EDSim: An Agentic Simulator for Emergency Department Operations

Jiajun Wu
University of Calgary
Hutton Ledingham
University of Calgary
Zirui Wang
University of Calgary
Braden Teitge
University of Calgary
Alexander Burn
University of Calgary
Oussama Ouadihi
University of Calgary
Mohamad Ghattas
University of Calgary
Darin Vicaldo
University of Calgary
Sergiu Cociuba
University of Calgary
Megan Harmon
University of Calgary
Tanvir Chowdhury
University of Calgary
Zack Marshall
University of Calgary
Tyler Williamson
University of Calgary
Tracie Risling
University of Calgary
Eddy Lang
University of Calgary
Jiayu Zhou
University of Michigan
Jessalyn Holodinsky

University of Calgary

Steve Drew

steve.drew@ucalgary.ca

University of Calgary

Article

Keywords: Emergency departments, simulation, large language model, agentic AI

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8960989/v1

Additional Declarations: No competing interests reported.

EDSim: An Agentic Simulator for Emergency Department Operations

Jiajun Wu[1], Hutton Ledingham[1], Zirui Wang[1], Braden Teitge[2], Alexander Burn[7], Oussama Ouadihi[1], Mohamad Ghattas[1], Darin Vicaldo[1], Sergiu Cociuba[3,4], Megan Harmon[2], Tanvir Chowdhury[3,8], Zack Marshall[3], Tyler Williamson[3,4], Tracy Risling[6], Eddy Lang[2], Jiayu Zhou[5], Jessalyn Holodinsky[2,4*], Steve Drew[1*]

[1]Department of Electrical and Software Engineering, University of Calgary, Calgary, AB, Canada.
[2]Department of Emergency Medicine, University of Calgary, Calgary, AB, Canada.
[3]Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada.
[4]Centre for Health Informatics, University of Calgary, Calgary, AB, Canada.
[5]School of Information, University of Michigan, Ann Arbor, MI, USA.
[6]Faculty of Nursing, University of Calgary, Calgary, AB, Canada.
[7]Department of Biomedical Engineering, University of Calgary, Calgary, AB, Canada.
[8]Department of Family Medicine, University of Calgary, Calgary, AB, Canada.

*Corresponding author(s). E-mail(s): jessalyn.holodinsky@ucalgary.ca; steve.drew@ucalgary.ca;
Contributing authors: jiajun.wu1@ucalgary.ca; hutton.ledingham@ucalgary.ca; zirui.wang1@ucalgary.ca; braden.teitge@albertahealthservices.ca; alexander.burn@ucalgary.ca; oussama.ouadihi@ucalgary.ca; mohamad.ghattas@ucalgary.ca; darin.vicaldo@ucalgary.ca; sergiu.cociuba@ucalgary.ca; megan.harmon@ucalgary.ca; chowdhut@ucalgary.ca;

zack.marshall@ucalgary.ca; tyler.williamson@ucalgary.ca;
tracie.risling@ucalgary.ca; eslang@ucalgary.ca; jiayuz@umich.edu;

Abstract

Emergency departments (EDs) face chronic crowding and complex patient flow challenges that traditional simulations struggle to capture. Conventional discrete-event or agent-based simulation can match high-level metrics, such as wait time distributions and throughput. However, they cannot reproduce the fine-grained behaviors, communications, and dynamic decision-making of real clinicians and patients. These micro-level interactions are often key to showing ED performance. We propose EDSim, an agentic ED flow simulator that uses large language model (LLM) agents to drive realistic, environment-aware interactions among artificial intelligence agents. EDSim offers a modularized patient journey from triage through treatment and discharge that supports customization, with virtual patients and healthcare providers that converse and make decisions in natural language based on dynamic conditions in the ED. The LLM agents are constrained by clinical rules and global ED states, enabling both macro-level fidelity and micro-level insight. We have parameterized EDSim with aggregate historical data to reproduce daily arrival patterns and acuity mixes. Results show that EDSim's baseline outputs align with historical wait time distributions stratified by triage acuity. Moreover, EDSim generates convincing individual conversations and behavior under potentially new workflows. This illustrates a new paradigm for healthcare operations research, combining data-driven agent-based modeling with LLM-generated behavior to provide a realistic, versatile testbed for improving ED operations. We demonstrate how analysts and hospital managers can use our tool to conduct what-if experiments in minutes, such as reallocating beds or staff to uncover bottlenecks and evaluate interventions. Our simulator can be used by both researchers and practitioners to explore ways to improve emergency care delivery.

Keywords: Emergency departments, simulation, large language model, agentic AI

1 Introduction

The Emergency Departments (ED) are high-stake, high-variability environments constantly dealing with unscheduled and urgent patient visits under extreme time pressure. In recent years, Canadians have experienced steadily increasing wait times in the ED, with little to no improvement in the overall patient experience [1]. In Canada, between April and September 2024, the median ED length of stay for admitted patients was 15.7 hours, and the 90th percentile was 48.4 hours [2]. With such increasingly unacceptable wait times, improving ED operations efficiency has become more pressing than ever. However, the ED environment is highly dynamic, unpredictable, and filled with nuanced scenarios, making such improvements incredibly challenging [3, 4].

Variations in patient flow and service times can negatively impact operational efficiency [5]. In addition, the quality of work life, burnout, stress, and other conditions of physicians, nurses, and patients can also contribute to prolonged wait times [6–8]. With all constraints above, the ED must maintain effective care and perform resiliently under high stress [9]. Understanding these complex interactions is key to solving the problem.

