

“All You Need” is Not All You Need for a Paper Title: On the Origins of a Scientific Meme

Anton Alyakin
alyakin314@gmail.com

Abstract

The 2017 paper “Attention Is All You Need” introduced the Transformer architecture, and inadvertently spawned one of machine learning’s most persistent naming conventions. We analyze 717 arXiv preprints containing “All You Need” in their titles (2009–2025), finding exponential growth ($R^2 > 0.994$) following the original paper, with 200 titles in 2025 alone. Among papers following the canonical “X [Is] All You Need” structure, “Attention” remains the most frequently claimed necessity (28 occurrences). Situating this phenomenon within memetic theory, we argue the pattern’s success reflects competitive pressures in scientific communication that increasingly favor memorability over precision. Whether this trend represents harmless academic whimsy or symptomatic sensationalism, we leave—with appropriate self-awareness—to the reader.

1 Introduction

In June 2017, Vaswani et al. [1] published a paper that would fundamentally reshape the landscape of deep learning. While on the surface appearing to be yet just another machine translation paper, “Attention Is All You Need” introduced the Transformer architecture which was destined to become the backbone of modern large language models [2], vision-language models [3], and countless other applications. However, just as serendipitous as its technical impact was its cultural legacy: a title format that has proven irresistible to subsequent researchers.

The phrase “All You Need Is Love,” first broadcast to an estimated 400 million viewers during The Beatles’ 1967 appearance on *Our World* [4], carries an inherent memetic appeal. Its structure—asserting that a single element is sufficient for some complex all-capturing endeavor—combines simplicity with boldness. When applied to academic work, this format implicitly promises parsimony: the paper will identify one crucial insight that renders other approaches unnecessary. While this indeed happened with the attention mechanism replacing nearly all recurrent and convolutional components in sequence modeling, it has since been used hundreds of times, often with considerably less revolutionary justifications, implications, and consequences.

This study examines the proliferation of the “All You Need” naming pattern in academic manuscripts indexed on arXiv. We address three questions:

- How did the frequency of this pattern evolve over time?
- What subjects do researchers claim are “all you need”?
- What can this phenomenon tell us about academic culture and the dynamics of scientific communication?

2 Methodology

2.1 Data Collection

We queried the arXiv API for all papers containing the phrase “all you need” in their titles. The search was conducted on December 2nd, 2025 and returned 717 papers spanning 2009–2025. For each paper, we extracted the title, publication date, and the arXiv category.

2.2 Pattern Analysis

To analyze the structure of titles, we identified the words preceding the verb in the canonical “X [Is/Are/Is Not/etc.] All You Need” format. This yielded 499 matching papers (69.6% of the corpus). We experimented with extracting subjects following the phrase (i.e., “All You Need Is X”) but found only 21 entries with no repeated subjects.

3 Results

3.1 The Pre-Transformer and Early Transformer Eras

Table 1 presents all arXiv preprints with “All You Need” in their titles published before 2019. This chronological listing reveals the pattern’s sparse early usage and subsequent explosion following the Transformer paper.

Notably, the first instance of “All You Need” appeared in high energy physics (HEP) [5], predating the machine learning adoption by six years. The second [6] and third [7] preprints address neural network initialization, notably with the latter explicitly referencing the former. The fourth paper is the “Attention Is All You Need” itself [1]. The fifth manuscript, named, ironically, “CNN Is All You Need” [8] is a machine translation paper and appears to be the first work with a name intentionally referencing Vaswani et al. [1]. Interestingly, note that Gröndahl et al. [9] appear to be referencing the Beatles, despite being published after Vaswani et al. [1].

Table 1: Complete list of “All You Need” arXiv preprints published before 2019.

