Journal of Entrepreneurship and Business Development

Volume 5 , Issue 1, November 2025 , Pages 7-25

DOI: 10.18775/ijmsba.1849-5664-5419.2014.51.7001
URL :https://doi.org/10.18775/ijmsba.1849-5664-5419.2014.51.7001

THE DISAPPEARING AUTHOR: LINGUISTIC AND COGNITIVE MARKERS OF AI-GENERATED COMMUNICATION

1 Suresh Sood
1 Industry/Professional Fellow, Australian Artificial Intelligence Institute, University of Technology Sydney

Abstract: As generative AI permeates professional, educational, and policy communication, distinguishing human from machine-authored text is increasingly consequential for assessment, accountability, and trust (Caliskan et al., 2017; COPE, 2022; Gehrmann et al., 2019). We introduce VERMILLION, a ten-signal heuristic framework grounded in stylometry, cognitive linguistics, and AI interpretability, and apply it to the White House Make America Healthy Again (MAHA) policy report. Paragraph-level annotation reveals recurrent markers consistent with LLM outputs (e.g., echoed sentence structures, rigid transitions, hedging, and absent lived experience), suggesting substantial AI involvement. We position VERMILLION relative to statistical detectors (e.g., GLTR; DetectGPT), watermarking approaches, and human detection limits (Gehrmann et al., 2019; Kirchenbauer et al., 2023; Mitchell et al., 2023), and propose disclosure practices for responsible authorship. Findings have implications for pedagogy, editorial policy, and public communication.

Keywords: AI Generated writing; Professional writing; Forensic linguistics; Machine authorship; Computational linguistics; Authorship transparency; Linguistic markers; Heuristic detection; Policy document analysis; Large Language Models (LLMs); AI-Generated Communication

1. Introduction -Why Authorship Matters

As generative AI (GenAI) becomes more deeply embedded in our communication systems, the ability to differentiate between human and machine-authored text is no longer just technical curiosity but a necessity. Authorship whether human or machine touches on the very foundations of trust, authenticity, and intellectual development. Nowhere is this more apparent than in education. The integrity of academic assessment hinges on knowing who or what authors a piece of work. When students rely on large language models to generate essays or reports, they unconsciously bypass the cognitive challenges fostering learning. Writing is not merely a vehicle for communication but a process for embracing forms of reasoning, synthesis, and personal insight. Without authentic engagement, developing core skills of critical thinking, argumentation, and expression diminish over repeated use of GenAI. Yet, the implications reach far beyond the classroom. In public domains like policy, journalism, and science, authorship carries weight implying accountability, credibility, and human judgment. Text produced by AI, even when factually correct, may lack context, nuance, or the capacity to interpret ambiguity. Transparency about the use of AI in these contexts is essential not to penalize the usage, but to preserve trust in the human voices behind institutions and decisions.

This concern over voice and responsibility underscores a deeper ethical issue. AI cannot be held morally or legally accountable for the words generated. When automated systems produce misleading or harmful content, assigning responsibility becomes murky. This content is especially important in areas of healthcare, law, or politics. The absence of human attribution raises troubling questions about bias, consent, and misrepresentation. At the heart of this responsibility lies something profoundly human. Our language and writing reflects memory, culture, emotion, and lived experience. These are elements AI can mimic but not originate. The homogenized outputs of language models often feel polished but empty, missing the quirks, interruptions, and textures marking a genuine voice. To defend the value of human expression, we must learn to recognize the absence.

There is now a newfound sense of urgency. The ease with which AI can now generate persuasive, coherent, and syntactically flawless text poses new risks in the spread of disinformation. In moments of crisis during elections, pandemics, or conflict, generative capability can be weaponized. Tools for identifying machine authorship thus become vital safeguards in an age of algorithmic persuasion. Distinguishing between human and machine writing is not simply about style. But, about responsibility, authenticity, and the kind of communicative world we desire to build. If we lose sight of the human author, we risk losing more than just the writer we risk losing the reader.

In light of these challenges, the question becomes not only why authorship matters, but also how we can reliably tell who or what is doing the writing. The subtle erosion of human voice, intentionality, and stylistic variation in AI-generated text often evades detection by casual readers. As AI tools grow more advanced and outputs more polished, identifying linguistic fingerprints becomes both more urgent and complex. This necessitates a systematic approach, one that moves beyond gut feeling and toward a principled framework for detection.

