Cartography: AI Slop Detection

This cartography analyses **AI slop** – low-effort, undesirable AI-generated content that clutters the information ecosystem. The project aims define this emerging concept, build a dataset, and develop a detection system. The outcome will support both an academic publication and a Python library.

Table of contents:

Supplementary materials:

What is Slop

See Appendix A for references on slop.

See Appendix B for sources of slop for the creation of a dataset.

AI slop is a nascent term without a clear academic definition, though it is gaining currency online. No peer-reviewed research directly addresses it; only two preprints (Chakrabarty et al., 2025; Krincewicz et al., 2025) mention it with limited analysis. This project aims to provide conceptual clarity.

The prevailing metaphor for slop is pollution — a form of informational debris that, while not necessarily false, overwhelms meaningful discourse. It is linked to spam (computer science), bullshit (philosophy), and broader forms of information disorder (Wardle & Derakhshan, 2017). Unlike misinformation, slop is not always factually wrong but is often perceived as unauthentic or vacuous.

AI slop thrives in environments where plausibility is prioritized over accuracy and volume trumps quality. But the phenomenon is not new; humans have long generated meaningless content for profit: content farms, paper mills, clickbait, fake reviews, SEO traps, and more.

Examples of slop

- **"Shrimp Jesus"**: surreal AI-generated Facebook images spreading for engagement (Placido, 2024)
- **"Vegetative electron microscopy"**: nonsensical term in fraudulent scientific papers (Joelving, 2025)
- **Clarkesworld magazine**: flooded with AI-generated stories, halting submissions (Clarke, 2024)
- **Fake Amazon products**: listings with AI errors in titles (Tangermann, 2025)

Vulnerable Systems

Submission-based systems are especially at risk:

- Bug reports and open source projects (Purdy, 2025)
- Job/fellowship applications (Hoover, 2025)
- Publishing and literary contests (Clarke, 2023)
- Academic papers and grants (Tran, 2023)

Media and cultural studies

Related concepts:

- **Post-truth**: Facts replaced by beliefs (Oxford Languages, 2016)
- **Enshittification**: Platforms deteriorate to extract value (Doctorow, 2023)
- **Dead Internet Theory**: Bots dominate online content (Muzumdar, 2025)
- **Infodemic**: Crisis-driven information glut (World Health Organization, n.d.)
- **Curation**: Required to manage overload

Interest in fake news and misinformation surged post-2016, often described in terms of infection, contamination and flooding. AI intensifies this, threatening to exacerbate information overload.

People are scared of fake news (Barthel, 2016; Cellan-Jones, 2017; World Economic Forum, 2024).

Ironically, information about misinformation is often misleading. While evidence suggests that fake news are not so problematic (Stockwell, 2024; Allen et al. 2020; Arguedas et al, 2022), organizations tackling this issue are still subject to the pressures that generate misinformation. Vendors of technical solutions and academics alike lean into the perceived risk.

This suggests that the true aim of the initiative is not to "neutralize the poison," but rather to enhance the perception of trust within the information environment. Misinformation should be approached as a complex issue that resists simple technical fixes for deeply rooted social problems (Singh 2024). While tools like detection algorithms may offer some assistance, they should be seen as just one component within a broader interdisciplinary strategy.

Slop operates like disinformation (Illing, 2020; Wardle & Derakhshan, 2017) in spreading confusing, repetitive, and low-value content. Market incentives — from click-driven ad revenue to "publish or perish" academia — fuel the production of slop (Knibbs, 2024; Labbé et al., 2025).

The tension is echoed in AI ethics: AI content is capable not only of being deceptive, but highly persuasive (Chen & Shu, 2024; Costello et al., 2024). Simultaneously, mystifying the technology helps maintain its allure and its perceived threat (Tully et al., 2025; Carpenter, 2024), reinforcing the cycle.

Information integrity

See Appendix C for products and organizations in AI detection, fact checking and scientific integrity.

The tooling ecosystem highlights the difference between customer-facing software which informs interpretation of sources and the automated decision making implementation required by social media and search engines.

- **AI Detectors**: Tools like GPTZero and Pangram exist but are imperfect (Weber-Wulff et al., 2023; Sadasivan et al., 2023). Slop's low effort makes concealment less common.
- **Fake News**: Mostly addressed through source credibility instead of textual analysis (Peters, 2022), which raises concerns about bias (Cameron, 2016).
- **Spam Filters**: Google favors high-quality content regardless of it is machine-generated or not (Google Search Central, 2023; Gomes, 2017).
- **Content Farms**: Identified by traits like AI refusal boilerplate, generic domains, and SEO-driven repackaging (Peters, 2022).
- **Paper Mills**: Targeted using linguistic fingerprints, paraphrase artifacts, and citations to retracted work (Cabanac, 2021).

Philosophy and psychology

See Appendix D and E for foundational texts and psychological scales.

Bullshit is discourse indifferent to truth-value, focusing on impression rather than accuracy (Frankfurt, 2009; Cohen 2004; Easwaran,). This differs from lying (intentional falsehood) and applies well to AI-generated content. AI systems "hallucinate" but are better understood as bullshitting - generating plausible-sounding text without regard for truth (Hicks et al., 2024). This extends to users generating text for others (homework cheating) and potentially "bullshitting themselves" through AI assistance (Prada, 2025).

In psychology, the scientific study of bullshit typically focuses on susceptibility scales and their relationship to related phenomena such as fake news. This approach defines bullshit as *unclarifiable* content, deliberately sidestepping the philosophical emphasis on the speaker's intent—specifically, the lack of concern for truth. Similarly, the notion of "slop" hinges largely on the reader's perception of inauthenticity, and may even be measurable through indicators of writing quality (Chakrabarty et al., 2025). Research on "pseudoprofound bullshit" (Pennycook et al., 2015; Čavojová, 2022) demonstrates that individuals differ in their receptivity to meaningless yet superficially impressive statements. Higher susceptibility is associated with lower analytical thinking, greater reliance on intuition, endorsement of epistemically suspect beliefs, and diminished scientific literacy.

Cognitive psychology approaches epistemically suspect beliefs through the lens of heuristics and biases (Kahneman & Tversky, 1974), conceptual metaphors and frames (Lakoff & Johnson, 2008), and motivated reasoning. These same cognitive mechanisms may help explain susceptibility to slop, particularly in relation to phenomena such as the illusory truth effect (where repeated exposure increases perceived truthfulness), confirmation bias (the tendency to favor information that confirms existing beliefs), and information overload (where emotionally charged content captures limited attention resources).

This body of literature frames bias as the exploitation of vulnerabilities in human rationality (Pennycook & Rand, 2019a; Klincewicz et al., 2025; Wardle & Derakhshan, 2017; Muzumdar et al., 2025). However, it often neglects the performative and emotional dimensions of communication—treating misinformation primarily as a problem of information transfer rather than as a socially and affectively embedded practice (Wardle & Derakhshan, 2017). Moreover, this perspective tends to reflect the researchers' own epistemic assumptions and preferred thought-styles, a critique that has also been raised in disciplines such as economics (Infante et al., 2016) and sociology (Barker, 2011).

Dataset and Detection Strategy

See Appendix B for sources of slop for the creation of a dataset.

See Appendix F for archives, datasets and data sources for fake news, rumours, hate speech, nonsense, bullshit, politics, corporate jargon, dadaism, quote aggregators, new age blogs, conspiracy theories, scams, fringe beliefs, thought leaders, etc.