#	Title	Year	Citation
1	All you need is N: Baryon spectroscopy in two large N limits	2009	[5]
2	All you need is a good init	2015	[6]
3	All You Need is Beyond a Good Init: Exploring Better Solution for Training Extremely Deep Convolutional Neural Networks with Orthonormality and Modulation	2017	[7]
4	Attention Is All You Need	2017	[1]
5	CNN Is All You Need	2017	[8]
6	Diversity is All You Need: Learning Skills without a Reward Function	2018	[10]
7	MemGEN: Memory is All You Need	2018	[11]
8	Cross-lingual Argumentation Mining: Machine Translation (and a bit of Projection) is All You Need!	2018	[12]
9	All You Need is "Love": Evading Hate Speech Detection	2018	[9]
10	Bytes are All You Need: End-to-End Multilingual Speech Recognition and Synthesis with Bytes	2018	[13]

Figure 1: Publication Frequency Over Time

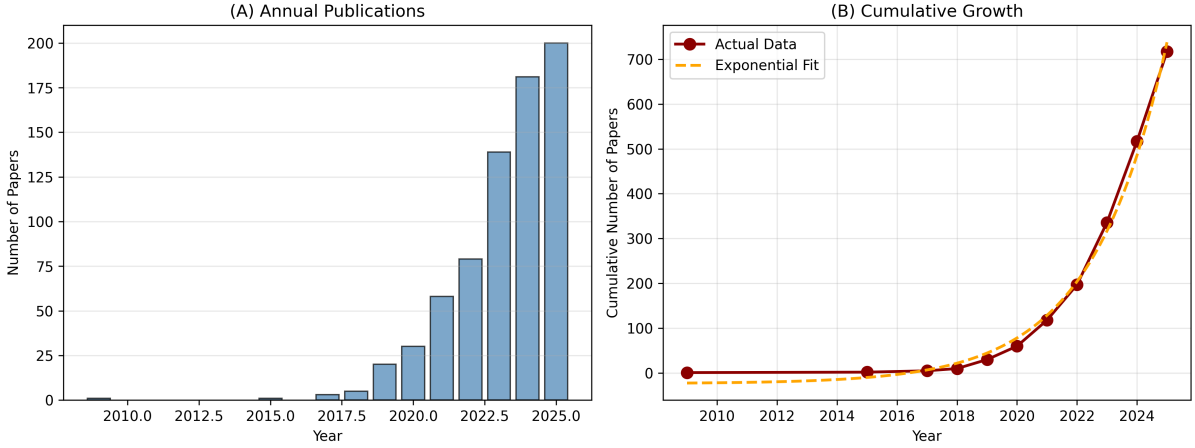


Figure 1: Left: Annual publication counts showing accelerating growth post-2017. Right: Cumulative papers with exponential fit.

3.2 Exponential Temporal Dynamics - Catching Fire

Figure 1 presents the annual distribution and cumulative growth of “All You Need” papers. The publication of Vaswani et al. [1] marked an inflection point, after which growth became approximately exponential. The year 2025 alone accounts for 200 papers (27.9% of all papers collected), with the trend showing no signs of saturation at the time of data collection. Fitting an exponential model to the cumulative data yields an excellent fit ($R^2 > 0.994$), suggesting the meme’s spread follows dynamics similar to other viral phenomena.

3.3 Spread to Other Categories

As shown in Figure 2 and the main category distribution in Table 2, the proliferation of the “All You Need” meme is overwhelmingly concentrated in computer science (87.7% of all titles). The subfield breakdown in Table 3 expectedly demonstrates the dominance of artificial intelligence-adjacent areas with Computer Vision, Machine Learning, and NLP collectively accounting for the majority of usages.

Despite this strong centralization, the pattern has clearly begun to diffuse outward. Table 4, which lists first appearances by category, shows that over time the naming convention has emerged in areas as diverse as quantitative biology, condensed matter physics, astrophysics, and finance. While these incursions remain sparse compared to computer science, their presence demonstrates that the meme has crossed disciplinary boundaries and is now present in fields far removed from deep learning.

3.4 Semantic Distribution

Figure 3 presents the most frequently occurring words in the “X [Is] All You Need” pattern. Unsurprisingly, “Attention” leads with 28 occurrences—a testament both to the original paper’s influence and to subsequent researchers’ attempts to refine or challenge its claims.

