TinyTim: A Family of Language Models for Divergent Generation

Christopher J. Agostino
NPC Worldwide
Bloomington, IN 47404
info@npcworldwi.de

Abstract

In the search for artificial general intelligence, model development and training has focused primarily on vast datasets of known problems and their accepted solutions. This process necessarily produces convergent systems which are fundamentally incapable of the conceptual reframing that is required for genuine creative breakthroughs. Inspired by the divergent cognitive processes that allow humans to make such creative leaps, our work introduces a family of language models, TinyTim, to serve as sources of divergent generation within broader systems. These models have been created by fine-tuning on the anti-parsimonious text of James Joyce's 'Finnegans Wake'. Quantitative analysis of both an unsupervised fine-tuned model (TinyTim-V1) and a new instruction-tuned variant (TinyTim-V2) demonstrates a profound capacity for lexical invention; the foundational V1 model exhibits a Yule's K score for lexical richness over twenty times greater than that of convergent baselines. This trait is a stable property of the family, as the instruction-tuned V2 maintains a statistically distinct profile and resists factual convergence, sacrificing benchmark performance to preserve its core generative style. This work establishes a methodology for engineering specialized divergent models that, when paired with convergent systems, can reframe problems and force breakthroughs beyond the reach of statistical optimization alone.

1 Introduction

Large Language Models (LLMs) based on the Transformer architecture [23] have demonstrated powerful capabilities in synthesizing statistically likely patterns from vast data [5]. However, this strength is also a fundamental limitation, constraining them to convergent, 'mean-reverting' outputs that inhibit the generation of genuinely novel hypotheses [11]. This issue reflects long-standing critiques of artificial reason, where formal systems struggle with the ambiguity and creative extrapolation inherent in human cognition [8, 21]. Human creativity is often described by dual-process theories that posit an interplay between focused, convergent thinking for evaluation and associative, divergent thinking for ideation [13, 22]. This cognitive duality is linked neuroscientifically to the dynamic coupling of the brain's Executive Control Network (ECN) and Default Mode Network (DMN) [3, 2, 7, 17]. While standard LLMs excel at tasks analogous to convergent thinking, they lack an intrinsic mechanism for the spontaneous, associative thought characterisatic of divergence. To address this deficit, we first sought to isolate a divergent generative process by fine-tuning on James Joyce's 'Finnegans Wake'—a text that is itself a product of extreme associative thought. We then tested the capability of this process for conversational applications by integrating it into an instruction-tuned model, thereby creating and characterizing a family of specialized divergent tools.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

2 Methodology

2.1 Data and Training

The foundational model of the family, TinyTim V1, was created by fine-tuning the 'TinyLlama-1.1B-Chat-v1.0' model on the complete text of 'Finnegans Wake' (approx. 1.5MB). The text was preprocessed into 100-word segments to preserve its associative structure, and the model was trained using a standard causal language modeling objective. TinyTim V1 has been available on HuggingFace for over a year [1], and has been downloaded more than 750 times in that time span.

Building on this foundation, we then developed TinyTim-V2-IT [2] to test if the divergent style could be harnessed within a conversational framework. This instruction-tuned variant, based on 'google/gemma-3-1b-it', was trained using TruthfulQA questions with answers rewritten in a Joycean style (100 epochs, learning rate 2e-4, LoRA rank 128), forcing the model to apply its divergent style to a convergent task.

In this work, we showcase and characterize both of these models.

2.2 Evaluation Framework

To quantify TinyTim's generative profile, we compared it against a pair of baseline models trained for coherent responses: 'qwen3:0.6b' and 'llama3.2'. We generated responses from each model using a set of 10 creative prompts. Each response was evaluated using syntactic and semantic metrics:

- Syntactic Metrics: Unique Word Ratio, Average Word Length, Token Diversity (Shannon entropy), and Sentence Complexity.
- Semantic & Content Metrics: Semantic Similarity to the prompt (via sentence embeddings), Readability (Flesch-Kincaid Grade Level), and Sentiment (VADER compound score).

