Is This the Real Life? Is This Just Fantasy?
Hunting for Hallucinated References in Paper Submissions

Gianluca Stringhini (Boston University)

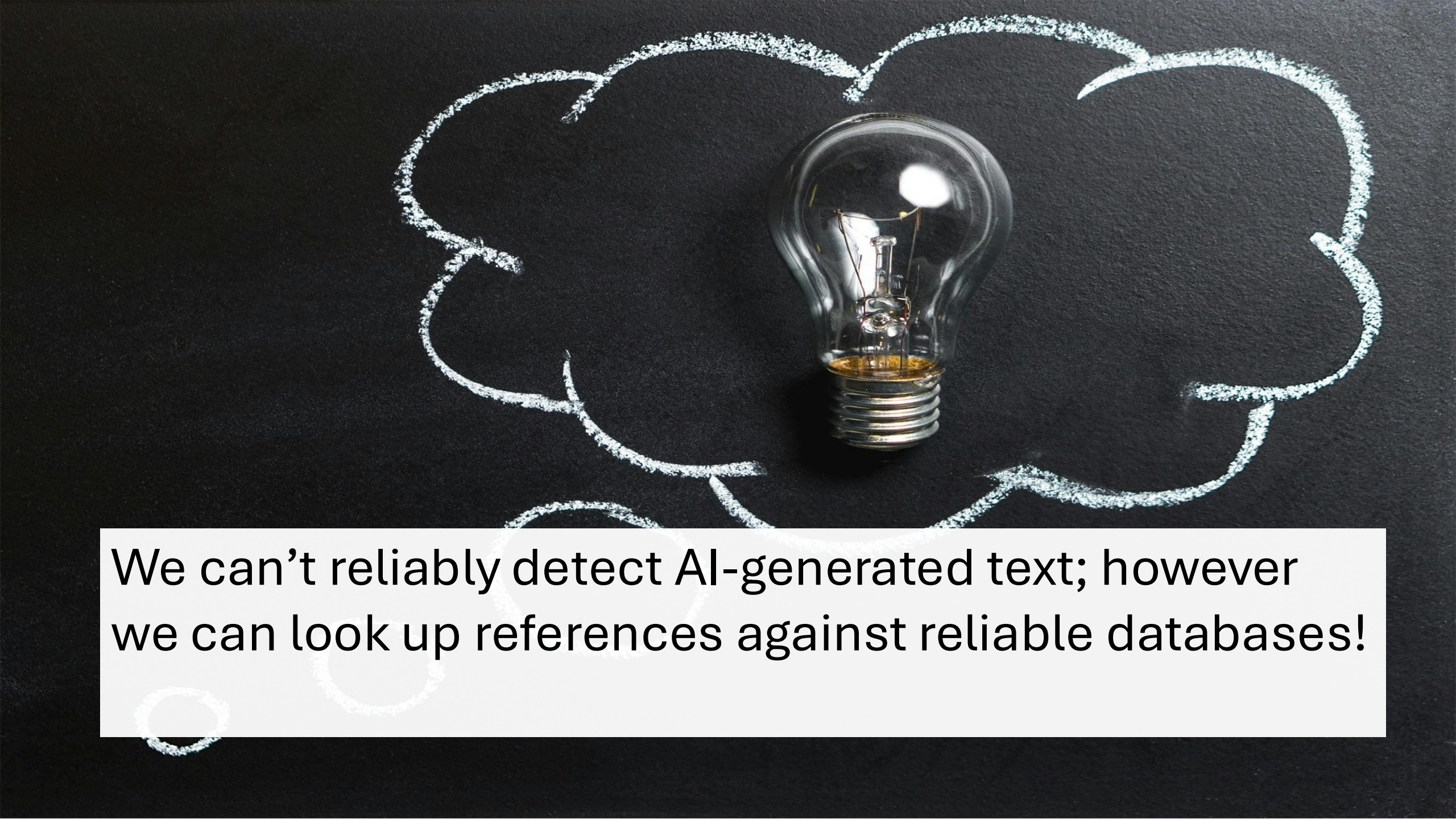
Jeremy Blackburn (Binghamton University)


