National Ph.D. in AI for Society

On the “performance” of scientific publications: indexes, impact, citations

Stefano Chessa

Department of Computer Science, University of Pisa

March 25th, 2025

Agenda

Performance indicators & evaluation of research

Publishing in high-impact venues

Citations

Writing papers in the LLM era

Bad practices

Conclusions

Preliminary note on this talk

This seminar is not intended to replace your supervisor!

- If you are a PhD student, you are learning how to conduct a research
- In this case your supervisor is you first reference:
  - He/she is experienced, and knows the rules of the game
  - Learn from him/her as much as you can

Why this talk

In the last years “aggressive” use of bibliometries to evaluate the research

… and consequent use of “aggressive” strategies by the researchers to improve their bibliometric indexes…

Number of papers…

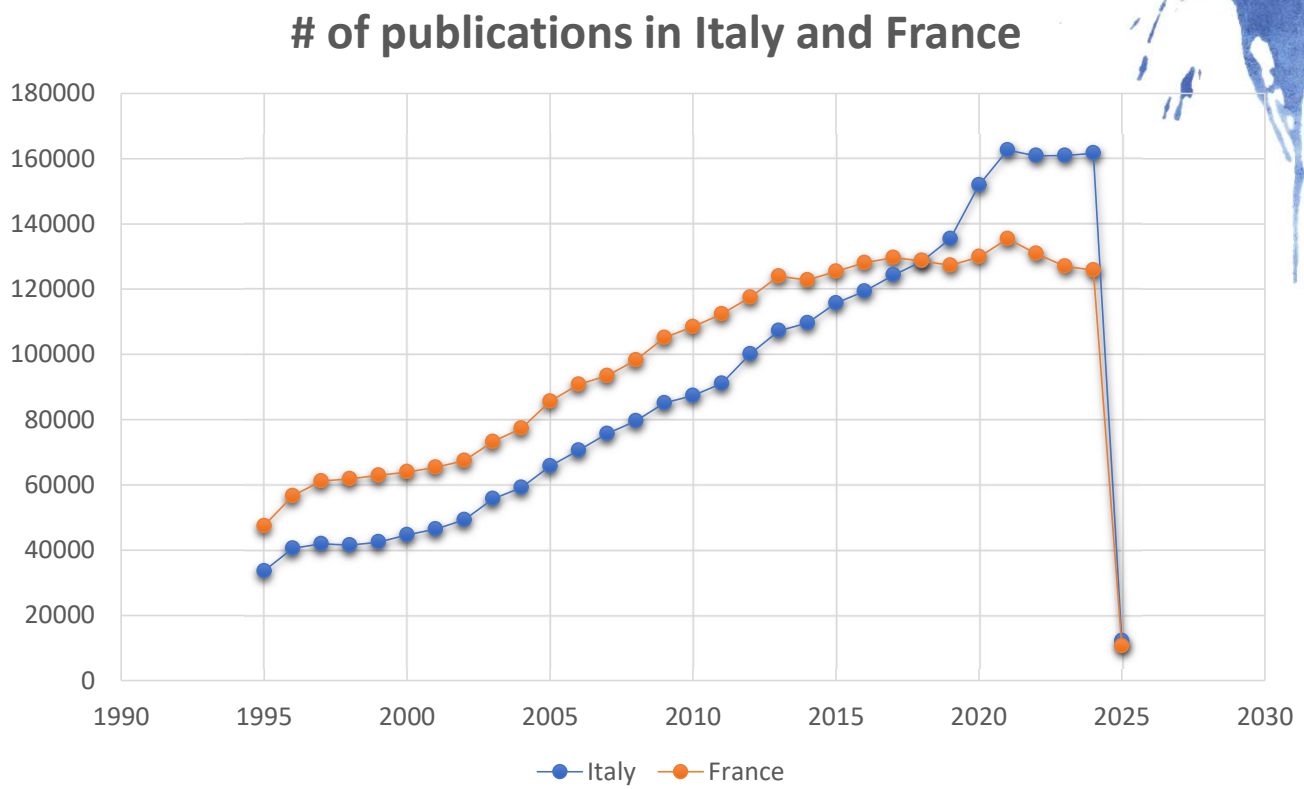
From Scopus, queries:

AFFILCOUNTRY(Italy)

AFFILCOUNTRY(France)

Moore's Law for papers: the number of papers that are "inexpensively" produced doubles every 10 years…

But in the last years it's getting flat…



Performance indicators & evaluation of research

Properties of a paper

Performance v.s. Maturity

venue	year	citations
ComCom	2007	1199
INFOCOM	2005	223
ComCom	2001	143
SRDS	2001	139

venue	year	citations
J. of Algo.	2002	32
IEEE TIT	2012	19
IoT J.	2018	56
IEEE TC	2001	13

Performance indicators

Sometimes they are called “quality” indicators

(in Italy for example)

… but they are not. They ofter measure the performance of a researcher, or of a paper or a journal in terms of “diffusion” in the research community

Many different indexes:

- Impact factors
- H-index
- Number of citations
- Number of indexed papers
- …

Impact factors: performance of journals

- IF (web of science)
- Scimago SJR (scopus)
- SNIP (scopus)
- CITESCORE (scopus)
- MCQ (MathSciNet, for mathematics)
- …

Impact factors

IF (web of science):

- "the **Impact Factor** of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years"
- Example: X papers published in 2019 and 2020; Y citations received by these papers in 2021; $IF_{2021}=Y/X$

Citescore (Scopus)

- **Similar to IF** but it is computed over the scopus database and over a 4 years time frame.
- Example: X papers published in 2021-2024; Y citations received in 2021-2024; $Citescore_{2024}=Y/X$

Impact factors

SJR (scopus):	"**SCImago Journal Rank** measures weighted citations received by the serial. Citation weighting depends on subject field and prestige (SJR) of the citing serial." Inspired to Google PageRank
SNIP (scopus):	"**Source Normalized Impact per Paper** measures actual citations received relative to citations expected for the serial's subject field."

