ATLAS MAS Evidence Aware Trauma Triage and Capacity Constrained Hospital Routing

J Siddartha Varma[1], Dr. S. Jhansi Rani[2]
[1,2]*Department of Computer Science and Systems Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India*

Abstract: **Hospital availability and specialist services can change the destination selected for a trauma incident, even when a nearer hospital has a shorter estimated journey. We evaluated this effect in ATLAS-MAS, an n8n workflow that combines rule-based urgency assessment with hospital eligibility checks and road routing. Four stages handle input validation, assessment, coordination simulation and decision-support outputs. Road duration and distance are obtained from Open-Source Routing Machine. Nine scripted tests assessed routine and failure branches; a separate set of 12 controlled scenarios compared destination-selection policies using the same hospital candidates. All 22 scripted assertions passed. Availability only selection produced nine eligible recommendations, whereas adding capability checks produced 12. The three changed destinations increased mean estimated travel by 0.88 minutes across the scenario set. Mean pipeline time was 614.9 ms, and five cases exceeded the separate 250 ms core-processing target. The main finding is that service requirements affected hospital choice in cases that availability filtering alone accepted. The results verify this behaviour in the implemented workflow; repeated execution and clinical assessment remain necessary before operational use.**

Index Terms: Capacity constraints, hospital selection, multi-agent workflow, road routing, trauma triage, workflow evaluation.

I. INTRODUCTION

A hospital can be geographically close yet unavailable or unable to provide the services required for a trauma case. An urgency category alone does not resolve this destination decision. Studies of severe-trauma prediction [1], surgical need [2] and prehospital prediction [3], [4] address the assessment component. Work on hospital strategies and ambulance dispatch [5] considers hospital operations. ATLAS-MAS examines how an incident assessment can be carried through to a hospital recommendation using explicit availability and service checks.

We implemented the workflow for six hospitals around Visakhapatnam and Anakapalli. Each execution retains the urgency category, candidate hospitals, exclusion reasons, selected destination and processing intervals. These records allow us to examine whether capability checks change the choice made by an availability-only rule, and what additional travel those changes imply. Separate scripted tests exercise invalid input, delayed processing and coordination fallback. The study therefore evaluates software behaviour and destination selection, with each reported outcome linked to a recorded execution.

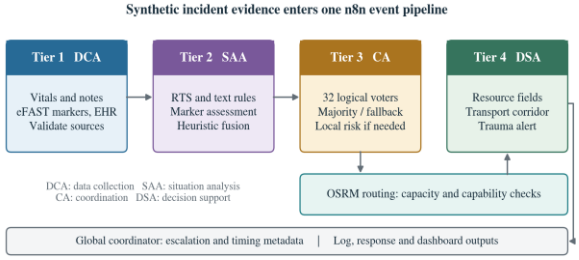

Fig. 1. Four-tier architecture. Each tier performs a separate task, while the event record carries its inputs, decisions and timing information

II. RELATED WORK AND RESEARCH RATIONALE

Chen et al. predicted severe trauma from prehospital information [1]. Millarch et al. considered a different endpoint, the need for major and specialty-specific surgery [2]. Although both concern early assessment, their performance measures depend on the clinical outcome being predicted. Neither endpoint is measured in the ATLAS-MAS records. These studies link prehospital information to subsequent care. Our

evaluation examines the next decision: hospital selection after an urgency category has been assigned. The ShockMatrix studies provide an implementation perspective. The pilot addressed collection of predictors during care [3], and the later prospective study compared human and machine predictions [4]. Their distinction between implementation and prediction is relevant to ATLAS-MAS. Our records support evaluation of input acceptance, execution paths and recorded outputs; they do not establish clinical prediction accuracy.

Language models have been studied for emergency triage and related decisions [6]– [8]. Defilippo et al. instead represented patient similarity as a graph and used a graph neural network for triage classification [9].

These studies concern trained assessment methods. ATLAS-MAS uses physiological scoring and keyword rules. Although its exported nodes retain the names LLM and DistilBERT, neither model runs in the current version. Consequently, the fallback experiment tests a change in execution path, not the performance of a language model.

Table I. Related approaches and their relevance to the proposed workflow.

