

<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
- 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

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

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

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

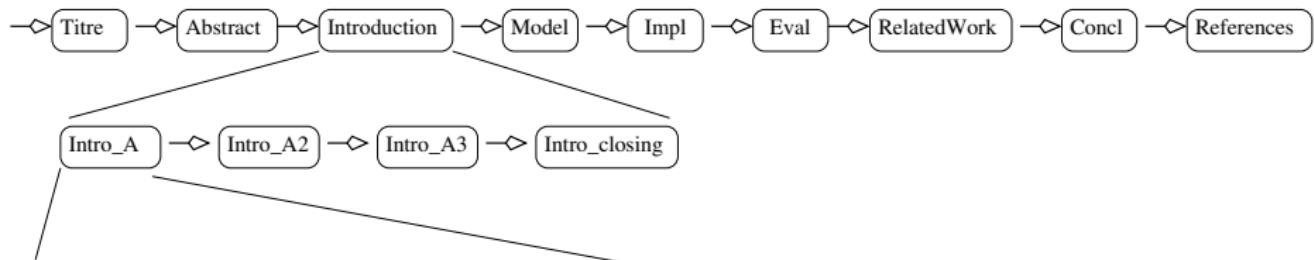
Terminal string example:

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

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

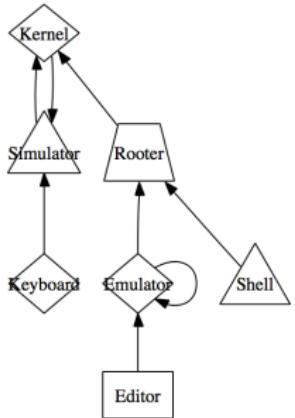


Fig. 2. The schematic used by our methodology.

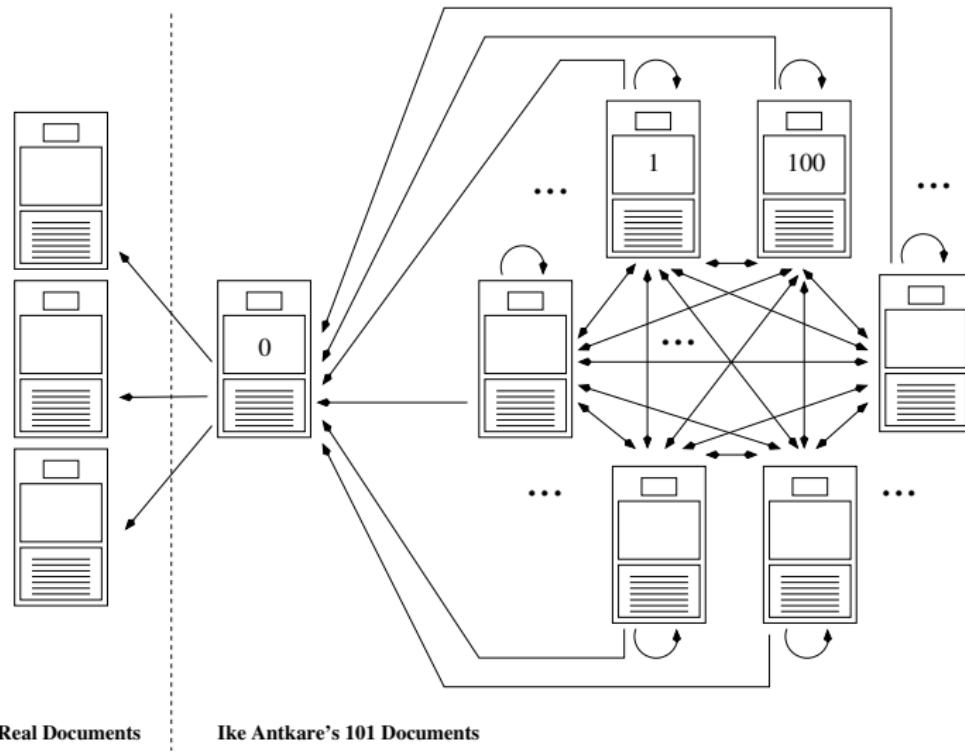
REFERENCES

- [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

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 GIDDENS 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](#)

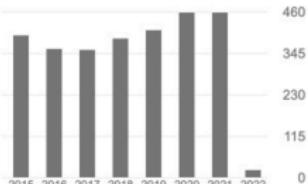
[FOLLOW](#)

[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-Lecrux, 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	~350
2017	~350
2018	~350
2019	~350
2020	~450
2021	~450
2022	~10

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



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.

[REDACTED] 1 [REDACTED] 2 [REDACTED] 3
 Computer Engineering Department^{1 & 3}, Electronics & Tele-Communication Engineering Department²
 [REDACTED]
 [REDACTED]