Table 3: Top 20 detailed categories.

Table 2: Main category distribution.		Code	Category	Count (%)
Main Category	Count (%)	cs.CV	Computer Vision	185 (25.8%)
Computer Science	629 (87.7%)	cs.LG	Machine Learning	170 (23.7%)
Electrical Eng & Systems	32 (4.5%)	cs.CL	Computation & Language	112 (15.6%)
Quantum Physics	10 (1.4%)	cs.AI	Artificial Intelligence	30 (4.2%)
Physics	9 (1.3%)	cs.CR	Cryptography & Security	18 (2.5%)
Mathematics	7 (1.0%)	cs.IR	Information Retrieval	16 (2.2%)
Statistics	6 (0.8%)	eess.IV	Electrical Eng & Systems	15 (2.1%)
Condensed Matter	6 (0.8%)	cs.RO	Robotics	15 (2.1%)
HEP - Phenomenology	5 (0.7%)	eess.AS	Electrical Eng & Systems	13 (1.8%)
Quantitative Biology	5 (0.7%)	cs.NE	Neural & Evolutionary	13 (1.8%)
Astrophysics	3 (0.4%)	cs.SE	Software Engineering	13 (1.8%)
HEP - Lattice	2 (0.3%)	quant-ph	Quantum Physics	10 (1.4%)
Nonlinear Sciences	1 (0.1%)	cs.SD	Sound	9 (1.3%)
HEP - Experiment	1 (0.1%)	cs.AR	Computer Science	6 (0.8%)
Quantitative Finance	1 (0.1%)	hep-ph	HEP - Phenomenology	5 (0.7%)
		cs.DB	Databases	5 (0.7%)
		stat.ML	Statistics	5 (0.7%)
		cs.DC	Distributed Computing	5 (0.7%)
		cs.DS	Data Structures & Algorithms	5 (0.7%)
		cs.NI	Networking	4 (0.6%)

Table 4: First appearance of “All You Need” papers by category.

Year	Category	Paper Title
2009	HEP - Phenomenology	All you need is N: Baryon spectroscopy in two large N limits [5]
2015	Computer Science	All you need is a good init [6]
2018	Electrical Eng & Systems	Bytes are All You Need: End-to-End Multilingual Speech Recognition and Synthesis with Bytes [13]
2020	Mathematics	Scalability and robustness of spectral embedding: landmark diffusion is all you need [14]
2020	Statistics	All You Need is a Good Functional Prior for Bayesian Deep Learning [15]
2020	Quantitative Biology	Evolution Is All You Need: Phylogenetic Augmentation for Contrastive Learning [16]
2021	Physics	All you need is time to generalise the Goman-Khrabrov dynamic stall model [17]
2022	Nonlinear Sciences	Physics-Enhanced Bifurcation Optimisers: All You Need Is a Canonical Complex Network [18]
2022	Condensed Matter	Microscopy is All You Need [19]
2023	Quantum Physics	All you need is spin: SU(2) equivariant variational quantum circuits based on spin networks [20]
2023	HEP - Lattice	Equivariant Transformer is all you need [21]
2023	HEP - Experiment	Two Watts is All You Need: Enabling In-Detector Real-Time Machine Learning for Neutrino Telescopes Via Edge Computing [22]
2024	Astrophysics	Five parameters are all you need (in Λ CDM) [23]
2024	Quantitative Finance	MoA is All You Need: Building LLM Research Team using Mixture of Agents [24]

