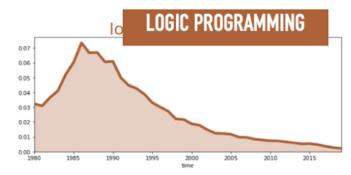


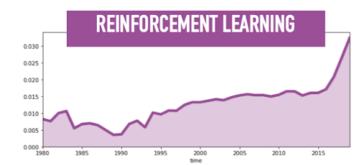


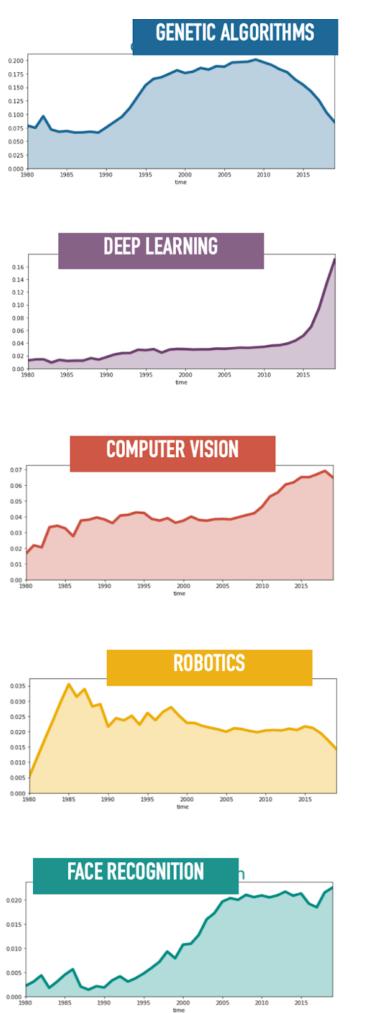














#### We identify 15 AI sub-categories with different temporal patterns

## Al from development to application...

SET	DOCUMENT1 <u>AI Keywords</u> : [Kw1,kw2]	DOCUMENT2 <u>AI Keywords:</u> [Kw3]	DOCUMENT3 <u>AI Keywords:</u> [Kw1,kw3,kw4,kw5]	[
DATAS	<u>Year</u> : Y1	<u>Year</u> : Y1	<u>Year</u> : Y2	
MAG DA	<u>Authors:</u> [Au1,Au2]	<u>Authors:</u> [Au1,Au3]	<u>Authors:</u> [Au4]	
	<u>Journal</u> : J1	<u>Journal</u> : J2	<u>Journal</u> : J1	
	Discipline: d1	Discipline: d2	Discipline: d1	

For each year we take the list of disciplines in which AI keywords are used

> Y1: {d1: 145, d2:364,...} Y2: {d1: 137, d2:789,...} Y3: {d1: 56, d2:1034,...}

 $\bullet \bullet \bullet \bullet$ 

DOCUMENT/	
DOCUMENT4 <u>AI Keywords:</u> [Kw2,kw3,kw4]	We define
<u>Year</u> : Y3	<b>Mathematics and</b>
<u>Authors:</u> [Au1, Au2,Au5] <u>Journal</u> : J3	Computer science the "originating
Discipline: d3	disciplines" of Al

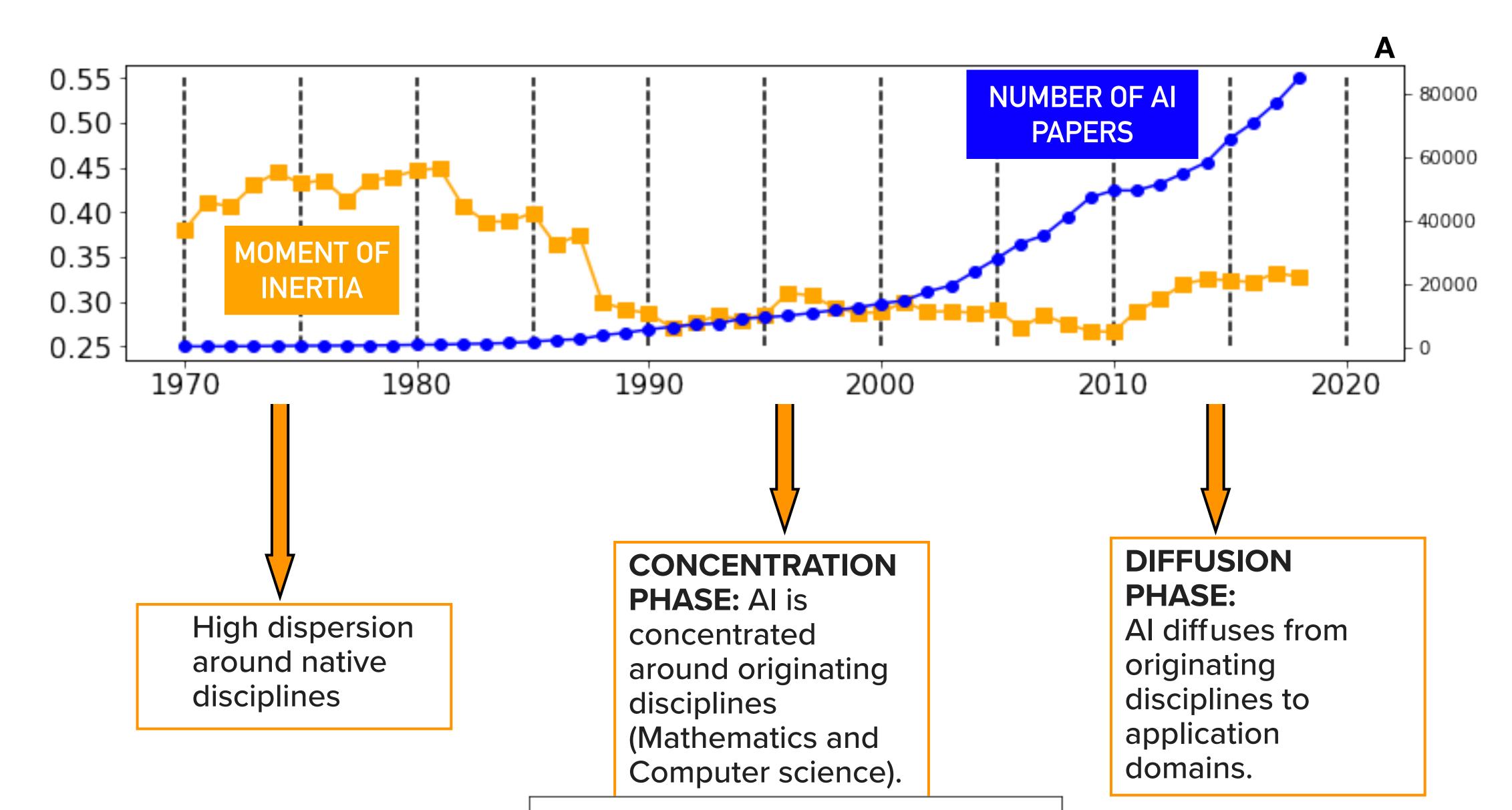
Inertia moment around originating disciplines