To address these challenges, hospital administrators and researchers have historically turned to operations research methods to model patient flow and test interventions. Discrete event simulation, traditional agent-based simulation, and system dynamics have commonly been used to model the ED [10]. These simulation methods allow for testing interventions like adding a fast-track or altering staffing and predict their effects on length of stay (LOS) and waiting time [11], allowing simulations of patient flows, resource allocations, and care pathways [12].

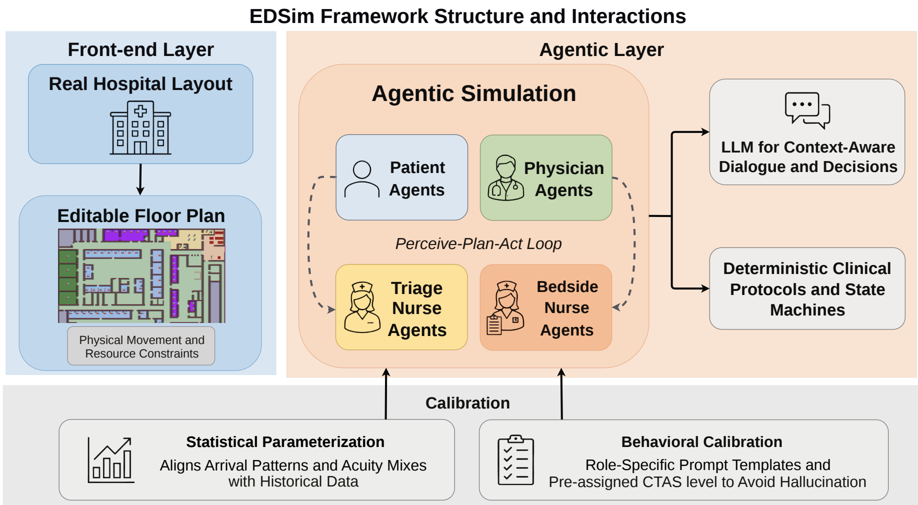

Fig. 1 The proposed EDSim framework.

Discrete Event Simulation (DES) has been used to study ED operations [10] and evaluate interventions that affected patient flow, such as adding a fast track for patients with low acuity or altering staffing schedules [13, 14]. For example, Hurwitz et al. [15] developed a flexible DES platform for ED crowding and validated it on both community and academic hospital settings. ED managers can use such tools to test changes like adding beds or staff and observe the impact on metrics such as length of stay and left without being seen rates [13]. Subsequent analysis showed that bottlenecks are context-specific: some EDs have limited physician availability, while others have bed capacity as the primary constraint [14].

Agent-Based Simulation (ABS) is another popular simulation technique in healthcare due to its ability to represent heterogeneous roles as agents with complex interactions [16]. Cabrera et al. [17] proposed an ABS model to optimize patient flow and resource allocation in the ED and demonstrated its applicability using real data from a large academic hospital in Spain. Liu et al. [18] introduced an ABS calibrated from 12-month data from the Hospital of Sabadell [19], demonstrating that their ABS can represent the emergent behavior of the ED. Additional studies have explored human factors: Son et al. [9] analyzed how staff behaviors affect ED performance under stress, while Zou et al. [20] quantified the impact of patient-flow variability on throughput in large-scale outpatient settings.

System Dynamics (SD) is a less commonly used simulation paradigm for modeling ED operations [10]. SD models the ED at an aggregate level using stocks, flows, and feedback loops, making it well-suited for analyzing system-wide pressures Schiff [21]. Romano et al. [22] developed an SD-based decision support tool for ED management to test operational policies and identify interventions that improve ED effectiveness under varying demand and capacity conditions. Wong et al. [23] showed via SD that smoothing the timing of inpatient discharges can reduce ED congestion by easing downstream bed blocking and mitigating boarding-driven delays. McAvoy et al. [24] built an SD model to evaluate which operational levers most effectively reduce bottlenecks and waiting delays, supporting access-to-care improvements through policy testing.

Despite prior works, traditional simulation methods remain challenging to capture more fine-grained, nuanced human interaction [16, 25]. While DES-based methods can provide valuable insight into ED bottlenecks, they are heavily reliant on pre-defined rules and static logic, which limits their ability to capture more nuanced aspects of ED. A small change to ED workflows could result in major refactoring of the underlying DES framework. ABS offers more flexibility in simulating agents with statuses and complex workflows, but it still relies on simplified decision logic and assumptions that limit adaptability to unseen scenarios. SD simulation focuses on understanding behavior over time in complex systems. In practice, many SD models are custom-built for a particular scenario, and generalizing them to new settings requires significant tuning. Neither of the approaches can 1) achieve human-like behavior, 2) directly visualize ED operations in a spatially accurate setting, and 3) enable realistic communication and complex interactions between agents. As a result, critical nuances of ED operations, such as physician differences in efficiency (e.g., faster vs. slower physicians), variation in patient treatment styles, and how new protocols play out, are lost. individuals characterized by complicated psychological and behavioral attributes [25].

The rapid advancement of Large Language Models (LLMs) in simulation has unlocked a new type of simulation, which Wu et al. [26] termed as Smart Agent-Based Modeling (SABM). LLMs have recently made waves in the simulation field thanks to their ability to mimic realistic human behavior [27]. A growing body of work now explores LLM-driven agent simulations, particularly in the social domain, where nuanced interaction is essential [28]. Gao et al. [28] surveyed how LLM-powered agents can be embedded in simulation frameworks, highlighting challenges in environment perception, alignment, and evaluation. A more frequently seen term would be LLM-powered agentic simulation, where agents are no longer rule-based but draw on generative models to simulate believable human behavior [29].

Table 1 Types of simulation for emergency department modeling.