Study	Published approach	Relevance to ATLAS-MAS
Chen [1]; Millarch [2]	Clinical models for trauma severity or surgical need	Early assessment with outcome-based validation
ShockMatrix [3], [4]	Pilot integration and prospective prediction assessment	Assessment of workflow integration and decision performance
Masanneck [6]; Shekhar [7]; Gaber [8]	Trained language models assessed on clinical decision tasks	Separate validation of text-based assessment
Defilippo et al. [9]	Patient-similarity graph and neural triage classification	Learned triage classification; ATLAS-MAS uses explicit rules
Fu et al. [5]	Hospital-strategy and ambulance-dispatch simulation	Hospital conditions as selection criteria
Murarețu et al. [10]; Sivagnanam et al. [11]	Learning-based resource allocation and responder stationing	Broader allocation scope than incident-level routing

A. Research Scope and Design Rationale

Hospital destination selection also differs from allocating resources across an emergency service. Fu et al. examined hospital strategies and ambulance dispatch through simulation [5]. Murarețu et al. studied resource allocation during mass-casualty incidents [10], whereas Sivagnanam et al. considered emergency-responder stationing [11]. Those decisions involve a wider operational scope than the selection of a receiving hospital for one incident. Table I summarises the relationship between these approaches and the implemented workflow. For this study, hospital selection is evaluated separately from clinical prediction. Every candidate in the six-hospital registry can be inspected, including the route estimate and reason for exclusion. This permits an availability-only decision to be compared with a decision that also requires specialist capability. The narrower scope makes the comparison reproducible from the stored candidates, but it provides no basis for a predictive-performance comparison with the studies in Table I.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The four agent groups in Fig. 1 are implemented as sequential n8n nodes. Data Collection Agents (DCA) check the submitted evidence before Situation Analysis Agents (SAA) calculate urgency. Coordination Agents (CA) apply voting and fallback rules to that assessment. The destination and resource fields are prepared by Decision Support Agents (DSA), while the Global Coordinator (GC) records escalation and elapsed time. Fig. 2 traces the normal path and the branches reached when validation or coordination conditions are not met.

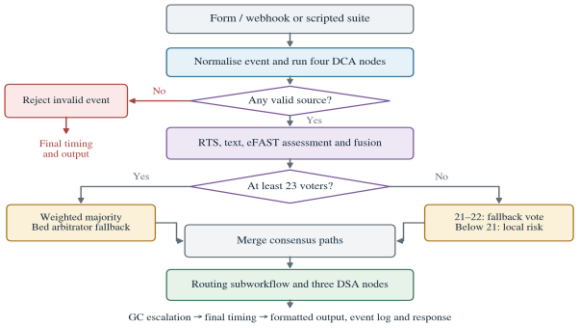

Fig. 2. n8n execution sequence with validation and coordination branches. Invalid input bypasses assessment and GC processing; consecutive nodes are grouped for readability.

A. Input Validation and Urgency Assessment

Four input types are supported: vital signs, paramedic notes, eFAST markers and electronic health record (EHR) fields. The eFAST markers represent findings from extended focused assessment with sonography for trauma, an ultrasound examination used in trauma care. They enter the workflow as structured values; image interpretation is outside the implementation. EHR fields supply history information in the event record. The fusion calculation uses the assessment sources described below and does not include EHR values.

Physiological assessment uses the Revised Trauma Score (RTS) [12], calculated as:

$$RTS = 0.9368G + 0.7326S + 0.2908R \quad (1)$$

Here, G, S and R are the coded Glasgow Coma Scale, systolic blood pressure and respiratory rate values, respectively. The Glasgow Coma Scale measures level of consciousness. The implemented categories are High for RTS < 5, Medium for 5 ≤ RTS < 7, and Low for RTS ≥ 7. Validation requires at least two valid numeric vital signs. Heart rate and SpO_2, the estimated oxygen saturation, are checked but are not terms in RTS. Missing RTS inputs can receive default values.

B. Evidence Fusion and Coordination

The assessment nodes return Low, Medium, High or Critical. For fusion, these categories are assigned scores of 0.15, 0.50, 0.85 and 0.97. The implemented equation is:

$$F = 1 - \prod(1 - s_i) \quad (2)$$

The rule uses up to three assessment scores, s_i, from the available assessment sources. With no assessment score, the implementation assigns $F = 0.20$. Category boundaries are $F < 0.40$ for Low, $0.40 \le F < 0.80$ for Medium, $0.80 \le F \le 0.90$ for High, and $F > 0.90$ for Critical. F is a heuristic score. Text assessment counts keywords and includes a fallback condition at 120 ms of simulated latency.

