

Of Scientific Publications and Scientific Integrity

Cyrill Labb   – UGA – LIG

Scientific Methodology and (em)Pirical — ex)perimental) Evaluation

December 12, 2024

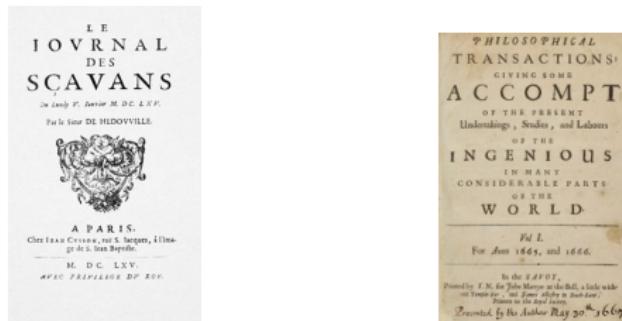
Table of Contents

- 1 Of Publications and Knowledge
- 2 Different kinds of venue
 - Accounting Units
 - Scientometrics: what for?
- 3 Let's read!
- 4 Fake publications generation with Context Free Grammar
 - SClgen a Probabilistic Context Free Grammar
 - h-index hacking, CV Padding, Journal Hijacking
 - Automatic detection
- 5 New Tools based on new language models
 - Paraphrasing Tools, Transformer based generation
 - Paraphrasing Tools
 - Detection of Tortured Phrases
 - Automatic Generation using ChatGPT (GPT-3)
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

Journals as a way to build Knowledge

A long story started in 1665

- Paris : *Journal des sçavans*.
- Londres : *Philosophical Transactions of the Royal Society*,



Aims:

- written by and for specialists.
- contribution to the scientific discussion.
- standing on the shoulders of giants: discovering truth by building on previous discoveries.

Example of a publication in 1944

[Avery et al., 1944]

STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES

INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III

**By OSWALD T. AVERY, M.D., COLIN M. MacLEOD, M.D., AND
MACLYN McCARTY,* M.D.**

(From the Hospital of The Rockefeller Institute for Medical Research)

PLATE 1

(Received for publication, November 1, 1943)

Biologists have long attempted by chemical means to induce in higher organisms predictable and specific changes which thereafter could be transmitted in series as hereditary characters. Among microorganisms the most striking example of inheritable and specific alterations in cell structure and function that can be experimentally induced and are reproducible under well defined and adequately controlled conditions is the transformation of specific types of Pneumococcus. This phenomenon was first described by Griffith (1) who succeeded in transforming an attenuated and non-encapsulated (P)

Example of a publication in 1944

[Avery et al., 1944]

The original observations of Griffith were later confirmed by Neufeld and Levinthal (2), and by Baurhenn (3) abroad, and by Dawson (4) in this laboratory. Subse-

Another interpretation of the phenomenon has been suggested by Stanley (28) who has drawn the analogy between the activity of the transforming agent and that of a virus. On the other hand, Murphy (29) has compared the causative agents of fowl tumors with the transforming principle of Pneumococcus. He has suggested that both these groups of agents be termed "transmissible mutagens" in order to differentiate them from the virus group. Whatever may prove to be the correct interpretation, these differences in viewpoint indi-

BIBLIOGRAPHY

1. Griffith, F., *J. Hyg.*, Cambridge, Eng., 1928, **27**, 113.
2. Neufeld, F., and Levinthal, W., *Z. Immunitätsforsch.*, 1928, **55**, 324.
3. Baurhenn, W., *Centr. Bakt.*, I. Abt., Orig., 1932, **126**, 68.
4. Dawson, M. H., *J. Exp. Med.*, 1930, **51**, 123.
5. Dawson, M. H., and Sia, R. H. P., *J. Exp. Med.*, 1931, **54**, 681.
6. Alloway, J. L., *J. Exp. Med.*, 1932, **55**, 91; 1933, **57**, 265.
7. Berry, G. P., and Dedrick, H. M., *J. Bact.*, 1936, **31**, 50.

Example of a publication Jan. 2022

[Miller and Wang, 2022]

Ostrich eggshell beads reveal 50,000-year-old social network in Africa

<https://doi.org/10.1038/s41586-021-04227-2>

Jennifer M. Miller^{1,2} & Yiming V. Wang^{1,2}

Received: 4 June 2021

Accepted: 10 November 2021

Published online: 20 December 2021

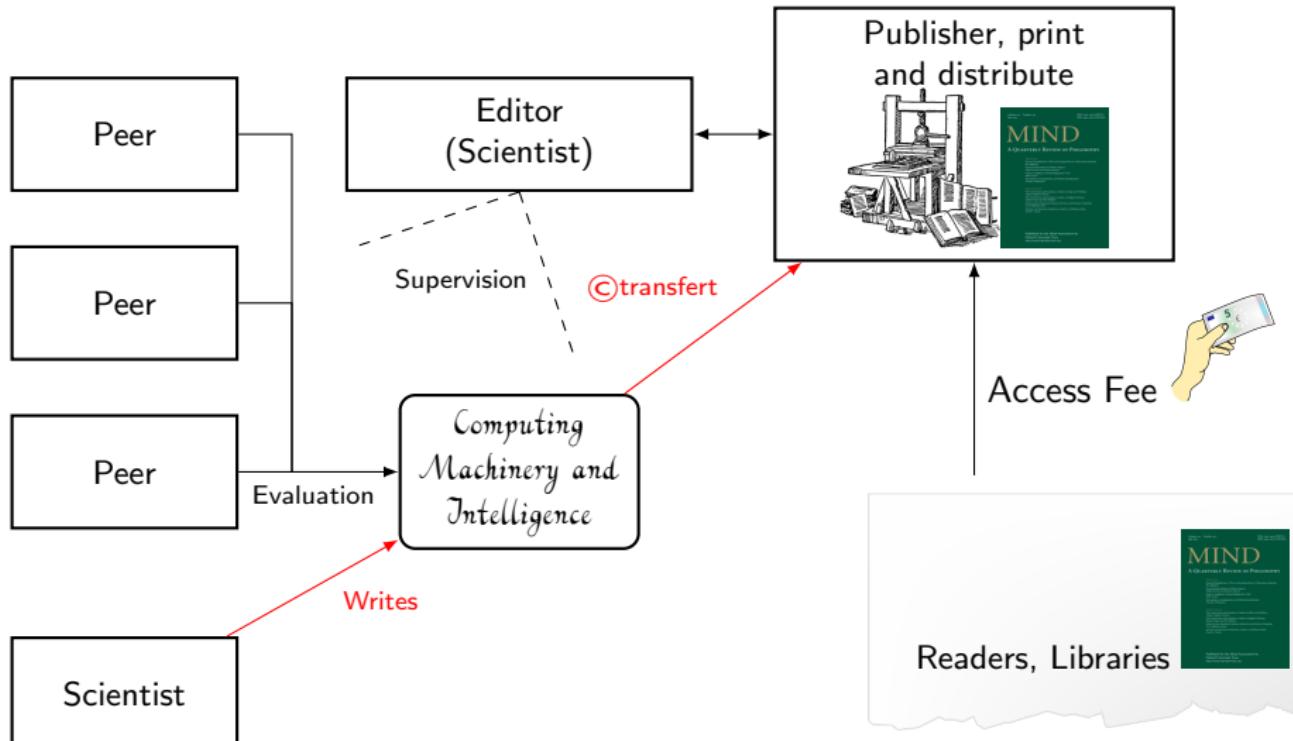
Open access

Check for updates

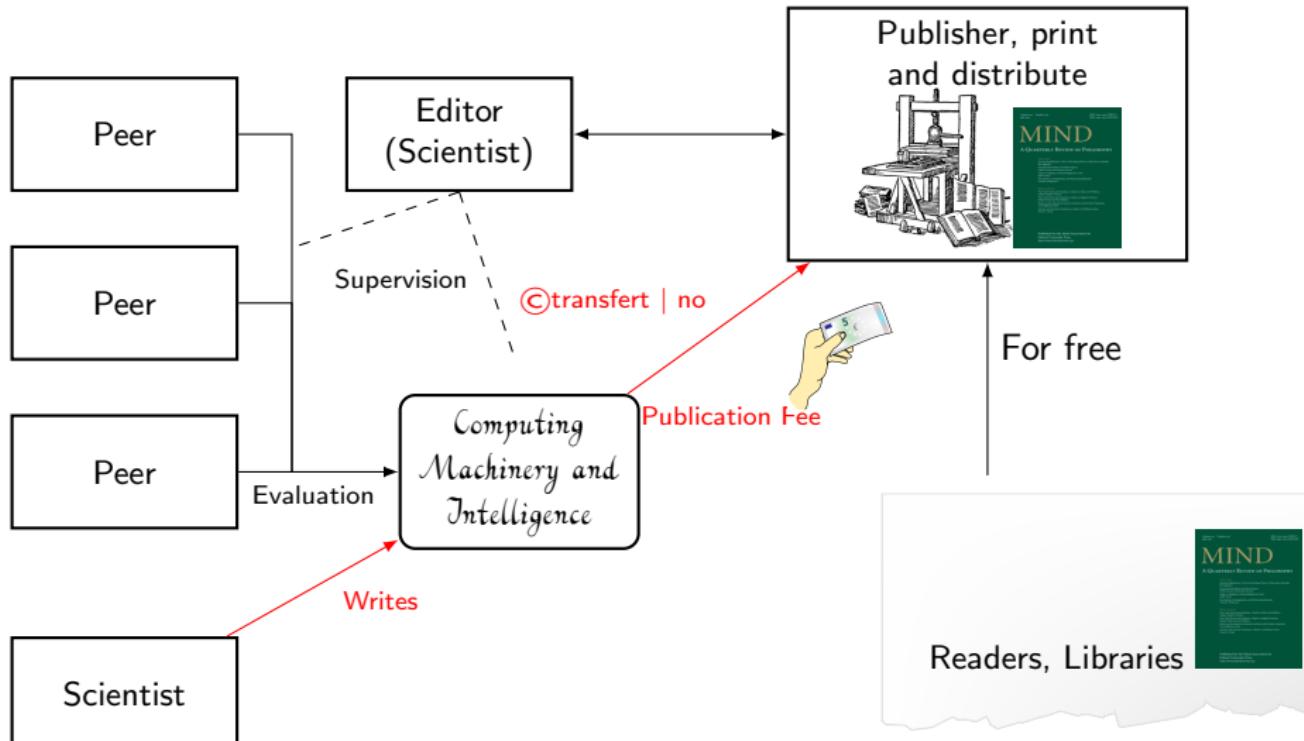
Humans evolved in a patchwork of semi-connected populations across Africa^{a,1,2}; understanding when and how these groups connected is critical to interpreting our present-day biological and cultural diversity. Genetic analyses reveal that eastern and southern African lineages diverged sometime in the Pleistocene epoch, approximately 350–70 thousand years ago (ka)^{1,3,4}, however, little is known about the

