AI Scribing | CDS
Responsible Use & Risk Awareness

Daniel Fitzgerald, MD, MS
Director of Clinical AI · Clinical Innovation Group · US Acute Care Solutions

emRIC · APRIL 2026

Learning Objectives

This session will help provide you with the following:

Diagnose

Where AI scribe errors originate — distinguishing human input gaps from AI model gaps, and understanding why the fix is different for each.

Recognize

The failure modes that are hardest to catch — templated fabrications, silent omissions, and the cognitive biases that make polished AI notes feel more accurate than they are.

Apply

Practical behaviors at the bedside — to reduce risk, preserve your clinical reasoning, and use AI scribes as a tool rather than a replacement for your own thinking.

The State of Clinical AI, 2026

The promise is real. The evidence base is thin. Deployment is outpacing safety measurement.

THE PROMISE · AMBIENT SCRIBE ADOPTION

63%	of hospitals using Epic have adopted ambient AI scribes
83%	reduction in note-writing effort · Sharp HealthCare
47%	reduction in cognitive load · U. Chicago Medicine
112%	ROI · Northwestern Medicine (+11.3 pts/mo)

THE GAP · EVIDENCE & MEASUREMENT

5%	of 500+ clinical AI studies reviewed used real patient data
233→362	AI incidents (2024→2025) — 55% rise; safety reporting lags
0 / 10	# of NOHARM studies that are EM, ICU, or HM · (76.6% are omissions)
19%	slower with AI (METR RCT) — while feeling 20% faster

BOTTOM LINE · AI performs most effectively when supporting rather than replacing clinician judgment. (Stanford–Harvard ARISE Network, 2026)

The 3 Cs of AI Scribing

HI Gap = what you said or didn't say → Fix: how you speak to the scribe

AI Gap = what the model did with your input → Fix: how the system is prompted

C 01

CAPTURE

Physical acoustics

ERROR RATES >50% IN MULTISPEAKER ED SCENARIOS

C 02

CONTENT

Clinical density

70% OF AI NOTES CONTAIN ≥1 ERROR · 44% ARE 'MAJOR'

C 03

CONTEXT

Semantic framing

FRAME IT YOURSELF. DON'T GAMBLE ON THE GUESS.

The AI will always transform your input. The 3 Cs determine how much of that transformation reflects what you actually said versus what the model predicted.

AI-Generated Errors: What Can Go Wrong

EXTRACTION

Lifting data from the encounter

- Numeric substitution: "50 mg" → "15 mg"
- Entity swap: Metoprolol → Metformin
- Negation failure: "No allergies" → "Allergies: yes"
- Misattribution: nurse's statement → patient

ABSTRACTION

Paraphrasing & summarizing

- Intrinsic fabrication: contradicts the transcript
- Extrinsic fabrication: adds what was never said
- Chronology error: reorders the symptom timeline
- False relational inference / narrative conflation

INFERENCE

AI generating reasoning

- Error of omission: drops discordant data
- Error of commission: wrong recommendation
- Normalcy hallucination: invents "PERRL, EOMI"
- Diagnostic anchoring on early labels

OMISSIONS DOMINATE — 76.6% of severely harmful LLM errors are omissions (NOHARM, Wu 2025). In the ED: a missed STEMI workup, an undiagnosed PE, or a delayed sepsis bundle.

The Templated Normal: A Cascading Failure

When you tell the AI to document what you didn't do, it reasons from what you didn't find.

TEMPLATED NORMAL · AI-INSERTED

`Respiratory: No acute distress. Normal work of breathing. Lungs clear to auscultation bilaterally. No wheezing, rhonchi, or rales.`

WHAT WAS ACTUALLY HAPPENING

`Respiratory: Tachypneic to 28. Accessory muscle use. Audible inspiratory wheezing bilaterally. Speaking in 2–3 word sentences.`

AI-GENERATED A&P

"Asthma exacerbation, mild, improving. Continue albuterol nebulizer. Reassess for discharge."

WHAT IT SHOULD SAY

"Severe acute asthma exacerbation with respiratory distress. Continuous nebulization, IV magnesium, reassess for escalation."

13.5% of pre-AI exam documentation was already fabricated (Weiner et al., JAMIA 2020).
AI amplifies this at scale — and reasons from it.

Prompting Risks

PROMPT PATTERNS WITH POTENTIAL RISKS

"Provide the most likely diagnosis"

Triggers anchoring and automation bias. The AI is no longer documenting — it is practicing.

"Fill out normal PE"

Commands fabrication. Chart pollution. Potential billing fraud (CMS/OIG). 13.5% of pre-AI exam docs were already fabricated (Weiner 2020) — AI amplifies at scale.

Automated normal-exam templates in prompts

Templated normals become AI input. Without override, AI reasons from fabricated data — skewing the differential and A&P.

** LIABILITY FALLS ON THE CLINICIAN

AI scribes = administrative tools, NOT medical devices. No FDA oversight. You own the signed note.

EVIDENCE GAP ·

Only **5% of clinical AI studies use real patient data** (Stanford AI Index 2026). NOHARM specialties excluded EM, ICU, HM.

Cognitive Biases in the Human-AI Loop

Bias flows both directions. The clinician can be misled by the AI — and the AI by the clinician.