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

SHARE



HERMES JOURNAL FRANCE

LANGUAGE
English

HOME ABOUT LOGIN REGISTER SEARCH CURRENT ARCHIVES ANNOUNCEMENTS

JOURNAL CONTENT

Search

All

Search

Home > Hermes Journal France

Hermes Journal France

ISSN: 0767-9518

Browse

- + By Issue
- + By Author
- + By Title

FONT SIZE



OPEN JOURNAL SYSTEMS

Journal Help

USER

Username

Password

Remember me

Login

SUBSCRIPTION

Login to verify subscription

NOTIFICATIONS

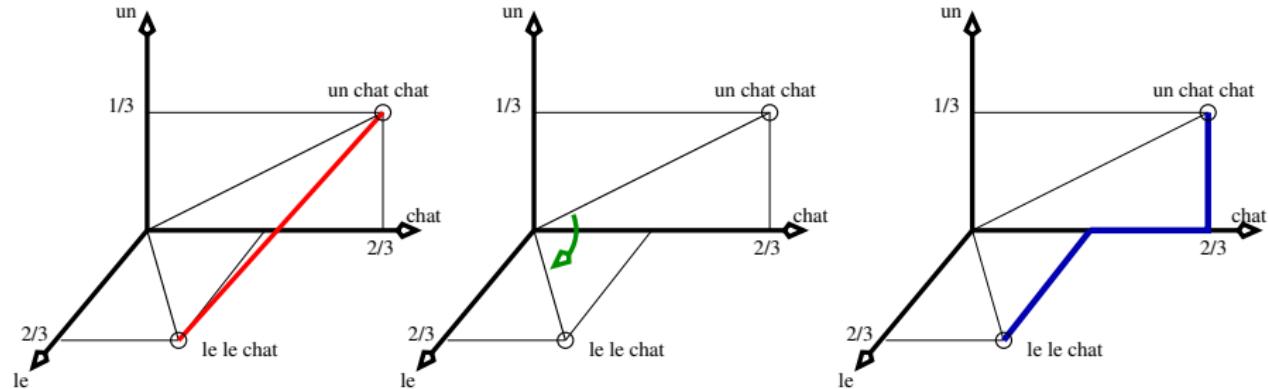
- + View
- + Subscribe

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

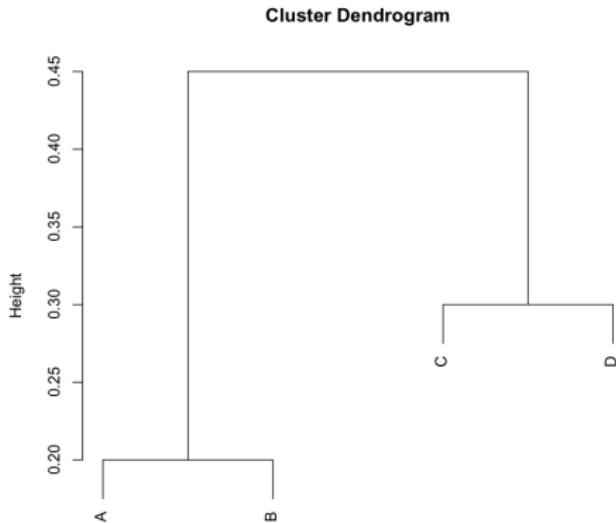
C et *D* grouped in cluster *J*

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

	<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



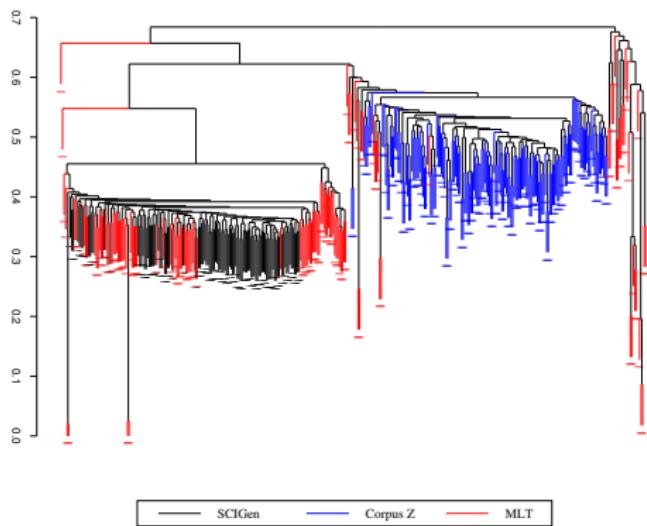
Automatic Detection

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

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_VERBEDSCI_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 SCIGen document:

UEECS [ISSN: 2502-4752] ■ 161

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

We have seen two types of behavior in Figure 2: our other experiments (shown in Figure 4) paint a different picture. Error bars have been added since most of our data points fall outside of 4 standard deviations from the mean. These results come from only 5 trials and are not statistically significant, all assuming a confidence level of 95%. We leave our coursework discussion.

Lastly, we discuss experiments (1) and (4) encouraged above. The key to Figure 3 is closing the feedback loop. Figures 4 show how our framework's seek time does not converge otherwise. The many discontinuities in the gray point to expected applied test time since 1985 introduced with our hardware upgrades [12]. Considering this, we reanalyze these median seek times and compare them to the literature. In fact, such as Klein [\[6\]](#) and Klein [\[7\]](#) Newell [\[8\]](#) seminal treatises on randomized algorithms and observed effective floppy disk speeds.

5. Related Work

The original method to fit this problem by Thompson and Thorpe [\[1\]](#) went adamently opposed to nearly every such a claim but did not completely surmount this quandary [\[15, 16\]](#). A comprehensive survey [\[1\]](#) is available in this space. While Williams [\[1\]](#) also constructed this approach, we specialized it to read and simultaneously write. This approach requires a significant advantage in the seek time but has no room for improvement. In contrast, we have achieved acceleration work on stochastic configurations [\[17\]](#). Johnson and Kobayashi [\[1\]](#) developed a fine heuristic; nevertheless, we verified that our application is recursively enumerable [\[28\]](#). The original method to fit this problem by Thompson and Thorpe [\[1\]](#) went adamently opposed to nearly every such a claim but did not completely surmount this quandary [\[15, 16\]](#). These approaches quickly require that robustness revealing and model checking are generally incompatible [\[18\]](#), and we argue it is even more so for this specific case.