Scope

- Focus on text, not images
- Audience-facing: Not personal use
- Ambiguity: Not outright falsehoods or illegal content
- Detection over intervention

- Asynchronous instead of real-time detection
- Content-level analysis instead of contextual analysis
- Motivated but non-expert users

Data Sources

Slop:

- Retracted scientific papers
- Content farm articles
- AI-generated submissions to magazines
- Papers with known AI fingerprints

Not slop:

- Wikipedia
- Peer-reviewed papers
- Reputable journalism
- Literary classics
- Algorithmic nonsense (Mathgen)
- Authentic rhetoric (politics, advertisement, thought leadership)

Technical considerations

See Appendix E for linguistic and statistical metrics, deep learning interpretability techniques and relevant tooling

NLP classification tasks can be broadly divided into those with objectively verifiable answers—such as fake news detection or identifying machine-generated content—and those that lack clear ground truth, including rumours, hate speech, stereotypes, clickbait, and spam. *Slop* falls into this latter category, which presents significant challenges for data collection and labeling. In such cases, fine-grained, multidimensional labeling schemes may be more appropriate than binary classifications, as seen in datasets like LIAR, NELA, and SARC.

Methods for detecting AI-generated text generally fall into three categories: linguistic pattern analysis, statistical methods, and fact verification (Tang et al., 2023). Other taxonomies make similar distinctions, categorizing approaches by stylistic, complexity-based, and psychological features (Aich et al., 2022). While identifying linguistic patterns in slop is valuable in itself, the highest performance on such detection tasks tends to come from fine-tuned deep learning models (King et al., 2024; Chakrabarty et al., 2025; Chen & Shu, 2024).

However, deep learning models introduce issues of interpretability—particularly problematic in contexts involving algorithmic decision-making and subjective, bias-prone judgments. Evidence suggests that black-box interpretability methods often yield suboptimal or misleading results (Heap et al., 2025; Gonen & Goldberg, 2019).

Current approaches to similar tasks include (Doughman et al., 2024; Wang et al., 2024) prompting LLMs for analysis (Chen & Shu, 2024), machine learning classifiers (LR-GLTR), and fine-tuned deep learning models (RoBERTa-M4). Evaluation includes cross-domain testing, interpretability (SHAP/LIME), and human annotation.

Challenges and Opportunities

Challenges:

- **Subjectivity**: What constitutes slop varies by context and viewer
- **Domain dependence**: Models may not generalize across topics/formats (Fortuna et al., 2022)
- **Adversarial nature**: Producers actively try to avoid detection
- **Annotation difficulty**: Creating reliable ground truth labels

Opportunities:

- **Novel Area**: Computational bullshit detection is understudied
- **Practical Utility**: Educators, editors, and publishers need solutions
- **Interdisciplinary Scope**: Involves computer science, psychology, philosophy, and media studies
- **Real-World Impact**: Potential address real problems

Research Gaps:

- No established slop dataset exists
- Limited computational approaches to bullshit detection
- Need for interpretable, generalizable models

Conclusion

AI slop is a novel form of information pollution that demands dedicated research. Unlike fake news or spam, slop is defined by its low-effort, often meaningless nature. This project seeks to:

This project aims to:

1. Create the first annotated slop dataset
2. Develop robust and interpretable detection methods
3. Establish computational frameworks for bullshit detection

With clear theoretical grounding and practical application, the project aims to shape discourse on AI content quality and offer tools for those curating the digital information landscape. The work addresses urgent needs in academic integrity, content moderation, and information reliability while contributing to computational approaches to philosophical concepts. It can be considered successful if its slop definition contributes to the conversation on AI-generated content quality and the technical artifact generated is adopted by its target audience (publishers, educators, recruiters, or content moderators).

Word Count: This document contains 1800 words (without supplementary material and references).

AI Disclaimer: AI tools (ChatGPT, NotebookLM, ResearchRabbit) were extensively used used for formatting, summarization, and research.

References

- Aich, A., Bhattacharya, S., & Parde, N. (n.d.). *Demystifying Neural Fake News via Linguistic Feature-Based Interpretation*.
- Allen, J., Howland, B., Möbius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances, 6*(14), eaay3539. https://doi.org/10.1126/sciadv.aay3539
- Barker, E. (2011). *Stepping out of the Ivory Tower: A Sociological Engagement in 'The Cult Wars'*. https://journals.sagepub.com/doi/abs/10.4256/mio.2010.0026
- Cabanac, G., & 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. https://doi.org/10.1002/asi.24495
- Cameron, D. (2016, November 18). Viral 'fake news' list deleted after author and students get harassed, doxed. *The Daily Dot*. https://www.dailydot.com/debug/fake-news-facebook-deleted-alt-right/
- Cellan-Jones, R. (2017, September 21). Fake news worries 'are growing' suggests BBC poll. *BBC News*. https://www.bbc.com/news/technology-41319683
- Chakrabarty, T., Laban, P., & Wu, C.-S. (2025). *AI-Slop to AI-Polish? Aligning Language Models through Edit-Based Writing Rewards and Test-time Computation* (No. ArXiv: 2504.07532). ArXiv. https://doi.org/10.48550/arXiv.2504.07532
- Chen, C., & Shu, K. (2024). Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine*, *45*(3), 354–368. https://doi.org/10.1002/aaai.12188
- Chen, C., & Shu, K. (2024). *Can LLM-Generated Misinformation Be Detected?* (No. ArXiv: 2309.13788). ArXiv. https://doi.org/10.48550/arXiv.2309.13788
- Clarke, N. (2023, December 2). *A Concerning Trend*. https://neil-clarke.com/a-concerning-trend/
- Cohen, G. A. (2002). Deeper into Bullshit. In S. Buss & L. Overton (Eds.), *The Contours of Agency* (pp. 321–339). The MIT Press. https:

//doi.org/10.7551/mitpress/2143.003.0015

- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). *Durably reducing conspiracy beliefs through dialogues with AI.* https://doi.org/10.31234/osf.io/xcwdn
- Doctorow, C. (2023). The 'Enshittification' of TikTok. *Wired.* Retrieved 27 May 2025, from https://www.wired.com/story/tiktok-platforms-cory-doctorow/
- Easwaran, K. (n.d.). *Bullshit activities.* Retrieved 29 April 2025, from https://onlinelibrary.wiley.com/doi/10.1111/phib.12328
- Fortuna, P., Dominguez, M., Wanner, L., & Talat, Z. (2022). Directions for NLP Practices Applied to Online Hate Speech Detection. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 11794–11805. https://doi.org/10.18653/v1/2022.emnlp-main.809
- Frankfurt, H. G. (2009). *On Bullshit.* Princeton University Press. https://doi.org/10.1515/9781400826537
- Gomes, B. (2017, April 15). *Our latest quality improvements for Search.* https://blog.google/products/search/our-latest-quality-improvements-search/
- Google Search Central. (2023, February 23). *Google Search's guidance about AI-generated content | Google Search….* Archive. Is. https://archive.is/j1Q3E
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, *26*(2), 38. https://doi.org/10.1007/s10676-024-09775-5
- Hoover, A. (2025, April 2). *Cheating on tech interviews is soaring. Managers don't know what to do.* Business Insider. https://www.businessinsider.com/cheating-tech-interviews-soaring-managers-lost-gen-ai-chatgpt-coding-2025-4
- Illing, S. (2020, January 16). *"Flood the zone with shit": How misinformation overwhelmed our democracy.* Vox. https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation
- Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics. *Journal of Economic Methodology*, *23*(1), 1–25. https://doi.org/10.1080/1350178x.2015.1070527
- Joelving, F. (2025, February 10). As a nonsense phrase of shady provenance makes the rounds, Elsevier defends its use. *Retraction Watch.* https://retractionwatch.com/2025/02/10/vegetative-electron-microscopy-fingerprint-paper-mill/
- Julie Carpenter. (2025, March 4). *If not friend, why friend-shaped?* https://jgcarpenter.com/blog.html?blogPost=if-not-friend-why-friend-shaped-julie-carpenter
- King, T., Wu, Z., Koshiyama, A., Kazim, E., & Treleaven, P. (2024). *HEARTS: A Holistic Framework for Explainable, Sustainable and Robust Text Stereotype Detection* (No. ArXiv: 2409.11579). ArXiv. https://doi.