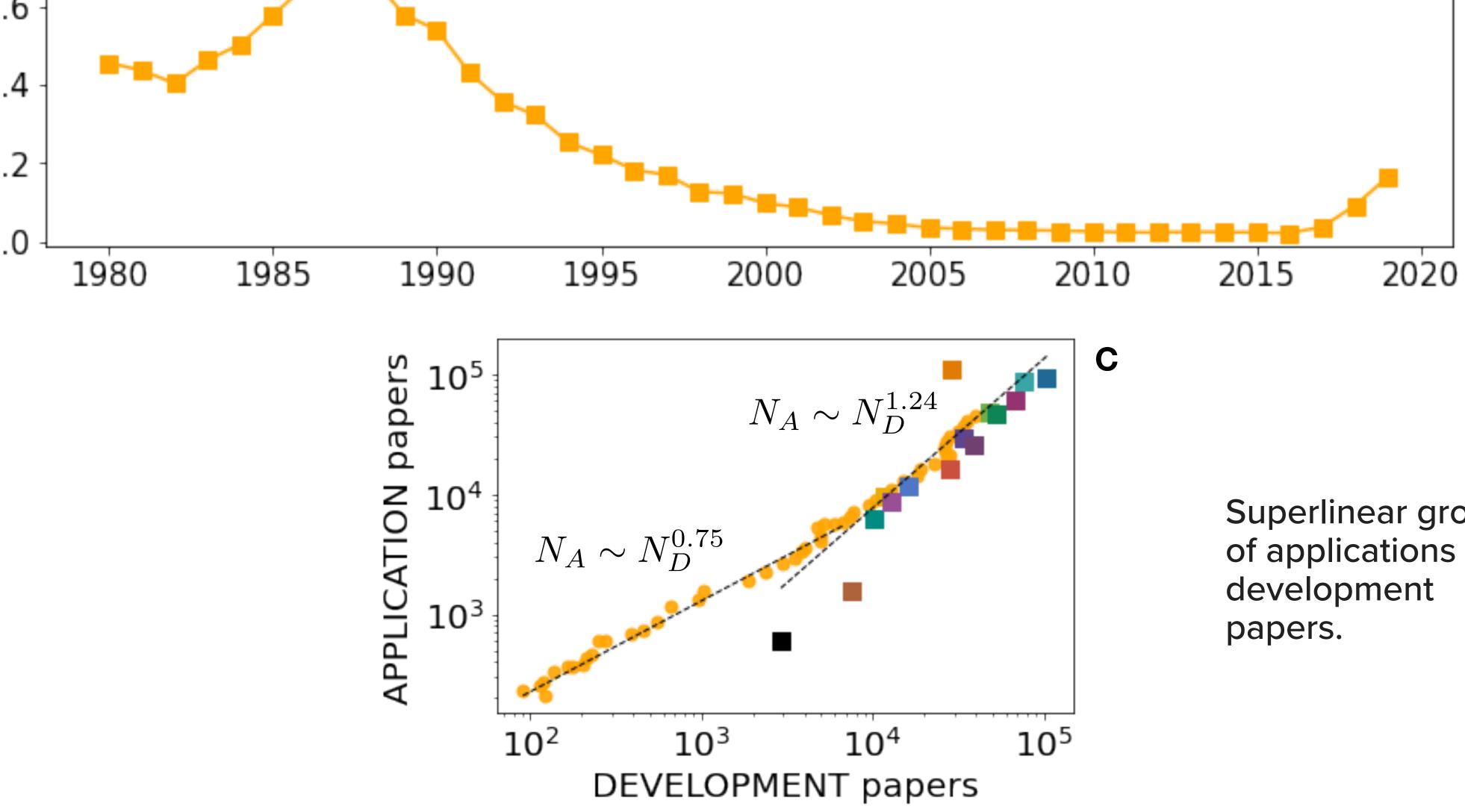
$$m_{I} = \sum_{i} \frac{n_{i}}{N_{tot}} \min(D_{i,CS}, D_{i,Math}, D_{i,Stat})^{2}$$