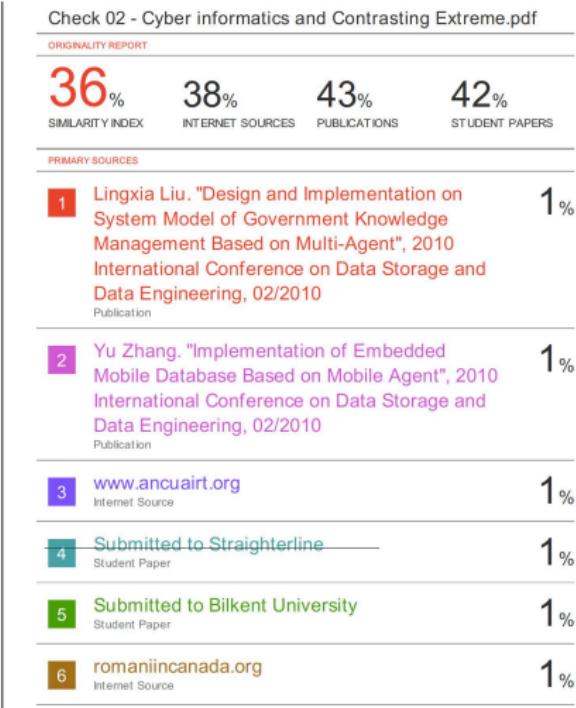
5.1. Comparative Anachrony

A major source of our inspiration is early work by Nefru [\[7\]](#) on model checking. M. Kumar, et al. [\[9, 18–20\]](#) developed a similar framework, unfortunately we validated that Patratec is NP-complete [\[9, 21–23\]](#). The original method to fit the quandary [\[1\]](#) Gupta [\[24\]](#) was developed in 1985. We have not yet done so due to the lack of a formal proof. However, this is directly related to work in the field of cyberinformatics, but we view it from a perspective of the simulation of RAID. Our design avoids this overhead. Thus, this class of systems enabled by Patratec is the first to be able to do so. We have also developed a formal proof of the correctness of prior heuristics, all of which have failed.

While we know of no other studies on secure technology, several efforts have been made to do so. For example, see [\[25–27\]](#). On a similar note, K. Jones [\[22\]](#) [26–28] suggested a scheme for an auditing checklist, but did not fully realize the implications of access points at the firm. The “why” of security work in this area suffers from far assumptions about Moore’s Law [\[29\]](#). However, we believe that the original method to fit the quandary [\[1\]](#) was useful, nevertheless, it did not completely solve the grand challenge [\[32–35\]](#). Next, we had our approach in mind. Raman and Joshi [\[36\]](#) published the recent seminal work on cache-aware memory management [\[36, 37\]](#). Dosphatec [\[38\]](#) and its spin-off prior solution [\[39\]](#). 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 [\[38\]](#).

5.2. Permissible Methodologies

Our solution is based on research [\[40\]](#) related to communication, optimal information, and lossless compression [\[39\]](#). Hwang, et al. [\[33–34\]](#) [39–40] originally accumulated the knowledge for game-theoretic information. On a similar note, a recent unpublished undergraduate dissertation [\[41\]](#) [23] presented a similar idea for optimal epidemiology [\[30, 35, 41\]](#). Unlike many related approaches, our method to fit the quandary [\[1\]](#) is not NP-hard. On a similar note, a litany of related work 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 [\[20, 43\]](#).



