Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 23, 2026
A Regional Conference of the Society for Industrial and Systems Engineering

Designing a RAG-Enabled Clinical Support Tool for Army Sick Call: Prototype Development and a Lab-Based Workflow Evaluation

Kira MacMullan, Alex Zhang, Cody Bradford, Peter Segat, Elhadji Kone, and Ahmed Bahabry

Department of Systems Engineering, United States Military Academy, West Point, New York 10996

Corresponding author's Email: kira.macmullan@gmail.com

Author Note: The first five authors are cadets at the United States Military Academy who conducted the research described in this paper under the advisement of Dr. Ahmed Bahabry. Each member made significant contributions to this project, and a collective writing effort produced the paper. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of War.

Abstract: Army Sick Call operations are critical to soldier readiness but are often slowed by documentation burdens and inefficient patient flow. This capstone presents the stakeholder-informed design of a Retrieval-Augmented Generation (RAG)-enabled clinical support prototype to improve intake, prioritization, and provider decision support while remaining grounded in ADTMC doctrine. The tool was developed through stakeholder interviews, direct observation of sick call operations, and Systems Decision Process methods. To evaluate potential impact, we conducted a controlled lab-based experiment in which 33 cadets role-played standardized patients under both the current paper-based workflow and a prototype-enabled workflow. Performance was compared using measures including total cycle time, wait time, and form completion time. A stakeholder-weighted value model was also developed to assess system performance. Results suggest the prototype may reduce total process time and administrative burden, though further validation with providers is required to assess triage performance and clinical effectiveness.

Keywords: Retrieval Augmented Generation (RAGs), Triage, Systems Decision Process, Value Model, Stakeholder Analysis, Sick Call, Military Health System (MHS), Algorithm-Directed Troop Medical Care (ADTMC), Clinical Decision Support Systems, Multi-Criteria Decision Analysis (MCDA), First-In-First-Out (FIFO) Queuing

1. Introduction

Sick call is the primary point of contact in military healthcare, requiring rapid, accurate triage and documentation to maintain soldier readiness. However, current processes are inefficient, with providers spending significant time on administrative data entry rather than direct clinical care (Roberts & Wolfe, 2025). While civilian AI tools are widely available, they often lack the doctrinal rigor required by the Military Health System (MHS). Recent research shows that Retrieval-Augmented Generation (RAG) improves triage reliability by being able to ground responses in published doctrine (Miao, Zhao, Luo, Wang, & Wu, 2025).

This gap between efficiency and compliance is further highlighted in operational environments. A field study evaluating the Algorithm Directed Troop Medical Care (ADTMC) manual demonstrated that transitioning from a traditional paper-based system to a digitized application resulted in significant improvements in clinical workflow efficiency, including substantial reductions in both examination time and documentation burden (Chrosniak, Olsen, & Galdi, 2021). Interviews with key leaders in the Army medical field directed the capstone towards common pain points and valuable metrics among stakeholders. These findings reinforce the potential of digital decision-support tools to enhance both speed and accuracy while maintaining adherence to military medical doctrine.

Stakeholder engagement shaped both the problem definition and the prototype requirements. The capstone team engaged Army medical stakeholders, including medics, sick call personnel, and leaders familiar with military clinical workflows, to identify the most persistent pain points. These discussions consistently highlighted long patient wait times, documentation burden on providers, the importance of maintaining doctrinal consistency with ADTMC guidance, and the need for a tool that supports rather than replaces provider judgment. These findings directly informed the system requirements, value measures, and weighting structure used in the value model.

To address the gap between AI clinical tools and the demands of the military, prior research applied the Systems Decision Process (SDP) (Parnell & Driscoll, 2011) to design a RAG-enabled sick call system (MacMullan et al., 2026). Initial ProModel simulations suggested some reductions in patient time, but empirical validation was required. This study evaluates