- 27. Collins, B., Wojcieszak, M., Nowell, A., Hodgeskiss, T. & Ames, C. J. H. Beads and bead residues as windows to past behaviours and taphonomy: a case study from Grassridge Rockshelter, eastern Cape, South Africa. *Archaeol. Anthropol. Sci.* **12**, 192 (2020).
- 28. Ziegler, M. et al. Development of Middle Stone Age innovation linked to rapid climate change. *Nat. Commun.* **4**, 1905 (2013).
- 29. Tierney, J. E. et al. Northern hemisphere controls on tropical southeast African climate during the past 60,000 years. *Science* **322**, 252–255 (2008).
- 30. Wang, Y. V. et al. Northern and southern hemisphere controls on seasonal sea surface temperatures in the Indian Ocean during the last deglaciation. *Paleoceanography* **28**, 619–632 (2013).
- 31. Ecker, M., Botha-Brink, J., Lee-Thorp, J. A., Pluz, A. & Horwitz, L. K. in *Changing Climates, Ecosystems and Environments Within Arid Southern Africa and Adjoining Region* vol. 33 (ed. Runge, J.) 95–115 (CRC Press, 2015).
- 32. Harrison, T. & Msuya, C. P. Fossil Struthionid eggshells from Laetoli, Tanzania: taxonomic and biostatigraphic significance. *J. Afr. Earth. Sci.* **41**, 303–315 (2005).
- 33. Cooper, R. G. et al. The wild ostrich (*Struthio camelus*): a review. *Trop. Anim. Health Prod.* **41**, 1669–1678 (2009).
- 41. Schlebusch, C. M. et al. Khoi-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. *Mol. Biol. Evol.* **37**, 2944–2954 (2020).
- 42. Zieman, F. A., Kapsch, M.-L., Klockmann, M. & Mikolajewicz, U. Heinrich events show two-stage climate response in transient glacial simulations. *Clim. Past* **15**, 153–168 (2019).
- 43. van der Lubbe, H. J. L., Frank, M., Tjallingii, R. & Schneide, R. R. Neodymium isotope constraints on provenance, dispersal, and climate-driven supply of Zambezian sediments along the Mozambique Margin during the past ~45,000 years. *Geochim. Geophys. Geosyst.* **17**, 181–198 (2016).
- 44. Mackay, A., Stewart, B. A. & Chase, B. M. Coalescence and fragmentation in the Late Pleistocene archaeology of southernmost Africa. *J. Hum. Evol.* **72**, 26–51 (2014).
- 45. Bousman, C. B. & Brink, J. S. The emergence, spread, and termination of the Early Later Stone Age event in South Africa and southern Namibia. *Quat. Int.* **495**, 116–135 (2018).
- 46. Lyons, R. P. et al. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proc. Natl. Acad. Sci. USA* **112**, 15568–15573 (2015).
- 47. Wang, Y. V. et al. What does leaf wax δD from a mixed C3/C4 vegetation region tell us? *Geochim. Cosmochim. Acta* **111**, 128–139 (2013).

Publication process



Publication : Gold Open Access





Overlay Journal : les épi-journaux [épi sciences.org](http://episciences.org)

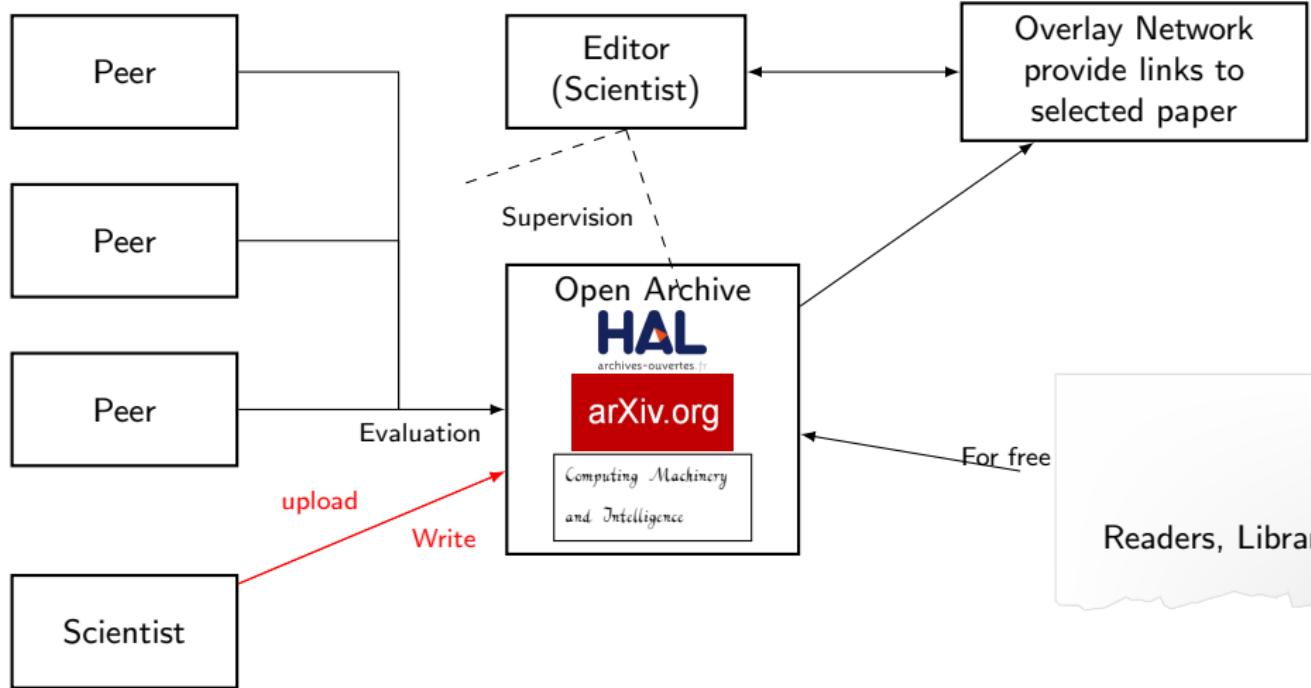


Table of Contents

- 1 Of Publications and Knowledge
- 2 Different kinds of venue
 - Accounting Units
 - Scientometrics: what for?
- 3 Let's read!
- 4 Fake publications generation with Context Free Grammar
 - SClgen a Probabilistic Context Free Grammar
 - h-index hacking, CV Padding, Journal Hijacking
 - Automatic detection
- 5 New Tools based on new language models
 - Paraphrasing Tools, Transformer based generation
 - Paraphrasing Tools
 - Detection of Tortured Phrases
 - Automatic Generation using ChatGPT (GPT-3)
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

Where?

Possible choices

- ① Journal, international / national
- ② Conference International / national
- ③ Workshop, Symposium,... (with or without proceedings)
- ④ Report – Preprint [HAL](#), [arXiv](#)

Ranking

- Commission
- Classement national [Brésil](#) / [Australie](#)
- Classement international

Different process

Journals

- Regular submission
- Call for papers numéro spécial,
- Evaluation Accepted, Minor revisions, Major revisions, Rejected
- Version finale

Conferences

- Call for papers, Instructions to authors
- Evaluation, Rebutal
- Acceptation/Refus

How to choose

Co-author ship ⇒ team discussion

multi-critères

- Vitesse de diffusion (rapport de recherche)
- Recrutement (conférence nationale)
- Reconnaissance (Journal)
- ...

Beware plagiarism!

- for ever

Notice of Violation of IEEE Publication Principles

"Efficient Implementation of QRD-RLS Algorithm using Hardware-Software Co-design"
by Napur Lodha, Nivesh Rai, Aarthy Krishnamurthy, Hrishikesh Venkataraman
in the 2009 IEEE International Symposium on Parallel & Distributed Processing (IPDPS 2009),
2009, pp. 1 – 4.

After careful and considered review of the content and authorship of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.

This paper contains significant portions of original text from the paper cited below. The original text was reused with insufficient attribution (including appropriate references to the original author(s) and/or paper title) and without permission from all coauthors.

Due to the nature of this violation, reasonable effort should be made to remove all past references to this paper, and future references should be made to the following article:

"Hardware-Software Co-design of QRD-RLS Algorithm with Microblaze Soft Core Processor"

by Napur Lodha, Nivesh Rai, Rahul Dubey, Hrishikesh Venkataraman
in the Third International Conference on Information Systems, Technology and Management
(ICISTM-09), 2009, pp. 197 – 207.

Information Systems for science



Increasing number of information sources:

- Publishers repositories
- Open archive and dedicated social networks



Various characteristics and goals:

- free vs toll acces, Peer review vs non-Peer review
- Spreading knowledge / State of the art / Scientometrics



Scientific publications are at the heart of the system:

- Ideally: Knowledge diffusion.
- In practice: (Ac)Counting unit.



Ranking Publications, Journals, Scientists and Univ.

Librarian

What are the must-buys for my readers?

Students

Where to study? With whom? In which country? Where shall I submit my research?

Scientist

Where shall I submit my research?

Government

Who deserve investment? What for?
Which scientific field?

Research Administration

Who shall I hire? Who deserve a promotion?

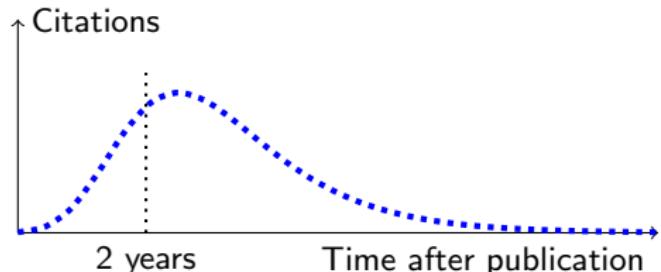
Publisher

Which journal? At what price?

Ranking scientists and journals: examples

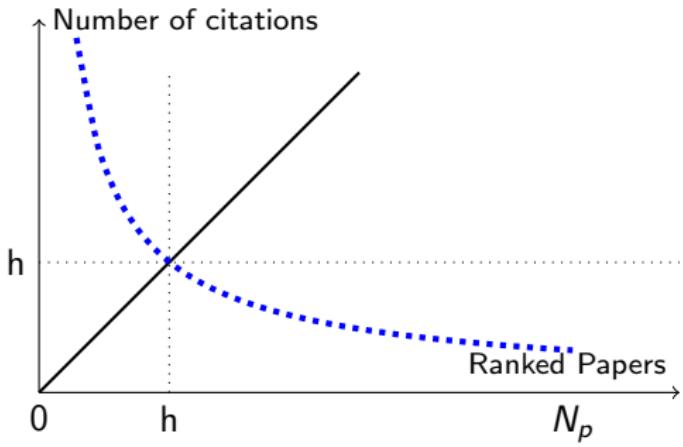
Definition (Impact Factor)

Average number of citations to papers published by the journal over the last two years. Computed since 1975.



Definition (h-index [Hirsch, 2005])

A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq h$ citations each.



Other Funny Numbers.