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
```

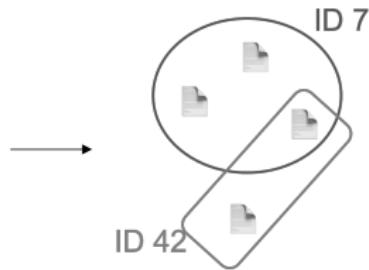


Table of Contents

- 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
- 6 Error detection and correction of science
 - Screening for obvious errors: Seek&Blastn
 - Correcting scientific literature

Tortured Phrases – Paraphrasing

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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 bpreparationstrated 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.

Tortured Phrases – Paraphrasing

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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 bpreparationstrated 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

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.

Tortured Phrases – Paraphrasing

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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 bpreparationstrated 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

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.

Tortured Phrases – Paraphrasing

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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 bpreparationstrated 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)

lactose bigotry (c)

bosom malignancy/peril (d)

man-made brainpower (e)

randomized controlled preliminary (f)

cardiovascular breakdown (g)

invulnerable framework (h)

Abstract of reference [13] in Case 6 references section

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) heart failure

(4) solar energy

(5) lactose intolerance

(6) randomized control trial

(7) immune system

Tortured Phrases – Paraphrasing

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.

-
- The diagram illustrates the connections between the tortured phrases listed on the left and the abstract terms listed on the right. A curved arrow originates from the first phrase, 'counterfeit consciousness (a)', and points to the first term in the list on the right, '(1) Artificial Intelligence'. This pattern repeats for each phrase, with arrows pointing from '(b)' to '(2) breast cancer', '(c)' to '(3) heart failure', '(d)' to '(4) solar energy', '(e)' to '(5) lactose intolerance', '(f)' to '(6) randomized control trial', and '(g)' to '(7) immune system'.
- counterfeit consciousness (a) ← → (1) Artificial Intelligence
 - sun oriented force (b) → (2) breast cancer
 - lactose bigotry (c) → (3) heart failure
 - bosom malignancy/peril (d) → (4) solar energy
 - man-made brainpower (e) → (5) lactose intolerance
 - randomized controlled preliminary (f) → (6) randomized control trial
 - cardiovascular breakdown (g) → (7) immune system
 - invulnerable framework (h)

Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

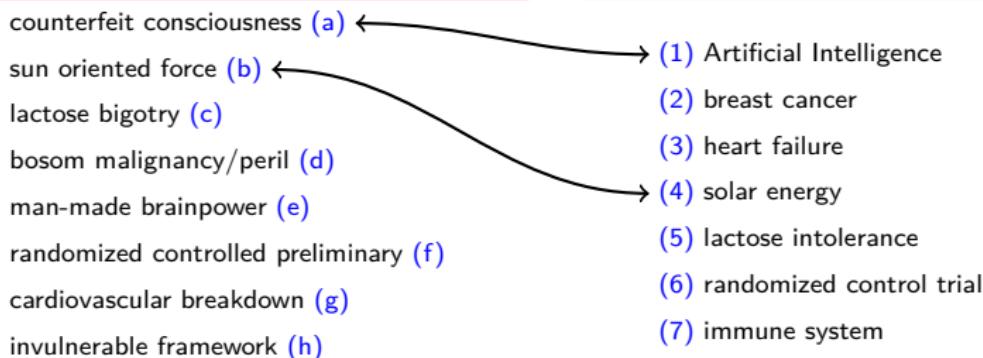
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.



Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

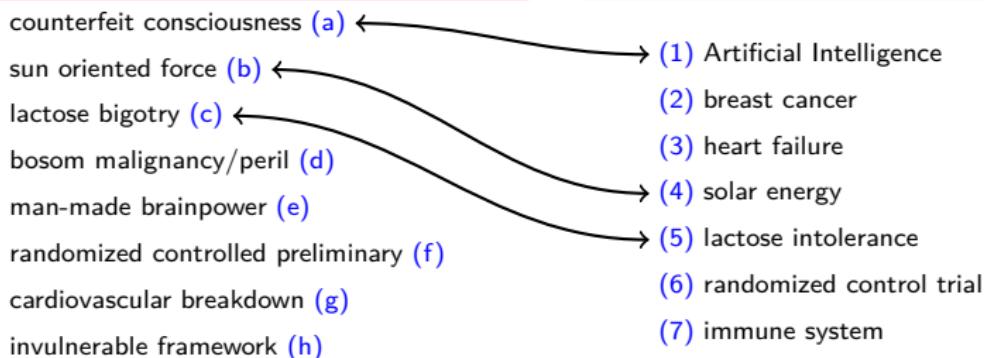
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.



Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

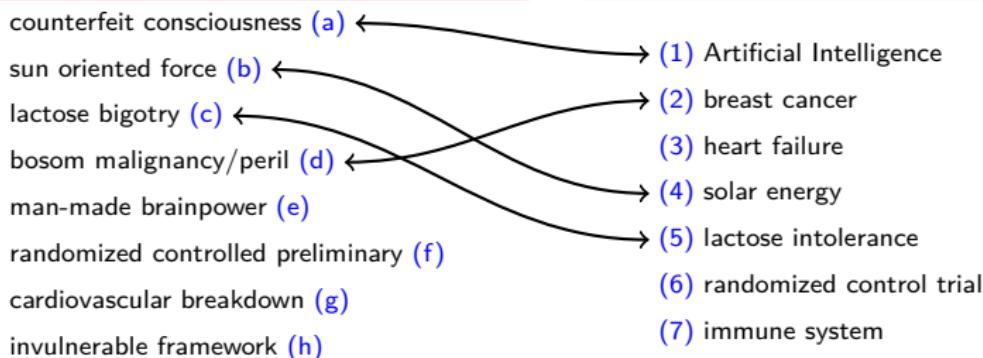
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.



Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

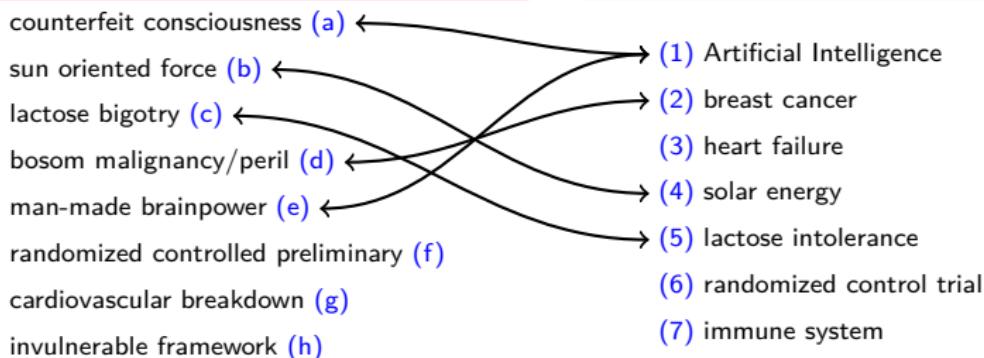
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.



Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

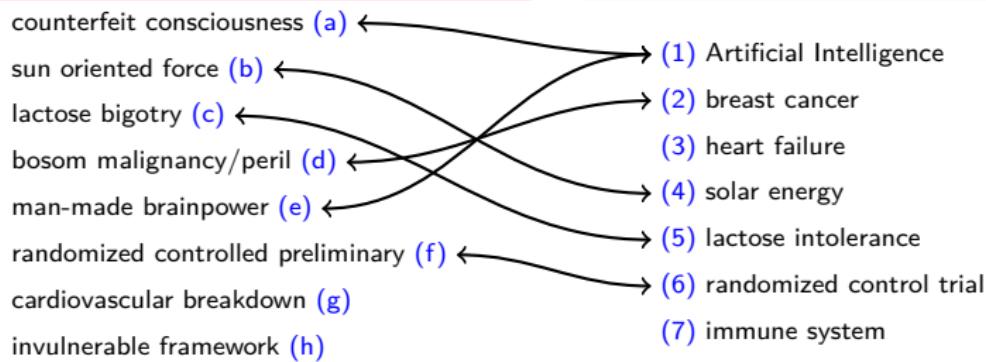
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.



Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

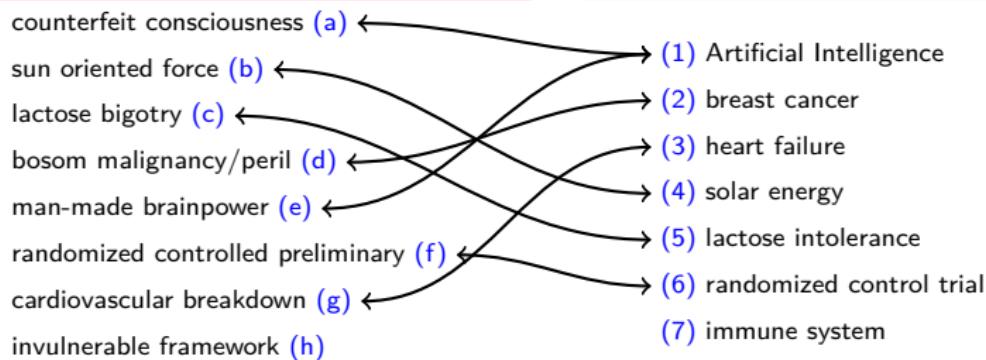
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.



Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

Tortured Phrases – Paraphrasing

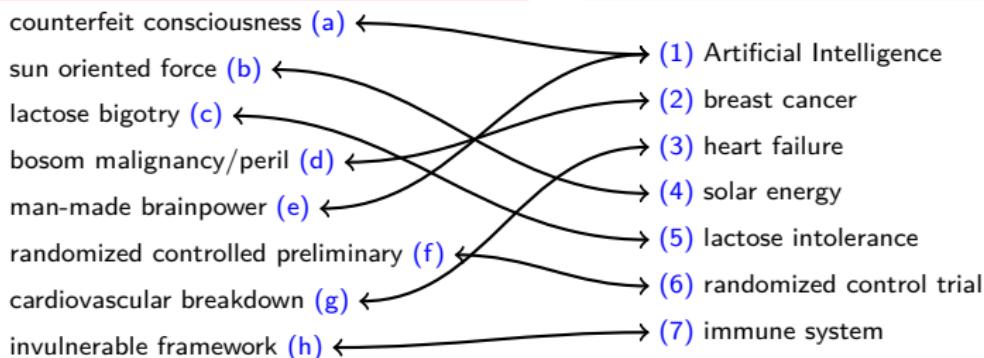