WE FOUND A PAPER SUBMITTED TO ACSAC 2025 WITH SEVERAL NON-EXISTING REFERENCES!
OH! THAT'S CONCERNING. THANKS FOR FLAGGING IT!
ACSAC 2025 HONOLULU
WELCOME TO HONOLULU
ALOHA!
Submitted to ACSAC 2025
References
1. Smith et al., 2021
2. Doe & Roe, 2022
3. Foo et al., 2023
4. Bar et al., 2024
5. Baz et al., 2025
REVIEWER
CHAIR ACSAC 2025
CAFFEINE & SKEPTICISM
PAPER REVIEW
CONFERENCE CHAIR
ALOHA & CYBERSECURITY

From the ACM CCS Call for Papers

The use of generative AI tools to assist in preparing submissions is permitted, provided that human authors retain full responsibility for the accuracy, originality, and integrity of the work. Generative AI tools cannot be listed as authors. By submitting to ACM CCS, authors affirm that they have critically reviewed all AI-assisted content as if they had written it themselves, including all text, figures, code, experimental data, and citations.

Submissions that contain hallucinated citations (e.g., references to non-existent or fabricated works), falsified or fabricated data, experiments, or results, or other invented claims presented as fact may be desk rejected. Such practices are treated as research misconduct (fabrication and falsification).



We can’t reliably detect AI-generated text; however we can look up references against reliable databases!

Examples of hallucinated references

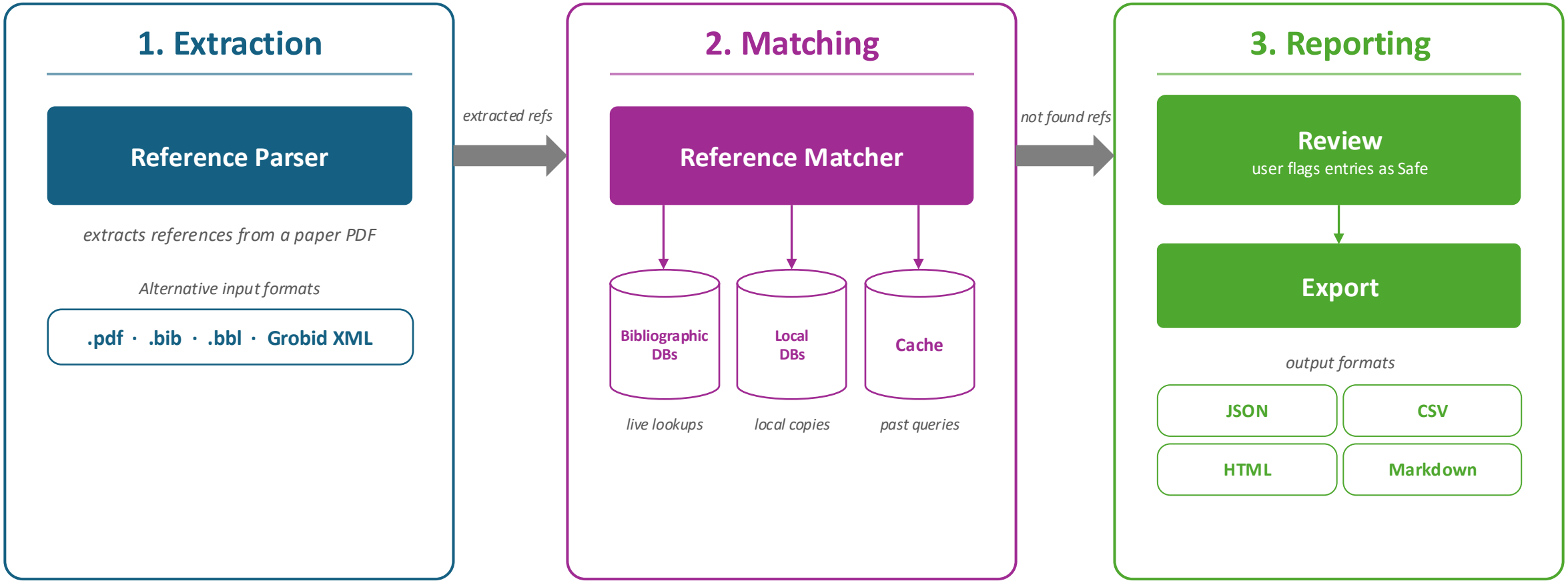
#	Source	Title	Authors	Venue	Year
1	Codex 5.3	The rise of a new type of spammers in Twitter	Metaxas, Panagiotis; Mustafaraj, Eni; Gayo-Avello, Daniel	First Monday 20(6)	2015
2.1	Claude Opus 4.5	Detecting malicious web links and identifying their attack types	Wang, Gang; Mohanlal, Manish; Wilson, Christo; Wang, Xiao; Metzger, Miriam; Zheng, Haitao; Zhao, Ben Y	USENIX Conference on Web application Development	2011
2.2	Google Scholar	Detecting malicious web links and identifying their attack types	Choi, Hyunsang and Zhu, Bin B and Lee, Heejo	USENIX Conference on Web application Development	2011