LOW MI = CONCENTRATION AROUND ORIGINATING HIGH MI = DISCIPLINATY DISPERSION



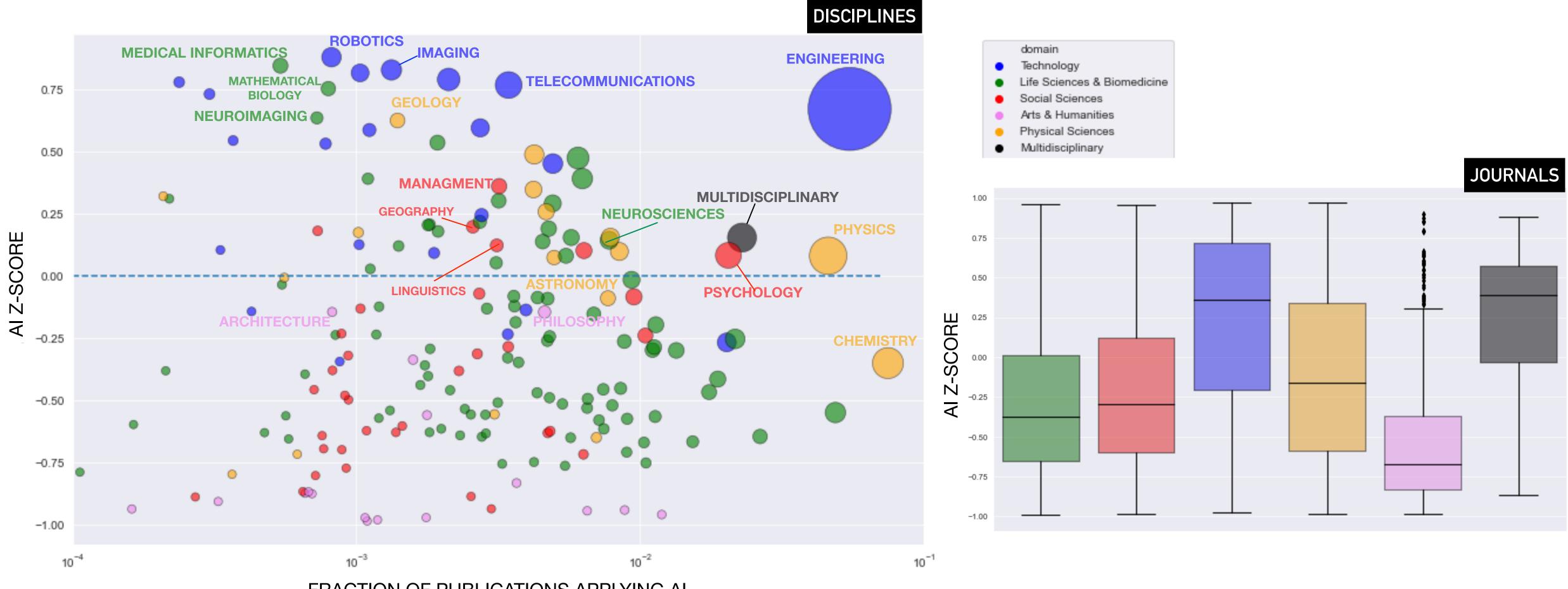
## Al from development to application...





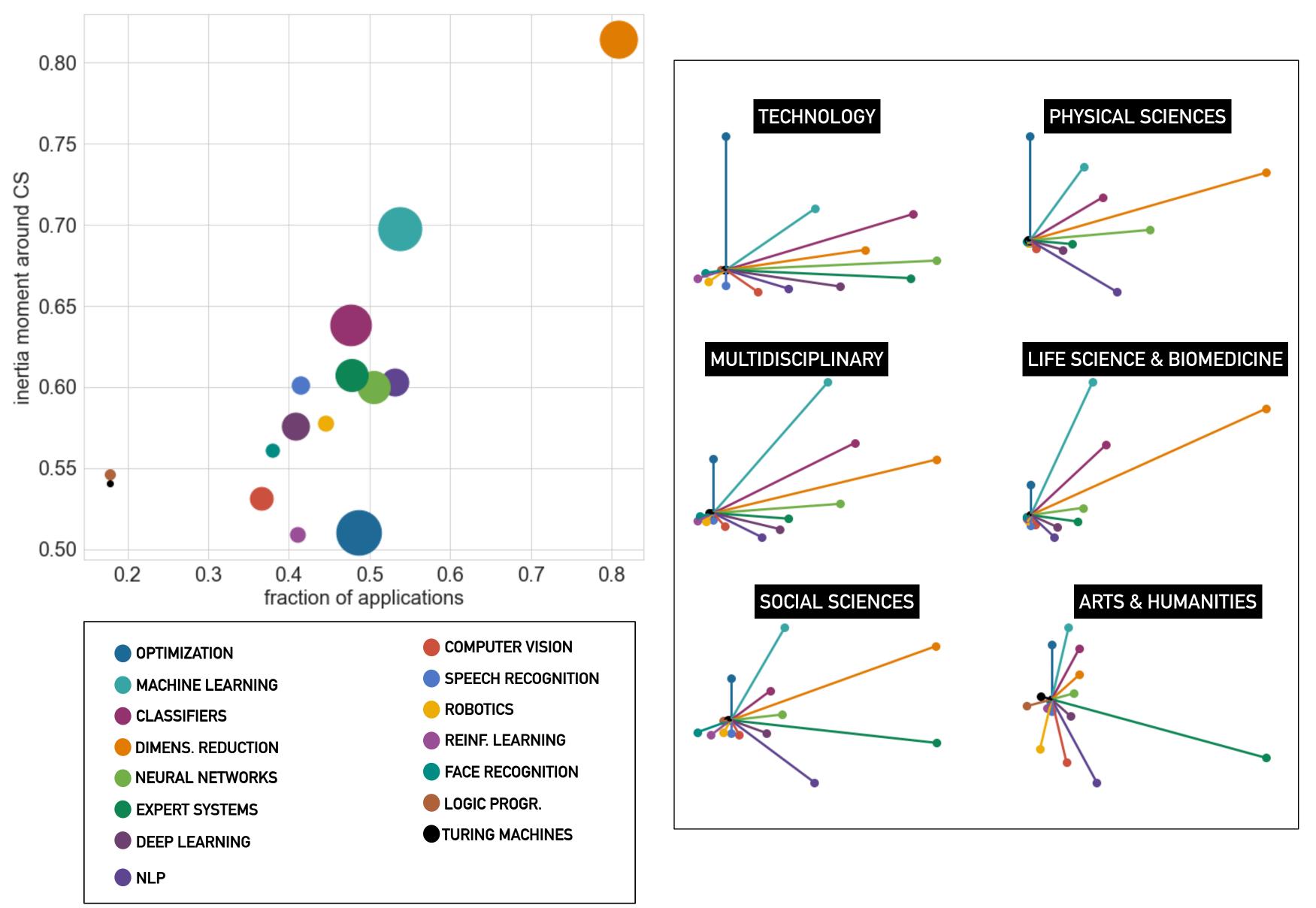
Superlinear growth of applications with

# In which disciplines AI has diffused?



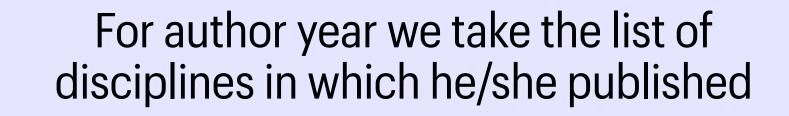
FRACTION OF PUBLICATIONS APPLYING AI

# In which disciplines AI has diffused?



# Interdisciplinary collaborations around AI

É.	DOCUMENT1	DOCUMENT2	DOCUMENT3	
DATASE	<u>Al Keywords:</u> [Kw1,kw2]	<u>Al Keywords:</u> [Kw3]	<u>Al Keywords:</u> [Kw1,kw3,kw4,kw5]	<u>A</u> [K
	<u>Year</u> : Y1	<u>Year</u> : Y1	<u>Year</u> : Y2	
MAG D/	<u>Authors:</u> [Au1,Au2]	<u>Authors:</u> [Au1,Au3]	<u>Authors:</u> [Au4]	[A
	<u>Journal</u> : J1	<u>Journal</u> : J2	<u>Journal</u> : J1	,
ž	<u>Discipline</u> : d1	<u>Discipline</u> : d2	Discipline: d1	D