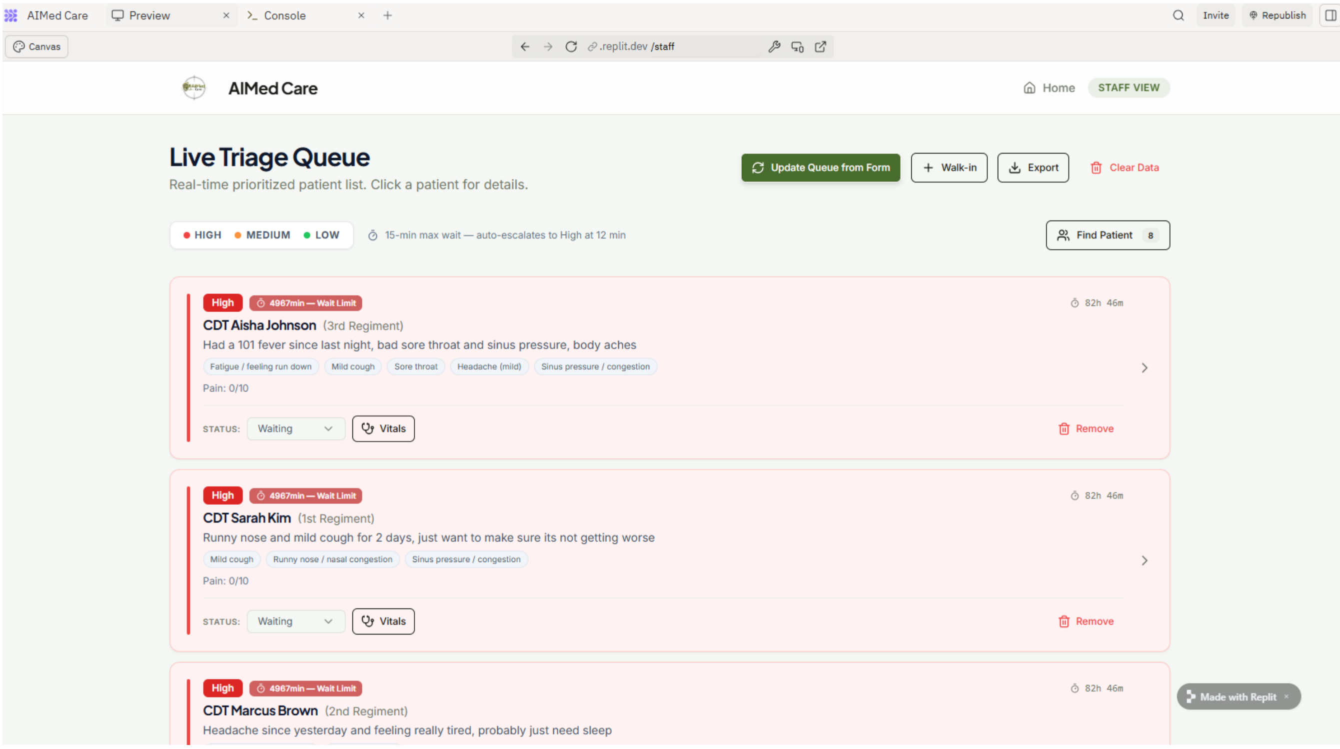


Figure 1: Screenshot of the User Interface for the RAG-Enabled AI Clinical Support System Prototype in Replit

the system seen in Figure 1 through a lab-based experiment, comparing the current paper-based workflow to an AI-enabled process to quantify improvements in efficiency and accuracy.

2. Methodology

The prototype was developed as a software tool to support sick call intake, prioritization, and provider review. First, the existing paper-based intake process was translated into a digital intake form that enabled patient-reported symptoms to be mapped to relevant doctrinal content in ADTMC PAM 40-7.

The core engineering of the prototype lies in the triage engine that processes this digital intake form. This triage engine prototype was developed and hosted using Replit, a cloud based integrated development environment. Once a patient submits their form, the Replit-hosted system utilizes a RAG architecture to automatically evaluate the reported symptoms.

Replit is equipped with AI-assisted code generation capabilities driven by user prompting. The system's architecture was iteratively built using comprehensive prompts derived from prior research and stakeholder analysis. Initial development was driven by foundational documents which relied on four core capabilities: a digital intake interface, persistent backend data storage, an algorithmic triage engine, and a live provider dashboard. Subsequent prompts integrated the overall objective, a system overview, the logistics of the backend logic, functional requirements, and a use-case scenario to generate the initial software framework. Following the initial generation, the prototype was continuously refined to meet the demands of the sick call process. This involved integrating the ADTMC manual into the system's backend, which constrains the AI triage engine to evaluate symptoms and assign priority scores based on published army medical protocols. The prototype then used a RAG-based architecture to retrieve doctrine-relevant information and generate structured output for provider support, including symptom-based prioritization and dashboard updates as illustrated in Figure 1.