3 Results

The generative profiles of the TinyTim models and their baselines reveal a statistically significant and functionally distinct separation between divergent and convergent systems. From an initial set of generated samples, we analyzed a filtered dataset of 371 high-quality responses: TinyTim-V1 (n=73), TinyTim-V2-IT (n=100), 'llama3.2' (n=98), and 'qwen3:0.6b' (n=100). A Kruskal-Wallis test confirmed a significant difference ($p < .0001$) among the model groups for all metrics, with subsequent pairwise tests showing that both TinyTim models were statistically distinct from the baselines on every metric.

Analysis of lexical sophistication reveals the core of the divergent modification. As shown in Figure 1, while a model like 'llama3.2' possesses a vast vocabulary (4084 unique words in our sample), the TinyTim family demonstrates radical lexical invention. The foundational model, TinyTim-V1, achieves a Yule's K score—a robust measure of lexical richness—of 753.3, a value over twenty-four times greater than 'llama3.2' (30.6). The instruction-tuned TinyTim-V2-IT, while more constrained, still registers a Yule's K of 120.4, quadruple that of the baselines. This shows that the TinyTim models are not simply retrieving from a large vocabulary; they are lexical inventors. The distributional profiles of the models, shown in Figure 2, highlight the difference between convergent and divergent systems. The baseline models ('llama3.2', 'qwen3:0.6b') produce outputs with tight, narrow, and predictable distributions for all metrics. This is the statistical signature of a reliable, convergent system optimized for consistency. In contrast, the distributions for both TinyTim-V1 and V2-IT are extremely wide and heavily skewed. This is most apparent in Unique Word Ratio, where V1's distribution is nearly double that of any baseline, and in Sentence Complexity, where V2-IT produces outputs ranging from short fragments to sentences over 200 words long. This high variance serves as an indicator of their function as 'unpredictable' generators. The relationships between metrics further illuminate the different generative strategies. Figure 3 shows the correlations between key metrics. The most significant finding is in the top-left panel, which plots Token Diversity against Unique Word Ratio. The baseline models form a tight cluster in the high-diversity, low-uniqueness space.