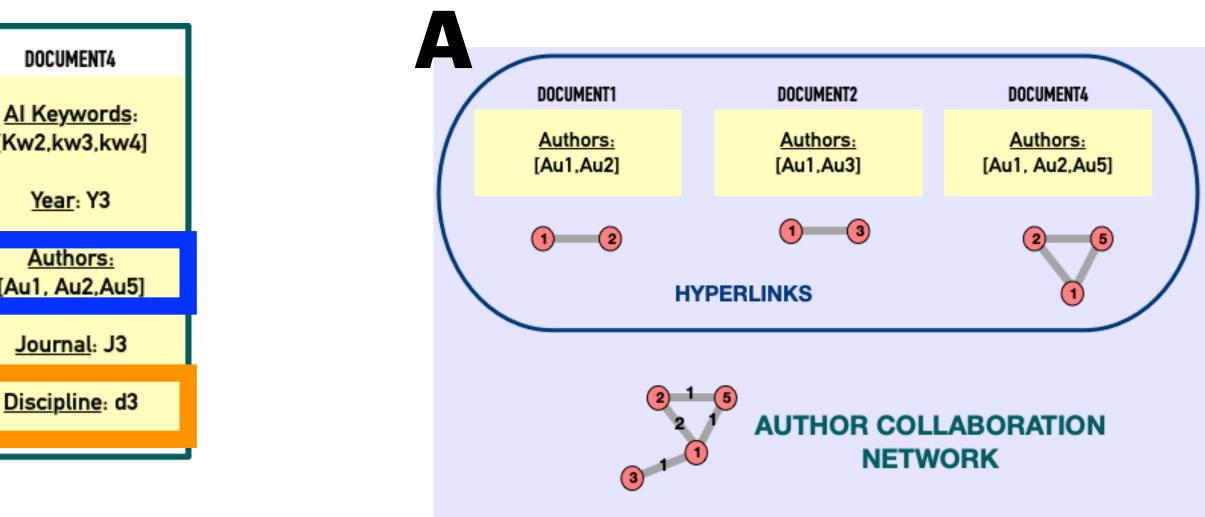


B

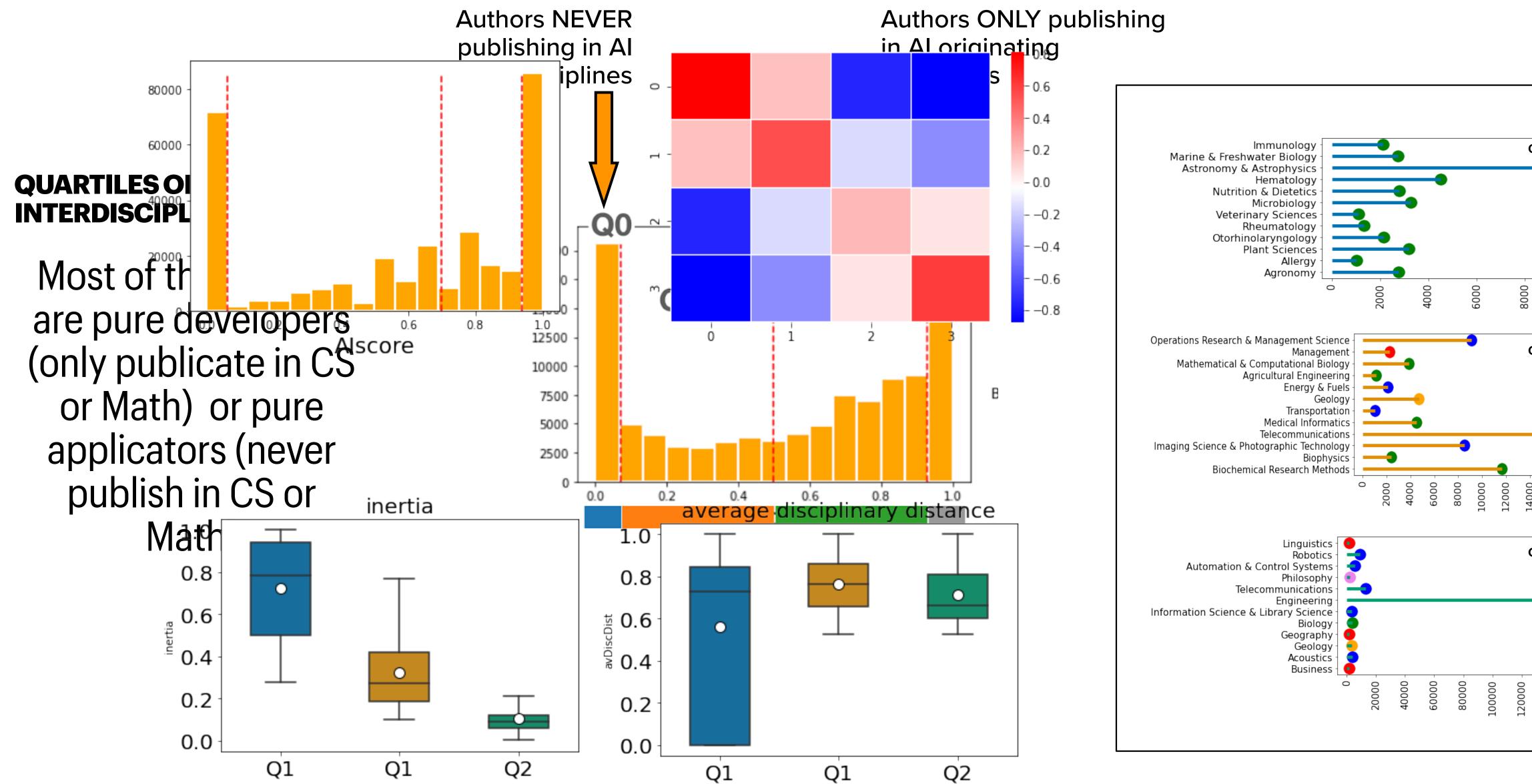
Au1: {d1:4, d2:6,...} Au2: {d1: 1, d2:6,...} Au3: {d1: 10, d2:0,...}

