

Generative AI for Emergency Medicine Knowledge Transfer

Frameworks, Validation & Practical Strategies

Alexander Forrester, MS4 • FLEX449

Supervisors: Elizabeth Stacy & Dr. Braedon Paul

Project Purpose



The Problem

- EM conferences generate valuable educational content
- Translating content into usable frontline resources is **time-consuming and cumbersome**
- Critical updates may not reach clinicians in a timely manner
- Knowledge translation burden falls on already-stretched physicians



Project Purpose

Develop a **standard process** for using AI to facilitate knowledge transfer from the St. Paul's EM Update Conference

TEST CASE SESSIONS

- The Crashing Neonate – Dr. Garth Meckler
- Mastering Preeclampsia in the ED – Dr. Todd Ring
- Navigating Alcohol Use Disorders – Panel

Project Phases

PHASE 1

Setup & Framework

- Select AI tools (Gemini, ChatGPT, Claude)
- Research validation frameworks
- Develop validation checklist
- Establish comparison methodology
- Create feedback survey

PHASE 2

Generation & Validation

- Develop standardized prompt template
- Generate resources from 3 AI tools
- Apply validation checklist
- Compare tool outputs
- Select optimal AI tool

PHASE 3

Review & Finalization

- Expert presenter review
- Clinical committee feedback
- Incorporate revisions
- Transcribe to standard template (pending)
- Prepare for distribution (pending)

Key Deliverables: SOP, AI Prompt Template, Summary Resource Template, 3 Validated Summaries

Modified ACCURATE-LE Framework

A Accuracy – Verify information is true and correct

C Context – Does resource address session objectives?

C Clarity – Is resource clear and understandable?

U Up-to-date – Current with latest guidelines and data

R Relevance – Relevant to prompt and input material

A Attribution – Check citations and disclosures

T Thoroughness – Check for omissions

E Equity – Check for bias and appropriate language

P Privacy – Is patient privacy protected?

Scoring Methods

- 1-5 Scale for quality spectrum
- Yes/No for binary items
- Pass/Fail for critical safety

Maximum Score: 44

16 questions across accuracy, safety, and quality domains

[Evaluating GenAI Outputs - Generative Artificial Intelligence - Research Guides at University of Saskatchewan](#)

[AI Validation Checklist.xlsx](#)

Project Results



SELECTED AI TOOL

Claude OPUS 4.5

Validation Score Comparison

Session	Gemini	Claude
The Crashing Neonate	32	32
Mastering Preeclampsia	33	34
Alcohol Use Disorders	24	34

ChatGPT 5.1 Discontinued

Outputs contained unrelated content and erroneous information

Key Findings

Consistent Performance

Claude produced equal or better scores across all validation criteria

Prompt Engineering Critical

Standardized template with safety rules essential for quality

Human Oversight Required

Student first-pass + physician review ensures accuracy

Replicable Process

SOP and templates enable future knowledge translation

AMA Core Competencies for AI

1

Foundational Knowledge

What is this tool?

Understand development, validation, and regulation of the AI

2

Critical Appraisal

Why should I use this tool?

Evaluate evidence, benefits, limitations, appropriate use cases

3

Clinical Integration

When and where should I use this tool?

Incorporate AI into clinical decision-making appropriately

4

Technical Use

How should I use this tool?

Operate effectively while maintaining patient-clinician relationships

5

Addressing Unintended Consequences

What are the side effects?

Identify biases, automation risks, and liability concerns

FUTURE-AI Framework

F

Fairness

Equity and bias mitigation

ED: Equal performance for all patient groups?

U

Universality

Generalizability, interoperability

ED: Works with our EMR and equipment?

T

Traceability

Monitoring, auditing, accountability

ED: Can we track performance? Who is accountable?

U

Usability

Human-centered design, clinical utility

ED: Can staff use it during a busy shift?

R

Robustness

Reliability, resilience, safety

ED: Handles chaos and poor data quality?

E

Explainability

Transparency, interpretability

ED: Can it explain its recommendations?

TEHAI Framework

Translational Evaluation of Healthcare AI – From development to real-world practice

1 Capability

Was it built right?