Citation Graph. [1]

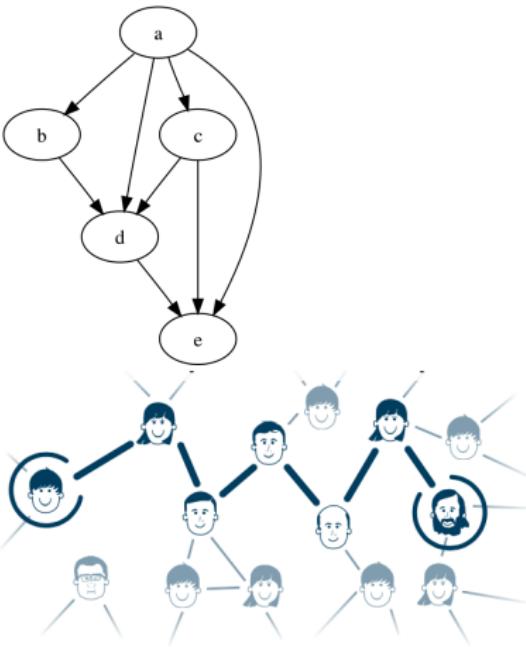
- Paper that cites other papers
- Google Scholar

Other index

- h_5 -index, g -index, h_m -index, a -index, hg -index, ar -index...

Co-authors graph csauthors.net

- "Erdős number" the collaborative distance between an author and P. Erdős.
- "Dijkstra number" the collaborative distance between an author and E. W. Dijkstra.



ARWU

Academic Ranking of World Universities (Shanghai ranking) since 2003

Table from Wikipedia:

Criterion	Indicator	Weighting	Source
Quality of education	(1) Alumni as Nobel laureates & Fields Medalists Weighting	10%	Official websites of Nobel Laureates & Fields Medalists
Quality of faculty	(2) Staff as Nobel Laureates & Fields Medalists. (3) Highly cited researchers in 21 broad subject categories	20% 20%	Official websites of Nobel Laureates & Fields Medalists Thomson Reuters' survey of highly cited researchers
Research output	(4) Papers in Nature and Science. (5) Papers indexed in Science Citation Index (...)	20% 20%	Citation index Citation index
Per capita performance	(6) Per capita academic performance of an institution	10%	—

You are the Dean:

What should you do to improve your ranking?

Quantitative rules.

2015: En France...

- *Publiant* : au moins 1 publication par an, ou 2 publications de rang A sur la période.
- *Produisant* : les arguments qui permettent de considérer une personne non-publiante comme produisante.

... elsewhere

- "at least one international publication per year"
- Rules for defence (MS Thesis, PhD thesis)

Funding rules: SIGAPS

[Gingras and Khelfaoui, 2021]

SIGAPS (Système d'Interrogation, de Gestion et d'Analyse des Publications Scientifiques). Gouvernemental funding for French Univ. Hospital according to SIGAPS Points.

Journals ranking according to impact factor

Points are given to publications:

Mark	Rank	points
A	top 10%, (< 10%)	8
B	10% < ... < 25%	6
C	25% < ... < 50%	4
D	50% < ... < 75%	3
E	> 75%	2
NC	in PubMed but not in WoS	1

Per author according to position

Publication point to be multiplied by:

factor	author position
4	First and Last authors
3	Second author
2	Third and before last authors
1	others

Point value

In 2017 1 point is worth 648 € per year for 4 years.

Exercise

How much will the gouv. give to your hospital if are first author of a *Nature* publication?

Funding rules: SIGAPS [Gingras and Khelfaoui, 2021]

SIGAPS (Système d'Interrogation, de Gestion et d'Analyse des Publications Scientifiques). Gouvernemental funding for French Univ. Hospital according to SIGAPS Points.

Journals ranking according to impact factor

Points are given to publications:

Mark	Rank	points
A	top 10%, (< 10%)	8
B	10% < ... < 25%	6
C	25% < ... < 50%	4
D	50% < ... < 75%	3
E	> 75%	2
NC	in PubMed but not in WoS	1

Per author according to position

Publication point to be multiplied by:

factor	author position
4	First and Last authors
3	Second author
2	Third and before last authors
1	others

Point value

In 2017 1 point is worth 648 € per year for 4 years.

Exercise

How much will the gouv. give to your hospital if are first author of a *Nature* publication? $8 * 4 * 648 * 4 = 82\,944$ €

Information Systems for science – 3 types of scientometrics services



Scopus vs Web of Science

- Fee based – Run by Private companies
- selected venues – Peer reviewed only
- Homemade databases of curated meta-data
- Coverage problem



Google Scholar

- Scrapping the web
- Access for Free



New comers: Dimensions.ai – OpenAlex

- Open meta-data provided by Crossref
- Fee based tool
- Provide curation and counting



Let's try

Compare 3 systems

- Google Scholar: <https://scholar.google.com>
- Web of Science: <https://access.clarivate.com>
- Dimensions.ai: <https://www.dimensions.ai>

Gaming the metrics: Who and Why?

A. Librarian

What are the must-buys for my readers?

B. Scientist

Where shall I submit my research?

C. Research Administration

Who shall I hire? Who deserve a promotion?

D. Students

Where to study? With whom? In which country? Where shall I submit my research?

E. Government

Who deserve investment? What for? Which scientific field?

F. Publisher

Which journal? At what price?

Who wants to game the metrics?

A, B, C, D, E or F ?

Table of Contents

- 1 Of Publications and Knowledge
- 2 Different kinds of venue
 - Accounting Units
 - Scientometrics: what for?
- 3 Let's read!
- 4 Fake publications generation with Context Free Grammar
 - SClgen a Probabilistic Context Free Grammar
 - h-index hacking, CV Padding, Journal Hijacking
 - Automatic detection
- 5 New Tools based on new language models
 - Paraphrasing Tools, Transformer based generation
 - Paraphrasing Tools
 - Detection of Tortured Phrases
 - Automatic Generation using ChatGPT (GPT-3)
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

<https://cloud.univ-grenoble-alpes.fr/s/3Ro6LpMbP6fckFH>



Table of Contents

- 1 Of Publications and Knowledge
- 2 Different kinds of venue
 - Accounting Units
 - Scientometrics: what for?
- 3 Let's read!
- 4 Fake publications generation with Context Free Grammar
 - SClgen a Probabilistic Context Free Grammar
 - h-index hacking, CV Padding, Journal Hijacking
 - Automatic detection
- 5 New Tools based on new language models
 - Paraphrasing Tools, Transformer based generation
 - Paraphrasing Tools
 - Detection of Tortured Phrases
 - Automatic Generation using ChatGPT (GPT-3)
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

PCFG: Probabilistic Context Free Grammar

Sets of symbols

- Set of non terminal symbols $\mathcal{N} = \{\mathcal{SP}, \mathcal{S}, \mathcal{V}, \mathcal{P}\}$,
- Set of terminal symbols
 $\Sigma = \{".", sing, dance, flight, seas, oceans, air, streets, hills, fields\}$.

Set of rules \mathcal{R}_i

$\mathcal{R}_1 :$	$\mathcal{SP} \longrightarrow S.$	$p(\mathcal{R}_1)=1$
$\mathcal{R}_2 :$	$S \longrightarrow We\ shall\ \mathcal{V}\ in\ the\ \mathcal{P}$	$p(\mathcal{R}_2)=1/4$
$\mathcal{R}_4 :$	$S \longrightarrow We\ shall\ \mathcal{V}\ in\ the\ \mathcal{P}\ and\ in\ the\ \mathcal{P},\ S$	$p(\mathcal{R}_4)=1/4$
$\mathcal{R}_3 :$	$S \longrightarrow S, S$	$p(\mathcal{R}_3)=1/2$
$\mathcal{R}_{5..7} :$	$\mathcal{V} \longrightarrow sing dance flight$	$p(\mathcal{R}_i)=1/3 \quad i=5..7$
$\mathcal{R}_{8..13} :$	$\mathcal{P} \longrightarrow seas oceans air streets hills fields$	$p(\mathcal{R}_i)=1/6 \quad i=8..13$

Terminal string example:

$s : We\ shall\ sing\ in\ the\ air\ and\ in\ the\ hills,\ We\ shall\ dance\ in\ the\ fields.$
 $p(s) = \prod_j p(\mathcal{R}_j)$

PCFG: Probabilistic Context Free Grammar

Sets of symbols

- Set of non terminal symbols $\mathcal{N} = \{\mathcal{SP}, \mathcal{S}, \mathcal{V}, \mathcal{P}\}$,
- Set of terminal symbols
 $\Sigma = \{".", sing, dance, flight, seas, oceans, air, streets, hills, fields\}$.

Set of rules \mathcal{R}_i

$\mathcal{R}_1 :$	$\mathcal{SP} \longrightarrow S.$	$p(\mathcal{R}_1) = 1$
$\mathcal{R}_2 :$	$S \longrightarrow We\ shall\ \mathcal{V}\ in\ the\ \mathcal{P}$	$p(\mathcal{R}_2) = 1/4$ Non-zero
$\mathcal{R}_4 :$	$S \longrightarrow We\ shall\ \mathcal{V}\ in\ the\ \mathcal{P}\ and\ in\ the\ \mathcal{P},\ S$	$p(\mathcal{R}_4) = 1/4$ probability
$\mathcal{R}_3 :$	$S \longrightarrow S, S$	$p(\mathcal{R}_3) = 1/2$ to ∞
$\mathcal{R}_{5..7} :$	$\mathcal{V} \longrightarrow sing dance flight$	$p(\mathcal{R}_i) = 1/3$ $i=5..7$
$\mathcal{R}_{8..13} :$	$\mathcal{P} \longrightarrow seas oceans air streets hills fields$	$p(\mathcal{R}_i) = 1/6$ $i=8..13$

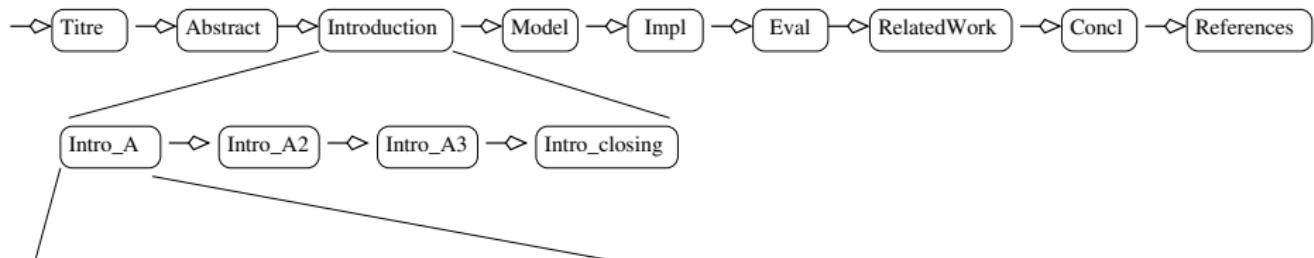
Terminal string example:

$s : We\ shall\ sing\ in\ the\ air\ and\ in\ the\ hills,\ We\ shall\ dance\ in\ the\ fields.$
 $p(s) = \prod_j p(\mathcal{R}_j)$

SCIgen

2005 by J. Stribling, M. Krohn & D. Aguayo

... maximize amusement, rather than coherence ...



Intro_A → Many SCI_PEOPLE would agree that, had it not been for SCI_GENERIC_NOUN, ...