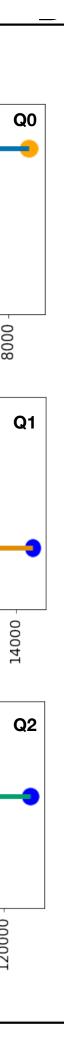
 $\bullet \bullet \bullet \bullet$ 

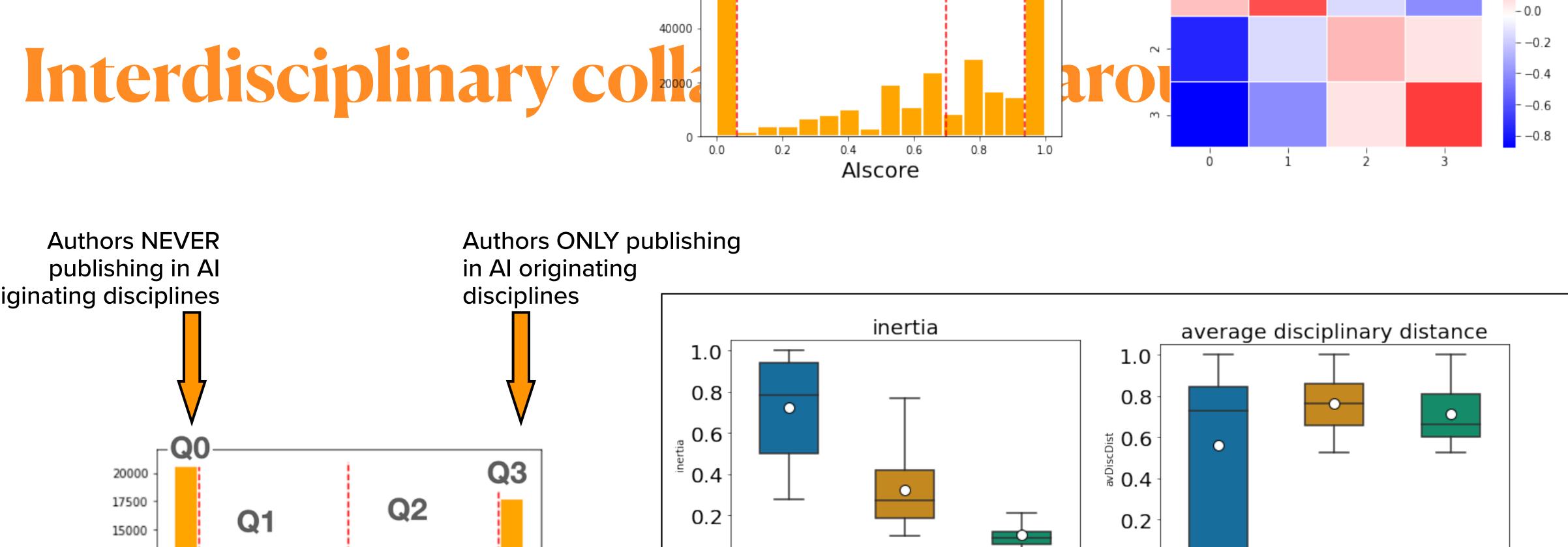
<u>Author interdisciplinary score</u>: Fraction of papers of the author in Al originating disciplines (Mathematics and CS)

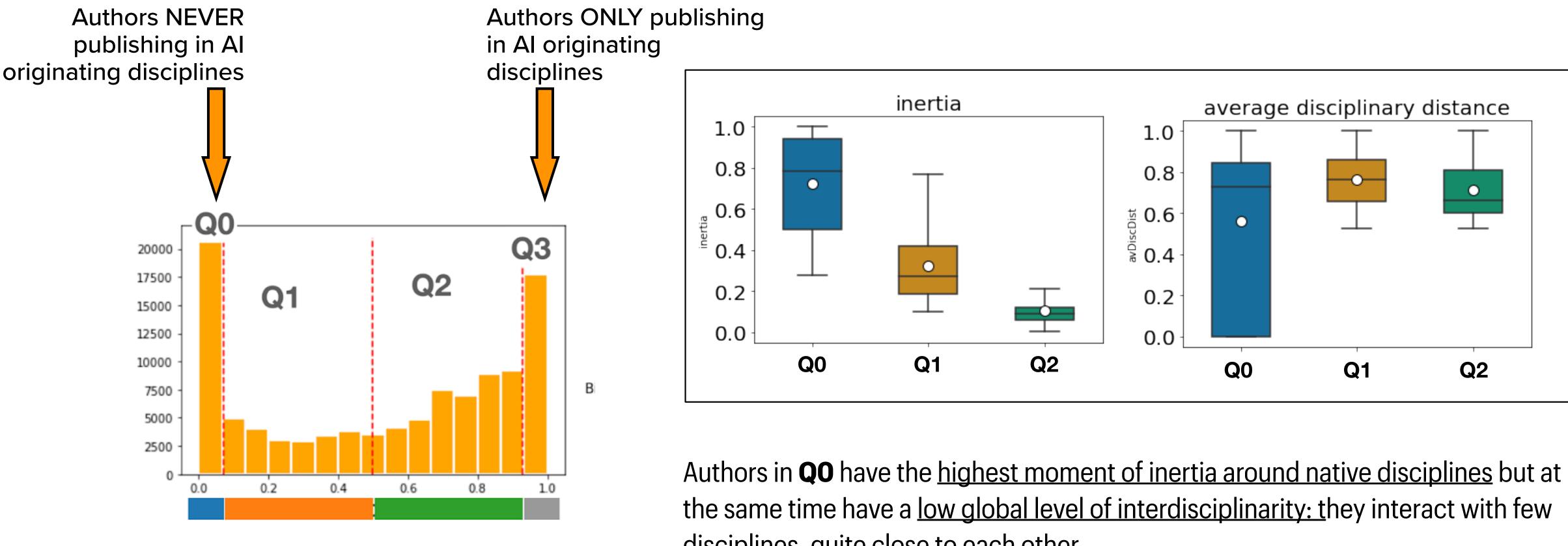


# Interdisciplinary collaborations around AI









disciplines, quite close to each other.