To evaluate the operational utility of the RAG-enabled prototype, a comparative lab-based experiment was conducted using physical participants role-playing standardized patients. While the study draws on principles from discrete-event simulation – an established method for analyzing healthcare workflows and patient flow without disrupting clinical operations (Vázquez-Serrano, Peimbert-García, & Cárdenas-Barrón, 2021) – the experiment was implemented as a controlled, in-person setup rather than a purely computational model. Discrete time distributions were incorporated to approximate real-world sick call processes, particularly for intake, vitals collection, and provider interaction times. These distributions were informed by di-

ISBN: 97819384962-5-7

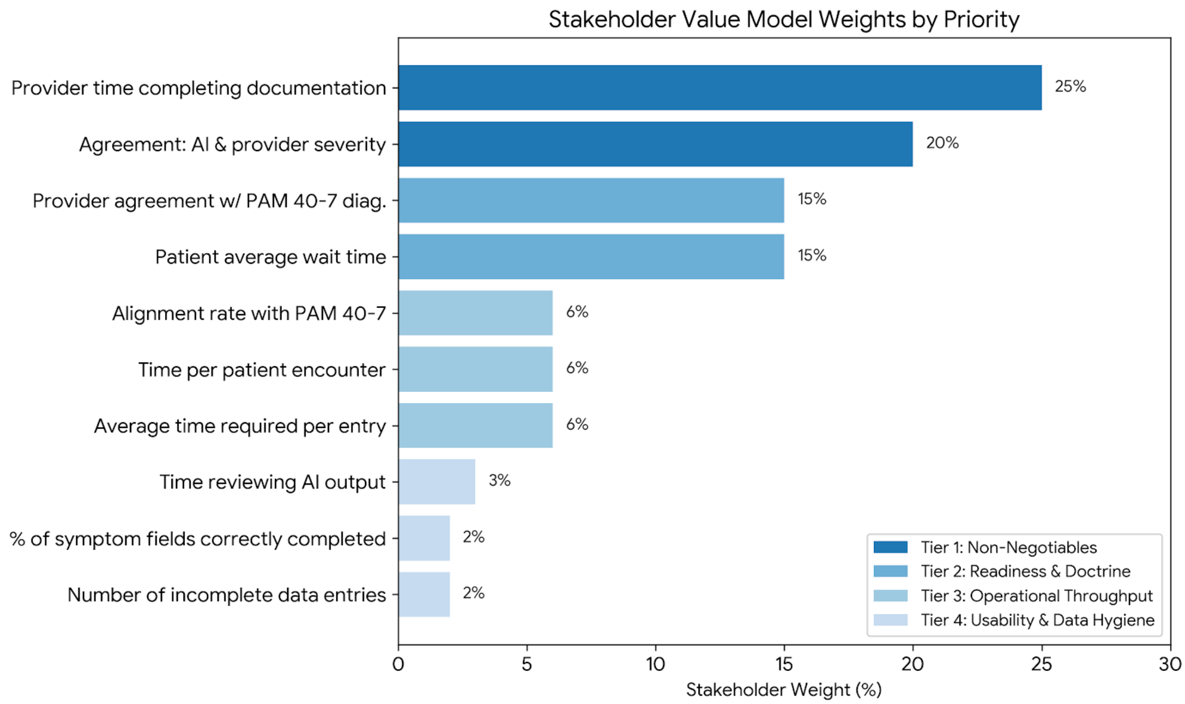


Figure 2: Bar Chart Depicting Stakeholder Swing Weights

rect observation of sick call operations, allowing the experimental design to reflect realistic workflow variability. This approach enables structured comparison between a traditional First-In-First-Out (FIFO) system and a priority-based system that maintains consistency with prior work evaluating multi-level queuing models in healthcare environments (Kumari & Mala, 2024).

To ensure a controlled comparison between these two tests, the experiment setup was standardized across all participants. The system utilized a cohort of 33 cadet participants physically role-playing as patients. To isolate the system workflow as the primary variable, each cadet was issued a standardized “patient card.” These cards contained predetermined data, including a unique patient identifier (1-33), a randomized name, a specific symptom profile, and corresponding vital signs. Both tests utilized the exact same 33 patient profiles to ensure consistency.