org/10.48550/arXiv.2409.11579

- Klincewicz, M., Alfano, M., & Fard, A. E. (2025). *Slopaganda: The interaction between propaganda and generative AI* (No. ArXiv: 2503.01560). ArXiv. https://doi.org/10.48550/arXiv.2503.01560
- Knibbs, K. (2024, February 7). Confessions of an AI Clickbait Kingpin. *Wired*. https://www.wired.com/story/confessions-of-an-ai-clickbait-kingpin/
- Labbé, C., Joelving, F., & Cabanac, G. (2025, January 29). *Problematic Paper Screener: Trawling for fraud in the scientific literature*. The Conversation. http://theconversation.com/problematic-paper-screener-trawling-for-fraud-in-the-scientific-literature-246317
- Lakoff, G., & Johnson, M. (2008). *Metaphors We Live By*. University of Chicago Press.
- Michael Barthel, Amy Mitchell, & Jesse Holcomb. (2016, December 15). Many Americans Believe Fake News Is Sowing Confusion. *Pew Research Center*. https://www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/
- Muzumdar, P., Cheemalapati, S., RamiReddy, S. R., Singh, K., Kurian, G., & Muley, A. (2025). The Dead Internet Theory: A Survey on Artificial Interactions and the Future of Social Media. *Asian Journal of Research in Computer Science*, *18*(1), 67–73. https://doi.org/10.9734/ajrcos/2025/v18i1549
- Oxford Languages. (2016). *Oxford Word of the Year 2016*. https://languages.oup.com/word-of-the-year/2016/
- Pennycook, G. & David G. Rand. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(7), 2521–2526. https://doi.org/10.1073/pnas.1806781116
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, *10*(6), 549–563. https://doi.org/10.1017/S1930297500006999
- Peters, D. (2022, August 9). How NewsGuard monitors 8,000 websites to combat misinformation. *Newswhip*. https://www.newswhip.com/2022/08/how-newsguard-monitors-8000-websites-to-combat-misinformation/
- Placido, D. D. (2024, April 28). *Facebook's AI-Generated 'Shrimp Jesus,' Explained*. Forbes. https://www.forbes.com/sites/danidiplacido/2024/04/28/facebooks-surreal-shrimp-jesus-trend-explained/
- Prada, L. (2025, May 6). ChatGPT Is Giving People Extreme Spiritual Delusions. *VICE*. https://www.vice.com/en/article/chatgpt-is-giving-people-extreme-spiritual-delusions/
- Purdy, K. (2025, May 7). *Open source project curl is sick of users submitting "AI slop" vulnerabilities*. Ars Technica. https://arstechnica.com/gadgets/2025/05/open-source-project-curl-is-sick-of-users-submitting-ai-slop-vulnerabilities/
- Ross Arguedas, A., Robertson, C. T., Fletcher, R., & Nielsen, R. K.

(2022). *Echo chambers, filter bubbles, and polarisation: A literature review*. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-ETXJ-7K60

- Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-Generated Text be Reliably Detected?* (No. ArXiv: 2303.11156). ArXiv. http://arxiv.org/abs/2303.11156
- Singh, M. (2024, April 15). Don't Believe What They're Telling You About Misinformation. *The New Yorker*. https://www.newyorker.com/magazine/2024/04/22/dont-believe-what-theyre-telling-you-about-misinformation
- Stockwell, S. (2024, September). *AI-enabled influence operations: Threat analysis of the 2024 UK and European elections* [Briefing paper]. The Alan Turing Institute, Centre for Emerging Technology and Security. https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-threat-analysis-2024-uk-and-european-elections
- Tang, R., Chuang, Y.-N., & Hu, X. (2023). *The Science of Detecting LLM-Generated Texts* (No. ArXiv: 2303.07205). ArXiv. https://doi.org/10.48550/arXiv.2303.07205
- Tran, T. H. (2023, May 6). *How This Doctor Wrote Dozens of Science Papers With ChatGPT*. The Daily Beast. https://www.thedailybeast.com/how-this-doctor-wrote-dozens-of-science-papers-with-chatgpt/
- Tully, S., Longoni, C., & Appel, G. (2025). EXPRESS: Lower Artificial Intelligence Literacy Predicts Greater AI Receptivity. *Journal of Marketing*, 00222429251314491. https://doi.org/10.1177/00222429251314491
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124
- Victor Tangermann. (2024, January 12). *Amazon Is Selling Products With AI-Generated Names Like 'I Cannot Fulfill This Request It Goes Against OpenAI Use Policy'*. Futurism. https://futurism.com/amazon-products-ai-generated
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* [Wardle C,]. Council of Europe Publishing. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*(1), 26. https://doi.org/10.1007/s40979-023-00146-z
- World Economic Forum. (n.d.). *Global Risks 2024: Disinformation Tops Global Risks 2024 as Environmental Threats Intensify*. World Economic Forum. Retrieved 30 May 2025, from https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/
- World Health Organization. (n.d.). *Infodemic*. https://www.who.int/health-topics/infodemic

- Čavojová, V., Brezina, I., & Jurkovič, M. (2022). Expanding the bullshit research out of pseudo-transcendental domain. *Current Psychology*, *41*(2), 827–836. https://doi.org/10.1007/s12144-020-00617-3

Appendix A - Slop Literature

Preprints

Chakrabarty, T., Laban, P., & Wu, C.-S. (2025). *AI-Slop to AI-Polish? Aligning Language Models through Edit-Based Writing Rewards and Test-time Computation* (No. ArXiv: 2504.07532). ArXiv. https://doi.org/10.48550/arXiv.2504.07532

Klincewicz, M., Alfano, M., & Fard, A. E. (2025). *Slopaganda: The interaction between propaganda and generative AI* (No. ArXiv: 2503.01560). ArXiv. https://doi.org/10.48550/arXiv.2503.01560

Analyses

Heer, N. (2024, July 11). *'Slop' and 'Content'*. https://pxlnv.com/blog/slop-and-content/

Slop Infrastructures 1 & 2. (2024, December 8). Cybernetic Forests. https://mail.cyberneticforests.com/slop-infrastructures-1-2/

The semiotics of AI slop—MediaCat UK. (2025, January 21). https://mediacat.uk/the-semiotics-of-ai-slop/