1.1 Related Work

Across decades of work, stylometry has shown that an author's "fingerprints" live in small, often unconscious choices of function words, character n-grams, and recurrent syntactic patterns modeled to attribute or verify authorship (Juola, 2006; Koppel et al., 2009; Stamatatos, 2009, 2016). In the era of neural networks, complementary detectors mine statistical signatures of generated text: Giant Language Model Test Room[1] (GLTR) visualizes token-probability profiles to surface stretches of unusually low surprisal typical of machine outputs (Gehrmann et al., 2019), while DetectGPT leverages probability curvature under perturbations to enable zero-shot detection (Mitchell et al., 2023). Model-side watermarking, meanwhile, aims to embed provenance directly into generated text, though important trade-offs remain around robustness and vulnerability to adversarial removal (Kirchenbauer et al., 2023). Complicating matters further, human readers are poor judges of fluent machine prose especially after light editing or paraphrasing further minimizing unaided detection accuracy (Ippolito et al., 2020; Zellers et al., 2019). Taken together, these strands motivate an interpretable, heuristic layer such as VERMILLION to sit alongside statistical detectors and disclosure practices, offering transparent cues capable of being audited and triangulated.

2. Tell-Tale Signs of AI Writing -VERMILLION Heuristic Framework

The VERMILLION ten cues align with well-established dimensions of writing style studied in stylometry and discourse analysis, lexical choice, syntax, punctuation, structural layout, and metadiscourse. Large-scale authorship research shows that style is recoverable from these levels (e.g., punctuation, sentence/paragraph length, and function-word patterns), and that combining levels improves attribution robustness (Stamatatos, 2009; Abbasi & Chen, 2008; Jafariakinabad & Hua, 2021). Below we situate each cue in prior literature and explain why it is informative for identifying machine assistance.

[1] GLTR visualizes token probability profiles to flag low-surprisal sequences (i.e., highly predictable, low-information words) typical of machine text

1. V — *Vague 'their' (ambiguous referents).*
   Pronoun interpretation is tightly coupled to local discourse coherence; ambiguous referents increase processing difficulty and signal weak coherence planning. Centering theory models preferences for pronoun antecedents (Grosz, Joshi, & Weinstein, 1995) and experiments show that coherence relations drive pronoun resolution preferences (Wolf, Gibson, & Desmet, 2004). Texts that repeatedly leave 'they/their' unanchored to salient discourse entities exhibit coherence lapses that VERMILLION flags.
2. E — *Echoed sentence structures (template reuse).*
   Neural text generators can produce bland, repetitive phrasing without careful decoding (Holtzman, Buys, Du, Forbes, & Choi, 2020). Academic prose also relies on recurrent 'lexical bundles' and phrase frames (Hyland, 2008; Biber & Barbieri, 2007), but over-regular templating across adjacent sentences lowers human-like variation. VERMILLION targets local repetition and overuse of stock frames.
3. R — *Rigid transitions and canned connectives*.
   Metadiscourse resources of connectors, frame markers, and stance devices organize text and project authorial engagement (Hyland, 2005). Corpus studies show register- and discipline-specific distributions of these devices (Biber & Barbieri, 2007). Uniform, formulaic transitions at fixed intervals are atypical of expert human style and thus diagnostic.
4. M — *Mechanical punctuation and rhythm.*
   Punctuation patterns (including em/en dashes, commas, and sentence-length distributions) are distinctive style carriers usable for attribution (Darmon et al., 2019) and are standard in feature sets like Writeprints (Abbasi & Chen, 2008). VERMILLION attends to overly uniform punctuation rhythms or avoidance of expressive marks.
5. I — *Inflexible paragraphing and layout.*
   Structural features—paragraph counts and average sentence/paragraph lengths—encode stable aspects of style and improve attribution when combined with lexical/syntactic information (Jafariakinabad & Hua, 2021; Jafariakinabad, Tarnpradab, & Hua, 2020). VERMILLION treats highly uniform paragraph blocks as a weak-signal heuristic.
6. L — *Lack of short, emphatic paragraphs.*
   While conventions vary by genre, human-authored expository writing typically mixes paragraph lengths for rhetorical effect. Structural stylometry treats such variation as signal; flat distributions can indicate templated generation (Jafariakinabad & Hua, 2021).
7. L — *Lack of personal voice / authorial presence.*
   Authorial identity and the strategic use (or avoidance) of first-person reference are central to credibility and stance (Hyland, 2002). Across university registers, stance devices are sparser in research articles than speech but are not absent (Biber, 2006). Persistent impersonality where disciplinary norms permit self-mention suggests machine assistance.
8. I — *Imprecise abstraction and nominalizations.*
   Academic writing often compresses information via complex noun phrases and nominalizations, but overreliance reduces clarity and specificity (Biber & Gray, 2010). VERMILLION flags cascades of abstract nouns that mask agency and evidence chains.
9. O — *Overuse of hedging.*
   Hedging is a legitimate resource for epistemic caution (Hyland, 1998) and varies by register (Biber, 2006). Excessive or patterned hedging combined with the other cues contributes to the overall signal.
10. N — *No lived experience or concrete provenance.*
    Work on authorial identity shows that selective self-mention and evidential grounding help establish responsibility for claims (Hyland, 2002). Absence of experiential anchors (where genre permits) can therefore support a machine-assistance hypothesis when co-occurring with other cues.