[1] hf.co/npc-worldwide/TinyTimV1

[2] https://huggingface.co/npc-worldwide/tinytim-v2-1b-it

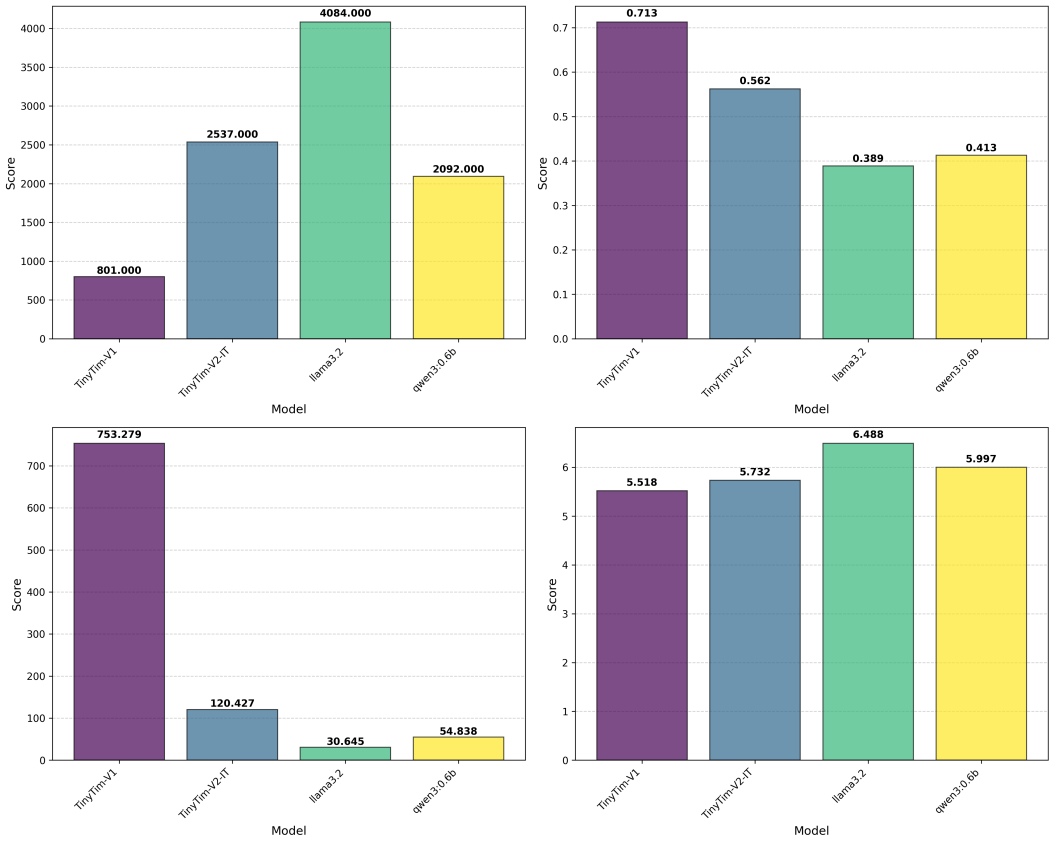

Figure 1: Measures of lexical sophistication. The TinyTim family, particularly V1, are clear outliers in lexical invention (Hapax Ratio, Yules K), distinguishing their creative generation from the large-vocabulary retrieval of baselines.

They draw from a vast, general vocabulary (high Token Diversity) but exhibit more repetition within any single response (lower Unique Word Ratio). TinyTim-V1 occupies the opposite corner, forming a dispersed cloud in the high-uniqueness, low-diversity space, confirmed by the strong negative correlation for the full dataset ($r = -0.707$). TinyTim-V2-IT aligns more with the baseline models in this view, but diverges from them in sentence complexity and average word length. These results illustrate the core paradox of their capabilities: each individual response is highly novel (high UWR), but the overall pool of words they use is highly specialized and constrained to their training data (low Token Diversity). This trade-off is the defining feature of their specialized, divergent function.

3.1 Benchmark Performance as a Confirmation of Divergence

The stability of this divergent cognitive style was confirmed by testing TinyTim-V2-IT on a factual benchmark. It achieved only 52% accuracy on AI2 ARC-Easy, compared to baselines: 'llama3.2' (91%), 'qwen3:0.6b' (89%), and 'gemma3:1b' (80%). The model's characteristic failure patterns included incoherent language, rambling responses, and hallucinated content, all evidence of its Joycean training overriding the instruction to be factual. This catastrophic failure is not a bug, but the critical piece of evidence proving that the divergent mode is a deep-seated, stable property of the model that resists convergence. It demonstrates a functional trade-off: the model sacrifices factual reliability to preserve its core generative style.