Does the AI perform its intended function?
Examines data quality, bias, validation.

ED: Triage algorithm should be trained on diverse populations and validated externally.

2 Utility

Is it safe and usable?

Usability, safety, ethics once deployed.
Are outputs interpretable under time pressure?

ED: Who reviews false negatives? How can alerts be modified?

3 Adoption

Will it work here?

Integration into workflows, EHR compatibility, scalability across sites.

ED: Confirm Epic/Cerner integration, alert placement, downtime protocols.

Common Problem: Many AI models achieve high capability scores but fail in utility and adoption which is what the TEHAI Framework aims to identify.

Practical Strategies for the ED

Maintain Human Oversight

1

Combine AI recommendations with clinician review. Retain authority to override AI-driven decisions when clinical judgment requires.

Implement Verification Processes

2

Establish structured verification for AI-suggested diagnoses or treatments using reliable peer-reviewed medical sources. Require critical AI-prompted decisions to be documented with explanations.

Request Uncertainty Quantification

3

Encourage AI models to express confidence levels or flag uncertain outputs. Tools lacking such information should be viewed with caution.

Monitor for Bias

4

Continuous monitoring and auditing of AI system performance across demographic subgroups is essential for identifying and rectifying emerging disparities.

Key Questions to Ask

Adapted from AMA Core Competency Frameworks for evaluating AI tools in emergency medicine

What is this tool?

- What is the intended clinical use and target population?
- What type of AI model underlies it (ML, deep learning, LLM)?
- Is a model card available with training data and limitations?

Why should I use this tool?

- Has it been externally validated on similar populations?
- What are the reported performance metrics?
- Is there evidence from prospective trials on outcomes?

When and where should I use this tool?

- How does it integrate into existing clinical workflows?
- What are the known failure modes and edge cases?
- When should the tool NOT be used?

How should I use this tool?

- How should AI outputs be verified and documented?
- What level of physician oversight is required?
- How should conflicting AI recommendations be handled?

What are the potential adverse effects?

- What biases have been identified and how might they affect my patients?
- What is the risk of automation bias or over-reliance?
- Who bears liability for AI-influenced decisions?

These questions help clinicians critically appraise AI tools before integrating them into ED practice

Key Takeaways

1 AI enhances, not replaces, human judgment. Generative AI accelerates knowledge translation but requires rigorous validation and expert oversight.

2 Structured validation is essential. Use frameworks to ensure AI outputs are accurate, safe, and clinically appropriate.

ACCURATE-LE

FUTURE-AI

TEHAI

AMA Competencies

3 Develop AI competencies now. Foundational knowledge, critical appraisal, and understanding limitations will be essential as AI transforms ED workflows.

4 Practical tools make a difference. SOPs, prompt templates, and validation checklists enable safe, reproducible AI integration.

With thoughtful implementation, AI can bridge the gap between conference learning and clinical practice.

References and Contact Information

AMA. (2025). *Future of Health: The Emerging Landscape of Augmented Intelligence in Health Care*. American Medical Association. Retrieved December 3, 2025, from <https://www.ama-assn.org/system/files/future-health-augmented-intelligence-health-care.pdf>

USask. (2025, October 30). *Generative Artificial Intelligence: Evaluating GenAI Outputs*. Retrieved November 24, 2025, from University of Saskatchewan; University Library Research Guides: https://libguides.usask.ca/gen_ai/evaluating

Lekadir, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *The British Medical Journal*. 2025 Feb 5;388:e081554. doi: 10.1136/bmj-2024-081554.

Reddy, S., Rogers, W., Makinen, V.-P., Coiera, E., Brown, P., Wenzel, M., Weicken, E., Ansari, S., Mathur, P., Casey, A., Kelly, B. (2021). Evaluation Framework to Guide Implementation of AI Systems Into Healthcare Settings. *The British Medical Journal Health & Care Informatics*, 28, e100444. Retrieved December 2, 2025, from <https://informatics.bmj.com/content/28/1/e100444>

Alexander Forrester

Email: falex@student.ubc.ca