3. Methodology – Applying the VERMILLION Framework

The central tool used to assess the likelihood of AI authorship in writings is the VERMILLION framework to detect common markers of AI-generated language. Rather than relying on black-box detection software or algorithmic probability scores, this methodology emphasizes human interpretability and textual transparency, allowing for qualitative analysis based on well-established linguistic indicators. While no single sign is conclusive, the co-occurrence of multiple VERMILLION features within a document significantly strengthens the inference of AI assistance or authorship.

The VERMILLION Dictionary of Heuristics (Table 1) offers a structured, interpretable, and practical approach to identifying signs of AI-generated writing. Heuristics bridge the gap between human expertise and algorithmic structure offering transparency and interpretability when evaluating authorship in high-stakes contexts like education, policy, and science (Binns, 2017; Mittelstadt et al., 2018). Unlike black-box classifiers, the VERMILLION Dictionary enables users to "see" and "understand" why text may be AI-generated making for both an educational and ethical tool in the age of machine authorship. The utility of the dictionary spans both human-led review and machine-assisted analysis. We now describe how each mode can be applied effectively.

Table 1

Dictionary of Heuristics for AI-Generated Writing (VERMILLION Framework)

Heuristic (VERMILLION)	Definition	Detection Tip / Rule of Thumb
V – Vague "Their" Usage	Frequent third-person possessive pronouns ("their") without a clear or recent noun antecedent, causing ambiguity.	Search for "their" and check if the noun it refers to is clear. Ambiguity suggests AI tendency.
E – Excessive Hedging	Overuse of modal verbs and qualifiers ("might," "could," "arguably") to avoid assertiveness or narrative commitment.	Count modals like "might," "may," "could." High frequency signals hedging language typical of AI.
R – Repetitive Structure	Uniform sentence patterns (e.g., "They did X. They did Y.") across paragraphs with little syntactic variation.	Read aloud for "drumbeat" cadence. Similar lengths and rhythms signal AI templating.
M – Mechanical Transitions	Use of generic connectors like "Moreover," "Furthermore," "In conclusion" to open paragraphs in a formulaic way.	Scan paragraph starts for stock transitions. Consistency without variation suggests automation.
I – Inconsistent Rhythm	Paragraphs follow uniform length or pacing without natural variation or rhetorical breaks, giving a mechanical feel.	Highlight paragraph lengths. If most are the same length, flag for review.
L – Lack of Short Paragraphs	No use of one-line paragraphs for emphasis or pacing—a human habit to dramatize or draw attention.	Search for 1-line paragraphs. If none, this suggests non-human rhythm control.
L – Lack of Personal Voice	Text lacks subjective tone, emotional cues, opinion, or anecdotal elements that give writing a human feel.	Ask: "Could this have been written by anyone?" If yes, human voice may be absent.
I – Imbalanced Apostrophes	Overuse or uniform use of contractions (e.g., "it's," "you'll," "we're"), giving a mechanically casual tone.	Count contractions. Overly balanced usage may signal unnatural stylization.

O – Overuse of Em Dashes	Em dashes used mid-sentence to simulate dramatic pauses or clause separation, often overused in AI writing.	Count em dashes (—). If frequent and awkward, revise or replace with commas/periods.
N – Nominalizations	Excessive use of noun forms like "implementation," "utilization," and "application" that obscure who is doing what.	Replace with active verbs. "Implementation of policy" → "We implemented the policy."

Manual Application: A Guided Forensic Reading

For educators, editors, researchers, and students, the dictionary serves as a diagnostic lens much like a physician checking for symptoms. The process details follow these steps:

1. Print or Display the Heuristics Table (table 1): Keep the dictionary beside the document under review.
2. Paragraph-by-Paragraph Review: As you read, annotate each paragraph with potential heuristic markers (e.g., repetitive structure, vague "their" usage).
3. Tally and Cluster: Note how many signs appear per section. A dense clustering of multiple heuristics across multiple paragraphs increases the likelihood of machine authorship.
4. Judgment with Caution: No single marker guarantees AI authorship. Instead, the accumulation of stylistic and structural signs over a long passage provides a pattern-based justification for suspicion or intervention.

