HUMAN vs. AI WRITING STYLE: A LINGUISTIC AND QUANTITATIVE FRAMEWORK

1. OVERVIEW Human and AI writing diverge systematically across measurable linguistic and rhetorical dimensions. Even when both are grammatically correct, AI-generated text displays statistically distinctive distributions of grammatical, lexical, and stylistic features. The HAP-E (Human–AI Parallel English) corpus identifies these features across 66 linguistic dimensions and 10 genres.

AI writing tends to be: overly formal (few contractions, high preposition density), uniform (low variance across tone and genre), structurally dense (complex clause stacking), impersonal (few pronouns, low emotional register), consistent and symmetrical in rhythm and syntax. Human writing is adaptive and expressive, rhythmically varied, context-sensitive, and characterized by imperfection, emotion, and narrative variation.

2. QUANTITATIVE METRICS (PER 1,000 WORDS) Contractions: Human 18.1 | AI 10.9 (-39%) 1st-Person Pronouns: Human 35.3 | AI 21.8 (-38%) Prepositions: Human 98.0 | AI 115.3 (+18%) Past-Tense Verbs: Human 41.9 | AI 34.6 (-17%) Mean Word Length: Human 4.4 | AI 4.6 (+5%)

3. LINGUISTIC SIGNATURES BY GENRE General Writing: Human 15–35 contractions, variable tone. AI: Fewer contractions, steady tone. Legal: Human idiomatic phrasing; AI: templated clauses. Business: Human adaptive tone; AI: over-structured phrasing. Academic: Human varies by discipline; AI: excessive precision. Technical: Human clear and stepwise; AI: over-explanatory. Fiction: Human emotional rhythm; AI: monotone prose. Nonfiction: Human narrative-reflective mix; AI: factual but flat. Journalism: Human past tense, direct; AI: literary diction. Narrative: Human episodic pacing; AI: uniform exposition. Script: Human dialogue rhythm; AI: perfect grammar, even cadence.

4. QUANTITATIVE SEPARABILITY (HAP-E CLASSIFIER) AUC = 0.98. AI-likeness by genre: Legal 1.00, Technical 0.99, Academic 0.98, Business 0.65, Journalism 0.22, Nonfiction 0.12, General 0.04, Fiction 0.01, Narrative 0.00, Script 0.00.

5. CORE LINGUISTIC DIMENSIONS Formality: Humans adapt contractions; AI uniformly formal. Complexity: Humans vary syntax; AI over-dense. Narrativity: Humans use past tense; AI overuses present. Personal Voice: Humans use pronouns; AI detached. Variance: Humans show rhythm shifts; AI uniform.

6. RHETORICAL & VOCABULARY DIFFERENCES AI overuses participial clauses, nominalizations, and ornate words ("camaraderie," "amidst," "tapestry"). Humans vary diction and tone naturally, include imperfections, idioms, and digressions.

7. QUANTITATIVE TARGET RANGES (PER 1,000 WORDS) Informal (Blog): 25–40 contractions, 40–70 pronouns, 60–90 prepositions. Moderate (News): 8–20 contractions, 10–30 pronouns, 90–120 prepositions. Formal (Academic): 0–5 contractions, 0–15 pronouns, 110–140 prepositions.

8. MULTI-DIMENSIONAL DIAGNOSTIC FRAMEWORK Formality (Contractions <10 = AI), Complexity (Prepositions >130 = AI), Narrativity (Past <30 = AI), Personal Voice (Pronouns <20 = AI), Variance (low = AI).

9. CHECKLIST OF AI WRITING DETECTION CUES Predictable syntax, redundancy, even tone, recycled phrases, perfect structure, impersonal voice.

10. INTERPRETIVE MODELS PCA: PC1 = stylistic variance (human diversity vs AI uniformity); PC2 = formality ↔ narrativity. Radar profile: Humans high in narrativity, voice, variance; AIs high in formality, complexity.

11. PRACTICAL APPLICATIONS Use contraction, pronoun, and preposition ratios for classification. Flag overly perfect or monotonous writing as likely AI.

12. SYNTHESIS: THE HUMAN SIGNATURE Human writing breathes; AI writing balances. Human: irregular rhythm, adaptive voice, flexible tense, emotional nuance. AI: uniform tone, dense syntax, formal vocabulary, low variation.

13. QUANTITATIVE SUMMARY TABLE Contractions: 18.1 vs 10.9 (-39% AI) 1st-Person Pronouns: 35.3 vs 21.8 (-38% AI) Prepositions: 98.0 vs 115.3 (+18% AI) Past Tense: 41.9 vs 34.6 (-17% AI) Word Length: 4.4 vs 4.6 (+5% AI)

14. CONCLUSION Human writing shows variance, voice, and adaptability. AI writing shows consistency, density, and detachment. Linguistic distribution metrics—contractions, pronouns, prepositions, and sentence variance—form the foundation for identifying authorship. When text feels too perfect or balanced, it's AI; when it shifts tone, shows emotion, and breaks rhythm, it's human.