	Discrete Event Simulation	**Agent-Based Simulation**	**System Dynamics**	**EDSim (proposed)**
Representation	Models the system as a sequence of discrete events; well-suited for analyzing patient flow, queues, and resource utilization.	Models the system as interacting agents with states and behavioral rules; captures emergent effects from interactions.	Models the system at an aggregated level using stocks, flows, and feedback loops; emphasizes system-wide dynamics.	Models the system as LLM agents with memory and spatial awareness operating within a virtual environment; captures nuanced clinical reasoning and emergent social behaviors.
Core logic	Event-scheduled: state changes occur at timestamped events on a simulation clock, subject to resource constraints and routing logic.	Agent-based: agents update state and act according to specified decision policies and interaction rules (often rule or state-based).	Continuous-time (or discretized): differential (or difference) equations define rates and feedback structure governing stock accumulation and depletion.	Hybrid architecture: LLM Agents follow a perceive-plan-act loop where LLM-generated decisions are constrained by deterministic state machines, role-specific prompt templates, and clinical protocols.

Agentic AI describes AI systems structured as agents that maintain an internal state, perform reasoning, plan, and take actions to achieve objectives under constraints. Compared to traditional DES, ABS, and SD, agentic AI, especially LLM-powered agents, can simulate real-world scenarios with significantly enhanced nuance and realism through their inherent reasoning capabilities [27, 29]. We show the detailed comparison of the three types of simulation in Table 1.

Almutairi et al. [30] shows that LLM agents can be placed in a 2D interactive environment, and the resulting team simulation demonstrates robustness for studying team behavior. AgentSociety [31] uses 10k+ LLM-driven agents and shows that agentic simulation can work at a large scale. Park et al. [29] simulated a virtual town environment with realistic individual and social behavior in a sandbox environment. Li et al. [32] demonstrate that LLM agents are capable of evolving medical knowledge, clinical reasoning, and interpersonal behavior over time, closely simulating real-world clinical staff and patient workflows. More recently, Teitge et al. [33] suggests that human-aware agentic AI could transform emergency care workflows by augmenting clinicians and coordinating complex, time-sensitive ED tasks. Despite existing efforts in both LLM-powered agents in hospitals [34] and traditional agent-based simulation in ED [16–18], the integration of LLM-powered agents into an agentic ED simulation remains largely unexplored.

In this work, we present EDSim, a novel agentic ED simulation framework powered by LLMs that can simulate ED roles, resources, and workflows. As Figure 1 shows, these simulated ED roles can plan, perceive the environment, and take actions. EDSim is designed to serve as a clinical sandbox for decision-makers, allowing for the testing of operational interventions in a risk-free virtual environment before they are deployed in the real world. Fundamentally distinct from traditional simulation methods, EDSim enables more realistic, adaptive, and context-sensitive modeling of clinical behavior through the use of LLM agents. To the best of our knowledge, this is the first ED simulator incorporating LLMs for ED decision logic and allowing map reconfiguration of a real hospital floor plan by users.

2 Methodology

We build EDSim from an open-sourced project [35], offering better accessibility over commercial software from Arena, FlexSim, CreateASoft, and Simio [36–39]. In addition to achieving similar accuracies for simulation results compared with traditional simulations, EDSim offers a visualized sandbox environment where interactions and agent conversations, and interventions of resources can both be viewed. EDSim is different from the work by Li et al. [32] employing a simulacrum mainly to evolve medical agents, and the framework by Du et al. [40] for achieving realistic patient-physician interactions. Instead, it focuses exclusively on simulating realistic ED operations. Our framework serves as a testbed which will ingest real ED data to study crowding, layout, and staffing interventions. We plan to develop a more faithful model of ED dynamics powered by agentic AI. We do not intend to use LLMs directly for clinical decision-making, for instance, diagnostics [41, 42] or triage [43, 44], since their performance and reliability on these tasks remain unclear; instead, we leverage their reasoning and

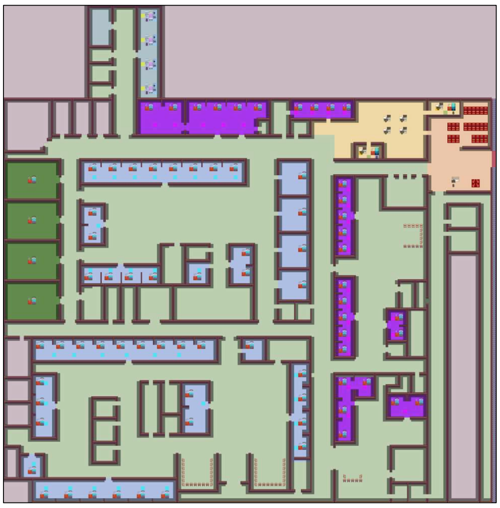

Fig. 2 A visual representation of a major Level-1 Trauma Center in Canada. The yellow area is the triage area. The purple area is minor treatment. The light blue is higher acuity and/or monitored bed spaces. The orange area is the waiting room. At the top of the layout is the diagnostic imaging area.

human-behavior emulation capabilities to create more realistic, adaptive agents for ED simulation.

The patient journey is calibrated at every step, from triage through diagnostics and discharge, based on historical aggregated ED flow data including mean daily presentations, proportion of high acuity visits, and mean wait times. No individual patient data was used in this study. This information is used to determine our arrival rates, triage acuity mix, service times, and transfer delays.

EDSim creates multiple staff roles, including triage nurses, bedside nurses, and physicians, each with a set of parameters that can represent different behavioral profiles and priorities. By capturing these human factors, EDSim can surface queue dynamics that simpler models might miss. The simulator records detailed events and metrics for each agent, including daily patient volumes and length-of-stay by acuity level, time spent in waiting rooms, treatment durations, idle times for staff, and throughput per hour. These metrics can be easily exported for visualization and statistical analysis. We acknowledge that not all ED roles and procedures have been programmed. However, the preliminary version of the simulation shows the potential of such situation.