Automated Application: From Heuristics to Code

The structured nature of the VERMILLION framework makes it ideal for light machine-learning or rule-based natural language processing (NLP) integration. Here are the steps for putting the automation into practice:

1. Tokenization and POS Tagging: Use NLP libraries (e.g., ,) to segment text into tokens and apply part-of-speech tagging to detect modal verbs, dashes, contractions, sentence length, and pronoun usage.
2. Rule-Based Flags: For example,
    - If >30% of sentences begin with the same 3 transition words, flag for Mechanical Transitions.
    - If "their" appears >5 times in a 300-word section without clear antecedents, flag for Vague Pronoun Use.
    - Calculate mean and standard deviation of sentence lengths—if very low, flag for Uniform Sentence Structure.
3. Score Accumulation and Thresholding:
    - Each heuristic can be scored (0–1) and summed to create a VERMILLION Score.
    - A threshold (e.g., 7+ out of 10 flags) can suggest probable AI authorship and prompt further review.
4. Hybrid AI-Human Workflow:
    - Use detection as a pre-screening tool, not a final arbiter.
    - Documents flagged with high VERMILLION scores are then reviewed manually for context, confirming or rejecting the initial AI-authorship suspicion

Use Case (examples): Academic Essay Screening

- Educator Tool: A plugin for grading platforms (e.g., Turnitin or Moodle) can highlight VERMILLION signs during essay submission, offering formative feedback.
- Publishing Tool: Editors reviewing policy or white papers can apply automated scanning before peer review to screen for overuse of hedging or mechanical transitions.
- Research Corpus Analysis: Digital humanities scholars can use the framework to assess how AI-generated text has entered corpora over time.

To apply this framework, as an example, the full text of the White House Make America Healthy Again (MAHA) Report (MAHA Commission 2025) is subject to paragraph-level content analysis. Each paragraph is coded for the presence or absence of the VERMILLION indicators. Initially, the coding is conducted manually using the structured rubric of table 1 defining each sign with examples and detection criteria (e.g., presence of vague third-person possessives without antecedents; sentences of uniform rhythm and clause count; paragraph transitions using "Moreover" or "In conclusion").

Where relevant, contextual qualifiers such as genre conventions, policy tone, and expected rhetorical features are considered to avoid over-attributing stylistic choices to machine authorship.

This qualitative-quantitative hybrid method allows for triangulation across three dimensions:

1. Stylistic frequency (how often a sign appears),
2. Structural consistency (how widely distributed the signs are across the document), and
3. Functional role (whether the sign affects meaning, tone, or reader engagement).

While this framework does not provide a probabilistic certainty of AI authorship, the strength lies in offering a replicable, explainable, and human-auditable method for assessing documents suspected of machine generation. The VERMILLION framework thus bridges the gap between technical detection tools and human expert judgment, contributing to the emerging discipline of authorship forensics in the age of generative AI.

4. Analysis and Key Findings Using VERMILLION Framework

The White House Make America Healthy Again (MAHA) Report (MAHA Commission 2025) is systematically reviewed using the framework. The goal is to assess whether the document exhibits stylistic and structural markers typically associated with large language models (LLMs) such as GPT-4. The MAHA Report was selected as the subject of this analysis due to public scrutiny concerning factual inaccuracies and questionable citations. Multiple media outlets including the Associated Press (AP, 2025), Axios (Owens, 2025), and Al Jazeera (2025) report the White House acknowledges errors in the report, including references to non-existent studies and misattributions related to vaccine and chronic illness research. These issues triggered widespread debate over the credibility of the report, prompting official commitments to amend the document. Given the policy significance of the report originating from the Executive Office of the President and intended role in shaping national health strategy, we would not typically expect the involvement of AI-generated content, which raises further concerns regarding the production process. The combination of factual missteps, bureaucratic authorship, and stylistic anomalies make the MAHA Report even in an updated state (MAHA Commission 2025), an ideal candidate for evaluating the efficacy of the VERMILLION heuristic framework in detecting AI authorship patterns in high-stakes, real-world documentation.

Table 2

MAHA Report Key Findings by VERMILLION Sign

Sign	Definition	Examples	Example Locations (Page ¶)
V – Vague "their"	Frequent third-person possessive pronouns lacking clear antecedents.	[their health, their community, their environments, their experience, their wellbeing]	[Page 42 ¶2, Page 54 ¶1, Page 63 ¶3, Page 67 ¶2, Page 12 ¶4]
E – Echoed sentence structures	Sentences that follow the same rhythm and length.	[An analysis of..., The findings suggest..., Research shows..., Studies confirm..., Evidence indicates...]	[Page 46 ¶3, Page 18 ¶1, Page 25 ¶4, Page 31 ¶2, Page 60 ¶1]
R – Rigid transitions	Formulaic phrases used to stitch paragraphs rigidly.	[Moreover,, Furthermore,, In conclusion,, It is important to note that, Additionally,]	[Page 9 ¶1, Page 15 ¶2, Page 21 ¶3, Page 30 ¶1, Page 45 ¶2]