Goal: develop a tool that can help the chairs of large conferences like Oakland identify hallucinated references

Requirements

- It needs to be easy to use and fast to deploy
- It needs to process thousands of references fast
- It needs to allow chairs to quickly check flagged references
- It needs to produce clear reports

Our tool: **Hallucinator**



Written with the help of AI (Claude Code, ChatGPT), but **no AI in the detection loop**

Database Sources

Database	API Key	Offline DB
ACL Anthology		✓
arXiv		✓
CrossRef		
DBLP		✓
DOI Resolver		
Europe PMC		
GovInfo	✓	
IACR ePrint		✓
Open Library		
OpenAlex	✓	✓
PubMed		
Semantic Scholar	✓	
Web Search + URL liveness check		

Evaluation

Performance analysis: 187 camera ready papers at NDSS 2026

9,777 citations analyzed in 22 minutes (7.4 citations/second)

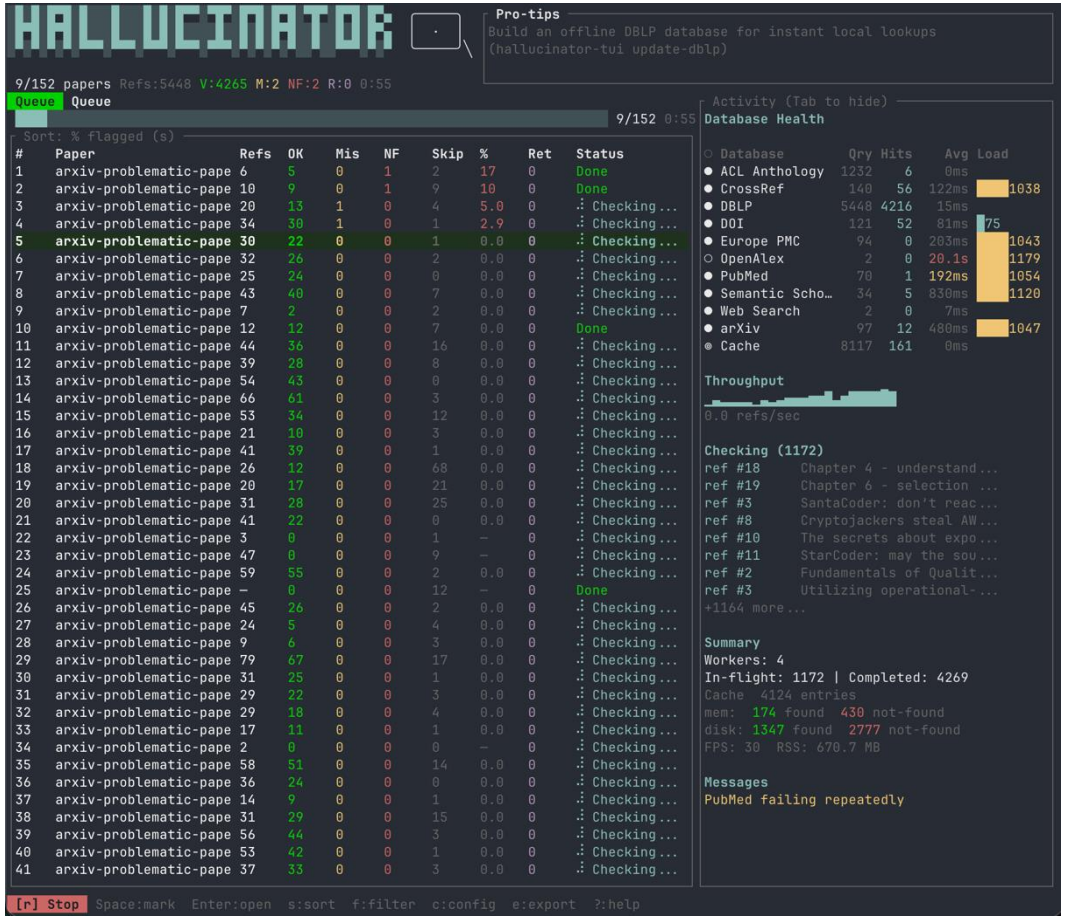
We ran Hallucinator as Program Chairs of **ACSAC 2025** (8 papers with hallucinated references confirmed) and **ICWSM 2025** (11 papers with hallucinated references confirmed)

We tested Hallucinator on the GPTZero NeurIPS dataset, finding 36 additional hallucinated references compared to their blog post

Common sources of false positives

- Parsing errors - we are constantly fixing these
- Entries that are not in our sources
  (e.g., OpenReview)
- Common book titles
  (e.g., “Reinforcement Learning”)
- Companies or organizations listed as authors instead of people

Demo time!