EDSim is built as a modular, multi-layer application designed to support the complex requirements of simulating multiple ED cycles with high fidelity. The architecture is divided into three primary layers: User Interface Layer, Simulation Core Layer, and Data Management Layer. In EDSim, multiple agents interact within a shared virtual ED environment, which is built based on a real hospital layout, shown in Figure 2. We retain agents with internal state and memory that operate in a loop: they perceive the current status in the virtual ED, form a plan, select and execute the next action before repeating the cycle. In EDSim, an AI agent can test numerous factors affecting

ED throughput, allowing administrators to test-run proposed changes to see how they might affect their departments prior to adoption.

Dynamic agent creation. The ED environment follows a dynamic pattern of arrivals and departures for all roles. New patients are continuously being admitted, and staff shifts constantly change. We extended the simulation back-end to support creating and removing agents at run-time. EDSim supports adding new patient agents, adjusting the number of medical agents, and adjusting resources, all on the fly.

Spatial constraints. In EDSim, the AI agents are subject to realistic spatial awareness and movement constraints. We constrained each agent's knowledge of the environment to only those areas relevant to their role by setting the appropriate memory and location-specific rules in their action selection. Agents can only move following the settings of the map to access areas relevant to their objectives and tasks. These spatial constraints ensure the agents operate within the ED and ground the agents' behaviors in the physical reality of an ED.

ED Flow data integration. We integrated ED flow data to drive simulation parameters and agent decision contexts. This includes historical mean arrival rates by hour, distribution of patient acuity levels, and common chief complaints and diagnoses. Only historical aggregated flow data were used in the EDSim; individual patient data were not used in this study. It is important to note that not every type of data is incorporated in the simulation, since not all procedures in the ED have been implemented yet. Variables are as explained below:

- We calibrated the patient agent arrival rate by using an average patients-per-hour statistic. During each simulation step, we model this statistical arrival rate as a probability to determine whether a new patient will arrive, ensuring that over time, the arrival patterns will mirror those observed in reality.
- Each time a new patient agent is added, their illness or injury is sampled from the diagnosis distribution developed based on historical aggregate data.
- To ensure realism without overcomplicating the triage process, we assign CTAS scores based on standard mappings for each presentation type.

By combining data-driven elements with generative agents, the system can both quantitatively match ED flow patterns and qualitatively produce realistic individual interactions.

2.1 Editable floor plan

We implemented a small-scale ED layout used for testing and demonstration, and a full replica of a real Hospital's ED for high-fidelity scenarios. To build and manage these maps, we used an open-source map editor [45]. The floor plans are editable to support layouts of additional EDs from different hospitals and for potential resource placements, such as changes to the number of beds and open trauma bays.

2.2 Agent roles and behavior

EDSim currently models four types of agents in the ED: Patient, Triage Nurse, Bedside Nurse, and Physician. Each agent type is encoded with distinct behaviors, priorities, and decision-making heuristics that reflect their real-world responsibilities. We leverage different prompts and agent profile templates 2 to guide each agent's actions.

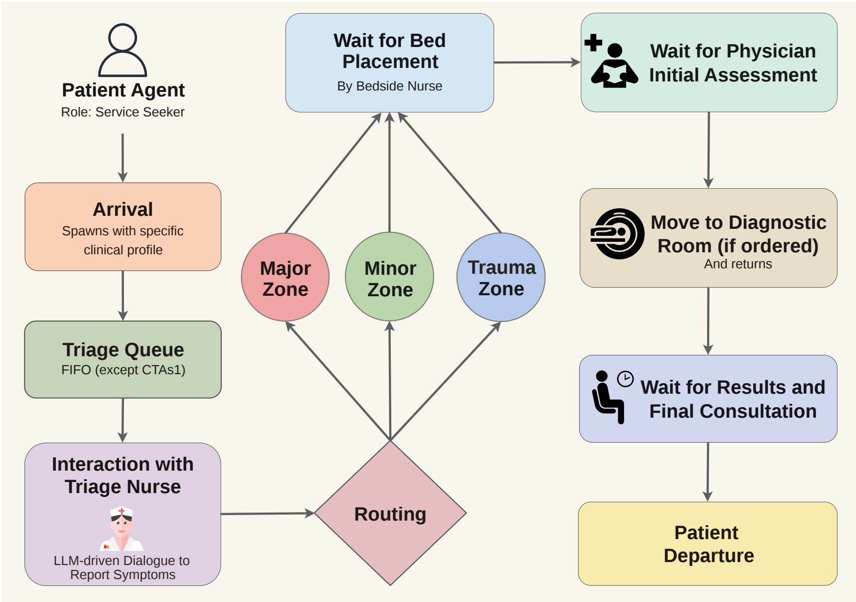

Fig. 3 A typical patient journey in the emergency department.

2.2.1 Patient

Patient agents represent individuals seeking care. They arrive with a set of complaints, generated based on aggregate historical presentation distributions. Patients largely behave passively, following directions from clinicians. Most of the time, they stay in the waiting areas until attended to, follow nurses or physicians as instructed, and provide information about their condition when asked. In the simulation, a patient's primary goal is to receive treatment, interact with staff, and spend most of the time waiting. However if a patient wait too long without being seen, the patient agent may decide to leave to simulate the real situation of Left Without Being Seen (LWBS) [46].

Role. Represents the ED service seeker.

Initialization. Upon creation, the patient is assigned a Clinical Profile sampled from aggregate historical distributions. This Clinical Profile includes a Chief Complaint, an initial acuity level, and a patience attribute.