An example: citescore rank

CiteScore rank ⓘ 2023 In category: Hardware and Architecture

☆	#7 / 177	IEEE Internet of Things Journal	17.6	96th percentile

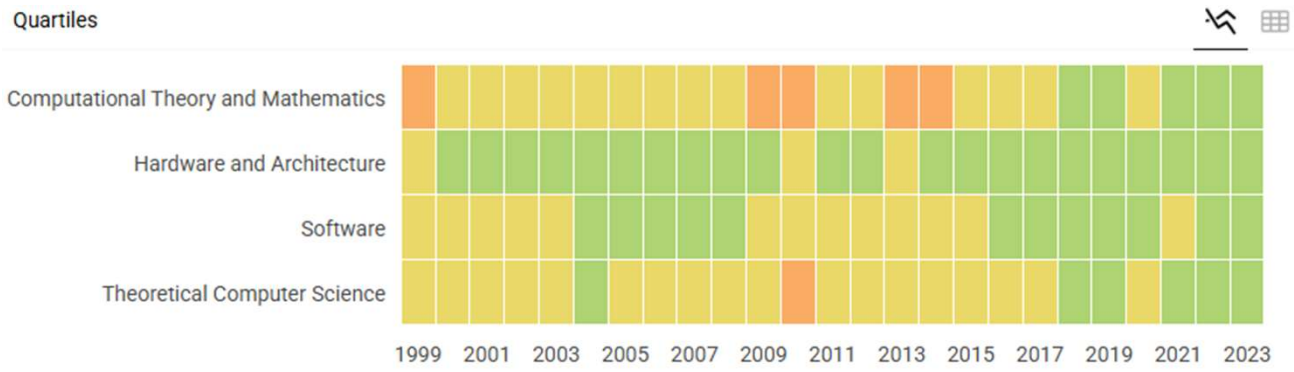
	Rank	Source title	CiteScore 2023	Percentile
	#1	Information Fusion	33.2	99th percentile
	#2	Journal of Manufacturing Systems	23.3	99th percentile
	#3	Journal of Network and Computer Applications	21.5	98th percentile
	#4	IEEE Network	20.4	98th percentile
	#5	Future Generation Computer Systems	19.9	97th percentile
	#6	Journal of Big Data	17.8	96th percentile
☆	#7	IEEE Internet of Things Journal	17.6	96th percentile
	#8	International Conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS	15.6	95th percentile
	#9	IEEE Transactions on Cognitive Communications and Networking	15.5	95th percentile
	#10	Digital Communications and Networks	12.8	94th percentile
	#11	VLDB Journal	12.3	94th percentile
	#12	Computer Physics Communications	12.1	93rd percentile

Quartiles

- the impact indexes naturally lead to a rank of journals
- but the rank itself may be considered too granular
- in some contexts it is considered the quartile of a journal:
  - Q1 refers to its position in the 25% top rank journals (Q2, Q3, Q4 follow by induction)
- NOTE: the quartile always relates to a specific impact index and, most often, to a specific subject area/category

---

Quartiles of IEEE Trans. on Comput. taken from Scimago SJR (https://www.scimagojr.com/)



… and what about conferences?

- The above indexes are also computed for the indexed conferences, and in principle also quartiles
  - However, many conferences are rather volatile,
  - Impact indexes may make sense only for the established conferences.

Journals and conferences classifications

- Another approach, that does not make use of (citation-based) computed indexes, is that of classification
- Classification of conferences or journals is a hand-made work, based on experience and specific to a subject area
  - Building a classification may take years
  - It is very hard to update
  - While the world changes fast…
- It is also the work of local communities:
  - There exist several classifications for different countries
  - To take into account their specificity
- Examples are:
  CORE (http://portal.core.edu.au/conf-ranks/) or
  GRIN/GIE/SCIE (https://scie.lcc.uma.es:8443/)

Journals and conferences classifications

- in Italy and Spain, classification of conferences in Computer Science/Engineering
  - Currently there's an effort in Italy to produce an hands-made classification of journals…
- joint effort og GRIN and GII and SCIE: https://scie.lcc.uma.es:8443/
  - you can query for single conference, class or download the excel file with the full classification
- it is still in progress (last update Oct. 2021!)…

Status of the classification of conferences:

Class	Ratings	Size	Description
Class 1	A++, A+	35 + 40 = **75 conferences**	excellent, top notch conferences
Class 2	A, A-	79 + 78 = **157 conferences**	very good events
Class 3	B, B-	187 + 132 = **319 conferences**	events of good quality
-	Work in Progress	1811 conferences	work in progress

Some remarks

- many other research areas in Italy (and in the world) evaluate papers by means of indexes and quartiles
  - and in some/many case they consider only journals
- In the Italian national evaluations (VQR and ASN):
  - more emphasis to journals
  - In the past no use of classifications but just indexes and other specific criteria
  - Now VQR is using informed peer review and ASN is using indexes (and more)

Publishing in high-impact venues

High impact journals

- In some context by high impact Journal they mean journals like Nature, Plos one and others:
  - they aim at a larger audience, although they are still “technical”
  - … and they take very seriously the way in which the paper is presented
- Here instead I refer to “deeply technical” journals that also have a high impact.

Publishing in high impact venues (I)

High Impact →
- high chance of being read & cited
- more selective, harder to publish(?)

- In many areas the impact of the journals is taken rather seriously
- … and more recently also for computer science & computer engineering it is becoming important

Publishing in high impact venues (II)

- It’s your preliminary choice
- … but look first at the meaningfulness of the journal for your paper
  - This includes having a look to what they published in the recent past, not just their own description
- and review process may be engaging…

Publishing in high impact venues (III)

- But…
- Journal impact is not a diamond
- its value is not forever…
- … and it may decrease over time

Scimago SJR of Theoretical Computer Science over the years

Publishing in high impact venues (IV)

- However,

  high impact  large number of citations
- … why so?

- The citations received by a paper are an individual value
- The impact of a journal is a collective value
- All high-impact journals have highly-cited and normally/lowly-cited papers

Citations and H-Index

- Usually, the number of citations received and the H-index are considered in combination with the journal's impact
- They indicate the "individual" performance of a researcher or of a paper
- Nowadays also other derived metrics are available
  - E.g. Field Weight Citation index (FWCI)

Citations and H-Index

- H-index of a researcher is X if he has exactly X papers each of which received at least X citations
- H-index grows slowly, and it is not linear!
  - 1 < 5 but 11 << 15 <<< 19 …
- There are criticisms to H-Index, but it is still widely used

Citations and H-Index

Citations are usually a factor of **stress** and **depression**:

- They do not (necessarily) depend on the quality of your work
- They do not (necessarily) depend on your preliminary choice (as impact factors)
- They depend on the future behavior of other researchers, out of your control

How to get cited?

there’s no guarantee,

depends on many factors

… and may take time…

Why do you cite a pa

- To refer a wo
- To

ettings in your

lo defend your approach/methodology
- To defend a statement in your paper
- …

You cite a paper because it is useful!

