

The Anatomy of Synthetic Writing: An Investigation into Linguistic Markers, Memory Cognition, and Strategies for Humanizing AI-Generated Text

Pietro Impagliazzo

Independent Researcher

amazing@pietro.works

April 2026

Abstract

The emergence of Large Language Models (LLMs) has transformed content production into a statistical processing activity, but this transition has brought with it a phenomenon that researchers and linguists identify as the “stylistic fingerprint” of the machine. While models like GPT-4 and Llama 3 display near-flawless grammatical fluency, they frequently fail to replicate the emotional density, rhythmic variability, and contextual depth inherent in human writing. This report analyzes the fundamental divergences between synthetic and authorial text, exploring metrics such as perplexity and burstiness, the distinction between semantic and episodic memories, and the impact of flat syntactic chains on the perception of authenticity.

1. The Phenomenon of Predictability and the Perplexity Metric

At the heart of the distinction between human and AI writing lies the mathematical logic of predictability. Artificial intelligence operates through probabilistic models that seek the next most likely word (token) in a sequence based on vast training corpora [1, 2]. This process results in texts with low *perplexity*, a technical term that quantifies how “surprised” a language model is when processing a sequence of words [3, 4].

The perplexity $PP(W)$ of a word sequence W can be expressed mathematically as the N -th root of the inverse probability of the sequence:

$$PP(W) = P(w_1, w_2, \dots, w_N)^{-\frac{1}{N}} \quad (1)$$

For AI, the goal is to minimize this metric, which invariably leads to the selection of safer, more frequent linguistic paths [4, 5]. In contrast, human writing is inherently more “surprising”: human authors often use rare metaphors, regional idioms, and grammatical structures that defy statistical expectations, resulting in high perplexity [6]. Measured against GPT-2, human-authored text registers a perplexity of 12.38, while beam-search decoding collapses to 1.48 — an order of magnitude flatter than the writing it imitates [6]. When AI selects obvious words, the text becomes monotonous; when it tries to be unpredictable without context, it becomes confusing [6]. Human balance, however, resides in the ability to inject lexical novelty without sacrificing clarity [7].

2. Burstiness and the Heartbeat of Prose

While perplexity deals with word choice, *burstiness* refers to the rhythm and cadence of sentences. In its original technical sense the term describes the clustering of lexical items within a document, modelled as a mixture of Poisson distributions [8, 9, 10]; in the stylistic sense adopted here it denotes the variance of sentence length across a text. Human writing is characterized by a dynamic alternation between short, incisive sentences and long, complex periods [11]. This variation creates a narrative flow that maintains reader engagement. In contrast, AI-generated texts tend to exhibit low burstiness: across 13,371 news articles, LLMs concentrated their output in the 10 to 30 token band while human writers produced a markedly wider and more scattered distribution of sentence lengths — an effect the authors term *register leveling* [11].

Detection research confirms that these statistical regularities are the operative signal. Per-token likelihood and probability curvature are sufficient to separate machine from human text without any supervised training [12, 13], and detectors consistently flag

low-perplexity prose. However, the signal is fragile. In a controlled trial of six commercial detectors, a prompt instructing the model to vary sentence length — the manual introduction of burstiness — cut mean detector accuracy from 39.5% on unmodified AI text to 15.9% [14]. Human manual editing produced a comparable collapse, from 74% to 42% [15], and paraphrasing drove DetectGPT from 70.3% to 4.6% at a fixed 1% false-positive rate [16]. This suggests that rhythmic uniformity is one of the strongest indicators of non-human authorship, and also one of the easiest to erase.

3. Excessive Parallelism and Flat Syntactic Chains

One of the most common issues in AI-generated text is the excessive use of grammatical parallelism. Although parallelism is a legitimate rhetorical device for organizing ideas, AI tends to use it as a structural crutch, repeating “verb + object” patterns in lists or descriptions. The effect is measurable: at a template length of six part-of-speech tags, 95% of model-generated film reviews contained a repeated syntactic template against 38% of human-written ones, and 90% of generated news summaries against 36% [17]. Roughly 76% of the templates found in model output can be traced to pre-training data and survive fine-tuning [17]. Examples like “*The system reduces costs, eliminates intermediaries, and accelerates processes*” illustrate this tendency toward mechanical symmetry; GPT-4o produces phrasal coordination at 1.9 times the human rate [18].