Coordination depends on the number of reachable voters. Of the 32 represented voters, 23 are required for the normal weighted vote; 21 or 22 activate the fallback. Fewer than 21 leave the local assessment in place. A quorum is the minimum number of voters required for a decision. All voters copy the same fused category, so the test exercises branch selection rather than agreement among independent assessors. Practical Byzantine Fault Tolerance (PBFT) [13] informed the fallback counts, but its distributed protocol is not implemented.

C. Hospital Eligibility and Route Selection

The hospital registry contains six destinations. Open-Source Routing Machine (OSRM) supplies road distance and duration from the incident coordinates to each one. The ETA field, named for estimated time of arrival, stores this duration. Hospital service and capacity checks then determine which routes remain eligible. Diversion status makes a destination unavailable even when OSRM returns a valid route.

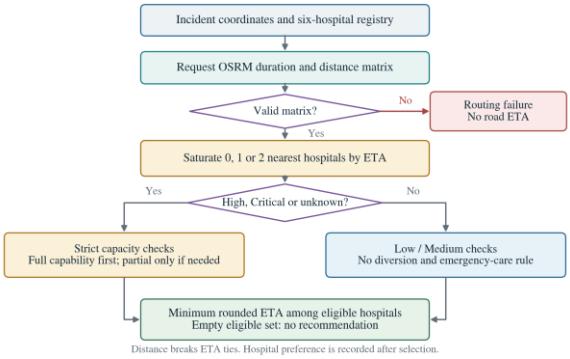

Fig. 3. Hospital eligibility filtering and route ranking. The selected destination has the minimum rounded ETA among eligible candidates; a routing failure yields no road ETA.

The three capacity settings modify the nearest destinations, ordered by road ETA. Normal capacity leaves them unchanged; nearest saturated disables one; regional surge disables two. Disabling sets available intensive care unit (ICU) beds and trauma bays to zero and enables diversion. A trauma bay is the treatment space used for initial emergency care. These settings control the availability conditions used in the evaluation.

For High, Critical and unknown severity, eligibility requires a valid route, diversion disabled, and positive counts for both ICU beds and trauma bays. The service check first seeks a complete match to the requirements extracted from the notes. A partial match is labelled and permitted only if no complete match exists. Low and Medium categories follow the broader emergency-care rules. Eligible destinations are ranked by ETA, then by distance to resolve ties. Both quantities are rounded to one decimal place before this comparison.

Hospital preference is retained in the output but has no effect on ranking. If the routing request fails, an allocation made earlier in the parent workflow may remain available without an ETA. This allocation exists only on the normal majority path; the coordination fallback does not guarantee it. Table II identifies which outputs are recommendations or descriptions and which external confirmations would be needed for operational use.

Table II. Decision-support outputs and requirements for operational integration.

Output	Interpretation	Integration requirements
Hospital recommendation	A destination meeting the applicable configured rules	Verified hospital services and current availability
Resource allocation	A request for 1, 2 or 5 ICU beds using a fixed nominal pool of 15	An actual reservation and hospital acknowledgement
Road ETA	An OSRM estimate of the journey	Observed ambulance travel and traffic validation
GC and alert fields	Escalation, timing and notification descriptions	External systems that carry out those actions

IV. EXPERIMENTAL DESIGN AND FUNCTIONAL VERIFICATION

We analysed the 23 event records logged on 1st October 2026. The scripted suite accounts for nine records, and the scenario comparison uses 12 individual form submissions.

Two earlier submissions are retained as supplementary records (Table III). The scenario set pairs four incident profiles with three capacity settings. Patient evidence and hospital capacity were specified synthetically exercise those conditions; neither patient records nor operational hospital feeds were used.

Table III. Evaluation dataset and analytical scope.