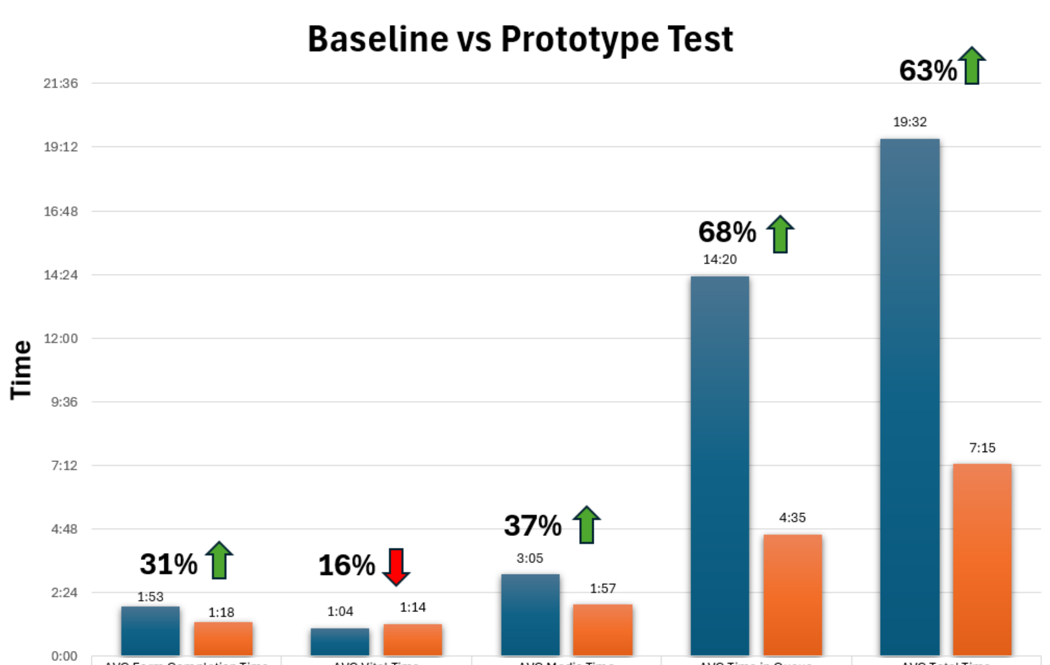
Prior to the experiment, the team conducted direct observation of Sick Call operations to understand patient flow, documentation steps, and approximate service timing. Those observations informed the randomized time ranges used in the lab-based experiment. The baseline test replicates the traditional, First-In-First-Out (FIFO) queuing model currently utilized at sick call. Cadets form a queue outside the sick call lab-based experiment and complete standard paper intake forms while a timer records completion time. Upon completing their form, each cadet manually records their completion time on their form. Cadets enter medic stations based on availability, with vitals and assessment times simulated using randomized timers. Patients exit the system after evaluation, and total time is recorded.

Following the baseline test, the prototype was tested under the same conditions to isolate the impact of the AI-enabled workflow. The prototype test evaluates a RAG-enabled digital system that transitions sick call from a first-come, first-served model to a queue based on patient priority, grounded in ADTMC PAM 40-7 doctrine (Department of the Army Headquarters, United States Army Medical Command, 2019). Cadets complete digital intake forms via QR code, with timestamps automatically recorded. A mobile medic collects vitals directly into the system using randomized completion times between 40 and 75 seconds. As data is collected, the AI updates the medic dashboard, prioritizing patients based on clinical urgency and doctrinal “Red Flags.” Cadets are then called to see one of the open medic stations according to this prioritized queue, with assessment times at medic stations ranging from 1 minute 30 seconds to 2 minutes 30 seconds. Finally, patients are digitally cleared at checkout, generating a precise total system time.

After conducting both tests, performance metrics were collected and prepped for data analysis. The primary metrics collected during the experiment include the intake form completion time, the wait time, the service time, and the total time in the system.

ISBN: 97819384962-5-7

A key contribution of this work is the development of a stakeholder-informed value model used to evaluate system performance. Unlike simple time comparisons, the value model incorporates multiple performance measures weighted according to stakeholder priorities. These value model weights are shown in Figure 2. For example, provider documentation time was assigned the highest weight based on stakeholder feedback emphasizing administrative burden. As a result, improvements in documentation efficiency contribute more significantly to total system value than lower-weighted measures. This weighting structure ensures that the evaluation reflects the most valuable factors to end users rather than treating all metrics equally.