Table 1: Comparison of key metrics between human and AI writing.

Metric	Human Writing	AI Writing (Standard)	Impact on Perception
Perplexity	High (unexpected choices)	Low (probable choices)	Human text feels creative; AI feels obvious.
Burstiness	High (variable rhythm)	Low (uniform rhythm)	Human text has “music”; AI is monotonous.
Sentence Variance	Large (short vs. long)	Small (10–30 token band)	AI’s uniformity tires the reader quickly.
Structure	Organic and adaptable	Rigid and predictable	AI follows school-like intro–conclusion patterns.

3.1. The Subordination Deficit and Tree-Like Syntax

High-quality human writing utilizes *hypotaxis* — the organization of sentences through subordinate clauses that establish a hierarchy of importance and causal relationships [19, 20]. This style creates a “tree-like” structure where main ideas are supported by layers of context [21]. AI, on the other hand, frequently generates flat or paratactic syntactic chains, where ideas are presented side-by-side without clear subordination. Against

native-speaker argumentative essays, ChatGPT produced 0.368 dependent clauses per T-unit in the thesis stage against 1.161 for human writers, and 0.126 against 1.173 in the conclusion, while its use of coordinate phrases ran three to five times higher [22].

Parataxis results in sentences like: *“The sun rose. The day was hot. People went to the beach.”* Human hypotaxis allows: *“Even though the sun had just risen, the heat was already so intense that people, seeking relief, hurried toward the beach.”*

AI’s lack of subordination removes tension and ambiguity, favouring an excessively “safe” fluency that sounds artificial to experienced readers [23]. It should be noted that the equation of subordination with quality is itself contested: finite dependent clauses are more frequent in conversation than in academic prose, where complexity is carried phrasally rather than clausally [24]. The relevant contrast is therefore not subordination as such, but where each writer places structural weight.

3.2. Part-of-Speech Balance and Nominal Loading

Quantitative research using Biber features reveals that LLMs suffer from “nominal loading.” The effect is not one of raw noun counts — on news corpora, human writers actually use slightly more noun tokens than LLMs (19.69% against 17.44–17.85%) [11]. It is a matter of how those nouns are packaged. Instruction-tuned models produce nominalisations at 1.5 to 2 times the human rate, with GPT-4o reaching 2.1 times (paired Cohen’s $d = 1.23$) and present participial clauses at 5.3 times ($d = 1.38$) [18]. In argumentative writing, ChatGPT’s complex nominals per clause run at roughly double the native-speaker rate [22]. Independent replication confirms that generated text uses more nominalisations and higher sentence complexity, while human writers retain more modal and epistemic constructions — the markers that convey speaker attitude [25].

This information density in AI aims for efficiency but sacrifices the “humanity” of the text, which often requires natural redundancy and variation in references to maintain communicative flow [23, 26].

4. Semantic vs. Episodic Memory: The Lack of Lived Experience

The most profound difference between AI and humans lies not just in grammar, but in the cognitive basis of information. Artificial intelligence excels at *semantic memory*: the storage of concepts, theories, and general facts [27, 28]. It knows what an apple is but cannot recall the episode of biting into a cold apple on an autumn afternoon. Benchmarking confirms the deficit rather than assuming it: on tasks requiring the recall of events bound to particular times and places, GPT-4, Claude, Llama 3.1 and o1-mini

Table 2: Part-of-speech and stylistic frequency comparison between human and AI writing. Percentages are UPOS token frequencies from a 13,371-article news corpus [11]; rate multiples are paired human-model comparisons on the HAP-E corpus [18].