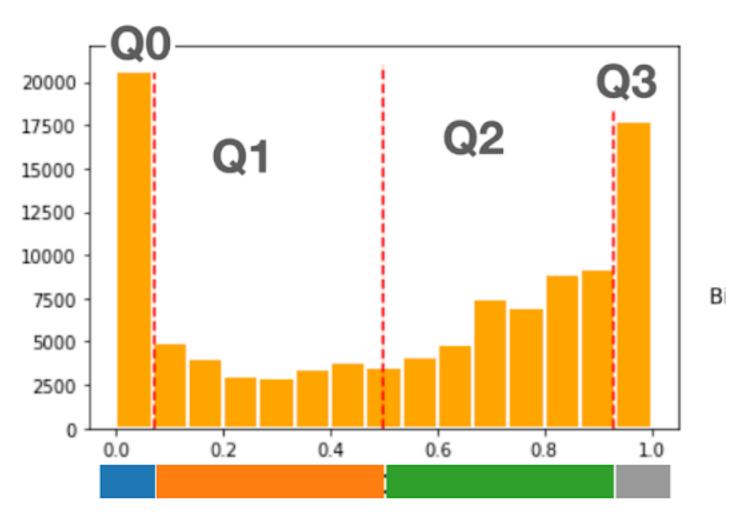
**Q1** includes most authors involved in interdisciplinary collaborations.

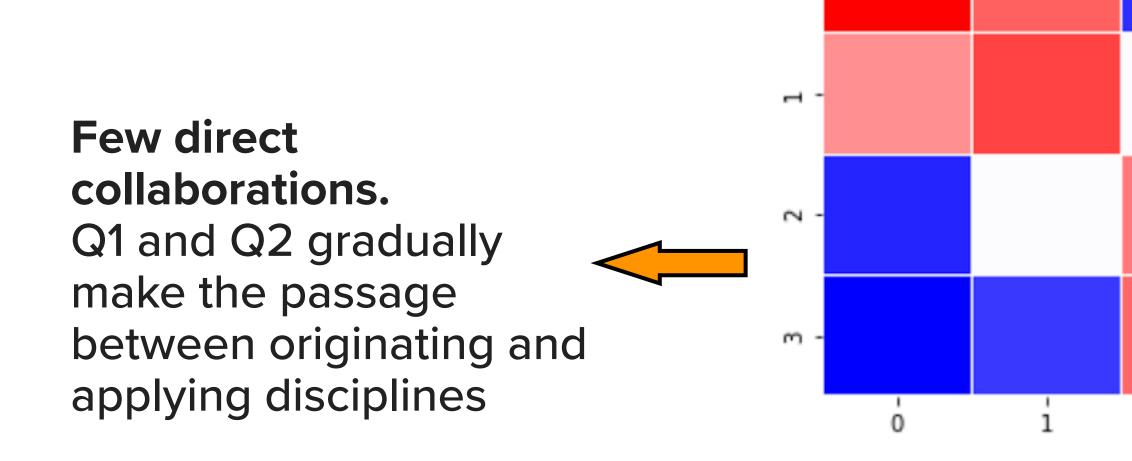
**Q2** authors publish in disciplines close to the native ones (low moment of inertia) but at a quite large distance from them.

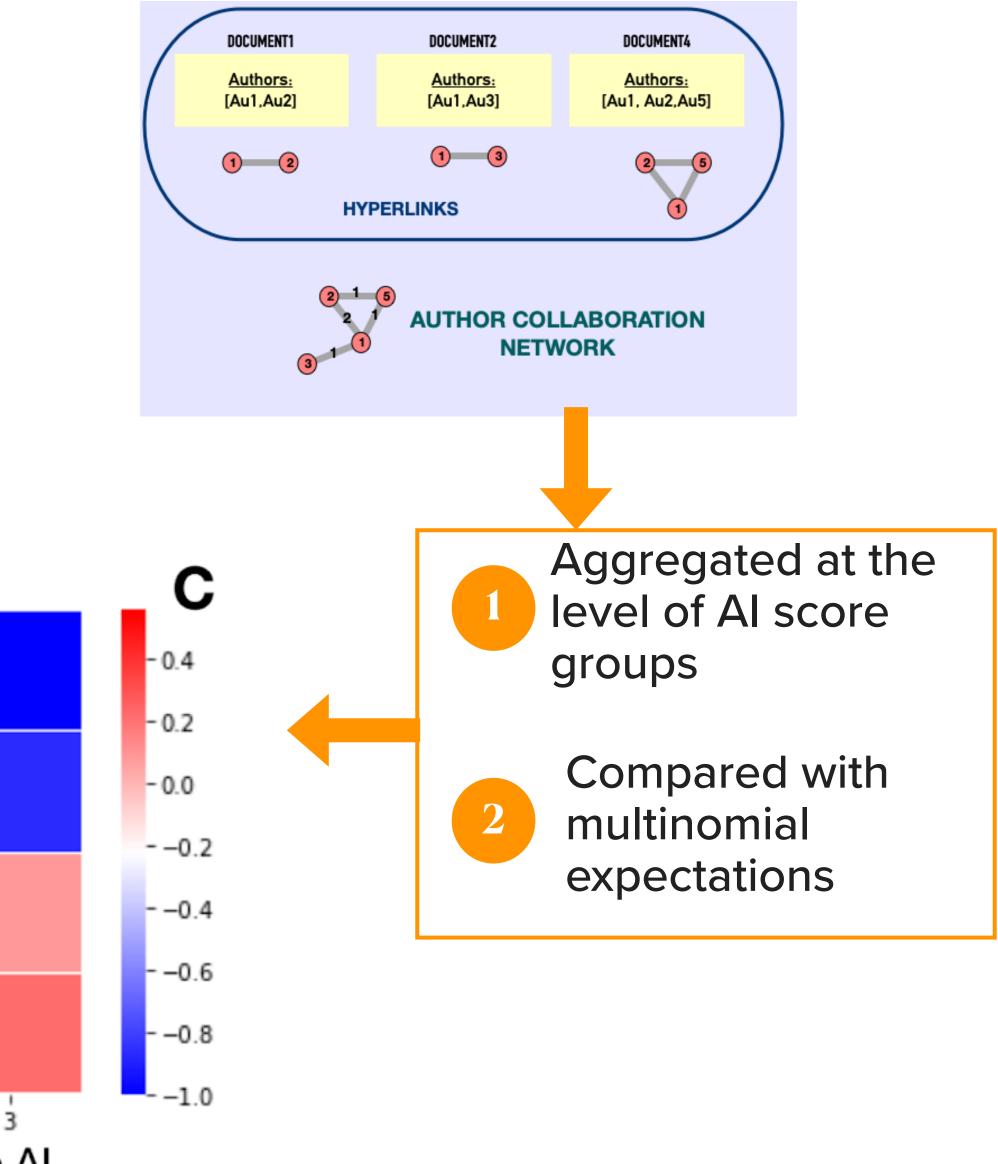
#### Interdisciplinary collaborations around AI