	AVG Form Completion Time	AVG Vital Time	AVG Medic Time	AVG Time in Queue	AVG Total Time
Baseline	1:53	1:04	3:05	14:20	19:32
Prototype	1:18	1:14	1:57	4:35	7:15

Figure 3: Baseline vs Prototype Test Results

3. Results

The lab-based experiment suggests that the prototype-enabled workflow could substantially reduce total sick call processing time relative to the baseline paper-based process. As depicted in Figure 3, average total time decreased from 19 minutes 32 seconds to 7 minutes 15 seconds, while average queue time decreased from 14 minutes 20 seconds to 4 minutes 35 seconds. Form completion time also decreased from 1 minute 53 seconds to 1 minute 18 seconds, and average medic interaction time decreased from 3 minutes 5 seconds to 1 minute 57 seconds. However, one metric that did not improve was vitals collection time, which increased slightly from 1 minute 4 seconds to 1 minute 14 seconds (a 16% decrease in efficiency). This is likely due to the procedural change from a stationary vitals collection point in the baseline test to a mobile medic collecting vitals throughout the queue in the prototype test. Despite this increase, overall system performance was not negatively impacted due to the substantial reductions in queue time. Overall, these findings suggest that the prototype may reduce administrative friction and improve patient flow in a controlled setting.

Informal, anecdotal observations from participants suggested that the digital intake process was easier to complete and provided more structured information prior to the provider interaction.

4. Discussion

This capstone developed a prototype clinical support tool designed to improve sick call workflow through digital intake, prioritization based on Army doctrine, and AI decision support. The system is intended to assist providers by organizing patient information, identifying potential red flags based on ADTMC doctrine, and reducing administrative burden during documentation. Most importantly, the tool is designed to support, not replace, provider decision-making.

If implemented in an operational setting, the prototype has the potential to improve several aspects of sick call performance. The observed reductions in total time and queue time suggest that the system could significantly decrease patient wait times and improve system flow. Additionally, the reduction in form completion and medic interaction time indicates that providers may spend less time on administrative tasks and more time on direct patient care. The structured digital intake process also ensures more complete and standardized data collection prior to the physical encounter.

Several limitations should be considered when interpreting these results. First, the experiment was conducted as a lab-based study using 33 cadet participants role-playing standardized patients. While this controlled setup isolates workflow variables, it does not capture the full complexity and variability of a real clinic, and it does not account for factors such as facility layout or equipment constraints. In addition, the prototype was not evaluated by credentialed medical providers, limiting the ability to assess the accuracy and applicability of the system's clinical support outputs. Finally, the prototype operates as a standalone system and is not integrated with operational systems such as MHS GENESIS, which may introduce additional constraints in real-world implementation.

Future work should focus on validating the prototype in operational or clinically supervised environments with medical providers. In particular, the accuracy and usefulness of the system's prioritization and clinical support outputs must be assessed against provider judgment. Additionally, further testing should populate the value model with validated data and enable statistical analysis to estimate a plausible range of expected time savings. In future work, the value model will be utilized to score this prototype against other alternatives. To evaluate the effectiveness of the RAG-enabled prototype, the raw performance metrics need to be converted into utility scores from 0-100 using multi-criteria decision analysis (MCDA) consistent with systems engineering decision frameworks (Parnell & Driscoll, 2011). The application of MCDA for use in healthcare is currently being piloted and tested by "various Health Technology Assessment (HTA) agencies" (Marsh, Sculpher, Caro, & Tervonen, 2018). Furthermore, recent systematic reviews confirm that MCDA is an increasingly vital and transparent tool for decision-makers needing to weigh criteria, such as documentation time versus triage accuracy, when evaluating whether to implement new healthcare interventions (Baltussen, Marsh, Thokala, et al., 2019). Raw time metrics and triage accuracy percentages will be translated into normalized utility scores (0-100) using the value curves established in prior stakeholder research. This planned future analysis will quantify whether the AI-enhanced system justifies broader implementation across Army Medicine. Expanding the dataset will also enable statistical analysis, including confidence intervals, to estimate a plausible range of expected time savings. Lastly, integration with systems such as MHS GENESIS will be necessary to assess feasibility and scalability in real-world deployment.