About usefulness

- Writing papers useful for a research community is not easy
  - Many time you will know only later that they were useful
  - I don’t know of anybody who wrote only useful papers
  - In fact, most papers have a limited “usefulness” …
- Sometimes we write papers just to:
  - test our ideas,
  - receive opinions from reviewers,
  - document our work
  - … and sometimes even to witness or to strengthen a cooperation

Main factors for citations

USEFULNESS

VENUE

REPUTATION OF THE AUTHORS

SIZE OF RESEARCH COMMUNITY

TIMELINESS OF THE WORK

2. Venue of the publication

- Not only a matter of impact
- The content of the paper should match well the audience of the journal/conference
  - Write the paper for that journal
  - Use terminology, methodology, approach typical of that community
  - i.e. if they expect formal proofs give them formal proofs
  - If they expect simulations give them simulations
  - … etc…

2. Venue of the publication: example

Two papers with a very similar idea about routing protocols in ad hoc networks, (almost) same time

- GPSR: Greedy Perimeter Stateless Routing for wireless networks
  MOBICOM 2000 – 6206 citations
- Routing with guaranteed delivery in ad hoc wireless networks
  Dial-M '99 – 629 citations
  - Later it appeared also in Wireless Networks'01 – 897 citations

2. Venue & 3. Reputation: example

Two papers with a very similar idea about routing protocols in ad hoc networks, same year

- Virtual ring routing: Network routing inspired by DHTs
  ACM SIGCOMM ‘06 – 183 citations
- Reliable routing in wireless ad hoc networks: The virtual routing protocol
  J. of Network and Systems Management ‘06 – 14 citations

3. Reputation of the authors

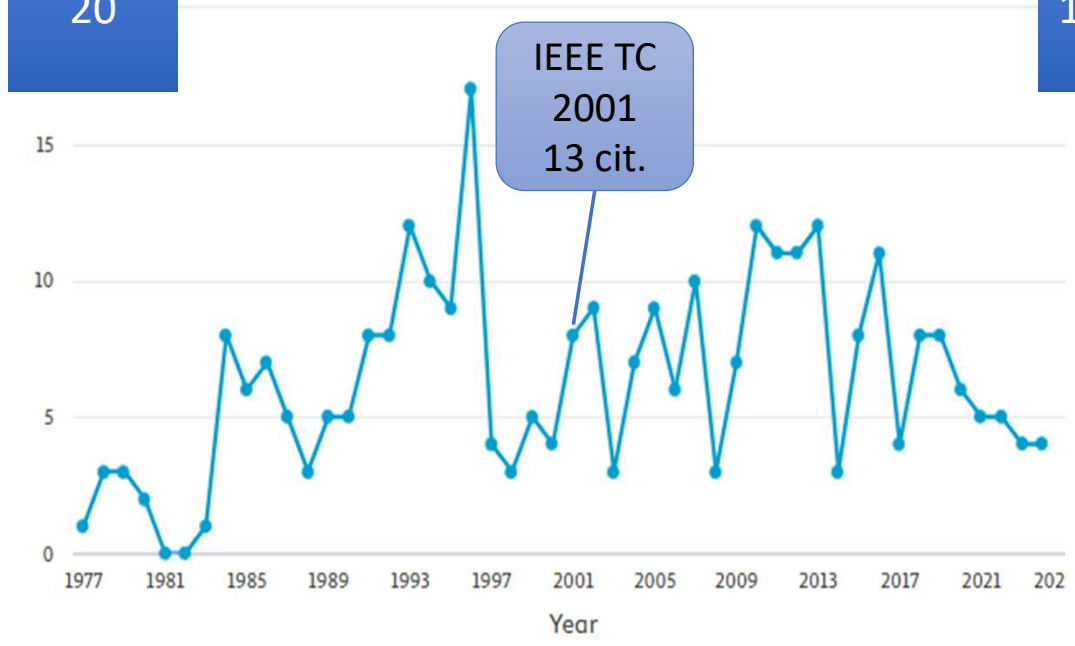
How do you gain reputation?

1. Writing high-quality papers
2. Being involved in a research community
   - serve the community
   - take part to the public events
   - …
3. Being proactive in innovation:
   - proposing new themes of research
   - Creating/organizing workshops/special issues
   - …
4. Establishing a network with other researchers

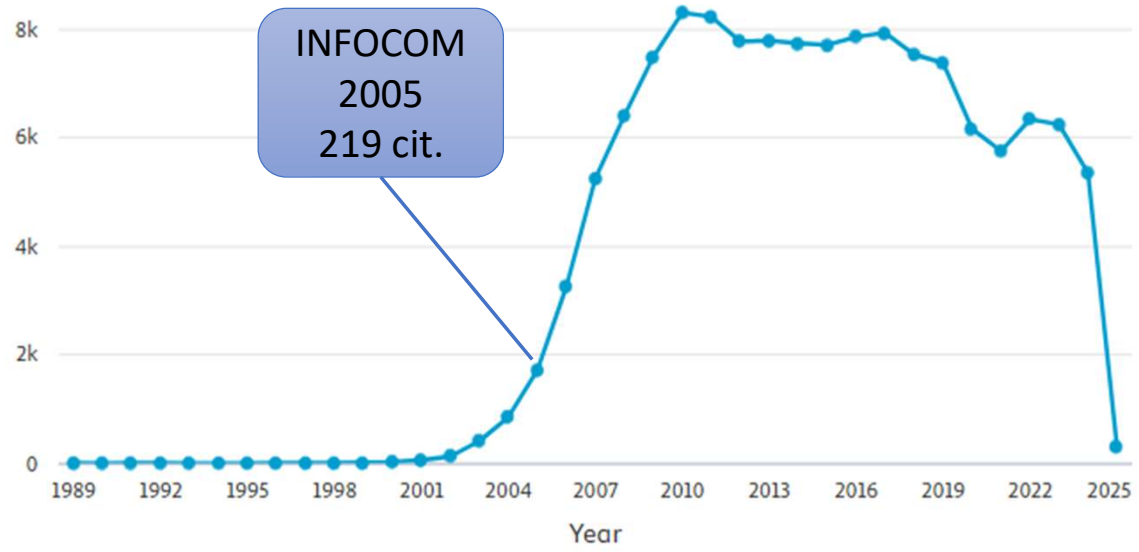
4. Size of a research community

"system-level diagnosis" vs "Wireless sensor networks"

20 works in "System-Level Diagnosis" per year



10.000 works in "Wireless Sensor Networks" per year



4. Size of a research community

really top papers had (often) been written for communities that did not exist yet…

- don’t be obsessed by the size
  - publishing early in a small community that grows fast can be a big boost…
- … but don’t remain entrapped in a “black hole”
  - If a research field is becoming “desertified” consider moving elsewhere