Group	Events	Purpose
Scripted suite, run_71	9	Check expected decisions, rejection and fallback behaviour
Scenario matrix, run_83–run_131	12	Compare four incident profiles across three capacity settings
Earlier submissions, run_32 and run_66	2	Supplementary records, excluded from the balanced comparison
Total	23	22 processed events and one correctly rejected malformed event

A. Scripted Test Coverage

Table IV reports the nine tests and the properties checked in each. The failure tests introduce different conditions: a 180 ms text-path delay in TC4, 21 reachable voters in TC5, invalid vital signs in TC7, and a 600 ms injected delay in TC9. All 22 assertions passed. An assertion is an expected property of execution, so its count differs from the event count. Across the full log, 22 events were processed and TC7 was rejected.

Table IV. Scripted test conditions and verification outcomes.

Case	Test input or condition	Observed outcome	Checks
TC1	Stable vital signs	Low risk; normal vote; no escalation	4/4
TC2	Low consciousness and blood pressure	High risk; normal vote; no escalation	4/4
TC3	Notes describing severe injuries	High risk; primary text path; normal vote	4/4
TC4	Forced text-path latency of 180 ms	Medium risk; simulated fallback path used	2/2
TC5	Only 21 voters reachable	PBFT-labelled fallback used; Medium risk	2/2
TC6	Chemical blast with multiple casualties	Critical risk; GC escalation recorded	2/2

Case	Test input or condition	Observed outcome	Checks
TC7	Out-of-range vital signs	Rejected before assessment	1/1
TC8	Vitals, injury notes and eFAST markers	Processed; fused risk Critical	1/1
TC9	Injected delay of 600 ms	Processed; core-time target breach detected	2/2
Total	Nine test cases	All expected checks passed	22/22

B. Performance Measures and Comparative Protocol

Timing summaries use the stored payloads from one execution of each scenario. They describe the observed sample, without repeated measurements for any scenario. Changes to default capacity settings and labels in later exports were not followed by reruns of these cases. The saved event payloads, rather than those later defaults, define the inputs associated with the reported results.

The measured pipeline ends at the final timing node. Output formatting, log storage and the HTTP response occur afterwards and are excluded. Core time is obtained by subtracting the routing interval from pipeline time; any injected delay remains in the core measurement. We apply the 250 ms target to that interval. Destination-policy comparisons reuse the stored candidate sets and do not generate new timing observations.

V. RESULTS

A. Scenario Outcomes and Destination Selection

The case identifiers in Table V combine an incident profile with a capacity setting. RA denotes road-accident trauma at Kurmannapalem; UV, unstable vitals at Port Road; PN, severe paramedic notes at Madhurawada; and SP, a stable patient at Anakapalli. Suffixes N, S and R identify normal capacity, nearest saturated and regional surge, respectively. All 12 submissions completed routing and returned a destination eligible under the configured rules.

Table V. Individual scenario outcomes and processing times. ETA denotes estimated travel duration.

Case	Run	Risk	Selected hospital	ETA (min)	Pipeline (ms)	Core (ms)
RA–N	83	Critical	KIMS ICON	8.5	617	272
RA–S	87	Critical	CARE	21.0	638	242
RA–R	92	Critical	CARE	21.0	447	243
UV–N	97	High	CARE	9.9	558	235
UV–S	102	High	KGH	11.5	561	258
UV–R	107	High	Apollo Health City	18.0	577	261
PN–N	111	High	CARE	18.4	796	264
PN–S	116	High	CARE	18.4	451	233
PN–R	120	High	CARE	18.4	894	264
SP–N	124	Low	District, Anakapalli	1.8	474	246
SP–S	128	Low	KIMS ICON	25.8	483	241
SP–R	131	Low	CARE	38.2	883	243

CARE was selected for all three severe-notes scenarios, with an ETA of 18.4 minutes. Nearer alternatives lacked the required services, so changing capacity did not change this profile's destination. The other three profiles had different destinations across the settings shown in Fig. 4. The widest ETA range occurred for SP, from 1.8 to 38.2 minutes. Nine of the 12 recommendations differed from the nearest hospital by road ETA.

The road-accident inputs also changed: RA–S and RA–R contain a positive eFAST marker that is absent from RA–N. Although all three received a Critical category, their destination differences cannot be attributed to capacity alone.