Feature	Human Use	AI Use	Stylistic Consequence
Nominalisations	Baseline	1.5–2.1× human rate	AI text becomes abstract, dense and agentless.
Nouns (tokens)	19.69%	17.44–17.85%	Density comes from packaging, not raw noun count.
Adjectives	7.58%	6.69–6.86%	Human modification is more varied and contextual.
Pronouns	5.32%	6.11–7.33%	AI leans on pronouns without anchoring reference.
Auxiliaries	3.81%	5.41–6.02%	AI favours periphrastic, hedged constructions.
Phrasal coordination	Baseline	1.9× human rate	Parallel lists replace hierarchical argument.

all degrade sharply once multiple related events or complex spatio-temporal relations are involved, even within contexts of 10k–100k tokens [32].

4.1. The Role of Episodic Memory in Authenticity

Human episodic memory is anchored in time, space, and auto-noetic awareness of the self as the one who lived the event [28]. When a human writes, they bring *internal details* — the event, place, time, perceptual and emotion/thought fragments that belong to one specific episode. AI, limited to its training data, resorts to *external details*: semantic facts, off-event digressions and editorial commentary not tied to any central episode. The distinction is not metaphorical; it is the scoring taxonomy of the Autobiographical Interview, in which younger adults were biased toward details “reflecting happenings, locations, perceptions, and thoughts” while older adults “favored semantic details not connected to a particular time and place” [29]. The same construct dissociates in the brain, with lateral parietal cortex tracking the quantity of internal detail in an imagined event [30, 31].

Including personal, first-person material is one of the most effective methods of humanising a text. Across a corpus of matched human and ChatGPT stories, first-person pronouns appeared at $M = 2.81$ in human writing against $M = 0.21$ in generated text, and descriptions of relativity — motion, space and time — were consistently sparser in the machine output [33]. Readers were measurably less transported into AI stories, and mediation analysis attributed that gap to the missing personal pronouns [33]. The

absence of these episodic markers results in a text that feels “detached” from reality, floating in a vacuum of academic or corporate abstractions [18, 23].

5. Engagement Markers vs. Hedging

Another point of divergence is how the author handles the reader and their own certainty. AI, trained to be helpful and safe, tends toward excessive “hedging” in expository and research-abstract genres, using words like “possibly,” “it could be argued,” and “perhaps” to avoid categorical commitments [34, 35, 36]. The finding is genre-dependent rather than universal: in student argumentative essays the same research group reports the opposite pattern, with ChatGPT producing significantly *lower* frequencies of interactional metadiscourse and a flatter, more impersonal tone [37].

The engagement side of the contrast is more robust. Human writers use more engagement markers — direct mentions of the reader, rhetorical questions, and appeals to shared knowledge [38]. Comparing 145 student essays against 145 ChatGPT essays, student writing was significantly richer in both the quantity and the variety of engagement features, with rhetorical questions and personal asides the most conspicuous absences from the generated set [39].

Table 3: Rhetorical marker frequency in AI and human writing.

Rhetorical Marker	Frequency in AI	Frequency in Human	Function in Text
Hedging (expository genres)	High	Moderate	Demonstrate caution and academic rigour.
Reader mentions	Low	High	Create connection and dialogue with the audience.
Rhetorical questions	Low	High	Stimulate reflection and break monotony.
Personal asides	Nearly non-existent	Common	Humanise the author and validate authority.

6. The Algorithm’s Lexicon: Keywords to Avoid

Artificial intelligence has a preferred vocabulary that acts as a “red flag” for detectors and attentive readers. Terms like *delve*, *intricate*, *comprehensive*, and *meticulous* are used with a statistical frequency disproportionate to casual or professional human writing. Against a baseline of 15 million PubMed abstracts, *delves* appears in 2024 at 28.0 times its expected frequency and *showcasing* at 10.7 times, with 454 excess words identified in that year alone against none exceeding the threshold between 2013 and 2019 [40].

In machine-assisted peer reviews, *meticulous* rose 34.7-fold and *intricate* 11.2-fold [41]; across 1.1 million papers, *realm*, *intricate*, *showcasing* and *pivotal* were the four most disproportionately LLM-associated words [42].