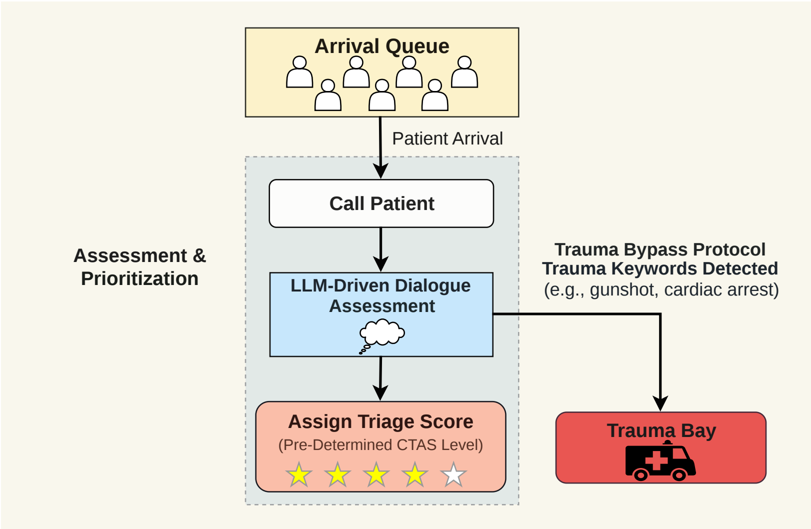

Fig. 4 A typical patient journey in the emergency department.

Logic. Patients primarily follow a linear flow: Wait → Triage → Registration → Treatment → Disposition, shown in Figure 3. Meanwhile, they possess the LWBS Logic. If their wait time exceeds their Patience Limit, the agent triggers a decision check. The LLM evaluates its current state and decides whether to leave the ED, simulating a critical metric of ED performance. Patients respond to staff inquiries but do not initiate any medical tasks.

2.2.2 Triage Nurse

The triage nurse agent is the first medical staff member patients interact with. We assign triage nurse agents the behavior of prioritizing incoming patients for initial assessment. Triage nurse monitors the waiting room for the next patient to call. When a patient arrives, and the triage station is free, the nurse will call the next patient, typically in order of arrival, but potentially jumping the queue if someone appears in critical condition, shown in Figure 4.

During the triage interaction, the nurse agent asks the patient about their issues and performs a brief observation; however, we do not use LLMs to evaluate patients and determine CTAS scores. Instead, each simulated patient is assigned a pre-determined CTAS level, which lets us focus on downstream flow and routing without explicitly modeling the full CTAS decision process. This design choice is also motivated by evidence that current automated approaches, especially general-purpose LLMs current do not yet provide sufficiently reliable CTAS assignments for clinical use [47–49]. Our Triage Nurse Agent, therefore, uses pre-assigned CTAS scores to determine the appropriate treatment area, and our implementation can fast-track CTAS-1 patients directly to a trauma bay.

Role. Gatekeeper of the ED.

Logic. The agent monitors the arrival queue. Its primary loop involves calling a patient, performing an assessment (LLM-driven dialogue), and assigning a priority score. We implemented a Trauma Bypass Protocol. If the LLM detects keywords associated with an immediate life threat (e.g., gunshot, cardiac arrest), the agent bypasses the standard registration process and routes the patient immediately to the Trauma Bay, triggering a Code Blue event in the simulation.

2.2.3 Bedside Nurse

Bedside nurses are the agents managing patient care within the treatment zones. Their responsibilities include transferring patients from the waiting area to treatment beds, carrying out or assisting with procedures and tests, and generally monitoring patients in the zone. Detailed workflow for the bedside nurse is shown in Figure 5.

Role. Manager of patient logistics and care execution.

Logic. The Bedside Nurse manages a task queue for their assigned zone. Tasks include transport to bed, administering medications, patient monitoring, assessment, and transport to diagnostic imaging (in lieu of having a porter built into this proof of concept simulation). The agent uses a weighted algorithm to sort tasks. Priority is determined by CTAS score, wait time, and Task urgency.

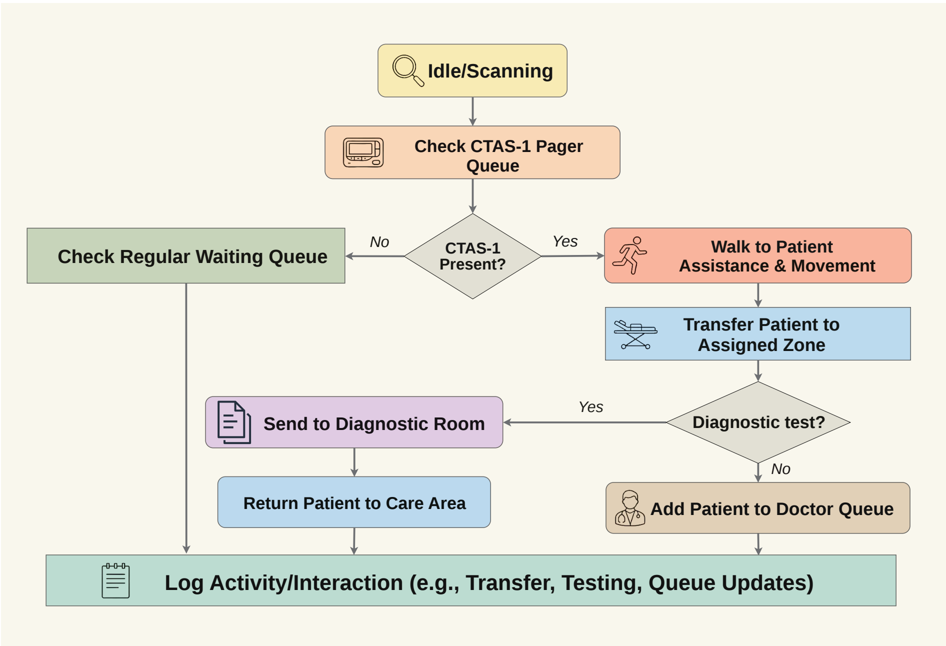

Fig. 5 Full cycle of bedside nurse.

