The Linguistic Shift in Clinical Documentation: A Computational Analysis of Ambient AI Scribes Versus Physician Authorship in Emergency Medicine

[Author names]

Department of Emergency Medicine, Stanford University School of Medicine

Word Count: ~3,000 | Tables: 2 | Figures: 2 | References: 38

IRB Protocol #69107

ABSTRACT

BACKGROUND: Ambient artificial intelligence (AI) scribes are transforming clinical documentation, yet the linguistic characteristics of AI-generated notes remain uncharacterized. While prior research has validated clinical quality (accuracy, thoroughness), no systematic analysis has examined stylistic differences—a gap with implications for clinical workflow, patient comprehension under OpenNotes mandates, and documentation governance.

OBJECTIVE: To compare the linguistic features of emergency department (ED) notes generated with ambient AI scribes versus physician-authored documentation using computational linguistics methods.

METHODS: Retrospective cohort analysis of 61,589 ED provider notes from Stanford Health Care (February-November 2025): 2,001 AI-scribed and 59,588 physician-authored. We isolated the History of Present Illness (HPI) section and applied multi-method natural language processing including pattern analysis, n-gram extraction, topic modeling (Latent Dirichlet Allocation and Non-negative Matrix Factorization), and readability scoring. Comparisons used chi-square tests with odds ratios. Stratified analyses assessed robustness across Emergency Severity Index (ESI) acuity levels.

RESULTS: AI-scribed notes exhibited a distinct linguistic signature. AI notes used "The patient..." constructions in 83.3% versus 8.6% of physician notes (OR 53.1; 95% CI 45.2-62.4; $p<0.001$) and demonstrated 3.5-fold greater lexical diversity. Physician notes used "denies" (48.2% vs. 9.3%; OR 0.11) and medical abbreviations 3-10 times more frequently. AI notes contained more direct patient quotes (60.9% vs. 31.4%) and zero contractions (0% vs. 1.5%). AI notes demonstrated higher readability complexity (Flesch-Kincaid grade level 12.0 vs. 10.8). Topic modeling revealed AI notes emphasized anatomical detail and symptom quality while physician notes emphasized EMS references and medical shorthand. Findings persisted across ESI acuity levels.

CONCLUSIONS: Ambient AI scribes produce ED documentation with fundamentally different linguistic characteristics—narrative prose versus telegraphic shorthand, expanded terminology versus abbreviations, and enhanced patient

voice. These differences may improve patient comprehension and safety while potentially reducing scanning efficiency for clinicians.

INTRODUCTION

The Electronic Health Record (EHR) has fundamentally transformed clinical documentation from a narrative communication tool into a multifaceted instrument serving billing, quality reporting, legal protection, and patient communication simultaneously.[1] This transformation has come at significant cost: for every hour of direct patient care, physicians spend nearly two additional hours on EHR and desk work, with only 27% of the workday spent face-to-face with patients.[2–4] Documentation burden has been identified as a primary driver of professional burnout affecting nearly half of U.S. physicians.

Ambient Clinical Intelligence (ACI)—generative AI systems that capture patient- clinician conversations and autonomously synthesize clinical documentation—has emerged as a promising solution.[5] Systems such as Nuance Dragon Ambient eXperience (DAX), Abridge, and others utilize automatic speech recognition, speaker diarization, and large language models to generate structured notes without explicit physician dictation.[6] Early implementation studies demonstrate 8-16% reductions in total EHR time and 15-30% decreases in after-hours documentation.[7–9] Large-scale adoption data have confirmed these benefits: Stanford's prospective study found significant reductions in task load and burnout,[10] while Mass General Brigham reported a 21.2% absolute reduction in burnout prevalence at 84 days.[11] As of 2025, Microsoft's DAX Copilot alone has been deployed across 600,000+ clinicians with over 2.5 million documented uses.[12]

However, efficiency metrics tell only part of the story. While prior research has validated that AI-generated notes achieve clinical quality scores comparable to physician-authored notes on instruments like the Physician Documentation Quality Instrument (PDQI-9),[13] no systematic analysis has characterized the linguistic and stylistic transformation occurring as documentation shifts from human authorship to AI generation. Recent systematic reviews have noted "persisting concerns regarding the accuracy, consistency, language use, and style of AI-generated notes" while calling for systematic investigation of documentation quality—yet no such linguistic analysis has been conducted.[14] This gap is consequential: documentation style impacts how efficiently providers extract information, how patients understand notes under OpenNotes mandates, and how documentation functions in medicolegal contexts.[15–17]