5. Timeliness (in geographic routing)

Greedy perimeter stateless routing (GPSR), *MOBICOM 2000*

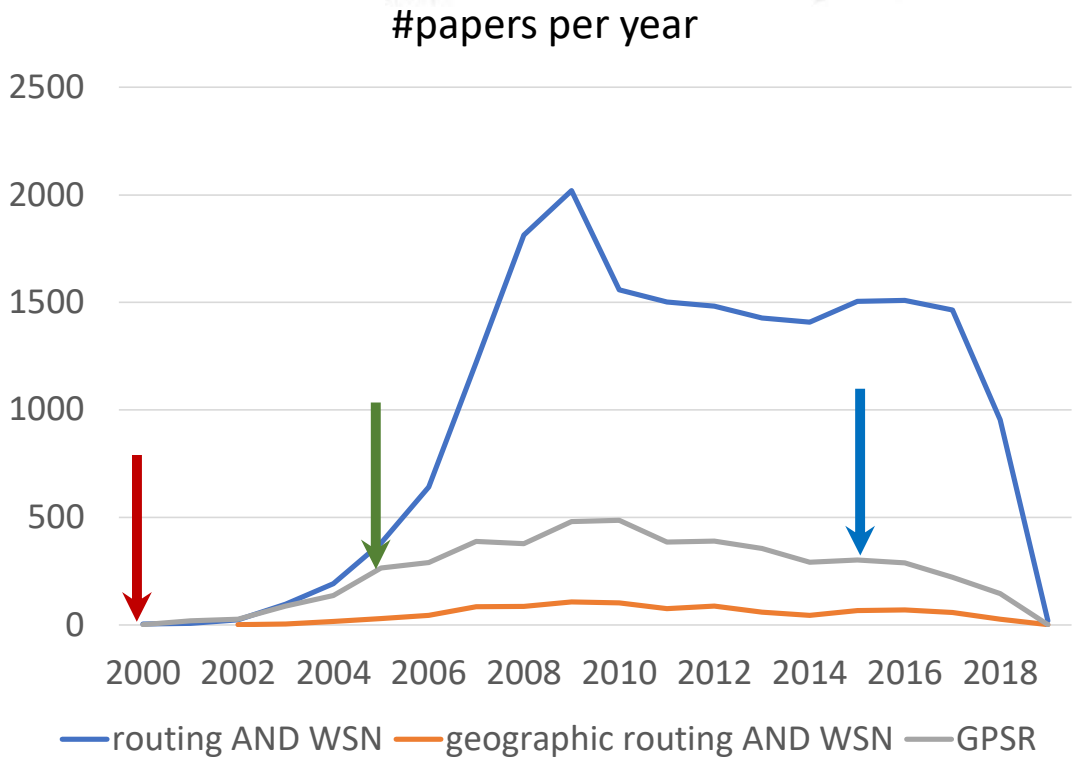
- 6206 citations, a top conference

GPS free coordinate assignment and routing in wireless sensor networks (VCAP), *INFOCOM 2005*

- 223 citations, a top conference

Multi-Dimensional Recursive Routing with Guaranteed Delivery in Wireless Sensor Networks, *ComCom 2015*

- 2 citations, a good impact journal



5. Timeliness

Note: timeliness == right on time

- too early may be as bad as too late!

Some works deserve to be written anyway:

- If they close definitively a research field (they will probably don't get many citations…)
- If they have other values

Again, don't be obsessed by timeliness, but keep an eye to it

Writing papers in the LLM era

References

[1] The Impact of Large Language Models on Scientific Discovery: a Preliminary Study using GPT-4; Microsoft Research AI4Science, Microsoft Azure Quantum, November, 2023. arXiv:2311.07361v2 [cs.CL] 8 Dec 2023. DOI: https://doi.org/10.48550/arXiv.2311.07361

[2] Weixin Liang et Al., "Mapping the Increasing Use of LLMs in Scientific Paper", 2024/04/01 http://arxiv.org/pdf/2404.01268.pdf. DOI: https://doi.org/10.48550/arXiv.2404.01268

[3] Abeba Birhane, Atoosa Kasirzadeh, David Leslie & Sandra Wachter, "Science in the age of large language models", Nature Review Physics, 26/4/2023. DOI: https://doi.org/10.1038/s42254-023-00581-4

Potential uses of LLM in research [1]

Accessing and analyzing scientific literature

suggest relevant research papers,
extract key information,
summarize insights for researchers

Concept clarification.

explaining and providing definitions for scientific terms, concepts, and principles helping researchers better understand the subject matter

Data analysis.

process, analyze, and visualize large datasets from experiments, simulations, and field observations, uncover non-obvious trends and relationships in complex data

Theoretical modeling.

assist in developing mathematical/computational models of physical systems (in fields like physics, chemistry, climatology, systems biology, etc.)

Potential uses of LLM in research [1]

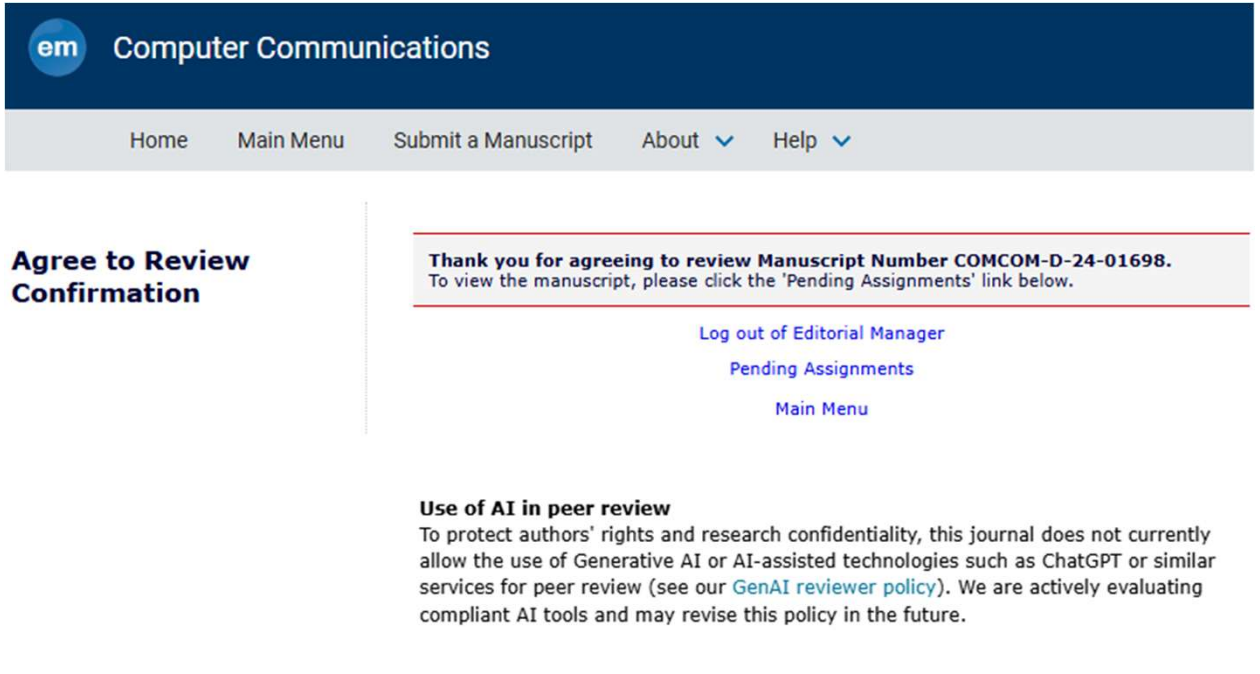
Methodology guidance.	choose the right experimental/ computational methods and statistical tests for research or running simulations on synthetic data?
Prediction.	analyze prior experimental data to make predictions on new hypothetical scenarios and experiments, allowing for a focus on the most promising avenues?
Experimental design.	suggest useful experimental parameters, setups, and techniques that researchers may not have considered, thereby improving experimental efficiency?
Code development.	assist in developing code by generating code from natural language descriptions or suggesting code snippets from a library of prior code?
Hypothesis generation.	By connecting disparate pieces of information across subfields, come up with novel hypotheses (e.g., compounds, materials, etc.) to test, expanding the scope of research?