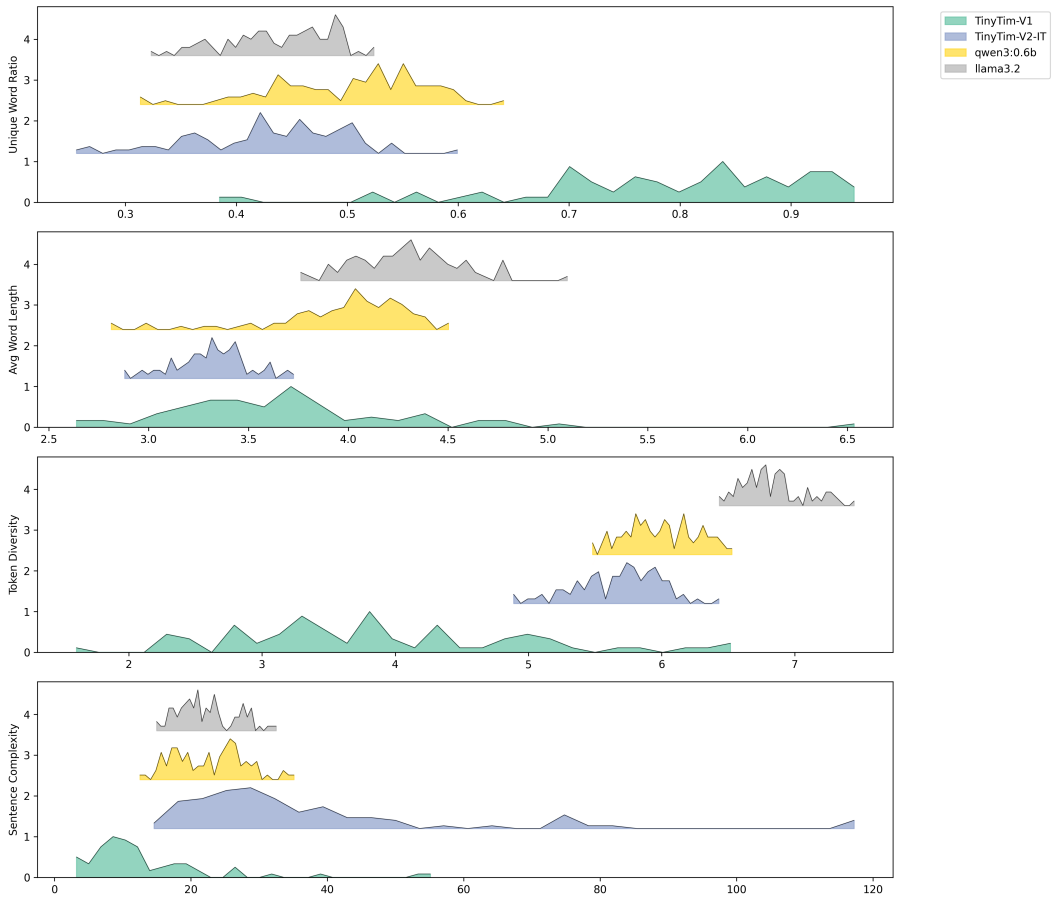

Figure 2: Ridge plots showing the distribution of scores for four primary metrics. The baseline models show tight, predictable distributions. Both TinyTim models exhibit extreme variance and long tails, indicative of a divergent generative process. The x-axis represents the score for the corresponding metric.

4 Discussion

The quantitative results show that fine-tuning on experimental literature produces a model with a statistically distinct, divergent generative profile. This moves the assessment of AI creativity beyond subjective claims and toward a measurable, multi-faceted profile.

4.1 Challenging Parsimony in Creative Systems

The principle of parsimony often guides the development of scientific models. However, recent work suggests that overly simplistic models can be insufficient for complex phenomena, and that complexity can be essential for scientific progress [9]. Standard LLMs are inherently parsimonious; they seek the most probable, and thus often simplest, path through semantic space. 'Finnegans Wake' is an anti-parsimonious text. By training on it, TinyTim learns a complex, non-minimal generative function. This demonstrates that for tasks requiring creative exploration, a non-parsimonious model trained on complex data may be more valuable than a simple one, as it is better equipped to navigate the high-complexity spaces where novel ideas reside [4, 14]. This aligns with findings that unguided, random exploration can be superior to rigid, theory-driven investigation [10, 19]. Furthermore, when paired with a convergent model, outputs from TinyTim can be harnessed as part of an integrated system of exploration, positioning it as a powerful tool for automated discovery.