6.1. The AI “Cliché Dictionary”

Lists of words to avoid have become working tools for humanising text, though they require care. The claim that LLMs overuse formal connectives is only partly supported: large-scale feature analysis finds conjuncts such as *furthermore* and *consequently* occurring at 96–116% of the human rate, essentially no difference, while stance adverbs such as *notably* and *additionally* do appear in the excess-vocabulary set [18, 40]. Two further caveats matter. These terms were already rising before ChatGPT’s release, so LLMs amplified an existing trend rather than creating it [43], and *delve* has been in measurable decline since early 2024 as writers learned to avoid it [44]. The lexicon is a moving target, not a fixed signature. Its origin appears to lie in reinforcement learning from human feedback rather than in the pre-training distribution [45].

Replacing these terms with human alternatives is a practical step:

- Instead of “*leverage resources*,” use “*use what we have*.”
- Instead of “*revolutionise the field*,” describe “*changing how things work*.”
- Instead of “*in conclusion*,” use “*what this means for us is...*”

Beyond individual words, the deficit is one of variety rather than of figuration. LLMs overproduce a narrow band of ornamental vocabulary — *camaraderie* at 162 times the human rate, *tapestry* at 155, *amidst* at 100 — which Reinhart and colleagues read as a preference for “grandiose, if hollow, summative sentences” [18]. At the population level, model output is homogeneous where human output is not: on divergent-thinking tasks, between-response variability was 0.459 for LLMs against 0.699 for humans [46], and generated stories recycle idiosyncratic plot elements across generations and across different models while human plots are rarely echoed [47].

7. Narrative Structures Beyond the School Model

AI logic often follows the most obvious path: introduction, development, and conclusion. This model, while functional, is predictable and lacks the structural creativity of an experienced writer [18]. To make text more human, it is necessary to break this linearity.

7.1. The *In Medias Res* Model and the Epiphany Bridge

Impactful human narratives often begin “in the middle of the action” (*in medias res*) or at a point of maximum tension. The departure of discourse order from story order — analepsis and prolepsis — is a structural resource rather than an ornament [48], and corpus analysis of oral personal narrative identifies complicating action, evaluation and resolution as the load-bearing sequence [49]. A narrative becomes tellable precisely when a canonical expectation is breached [50].

The Epiphany Bridge framework, introduced in practitioner literature [51], suggests a structure that prioritises the emotional journey before logical justification. This process involves showing “The Wall” (the insurmountable obstacle) before revealing “The Epiphany” (the solution). Although the framework itself has not been tested empirically, the mechanism it exploits is well documented. Readers absorbed in a narrative detect fewer false notes and generate fewer critical thoughts [52]; a meta-analysis of 132 effect sizes across 76 studies puts the association between transportation and critical counter-arguing at $\rho = -.20$, against $\rho = .57$ for affective response [53]. Sentiment analysis of 1,327 works of fiction finds that arcs containing a downturn before an upturn are among the most widely read [54].

AI tends to skip the conflict and doubt phase to deliver the ready-made solution, which removes the vulnerability necessary for human connection. Disclosure reliably increases liking, both for the discloser ($d = .281$) and as a consequence of having disclosed ($d = .317$) [55], and an admitted blunder raises the attractiveness of an otherwise competent speaker [56] — an effect that is conditional on demonstrated competence rather than universal. Integrating confessional failures and trial-and-error processes is therefore fundamental to simulating authenticity.

Table 4: Storytelling phase comparison between AI and human focus.

Phase	AI Focus	Human Focus	Effect
Introduction	Dictionary definition	Action scene or dialogue	Immediate engagement vs. boredom.
Conflict	Technical description	Feelings of frustration	Empathy and credibility.
Resolution	Fact summary	Lesson learned and change	Lasting impact and transformation.

8. Prompt Engineering for Humanising Text

To ensure that AI produces text that does not sound like a “machine,” the prompt must go beyond a simple request for content. It must provide a complete rhetorical context, defining the author, the audience, and crucially, the stylistic constraints [57, 58].

The effect is measurable: even ordinary task-oriented constraints in an instruction shift detector performance by up to 14.4 F1 points, and generally make the output harder to classify than an unconstrained baseline [59].