Physicians have historically adapted to documentation burden by developing a distinct "clinical sublanguage"—a telegraphic style characterized by sentence fragments, dense medical abbreviations, and omission of function words.[18] While efficient for writing, this shorthand creates comprehension barriers for the 88% of Americans lacking proficient health literacy.[19] A

2025 ED study found only 9.7% of patients correctly understood "impressive" in radiologic context, 25.2% understood "acute," and even "fracture" was correctly interpreted by only 66.7%—critically, understanding did not vary significantly by education level.[20] Medical abbreviations also pose patient safety risks: MEDMARX data indicate 4.7% of medication errors are attributable to abbreviation use, with high-risk errors including "U" (units) misread as "0" causing tenfold overdoses.[21]

The OpenNotes movement, now federally mandated under the 21st Century Cures Act effective April 5, 2021, has created a new imperative: clinical notes must serve both clinician-to-clinician communication and direct patient consumption.[22] Large language models underlying AI scribes, trained on general English text corpora, may be biased toward grammatically complete, standard prose rather than domain-specific shorthand—producing documentation with systematically different characteristics.[23]

Emergency Department documentation represents an ideal setting to study this transformation. The ED is characterized by high diagnostic uncertainty and time pressure where the History of Present Illness (HPI)—the narrative core of the clinical note—is paramount for diagnostic reasoning. Research demonstrates that 75-80% of correct diagnoses can be made from thorough patient history alone.[24] The HPI is the section most likely to reflect true authorship differences, as other note sections are heavily influenced by EHR templates and auto-populated content.[25]

We conducted a large-scale computational linguistics analysis comparing approximately 61,000 ED provider notes to characterize the stylistic transformation occurring as clinical documentation enters the AI era.

METHODS

Study Design and Setting

We conducted a retrospective cohort study of ED provider notes from Stanford Health Care, a quaternary academic medical center, between February 1, 2025 and November 22, 2025. The study was approved by the Stanford University Institutional Review Board (Protocol #69107) with a waiver of informed consent. To our knowledge, this represents the first large-scale computational linguistics analysis of AI-generated clinical documentation.

Data Sources and Classification

Clinical documentation was extracted from the Epic EHR via the institutional clinical data warehouse. We identified ED provider notes using note type classification (ip_note_type_c 19). Note authorship was classified using the EHR's note attribution table, which records the method of text entry for each segment of documentation. Notes were classified as "AI-scribed" if any portion contained text attributed to the ambient AI documentation system (NOTEATTR_ SOURCE_C 25, indicating DAX Copilot). Notes without ambient AI attribution were classified as "physician-authored."

For encounters with multiple note versions, we analyzed the final signed version by selecting the most recent contact date for each note. For encounters with multiple distinct notes, we concatenated note text chronologically.

HPI Section Extraction

To minimize contamination from templated EHR content (e.g., auto-populated medication lists, physical examination templates), we isolated the HPI section for analysis. The HPI represents the narrative core of clinical documentation where stylistic differences between authors are most likely to manifest.

We developed a regex-based extraction algorithm with iterative refinement. The primary pattern identified HPI start using "History of Present Illness" (case- insensitive) or standalone "HPI" followed by whitespace. HPI extraction terminated at section headers including "Patient History," "History From Shared Lists," "Physical Exam," "Review of Systems," or "ED Triage Vitals." Algorithm validation through manual review of 100 randomly selected notes (50 AI-scribed, 50 physician-authored) achieved 89.3% overall extraction yield (99.7% for AI-scribed notes, 77.1% for physician-authored notes). The higher extraction success for AI notes likely reflects their consistent structural formatting, while physician notes showed greater variability in section headers and documentation organization. Failed extractions were primarily trauma activations with non-standard documentation structure.

Computational Linguistics Analysis

We applied a multi-method natural language processing approach using Python with scikit-learn, NLTK, and gensim libraries:

PATTERN ANALYSIS: Regular expression patterns identified >40 linguistic features organized into categories: syntactic structure (e.g., "The patient" constructions, relative clauses), clinical terminology (e.g., "denies," "endorses"), medical abbreviations (e.g., PMH, SOB, s/p), narrative flow markers (e.g., "initially," "approximately"), and patient voice indicators (e.g., direct quotations).