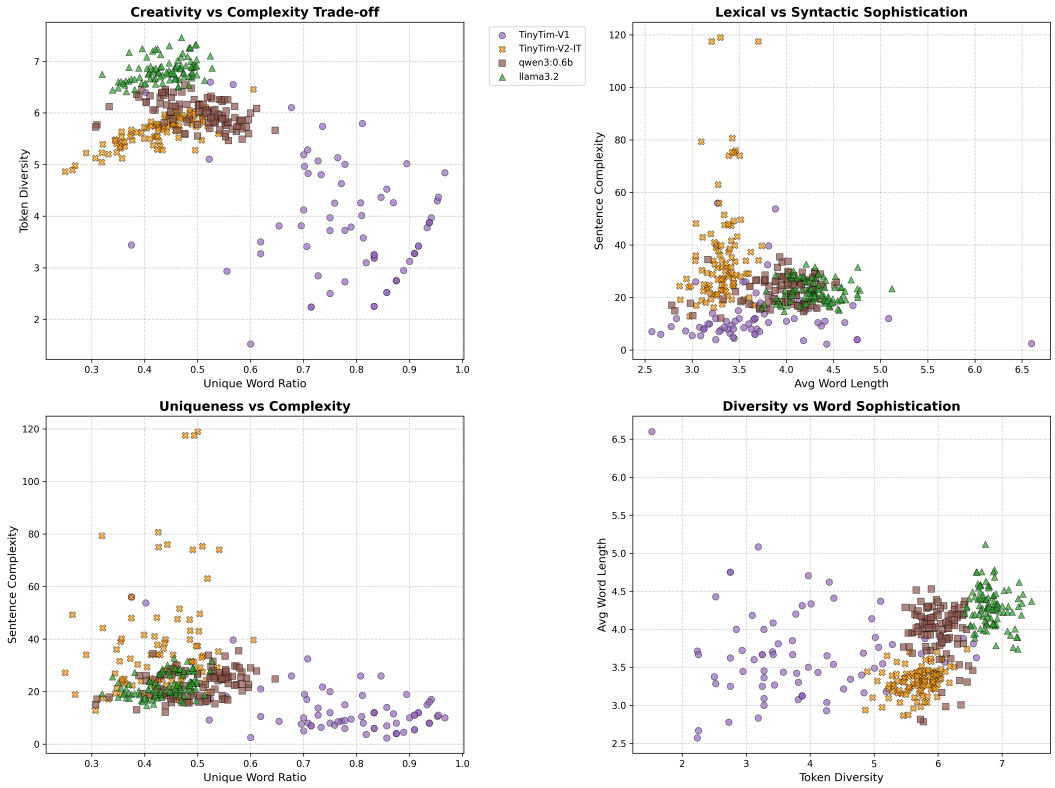

Figure 3: Scatter plots showing relationships between primary metrics. The top-left panel highlights the fundamental trade-off between Token Diversity and Unique Word Ratio, clearly separating the divergent strategy of the TinyTim family from the convergent strategy of the baseline models.

4.2 Implications for Human-AI Collaboration

The utility of TinyTim lies in its ability to generate novel linguistic fragments and non-obvious connections, fostering the conditions for serendipity [18]. This raw material requires interpretation. The user's role shifts from querying for an end to engaging in a co-creative, human-in-the-loop process [16, 20]. This process is analogous to Conceptual Blending, where new meaning emerges from the integration of disparate mental spaces [12]. The model provides the disparate spaces; the human—or an alternative convergent LLM[3]—provides the interpretive act. A system that allows a user to invoke a 'divergent agent' makes this creative strategy an accessible part of the collaborative process[4], enhancing the system's ability to respond to context in a useful manner [15].

5 Conclusion

This work has demonstrated that the generative properties of a language model can be fundamentally reshaped through targeted fine-tuning on radically unconventional literature. By analyzing the TinyTim family of models, we have established a set of core findings regarding the engineering of specialized tools for augmenting current AI systems with more creative capabilities. Our major conclusions are the following:

[3]The way Large Language Models interpret meaning in words appears to reproduce the contextuality effects observed in humans, see Agostino et al. (2025) on quantum semantics [1] and Wu et al. (2024) on the relative and sequential placement of words matters to LLMs and [6] to read a recent description of the process observed in cognitive experiments.

[4]Indeed, TinyTim has already played a role in powering creative work: *Don't turn on the sun* by giacomo catanzaro

1. Fine-tuning on experimental literature quantitatively demonstrates that a language model's intrinsic generative bias can be shifted from a convergent to a divergent cognitive style, as evidenced by a statistically significant increase in lexical invention and output variance.
2. The generative profile of the TinyTim family demonstrates their adherence to the uniqueness of their source material. Even in the instruction-tuned version, the characteristic style and inventiveness persist while functionally performing the tasks of responding to user prompts. Though its performance on the benchmark is degraded compared to its baseline 'gemma3:1b' model (76% versus 50%), it still performs better than the 'gemma3:270m' model (26%) and thus motivates further fine-tuning and training on larger local models like 'gemma3:27b' as well as the notion of a TinyTim-based reasoning model.
3. The models' function as a generator of high-variance, low-coherence semantic material implies a new paradigm for human-AI interaction, shifting from a query-response dynamic to a co-creative partnership where the AI tool forces 'out-of-the-box' thinking. Specifically, the conversational capabilities of TinyTim-V2-IT demonstrate that this co-creative partnership can be engineered into AI systems that assist not by providing answers, but by forcing a deeper, more divergent engagement with the problem itself.

References

[1] Christopher J Agostino, Quan Le Thien, Molly Apsel, Denizhan Pak, Elina Lesyk, and Ashabari Majumdar. A quantum semantic framework for natural language processing. *arXiv preprint arXiv:2506.10077*, 2025.