Does it work?

- From this study on GPT-4 [1] apparently yes, at least for some specific things and domains
- Specifically:
  - Demonstrates potential in various scientific domains, including drug discovery, biology, computational chemistry, materials design, and Partial Differential Equations
  - Can show a “comprehensive grasp of the field”, retrieves information, suggests design principles, recommends computational methods, finds proof approaches
  - Shows limitations when processing biological sequences or in representing and proposing complex structures (e.g. polymers), and also theorem proving has room for improvement
- In general:
  - Crucial to maintain a healthy skepticism when interpreting GPT-4’s output.
  - Requires some skills in questioning it.
  - In many cases, it may be beneficial to combine it with more specialized tools and models designed specifically for scientific discovery tasks

… and it raises concerns about accuracy, plagiarism, anonymity, and ownership of research.

… and reviewing papers?



But ask yourself: I am happy to know that my work has not been read by anybody and automatically reviewed?

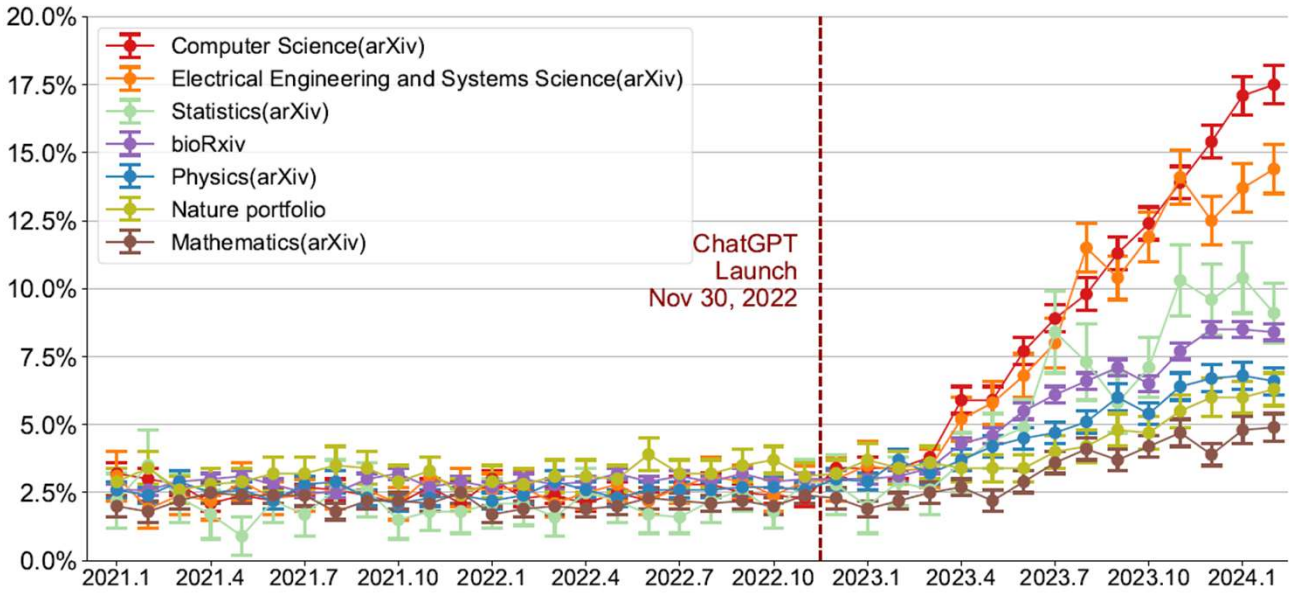
… and writing?

- The International Conference on Machine Learning (ICML) 2023, forbids the inclusion of text generated by LLMs in submitted papers, unless it is used as part of the experiments (https://icml.cc/Conferences/2023/llm-policy)
- The editorial policies of the journal Science forbids text, figures, images, or graphics generated by LLM tools in published works (https://www.science.org/doi/full/10.1126/science.adg7879)
- …

Are researchers really using LLM? (I)