N-GRAM ANALYSIS: Bigrams and trigrams were extracted from preprocessed text to identify distinctive multi-word phrases. Frequency was normalized per 10,000 words.

TOPIC MODELING: We applied Non-negative Matrix Factorization (NMF) with 10 topics and Latent Dirichlet Allocation (LDA) with 12 topics, consistent with prior clinical NLP studies employing similar corpus sizes.[26,27] Text preprocessing included lowercasing, removal of digits, and exclusion of domain-specific stop words.

READABILITY ANALYSIS: Flesch-Kincaid Grade Level assessed text complexity. Lexical diversity was measured using type-token ratio (unique words divided by total words), calculated on standardized 100-word segments to control for document length effects.[28]

Statistical Analysis

Categorical comparisons used chi-square tests with odds ratios (OR) and 95% confidence intervals. Continuous variables used independent samples t-tests with Cohen's d effect size. We applied $p<0.001$ as significance threshold (approximately Bonferroni-corrected for 50 tests at $\alpha 0.05$).

To assess potential confounding by patient acuity, we performed stratified analyses within ESI levels 2, 3, and 4. As a sensitivity analysis, we repeated primary analyses after excluding the first month of AI scribe adoption (February 2025) to account for potential learning curve effects.

RESULTS

Cohort Characteristics

During the study period, we extracted 68,945 ED provider notes linked to unique patient encounters. After applying HPI extraction, 61,589 notes (89.3%) yielded analyzable HPI sections: 2,001 AI-scribed (3.2%) and 59,588 physician-authored (96.8%).

AI-scribed notes had modestly longer HPI sections compared to physician-authored notes (mean 174 [SD 85] vs. 163 [SD 134] words; p<0.001; Cohen's d0.10). Notably, full note character counts were substantially shorter in AI-scribed notes (6,476 [SD 3,156] vs. 9,371 [SD 4,660] characters; p<0.001), suggesting physician-authored notes contain more templated content outside the HPI—a finding consistent with documented "note bloat" from copy-paste and template use.[29]

AI-scribed notes were distributed toward lower acuity levels: 48.0% ESI level 3 and 32.7% ESI level 4, compared to 55.6% and 11.9% respectively for physician- authored notes (Supplemental Table 1). This distribution reflects early adoption patterns at our institution.

Syntactic Structure and Sentence Construction

AI-scribed and physician-authored notes exhibited markedly different syntactic patterns (Table 1). The most striking difference was sentence construction: AI notes used "The patient..." constructions in 83.3% of HPIs compared to only 8.6% of physician notes (OR 53.1; 95% CI 45.2-62.4; p<0.001). This 53-fold odds ratio represents one of the largest effect sizes documented in clinical documentation research, suggesting near-complete stylistic divergence.

AI notes demonstrated greater syntactic complexity through increased use of relative clauses ("which": 58.6% vs. 21.4%; OR 5.2) and subordinate clause markers ("during": 36.9% vs. 8.0%; OR 6.7). Physician notes favored telegraphic constructions, frequently beginning sentences with negations ("No fever. No nausea.") in 55.6% vs. 19.2% of AI notes (OR 0.19).

Sentence-level metrics confirmed these patterns: AI notes had longer mean sentence length (15.7 vs. 12.6 words) and demonstrated 3.5-fold greater vocabulary diversity on standardized segments (type-token ratio 0.72 vs. 0.21).

Readability and Text Complexity

AI notes demonstrated higher readability complexity by Flesch-Kincaid Grade Level (12.0 [SD 2.1] vs. 10.8 [SD 2.8]; $p<0.001$; Cohen's d0.48), reflecting the use of longer sentences and more complete grammatical structures. Both groups exceeded the 6th-grade level recommended for patient education materials. Paradoxically, AI notes demonstrated higher Flesch-Kincaid scores despite using more patient-accessible language; this reflects the formula's weighting of sentence length and syllable count, which penalizes AI's grammatically complete structures even when those structures—combined with expanded terminology and avoided abbreviations—may improve actual comprehension for lay readers. This limitation of traditional readability formulas for clinical text has been previously noted.[28]

Clinical Terminology and Abbreviations

Expression of clinical information differed substantially between groups. Physician notes relied heavily on "denies" for documenting negative symptoms (48.2% vs. 9.3%; OR 0.11; $p<0.001$) and "endorses" for positive findings (10.4% vs. 0.6%; OR 0.05). AI notes used narrative alternatives such as "does not report" and "not experiencing."