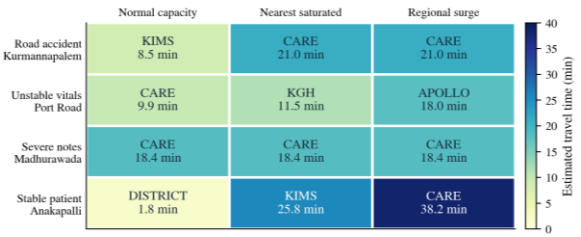

Fig. 4. Selected destinations and estimated travel duration by incident profile and capacity condition. DISTRICT denotes District Hospital, Anakapalli; KGH denotes King George Hospital.

The comparison in Section V-C therefore applies each selection policy to the same candidates within a recorded case.

B. Processing Latency

Table VI reports pipeline, routing and core-processing intervals separately. Across the 12 cases, the pipeline ranged from 447 to 894 ms. Core processing varied by 39 ms, from 233 to 272 ms; routing varied by 436 ms, from 204 to 640 ms. Fig. 5 shows these intervals for each case, and Fig. 6 identifies the executions above the 250 ms core target.

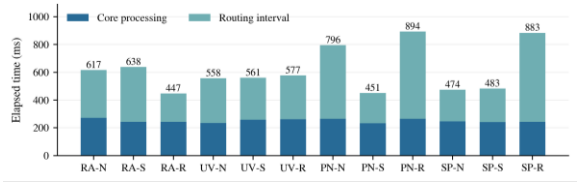

Fig. 5. Pipeline time for each case in Table V, divided into core processing and the routing interval.

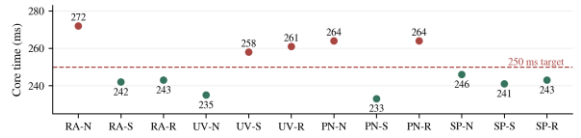

Fig. 6. Core-processing time against the 250 ms target. Seven cases met the target and five exceeded it.

Table VI. Descriptive processing-time statistics for the 12 evaluation scenarios.

Measure	Mean	Median	Range
Pipeline time (ms)	614.9	569.0	447–894
Routing interval (ms)	364.8	319.5	204–640
Core processing (ms)	250.2	244.5	233–272
Core target exceeded	5/12 (41.7%)	Not applicable	Target: 250 ms

The largest pipeline measurement occurred in PN–R: 894 ms, comprising 630 ms of routing and 264 ms of core processing. RA–R had the smallest pipeline measurement, 447 ms, with corresponding intervals of 204 and 243 ms. Routing accounted for 59.3% of the summed time across all cases. The logs do not separate network communication time from OSRM computation time.

The core target was exceeded by RA–N, UV–S, UV–R, PN–N and PN–R. Their core intervals were 272, 258, 261, 264 and 264 ms, respectively. The other seven cases met the target. A single execution per scenario cannot establish how often these breaches would recur. Repetition and concurrent-load testing are therefore required; batch-test timings are excluded because the execution conditions differ.

C. Comparative Evaluation of Selection Policies

We reapplied three selection policies to each stored candidate set. The first selects the shortest ETA without checking eligibility. The second requires operational availability, but omits the specialist-service restriction. The third is the ATLAS-MAS policy, which includes both requirements. Fig. 7 counts selections that violate the configured rules. Fig. 8 and Table VII give the estimated travel associated with each policy.

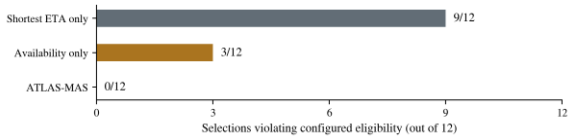

Fig. 7. Selections that failed the configured eligibility rules when the same 12 candidate sets were reconsidered.

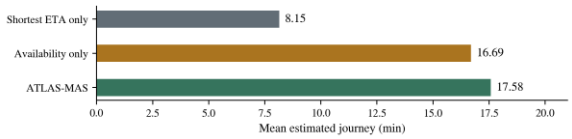

Fig. 8. Mean estimated journey for each selection rule. The shortest journey is not always to an eligible hospital.

The availability-only policy failed the service check in RA–S, PN–N and PN–S. For RA–S, the capability checks excluded District Hospital at 18.4 minutes and selected CARE at 21.0 minutes.