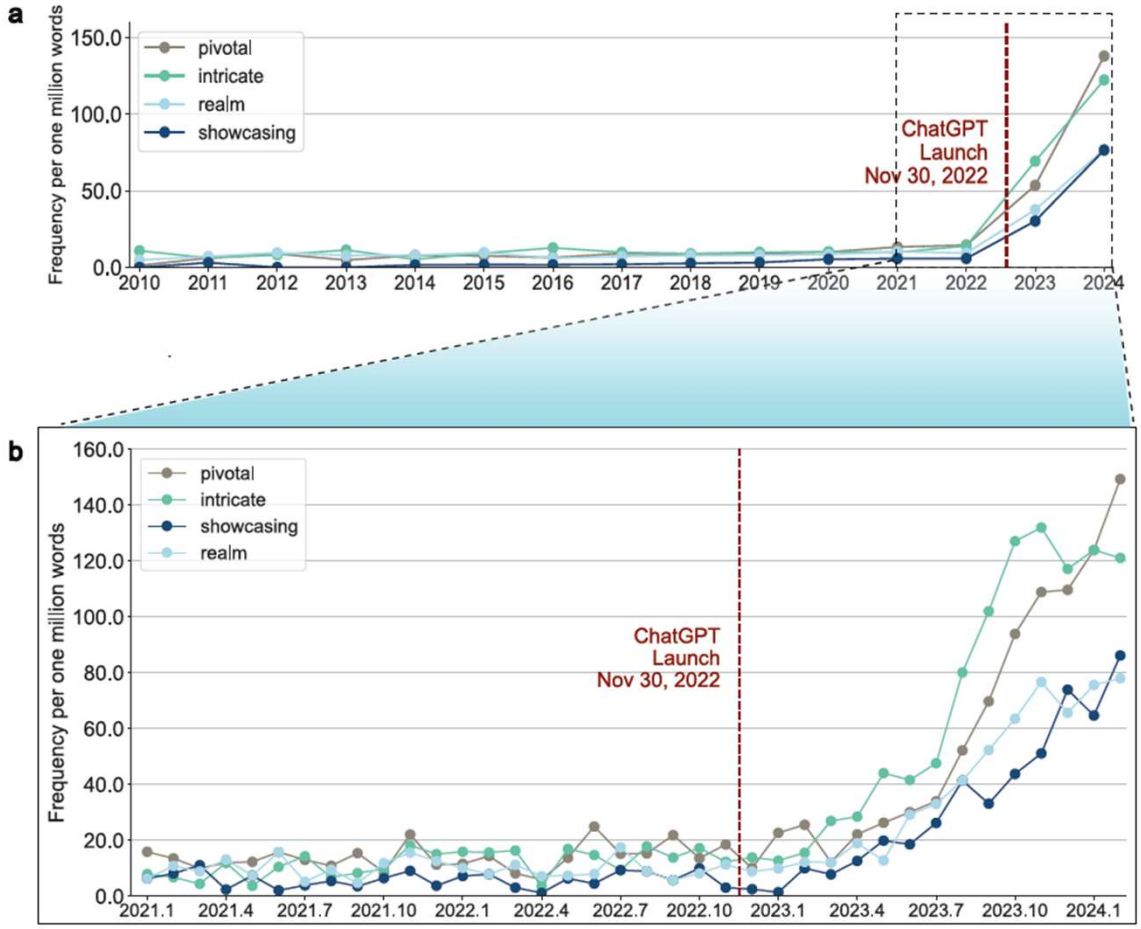
Estimated Fraction of LLM-Modified Sentences across Academic Writing Venues over Time. [2]

- estimate of fraction of sentences that have been substantially modified by LLM in abstracts from various academic writing venues (arXiv, bioRxiv and 15 journals within the Nature portfolio).
- It's a population-level estimates rather than individual document analysis.

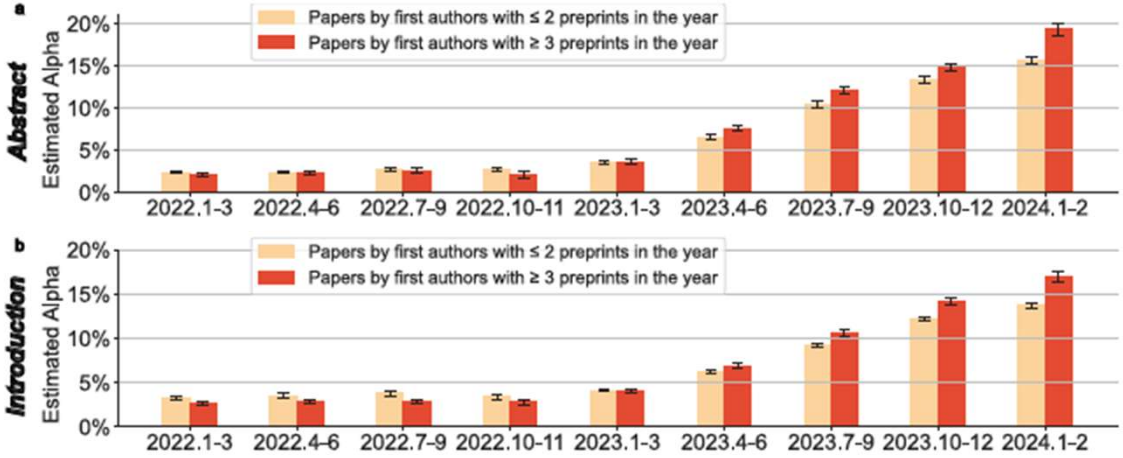


Word Frequency Shift in arXiv Computer Science abstracts over 14 years [2].

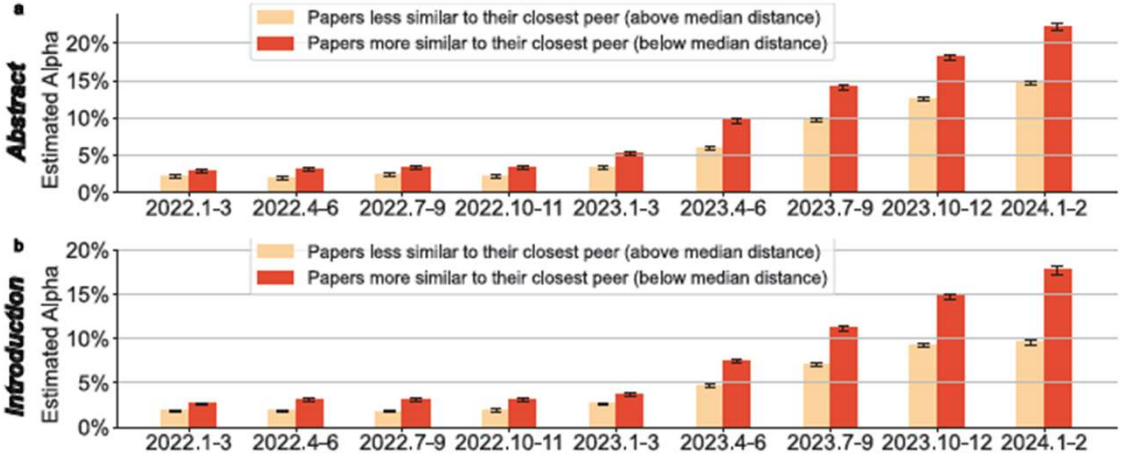
- frequency over time for the top 4 words most disproportionately used by LLM compared to humans
- The words are: realm, intricate, showcasing, pivotal
- a sudden surge in usage starting in 2023…



Are researchers really using LLM? (II)



Papers authored by first authors who post preprints more frequently tend to have a higher fraction of LLM-modified content



Papers in more crowded research areas tend to have a higher fraction of LLM modified content

Who? How?