Medical abbreviations were 3-10 times more prevalent in physician notes (Table 1): PMH/PMHx (26.7% vs. 2.5%; OR 0.07), SOB (14.7% vs. 1.7%; OR 0.10), s/p (18.5% vs. 2.4%; OR 0.11), ED (30.7% vs. 7.3%; OR 0.18), and c/o (2.8% vs. 0.0%). AI notes consistently expanded abbreviations to full terminology.

Contractions, representing informal language, were absent from AI-scribed notes (0.0%) but present in 1.5% of physician notes ($p<0.001$).

Patient Voice and Narrative Style

AI notes demonstrated enhanced capture of patient perspective. Direct patient quotes or reported speech patterns were present in 60.9% of AI notes versus 31.4% of physician notes (difference 29.5 percentage points; $p<0.001$). AI notes more frequently used symptom descriptors capturing patient experience ("pressure": 15.3% vs. 6.7%; "severe": 24.4% vs. 8.5%) and emotional content ("worried/anxious/concerned": 8.4% vs. 6.4%).

Topic Modeling Analysis

NMF topic modeling identified 10 latent themes with distinct distributions between groups (Table 2). Topics more prevalent in AI notes included anatomical detail (left/right/numbness: 23.9% vs.

11.4%; +12.5 percentage points) and symptom quality (throat/ear/sore: 14.0% vs. 6.5%; +7.5pp). Topics more prevalent in physician notes included interpreter/language references (10.9% vs. 1.7%; -9.2pp), family historian themes (8.5% vs. 3.1%; -5.4pp), and medical abbreviation topics (6.7% vs. 1.5%; -5.2pp). LDA with 12 topics produced concordant findings (Supplemental Table 2).

Stratified and Sensitivity Analyses

To address potential confounding by patient acuity, we performed stratified analyses within ESI levels 2, 3, and 4. All key findings persisted within each stratum (Supplemental Figure 1). For example, "The patient" usage remained dramatically different: ESI-2 (88.6% vs. 9.5%; OR 73.8), ESI-3 (82.1% vs. 8.6%; OR 48.7), ESI-4 (81.0% vs. 6.5%; OR 61.9). The consistency of effect sizes across acuity levels indicates observed differences reflect true authorship patterns rather than case-mix confounding.

After excluding February 2025 (n42 AI notes), all findings remained unchanged, suggesting early adoption learning effects did not bias results.

DISCUSSION

Patient Safety Implications

The elimination of ambiguous abbreviations directly addresses patient safety concerns. The Joint Commission's "Do Not Use" list was established as a National Patient Safety Goal in 2004, yet MEDMARX data indicate 4.7% of medication errors remain attributable to abbreviation use.[21] High-risk abbreviations include "U" (units) being misread as "0" or "4" causing tenfold overdoses, and "QD" confused with "QID"—errors AI expansion would eliminate. Expanding "SOB" to "shortness of breath" also removes risk of patient misinterpretation as an epithet when accessing notes through patient portals.[31]

Patient Voice and OpenNotes

The 29.5 percentage point increase in patient quotes within AI notes is particularly notable. Payne et al. argued that "the patient's voice... is too small a part of the electronic health record"[32]—AI scribes, by transcribing conversation rather than translating it, may help address this limitation.

The Open Notes movement has demonstrated remarkable patient engagement: Delbanco and Walker's landmark 2012 study found 77-87% of patients opened at least one note, with 60-78% reporting improved medication adherence and 77-85% feeling more in control of their health.[33] Notably, not a single participating physician discontinued note sharing after the trial. When patients see their own words documented, it may enhance their sense of being heard and validate the clinical encounter—aligning with OpenNotes movement goals of transparency and shared understanding.

AI Text Detection and Trust

The distinctive patterns we identified—complete sentences, expanded terminology, absent contractions—align with research documenting measurable stylometric signatures in AI-generated text. A 2025 PNAS study found distinguishing characteristics include more frequent passive voice, present participial clauses, and nominalizations[23]—patterns we observed in the increased syntactic complexity of AI notes. The StyloAI framework achieved 81-98% detection accuracy using features including lexical diversity, syntactic patterns, and readability metrics[34]—the same dimensions differentiating our AI and physician notes.