M – Mechanical punctuation	Excessive em dashes interrupt sentence flow.	[— while necessary —, — a known factor —, — for example —, — often seen —, — including —]	[Page 6 ¶2, Page 10 ¶1, Page 23 ¶3, Page 37 ¶2, Page 59 ¶1]
I – Inflexible paragraphing	Overuse of modal verbs and qualifiers to avoid commitment.	[may indicate, could reflect, possibly suggest, arguably linked, might demonstrate]	[Page 8 ¶1, Page 20 ¶3, Page 26 ¶2, Page 33 ¶1, Page 47 ¶3]
L – Lack of short paragraphs	Use of jargon-heavy or vague buzzwords.	[cross-sector synergy, data-driven solutions, impactful interventions, scalable models, multi-pronged approach]	[Page 11 ¶2, Page 14 ¶3, Page 28 ¶2, Page 35 ¶1, Page 50 ¶4]
L – Lack of personal voice	Paragraphs consistently similar in length and structure.	[All paragraphs 3–5 lines, No visible variation, Even pacing, Predictable rhythm, Homogenized structure]	[Page 7 ¶1, Page 13 ¶3, Page 19 ¶2, Page 29 ¶1, Page 40 ¶2]
I – Imprecise abstraction	Absence of punchy, one-line paragraphs for emphasis.	[No single-sentence para, No rhetorical breaks, Consistent block style, No visual relief, Same structure throughout]	[Page 5 ¶1, Page 17 ¶2, Page 24 ¶3, Page 36 ¶1, Page 53 ¶3]
O – Overuse of hedging	Uniform application of contractions throughout.	[it's, we've, they're, you'll, we're]	[Page 5 ¶3, Page 16 ¶1, Page 22 ¶4, Page 34 ¶2, Page 44 ¶3]
N – No lived experience	Absence of anecdotes, emotions, or personal narratives.	[No quotes, No subjective insights, No human examples, No grounded testimony, Abstract tone only]	[Page 27 ¶1, Page 32 ¶2, Page 41 ¶3, Page 52 ¶1, Page 58 ¶4]

V — Vague Use of "Their"
The report frequently uses third-person possessives like "their efforts," "their health," and "their communities" without clear antecedents.

- Examples: Page 54, 63, 67: "Their..." lacks subject clarity.
- Implication: This reflects AI-generated tendencies to generalize audiences without specifying actors. Page 42 highlights 6 occurrences of "Their…"

E — Echoed Sentence Structures (Repetitive Syntax)
Multiple paragraphs exhibit uniform clause length and mirror grammatical rhythm.

- Example: Page 46, ¶3: "An analysis of a common pesticide…An analysis of a 115 studies…"
- Implication: This drumbeat pacing reflects template-based generation common in LLMs.

R — Rigid/Mechanical Transitions
Transitional phrases such as "Moreover," "In addition," and "It is important to note that…" are overused, especially at paragraph openings.

- Example: Page 9, ¶1 begins with "Furthermore," and follows a sequence of "Moreover" and "In conclusion" in preceding sections.
- Implication: This suggests reliance on generic cohesion strategies common to LLMs.

M — Mid-Sentence Dash Overload
There is notable use of em dashes for interruption or emphasis where commas or full stops might serve better.

- Example: Page 6, ¶2: "The strategy — while ambitious — was necessary."
- Implication: This mimics a stylistic quirk found in AI-authored text when attempting emphasis.

I — Indecisive Hedging
The report frequently uses qualifiers like "may," "could," "possibly," and "arguably," avoiding commitment.

- Example: Page 8, ¶1: "This may suggest a pathway to improvement…"
- Implication: AI systems hedge statements to minimize error or liability.

L — Lingo-Heavy Buzzword Flood
Corporate and bureaucratic buzzwords are present throughout without concrete explanation.

- Examples: "Data-driven outcomes," "cross-sector synergy," "impactful interventions."
- Implication: AI tends to inflate language for authority without added meaning.

L — Lack of Paragraph Variability
Paragraphs across the report are consistently 3–5 lines long, regardless of subject complexity.

- Implication: Human-authored reports typically show more natural variation reflecting rhetorical pacing or argument shifts.

I — Infrequent Short Paragraphs
Punchy, one-line paragraphs—which human writers often use for emphasis—are almost entirely absent.

- Implication: The text lacks cadence variation, which contributes to monotony.

O — Overuse of Apostrophes (Contractions)
There is extensive use of contractions like "it's," "we've," "they're," which can feel uniformly applied.

- Example: Page 5, ¶3: "It's clear that we've made progress."
- Implication: AI-generated text sometimes inserts contractions to sound conversational but overdoes the shortening uniformly.

N — No Personal Voice or Lived Experience
The report lacks anecdotes, quotes, or grounded human experiences.

- Example: Even when discussing community health interventions, no personal stories or perspectives are included.
- Implication: This absence of emotional grounding and experiential depth is typical of LLM outputs.