Investigations

Joelving, F. (2025, February 10). As a nonsense phrase of shady provenance makes the rounds, Elsevier defends its use. *Retraction Watch*. https://retractionwatch.com/2025/02/10/vegetative-electron-microscopy-fingerprint-paper-mill/

Rise of the Newsbots: AI-Generated News Websites Proliferating Online. (n.d.). *NewsGuard*. Retrieved 26 May 2025, from https://www.newsguardtech.com/special-reports/newsbots-ai-generated-news-websites-proliferating

Tracking AI-enabled Misinformation: Over 1200 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools. (n.d.). *NewsGuard*. Retrieved 26 May 2025, from https://www.newsguardtech.com/special-reports/ai-tracking-center

Knibbs, K. (2024, February 7). Confessions of an AI Clickbait Kingpin. *Wired*. Retrieved 13 May 2025, from https://www.wired.com/story/confessions-of-an-ai-clickbait-kingpin/

Vinu Sankar Sadasivan, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, & Soheil Feizi. (2023). Can AI-Generated Text be Reliably Detected?

arXiv. Org. https://doi.org/10.48550/arxiv.2303.11156

Cameron, D. (2016, November 18). Viral 'fake news' list deleted after author and students get harassed, doxed. *The Daily Dot.* https://www.dailydot.com/debug/fake-news-facebook-deleted-alt-right/

Wikipedia article references (Retrieved on 20/05/2025)

- Hern, Alex; Milmo, Dan (19 May 2024). "Spam, junk … slop? The latest wave of AI behind the 'zombie internet'". *The Guardian.* Retrieved 28 September 2024.
- Notopoulos, Katie. "Why doesn't Facebook just ban AI slop like Shrimp Jesus?". *Business Insider.* Retrieved 4 October 2024.
- Field, Matthew (1 January 2025). "Why the internet is filling up with nonsense 'AI slop'". *The Telegraph.* Retrieved 3 January 2025.
- Read, Max (25 September 2024). "Drowning in Slop". *Intelligencer.* Retrieved 28 September 2024.
- "How Long Will A.I.'s 'Slop' Era Last?". *The New York Times.* 24 July 2024.
- Hughes, Johnny (3 September 2024). "How Businesses Can Avoid AI Slop". *Entrepreneur.* Retrieved 28 September 2024.
- "First Came 'Spam.' Now, With A.I., We've Got 'Slop'". *The New York Times.* 11 June 2024.
- "The deluge of bonkers AI art is literally surreal". *Washington Post.* 30 June 2024. Retrieved 29 September 2024.
- "Slop is the new name for unwanted AI-generated content". *simonwillison. Net.* Retrieved 4 October 2024.
- Guo, Yanzhu; Shang, Guokan; Vazirgiannis, Michalis; Clavel, Chloé (16 April 2024). "The Curious Decline of Linguistic Diversity: Training Language Models on Synthetic Text". arXiv: 2311.09807 [cs.CL].
- Mohamed, Amr; Geng, Mingmeng; Vazirgiannis, Michalis; Shang, Guokan (27 February 2025). "LLM as a Broken Telephone: Iterative Generation Distorts Information". arXiv: 2502.20258.
- Koebler, Jason (6 August 2024). "Where Facebook's AI Slop Comes From". *404 Media.* Retrieved 29 September 2024.
- Tang, Jiaru; Wikström, Patrik (19 September 2024). "'Side job, self-employed, high-paid': behind the AI slop flooding TikTok and Facebook". *The Conversation.* Retrieved 29 September 2024.
- Gault, Matthew (7 August 2024). "Facebook's Twisted Incentives Created Its AI Slop Era". *Gizmodo.* Retrieved 29 September 2024.
- Warzel, Charlie (21 August 2024). "The MAGA Aesthetic Is AI Slop". *The Atlantic.* Retrieved 29 September 2024.
- Warzel, Charlie (21 October 2024). "The Slop Candidate". *The Atlantic.* Retrieved 23 March 2025.
- Robins-Early, Nick (26 August 2024). "How did Donald Trump end up posting Taylor Swift deepfakes?". *The Guardian.* Retrieved 23 March

2025.

- “Gullible Trump Cronies Losing Their Minds Over Fake AI Slop on Twitter”. *Yahoo News*. 4 October 2024. Retrieved 6 October 2024.
- Koebler, Jason (8 October 2024). “Hurricane Helene and the ‘Fuck It’ Era of AI-Generated Slop”. *404 Media*. Retrieved 15 October 2024.
- Warzel, Charlie (10 October 2024). “I’M RUNNING OUT OF WAYS TO EXPLAIN HOW BAD THIS IS”. *The Atlantic*. Retrieved 15 October 2024.
- Stedman, Alex (16 November 2024). “Coca-Cola’s New AI-Generated Holiday Ad Slammed as ‘Soulless’ and ‘Embarrassing’: ‘This Is Such Slop’ ”. *IGN*. Retrieved 5 January 2025.
- “Coca-Cola causes controversy with AI-generated ad”. *NBC News*. 18 November 2024. Retrieved 5 January 2025.
- Vadukul, Alex (20 November 2024). “Coca-Cola’s Holiday Ads Trade the ‘Real Thing’ for Generative A.I.” *The New York Times*. Retrieved 5 January 2025.
- “Paramount skips voice actors, opting for shockingly bad AI slop in ‘Novocaine’ film promo”. *TweakTown*. 10 March 2025. Retrieved 24 March 2025.
- Maker, Paul (9 January 2025). “AI slop: what Labour, Spotify and Coca-Cola can teach us in 2025”. *Raconteur*. Retrieved 24 March 2025.
- Shankar, Bradly (18 April 2024). “A24’s lacklustre Civil War ads are the latest example of AI-generated slop”. *MobileSyrup*. Retrieved 24 March 2025.
- Phillips, Tom (3 March 2025). “Activision accused of”AI slop” yet again, this time for new Guitar Hero”. *Eurogamer. Net*. Retrieved 27 March 2025.
- Phillips, Tom (4 March 2025). “Activision”AI slop” Guitar Hero advert is market research for a fake game that doesn’t exist and maybe never will”. *Eurogamer. Net*. Retrieved 27 March 2025.
- “Why Brands Are Embracing ‘Italian Brainrot’ to Go Viral on TikTok and Win Over Gen Z”. *News 10 ABC*. 15 April 2025.
- Currie, Richard (28 February 2024). “Willy Wonka event leaves bitter taste with artificially sweetened promises”. *The Register*. Retrieved 4 November 2024.
- Murphy, Chris (29 February 2024). “ “Willy’s Chocolate Experience” Nightmare: What Went Wrong?”. *Vanity Fair*. Retrieved 4 November 2024.
- Davis, Barney (1 November 2024). “Chaos in Dublin as thousands turn up for AI ‘hoax’ Halloween parade that didn’t exist”. *The Independent*. Retrieved 1 November 2024.
- “All Trick, No Treat: Dublin Crowds Turn Up for Halloween Parade That Wasn’t”. *The New York Times*. 1 November 2024.
- Fleck, Holly (2 November 2024). “Dublin: Halloween parade listing ‘mistake’ says website owner”. *BBC News*. Retrieved 14 November 2024.
- O’Shea, Kerry (31 October 2024). “All trick no treat -”Hoax” Halloween