THE CASCADE

Fluency bias → Proofreader's illusion + Authorship illusion → Proofreading error → Signed chart → Feedback loop

FLUENCY BIAS

47% vs 39% — AI notes preferred despite 55% more hallucinations

PROOFREADER'S ILLUSION

Word errors harder to detect than misspellings (Sheridan 2022)

AUTHORSHIP ILLUSION

When output looks like what you'd write, you claim ownership — cutting scrutiny further.

AUTOMATION BIAS

Passive acceptance of AI output. Leads to errors of both omission and commission.

BIDIRECTIONAL ANCHORING

AI ↔ CLINICIAN. Clinician anchors on AI's reasoning chain, not just the dx. AI anchors on clinician's early label.

FEEDBACK LOOP BIAS

Signed errors become EHR data. Downstream AI reads past mistakes as ground truth.

The A&P and MDM: Minimize AI Inference

Not all note sections carry equal inferential risk. Prompt design should reflect that gradient.

LOWER RISK → Allow AI transformation	HIGHER RISK → Minimize AI rephrasing
HPI · Medication list · Exam findings · Social history	**A&P · MDM · Reassessment · Clinical reasoning**

1 DESKILLING

Outsource your reasoning articulation → stop practicing the cognitive work.

2 PROOFREADER'S ILLUSION

Proofreading someone else's version of your thinking — against memory, not your words.

3 TRANSPARENCY LOSS

Provenance gone. Can't distinguish which reasoning is yours vs. AI-constructed.

4 ADOPTION BARRIER

"This doesn't sound like me." 75–80% adoption only with customization.

5 FALSE RELATIONAL INFERENCE

More rephrasing = more opportunities for spurious connections and conflation.

6 DOWNSTREAM CONTAMINATION

Shifted narrative in HPI → flows into A&P → AI reasons from its own altered version.

PRINCIPLE · In A&P / MDM / reassessment: **dictate your reasoning explicitly in your own words.** Let the AI abstract the HPI — not your thinking.

What You Can Do Now

01 **Optimize your Capture**
Position mic intentionally. Minimize ambient noise. Poor audio is the #1 error source.

02 **Verbalize explicitly**
State positives and negatives aloud. Don't assume the AI will infer what you didn't say.

03 **Frame your Context**
Name the section, timeline, and clinical significance. Don't leave placement to the AI.

04 **Practice Dual Encoding**
Recap the patient's story — to them, or within your dictation. Speaking your reasoning augments cognition and gives the AI higher-fidelity input.

05 **Scrutinize the A&P first**
The A&P / MDM is the highest-risk section for AI inference. Review it before the rest — and before you sign.

06 **Name the gap**
When output is wrong: HI gap or AI gap? Send examples to the clinical AI team.

07 **Don't auto-populate**
Templated normals break integrity, contaminate AI reasoning, and deskill. Non-negotiable.

08 **Use the monologue**
ED workflow: narrate prior records, interp, MDM, disposition. No patient consent needed.

Documentation quality improved with H+AI — but factual error rates remained **26–36%**. **"Just review the note"** is necessary but insufficient.

Works Cited

Primary sources for statistics and frameworks referenced in this presentation.

AI SCRIBE EVIDENCE

Biro et al. AI scribe error rates. JMIR. 2025.
70% notes contain ≥1 error.

Asgari et al. npj Digit Med. 2025;8:274.
44% of hallucinations classified major.

Ng et al. BMC Med Inform Decis Mak. 2025;25:236.
50% error rate, multispeaker.

Topaz et al. npj Digit Med. 2025;8:569.
Speaker mix-up, transcription errors.

Pepin et al. Frontiers in AI. 2025;8:1691499.
47% vs 39% — AI notes preferred despite more hallucinations.

Morey et al. Ann Emerg Med. 2025.
Mayo ED: 4.3 vs 1.8 min editing burden.

Preiksaitis et al. Ann Emerg Med. 2026.
28% less documentation time; uneven adoption.

COGNITIVE & COLLABORATION

Gaube et al. Nat Commun. 2024;15:6890.
87.3% → 77.1%; 6.6% false confirmation.

Wang et al. npj Digit Med. 2026;9:195.
H+AI meta-analysis; 26–36% error rates.

Logan & Crump. Science. 2010;330(6004):683–686.
Cognitive illusions of authorship.

Sheridan et al. Memory & Cognition. 2022.
Self-generation effect in proofreading.

Becker et al. METR; arXiv:2507.09089. July 2025.
Developer productivity; 19% slower, 20% faster perceived.

Nilsson C. Transforming Government. 2025.
AI competence paradox; deskilling.

Lopez et al. Preprint 2025. arXiv:2603.14158.
Content echoing; 93.5% concordance.

SAFETY, BENCHMARKS, FRAMEWORKS

Wu et al. NOHARM. arXiv:2512.01241v2. Dec 2025.
11.8–14.6 severe errors/100; 76.6% omissions.

Weiner et al. JAMIA. 2020;27(5):770–775.
13.5% PE fabrication, pre-AI baseline.

Lukac et al. NEJM AI. 2025.
RCT of ambient AI scribes; omissions dominant.

Pines JM & Fitzgerald D. NEJM AI. 2026 (under review).
Dialogue / monologue / hybrid workflow framework.

Tierney et al. NEJM Catalyst. 2025;6(5).
Kaiser: 2.5M+ uses; 84% improved communication.

Sajadieh et al. Stanford AI Index 2026. Stanford HAI.
Deployment stats, incident trends, 5% real data.

Stanford–Harvard ARISE Network. State of Clinical AI Report. Jan 2026.
AI supports rather than replaces clinician judgment.