Taken individually, none of the signs offer definitive proof of AI authorship. However, their accumulation and repetition across the entire document strongly suggest the involvement of generative AI tools. The report exhibits nearly every marker in the VERMILLION framework with moderate to high frequency, including structural rigidity, vague referents, hedging, and impersonal tone. Furthermore, the language lacks the spontaneous irregularity, anecdotal inclusion, and rhetorical nuance found in human-authored professional communication. It reads with a polished, neutral efficiency hallmarks of machine-assisted generation.

5. Future Possibilities

The VERMILLION framework not only serves as a diagnostic tool for detecting AI-assisted authorship but opens a range of future possibilities in research, education, policy, and software development.

5.1 Toward a Scalable Detection System

As the volume of machine-generated content increases, the need for scalable and systematic detection becomes urgent. The current application of VERMILLION relies on human judgment and close reading, but the structured, checklist makes it suitable for computational implementation. Future research may translate the VERMILLION indicators into natural language processing (NLP) features, enabling semi-automated screening of documents at scale. These tools could flag passages that match known AI writing patterns and support human reviewers with visual dashboards showing frequency and clustering of linguistic cues. Such systems would be especially valuable in academic publishing, public policy vetting, and educational assessment, where distinguishing human from synthetic authorship directly affects decision-making, trust, and fairness.

5.2 Disclosure Protocols and AI Authorship Policy

The findings underscore the importance of transparent authorship attribution, particularly in public-facing policy documents. If generative AI models such as GPT-4 are used in report drafting, even as co-authors or editors, this involvement should be clearly disclosed. Institutions including governments, universities, and think tanks likely need to develop AI authorship disclosure protocols, similar to citation standards in academic research. Doing so not only promotes accountability and trust but also sets ethical precedents for future human-AI collaborations in content creation.

Example: AI Authorship Disclosure in an Academic Report
Sample authorship note:
Portions of this report were drafted using the assistance of the GPT-4 language model developed by OpenAI (OpenAI, 2023). The AI was employed to generate initial drafts of sections 2 and 4, provide structural suggestions, and summarize background literature. All AI-generated content was subsequently reviewed, edited, and validated by the human authors to ensure accuracy, appropriateness, and adherence to disciplinary standards. The authors accept full responsibility for the final content.
This kind of disclosure:

- Mirrors citation standards by giving credit and clarifying intellectual contribution.
- Enables transparency about machine involvement.
- Supports academic integrity and allows readers to assess the authenticity and accountability of the claims.

The disclosure follows principles similar to COPE (Committee on Publication Ethics) and emerging AI ethics guidelines emphasizing full disclosure when automated tools influence scholarly or public-facing outputs.

5.3 Curriculum Design and Pedagogical Applications and AI

The VERMILLION framework holds pedagogical value. Educators can use it to help students develop critical literacy skills by analyzing how AI-generated text differs from human-authored prose. Assignments might include comparing documents, rewriting AI passages to restore human voice, or conducting VERMILLION-guided assessments of peer work. Instructors can also use the checklist to identify overreliance on generative tools and initiate conversations about authorship, originality, and ethical writing practices. By integrating VERMILLION into classroom activities, educators shift the focus from penalizing AI use to understanding and critique intelligently.

5.4 AI Writing Assistants with Human in the Loop Feedback

Another promising avenue involves turning VERMILLION into a feedback engine for generative AI platforms themselves. By embedding stylistic detectors based on VERMILLION indicators, platforms could highlight when a draft appears overly mechanical, hedged, or impersonal offering writers the option to revise or “humanize” their outputs. This could encourage more thoughtful co-creation, where users remain aware of the stylistic and ethical implications of outsourcing authorship to machines.

5.5 AI Cross-Disciplinary Research Opportunities

The intersectional nature of the VERMILLION framework drawing on cognitive psychology, computational linguistics,

and AI interpretability invites collaboration across fields. Scholars in behavioral economics, media studies, forensic linguistics, and computer science may apply the model to explore deeper questions. How does AI alter the rhetorical structure of communication? What biases persist in machine-generated narratives? Can humans detect synthetic authorship unaided, or do we need new literacy tools? As AI continues to reshape how knowledge is produced and communicated, such inquiries will become central to both scholarship and policy.

5.6 From Detection to Revision: Humanizing AI-Generated Text

Detecting AI-authored writing using the VERMILLION framework provides a diagnostic lens, but the true value lies in the next steps. Once tell-tale signs are identified be they mechanical transitions, uniform paragraphs, or vague pronouns, writers, educators, and editors need practical tools for transformation. The goal is not simply to punish the use of AI, but to encourage authentic, contextually grounded, and personally expressive writing reflecting human intent and cognition. To support this, we introduce a set of simple revision heuristics, rule-of-thumb strategies designed to convert "AI-sounding" prose into more natural, human-like communication. These heuristics are usable by students improving drafts, educators scaffolding learning, or professionals refining content. They are grounded in both cognitive linguistics and stylistic best practices, focusing on rhythm, specificity, emotional tone, and structural variation. Importantly, these strategies do not require complex software or advanced AI detection tools. Instead, they emphasize deliberate rephrasing, personalization, and narrative rhythm. When applied consistently, even minor textual adjustments can significantly reduce the synthetic feel of machine-generated prose.