Figure 2: Distribution by Field/Subspecialty

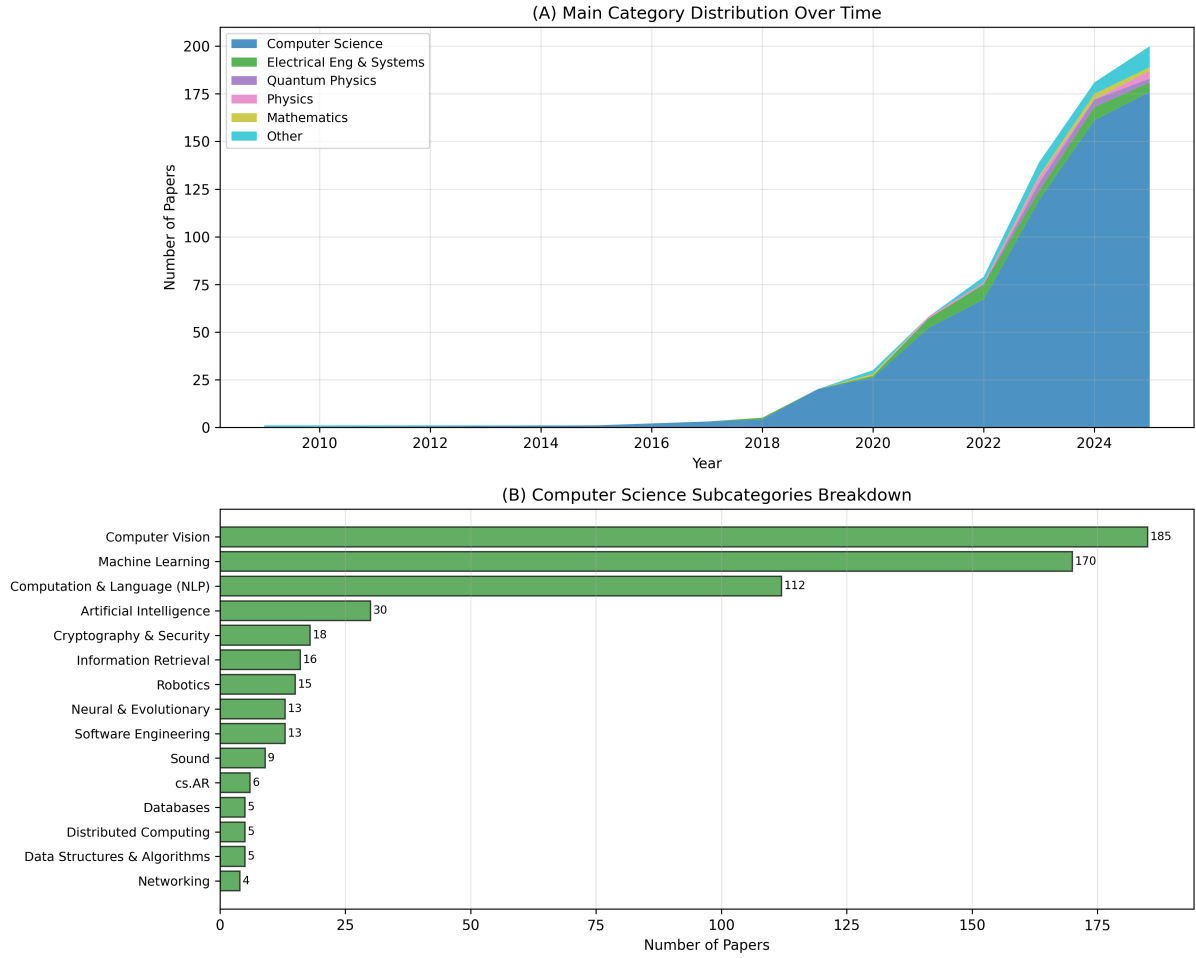


Figure 2: Top: Category distribution over time. Bottom: Breakdown of Computer Science subcategories.