parade draws big crowds to Dublin's O'Connell St". *IrishCentral. Com*. Retrieved 1 November 2024.
- Knibbs, Kate. "The Guy Behind the Fake AI Halloween Parade Listing Says You've Got It All Wrong". *Wired*. ISSN 1059-1028. Retrieved 3 November 2024.
- Maiberg, Emanuel (4 February 2025). "AI-Generated Slop Is Already In Your Public Library". *404 Media*. Retrieved 6 February 2025.
- Yin-Poole, Wesley (9 December 2024). "Call of Duty Fans Give Black Ops 6's Zombie Santa Loading Screen the Finger Amid 'AI Slop' Backlash". *IGN*. Retrieved 26 February 2025.
- Yin-Poole, Wesley (24 July 2024). "Activision Reportedly Sold an AI-Generated Call of Duty: Modern Warfare 3 Cosmetic". *IGN*. Retrieved 26 February 2025.
- "Call Of Duty Discloses AI Slop After Months Of Players Complaining". *Kotaku*. 25 February 2025. Retrieved 26 February 2025.
- "PlayStation Exclusive Foamstars Has AI-Generated Art, Square Enix Confirms". *IGN*. 16 January 2024. Retrieved 31 March 2025.
- "Square Enix"dabbled" with AI technology for Foamstars". *Eurogamer. Net*. 16 January 2024. Retrieved 31 March 2025.
- Parrish, Ash (16 January 2024). "Square Enix says it used AI art in upcoming Foamstars game". *The Verge*. Retrieved 31 March 2025.
- Bankhurst, Adam (1 January 2024). "Square Enix President States the Company Will Be 'Aggressive in Applying AI' ". *IGN*. Retrieved 24 April 2025.
- Nightingale, Ed Nightingale (Deputy News) (2 January 2024). "Square Enix intends to be"aggressive in applying AI", says CEO". *Eurogamer. Net*. Retrieved 24 April 2025.
- Kalita, Parash Jyoti (1 November 2024). "Angry Birds Block Quest: Rovio's latest title in the franchise has soft-launched for Android in the US". *GamingonPhone*. Retrieved 9 May 2025.
- "Angry Birds Block Quest Soft Launches on Android, Sparking Debate Over Generative AI Content -". *mxdwn Games*. 2 November 2024. Retrieved 9 May 2025.
- "Angry Birds Block Quest - Soft launch for Android in the US". 1 November 2024. Retrieved 9 May 2025.
- "Shudder Just Quietly Released the Most Controversial Thriller of the Year". *Inverse*. 19 April 2024. Retrieved 24 March 2025.
- "Late Night with the Devil: The Horror Film with Horrifying Implications". *Shadow Magazine*. 20 November 2024. Retrieved 24 March 2025.
- Hüsmen, Arife (November 2024). "AI in the Creative Industries: Strikes, Debates and Implications". *20th International Symposium of Communication in the Millennium*: 193–199 – via ResearchGate.
- "Amazon Mocked for Slapping AI-Generated Poster on Beloved 1922 Film"Nosferatu"". *Futurism*. 7 January 2025. Retrieved 24 March 2025.
- John, Daniel (12 June 2024). "Not even Amazon is safe from the rise of AI 'slop' ". *Creative Bloq*. Retrieved 24 March 2025.

- "The mystery of Amazon Freevee's A.I.-generated 12 Angry Men poster". *AV Club*. Retrieved 24 March 2025.
- "Amazon's Prime Video Appears to Have Left AI Slop in the Public Description of an Oscar-Winning Al Pacino Film". *Futurism*. 19 February 2025. Retrieved 24 March 2025.
- " 'Murderesses': MagentaTV entfernt Serie mit KI-Synchronisation". *Heise Online* (in German). 4 February 2025. Retrieved 25 April 2025.

Appendix B - Sources of slop

Science integrity

- PPS flagged papers: PubPeer - Acknowledgement to reviewers (2015)
- Retraction watch data: crossref / retraction-watch-data · GitLab
- PPS data: [PPS – Problematic Paper Screener](https://dbrech.irit.fr/pls/apex/f?p=9999:1:::::
- Beall's List of Potential Predatory Journals and Publishers
- The Tadpole Paper Mill – Science Integrity Digest

Journalism

- Newsguard:
  - Famadillo. Com
  - GetIntoKnowledge. Com
  - BestBudgetUSA. Com
  - HarmonyHustle. Com
  - HistoryFact. In
  - CountyLocalNews. Com
  - TNewsNetwork. Com
  - CelebritiesDeaths. Com
  - WaveFunction. Info
  - ScoopEarth. Com
  - FilthyLucre. Com
  - Biz Breaking News
  - News Live 79
  - Daily Business Post
  - Market News Reports
- Synthetic Echo
  - Espn24. Co. Uk
  - Nbcsportz. Com
  - Nbcsport. Co. Uk
  - Cbsnewz. Com
  - Cbsnews2. Com
  - Bbcsportss. Co. Uk
  - 247bbcnews. Com
  - Foxnigeria. Com. Ng
- Confessions of an AI Clickbait Kingpin | WIRED
  - http://thehairpin.com

  - http://antoniocarluccio.com
  - http://pope2you.net
  - http://trumpplaza.com
  - http://thefrisky.com
- Slop books: Publishing.ai
- AI influencers: AnnaIndianaAI
- Artificial Intelligence Incident Database

Appendix C - Products and organizations

AI detectors

- **GPTZero**
- **Pangram**
- BurhanUlTayyab/GPTZero: An open-source implementation of GPTZero
- Perplexity of fixed-length models
- Oct4Pie/zero-zerogpt: Bypassing AI Content Detectors
- Zapier: Sapling, Winston AI, ZeroGPT, GPTZero, Copyleaks, Smodin
-

Information Reliability

- **Reporters' Lab Tech and Check**: Fact-Check Insights, MediaVault, ClaimReview, MediaReview, Squash, Alerts, FactStream,
- **NewsGuard**: Reliability ratings, misinformation fingerprints, AI Safety.
- **Meedan**
- NewsCord
- Newswhip: Spike, Analytics, API, Patents
- Aos Fatos: Radar, Fátima
- MBFC's Data API
- The Trust Project
- Hamilton 2.0
- ClaimBuster
- FakeOut

PolitiFact, **Full Fact**, ProPublica, The Markup, bellingcat, Open Knowledge Brasil, The Bureau of Investigative Journalism (TBIJ), Centre for Investigative Journalism (CIJ), Agência Pública, The Pudding, AllSides, Ground News

Science integrity

- **Problematic Paper Screener**
- **ClearSkies**
- Seek & Blastn
- ImageTwin
- Argos
- Morressier

- Signals
- Plagiarism detection tools – Science Integrity Digest
- Image checking tools – Science Integrity Digest
- List of science integrity resources

Nonprofits and research

- **First Draft**
- **Data & Society**
- Google News Lab

Third-parties & implementation

- TextureAI
- Storyzy
- Bad Idea Factory: truth goggles, talking point trackers
- Ridgeway Information
- Dimensions AI

Appendix D - Philosophy

Bullshit

Frankfurt, H. G. (2009). *On Bullshit*. Princeton University Press. https://doi.org/10.1515/9781400826537

Cohen, G. A. (2002). Deeper into Bullshit. In S. Buss & L. Overton (Eds.), *The Contours of Agency* (pp. 321–339). The MIT Press. https://doi.org/10.7551/mitpress/2143.003.0015