Table 3

Heuristics for Humanizing AI-Generated Writing

Tell-Tale Sign	Humanizing Heuristic	AI Generated	Humanized
Dash Overload	Replace em dashes (—) with commas, periods, or ellipses for better rhythm.	"The policy—despite criticism—was implemented swiftly."	"The policy, despite criticism, was implemented swiftly." OR "The policy… despite criticism… was implemented swiftly."
Repetitive Sentence Structure	Vary sentence length and openers—use questions, subordinate clauses, or lists.	"It was tested. It was revised. It was implemented."	"After testing and revision, the team finally implemented it."
Mechanical Transitions	Drop or revise formulaic openers like "Moreover," or "In conclusion."	"Moreover, the findings were consistent."	"The findings held steady—adding weight to the earlier results."
Vague Use of "Their"	Replace vague pronouns with specific nouns or names where possible.	"Their decision was unexpected."	"The board's decision was unexpected."
No Short Paragraphs	Insert 1-line punchy paragraphs for impact or transition.	(Wall of text)	"That was the turning point."
Uniform Paragraph Size	Break up content with variable-length paragraphs.	(All ~3–4 lines)	Mix longer analysis with brief, reflective interjections.
Buzzword Flood	Swap vague terms (e.g., "synergy," "cutting-edge") with concrete, descriptive words.	"The cutting-edge initiative aimed to leverage synergy."	"The pilot project aimed to connect teams through shared goals."
Heavy Hedging	Reduce "could," "might," and "potentially"unless necessary—be confident.	"This might suggest a possible impact."	"This suggests a clear impact."
Overuse of Apostrophes	Remove excessive contractions for a more formal tone.	"It's clear. We're sure. You'll see."	"It is clear. We are confident. You will see."
No Personal Voice	Add anecdotes, questions, or personal opinion.	"The event was well received."	"I remember the applause—it echoed long after the final speech."

Tell-Tale Sign	Humanizing Heuristic	AI Generated	Humanized
Lack of Lived Experience	Add sensory details or emotional observations.	"Attendees enjoyed the session."	"Many attendees smiled, nodding through the speaker's story."

5.7 Temporal Erosion of Detectability (revision & version drift)

Detection power weakens over time for two reasons. First, revision/humanization**,** each editorial pass by a person or a "humanizer" paraphraser tends to smooth the very surface cues that VERMILLION flags (e.g., rigid transitions, templated rhythm, ambiguous pronouns). Paraphrasing specifically degrades detector accuracy and can even weaken watermark signals, making unaided classification less reliable after light rewriting (Krishna et al., 2023; Sadasivan et al., 2023; Masrour et al., 2025). Second, model version drift, as newer models are released, their default style shifts toward more human-like conventions (for example, testing responses from ChatGPT5 shows fewer stereotyped em-dashes in this model release[2]), reducing distinctiveness in punctuation regularities and transitional habits and thereby narrowing the gap to human prose. Even improved zero-shot detectors (e.g., Fast-DetectGPT) experience performance drops under paraphrase or heavy editing (Bao et al., 2023), and watermark approaches themselves require enough intact tokens for confident detection under paraphrasing or text mixing (Kirchenbauer et al., 2024). Practical recommendation**,** when provenance matters (policy, assessment, public communication), archive and timestamp the earliest public draft and direct VERMILLION assessment and any statistical detection to that first-available version; treat later revisions as progressively harder to attribute and interpret any single cue in context.

6. Limitations/Discussion

Our analysis is probabilistic and heuristic rather than definitive. Genre conventions (e.g., policy prose), copy-editing processes, and paraphrasing tools can attenuate or erase surface cues. VERMILLION should be used in conjunction with statistical detectors and transparent disclosure practices; no single method suffices in isolation (Gehrmann et al., 2019; Mitchell et al., 2023; Kirchenbauer et al., 2023). The heuristics for humanizing text (Table 3) are not just cosmetic, they restore human rhythm, agency, and emotion to otherwise mechanical prose. By embedding variation, context, and reflection, revised texts move beyond the flatness of AI outputs and regain the literary and cognitive qualities defining human writing. In educational contexts, such revision strategies offer a powerful bridge to guide students from passive consumption of AI outputs toward conscious, ethical, and expressive authorship.

7. Conclusion - Toward Responsible Authorship in Age of AI

As generative AI tools increasingly mediate the creation of professional, academic, and policy documents, the distinction between human and machine authorship has become both technically and ethically urgent. This article introduces the VERMILLION framework, an interpretive, heuristic method for identifying linguistic and cognitive markers typical of AI-generated text. Through detailed application to the MAHA Report, we demonstrate how specific stylistic indicators such as mechanical transitions, repetitive structure, hedging, and absence of personal voice can cumulatively signal probable AI involvement.