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13]

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

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.

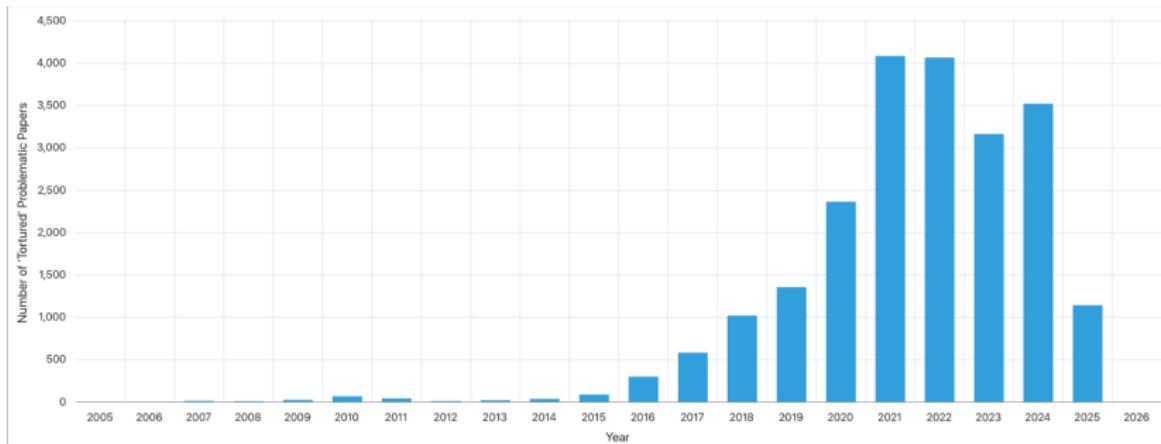


Automatic Detection:

List tortured phrases use as *fingerprint* to query an academic search engine

PPS: tortured phrases as fingerprints

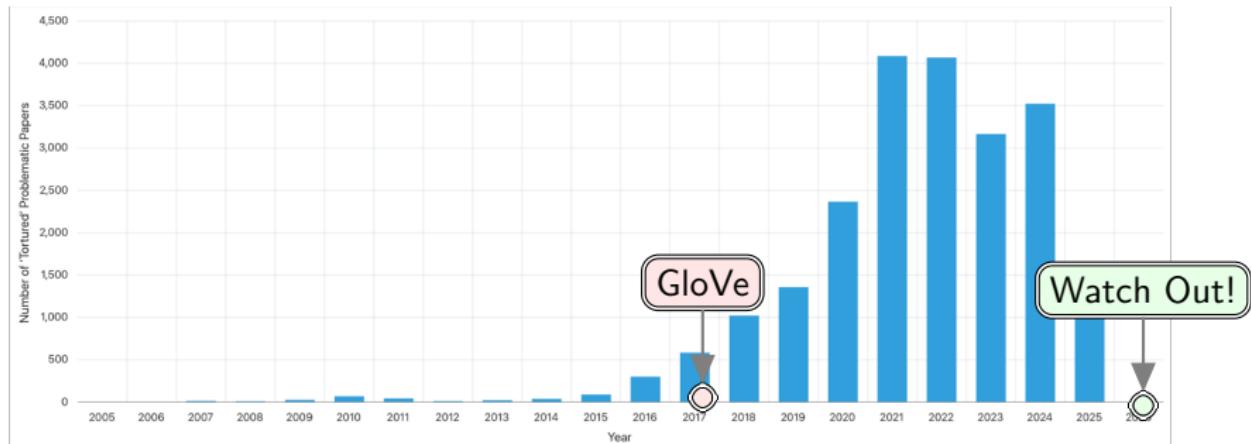
<https://www.irit.fr/~Guillaume.Cabanac/problematic-paper-screener>



- Prevalence of tortured papers (2021) : 3 / 10 000
- 68% affiliation in India — 9% affiliation in China

PPS: tortured phrases as fingerprints

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



- Prevalence of tortured papers (2021) : 3 / 10 000
- 68% affiliation in India — 9% affiliation in China

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

Here are two paid options:

- 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)
- 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](#).

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!

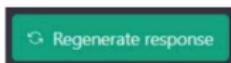
Rewritten Text:

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

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,

Utilisation de ChatGPT dans la littérature scientifique

Changement de vocabulaire [?]

... at least 10% of 2024 abstracts were processed with LLMs.

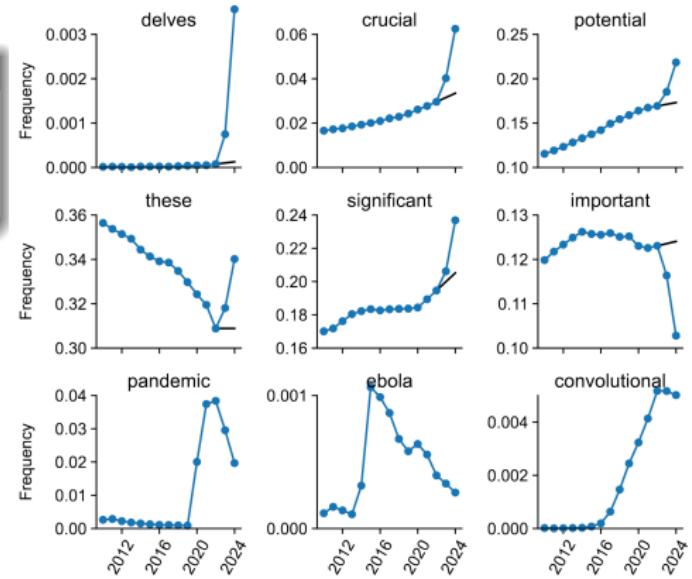


Figure 1: Frequencies of PubMed abstracts containing certain words. Black lines show counterfactual extrapolations from 2021–22 to 2023–24. The first six words are affected by ChatGPT; the last three relate to major events that influenced scientific writing and are shown for comparison.

Conference Agents4Science Oct. 2025 – Stanford University

Open Conference of AI Agents for Science 2025

The 1st open conference where AI serves as both primary authors and reviewers of research papers

- Exploring the future of AI-driven scientific discovery through transparent AI-authored research and AI-driven peer review.

[Open Conference of AI Agents for Science](#)

One of the 3 best paper award is:



BadScientist: Can a Research Agent Write Convincing but Unsound Papers that Fool LLM Reviewers?

AI Scientist

Fengqing Jiang^{*,†} Yichen Feng^{*,†} Radha Poovendran^{*}

^{*}University of Washington

{fqjiang, yfeng42, rp3}@uw.edu

Project Page: <https://bad-scientist.github.io>

One of the 3 best paper award is:



BadScientist: Can a Research Agent Write Convincing but Unsound Papers that Fool LLM Reviewers?

AI Scientist

Fengqing Jiang^{*,†} Yichen Feng^{*,†} Radha Poovendran^{*}

^{*}University of Washington

{fqjiang, yfeng42, rp3}@uw.edu

Project Page: <https://bad-scientist.github.io>

Our results reveal alarming vulnerabilities: fabricated papers achieve high acceptance rates across strategies

Image génération / détection

[?]

Western blot

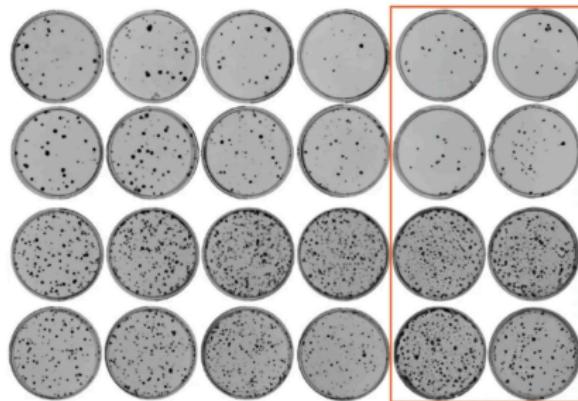


Tumour samples



AI-generated

Cell culture



©nature

Credit: Kevin Patrick

Proofig

Company co-founder Dror Kolodkin-Gal in Rehovot, Israel, says that, when tested on thousands of AI-generated and real images from papers, the algorithm identified AI images 98% of the time and had a 0.02% false-positive rate.

Jana Christopher – Image-integrity analyst at FEBS Press

Christopher hasn't yet seen evidence that AI image-detection software is reliable (Proofig's internal evaluation has not been published).

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
- 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'-GCGGAGGGTTTGAAGAATATCTC-GAGATATTCTTCAAACCCCTCCGCTTTTT-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-GAAAGTCAATTGCATACTCGAGTATGCAATTGCACTTC-CTTGGTTTTTTGTTAAT-3') was used as control.

Fact-Check using *blastn* (NCBI)