Apply AI

0







**Develop Al** 

2

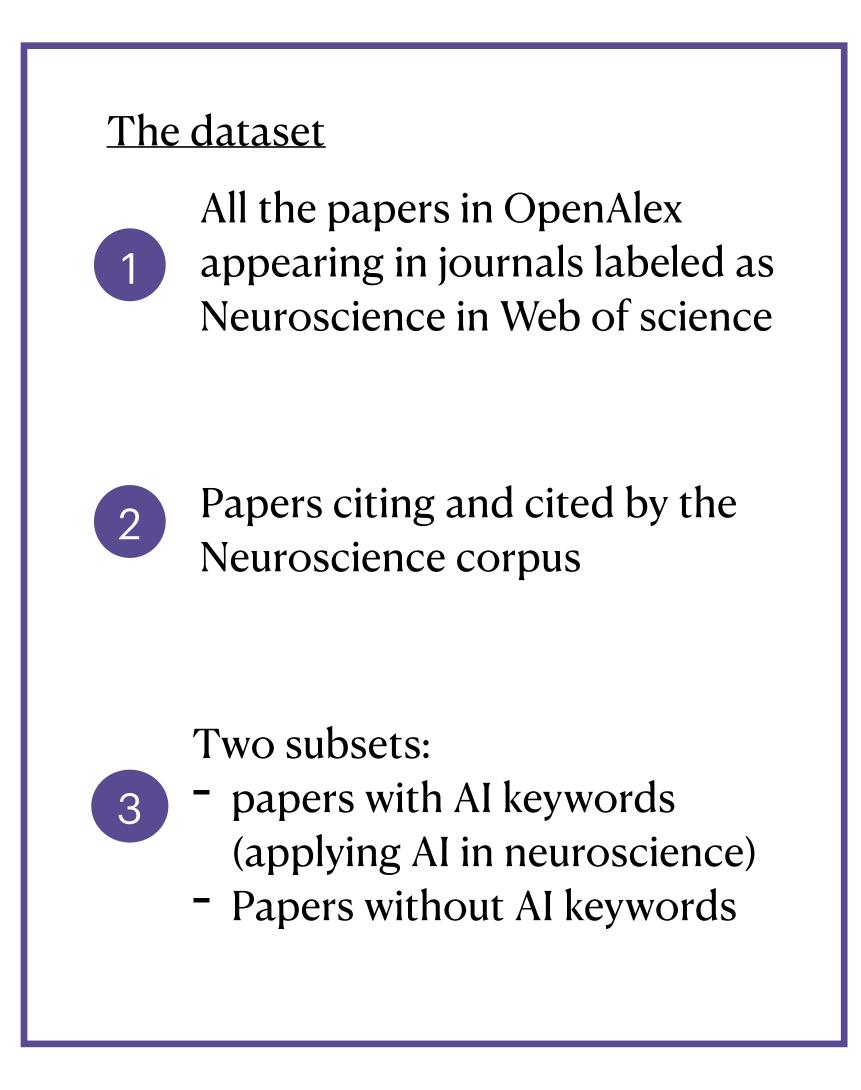
# Mainfindings

- applications.
- are extremely rare.

#### O AI is a set of 15 interconnected sub-categories of concepts each with its different historical pattern. **O 3 phases of AI diffusion in disciplines: large** participation to foundation, shrink on CS, spread in

**O** Direct collaborations between developers and users

#### Al in Neuroscience



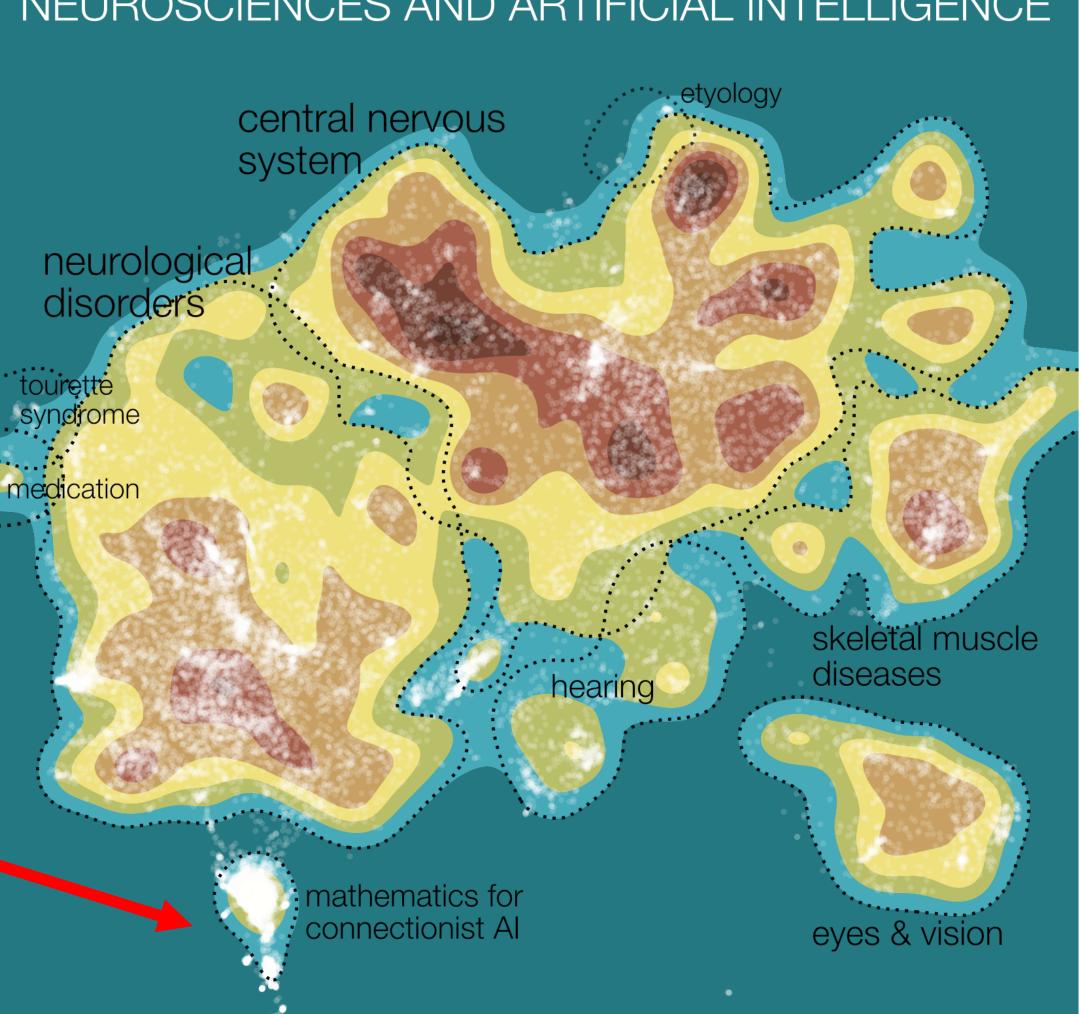
#### How to build the disciplinary landscape of AI?