Importantly, this approach does not aim to vilify AI authorship but to encourage transparency, accountability, and informed critique. As the boundaries of collaboration between humans and machines continue to blur, the responsibility lies with writers, educators, publishers, and policymakers to establish norms that preserve clarity about authorship. Detection frameworks like VERMILLION can be integrated into editorial review processes, educational curricula, and AI-assisted writing platforms to maintain epistemic trust and safeguard human originality.

[2] No peer-reviewed evidence that later model releases systematically reduce em-dash frequency

References

- Abbasi, A., & Chen, H. (2008). Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems, 26*(2), Article 7.
- Al Jazeera. (2025, May 30). White House to amend flagship health report citing phantom studies. https://www.aljazeera.com/news/2025/5/30/white-house-to-amend-flagship-health-report-citing-phantom-studies
- Associated Press. (2025, May 30). White House acknowledges problems in RFK Jr.'s 'Make America Healthy Again' report. https://apnews.com/article/maha-report-errors-rfk-health-studies-f382af8552dbc1729329a13e58f1f3c4
- Bao, G., Zhao, Y., Teng, Z., Yang, L., & Zhang, Y. (2023). Fast-DetectGPT: Efficient zero-shot detection of machine-generated text via probability curvature. *arXiv.* https://arxiv.org/abs/2310.05130
- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97–116.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263–286.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2–20.
- Birhane, A., Prabhu, V. U., & Kahembwe, E. (2022). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. *arXiv.* https://arxiv.org/abs/2110.01963
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency (FAT 2018)* (pp. 149–159). https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3086546_code2696457.pdf?abstractid=3086546&mirid=1
- Brennan, M., Afroz, S., & Greenstadt, R. (2012). Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Transactions on Information and System Security, 15*(3), 1–29.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186. https://doi.org/10.1126/science.aal4230
- COPE (Committee on Publication Ethics). (2022). *AI and authorship ethics: A position statement.*https://publicationethics.org
- Darmon, A. N. M., et al. (2019). Pull out all the stops: Textual analysis via punctuation sequences. *arXiv.*https://arxiv.org/abs/1901.00519
- Gehrmann, S., Strobelt, 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* (pp. 111–116). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-3019
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*(2), 203–225.
- Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). The curious case of neural text degeneration. In *International Conference on Learning Representations (ICLR).* https://openreview.net/forum?id=rygGQyrFvH

- Hyland, K. (1998). *Hedging in scientific research articles.* John Benjamins.
- Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*(8), 1091–1112.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing.* Continuum.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21.
- Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2020). Automatic detection of generated text is easiest when humans are involved? *arXiv.* https://arxiv.org/abs/1911.00650
- Jafariakinabad, F., & Hua, K. A. (2021). Unifying lexical, syntactic, and structural representations of written language for authorship attribution. *SN Computer Science, 2*, 481.
- Jafariakinabad, F., Tarnpradab, S., & Hua, K. A. (2020). Syntactic neural model for authorship attribution. In *Proceedings of FLAIRS-33.*
- Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A watermark for large language models. *arXiv.* https://arxiv.org/abs/2301.10226
- Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology, 60*(1), 9–26.
- 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 (NeurIPS).*https://proceedings.neurips.cc/paper_files/paper/2023/hash/575c450013d0e99e4b0ecf82bd1afaa4-Abstract-Conference.html
- MAHA Commission. (2025, May 22). *Make America Healthy Again report.* White House. https://www.whitehouse.gov/wp-content/uploads/2025/05/MAHA-Report-The-White-House.pdf
- Marcus, G., & Davis, E. (2020). *Rebooting AI: Building artificial intelligence we can trust.* Vintage.
- Masrour, E., Emi, B., & Spero, M. (2025). DAMAGE: Detecting adversarially modified AI-generated text. *arXiv.*https://arxiv.org/abs/2501.03437
- Mitchell, E., Lee, Y. T., Lin, C., Han, J., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-shot machine-generated text detection using probability curvature. *arXiv.* https://arxiv.org/abs/2301.11305
- Obrenovic, B., Asa, A. R., & Oblakovic, G. (2025). The use of ChatGPT in the workplace: A bibliometric analysis of integration and influence trends. *AI & Society*, 1–14.
- 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. https://doi.org/10.1186/s41239-024-00487-w
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology, 60*(3), 538–556.
- Stamatatos, E. (2016). Authorship verification: A review of recent work. In *CLEF 2016 Working Notes* (pp. 1–13).
- Wolf, F., Gibson, E., & Desmet, T. (2004). Discourse coherence and pronoun resolution. *Language and Cognitive Processes, 19*(6), 665–675.