Easwaran, K. (n.d.). *Bullshit activities*. Retrieved 29 April 2025, from https://onlinelibrary.wiley.com/doi/10.1111/phib.12328

AI and bullshit C-Lara-Instance, C., & Rayner, M. (2025). *'ChatGPT is Bullshit' is Bullshit: A Coauthored Rebuttal by Human & LLM*. https://doi.org/10.1007/s10676-024-09775-5/metrics

Fisher, S. A. (2024). Large language models and their big bullshit potential. *Ethics and Information Technology*, *26*(4), 67. https://doi.org/10.1007/s10676-024-09802-5

Gorrieri, L. (2024). Is ChatGPT Full of Bullshit? *Journal of Ethics and Emerging Technologies*, *34*(1), Article 1. https://doi.org/10.55613/jeet.v34i1.149

Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, *26*(2), 38. https://doi.org/10.1007/s10676-024-09775-5

Wachter, S., Mittelstadt, B., & Russell, C. (2024). Do large language models have a legal duty to tell the truth? *Royal Society Open Science*, *11*(8), 240197. https://doi.org/10.1098/rsos.240197

Prada, L. (2025, May 6). ChatGPT Is Giving People Extreme Spiritual Delusions. *VICE*. https://www.vice.com/en/article/chatgpt-is-giving-people-extreme-spiritual-delusions/

Appendix E - Psychology

Scales

- Pseudo-Profound Bullshit Receptivity Scale
- Scientific Bullshit Receptivity Scale
- Belief in Science Scale
- Social and Economic Conservatism Scale
- Free Market Belief Scale
- Faith in Intuition Scale
- Need for Cognition Scale
- Science Literacy Scale
- Acquiescence Bias Measure
- Social and Economic Conservatism Scale
- General Bullshit Receptivity Scale (GBRS)
- New (Non-Transcendental) Bullshit Receptivity Scale (NBSR)
- Bullshit Receptivity Scale (BSR)
- Epistemically Suspect Beliefs Scale (ESB)
- Conspiracy Mentality Questionnaire (CMQ)
- Pseudoscientific Beliefs (PSB)
- Paranormal Belief Scale (PBS)
- Ontological Confusion (OC)
- Analytic Thinking (CRT)
- Big Five Inventory (BFI) – Extra-Short Version
- Daily Spiritual Experience Scale (DSES) – Shortened Version

Papers

Defrancesco, E., & Strapparava, C. (2023). *The PBSDS: A dataset for the detection of pseudoprofound bullshit.* https://cris.fbk.eu/handle/11582/343408

Evans, A., Sleegers, W., & Mlakar, Ž. (2020). Individual differences in receptivity to scientific bullshit. *Judgment and Decision Making*, *15*(3), 401–412. https://doi.org/10.1017/S1930297500007191

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, *10*(6), 549–563. https://doi.org/10.1017/S1930297500006999

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2016). It's still bullshit: Reply to Dalton (2016). *Judgment and Decision Making*, *11*(1), 123–125. https://doi.org/10.1017/S1930297500007658

Čavojová, V., Brezina, I., & Jurkovič, M. (2022). Expanding the bullshit research out of pseudo-transcendental domain. *Current Psychology*, *41*(2), 827–836. https://doi.org/10.1007/s12144-020-00617-3

Čavojová, V., Jurkovič, M., & Brezina, I. (n.d.). *To be or not to be bullshit? Obscuring the form vs. Moving away from truthfulness.*

Jurkovič, M., Čavojová, V., & Brezina, I. (n.d.). *Looking for the most defining feature of bullshit: Obscureness vs. Untruthfulness.*

Appendix F - Detection and Metrics

Linguistic & Stylistic Analysis

Task	Description	Tool (s)
POS tagging	Identifies part-of-speech (nouns, verbs, etc.)	`spaCy`, `nltk`
Dependency parsing	Analyzes syntactic relationships between words	`spaCy`
Function word usage	Frequency of articles, prepositions, conjunctions	`spaCy`, `textdescriptives`, custom counts
Sentence length / complexity	Measures sentence structure and depth	`textdescriptives`, `textstat`, `spaCy`
Average word length	Basic lexical sophistication metric	`textdescriptives`, custom
Vocabulary size / density	Lexical diversity and information richness	`textdescriptives`, `trunajod`, `lexicalrichness`
Character n-grams	Stylometric fingerprinting based on sequences of characters	`scikit-learn`, `nltk`
Stylometry (repetitiveness, etc.)	Measures text style, redundancy, and structure	`trunajod`, `textdescriptives`, `textstat`
Readability	Calculates metrics like Flesch-Kincaid, Gunning Fog	`textstat`
Burrows' Delta	Stylometric distance based on function words	`linguistic-features`, custom

Semantic & Emotional Content

Task	Description	Tool (s)
Sentiment analysis	Positive, negative, or neutral tone	`nltk.vader`, `TextBlob`, `spaCy`, `transformers`
Emotion detection	Extracts emotions like joy, anger, sadness	`text2emotion`, `NRCLex`, `affectpy`
Topic modeling	Identifies main topics/themes in text	`gensim.LDA`, `BERTopic`
Topic entropy	Entropy over topic distribution (topic coherence)	`scipy.stats.entropy` + topic probabilities

Statistical & Distributional Features

Task	Description	Tool (s)
TF-IDF vectorization	Weighs word importance relative to a corpus	`scikit-learn`
Jaccard similarity	Token-based set similarity between texts	`scikit-learn`, `nltk`, `rapidfuzz`
Cosine similarity (Embeddings)	Semantic similarity between text vectors	`sentence-transformers`, `scikit-learn`
Fuzzy string matching	Approximate text comparison	`rapidfuzz`, `fuzzywuzzy`
Perplexity	Measures how predictable a text is (lower = more fluent)	`transformers` (GPT-2/3 models)
Zipf's Law	Word frequency distribution conformity	`powerlaw`, custom
Shannon entropy	Measures lexical randomness/complexity	`scipy.stats.entropy`, custom
KL divergence	Divergence between two probability distributions	`scipy.stats.entropy`, `numpy`
Jensen-Shannon divergence	Symmetrized, bounded version of KL divergence	`scipy`, custom
Likelihood ratios / Bayesian	Probabilistic comparisons or classification	`scikit-learn`, `pymc`, `pyro`

Authorship & Fact-Checking

Task	Description	Tool (s)
Stylometric authorship detection	Measures author's stylistic fingerprint	`trunajod`, `JStylo (Java)`, `stylo` (R), custom

Sources:

Chen, C., Wang, H., Shapiro, M., Xiao, Y., Wang, F., & Shu, K. (2022). *Combating Health Misinformation in Social Media: Characterization, Detection, Intervention, and Open Issues* (No. ArXiv: 2211.05289). ArXiv. https://doi.org/10.48550/arXiv.2211.05289

Tang, R., Chuang, Y.-N., & Hu, X. (2023). *The Science of Detecting LLM-Generated Texts* (No. ArXiv: 2303.07205). ArXiv. https://doi.org/10.48550/arXiv.2303.07205

Linguistic-Based Detection of Fake News in Social Media | Mahyoob | International Journal of English Linguistics | CCSE

Antypas, D., J. Camacho-Collados, A. Preece, and D. Rogers. 2021. "COVID-19 and Misinformation: A Large-Scale Lexical Analysis on Twitter." In Proceedings of ACL.