2.2.4 Physician

Physician agents oversee medical decision-making and definitive patient care, demonstrated in Figure 6. In the simulation, physicians are attached to the treatment zones. They will see patients who have been placed in beds in their zone. A physician agent's workflow involves selecting a patient to see next, performing an evaluation, ordering tests or treatments, and later following up.

Role. Clinical decision-maker.

Logic. The Physician agent has the most complex loop. They must constantly scan the tracking board (a list of all patients in their zone) and decide on the next action: Assess (new patient), Review (lab results returned), or Discharge. If the Physician determines a patient needs to see a specialist, such as a Psychiatrist or an Orthopedist, they generate a consultation request. This shifts the patient to a state, effectively pausing their flow until an external timer simulating the specialist's arrival expires.

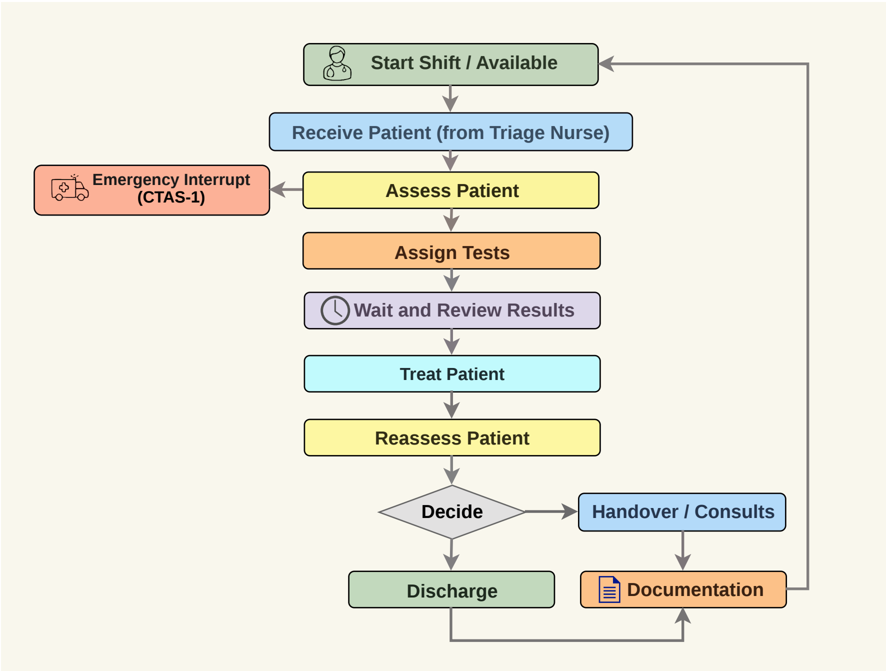

Fig. 6 Full cycle of a physician.

2.3 A Patient's Journey

Using the agents and resources defined above, our simulation models the full lifecycle of a patient agent moving through the ED, from arrival to discharge. In this section, we describe the following events:

2.3.1 Arrival and Waiting

Upon generation, a patient agent enters the waiting room area and joins the queue of patients waiting for triage. If a triage nurse is free, a triage interaction is initiated. If not, the patient agent waits in the waiting room. This waiting time is tracked for metrics.

2.3.2 Triage Assessment

When a triage nurse agent is available, the patient agent will be called and taken to the triage room, where the triage interview will take place. If no triage nurse is available, the patient agent will enter a separate queue for waiting. In this phase, the nurse gathers the patient's complaints and assesses their condition. The patient will be assigned an acuity score (CTAS-1 to CTAS-5) and a zone designation for treatment. If the patient's condition is critical (CTAS-1), the nurse can designate the Trauma Room instead and will immediately flag this for urgent handling. The patient's symptoms, CTAS score, and time spent are recorded in the simulation.

2.3.3 Bed Assignment and Transfer

Once triaged, the patient's next step is getting a bed space. Our simulation manages a bed queue for each zone. Triaged patients are placed into the queue corresponding to their assigned zone. If a bed is available immediately and a bedside nurse is free, a transfer happens. Otherwise, the patient waits in the waiting room until their turn and a bed frees up. Bedside nurse agents are continuously monitoring these queues. In each zone, bedside nurse agents will determine the highest-priority patient first according to CTAS, and then based on the patient's waiting time. Upon arrival at the bed, the patient is marked as "in bed" and now officially under the care of that zone's team. After securing a bed space, the patient agent enters a state of waiting for physician evaluation.

2.3.4 Initial Physician Evaluation

Each physician agent maintains a maximum caseload of active patients that is adjustable. Patient assignment follows a priority-based queuing mechanism: upon completing triage, patients are inserted into a global waiting queue sorted by their CTAS acuity score. Physicians with availability pull patients from this shared queue. Once assigned, patients are placed on the physician's internal queue, which employs a queue-aging algorithm that periodically decrements priority values to prevent lower-acuity patients from being indefinitely deferred. The physician then selects the most urgent patient from their personal wait list for assessment. The physician goes to the patient's bedside and begins the physician-patient consultation. Physician agents will order diagnostic test based on the symptom of the patients. In the simulation's data model, we then mark this patient as "awaiting tests". Our simulation currently doesn't detail clinical interventions, focusing more on flow, For the general flow, though, ordering tests is the main outcome of the initial exam.