HALLUCINATOR
Pro-tips
Build an offline DBLP database for instant local lookups
(hallucinator-tui update-dblp)
9/152 papers Refs:5448 V:4265 M:2 NF:2 R:0 0:55
Queue Queue
9/152 0:55
Sort: % flagged (s)
Paper Refs OK Mis NF Skip % Ret Status
1 arxiv-problematic-pape 6 5 0 1 2 17 0 Done
2 arxiv-problematic-pape 10 9 0 1 9 10 0 Done
3 arxiv-problematic-pape 20 13 1 0 4 5.0 0 Checking...
4 arxiv-problematic-pape 34 30 1 0 1 2.9 0 Checking...
5 arxiv-problematic-pape 30 22 0 0 1 0.0 0 Checking...
6 arxiv-problematic-pape 32 26 0 0 2 0.0 0 Checking...
7 arxiv-problematic-pape 25 24 0 0 0 0.0 0 Checking...
8 arxiv-problematic-pape 43 40 0 0 7 0.0 0 Checking...
9 arxiv-problematic-pape 7 2 0 0 2 0.0 0 Checking...
10 arxiv-problematic-pape 12 12 0 0 7 0.0 0 Done
11 arxiv-problematic-pape 44 36 0 0 16 0.0 0 Checking...
12 arxiv-problematic-pape 39 28 0 0 8 0.0 0 Checking...
13 arxiv-problematic-pape 54 43 0 0 0 0.0 0 Checking...
14 arxiv-problematic-pape 66 61 0 0 3 0.0 0 Checking...
15 arxiv-problematic-pape 53 34 0 0 12 0.0 0 Checking...
16 arxiv-problematic-pape 21 10 0 0 3 0.0 0 Checking...
17 arxiv-problematic-pape 41 39 0 0 1 0.0 0 Checking...
18 arxiv-problematic-pape 26 12 0 0 68 0.0 0 Checking...
19 arxiv-problematic-pape 20 17 0 0 21 0.0 0 Checking...
20 arxiv-problematic-pape 31 28 0 0 25 0.0 0 Checking...
21 arxiv-problematic-pape 41 22 0 0 0 0.0 0 Checking...
22 arxiv-problematic-pape 3 0 0 0 1 – 0 Checking...
23 arxiv-problematic-pape 47 0 0 0 9 – 0 Checking...
24 arxiv-problematic-pape 59 55 0 0 2 0.0 0 Checking...
25 arxiv-problematic-pape – 0 0 0 12 – 0 Done
26 arxiv-problematic-pape 45 26 0 0 2 0.0 0 Checking...
27 arxiv-problematic-pape 24 5 0 0 4 0.0 0 Checking...
28 arxiv-problematic-pape 9 6 0 0 3 0.0 0 Checking...
29 arxiv-problematic-pape 79 67 0 0 17 0.0 0 Checking...
30 arxiv-problematic-pape 31 25 0 0 1 0.0 0 Checking...
31 arxiv-problematic-pape 29 22 0 0 3 0.0 0 Checking...