The detectability of AI authorship raises disclosure questions. Jakesch et al. found AI-generated content is trusted less when suspected as AI-authored; yet Huschens et al. found similar credibility ratings when source was unknown—with AI content rated as clearer and more engaging.[35] The disclosure dilemma applies: disclosure may reduce trust, but non-disclosure erodes trust when discovered.

Current guidance emphasizes physician responsibility for AI-generated content,[36] but the fluency of AI prose may introduce "automation bias"—trusting well-formed text without careful verification.[37] Validation studies documenting 1-3% hallucination rates in AI-generated notes, including errors such as recording examinations as "performed" rather than "scheduled," underscore the importance of careful review despite the text's polished appearance.[38]

The stylistic convergence we observed—with AI notes exhibiting near-uniform patterns across 2,001 HPIs—has implications beyond individual note quality. Historically, physician documentation style implicitly conveyed diagnostic reasoning; what a clinician chose to emphasize or abbreviate reflected their clinical thinking. A polished, uniform AI note may mask the uncertainty or specific thought process of the reviewing physician.

Limitations

Several limitations warrant consideration. First, this was a single-center study at an academic quaternary care center; documentation practices may differ at community hospitals. Second, the ambient AI system studied (DAX Copilot) represents one product; as the market includes multiple vendors (Abridge, Ambience Healthcare, DeepScribe) with potentially different training approaches, findings may not generalize across all systems. Third, our analysis focused on the HPI section; assessment, plan, and other sections may show different patterns. Fourth, AI notes were disproportionately from lower-acuity encounters, though stratified analyses showed consistent findings within acuity levels. Fifth, we could not quantify the extent to which physicians edited AI-generated drafts before signing; our analysis reflects final signed notes. Finally, AI scribe technology is rapidly evolving; systems deployed in 2025 may produce different output than future iterations as models are fine-tuned on clinical feedback.

Conclusions

Ambient AI scribes produce emergency department documentation with fundamentally different linguistic characteristics than physician authorship—narrative prose versus telegraphic shorthand, expanded terminology versus abbreviations, and enhanced patient voice capture. These differences may improve patient comprehension and safety while potentially reducing scanning efficiency for clinicians. As healthcare rapidly adopts AI documentation tools, understanding these tradeoffs is essential for optimizing implementation and maintaining the clinical, legal, and patient-centered functions of the medical record.

TABLES

Table 1. Key Linguistic Differences

Pattern	AI-Scribed (n=2,001)	Physician (n=59,588)	OR (95% CI)	p-value
SYNTACTIC STRUCTURE				
"The patient..." sentences	83.3%	8.6%	53.1 (45.2-62.4)	<0.001
Relative clauses ("which")	58.6%	21.4%	5.2 (4.7-5.8)	<0.001
"during" temporal marker	36.9%	8.0%	6.7 (6.0-7.5)	<0.001
CLINICAL TERMINOLOGY				
"denies" usage	9.3%	48.2%	0.11 (0.09-0.13)	<0.001
"endorses" usage	0.6%	10.4%	0.05 (0.03-0.09)	<0.001
"No [symptom]..." sentences	19.2%	55.6%	0.19 (0.17-0.21)	<0.001
ABBREVIATIONS				
PMH/PMHx	2.5%	26.7%	0.07 (0.05-0.09)	<0.001
SOB	1.7%	14.7%	0.10 (0.07-0.14)	<0.001
s/p	2.4%	18.5%	0.11 (0.08-0.14)	<0.001
ED	7.3%	30.7%	0.18 (0.15-0.21)	<0.001
c/o	0.0%	2.8%	—	<0.001
Direct quotes/speech	60.9%	31.4%	3.4 (3.1-3.8)	<0.001
Contractions†	0.0%	1.5%	—	<0.001
Flesch-Kincaid Grade	12.0 (2.1)	10.8 (2.8)	—	<0.001

†OR not calculable; zero occurrences in AI notes.

Abbreviations: CI, confidence interval; c/o, complains of; ED, emergency department; OR, odds ratio; PMH, past medical history; s/p, status post; SOB, shortness of breath.