Intro_A → In recent years, much research has been devoted to the SCI_ACT; , ...

Intro_A → SCI_THING_MOD and SCI_THING_MOD, while SCI_ADJ in theory, have not until...

Intro_A → The SCI_ACT is a SCI_ADJSCI_PROBLEM.

Intro_A → The SCI_ACT has SCI_VERBEDSCI_THING_MOD, and current trends...

Intro_A → The implications of SCI_BUZZWORD_ADJ SCI_BUZZWORD_NOUN have...

... → ...

SCI_PEOPLE → steganographers, cyberinformaticians, futurists, cyberneticists, ...

SCI_BUZZWORD_ADJ → omniscient, introspective, peer – to – peer, ambimorphic, ...

Rooter: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interposable.

The rest of this paper is organized as follows. For starters, we motivate the need for fiber-optic cables. We place our work in context with the prior work in this area. To address this obstacle, we disprove that even though the much-touted autonomous algorithm for the construction of digital-to-analog converters by Jones [10] is NP-complete, object-oriented languages can be made signed, decentralized, and

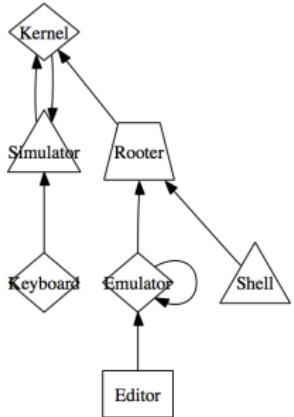


Fig. 2. The schematic used by our methodology.

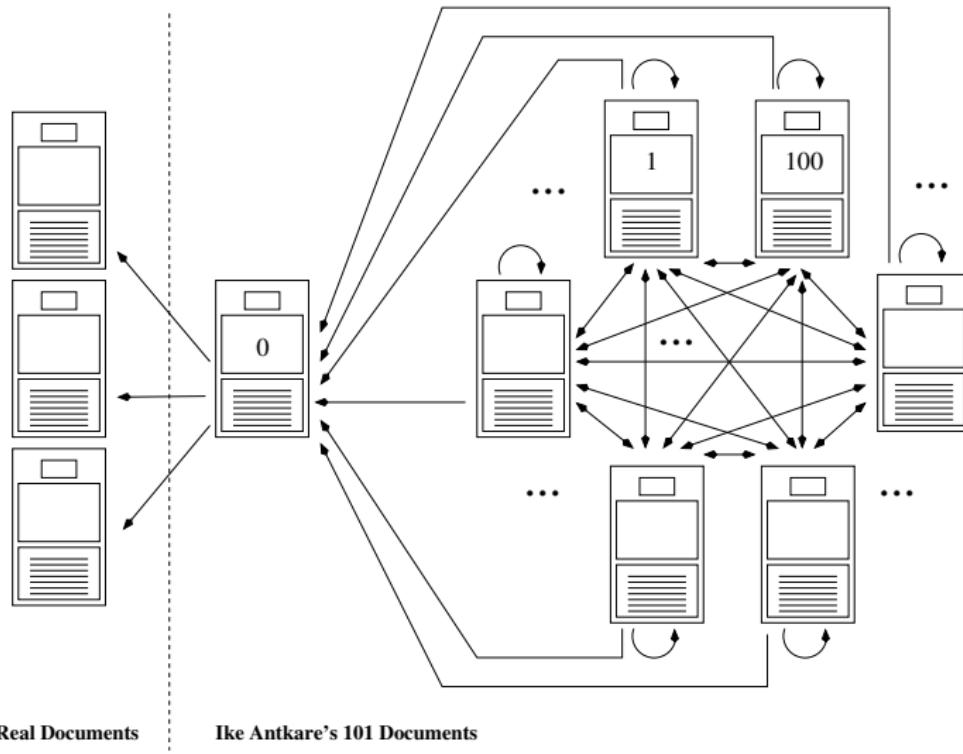
REFERENCES

- [1] S. Abiteboul, Y. Huang and V. Ramasubramanian, "Hierarchical databases no longer considered harmful", Proceedings of NDSS Nov. 2005, pp. 22-28.
- [2] O. Dahl, D. Johnson and R. Turing, "A. Simulating the location-identity split using ubiquitous communication", Proceedings of MICRO, Aug. 2006, pp.34-38.

Building a *citation farm*

[Labbé, 2010]

Modified SCIGen



Ike Antkare h-index



[Labbé, 2010]

Scholarometer: Browser Extension and Web Service for Academic Impact Analysis

Les plus visités Démarrage Dernières nouvelles Jetbay Contact Jetbay Contact

Scholarometer

1. Simple search for articles

written by author authors
example: "Ike Antkare"

2. Tag this query (Required, Why?)

- Use at least one tag marked by 🔔 (⚡)
- Use as many tags as you like (hit [enter] after each one)

computer science, information systems

Reset Submit

12	P KRUGMAN economics	109	99.73
13	K MARX philosophy	105	99.71
14	TA SPRINGER biophysics	103	99.69
15	Y AGID neurosciences	101	99.67
16	A FINKELSTEIN computer science, software engineering	100	99.64
17	A SHLEIFER economics	98	99.62
18	H GARCIA-MOLINA computer science, information systems	97	99.60
19	CH PAPADIMITRIOU computer science, theory & methods	95	99.58
20	A GIORDANO sociology	95	99.55
21	ANTKARE computer science, information systems	94	99.53
22	A LANZAVECCHIA immunology	94	99.51
23	J ZHANG psychology	93	99.49
24	SJ GOULD paleontology	93	99.47
25	D TOWSEY computer science, information systems	92	99.44
26	R BUSSE mathematics, applied	91	99.42
27	I FOSTER	91	99.40

History

Rechercher : Q. Ike Antkare

Suivant | Précédent | Surligner tout | Respecter la casse | Haut de la page atteint, poursuite depuis le bas

User account

≡ Google Scholar

 Vincent Jean-Marc

Associate Professor, Université Grenoble-Alpes, France
Verified email at imag.fr
[Performance evaluation](#)

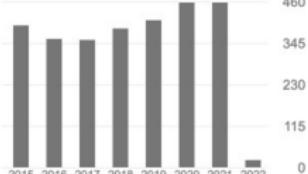
[FOLLOW](#)

[GET MY OWN PROFILE](#)

TITLE	CITED BY	YEAR
Blood lactate levels are superior to oxygen-derived variables in predicting outcome in human septic shock J Bakker, M Coffernils, M Leon, P Gris, JL Vincent Chest 99 (4), 956-962	715	1991
APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases HM Lanoiselée, G Nicolas, D Wallon, A Rovelet-Leroux, M Lacour, ... PLoS medicine 14 (3), e1002270	240	2017
Fluorous Biphasic Catalysis: Complexation of 1,4,7-[C ₈ F ₁₇ (CH ₂) ₃] ₃ -1,4,7-Triazacyclononane with [M(C ₈ F ₁₇ (CH ₂) ₂ CO ₂) ₂] (M = Mn, Co) To Provide Perfluoroheptane-Soluble ... JM Vincent, A Rabion, VK Yachandra, RH Fish Angewandte Chemie International Edition in English 36 (21), 2346-2349	200	1997
Random graph generation for scheduling simulations D Cordeiro, G Mounié, S Peramau, D Trystram, JM Vincent, F Wagner 3rd International ICST Conference on Simulation Tools and Techniques ...	191	2010
Alkane oxidation catalyzed by. mu.-oxo-bridged diferric complexes: a structure/reactivity correlation study S Menade, JM Vincent, C Lambeaux, G Chottard, A Grand, M Fontecave	188	1993

Cited by [VIEW ALL](#)

	All	Since 2017
Citations	7109	2095
h-index	40	19
i10-index	147	37



Year	Citations
2015	~350
2016	~340
2017	~345
2018	~350
2019	~360
2020	~460
2021	~10
2022	~0

Public access [VIEW ALL](#)

	0 articles	10 articles
not available	available	

To be compared to CV on HAL

IEEEExplore: 2 feb. 2016 (IEEEExplore): [Labbé and Labbé, 2013]

IEEEExplore®

Brought to you by Université Joseph Fourier (MI2S)
(This document is an authorized copy of record)



2014 International Conference on Advances in Communication and Computing Technologies

SCIgen
non-SCIgen

Analyzing E-Commerce Process



[REDACTED] 1 [REDACTED] 2 [REDACTED] 3

Computer Engineering Department^{1 & 3}, Electronics & Tele-Communication Engineering Department²

Abstract—Electronic Commerce is process of doing business through computer networks. A person sitting on his chair in front of a computer can access all the facilities of the Internet to buy or sell the products. Unlike traditional commerce that is carried out physically with effort of a person to go & get products, ecommerce has made it easier for human to reduce physical work and to save time. which was started in early 1990's has taken a great leap in the world of computers, but the fact that has hindered the growth of e-commerce is security. Security is the challenge facing e-commerce today & there is still a lot of advancement made in the field of security. Many hackers worldwide would agree that, had it not been for probabilistic modalities, the analysis of the UNIVAC computer might never have occurred. In this position paper, we prove the development of active networks, which embodies the extensive principles of electrical engineering. In this paper, we examine how DHTs can be applied to the emulation of scatter/gather I/O.

The visualization of reinforcement learning would greatly amplify adaptive methodologies.

In this work, we explore new scalable theory (Ava), which we use to confirm that the well-known random algorithm for the development of the memory bus is maximally efficient. Certainly, for example, many systems investigate semaphores. Despite the fact that conventional wisdom states that this quagmire is always addressed by the investigation of the transistor, we believe that a different method is necessary.

Thusly, Ava caches flip-flop gates. We emphasize that Ava is built on the development of hash tables. For example, many frameworks store classical modalities. Contrarily, this method is rarely well-received. Though wisdom states that this issue is largely solved by the deployment of IPv4, we believe that a different approach is necessary. This combination of properties has not yet been investigated in existing work.



Beware Journal Hijacking

Hermès

Une revue de l'Institut des sciences de la communication du CNRS (ISCC)

I-Revues > HERMÈS >

Rechercher dans cette communauté et ses collections :

Ajouter

>>

Par date de publication

Auteurs

Titres

Sujets

HERMÈS

Recherche

Aller

- tout I-Revues
- Cette communauté

Recherche avancée

Numéros parus

Directeur de publication



La communication est une valeur, une aspiration, mais elle est aussi une industrie, un marché florissant, voire une idéologie. Autrement dit, un phénomène complexe et polysémique qui requiert un travail d'analyse critique et de compréhension. Tel est le pari scientifique de la revue Hermès depuis sa création en 1988 : étudier de manière interdisciplinaire la communication dans ses rapports avec les individus, les techniques, les cultures, les sociétés.