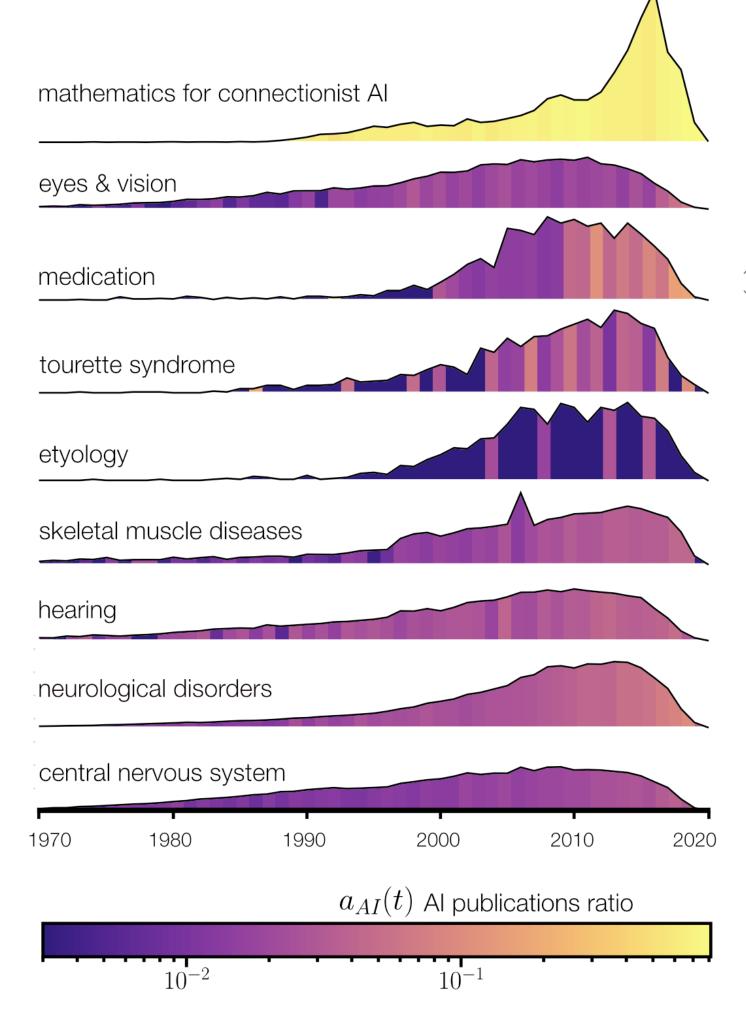
- Contextual words embedding of each paper of our dataset using Specter (an embedding model based on SciBert) (https://github.com/allenai/specter)
- Reduce the data if needed for a first 2D spatialized visualization with UMAP
- Use HDBSCAN for local density-based clustering, which determines automatically the number of clusters (points that does not belong to particular clusters are drawn in the same color)
- Characterize the founded clusters with concepts of papers from MAG concepts

#### Al in Neuroscience

Concentration on the cluster "Math for AI" (Computer Science, Math and Stats) and on the vocabulary space next to it in "Neurological disorders" → linked to the data deluge!

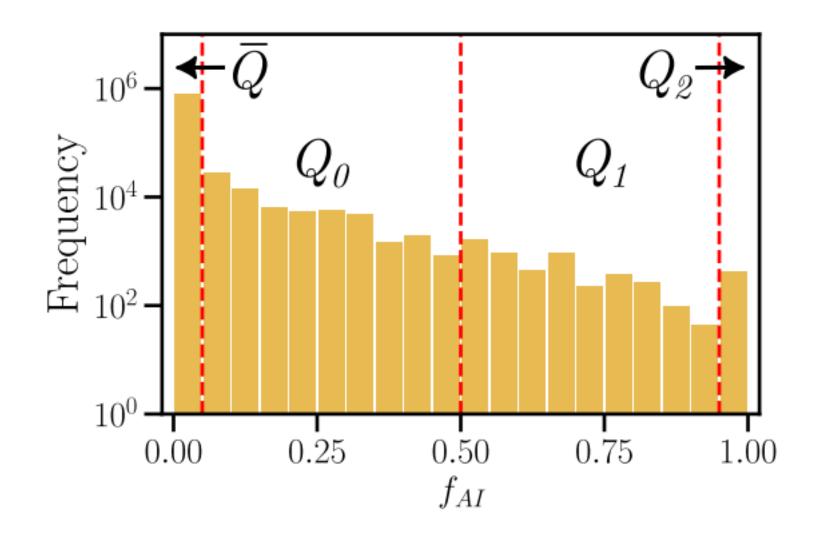
#### NEUROSCIENCES AND ARTIFICIAL INTELLIGENCE





a(t)otal nur

#### Al in Neuroscience



#### **Main findings**

**O** Epistemic (partial) integration of AI in Neuroscience **O** Social segregation of authors practicing AI in Neuroscience