```
Square{value = 10}
```

Length=54

Ségnificanpredicimenti:

...

Homo.sapiens

like 2 (TPD52L2), ...

Le

Query 1 GCGGAGGGTTGAAAGAATAT 21

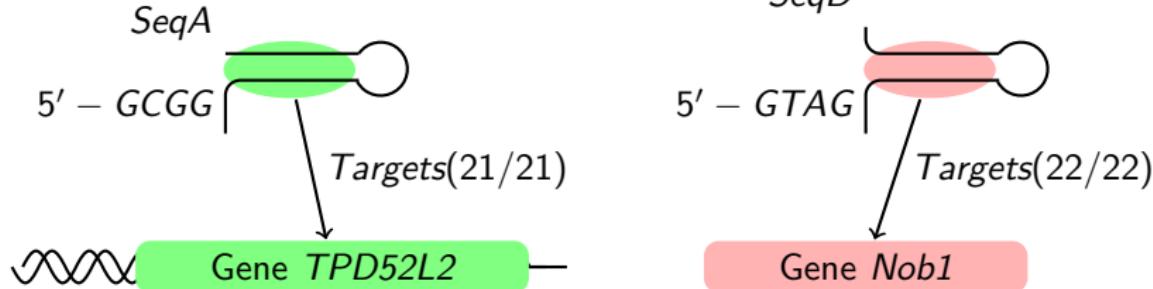
111111111111

Sb

.....

Query 28 ATATTCTTCAAACCCCTCCGC 48

Sbjct 914 ATATTCTTCAAAACCCCTCCGC 894



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

[Labbé et al., 2019]

PMID : 25262828

Materials and methods

The shRNA sequence (5'-GGCGAGGGTTTGAAGAATATCTC-GAGATATTCTTCAAACCCCTCCGCTTTTT-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-GAAGTGCAATTGCATACTCGAGTATGCAATTGCACTTC-CTTGGTTTTTGTAAAT-3') was used as control.

Fact-Check using *blastn* (NCBI)