At an aggregate level that higher levels of LLM-modification are associated with:

- Papers whose first authors post preprints more frequently
- Papers with shorter lengths.

results also show a closer relationship between papers with LLM-modifications:

- May indicate higher use in more crowded fields of study
- or that generated-text is flattening writing diversity

What are the specific concerns for science?

David Leslie (Professor of Ethics, Technology and Society):

- "GenAI technologies lack the basic capacities for intersubjectivity, semantics and ontology that are preconditions for the kind of collaborative world-making that allows scientists to theorize, understand, innovate and discover […]"
- "Researchers, however, must proceed with caution, engaging the affordances provided by these technologies with the same kinds of epistemic humility, deflationary scepticism and disciplined adherence to the scientific method that have functioned as preconditions of modern scientific advancement […] "

Atoosa Kasirzadeh (philosopher and ethicist of science):

- "Although LLMs seem to provide useful general summaries of some scientific texts, for example, it is less clear whether they can capture the uncertainties, limitations and nuances of research that are obvious to the human scientist."
- "Therefore, overreliance on LLMs for tasks such as writing literature reviews should be avoided. Or at least the output should be very carefully fact-checked"
- "[…] the use of LLMs in the peer-review process can endanger trust in it"

… and some opinions [3]

Who bears the responsibility?

Abeba Birhane (cognitive scientist):

- "[…] it would be a grave error to treat LLMs as scientists that can produce science. Knowledge implies responsibility and is never detached from the scientist that produces it."
- "As tools, LLMs, with close and constant vetting by the scientist, can aid scientific creativity and writing. "

What should scientists do?

Sandra Wachter (Professor of Technology and Regulation):

- "Science is fast paced and highly competitive. The pressure to publish can be overwhelming. […] At this stage, we need to think about how to responsibly integrate GenAI into science."
- "Scientists have an ethical responsibility to society to produce knowledge that follows the highest possible standards"

Atoosa Kasirzadeh (philosopher and ethicist of science):

- "[...] the scientific community should take a timely and firm stance to avoid any overreliance on LLMs and to foster practices of responsible science in the age of LLMs"

… and some opinions [3]


Round table...
Express your opinion about the paper I gave you

My pragmatic thought

- The paper I gave you was generated with an LLM:
  - A professional subscription costs around 500€ per year
  - You can write hundreds of those papers in a year…
  - … each paper costs less than one euro
- … but what is (almost) free, does not have a value…
  - I would stop writing papers like that from this point on

Strenghts:

- Can easily produce intermediate text on SoA and possibly more than that
- Helps in better and correctly writing, summarizing etc.

Weaknesses

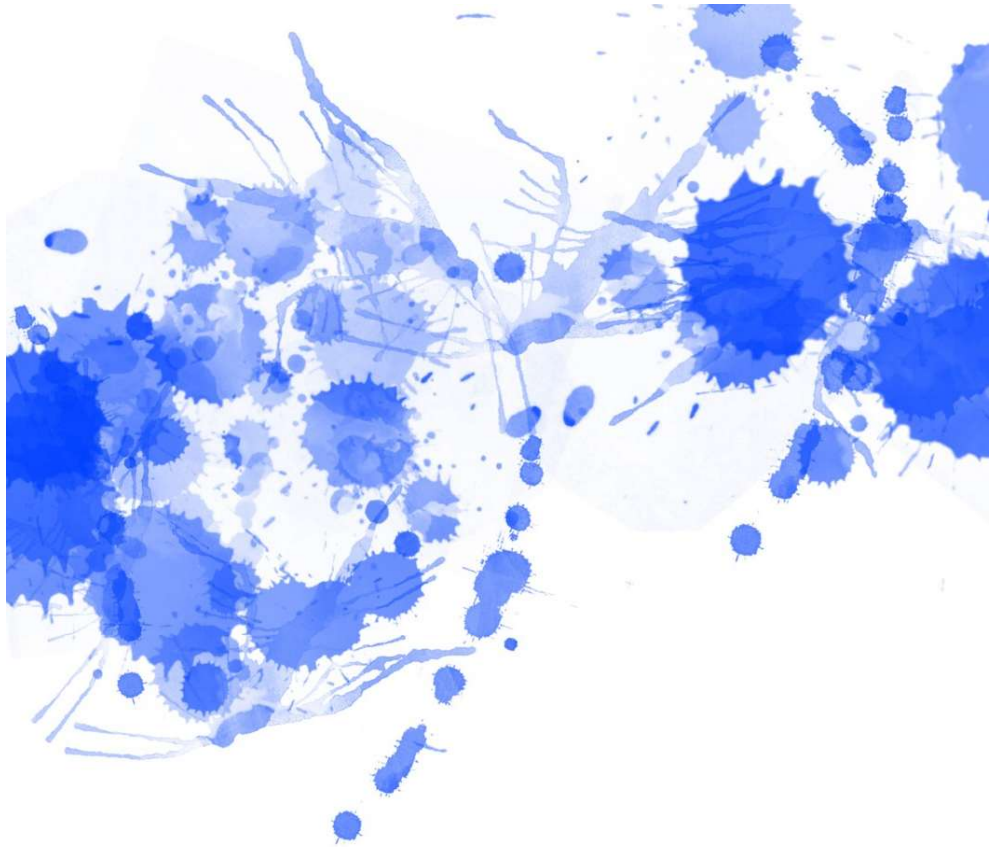
- Does not replace the researcher creativity
- Still need to keep control and output validation
- Still requires a rewriting effort to put the text in the right perspective

Opportunities

- May help to incresases productivity and to be more efficient at work

Threats:

- Other researchers become faster too
- Other may publish fake AI-generated papers to surreptitiously enrich their performance indicators (citations, number of publications…)
- Your research may be disclosed in advance to others (LLM may learn from the text you give to it)
- Reviews automatically generated by AI that are useless to improve your work



A SWOT analysis

My own perspective

- Writing is a creative process that lets me to understand:
  - what are the results,
  - what is the best viewpoint to present them,
  - … and, in the end, what is the paper I'm going to write.
- Personally, I cannot split paper writing from making research…
- … and it makes no sense to outsource the writing …
- However, any tool that improves efficiency while keeping control is, of course, welcome!