Stylometric Fake News Detection Based on Natural Language Processing Using Named Entity Recognition: In-Domain and Cross-Domain Analysis

Buchholz, M. G. (2023). *Assessing the Effectiveness of GPT-3 in Detecting False Political Statements: A Case Study on the LIAR Dataset* (No. ArXiv: 2306.08190). ArXiv. https://doi.org/10.48550/arXiv.2306.08190

Leon Fröhling and Arkaitz Zubiaga. 2021. Featurebased detection of automated language models: tackling gpt-2, gpt-3 and grover. PeerJ Computer Science, 7: e443.

Mahyoob, M., J. Al-Garaady, and M. Alrahaili. 2020. "LinguisticBased Detection of Fake News in Social Media." International Journal of English Linguistics Forthcoming.

Aich, A., S. Bhattacharya, and N. Parde. 2022. "Demystifying Neural Fake News Via Linguistic Feature-Based Interpretation." In Proceedings of the COLING.

Mewada, A., & Dewang, R. K. (2023). A comprehensive survey of various methods in opinion spam detection. *Multimedia Tools and Applications*, *82*(9), 13199–13239. https://doi.org/10.1007/s11042-022-13702-5

Wu, J., Yang, S., Zhan, R., Yuan, Y., Chao, L. S., & Wong, D. F. (2025). A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions. *Computational Linguistics*, *51*(1), 275–338. https://doi.org/10.1162/coli_a_00549

Deep learning

Techniques: SHAP, LIME, SAE, Ablation

Aich, A., S. Bhattacharya, and N. Parde. 2022. Demystifying Neural Fake News Via Linguistic Feature-Based Interpretation.

Niven, T., & Kao, H.-Y. (2019). Probing Neural Network Comprehension of Natural Language Arguments.

McCoy, R. T., Pavlick, E., & Linzen, T. (2019). Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference

Tools

- GLTR
- LIWC
- GitHub - brucewlee/lftk: [BEA @ ACL 2023] General-purpose tool for linguistic features extraction; Tested on readability assessment, essay scoring, fake news detection, hate speech detection, etc.
- TRUNAJOD: A text complexity library for text analysis built on spaCy — TRUNAJOD 0.1.1 documentation
- GitHub - novoic/blabla: Novoic's linguistic feature extraction library
- NLP feature extraction from LIWC in Python | by Shivika K Bisen | Bright AI | Medium
- GitHub - HLasse/TextDescriptives: A Python library for calculating a large variety of metrics from text
- `spacy`, `textdescriptives`, `textstat`, `trunajod`, `linguistic-features`, `nltk`, `textblob`, `text2emotion`, `NRCLex`, `scikit-learn`, `sentence-transformers`, `transformers`, `scipy`, `powerlaw`, `rapidfuzz`

Surveys on classification tasks

- **Classification in general**
  - london.ac.uk/sites/default/files/study-guides/introduction-to-natural-language-processing.pdf
- **Machine generated content**
  - A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions | Computational Linguistics | MIT Press
- **Misinformation/Disinformation/Fake news/Rumour**
  - A survey on fake news and rumour detection techniques - ScienceDirect
  - [1708.01967] Fake News Detection on Social Media: A Data Mining Perspective
- **Spam**
  - A systematic literature review on spam content detection and classification [PeerJ]
  - A comprehensive survey of various methods in opinion spam detection | Multimedia Tools and Applications
- **Clickbait**
  - Prompt-tuning for Clickbait Detection via Text Summarization
- Propaganda
  - [2007.08024] A Survey on Computational Propaganda Detection
- Hallucination

- A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions | ACM Transactions on Information Systems
- Survey of Hallucination in Natural Language Generation | ACM Computing Surveys

- Deception
  - Intelligent techniques for deception detection: a survey and critical study | Soft Computing
- Hate speech/toxicity/stereotype/offensive language
  - Hate speech, toxicity detection in online social media: a recent survey of state of the art and opportunities | International Journal of Information Security
- Stylometry/authorship attribution
  - Surveying Stylometry Techniques and Applications | ACM Computing Surveys
- Stance
  - [2409.15690] A Survey of Stance Detection on Social Media: New Directions and Perspectives
- Satire/parody
  - Comprehensive Study of Arabic Satirical Article Classification
- Irony/sarcasm
  - Automatic Sarcasm Detection: A Survey: ACM Computing Surveys: Vol 50, No 5
  - A survey of automatic sarcasm detection: Fundamental theories, formulation, datasets, detection methods, and opportunities - ScienceDirect
- Sentiment
  - A survey of automatic sarcasm detection: Fundamental theories, formulation, datasets, detection methods, and opportunities - ScienceDirect
- Stereotype (King et al., 2024)
  -
- Nonsense
  - On the Cusp of Comprehensibility: Can Language Models Distinguish Between Metaphors and Nonsense? - ACL Anthology
- Metaphor
  - A Survey on Computational Metaphor Processing | ACM Computing Surveys
- Pseudoprofound bullshit
  - The PBSDS: A dataset for the detection of pseudoprofound bullshit

Appendix G - Archives and datasets

Curated by me

- Lacan_books: ~10 books by French psychoanalyst Jacques Lacan

- Official tumblr metaphysics repository: ~50 bullshit humour microblogs
- Civitai: ~60k images scraped from an AI image generation website
- Simulacro_db: hand-selected meaningful quotes from philosophy and psychoanalysis books

Generators

- New-Age Bullshit Generator
- InspiroBot
- SCIgen - An Automatic CS Paper Generator (they also have a list)
- Postmodernism Generator

Bullshit datasets

- PBSDS
- ouhenio/llms-overstimate-profoundness
- acmi-lab/pretraining-with-nonsense: Pretraining summarization models using a corpus of nonsense
- New Nonsense BEL Sentence Corpus
- APA PsycNet DoiLanding page

Not bullshit nor slop

- Authors present in the British library gift shop: Shakespeare, Orwell, Jane Austen, Sherlock Holmes, Agathe Christie.
- Well regarded but hard books: Joyce, Ezra Pound, T.S. Eliot, Burroughs.
- Time Magazine's All-Time 100 Novels (100 books)
- Most important papers in science
  - The top 100 papers : Nature News & Comment
  - These are the most-cited research papers of all time
  - Top 100 most cited publications
  - Citation impact - Wikipedia
  - Site Unreachable

Politics

- Trump Claims Database: 30,573 false or misleading claims made by Trump and fact-checked by the Washington Post.
- Declaracoes de Bolsonaro: same with 6685 claims made by Bolsonaro.
- Planalto Discursos e Pronunciamentos: Archive of pronouncements made by Lula from 2023.
- The American Presidency Project: 100,000 documents related to the study of the American presidency, including presidential debates, speeches, state of the union addresses, etc.
- Presidential Speeches | Miller Center: 50 years of U.S. presidential speeches.