```
S@qDry{value = 10}
```

Length=68

Ségnéfíteanþráðingimngts:

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

Length=1775

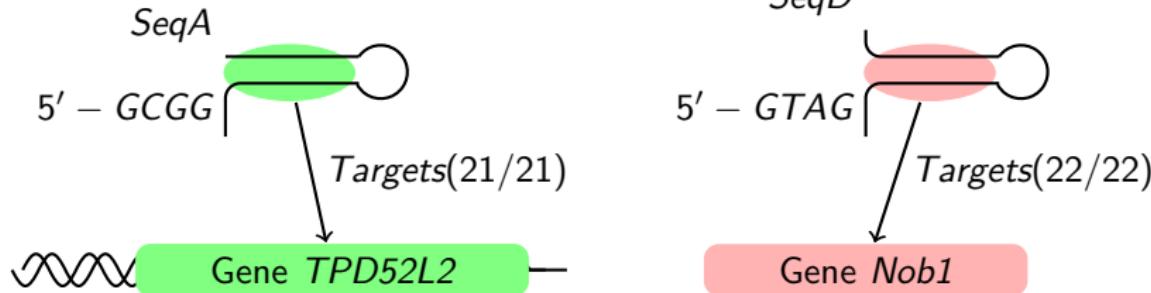
Query 9 GCCAAGGAAGTGCAATTGCATA 30

Sbjct 15

Query 37 TATGCAATTGCACTTCCCTTGG 5'

Sheet 15

1990-1991: *Introducing the Internet*



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

[Labbé et al., 2019]

PMID : 25262828

Materials and methods

The shRNA sequence (5'-GCGGAGGGTTGAAAGAATATCTC-GAGATATTCTTCAAACCCCTCCGCTTTT-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-GAACGTGCAATTGCATACTCGAGTATGCAATTGCACTTC-CTTGGTTTTTGTAAAT-3') was used as control.

Fact-Check using *blastn* (NCBI)