Table VII. Comparison using identical recorded hospital candidates.

Selection rule	Eligible choices	Availability failures	Capability-only failures	Mean ETA
Shortest ETA only	3/12 (25%)	8	1	8.15 min
Availability only	9/12 (75%)	0	3	16.69 min
ATLAS-MAS	12/12 (100%)	0	0	17.58 min

For PN–N, it excluded GITAM at 12.4 minutes; for PN–S, it excluded Apollo at 16.4 minutes. CARE was selected at 18.4 minutes in both PN cases. These changes account for the three additional eligible recommendations. The remaining nine decisions were unchanged. Mean differences were calculated before rounding. Across all 12 cases, the mean additional journey relative to availability-only selection was 0.88 minutes, although the PN–N increase was 6.0 minutes. Against ETA-only selection, the mean difference was 9.43 minutes. That comparison includes the effect of excluding nearby hospitals under the capacity settings, as well as the specialist-service requirement.

The observed advantage is confined to compliance with the configured requirements. Table VII does not measure patient benefit or establish the clinical validity of those requirements. Assessing whether the additional journey is justified would require patient outcomes and independent review of the eligibility rules.

VI. DISCUSSION

A. Behaviour of the Evidence Fusion Rule

The urgency category controls the routing requirements, making the fusion equation part of the destination decision.

To examine its behaviour, Fig. 9 combines one, two and three inputs of equal severity. These values are calculated from the equation; they are not additional execution results.

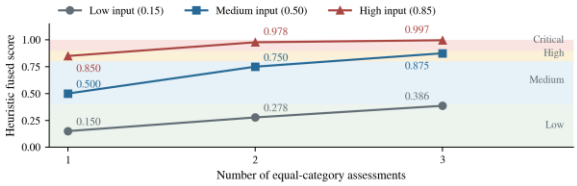

Fig. 9. Effect of combining equal-category assessments. The curves are calculated from the implemented rule; the shaded bands show its urgency categories.

The category changes depend on the accumulation of scores. Three Medium assessments give 0.875, above the High boundary of 0.80. Two High assessments give 0.9775, above the Critical boundary of 0.90. Three Low assessments do not reach Medium. The final category can therefore exceed each individual input category. The equation does not determine whether those inputs provide independent evidence. Repeated descriptions of one injury could therefore increase the score, which should be interpreted as a configured escalation measure rather than a clinical probability.

B. Limitations and Operational Integration

The current records support software testing but contain no patient outcomes for calculating clinical accuracy. Coordination testing is also limited by the use of copied assessments. Missing inputs also require further testing: defaults can enter RTS, and emergency capability can be assumed for lower-acuity routing when hospital metadata is absent. Neither condition has been evaluated comprehensively by the present suite.

A capacity mismatch remains between resource requests and routing eligibility. The configured request for a Critical case is five ICU beds, yet routing can accept a hospital with four available beds. Before operational use, these components must check the same reservable capacity and obtain hospital confirmation. The alert and scale-out outputs also require connections to services that perform the described actions. DECIDE-AI [14] provides guidance for subsequent clinical evaluation; TRIPOD+AI [15] applies to the reporting of any future trained prediction component.

VII. CONCLUSION

ATLAS-MAS was developed to examine destination selection when injury urgency, hospital capacity and service requirements must be considered together. The n8n workflow carries an incident through validation,

rule-based assessment, coordination simulation and decision support. Its hospital recommendation follows eligibility filtering and ranking by estimated road travel, with the excluded candidates retained for inspection.

In the 12 routing scenarios, checking specialist capability changed three destinations that had passed availability-only filtering. All 12 final choices met the configured eligibility rules, compared with nine under the simpler policy, at a mean additional journey of 0.88 minutes. The separate functional suite passed 22 assertions across nine tests. Mean pipeline time was 614.9 ms; five scenarios missed the 250 ms core target. The service check improved compliance with the configured eligibility rules, while the timing results identify a need for further optimisation.

The next steps are to verify hospital availability through operational feeds, reconcile allocation with confirmed reservations, and repeat the timing tests. Clinical review of the assessment and eligibility rules is needed before the recommendations can be used in patient care.