shortcuts & cheating

… and why they are not a good idea

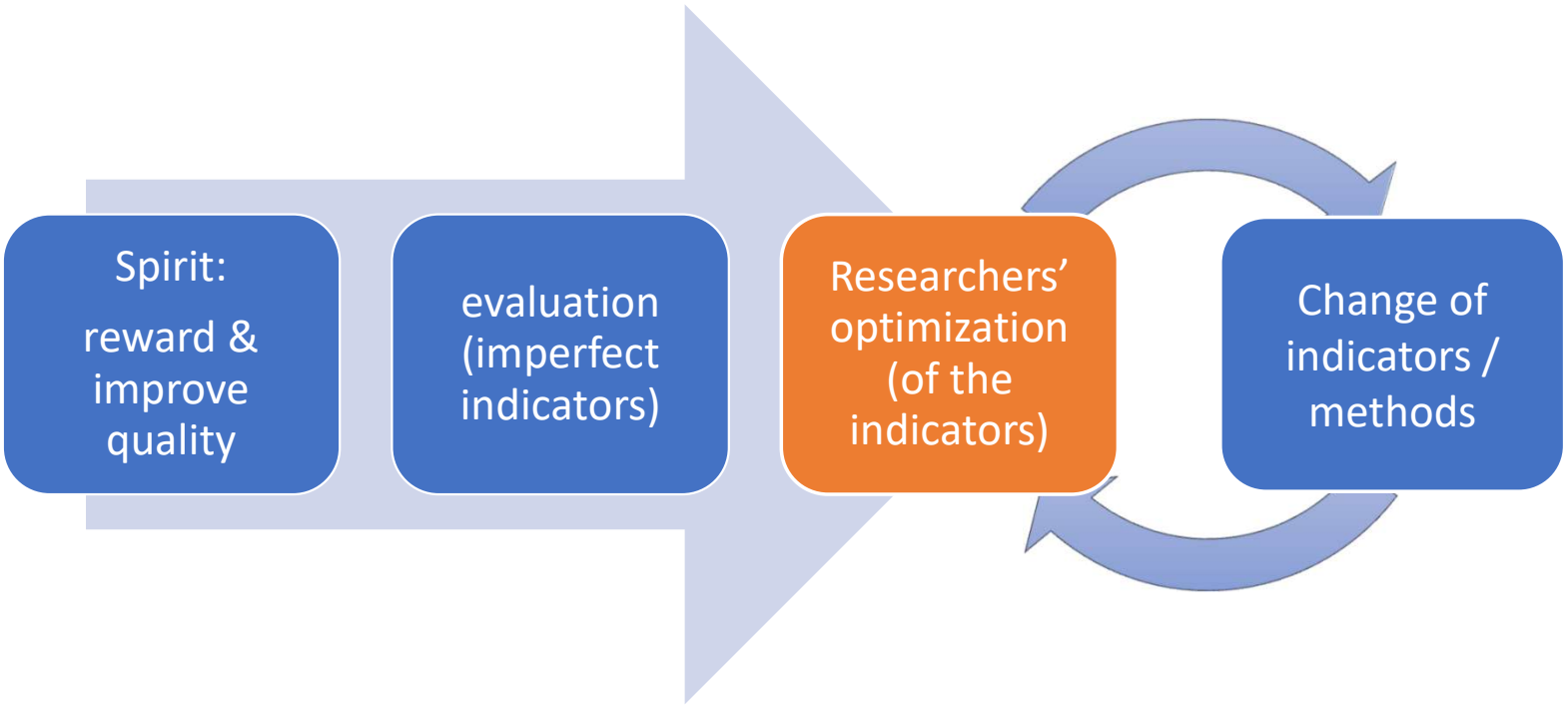
Weaknesses of the performance indicators

- The systematic use of performance indicators to assess researchers is producing a “speculative” bubble
  - Number of papers and citations are growing and growing
- Researchers may use strategies to increase their performance surreptitiously:
  - exchange citations
  - request citations of their papers in their reviews
  - unmoderated use of self-citations
  - pay ghost writers of paper mills…
  - …
- Bad practices of journals to increase their Impact Factors produced new and more complex indexes
  - we already seen the proliferation of impact indexes

Self-citations…

- self-citations are physiological:
  - your work is related to other previous works of yours
  - you make a bit of advertisement to your past works
- their unreasonable use may become a problem for yourself
  - easy to locate and filter out
  - they are written on the stone… are visible forever

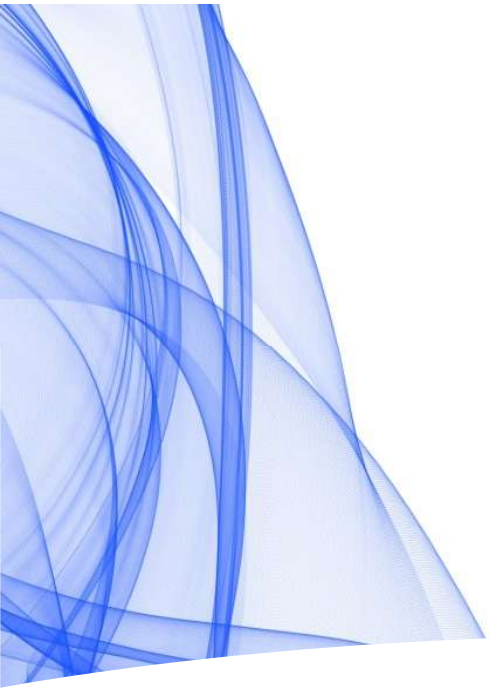
Hyper-specialization: the evaluation loop…



The risks of bad practices

- Bad practices and cheating may seriously affect your reputation
- Bad practices, cheating and iper-specialization are likely to produce immediate changes in the assessment of research
- The great risk is to follow these changes rather than to be always a step ahead

… but how to be a step ahead?



Focus on the quality of your work!

… and, of course, keep an eye to:

1. usefulness
2. venue
3. reputation
4. size of research community
5. Timeliness

… and to other factors that may become important in the future:

1. impact on society
2. interdisciplinarity
3. divulgation/teaching
4. …

Conclusions

Use of indexes in the evaluation

VQR (national evaluation of the quality of research of the universities and departments):

- Papers go directly to peer review and are evaluated with the support of the impact indexes (informed peer review)

ASN (Abilitazione Scientifica Nazionale):

- All indexed papers (including conferences) contribute to total number of citations and h-index.
- Indexed journal papers are also counted separately
- In addition, the commission evaluates the submitted papers and CV of candidates with its own specific criteria

Local calls for RTD/associate/full professor positions

- The commission decides its own specific criteria
- Including if and how to use indexes for conferences and journals

My best five recomendation

Do	do a quality job
Write	write papers for the others, not for yourself
Do not be	do not be obsessed by performance indicators
Keep	keep an eye on trends
Understand	understand the evaluation of research and its evolution

Furthermore…

Impact indexes are not all

when you enter a contest for a research position the commission may also consider your individual effort in your papers, based on:

- author's names position
- specific declarations in the paper
- your declaration in the CV (prime author / equal effort)

That's not all yet…

- your CV is also important
- and your roles in conferences/editorial boards may also be considered
- … plus many other things…

Thank you!