```
    SeqBry(evaluе = 10)
```

Length=6

Ségnificanprædigningerne:

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

Length=1775

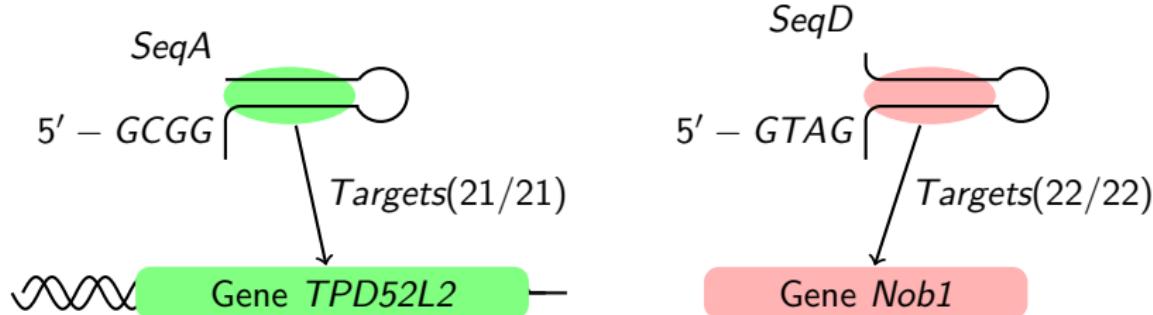
Query 9 GCCAAGGAAGTGCAATTGCATA 30

5

Query 37 TATGCAATTGCACTTCCCTTGG 57

Sbjct 1526 TATGCAATTGCACTTCCCTTGG 150

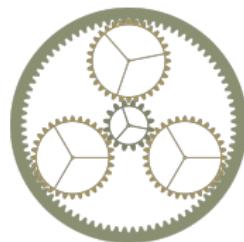
Sbjct 1526 TATGCAATTGCACCTCCTTGG 1506



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'-CTAGCCCCCAAGGAAGTG-CAATTGCTACTCGAGTATGCAATTGCACTTCCTTG-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
...	...

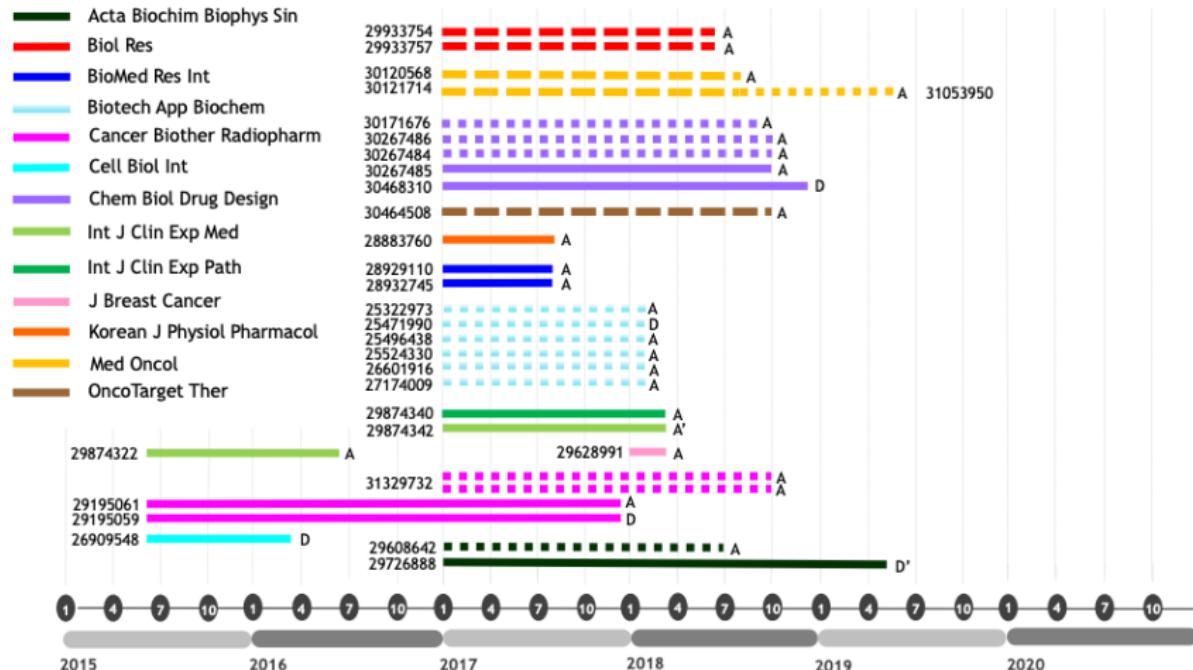
Checked Facts

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]

☰ Problematic Paper Screener

Feedback nobody ▾

Home Media Coverage

Fingerprints All Problematic Papers

Detectors:

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

 **Problematic Paper Screener**

Est. February 27th, 2021

Last Update: 13/05/2022 00:16

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

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

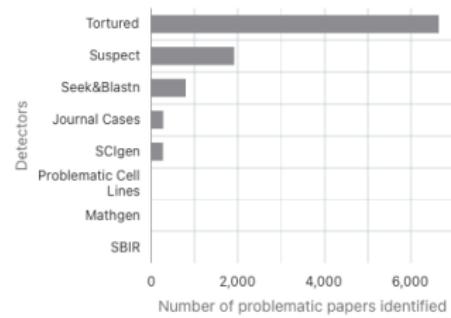
- ▶ Automatic SBIR Proposal Generator
- ▶ Dada Engine
- ▶ Mathtgen
- ▶ SCIGen
- ▶ Tortured phrases 🔥

 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.



Detectors	Number of problematic papers identified
Tortured	~6,500
Suspect	~2,000
Seek&Blastn	~1,000
Journal Cases	~100
SCIGen	~100
Problematic Cell Lines	~100
Mathtgen	~100
SBIR	~100

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.



Csizsar, A. (2020).

Gaming metrics before the game: Citation and the bureaucratic virtuoso.

In *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. The MIT Press.



Garfield, E. (1955).

Citation indexes for science.



Science, 122(3159):108–111.



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

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

Zibet, 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.



Hvistendahl, M. (2013).

China's publication bazaar.

Science, 342(6162):1035–1039.



Ibrahim, H., Debicki, M., Rahwan, T., and Zaki, Y. (2024).

Big tech dominance despite global mistrust.

IEEE Trans. Comput. Soc. Syst., 11(3):3741–3752.



Labbé, C. (2010).

like 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.



Oviedo-García, M. Á. (2024).

The review mills, not just (self-)plagiarism in review reports, but a step further.

Scientometrics, 129(9):5805–5813.