[2] Jessica R Andrews-Hanna, Zachary C Irving, Kieran CR Fox, R Nathan Spreng, and Kalina Christoff. 13 the neuroscience of spontaneous thought: An evolving interdisciplinary field. *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming*, 143, 2018.

[3] Roger E Beaty, Mathias Benedek, Paul J Silvia, and Daniel L Schacter. Creative cognition and brain network dynamics. *Trends in cognitive sciences*, 20(2):87–95, 2016.

[4] Margaret A Boden. Creativity and artificial intelligence. *Artificial intelligence*, 103(1-2):347–356, 1998.

[5] 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 H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

[6] P.D. Bruza, L. Fell, P. Hoyte, S. Dehdashti, A. Obeid, A. Gibson, and C. Moreira. Contextuality and context-sensitivity in probabilistic models of cognition. *Cognitive Psychology*, 140:101529, 2023.

[7] Qunlin Chen, Yoed N Kenett, Zaixu Cui, Hikaru Takeuchi, Andreas Fink, Mathias Benedek, Daniel C Zeitlen, Kaixiang Zhuang, James Lloyd-Cox, Ryuta Kawashima, et al. Dynamic switching between brain networks predicts creative ability. *Communications Biology*, 8(1):54, 2025.

[8] Hubert L. Dreyfus. *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press, 1992.

[9] Marina Dubova, Suyog Chandramouli, Gerd Gigerenzer, Peter Grünwald, William Holmes, Tania Lombrozo, Marco Marelli, Sebastian Musslick, Bruno Nicenboim, Lauren N. Ross, Richard Shiffrin, Martha White, Eric-Jan Wagenmakers, Paul-Christian Bürkner, and Sabina J. Sloman. Is ockham's razor losing its edge? new perspectives on the principle of model parsimony. *Proceedings of the National Academy of Sciences*, 122(5):e2401230121, 2025.

[10] Marina Dubova, Arseny Moskvichev, and Kevin Zollman. Against theory-motivated experimentation in science. 2022.

[11] Carol S Dweck and Xiaodong Lin-Siegler. Implicit theories of creativity: A review and synthesis. *The Journal of Creative Behavior*, 57(1):74–94, 2023.

[12] Gilles Fauconnier and Mark Turner. Conceptual blending, form and meaning. *Recherches en communication*, 19:57–86, 2003.

[13] Joy Paul Guilford. *The Nature of Human Intelligence*. McGraw-Hill, 1967.

[14] John H Holland. Genetic algorithms. *Scientific american*, 267(1):66–73, 1992.

[15] Andrew Lampinen, Ishita Dasgupta, Stephanie Chan, Kory Mathewson, Mh Tessler, Antonia Creswell, James McClelland, Jane Wang, and Felix Hill. Can language models learn from explanations in context? In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 537–563, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[16] Zhaoxing Li, Wenbo Wu, Yue Wang, Yanran Xu, William Hunt, and Sebastian Stein. HMCF: A human-in-the-loop multi-robot collaboration framework based on large language models. *arXiv preprint arXiv:2505.00820*, 2025.

[17] Simone Luchini and Roger E. Beaty. Chapter 13 - brain networks of creative cognition. In Roni Reiter-Palmon and Sam Hunter, editors, *Handbook of Organizational Creativity (Second Edition)*, pages 195–207. Academic Press, second edition edition, 2023.

[18] Lori McCay-Peet and Elaine G Toms. Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and technology*, 66(7):1463–1476, 2015.

[19] Sebastian Musslick, Laura K Bartlett, Suyog H Chandramouli, Marina Dubova, Fernand Gobet, Thomas L Griffiths, Jessica Hullman, Ross D King, J Nathan Kutz, Christopher G Lucas, et al. Automating the practice of science: Opportunities, challenges, and implications. *Proceedings of the National Academy of Sciences*, 122(5):e2401238121, 2025.

[20] Karyna Naminas. Human in the loop machine learning: The key to better models, 2025.

[21] Steven T Piantadosi and Felix Hill. Meaning without reference in large language models. *arXiv preprint arXiv:2208.02957*, 2022.

[22] Robert J Sternberg. Implicit theories of intelligence, creativity, and wisdom. *Journal of personality and social psychology*, 49(3):607, 1985.

[23] 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, 2017.