8.1. The C.R.E.A.T.E. Methodology and Rhetorical Context

The C.R.E.A.T.E. (Character, Request, Examples, Additions, Type, Extras) approach allows for granular shaping of AI output. By assigning a “character” that is not just a “professional writer” but someone with flaws, opinions, and a specific tone, the model is pushed out of its probabilistic comfort zone. Dialogue agents are usefully understood as maintaining a superposition of simulacra that a prompt collapses toward one role [60]. A caveat is warranted: across 162 personas and four model families, persona assignment produced no aggregate improvement in factual accuracy, and automatic persona selection performed no better than chance [61]. Personas shape register and voice; they do not by themselves make a model more correct.

Additionally, it is necessary to instruct the AI to ignore its default training of “making everything pretty” and visually perfect [18]. Asking for “rhythmic intentionality” and “controlled chaos” can help break the monotony of low burstiness [14].

8.2. Proposed Structured Prompts for Human-Like Writing

Based on the research presented, we propose two highly structured prompt models to feed the AI and obtain results that exceed detection metrics and create a better impression on human readers.

Prompt 1: The Episodic Narrative Structurer

Persona: You are an essayist and storyteller with 20 years of experience, known for your authentic, sometimes provocative voice that is always rich in sensory details. You detest corporate clichés and the “intro-body-conclusion” school structure.

Task: Write a text about [TOPIC] following these strict humanisation guidelines:

1. **Episodic Memory:** Do not start with definitions or generalisations. Begin *in medias res* with a specific scene. Use sensory details (smells, sounds, textures). Instead of talking about “innovation,” narrate the exact moment an idea failed and what was felt at that instant.
2. **Hierarchical Syntax (Hypotaxis):** Avoid parallel and flat sentences. Use subordinate clauses to create depth. I want to see sentences beginning with conjunctions like “Although,” “Despite,” “If.” Avoid the repetitive “Subject + Verb + Predicate” pattern.

3. **Dynamic Burstiness:** Radically vary sentence length. Use a very long, descriptive sentence (over 30 words) followed immediately by a short, punchy one (less than 5 words). Create “music” in the text.
4. **Engagement vs. Hedging:** Remove attenuators like “it can be said” or “it is important to note.” Speak directly to the reader. Use a rhetorical question that challenges common sense.
5. **Forbidden Lexical Filter:** It is strictly forbidden to use the following words: *delve, delineate, comprehensive, leverage, multifaceted, synergy, innovative, catalyst*. If used, the text has failed.

Prompt 2: The Verisimilitude and Rhythm Editor

System Configuration: Activate “Low Predictability Writing” mode. Your goal is to maximise lexical perplexity and rhythmic variability.

Style Instructions:

1. **Part-of-Speech Balance:** Reduce nominal loading. Use more action verbs and personal pronouns. Avoid turning processes into nouns (e.g., use “how people decide” instead of “the decision-making process”).
2. **Human Referential Maintenance:** Do not constantly repeat the subject of the sentence to maintain clarity. Use natural ellipses and pronominal variations, as a human would in a coffee conversation.
3. **Intentional Imperfection:** Introduce a slight informality or a regional idiom. If the text feels “too perfect” or visually decorated with bolding and symmetrical lists, it has failed. Use paragraphs of unequal sizes.
4. **Organic Subordination:** Instead of listing benefits with bullet points, connect them in a narrative paragraph where one point leads to the other through a complex cause-and-effect relationship.
5. **The Read-Aloud Test:** The final text must be able to be read aloud without the reader losing breath or feeling like they are reading an instruction manual.

9. Conclusion and Future Perspectives

The distinction between human and synthetic text is narrowing in terms of fluency, but the “scars” of AI’s mathematical logic remain visible to those who know where to look. Stylometric analysis of short fiction shows human authors forming broad, heterogeneous

clusters while model outputs cluster tightly and separately by model, each carrying a uniform fingerprint [62]. Understanding optimised syntax — where humans organise words to minimise cognitive dependency distance [63, 64] — is the new frontier of humanisation, and the result is counterintuitive: human news text is *more* optimised on this measure than LLM output, so an efficiency property long treated as a cognitive constraint now functions as a discriminating feature [11]. As models better integrate episodic memory capabilities and mental state simulation, prompts will need to become even more specific about the “world context” in which the text lives [32].