2.3.5 Diagnostic Imaging

After the physician's orders, the patient remains in their bed until the diagnostic imaging room is ready. The bedside nurse in that zone knows the patient needs, say, a diagnostic imaging, and a nurse agent will initiate transport when the diagnostic imaging room can take the patient. As of now, we do not simulate a complicated scheduling system for diagnostics; rather, if the tech is available, the nurse will move the patient right away, and if the tech is busy, the patient will wait until the tech becomes available. Once the test is done, the patient is escorted back to their bed by the nurse and waits for results. The physician is automatically notified when results are ready, at which point the patient will need a follow-up.

2.3.6 Follow-Up and Discharge

The final stage is when the physician returns to review the test results with the patient. The physician agent will see that the patient's test results are available and put that patient back into their "to-see" list if not already there. The physician returns to the patient's bedside for a follow-up consultation. In EDSim, most cases conclude with the patient being discharged with advice or treatment. The physician then formally marks the patient for discharge. While in real life, some patients remain in the ED after discharge, awaiting bedspace in the hospital, our current simulation does not yet model specific scenarios of post-disposition boarding delays, which we leave for future work. However we do provide the option to assign admitted patients a configurable boarding duration sampled uniformly from a user-defined range during which they continue to occupy their ED bed to capture the downstream impact of admission on ED throughput. We log the time of discharge to compute their total length of stay. The bed they occupied becomes free, which can immediately trigger the next waiting patient to be brought in by a nurse, thus continuing the cycle. Once this is done, the patient's journey in the ED is complete.

In summary, EDSim can produce metrics like total length of stay, waiting times at each stage, and resource utilization. Also included were chat histories and explanations of why each agent performs certain actions, such as selecting certain patients first.

2.4 Spatially explicit environment and editable Maps

We utilized the Tiled map editor to create a representation of a partner hospital's emergency department. The map defines:

Zones. High acuity zone, Low acuity zone, Trauma Bays, Triage, Waiting room, Diagnostics.

Resources. Specific coordinates for Beds, Chairs, and Trauma Bays.

Constraints. Walls and barriers are defined as collision layers, preventing agents from moving through them.

2.5 Adding/Removing AI agents

Our simulation framework supports agent addition and removal. This method assigns unique identifiers, initializes agent profiles based on predefined role-specific templates, and manages agent state initialization, such as patient symptoms or staff shift status. Agents are seamlessly added to the active simulation, which updates scheduling loops, environmental states, and indexing mechanisms essential for continuous logging and analytics. Conversely, agent removal is handled by flagging agents as inactive, immediately excluding them from scheduling and active environment tracking. Staff agents are removed during idle states to prevent disruption, with responsibilities smoothly reassigned as necessary.

2.6 Adding/Removing resources

Our simulation allows adjustments to physical resources such as beds and equipment. In EDSim, our map can contain a maximal set of resource locations. The map defines the maximum possible bed capacity for the ED. The maze status file controls how many of those beds are active by specifying a number to remove from each zone at initialization. To reduce capacity, users can increase the removal count; to restore capacity, they can decrease it. Since the map always starts with the full set of bed positions, adding beds does not require modifying the map itself, it simply means removing fewer beds from the available set.

Resource configuration is not limited to beds. Zone capacity in general is influenced by beds, since each zone's throughput is bed-limited in our ED model. We also allow configuration of the number of triage rooms. For example, if there were multiple triage nurse stations available, one could mark more than one triage area as active. We subsequently allow configuring the availability of the trauma room. For instance, if the ED has 2 trauma bays but we plan to simulate that only 1 is staffed, we could set the trauma capacity to 1. The simulation would then treat only one trauma bed as usable and ignore the other.

We also make sure the frontend always renders exactly what the backend has defined as active, so there is no discrepancy. To focus on the real-world ED scenario, we included an option to "max out" the ED at start, meaning all beds start occupied by patients actively being treated. This uses both the agent addition and resource configuration features: if the user sets an initial occupancy level equal to the number of beds, the simulation will spawn the same number of patient agents and immediately place them into each bed at the beginning of the simulation. This can simulate starting in a fully booked state, which is useful to test how the ED recovers from being at capacity or how it handles new arrivals when no beds are initially free. The dummy patients can either have random profiles or a standardized placeholder profile, and they will go through the motions of being treated and discharged over time, freeing beds for new arrivals.

By allowing the addition and removal of resources, our simulation can address questions often raised by ED managers, such as “What if we had 5 more beds in this zone?” or “How does closing Zone B temporarily affect waiting times?” These resource-focused experiments are made possible by the combination of an editable map and the initialization logic that respects scenario configurations for resources.

In summary, EDSim enables modeling of physical resources as configurable parameters, with changes reflected both in backend logic and frontend display.

2.7 Data Collection, analytics and output

As the simulation runs, we log events and state changes to build a detailed timeline for each patient and utilization statistics for each resource and agent. We structure data collection to compute key metrics common in ED operations research, and we provide the outputs in formats convenient for offline analysis. For the agent side, we record:

- Arrival time, triage start and end times, bed assignment time, time of first physician contact, times of start and end of any diagnostic tests, time of second physician contact, and discharge time.
- Total Length of Stay (LOS), Waiting room time, and bed wait time
- Treatment time, time waiting for physician, time with physician, time getting tests, physician-to-disposition time.
- CTAS (acuity) level and zone.

All these events and metrics are written to output files at the end of the simulation. We output a comprehensive CSV file where each row corresponds to one patient and contains columns for all the timestamps and durations, plus their attributes listed above. We also output summary statistics in either a JSON or a separate summary CSV. The JSON might contain an object with average wait times, max queue lengths observed, total arrivals, total discharges, and so on.