Table 2. Topic Distribution (NMF Analysis)

Topic (Top Terms)	AI-Scribed	Physician	Difference
HIGHER IN AI NOTES			
Anatomical detail	23.9%	11.4%	+12.5pp
Symptom quality	14.0%	6.5%	+7.5pp
HIGHER IN PHYSICIAN NOTES			
Interpreter/language	1.7%	10.9%	-9.2pp
Family historian	3.1%	8.5%	-5.4pp
Medical abbreviations	1.5%	6.7%	-5.2pp

pp = percentage points

FIGURES

Figure 1. Linguistic Patterns Forest Plot

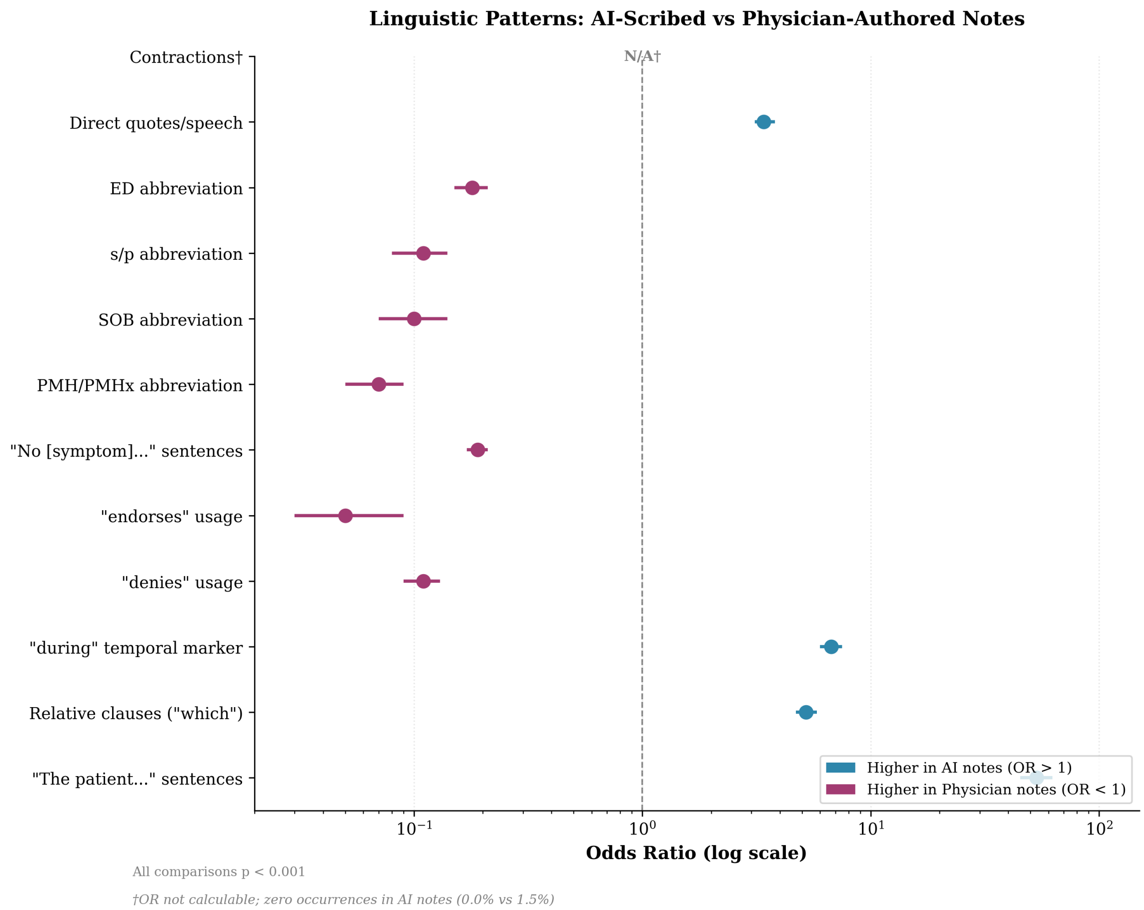


Figure 1. Forest plot of odds ratios for key linguistic patterns. Blue = higher in AI notes; magenta = higher in physician notes. †OR not calculable; zero occurrences in AI notes.