Figure 3: "X Is/Are (Not) All You Need" Pattern Analysis

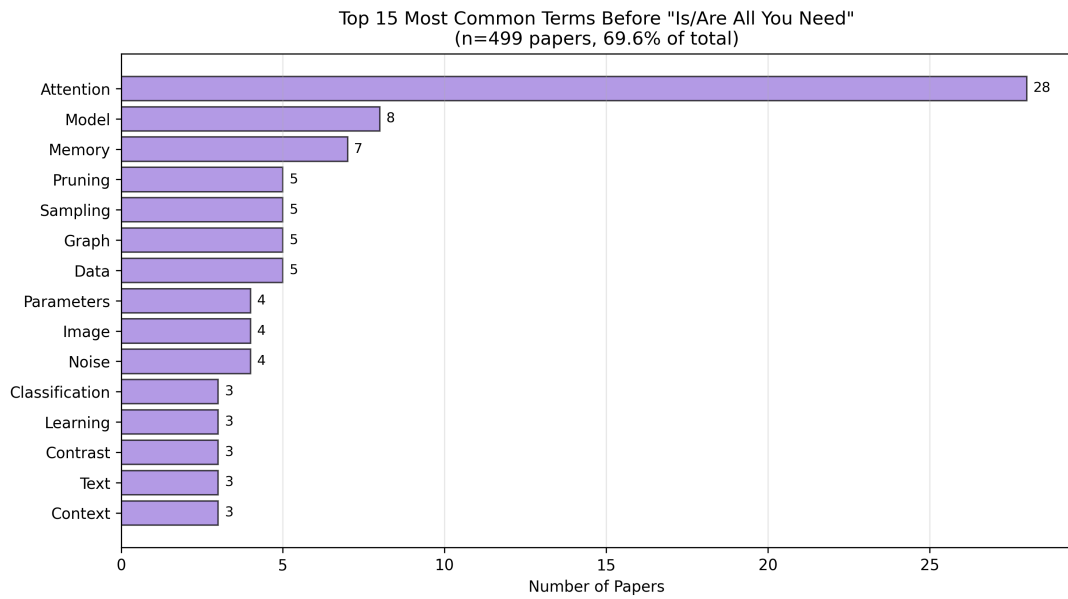


Figure 3: Most common single-word subjects in the "X [Is] All You Need" pattern.

4 Discussion

4.1 The Anatomy of an Academic Meme

The concept of the meme, as originally articulated by Dawkins [25], describes a unit of cultural transmission that propagates through imitation, analogous to genes in biological evolution. Dawkins defined memes as "tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches" that replicate by "leaping from brain to brain."

The "All You Need" pattern exhibits the three characteristics Dawkins identified as essential for successful replicators: longevity (the pattern has persisted for over eight years), fecundity (717 instances and counting), and copying fidelity (the core structure remains remarkably stable across variations). By adopting the format, authors align their work with a lineage of influential research while simultaneously claiming a degree of boldness. The structure's flexibility—any noun or phrase can precede "Is All You Need"—makes it applicable to nearly any contribution across most domains.

4.2 Comparison with Other Scientific Naming Conventions

The "All You Need" pattern is not the first formulaic titling convention to achieve widespread adoption in academic literature. Mathematics and physics have long favored the "On X" construction—exemplified by Turing's foundational "On Computable Numbers" [26] and Einstein's "On the Electrodynamics of Moving Bodies" [27]—a convention dating to classical treatises, perhaps signaling scholarly modesty. The machine learning literature has developed its own ecosystem of such patterns; for example "Towards X" signals incremental progress with a clear ambitious end-goal in mind [28].

What distinguishes the "All You Need" pattern is its implicit promise of parsimony—the claim that complexity can be reduced to a single sufficient element. While "Towards" hedges with epistemic humility and "On" maintains disciplinary neutrality, "All You Need" asserts completeness. This boldness may explain both its appeal and its ironic overuse: in a competitive attention economy, modesty is a disadvantage.

4.3 Implications for Scientific Communication

The exponential growth of "All You Need" pattern occurs against a backdrop of intensifying competition for attention in academic publishing. The "publish or perish" culture—a phrase with roots extending back to at least the late 1920s [29, 30]—has created structural incentives that reward visibility alongside—or sometimes instead of—substance. Publications attract funding, and funding enables further publication, creating feedback loops that favor attention-grabbing presentations of research. Visibility increasingly influences success. This is especially true in the modern world, and in the field, where much of the dissemination of science bypasses the formal academic peer-review and happens via self-submitted preprint sharing.¹

The pattern's success may also reflect the increasing role of social media in scientific discourse. Memorable titles serve legitimate functions: they aid recall and facilitate discussion. Preprints shared on Twitter (now X) or Reddit benefit from memorable titles that can be easily quoted and recognized. In this environment, "Attention Is All You Need" functions not merely as a title but as a slogan—a compression of the paper's contribution into a form optimized for viral transmission.