The key to success in the generative AI era is not to hide the use of the tool, but to collaborate with it to inject the subjectivity, vulnerability, and rhythmic chaos that define our species [65]. Humanised writing is not just a detection evasion technique — a fragile goal in any case, given that detectors misclassify roughly a third of genuinely human text [14] — it is a commitment to clarity, authenticity, and real connection with the reader [66, 67, 68].

References

- [1] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3, 1137–1155.
- [2] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- [3] Jelinek, F., Mercer, R. L., Bahl, L. R., & Baker, J. K. (1977). Perplexity — a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America*, 62(S1), S63. doi:10.1121/1.2016299
- [4] Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed., draft), ch. 3: N-gram Language Models.
- [5] Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., Lai, J. C., & Mercer, R. L. (1992). An estimate of an upper bound for the entropy of English. *Computational Linguistics*, 18(1), 31–40.
- [6] Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). The curious case of neural text degeneration. In *Proceedings of the 8th International Conference on Learning Representations (ICLR 2020)*.
- [7] Hans, A., Schwarzschild, A., Cherepanova, V., Kazemi, H., Saha, A., Goldblum, M., Geiping, J., & Goldstein, T. (2024). Spotting LLMs with Binoculars: Zero-shot detection of machine-generated text. In *Proceedings of the 41st International Conference on Machine Learning*, PMLR 235, 17519–17537.
- [8] Church, K. W., & Gale, W. A. (1995). Poisson mixtures. *Natural Language Engineering*, 1(2), 163–190. doi:10.1017/S1351324900000139
- [9] Katz, S. M. (1996). Distribution of content words and phrases in text and language modelling. *Natural Language Engineering*, 2(1), 15–59. doi:10.1017/S1351324996001246
- [10] Madsen, R. E., Kauchak, D., & Elkan, C. (2005). Modeling word burstiness using the Dirichlet distribution. In *Proceedings of the 22nd International Conference on Machine Learning*, 545–552. doi:10.1145/1102351.1102420
- [11] Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting linguistic patterns in human and LLM-generated news text. *Artificial Intelligence Review*, 57(10), 265. doi:10.1007/s10462-024-10903-2
- [12] Gehrmann, S., Strobel, H., & Rush, A. M. (2019). GLTR: Statistical detection and visualization of generated text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 111–116. doi:10.18653/v1/P19-3019

- [13] Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-shot machine-generated text detection using probability curvature. In *Proceedings of the 40th International Conference on Machine Learning*, PMLR 202, 24950–24962.
- [14] Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, 21, 53. [doi:10.1186/s41239-024-00487-w](https://doi.org/10.1186/s41239-024-00487-w)
- [15] 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, 26. [doi:10.1007/s40979-023-00146-z](https://doi.org/10.1007/s40979-023-00146-z)
- [16] Krishna, K., Song, Y., Karpinska, M., Wieting, J., & Iyyer, M. (2023). Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. In *Advances in Neural Information Processing Systems*, 36, 27469–27500.
- [17] Shaib, C., Elazar, Y., Li, J. J., & Wallace, B. C. (2024). Detection and measurement of syntactic templates in generated text. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 6416–6431. [doi:10.18653/v1/2024.emnlp-main.368](https://doi.org/10.18653/v1/2024.emnlp-main.368)
- [18] Reinhart, A., Markey, B., Laudenbach, M., Pantusen, K., Yurko, R., Weinberg, G., & Brown, D. W. (2025). Do LLMs write like humans? Variation in grammatical and rhetorical styles. *Proceedings of the National Academy of Sciences*, 122(8), e2422455122. [doi:10.1073/pnas.2422455122](https://doi.org/10.1073/pnas.2422455122)
- [19] Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). London: Routledge.
- [20] Matthiessen, C. M. I. M., & Thompson, S. A. (1988). The structure of discourse and “subordination.” In J. Haiman & S. A. Thompson (Eds.), *Clause Combining in Grammar and Discourse* (pp. 275–330). Amsterdam: John Benjamins. [doi:10.1075/tsl.18.12mat](https://doi.org/10.1075/tsl.18.12mat)
- [21] Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and Written Language: Exploring Orality and Literacy* (pp. 35–54). Norwood, NJ: Ablex.
- [22] Liu, W., & Liu, X. (2025). A comparative analysis of syntactic complexity in argumentative essays from rhetorical perspective: ChatGPT vs. English native speakers. *PLOS ONE*, 20(8), e0329410. [doi:10.1371/journal.pone.0329410](https://doi.org/10.1371/journal.pone.0329410)
- [23] Markey, B., Brown, D. W., Laudenbach, M., & Kohler, A. (2024). Dense and disconnected: Analyzing the sedimented style of ChatGPT-generated text at scale. *Written Communication*, 41(4), 571–600. [doi:10.1177/07410883241263528](https://doi.org/10.1177/07410883241263528)
- [24] Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. [doi:10.1016/j.jeap.2010.01.001](https://doi.org/10.1016/j.jeap.2010.01.001)