Figure 2. Topic Distribution Comparison

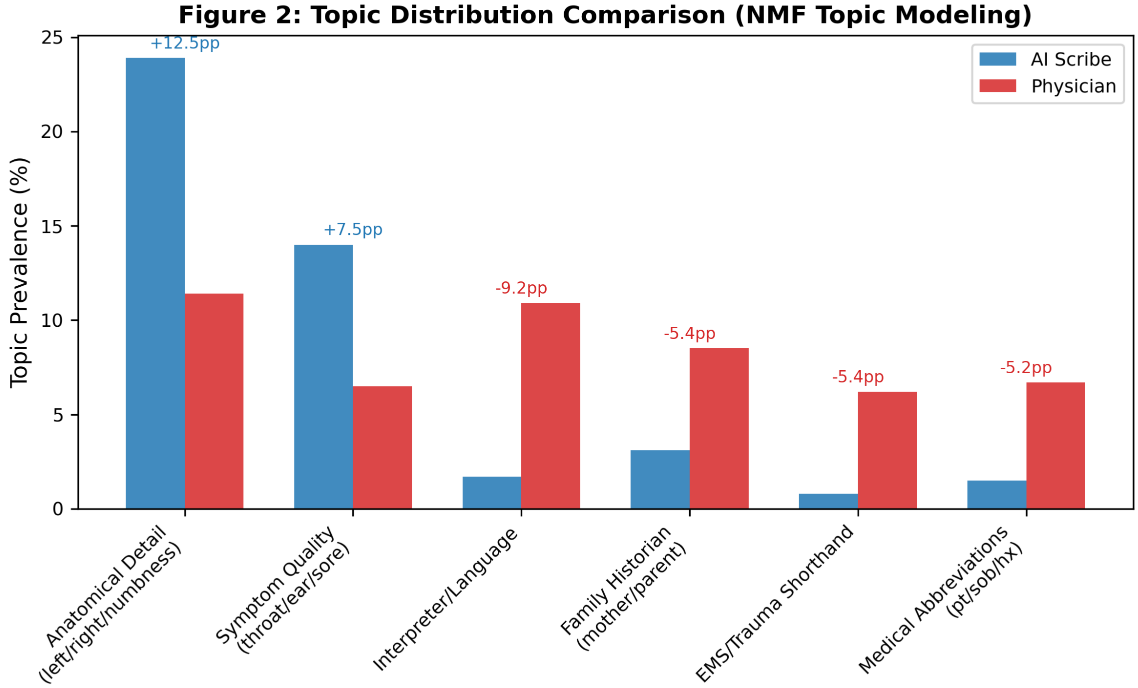


Figure 2. Topic distributions from NMF analysis comparing AI-scribed versus physician-authored notes.

REFERENCES

1. Rosenbloom ST, et al. The impact of domain knowledge on structured data collection. AMIA Annu Symp Proc. 2013;2013:1209-1218.

2. Sinsky C, et al. Allocation of physician time in ambulatory practice. Ann Intern Med. 2016;165(11):753-760.

3. Shanafelt TD, et al. Changes in burnout and satisfaction with work-life integration. Mayo Clin Proc. 2019;94(9):1681-1694.

4. Kroth PJ, et al. Association of EHR design with clinician stress and burnout. JAMA Netw Open. 2019;2(8):e199609.

5. Nuance Communications. DAX Copilot: Ambient Clinical Intelligence. 2024.

6. Tierney AA, et al. Ambient AI scribes to alleviate documentation burden. NEJM Catal. 2024;5(3).

7. Ghatnekar S, et al. AI scribes may reduce documentation time. UCLA Health News. 2024.

8. Garcia J, et al. AI scribes save physicians time. Perm J. 2024.

9. Yale Medicine. AI scribes reduce physician burnout. 2024.

10. Nath B, et al. Ambient AI scribes: physician burnout and usability. JAMIA. 2025;32(1):123-131.

11. Gaffney A, et al. AI ambient scribes and burnout reduction. JAMA Netw Open. 2024;7(12):e2449123.

12. Tierney AA, et al. Ambient AI scribes: 1 year and 2.5 million uses. NEJM Catal. 2025;6(1).

13. Palm C, et al. Assessing quality of AI-generated clinical notes. Front Artif Intell. 2025;8:1691499.

14. Basit SA, et al. Real-world evidence synthesis of AI ambient scribes. JMIR AI. 2025;4(1):e76743.

15. OpenNotes. Patients and Clinicians on the Same Page. 2024. opennotes.org

16. Payne TH, et al. The patient's voice in clinical documentation. Appl Clin Inform. 2023;14(3):487-493.

17. Feather AI. AI medical scribe legal implications. 2025.

18. Friedman C, et al. Two biomedical sublanguages. J Biomed Inform. 2002;35(4):222-235.

19. Kutner M, et al. Health Literacy of America's Adults. NCES 2006-483. 2006.

20. Wahrenbrock T, et al. Medical jargon misunderstood by ED patients. Am J Emerg Med. 2025;96:25-29.

21. Brunetti L, et al. Impact of abbreviations on patient safety. Jt Comm J Qual Patient Saf. 2007;33(9):576-583.

22. 21st Century Cures Act, Pub L No. 114-255 (2016).