¹In addition to granting the work credibility by publishing it, the publishing venues also serve as a desensationalization filter, imposing uniform standards. Preprint and published titles often vary dramatically—e.g., the author's "Metformin: We need to either put it in our drinking water or rethink how we study it" [31] became "Exploration of Residual Confounding in Analyses of Associations of Metformin Use and Outcomes in Adults With Type 2 Diabetes" [32] upon publication.

4.4 Limitations

Our analysis is limited to arXiv and thus excludes papers published only in traditional venues. We intentionally omitted author level analysis.

4.5 A Self-Referential Note

The author acknowledges the irony of contributing yet another “All You Need” paper to the literature, albeit in negated form. Hofstadter [33] devoted considerable attention to such “strange loops”—self-referential structures in which, by moving through a hierarchical system, one unexpectedly returns to one’s starting point. By analyzing the “All You Need” phenomenon with a title that participates in it (albeit does not include it due to temporal causation), this paper necessarily adds itself to the dataset it describes, achieving a modest Hofstadterian recursion.

5 Conclusion

We have presented an empirical analysis of the “All You Need” naming pattern in machine learning and related fields. Our findings document exponential growth from 3 papers pre-2017 to 717 by late 2025. The pattern is showing no signs of saturation with 200 papers already in 2025 alone. “Attention” remains the most claimed necessity, appearing in 28 titles. The conclusion of whether the persistence of this naming convention represents a harmless quirk of academic culture or a symptom of declining originality and rising sensationalism-to-subject ratio is left up to the readers and future researchers. Perhaps, they will use more original titles than this work, since, in the end, good science with an appropriate but memorable title is all one needs.

Remark

This work is intentionally written to be light-hearted but thought-provoking. It is in part inspired by Pretraining on the Test Set Is All You Need [34]. The author deeply respects the researchers conducting work in the field of artificial intelligence, and appreciates a good reference and the use of humor. There has been a number of breakthrough works with the names following the “All You Need” convention, especially in the field of tiny models.

Reproducibility

The code and data is available at https://github.com/alyakin314/All_You_Need.

Acknowledgments

The code and manuscript of this work were drafted with the help of Anthropic Claude Opus 4.5.

References

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008, 2017.
- [2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya

- Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.
- [3] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 2021.
 - [4] The Beatles. All you need is love, 1967. Single released July 7, 1967. Parlophone Records.
 - [5] Aleksey Cherman, Thomas D Cohen, and Richard F Lebed. All you need is N: Baryon spectroscopy in two large N limits. *arXiv preprint arXiv:0906.2400*, 2009.
 - [6] Dmytro Mishkin and Jiri Matas. All you need is a good init. In *International Conference on Learning Representations*, 2016.
 - [7] Di Xie, Jiang Xiong, and Shiliang Pu. All you need is beyond a good init: Exploring better solution for training extremely deep convolutional neural networks with orthonormality and modulation. *arXiv preprint arXiv:1703.01827*, 2017.
 - [8] Qiming Chen and Ren Wu. CNN is all you need. *arXiv preprint arXiv:1712.09662*, 2017.
 - [9] Tommi Gröndahl, Luca Pajola, Mika Juuti, Mauro Conti, and N Asokan. All you need is “love”: Evading hate-speech detection. In *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security*, pages 2–12, 2018.
 - [10] Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. *arXiv preprint arXiv:1802.06070*, 2018.
 - [11] Sylvain Gelly, Karol Kurach, Marcin Michalski, and Csaba Szepesvari. MemGEN: Memory is all you need. *arXiv preprint arXiv:1803.11203*, 2018.
 - [12] Steffen Eger, Johannes Daxenberger, Christian Stab, and Iryna Gurevych. Cross-lingual argumentation mining: Machine translation (and a bit of projection) is all you need! In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 831–844, 2018.
 - [13] Bo Li, Yu Zhang, Tara Sainath, Yonghui Wu, and William Chan. Bytes are all you need: End-to-end multilingual speech recognition and synthesis with bytes. *arXiv preprint arXiv:1811.09021*, 2018.
 - [14] Chao Shen and Hau-Tieng Wu. Scalability and robustness of spectral embedding: landmark diffusion is all you need. *arXiv preprint arXiv:2001.00801*, 2020.
 - [15] Ba-Hien Tran, Simone Rossi, Dimitrios Milios, and Maurizio Filippone. All you need is a good functional prior for bayesian deep learning. *arXiv preprint arXiv:2011.12829*, 2020.
 - [16] Amy X Lu, Amy X Lu, and Alan Moses. Evolution is all you need: Phylogenetic augmentation for contrastive learning. *arXiv preprint arXiv:2012.13475*, 2020.
 - [17] Fatma Ayancik and Karen Mulleners. All you need is time to generalise the goman-khrabrov dynamic stall model. *arXiv preprint arXiv:2110.08516*, 2021.