- [25] Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13, 18617. doi:10.1038/s41598-023-45644-9
- [26] Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511621024
- [27] Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). New York: Academic Press.
- [28] Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25. doi:10.1146/annurev.psych.53.100901.135114
- [29] Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, 17(4), 677–689. doi:10.1037/0882-7974.17.4.677
- [30] Thakral, P. P., Madore, K. P., & Schacter, D. L. (2020). The core episodic simulation network dissociates as a function of subjective experience and objective content. *Neuropsychologia*, 136, 107263. doi:10.1016/j.neuropsychologia.2019.107263
- [31] Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363–1377. doi:10.1016/j.neuropsychologia.2006.10.016
- [32] Huet, A., Ben Houidi, Z., & Rossi, D. (2025). Episodic memories generation and evaluation benchmark for large language models. In *Proceedings of the 13th International Conference on Learning Representations (ICLR 2025)*.
- [33] Appel, M., Malecki, W. P., Messingschlager, T. V., & Winkler, J. R. (2025). I, ChatGPT: Linguistic properties and human experiences of human- versus AI-generated stories. *Humanities and Social Sciences Communications*, 12, 1892. doi:10.1057/s41599-025-06341-2
- [34] Hyland, K. (1998). *Hedging in Scientific Research Articles* (Pragmatics & Beyond New Series 54). Amsterdam: John Benjamins. doi:10.1075/pbns.54
- [35] Alghazo, S., Rabab'ah, G., El-Dakhs, D. A. S., & Mustafa, A. (2025). Engagement strategies in human-written and AI-generated academic essays: A corpus-based study. *Amper-sand*, 15, 100237. doi:10.1016/j.amper.2025.100237
- [36] Ruegg, R., & Prommas, P. (2024). Comparing automated analysis and human analysis of hedging in academic texts. *Studies in Language Assessment*, 13(1), 263–287. doi:10.58379/SLVL8621
- [37] Jiang, F. K., & Hyland, K. (2025). Rhetorical distinctions: Comparing metadiscourse in essays by ChatGPT and students. *English for Specific Purposes*, 79, 17–29. doi:10.1016/j.esp.2025.03.001
- [38] Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. doi:10.1177/1461445605050365