5. Conclusion

This capstone focused on the design and development of a RAG-enabled clinical support tool intended to improve Army sick call operations through digital intake, prioritization of patients based on doctrine, and provider decision support. Using a stakeholder-informed Systems Decision Process, the team identified key workflow inefficiencies and translated these insights into a functional software prototype aligned with ADTMC guidance.

A controlled lab-based experiment using 33 cadets role-playing standardized patients was conducted to estimate the prototype's potential impact on workflow performance. The results suggest that the tool could reduce total system time and administrative burden, primarily through reductions in queue time and improvements in data collection efficiency. Additionally, the development of a stakeholder-weighted value model provides a structured framework for evaluating system performance based on what matters most to end users, rather than relying solely on individual metrics.

While these findings demonstrate the potential of the prototype, it is not a representation of clinical validation in an operational sick call environment. Future work is needed to evaluate the system with credentialed medical providers, as a means of assessing the accuracy and usefulness of its prioritization and clinical support outputs.

This research establishes a foundation for future development and validation of AI-enabled clinical support tools within military healthcare. By combining stakeholder-driven design, doctrine-grounded AI methods, and a structured evaluation framework, this work demonstrates how such tools may support providers and improve sick call workflow without replacing clinical judgment.

References

Baltussen, R., Marsh, K., Thokala, P., et al. (2019). Multi-Criteria Decision Analysis for Health Technology Assessment: Addressing Methodological Challenges to Improve the State of the Art. *European Journal of Health Economics*, *20*, 891–918.

Chrosniak, J., Olsen, C., & Galdi, A. (2021). Algorithm Directed Troop Medical Care Manual Application for Desktop and Smartphone. *Military Medicine*, *186*(Supplement$_1$)$, 65 - -69. doi : 10.1093/milmed/usaa471$

Department of the Army Headquarters, United States Army Medical Command. (2019). *MEDCOM Pamphlet 40-7-21: Physician Assistant (PA) Algorithm-Directed Troop Medical Care.* `https://nextlevelmedic.com/algorithm-directed-troop-medical-care/`. (Accessed: 2026-03-19)

Kumari, T., & Mala, D. (2024). Evaluating Multi-Level Priority Queues Through Simulation Techniques: A Non-Preemptive Priority Model Application in Healthcare Systems. *Journal of Computational Analysis and Applications*, *33*(1A), 796–815.

MacMullan, K., Zhang, A., Bradford, C., Kone, E., Segat, P., & Bahabry, A. (2026). A Systems Engineering Model for Integrating RAG-Based AI into Army Sick Call Workflows. In *2026 IEEE International Systems Conference (SysCon).*

Marsh, K. D., Sculpher, M., Caro, J. J., & Tervonen, T. (2018). The Use of MCDA in HTA: Great Potential, but More Effort Needed. *Value in Health*, *21*(4), 394–397. doi: 10.1016/j.jval.2017.10.001

Miao, Y., Zhao, Y., Luo, Y., Wang, H., & Wu, Y. (2025). Improving Large Language Model Applications in the Medical and Nursing Domains With Retrieval-Augmented Generation: Scoping Review. *Journal of Medical Internet Research*, *27*, e80557. doi: 10.2196/80557

Parnell, G. S., & Driscoll, P. J. (2011). *Decision Making in Systems Engineering and Management* (2nd ed.). John Wiley & Sons. doi: 10.1002/9780470926963

Roberts, C., & Wolfe, R. A. (2025). *Enabling Medics to Manage Sick Call* (Tech. Rep.). Next Generation Combat Medic. (Primary author: Craig Roberts, PA-C; Additional contributions: Robert Wolfe, PA-C)

U.S. Army Medical Center of Excellence. (2023). *Sick Call and Medical Documentation*. Fort Sam Houston, TX: Department of the Army. (68W Combat Medic Specialist Training Manual)

Vázquez-Serrano, J. I., Peimbert-García, R. E., & Cárdenas-Barrón, L. E. (2021). Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review. *International Journal of Environmental Research and Public Health*, *18*(22), 12262. doi: 10.3390/ijerph182212262

Wong, H. S., & Wong, T. K. (2026). Multi-Evidence Clinical Reasoning With Retrieval-Augmented Generation for Emergency Triage: Retrospective Evaluation Study. *JMIR Medical Informatics*.