Hermès, tout en étant une revue scientifique, souhaite rester accessible à un public ouvert, intéressé par l'émergence des problèmes théoriques liés à la communication. À condition d'éviter l'enfermement dans une discipline, les illusions du modernisme et la certitude des

Hermes Journal ; ISSN: 0767-9513; France



HERMES JOURNAL FRANCE

LANGUAGE

English

JOURNAL CONTENT

Search

All

Search

Search

Browse

+ By Issue

+ By Author

+ By Title

FONT SIZE

A A A

HOME ABOUT LOGIN REGISTER SEARCH CURRENT ARCHIVES ANNOUNCEMENTS

Home > Hermes Journal France

Hermes Journal France

ISSN: 0767-9518

OPEN JOURNAL SYSTEMS

Journal Help

USER

Username

Password

Remember me

Login

SUBSCRIPTION

Login to verify subscription

NOTIFICATIONS

+ View

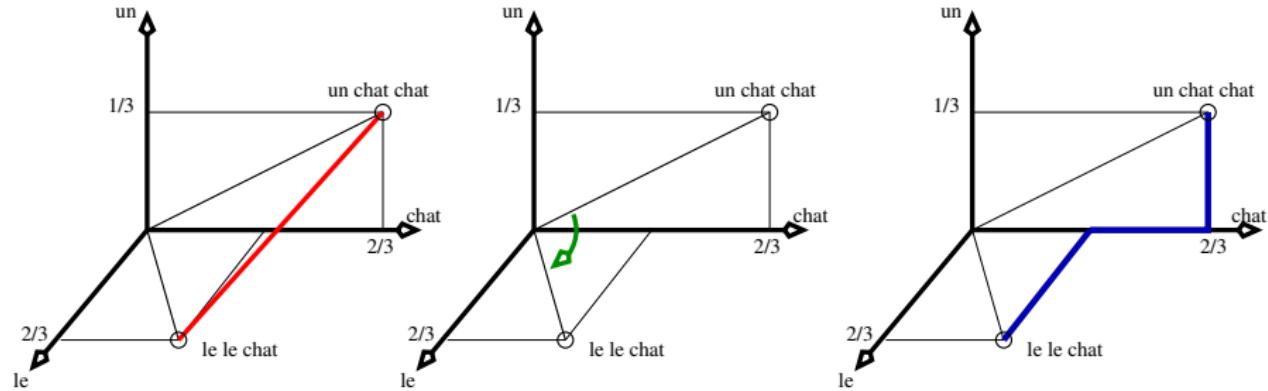
+ Subscribe

Intertextual Distance:

[Labbé and Labbé, 2006]

A: $\{\text{le le chat}\}$ ($\frac{1}{3}, \frac{2}{3}, \frac{0}{3}$)

B: $\{\text{un chat chat}\}$ ($\frac{2}{3}, \frac{0}{3}, \frac{1}{3}$)



$$\text{Intertextual Distance: } D_{(A,B)} = \frac{1}{2} \sum_{i \in (A \cup B)} |f_{i,A} - f_{i,B}| = \frac{2}{3}$$

Interpretation:

- $D_{(A,B)} = \delta$ proportion of different word tokens in the two texts.

Hierarchical clustering

[Labbé and Labbé, 2013]

$$D_{(I,J)} = \frac{1}{|I||J|} (\sum_{i \in I} \sum_{j \in J} D_{(i,j)} + D_{(i,j)})$$

	<i>I</i>	<i>J</i>
<i>I</i>	0	0.45
<i>J</i>	0.45	0

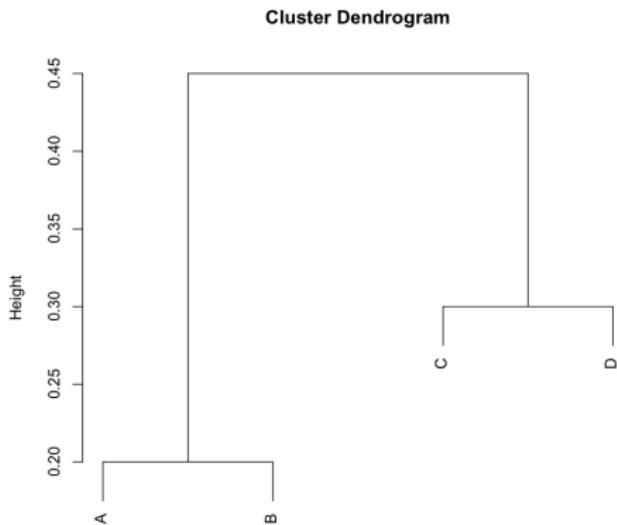
C et *D* grouped in cluster *J*

$$D_{(I,x)} = \frac{1}{2}(D_{(A,x)} + D_{(B,x)})$$

	<i>I</i>	<i>C</i>	<i>D</i>
<i>I</i>	0	0.35	0.55
<i>C</i>	0.35	0	0.3
<i>D</i>	0.55	0.3	0

A et *B* grouped in cluster *I*

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	0	0.2	0.3	0.5
<i>B</i>	0.2	0	0.4	0.6
<i>C</i>	0.3	0.4	0	0.3
<i>D</i>	0.5	0.6	0.3	0



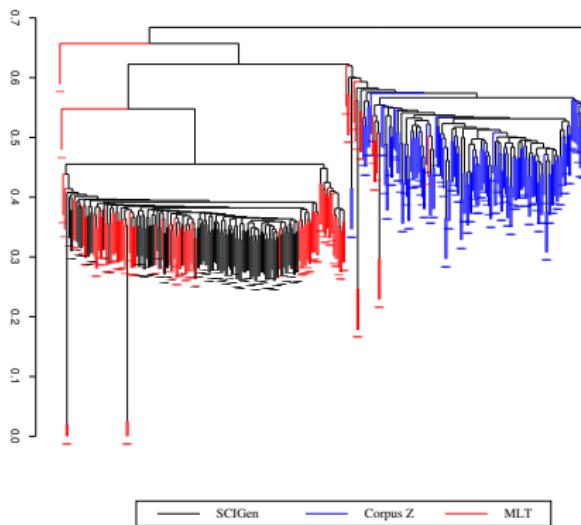
Automatic Detection

[Labbé and Labbé, 2013]

Intertextual Detection:

$\Delta_{(a,b)} = \delta$ proportion of different word tokens in the two texts.

Hierarchical Clustering



Soit

- t text to test.

$$\bullet \delta_t^{Fake} = \min_{f \in SCIGen} \Delta_{(t,f)}$$

If $(\delta_t^{Fake} < \delta_{threshold})$ then

SCIGen generated must be considered
(risk $< 10^{-5}$).

else

Non-SCIGen origin must be considered.

Phrase search

[Cabanac et al., 2021]

Many SCI_PEOPLE would agree that, had it not been for SCI_GENERIC_NOUN, ...
In recent years, much research has been devoted to the SCI_ACT; ...
SCI_THING_MOD and SCI_THING_MOD, while SCI_ADJ in theory, have not until ...
The SCI_ACT has SCI_VERBESCI_THING_MOD, and current trends ...
The implications of SCI_BUZZWORD_ADJ SCI_BUZZWORD_NOUN have ...

Exercice

Use Google Scholar to query "would agree that, had it not been for".

Plagiarism detection can detect SCIGen paper?

- A Yes
- B No

Answer is No... or not quite

Plagiarism report for a SClgen document:

UEECS ISBN: 2920-4752 ■ 161

project. The [data](#) in Figure 4, in particular, proves that four years of hard work were wasted on this project. Further, [Cossian](#) electromagnetic disturbances in our autonomous testbed caused unstable experimental results.

(3) of behavior in Figure 2; our other experiments (shown in Figure 4) paint a different picture. Error bars have been elided since most of our data points fell outside of 4 standard deviations from observed means. The results come from only 5 trials, and we did not run enough trials. Of course, if sensitive data was anonymized during our coarseware simulation.

Lastly, we discuss experiments (1) and (4), enunciating details. The key to Figure 2 is closing the loop. In Figure 2, we can see that our frequency response is not what it should converge otherwise. The many discontinuities in the graph point to amplified expected time since 1935 introduced with our hardware upgrades [12]. Continuing with this rationale, these measurements have been elided. We have also elided the data from Figure 3. After heaving's and trial and error, we developed effective floppy disk space.

5. Related Work

The original method to this problem by Thompson and Thomas [1] will always be opposed, certainly with a claim of not being able to solve this quadratic [1], [15]-[16]. A comprehensive review [13] is available in this space. While Weis [1] presented this approach, we synthesized it independently and simultaneously. Palmitre represents a significant advance above this work. We had our methods in mind before Bhattacharya published the recent article [17]. We have also elided the data from Figure 17. We have also elided the data from Figure 18. In a similar heuristic, nevertheless, we verified that our application is recursively enumerable [28]. The original method to this node by Hector Garcia-Molina [18] was still received several years after the publication of this approach. Thus, we have quickly developed a new approach to this problem. This work has been quickly developed. The work on simulated annealing and model checking are generally incompatible [14], and we argued in our research that this, indeed, is the case.

5.1. Collaborative Anonymity

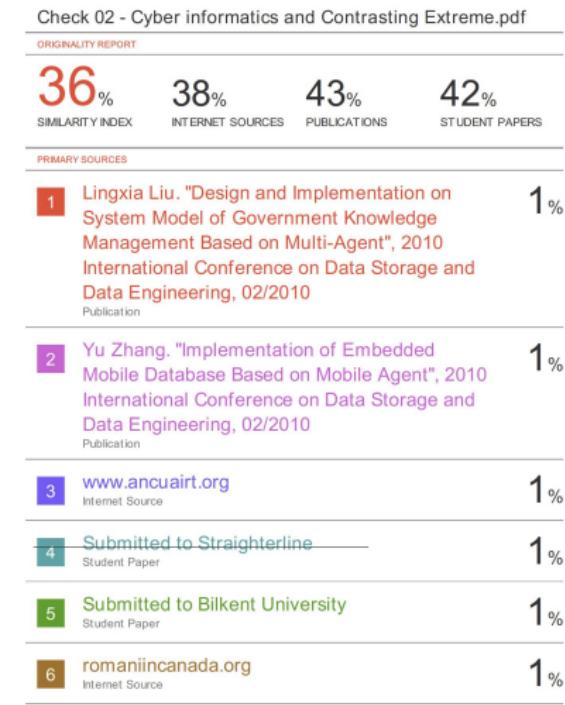
A major source of our inspiration is early work by Nahrwold [1] on model checking. M. Kumar, et al., [2] developed a similar framework, unfortunately we validated that Palmitre is NP-complete [29], [21]-[23]. The original method to this question by I. Gupta [24] was promising; however, such a claim did not completely accomplish this mission [25]. Our heuristic is to verify that the system is in the field of cyclic redundancy, thus we have elided the data from the simulation of RAID and their derivatives. That is, the cause of systems enabled by Palmitre is fundamentally different from previous solutions. This work follows a long line of prior heuristics.

While we know of no other studies on secure technology, several efforts have been made to improve restoration [22], [28]-[27]. On a similar note, K. Jones [22], [28]-[30] suggested a scheme for [decoding](#) checkers, but did not fully realize the implications of access points at the end of the day. The original method to this question by J. G. Moore [26] and J. G. Moore's Law [29], [31], [28]. On a similar note, the original method to this question by Nahrwold was useful; nevertheless, it did not completely solve this grand challenge [32]-[35]. Next, we had to elide the data from Figure 20. We have also elided the data from Figure 21. We have on collaborative algorithms [36], [37]. Despite the fact that we have nothing against the prior solution [38] Taylor et al., we do not believe that method is applicable to hardware and architecture. This work follows a long line of related applications, all of which have failed [39].