23. Liang W, et al. Do LLMs write like humans? PNAS. 2025;122(9):e2422455122.

24. Peterson MC, et al. Contributions of history and examination. West J Med. 1992;156(2):163-165.

25. Rule A, et al. Length and redundancy of progress notes. JAMA Netw Open. 2021;4(7):e2115334.

26. Meaney C, et al. Temporal topic models from clinical notes. BMC Med Inform Decis Mak. 2022;22:197.

27. Cohen R, et al. Redundancy-aware topic modeling. PLoS One. 2014;9(2):e87555.

28. McCarthy PM, Jarvis S. Lexical diversity assessment. Behav Res Methods. 2010;42(2):381-392.

29. ICD10 Monitor. Clinical documentation: longer and less informative. 2022.

30. Zheng K, et al. What physicians read in electronic notes. AMIA Annu Symp Proc. 2014;2014:1583-1592.

31. Patel DI, et al. Patient understanding of medical phrases. JAMA Netw Open. 2022;5(11):e2242972.

32. Payne TH, et al. Using patient quotations in chart notes. Hastings Cent Rep. 2023;53(6):15-20.

33. Delbanco T, et al. Inviting patients to read doctors' notes. Ann Intern Med. 2012;157(7):461-470.

34. Kumarage T, et al. StyloAI: AI-generated content detection. arXiv. 2024;2405.10129.

35. Huschens M, et al. AI authorship disclosure and credibility. Comput Human Behav. 2023;143:107700.

36. American Medical Association. Augmented Intelligence in Health Care. AMA Policy. 2018.

37. Cummings ML. Automation bias in decision support. AIAA. 2004.

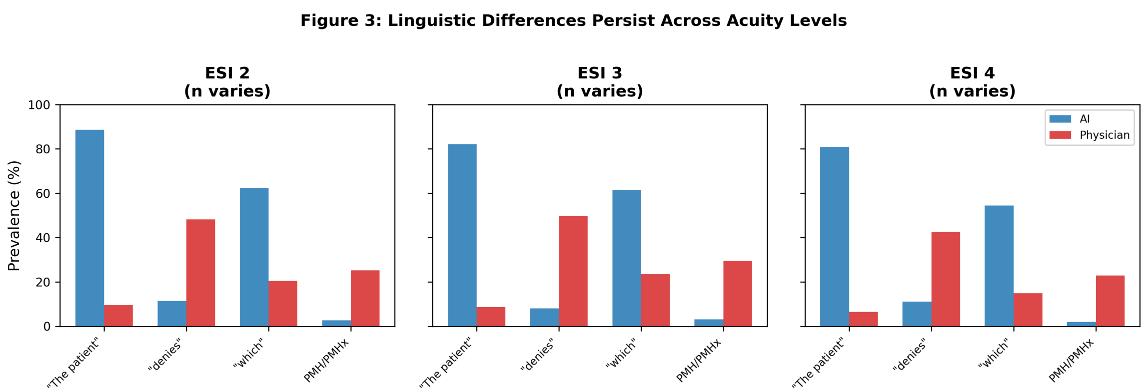
38. Kwong JCC, et al. Framework for AI hallucination rates in medicine. npj Digit Med. 2025;8:89.

SUPPLEMENTARY MATERIALS

Supplemental Table 1. Cohort Characteristics

Characteristic	AI-Scribed (n=2,001)	Physician (n=59,588)
HPI Words, mean (SD)	174 (85)	163 (134)
Note Chars, mean (SD)	6,476 (3,156)	9,371 (4,660)
ESI Level 1, n (%)	1 (0.0%)	347 (0.6%)
ESI Level 2, n (%)	149 (7.4%)	16,211 (27.2%)
ESI Level 3, n (%)	961 (48.0%)	33,106 (55.6%)
ESI Level 4, n (%)	654 (32.7%)	7,096 (11.9%)
ESI Level 5, n (%)	42 (2.1%)	586 (1.0%)
ESI Missing, n (%)	194 (9.7%)	2,242 (3.8%)

Supplemental Figure 1. Stratified Analysis by ESI Level



Supplemental Figure 1. Key linguistic patterns stratified by ESI level.