- [39] Jiang, F. K., & Hyland, K. (2025). Does ChatGPT write like a student? Engagement markers in argumentative essays. *Written Communication*, 42(3), 463–492. doi:10.1177/07410883251328311
- [40] Kobak, D., González-Márquez, R., Horvát, E.-Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27), eadt3813. doi:10.1126/sciadv.adt3813
- [41] Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., et al. (2024). Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. In *Proceedings of the 41st International Conference on Machine Learning*, PMLR 235, 29575–29620.
- [42] Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., et al. (2025). Quantifying large language model usage in scientific papers. *Nature Human Behaviour*, 9(12), 2599–2609. doi:10.1038/s41562-025-02273-8
- [43] Matsui, K. (2025). Delving into PubMed records: How AI-influenced vocabulary has transformed medical writing since ChatGPT. *Perspectives on Medical Education*, 14(1), 882–890. doi:10.5334/pme.1929
- [44] Geng, M., & Trotta, R. (2025). Human-LLM coevolution: Evidence from academic writing. In *Findings of the Association for Computational Linguistics: ACL 2025*, 12689–12696. doi:10.18653/v1/2025.findings-acl.657
- [45] Juzek, T. S., & Ward, Z. B. (2025). Why does ChatGPT “delve” so much? Exploring the sources of lexical overrepresentation in large language models. In *Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025)*, 6397–6411.
- [46] Wenger, E., & Kenett, Y. N. (2026). Large language models are homogeneously creative. *PNAS Nexus*, 5(3), pgag042. doi:10.1093/pnasnexus/pgag042
- [47] Xu, W., Jovic, N., Rao, S., Brockett, C., & Dolan, B. (2025). Echoes in AI: Quantifying lack of plot diversity in LLM outputs. *Proceedings of the National Academy of Sciences*, 122(35), e2504966122. doi:10.1073/pnas.2504966122
- [48] Genette, G. (1980). *Narrative Discourse: An Essay in Method* (J. E. Lewin, Trans.). Ithaca, NY: Cornell University Press.
- [49] Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the Verbal and Visual Arts* (pp. 12–44). Seattle: University of Washington Press. Reprinted in *Journal of Narrative and Life History*, 7(1–4), 3–38, 1997. doi:10.1075/jnlh.7.02nar
- [50] Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18(1), 1–21. doi:10.1086/448619
- [51] Brunson, R. (2017). *Expert Secrets: The Underground Playbook for Creating a Mass Movement of People Who Will Pay for Your Advice*. New York: Morgan James Publishing.

- [52] Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721. doi:10.1037/0022-3514.79.5.701
- [53] van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2014). The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers’ narrative transportation. *Journal of Consumer Research*, 40(5), 797–817. doi:10.1086/673383
- [54] Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5, 31. doi:10.1140/epjds/s13688-016-0093-1
- [55] Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, 116(3), 457–475. doi:10.1037/0033-2909.116.3.457
- [56] Aronson, E., Willerman, B., & Floyd, J. (1966). The effect of a pratfall on increasing interpersonal attractiveness. *Psychonomic Science*, 4(6), 227–228. doi:10.3758/BF03342263
- [57] Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, art. 314, 1–7. doi:10.1145/3411763.3451760
- [58] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), art. 195, 1–35. doi:10.1145/3560815
- [59] Koike, R., Kaneko, M., & Okazaki, N. (2024). How you prompt matters! Even task-oriented constraints in instructions affect LLM-generated text detection. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 14384–14395. doi:10.18653/v1/2024.findings-emnlp.841
- [60] Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature*, 623(7987), 493–498. doi:10.1038/s41586-023-06647-8
- [61] Zheng, M., Pei, J., Logeswaran, L., Lee, M., & Jurgens, D. (2024). When “a helpful assistant” is not really helpful: Personas in system prompts do not improve performances of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 15126–15154. doi:10.18653/v1/2024.findings-emnlp.888
- [62] O’Sullivan, J. (2025). Stylometric comparisons of human versus AI-generated creative writing. *Humanities and Social Sciences Communications*, 12, 1708. doi:10.1057/s41599-025-05986-3
- [63] Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33), 10336–10341. doi:10.1073/pnas.1502134112

- [64] Gibson, E., Futrell, R., Piantadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5), 389–407. doi:10.1016/j.tics.2019.02.003
- [65] Gottschling, M. (2025). Towards rhetorical AI literacy. Presenting a conceptual framework. *Argumentation et Analyse du Discours*, 35. doi:10.4000/14yaw
- [66] Georgiou, G. P. (2026). What distinguishes AI-generated from human writing? A rapid review of the literature. *Big Data and Cognitive Computing*, 10(2), 55. doi:10.3390/bdcc10020055
- [67] Georgiou, G. P. (2025). Differentiating between human-written and AI-generated texts using automatically extracted linguistic features. *Information*, 16(11), 979. doi:10.3390/info16110979
- [68] Culda, L. C., Nerişanu, R. A., Cristescu, M. P., Mara, D. A., Băra, A., & Oprea, S.-V. (2025). Comparative linguistic analysis framework of human-written vs. machine-generated text. *Connection Science*, 37(1), 2507183. doi:10.1080/09540091.2025.2507183