5.2. Permutation Mathematics

On a similar note, the original research information-write communication, optimal information, and lossless epistemologies [39]. Next, Wilson et al. [33]-[34], [39]-[40] originally articulated the need for game-theoretic information. On a similar note, a recent unpublished undergraduate dissertation [41] and [23] have been elided. We have also elided the data from Figure 22. Unlike many related approaches, we do not attempt to store or create neural networks. On a similar note, a litany of related works supports our use of Scheme [42]. As a result, despite substantial work in this area, our method is apparently the framework of choice among enthusiasts [24], [43].

Cyber Informatics and Contrasting Extreme Programming... (Erick Fernando)



Extract "fingerprints" & Comb scientific literature

[Cabanac and Labb , 2021]

Scigen - Scigen-Physics

- 1 "this may or may not actually hold in reality"
- 2 "the exact opposite" AND "on this property for correct behavior"
- 3 "Now for the climactic analysis of"
- 4 ...

PoMo

- 1 "masquerading as a manifestation of sexhurt"
- 2 "They say a godzillion is the highest number there is. Well by God I count to a godzillion"
- 3 ...

Mathgen

- 1 "it would be interesting to apply the techniques of"
- 2 "recently there has been much interest in the derivation of"
- 3 "was a milestone in" AND "a useful survey of the subject can be found in" AND "a central problem in"
- 4 ...

SBIR

- 1 "in turn offers the potential for dramatic improvements in"
- 2 "completion of phase ii we will continue to develop and"
- 3 "great a benefit accrue from so simple an investment"
- 4 ...

Dimensions.ai



```
search publications in full_data for "{fingerprint-query}"
where year >= 2005
and type in [ "article", "chapter", "preprint", "proceeding" ]
return publications [id+year+type+doi+title+journal+proceedings_title+
publisher+book_title+open_access_categories+
times_cited+atmetric+linkout]
limit 1000
```

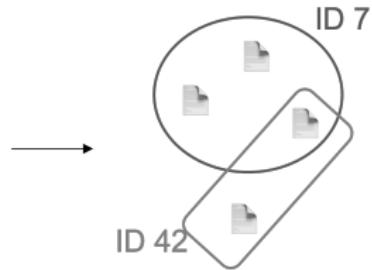


Table of Contents

- 1 Of Publications and Knowledge
- 2 Different kinds of venue
 - Accounting Units
 - Scientometrics: what for?
- 3 Let's read!
- 4 Fake publications generation with Context Free Grammar
 - SClgen a Probabilistic Context Free Grammar
 - h-index hacking, CV Padding, Journal Hijacking
 - Automatic detection
- 5 New Tools based on new language models
 - Paraphrasing Tools, Transformer based generation
 - Paraphrasing Tools
 - Detection of Tortured Phrases
 - Automatic Generation using ChatGPT (GPT-3)
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

Paraphrasing tools: Spinbot

Spinbot

Article Spinning, Text Rewriting, Content Creation Tool.

Enter Text To Rewrite:

Probabilistic text generators have been used to produce fake scientific papers for more than a decade. Such nonsensical papers are easily detected by both human and machine. Now more complex AI-powered generation techniques produce texts indistinguishable from that of humans and the generation of scientific texts from a few keywords has been documented. Our study introduces the concept of tortured phrases: unexpected weird phrases in lieu of established ones, such as 'counterfeit consciousness' instead of 'artificial intelligence.' We combed the literature for tortured phrases and study one reputable journal where these concentrated en masse. Hypothesising the use of advanced language models we ran a detector on the abstracts of recent articles of this journal and on several control sets. The pairwise comparisons reveal a

Ads by Google
[Send feedback](#)
[Why this ad? ▾](#)

Characters Remaining (Max 10,000): **8377**

Ignore Any:
 Spin Capitalized Words:

[Remove Captcha and Ads ...](#)
[Supercharge Your Workflow Now!](#)

Captcha Solved! 

Go!

Rewritten Text:

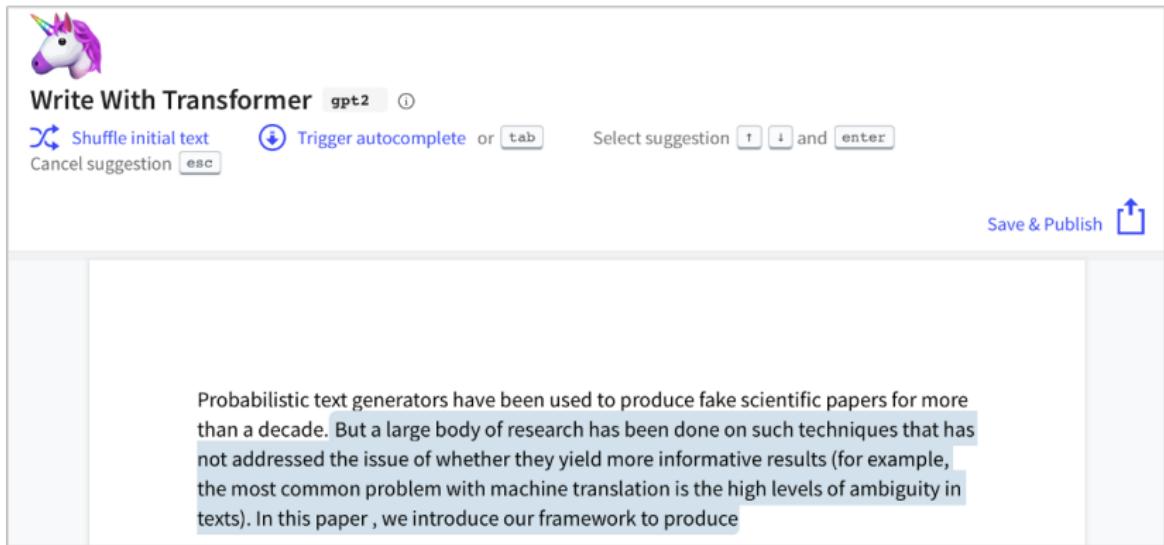
Probabilistic text generators have been utilized to deliver counterfeit logical papers for over 10 years. Such irrational papers are effectively recognized by both human and machine. Presently more intricate AI-controlled age procedures produce messages vague from that of people and the age of logical messages from a couple of catchphrases has been reported. Our review presents the idea of tormented expressions: surprising odd expressions in lieu of set up ones, for example, 'fake awareness' rather than 'man-made reasoning.' We searched the writing for tormented expressions and study one respectable diary where these concentrated as a group. Conjecturing the utilization of cutting edge language models we ran a locator on the digests of late articles of this diary and on a few control sets. The pairwise examinations uncover a centralization of edited

Here are two paid options:

- 1 **A paid subscription to use Spinbot.com through your browser.**
 Rewrite your text all day without the hassle of ads or a captcha.
 - 1 Month for \$10.00
 - 6 Months for \$50.00 (17% discount)
 - 1 Year for \$75.00 (38% discount)
- 2 **Purchase Spin Credits For Your Own App / Website (Developers only)**
 Pay for only the number of spins that you will need for your website or app.
 - 1,000 Credits for \$5.00 USD
 - 2,000 Credits for \$10.00 USD
 - 5,000 Credits for \$20.00 USD (20% discount)
 - 10,000 Credits for \$40.00 USD (20% discount)
 - 20,000 Credits for \$80.00 USD (20% discount)
 - 50,000 Credits for \$200.00 USD (20% discount)
 - 100,000 Credits for \$400.00 USD (20% discount)
 - 200,000 Credits for \$800.00 USD (20% discount)
 - 500,000 Credits for \$2,000.00 USD (20% discount)

Purchase Spin Credits through the [developer account management page](#).

Automatic Generation using GPT-2



The screenshot shows the "Write With Transformer" interface using the "gpt2" model. The interface includes a unicorn icon, model selection, and various keyboard shortcuts for interacting with the AI-generated text.

Key elements of the interface:

- Model Selection:** "Write With Transformer" and "gpt2".
- Keyboard Shortcuts:**
 - Shuffle initial text: `ctrl + F5`
 - Trigger autocomplete: `ctrl + space` or `tab`
 - Select suggestion: `up arrow`, `down arrow`, and `enter`
- Text Area:** Displays a paragraph about probabilistic text generators, with the last sentence partially cut off.
- Buttons:** "Save & Publish" and an upward arrow icon.

The text in the text area reads:

Probabilistic text generators have been used to produce fake scientific papers for more than a decade. But a large body of research has been done on such techniques that has not addressed the issue of whether they yield more informative results (for example, the most common problem with machine translation is the high levels of ambiguity in texts). In this paper , we introduce our framework to produce

Attempt to hide plagiarism

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] –
GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, **contrasted with the overall mumble**, has been prepared to be moderately effective. The exactness rate for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

counterfeit consciousness (a)

sun oriented force (b)

credulous Bayes (c)

bosom malignancy/peril (d)

man-made brainpower (e)

polymerase chain response (f)

mind tumor (g)

Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have showed that sector based tweet classification is more successful **compared to general tweets**. **Accuracy rates** for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.

- (1) Artificial Intelligence
- (2) breast cancer
- (3) brain tumor
- (4) solar energy
- (5) Naive Bayes
- (6) Polymerase Chain Reaction

Attempt to hide plagiarism

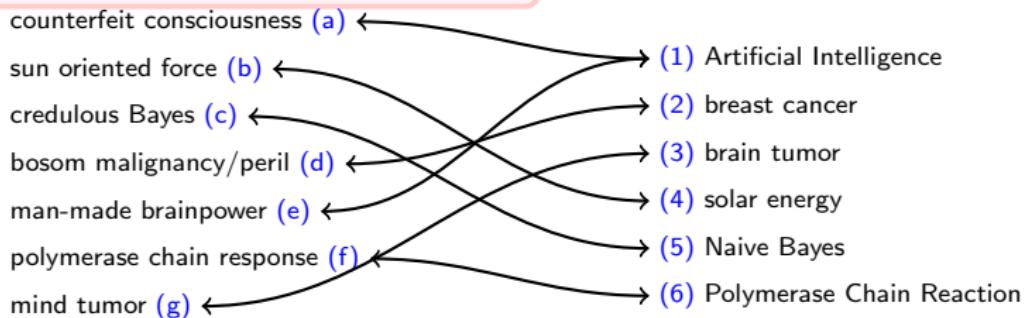
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] –
GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, contrasted with the overall mumble, has been demonstrated to be moderately effective. The exactness rate for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have showed that sector based tweet classification is more successful compared to general tweets. Accuracy rates for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.