REFERENCES

[1] Q. Chen et al., "Enhancing performance of the National Field Triage Guidelines using machine learning: Development of a prehospital triage model to predict severe trauma," Journal of Medical Internet Research, vol. 26, Art. no. e58740, 2024, doi: 10.2196/58740.

[2] A. S. Millarch, F. Folke, S. S. Rudolph, H. M. Kaafarani, and M. Sillesen, "Prehospital triage of trauma patients: Predicting major surgery using artificial intelligence as decision support," British Journal of Surgery, vol. 112, no. 4, Art. no. znaf058, 2025, doi: 10.1093/bjs/znaf058.

[3] T. Gauss et al., "Pilot deployment of a machine-learning enhanced prediction of need for hemorrhage resuscitation after trauma – the ShockMatrix pilot study," BMC Medical Informatics and Decision Making, vol. 24, Art. no. 315, 2024, doi: 10.1186/s12911-024-02723-9.

[4] T. Gauss et al., "Comparison of machine learning and human prediction to identify trauma patients in need of hemorrhage control resuscitation (ShockMatrix study): A prospective observational study," The Lancet Regional Health – Europe, vol. 55, Art. no. 101340, 2025, doi: 10.1016/j.lanepe.2025.101340.

[5] X. Fu, V. Krzhizhanovskaya, A. Yakovlev, and S. Kovalchuk, "Modelling diversity in hospital strategies in city-scale ambulance dispatching with coupled game-theoretic model and discrete-event simulation," Journal of Biomedical Informatics, vol. 162, Art. no. 104777, 2025, doi: 10.1016/j.jbi.2025.104777.

[6] L. Masanneck et al., "Triage performance across large language models, ChatGPT, and untrained doctors in emergency medicine: Comparative study," Journal of Medical Internet Research, vol. 26, Art. no. e53297, 2024, doi: 10.2196/53297.

[7] A. C. Shekhar, J. Kimbrell, A. Saharan, J. Stebel, E. Ashley, and E. E. Abbott, "Use of a large language model (LLM) for ambulance dispatch and triage," American Journal of Emergency Medicine, vol. 89, pp. 27–29, 2025, doi: 10.1016/j.ajem.2024.12.032.

[8] F. Gaber et al., "Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis," npj Digital Medicine, vol. 8, Art. no. 263, 2025, doi: 10.1038/s41746-025-01684-1.

[9] A. Defilippo, P. Veltri, P. Lió, and P. H. Guzzi, "Leveraging graph neural networks for supporting automatic triage of patients," Scientific Reports, vol. 14, Art. no. 12548, 2024, doi: 10.1038/s41598-024-63376-2.

[10] I. Murarețu, A. Vultureanu-Albiși, S. Ilie, and C. Bădică, "ABMS-driven reinforcement learning for dynamic resource allocation in mass casualty incidents," Future Internet, vol. 17, no. 11, Art. no. 502, 2025, doi: 10.3390/fi17110502.

[11] A. Sivagnanam, A. Pettet, H. Lee, A. Mukhopadhyay, A. Dubey, and A. Laszka, "Multi-agent reinforcement learning with hierarchical coordination for emergency responder stationing," in Proc. 41st International Conference on Machine Learning (ICML), PMLR, vol. 235, pp. 45813–45834, 2024. OnlineOnline. Available: https://proceedings.mlr.press/v235/sivagnanam24a.html

[12] H. R. Champion, W. J. Sacco, W. S. Copes, D. S. Gann, T. A. Gennarelli, and M. E. Flanagan, "A revision of the Trauma Score," Journal of

Trauma, vol. 29, no. 5, pp. 623–629, 1989, doi: 10.1097/00005373-198905000-00017.

[13] M. Castro and B. Liskov, "Practical Byzantine fault tolerance," in Proc. 3rd Symposium on Operating Systems Design and Implementation (OSDI '99), USENIX Association, 1999. OnlineOnline. Available: https://www.usenix.org/conference/osdi-99/practical-byzantine-fault-tolerance

[14] B. Vasey et al., "Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI," Nature Medicine, vol. 28, no. 5, pp. 924–933, 2022, doi: 10.1038/s41591-022-01772-9.

[15] G. S. Collins et al., "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods," BMJ, vol. 385, Art. no. e078378, 2024, doi: 10.1136/bmj-2023-078378.