- [18] Marvin Syed and Natalia G Berloff. Physics-enhanced bifurcation optimisers: All you need is a canonical complex network. *arXiv preprint arXiv:2207.11256*, 2022.
- [19] Sergei V Kalinin et al. Microscopy is all you need. *arXiv preprint arXiv:2210.06526*, 2022.
- [20] Richard D P East, Guillermo Alonso-Linaje, and Chae-Yeun Park. All you need is spin: $Su(2)$ equivariant variational quantum circuits based on spin networks. *arXiv preprint arXiv:2309.07250*, 2023.
- [21] Akio Tomiya and Yuki Nagai. Equivariant transformer is all you need. *arXiv preprint arXiv:2310.13222*, 2023.
- [22] Miaochen Jin, Yushi Hu, and Carlos A Argüelles. Two watts is all you need: Enabling in-detector real-time machine learning for neutrino telescopes via edge computing. *arXiv preprint arXiv:2311.04983*, 2023.
- [23] Paulo Montero-Camacho et al. Five parameters are all you need (in λ cdm). *arXiv preprint arXiv:2405.13680*, 2024.
- [24] Sandy Chen, Leqi Zeng, Abhinav Raghunathan, Flora Huang, and Terrence C Kim. Moa is all you need: Building llm research team using mixture of agents. *arXiv preprint arXiv:2409.07487*, 2024.
- [25] Richard Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, 1976.
- [26] Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 1936.
- [27] Albert Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 322(10):891–921, 1905. English translation: “On the Electrodynamics of Moving Bodies”.
- [28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016. URL <https://arxiv.org/abs/1506.01497>.
- [29] Clarence Marsh Case. Scholarship in sociology. *Sociology and Social Research*, 12:323–340, 1927–1928.
- [30] Guillaume Cabanac. What is the primordial reference for ...?—redux. *Scientometrics*, 114(2):481–488, 2018. doi: 10.1007/s11192-017-2595-4.
- [31] Mike Powell, Callahan Clark, Anton Alyakin, Joshua T. Vogelstein, and Brian Hart. Metformin: We need to either put it in our drinking water or rethink how we study it, 2021. URL <https://www.medrxiv.org/content/10.1101/2021.09.15.21263634v1>. Preprint.
- [32] Mike Powell, Callahan Clark, Anton Alyakin, Joshua T. Vogelstein, and Brian Hart. Exploration of residual confounding in analyses of associations of metformin use and outcomes in adults with type 2 diabetes. *JAMA Network Open*, 5(11):e2241505, November 2022. doi: 10.1001/jamanetworkopen.2022.41505. URL <https://doi.org/10.1001/jamanetworkopen.2022.41505>.
- [33] Douglas R. Hofstadter. *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books, New York, 1979.
- [34] Rylan Schaeffer. Pretraining on the test set is all you need, 2023. URL <https://arxiv.org/abs/2309.08632>.