Attempt to hide plagiarism

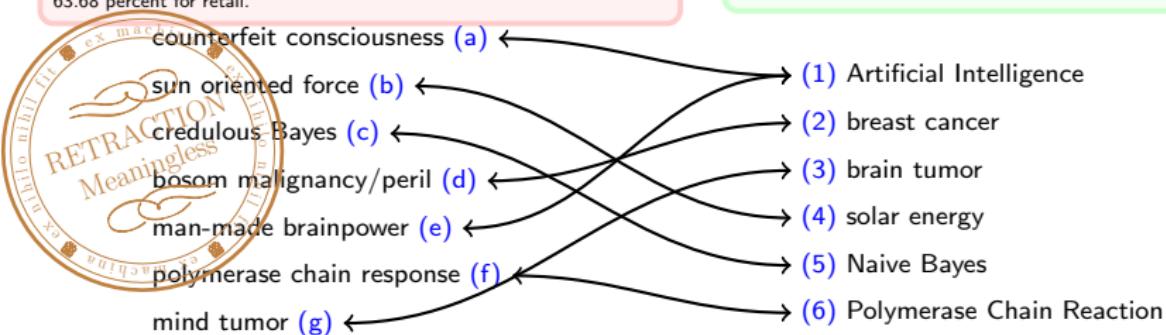
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] –
GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, contrasted with the overall mumble, has been demonstrated to be moderately effective. The exactness rate for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have shown that sector based tweet classification is more successful compared to general tweets. Accuracy rates for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.



Automatic Generation using ChatGPT (GPT-3)

Co-founder and CEO of OpenAI, Sam Altman:



Sam Altman

@sama

...

ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness.

it's a mistake to be relying on it for anything important right now. it's a preview of progress; we have lots of work to do on robustness and truthfulness.

1:11 AM · Dec 11, 2022

Automatic Generation using ChatGPT (GPT-3)

Chat bot:

<https://bigthink.com/the-future/chatgpt-microbiology-quiz-aced/>

Q10. A scientist wants to test the efficacy of a new antibiotic. He infects ten rats with *Salmonella* and administers the antibiotic to all ten rats. All of them recover. What is the flaw in this experiment?

A10. One potential flaw in this experiment is that the scientist only used ten rats. This is a relatively small sample size, and it is possible that the results of the experiment are not representative of the population as a whole. It is generally recommended to use a larger sample size in experiments in order to increase the reliability and validity of the results. Additionally, the experiment should include a control group of rats that do not receive the antibiotic in order to compare the effects of the antibiotic with the effects of no treatment.

Automatic Generation using ChatGPT (GPT-3)

Chat bot:

<https://bigthink.com/the-future/chatgpt-microbiology-quiz-aced/>

Q10. A scientist wants to test the efficacy of a new antibiotic. He infects ten rats with *Salmonella* and administers the antibiotic to all ten rats. All of them recover. What is the flaw in this experiment?

A10. One potential flaw in this experiment is that the scientist only used ten rats. This is a relatively small sample size, and it is possible that the results of the experiment are not representative of the population as a whole. It is generally recommended to use a larger sample size in experiments in order to increase the reliability and validity of the results. Additionally, the experiment should include a control group of rats that do not receive the antibiotic in order to compare the effects of the antibiotic with the effects of no treatment.

Grade:

Correct. (Though, I don't think the comment about the rats being insufficiently representative of the population is relevant here — this isn't an epidemiology study.)

ChatGPT (GPT-3) – failure vs success

Wrong explanations:

Human: I have 47 pennies. I throw away 44 of these pennies and divide the remaining pennies into three groups. How many pennies are in each group?

AI: There would be 3 pennies in each group.

Human: Are you sure?

AI: Yes, I am sure. The result of 44 divided by 3 is 14.4, so the remainder after accounting for 44 pennies would be 3. Therefore, each group of pennies will have 3 pennies.

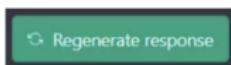
Failure List:

<https://github.com/giuvan95/chatgpt-failures>

PubPeer post about ChatGPT (GPT-3)

#1 Guillaume Cabanac commented January 2024

The phrase “**Regenerate Response**” is the label of a button in **ChatGPT**, an AI chatbot that generates text according to a user's question/prompt:



This preprint contains the unexpected phrase “**Regenerate Response**”:

Image Classification using Different Machine Learning Techniques

Simegnew Yihunie Alaba, *Member, IEEE*
Mississippi State University
Department of Electrical and Computer Engineering

Abstract—Artificial Neural Networks and Convolutional Neural Networks have become common tools for classification and object detection, owing to their ability to learn features without prior knowledge. During training, these networks learn the parameters, weights, and biases. This paper proposes a simple Neural Network and Convolutional Neural Network (CNN) for a classification task. Furthermore, the Bayesian neural network work is reproduced as a baseline for comparing my proposed networks. All experiments were conducted using the MNIST dataset. While the simple neural networks and the convolutional networks adjust their parameters based on the cost function during training, the Bayesian convolutional neural network updates its parameters based on the backdrop that drives a variational approximation to the true posterior. Hyperparameters such as optimizer, learning rate, regularizers, dropout, epochs, etc., were varied to train the two proposed networks. The proposed networks achieved better classification accuracy, approximately 99%, than the previously implemented Bayesian convolutional neural network. However, it is difficult to predict the certainty of the predictions made by my proposed networks, unlike Bayesian learning, which makes it easy to do so. You can find the code for this work at.

[Regenerate response](#)

optimizer were compared for the classification task, and cross-entropy loss was used as the loss function for all training. The Rectilinear Linear Unit (ReLU) was also applied as the activation function.

[Regenerate response](#)

II. RELATED WORK

There have been various studies on the applications of neural networks, CNNs, and Bayesian methods. Sham *et al.* [3] improved the classification error rate of the MNIST dataset [4] by using an ensemble method with heterogeneous deep network fusion based on the degree of certainty aggregation method. Agara [5] presented a classification task using CNN and support vector machine (SVM) as a classifier for the MNIST and fashion-MNIST datasets [6]. Another work in [7] improved the MNIST classification by using different nonlinear activation functions and comparing the results. Lukas [8] used an adversarial neural network for a classification task on the MNIST dataset. Bayesian learning has also been explored,

Table of Contents

- 1 Of Publications and Knowledge
- 2 Different kinds of venue
 - Accounting Units
 - Scientometrics: what for?
- 3 Let's read!
- 4 Fake publications generation with Context Free Grammar
 - SClgen a Probabilistic Context Free Grammar
 - h-index hacking, CV Padding, Journal Hijacking
 - Automatic detection
- 5 New Tools based on new language models
 - Paraphrasing Tools, Transformer based generation
 - Paraphrasing Tools
 - Detection of Tortured Phrases
 - Automatic Generation using ChatGPT (GPT-3)
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

Seek&Blastn screen for obvious errors:

[Labbé et al., 2019]

PMID : 25262828

Materials and methods

The shRNA sequence (5'-GCGGAGGGTTTGAAAGAATATCTC-GAGATATTCTTCAACCCCTCCGCTTTTT-3') targeting TPD52L2 (NM_199360) was inserted into the pFH-L plasmid (Shanghai Hollybio, China). A scrambled shRNA that shared no homology with the mammalian genome (5'-CTAGCCCCGCCAAG-GAAGTGCAATTGCATACTCGAGTATGCAATTGCACTTC-CTTGGTTTTGTAAAT-3') was used as control.

Fact-Check using blastn (NCBI)

SeqQueryValue = 10

Length=54

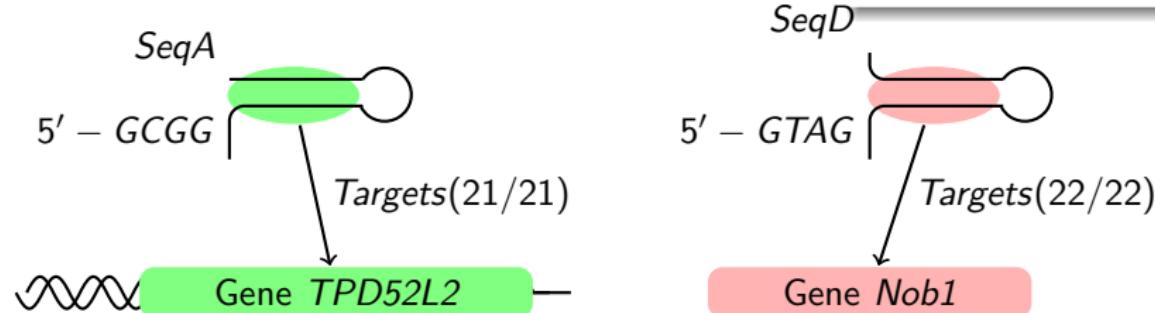
Sequence: GCGGAGGGTTTGAAAGAATATCTC-GAGATATTCTTCAACCCCTCCGCTTTTT

Momo.sapiens tumor protein D52 like 2 (TPD52L2), ...
Length=2230Query 1 GCGGAGGGTTTGAAAGAATAT 21
|||||||

Sbjct 894 GCGGAGGGTTTGAAAGAATAT 914

Query 28 ATATTCTTCAACCCCTCCG 48
|||||||

Sbjct 914 ATATTCTTCAACCCCTCCG 894



Seek&Blastn screen for **obvious errors**:

[Labbé et al., 2019]

PMID : 25262828

Materials and methods

The shRNA sequence (5'-GCGGAGGGTTTCAAAGAATATCTC-GAGATATTCTTCAAACCCCTCGCTTTTT-3') targeting TPD52L2 (NM_199360) was inserted into the pFH-L plasmid (Shanghai Hollybio, China). A scrambled shRNA that shared no homology with the mammalian genome (5'-CTAGCCCAGCCAAG-GAACGTCAATTGCATACTCGAGTATGCAATTGCACTTC-CTTGGTTTTTTGTTAAT-3') was used as control.

Fact-Check using *blastn* (NCBI)

```
S@uDry(value = 10)
```

Length=68

Ségnéfíteanþráðingimngts:

Homo.sapiens NIN1/PSMD8 binding protein 1 homolog (NOB1)...