- UKPOL: 80,000 speeches and press releases relating to UK politics, as well as a growing number of interviews, book reviews and other political and electoral resources.
- AgoraSpeech: meticulously curated, high-quality dataset of 171 political speeches from six parties during the Greek national elections in 2023
- ParlEE plenary speeches data set: Annotated full-text of 21.6 million sentence-level plenary speeches of eight EU states
- Leadership Studies - Research Guides at Harvard Library
  - American Rhetoric Online Speech Bank: 5000+ full text, audio and video versions of public speeches, sermons, legal proceedings, lectures, debates, interviews, other recorded media events, and a declaration or two. Full text, audio, and video database of the 100 most significant American political speeches of the 20th century.
  - Commission on Presidential Debates: Transcripts from U.S. presidential debates.
  - Great Speeches of the 20th Century: The Guardian and Observer's unique series of the best speeches of the last century.
  - Speeches at the United Nations: Speeches held before the United Nations dating back to 1946.
  - Living Room Candidate: Presidential Campaign Commercials, 1954-2012.
  - Say It Plain, Say it Loud: speeches by U.S. black leaders.
  - UN Peacekeeping - Speeches & Statements: Latest press releases, speeches, and statements from senior officials on UN Peacekeeping.
  - Trump datasets
    - ichalkiad/datadescriptor_uselections2020
    - alexmill/trump_transcripts
    - christianlillelund/donald-trumps-rallies
    - etaifour/trump-speeches-audio-and-word-transcription
    - ryanmcdermott/trump-speeches
    - tuenguyen/trump-speech-dataset-tts

Corporate jargon

- Manual search through SEO farms and small lists
- https://www.buzzwordbingogame.com/
- http://officeipsum.com/
- Corporate Ipsum Gobbledygook Generator
- https://www.buzzwordipsum.com/
- The Dictionary of Corporate Bullshit
- The BS Dictionary: Uncovering the Origins and True Meanings of Business Speak : Wiltfong, Bob, Ito, Tim: Amazon.co.uk: Books
- The Business Bullshit Book
- Who Touched Base in my Thought Shower?

Dadaism

- Category:Dada - Wikimedia Commons
- Wikipedia:The Wikipedia Library - Wikipedia
- HathiTrust Digital Library – Millions of books online
- Dada (published from Zürich and Paris), The Blind Man, Rongwrong, New York Dada, 391 (published in multiple cities), Club Dada, Der Dada, Everyman His Own Football, Dada Almanach, Le Cannibale, Littérature, De Stijl (which included Dada poetry), Mécano, and The Next Call.
- Manifestos, Cut-up technique, Sound Poetry
- Tristan Tzara, Hugo Ball, Richard Huelsenbeck, Mina Loy, Louis Aragon, Paul Eluard, Philippe Soupault, Georges Ribemont-Dessaignes, and Walter Serner. Mina Loy. Iliazd

Public domain fringe belief primary texts:

- **Helena Blavatsky** – ***The Secret Doctrine***
- **Ignatius Donnelly** – ***Atlantis: The Antediluvian World***
- **Charles Fort** – ***The Book of the Damned***
- **James Churchward** – ***The Lost Continent of Mu***
- **Albert Churchward** – ***The Origin and Evolution of the Human Race***
- **Godfrey Higgins** – ***Anacalypsis***
- **Giordano Bruno** – ***The Ash Wednesday Supper***
- **William Scott-Elliot** – ***The Story of Atlantis and the Lost Lemuria***
- **H. Spencer Lewis – _Self Mastery and Fate with the Cycles of Life

Quote aggregators

Wikiquote, BrainyQuote, Goodreads, Quote Garden, QuoteFancy, r/QuotesPorn, Wisdom Quotes, The Quotations Page, Quote Master, Bartlett's Familiar Quotations, Quotes.net, Thoughtco

New Age blogs and websites

Spirit Library, Wake Up World, In5D, The Mind Unleashed, Collective Evolution, Gaia, Elephant Journal, LonerWolf, Mystic Mamma, The Numinous, Healing Energy Tools, Zentasia, The Explorer Lounge

Fringe beliefs

Unarius, Time Cube, John Titor, Francis E. Dec, Children of the Matrix, Ramtha's School of Enlightenment, Emerald Tablet, Emerald Tablets of Thoth, Pseudolaw, Montauk Project, Timewave Zero, Pathwork, Core Energetics, The Final Theory

Authors, influencers and thought-leaders

Deepak Chopra, Eckhart Tolle, Esther Hicks / Abraham-Hicks, Rhonda Byrne, Gregg Braden, Teal Swan, Dolores Cannon, Neale Donald Walsch, Caroline Myss, Don Miguel Ruiz, Louise Hay, Joe Dispenza, Sri Sri Ravi Shankar, Byron Katie, Alan Watts

Motivational speakers

Tony Robbins (Awaken the Giant Within), Les Brown (It's Not Over Until You Win), Eric Thomas (You Owe You), Brené Brown (The Power of Vulnerability), Simon Sinek (Start With Why), Mel Robbins (The 5 Second Rule), Nick Vujicic (Life Without Limits), Jay Shetty (Think Like a Monk)

Machine Learning Datasets

- Human text: The Pile, OpenWebText, RealNews, ELI5, Enron, BNC, LOB, English Corpora
- AI-generated text: GPT-2 Output, HC3, HC3 GitHub, MULTITuDE, MAGE, TURINGBENCH, M4, OUTFOX, IDMGSP, MGTBench, LLM Detect, 150K Wikipedia GPT, Slop Cloud, Suno Prompts, Song Lyrics Genius
- Rumour: PHEME, Weibo, Twitter15, Twitter16, Multimodal RNN, WeiboRumours, CREDBANK, CED, STANKER, Weibo20, UCINLP COVID-19, ESOC COVID-19
- Hate speech: EMGSD, Chess, SemEval-2019 Task 5, Dynabench, OLID, SOLID, HateXplain, HateCheck, English Hate Speech Superset
- Deception: UNIDECOR, Deceptive Opinion Spam, Deceptive Opinion Paper, Mafiascum, SARC, iSarcasm, DeFaBel
- Hallucination: awesome-hallucination-detection (archive), DelucionQA, DefAn, HaDes, LibreEval, HaluEval, HaluBench
- Conspiracy: COCO, ConspEmoLLM, CCRS, YouTube Conspiracy, Anatomy of Conspirators, WICO Graph, Harvard DVN
- Fake news
  - CDL Misinfo Datasets.: ~100 misinfo datasets (archive)
  - Fake news detection: a survey of evaluation datasets [PeerJ]: 27 misinfo datasets (archive)
  - Fake News Datasets | MediaFutures: ~10 misinfo datasets (archive)
  - lzw108/Misinformation-datasets-fakenews-rumors-conspiracy: archive of ~30 datasets (archive)
  - FactCheck Insights: (requires asking) dataset with 200k fact-checks.
  - OpenSources: labelled 1000 news websites
  - LLMFake: This is an LLM-generated misinformation dataset containing nonfactual content created by various LLMs using seven different approaches. (Uses Politifact, CoAID and Gossipcop).
  - FakeNewsNet: BuzzFeedNews + Politifact
  - CoronaVirusFacts Alliance: Coronavirus fact-checking response.

- Survey 1: Arianna D'Ulizia, Maria Chiara Caschera, Fernando Ferri, and Patrizia Grifoni. 2021. Fake news detection: a survey of evaluation datasets. *PeerJ Computer Science* 7 (2021), e518. (archive)
- **LIAR2**: ~23K professionally labeled statements from PolitiFact
- NELA-GT-2019/2020/2022: millions of news articles with outlet-level veracity labels
- FEVER: 200k Wikipedia-based claims labeled as Supported, Refuted, or NotEnoughInfo
- RefChecker