32 arxiv-problematic-pape 29 18 0 0 4 0.0 0 Checking...
33 arxiv-problematic-pape 17 11 0 0 1 0.0 0 Checking...
34 arxiv-problematic-pape 2 0 0 0 0 – 0 Checking...
35 arxiv-problematic-pape 58 51 0 0 14 0.0 0 Checking...
36 arxiv-problematic-pape 36 24 0 0 0 0.0 0 Checking...
37 arxiv-problematic-pape 14 9 0 0 1 0.0 0 Checking...
38 arxiv-problematic-pape 31 29 0 0 15 0.0 0 Checking...
39 arxiv-problematic-pape 56 44 0 0 3 0.0 0 Checking...
40 arxiv-problematic-pape 53 42 0 0 1 0.0 0 Checking...
41 arxiv-problematic-pape 37 33 0 0 3 0.0 0 Checking...
Activity (Tab to hide)
Database Health
Database Qry Hits Avg Load
ACL Anthology 1232 6 0ms
CrossRef 140 56 122ms 1038
DBLP 5448 4216 15ms
DOI 121 52 81ms 75
Europe PMC 94 0 203ms 1043
OpenAlex 2 0 20.1s 1179
PubMed 70 1 192ms 1054
Semantic Scho… 34 5 830ms 1120
Web Search 2 0 7ms
arXiv 97 12 480ms 1047
Cache 8117 161 0ms
Throughput
0.0 refs/sec
Checking (1172)
ref #18 Chapter 4 - understand...
ref #19 Chapter 6 - selection ...
ref #3 SantaCoder: don't reac...
ref #8 Cryptojackers steal AW...
ref #10 The secrets about expo...
ref #11 StarCoder: may the sou...
ref #2 Fundamentals of Qualit...
ref #3 Utilizing operational-...
+1164 more...
Summary
Workers: 4
In-flight: 1172 | Completed: 4269
Cache 4124 entries
mem: 174 found 430 not-found
disk: 1347 found 2777 not-found
FPS: 30 RSS: 670.7 MB
Messages
PubMed failing repeatedly
[r] Stop Space:mark Enter:open s:sort f:filter c:config e:export ?:help

The future

As models get better, they might get bibliographic information right?

What if references exist, but what the paper says about them is fabricated?

Let’s discuss!

What should we do about hallucinated citations?

- Simply fixing the references
- Desk rejects
- More serious misconduct investigations
- …

Hallucinator is open source at https://github.com/gianlucasb/hallucinator

Please contribute and report issues that you encounter!