Length=1775

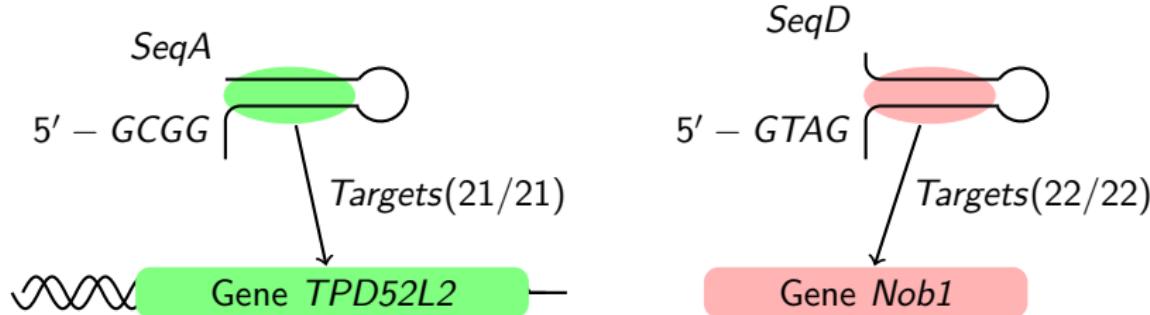
Query 9 GCCAAGGAAGTGCAATTGCATA 30

Sbjct 15

Query 37 TATGCAATTGCACTTCCCTTGG 5'

Sheet 15

1990-1991
1991-1992
1992-1993
1993-1994
1994-1995
1995-1996
1996-1997
1997-1998
1998-1999
1999-2000
2000-2001
2001-2002
2002-2003
2003-2004
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
2014-2015
2015-2016
2016-2017
2017-2018
2018-2019
2019-2020
2020-2021
2021-2022
2022-2023
2023-2024



Seek&Blastn screen for obvious errors:

[Labbé et al., 2019]

PMID : 25262828

Materials and methods

The shRNA sequence (5'-GCGGAGGGTTGAAAGAATATCTC
GAGATATTCTTCAAACCCCTCCGCTT1 TT1-3') targeting
TPD52L2 (NM_199360) was inserted into the pFH-L plasmid
(Shanghai Hollybio, China). A scrambled shRNA that shared no
homology with the mammalian genome (5'-CTAGCCCGCCAAG-
GAAGTGCATTGCATACTCGAGTATGCAATTGCACTTC-
CTTGGTTTTGTAAAT-3') was used as control.

Fact-Check using *blastn* (NCBI)

SugByfevalue = 10)

Length=6

Ségnéficeanprædigngngnts:

Homo.sapiens NIN1/PSMD8 binding protein 1 homolog (NOB1)...

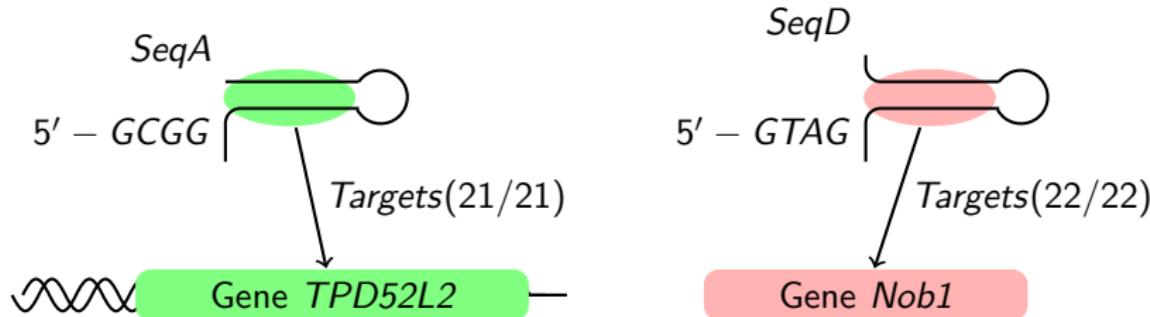
Length=1775

Query 9 GCCAAGGAAGTGCAATTGCATA 30

Sbjct 1505 GCCAAGGAAGTGCAATTGCATA 1526

Fig. 1. 1990. EMISSIONS FROM THE U.S.

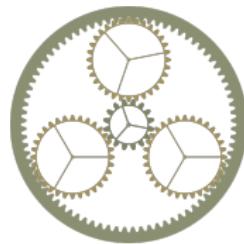
Digitized by srujanika@gmail.com



Seek & Blastn at a glance

Materials and methods
 The shRNA sequence (5'-GCGGAGGGTTGAAA-GAATATCTGAGATATTCTTCAAACCCCTCCGTTTTT-3') targeting TPD52L2 (NM_199360) was inserted into the pFH-L plasmid (Shanghai Hollybio, China). A scrambled shRNA that shared no homology with the mammalian genome (5'-CTAGCCCAGCCAAGGAAGTG-CAATTGCAACTCGAGTATGCAATTGCACTTCCTTG-GTTTTTTGTTAAT-3') was used as control.

(1) Facts extraction



Facts to check

Status	DNA Seq
...	...
Targeting	GCG...TTT
Non-Targ.	CTA...AAT
...	...

(2) Blatn call

Hit lists (Blastn results)

hit list	DNA Seq
...	...
TPD52L2, ...	GCG...TTT
NOB1, ...	CTA...AAT
...	...

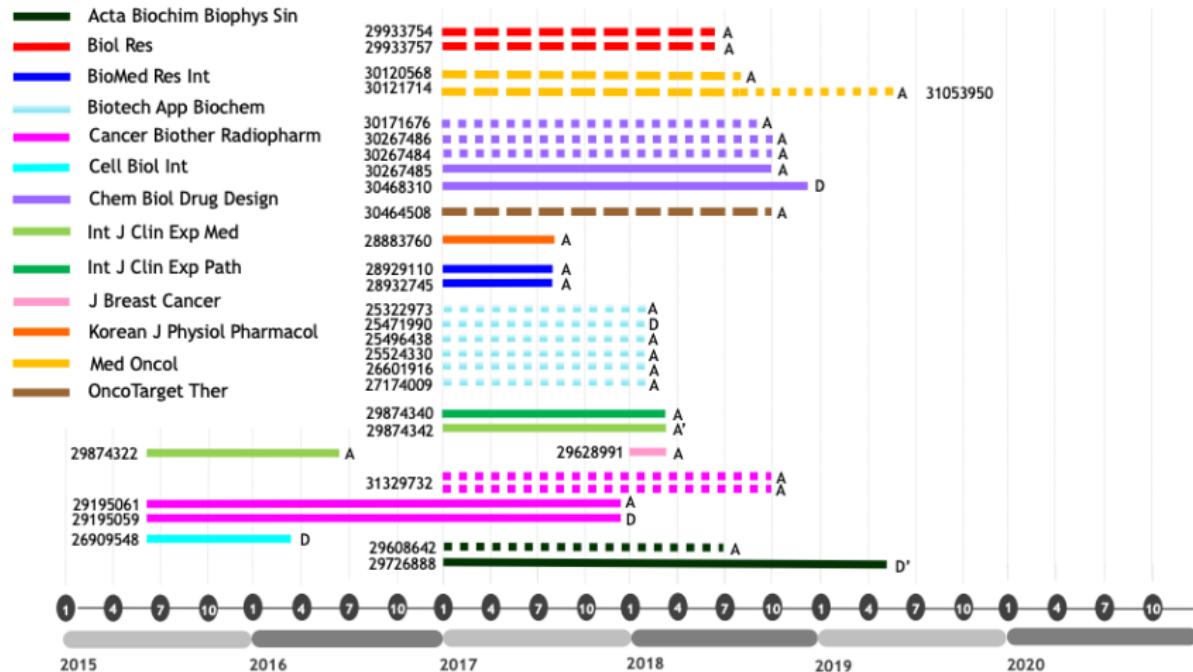
Checked Facts

Status	DNA Seq
Targ.	GCG...TTT
Non-Targ.	CTA...AAT
...	...

(3) Comparison

Papers Retraction and Correction

[?]



Published retractions solid lines (14), expressions of concern (5) broken lines and author corrections (7) are represented by dotted lines. Decisions to take no action shaded dotted lines (6).

Problematic Paper Screener

[Cabanac et al., 2021]

Feedback nobody ▾

Home

Media Coverage

Fingerprints

All Problematic Papers

Detectors:

- 1 Tortured
- 2 SCIGen
- 3 Mathten
- 4 SBIR
- 5 Suspect
- 6 Seek&Blastn
- 7 Problematic Cell Lines
- 8 Journal Cases

Stable URL:
<https://www.irit.fr/~Guillaume.Cabanac/problematic-paper-screener>

This website shows reports the daily screening of papers (partly) generated with:

- ▶ Automatic SBIR Proposal Generator
- ▶ Dada Engine
- ▶ Mathten
- ▶ SCIGen
- ▶ Tortured phrases 🔥

Harvesting data from these APIs:

- ▶ Dimensions
- ▶ PubPeer

PPS in a nutshell: problem tackled and method used —— video in English or in French.

Problematic papers are reported as PubPeer posts, see the [tortured phrases](#) collection and more.

Last Update: 13/05/2022 00:16

Detectors	Number of problematic papers identified
Tortured	~6,500
Suspect	~2,000
Seek&Blastn	~1,000
Journal Cases	~500
SCIGen	~500
Problematic Cell Lines	~500
Mathten	~500
SBIR	~500

Year	Doctype	Publisher	Venue	Title	Citations	Altmetric	Type	Tortured Phrases	Assessor
2021	proceeding	Institute of Electrical and Electronics Engineers (IEEE)	2021 International Conference on Advances in Computing, Communication, and Control (CAC3)	HVAC Hybrid Control methods for HEE in Buildings: Overview	0	0	9	back spread AND neural + counterfeit neural + enhancement calculation AND optimiz + fluffy rationale + hereditary calculation + profound neural organization + square bandler + warm solice + warmth move	Invitation for human assessors
2019	proceeding	Institute of Electrical and Electronics Engineers (IEEE)	2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RETECT)	Handling 3vs of Big Data Through Swarm Intelligence	1	0	7	computerized reasoning + enhancement calculation AND optimiz + figuring asset + genuine esteemed + hereditary calculation + molecular swarm + wellness	Invitation for human

Let's comment the scientific litterature

Start with the PPS:

Problematic Paper Screener

Post on PubPeer:

- The Online Journal Club
- Einstein Foundation Institutional Award 2024
- \$ 1,000 awards for selected PubPeer comments!

References



Avery, O. T., MacLeod, C. M., and McCarty, M. (1944).

STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III .

Journal of Experimental Medicine, 79(2):137-158.



Cabanac, G. and Labbé, C. (2021).

Prevalence of nonsensical algorithmically generated papers in the scientific literature.

Journal of the Association for Information Science and Technology, 72(12):1461-1476.



Cabanac, G., Labbé, C., and Magazinov, A. (2021).

Tortured phrases: A dubious writing style emerging in science. evidence of critical issues affecting established journals.



Gingras, Y. and Khelfaoui, M. (2021).

L'effet sigaps : la recherche médicale française sous l'emprise de l'évaluation comptable.

Zibsel, 8(1):144-174.



Hirsch, J. E. (2005).

An index to quantify an individual's scientific research output.

Proceedings of the National Academy of Science, 102:16569-16572.



Labbé, C. (2010).

Ike antkare, one of the great stars in the scientific firmament.

International Society for Scientometrics and Informetrics Newsletter, 6(2):48-52.



Labbé, C., Grima, N., Gautier, T., Favier, B., and Byrne, J. A.

(2019).

Semi-automated fact-checking of nucleotide sequence reagents in biomedical research publications: The seek & blastn tool.

PLOS ONE, 14(3):1-22.



Labbé, C. and Labbé, D. (2006).

A tool for literary studies. intertextual distance and tree classification.

Literary and Linguistic Computing, 21(3):311-326.



Labbé, C. and Labbé, D. (2013).

Duplicate and fake publications in the scientific literature: how many scieng papers in computer science?

Scientometrics, 94(1):379-396.



Miller, J. M. and Wang, Y. V. (2022).

Ostrich eggshell beads reveal 50,000-year-old social network in africa.

Nature, 601(7892):234-239.