Key time intervals of patient journey. A major focus of our study is on three key time intervals [50] that segment a patient’s journey through the ED. These intervals are Arrival-to-PIA (physician initial assessment) time, PIA-to-Dispo (disposition) time, and Dispo-to-Leave time, corresponding to phases of ED flow. Arrival-to-PIA time is the interval from when a patient arrives and registers in the ED to when a physician first begins their initial assessment. This essentially captures how long patients wait before being seen by a physician, and it is a standard performance metric. PIA-to-Dispo time is measured from the start of the physician’s initial assessment to the moment a disposition decision is made. This interval represents the active treatment and evaluation phase, including physician evaluation, tests, consultations, and any treatment up to the decision point. Finally, Dispo-to-Leave time is the interval from the disposition decision to the patient’s actual physical departure from the ED. This captures any “post-decision” delays, such as waiting for an inpatient bed after admission orders or lingering in the ED after discharge due to transportation or bed-cleaning delays. In other words, Dispo-to-Leave reflects boarding time for admitted patients and other departure delays. By observing these three components, we can observe from a finer granularity where waiting bottlenecks in the ED occur.

2.8 Calibration, parameterization, and behavioral control

To ensure EDSim functions as a realistic simulation of the ED, we developed a hybrid calibration framework. This process aligns the simulation with real-world operations through two distinct mechanisms: statistical parameterization of the environment using historical aggregate data and behavioral calibration of agents using a layered control architecture.

2.8.1 Statistical parameterization via historical data

The simulation environment was grounded using historical aggregate data from a major Level-1 Trauma Center. This historical aggregate data provided the ground truth for establishing arrival rates, triage acuity mixes, service times, and transfer delays. EDSim can also be parameterized given statistical data from another ED. We parameterized the simulation inputs to statistically mirror the stochastic nature of ED operations.

Arrival patterns. We calibrated patient generation using historical aggregate hourly arrival statistics. During each simulation step, these rates were converted into probabilities to determine new patient arrivals, ensuring the simulation reproduced the volume fluctuations observed in the real ED.

Clinical profiles. Upon initialization, each patient agent was assigned a clinical profile with chief complaint and diagnosis sampled directly from the historical aggregate distribution.

Acuity distribution. To maintain realistic acuity mixes without overcomplicating the triage logic, we assigned Canadian Triage and Acuity Scale (CTAS) scores based on standard mappings for each presentation type. This ensured the proportion of high-acuity vs. low-acuity patients aligned with historical distributions, bypassing the unreliability of purely LLM-generated acuity scores.

2.8.2 Behavioral calibration with layered control

While statistical parameters define the flow of the ED, the fidelity of individual agents depends on their decision-making logic. We observed that relying solely on LLMs for intricate, role-specific behaviors often resulted in inconsistent actions. Therefore, we implemented a layered calibration approach that combines hard behavioral constraints with role-specialized prompt templates.

To achieve consistent behavior, we engineered specialized controls directly into the agent classes. Rather than employing a single, generic prompt template for all agents, the specialized prompt templates are used for each distinct role within the simulated ED, as shown in Table 2. This approach is designed to ensure that each agent's actions and decisions are highly consistent with their designated professional responsibilities.

Each prompt template is a text file pre-structuring the instructions sent to an LLM. These templates are dynamically populated with contextual information specific to an agent's current state and the simulation's progress. This dynamic population allows the LLM to generate responses and actions that accurately reflect the agent's role-specific behavior. For example, templates dictate parameters for:

Table 2 Agent roles and persona constraints in the ED simulation.

Agent	**Daily plan**	**Core role**	**Hard constraints**	**Default area**
Patient	Visits ED	Receives symptoms, interacts with staff, and follows ED flow across waiting, triage, major, and low acuity zones as directed.	Avoid moving; follow staff instructions for treatment and movement.	Waiting room.
Physician	Works in ED	Provides information to patients about symptoms, diagnoses, and assigns a treatment plan; treats patients based on the preference queue.	Works in the injury zones; avoid other rooms.	Injury zones.
Triage Nurse	Works in ED	Assess patients at the triage computer, ask about symptoms, assign CTAS, then communicate key findings to the physician.	Works in the triage room; avoid other rooms.	Triage room.
Bedside Nurse	Works in ED	Nursing patient across injury zones: provides critical care in High acuity zone and checks patients in low acuity zone; works with physicians and triage.	Works in the injury zones; avoid other rooms.	Injury zones.

- **Movement and Navigation:** Guiding agents in their decisions regarding subsequent movements based on their assigned tasks and the evolving environment.
- **Communication Structure:** Defining the style, content, and objectives of conversations pertinent to their role, including patient intake, physician consultations, and nurse handovers.
- **Daily Planning and Task Prioritization:** Influencing how agents organize their simulated workday and prioritize tasks in alignment with their professional duties.

A key aspect involves the customization of prompt templates for specific actions, such as an agent's decision to move to a new location. For these critical decision points, we incorporate role-specific examples directly within the template. This provides the LLM with highly relevant and focused examples, enhancing the LLM's ability to generate behaviors that are not only consistent with the agent's job but also reduce hallucination. This granular control over agent prompts is instrumental in achieving high levels of behavioral accuracy and realism within the simulation.

2.8.3 AI agent behavior specialization and control

For instance, the Patient agent includes mechanisms to trigger state changes when interacting with specific ED staff members. We ensure their journey through the facility aligns with typical patient flows.

Similarly, the Bedside Nurse agent's action sequence includes selecting the next patient from a queue, navigating to the patient's location, engaging in a brief, role-appropriate conversation, and then escorting the patient to their designated room.

The physician agent class follows a comparable pattern for patient interactions, involving navigation to the patient's location and initiating conversations.

Furthermore, we've implemented explicit rules governing inter-agent communication to maintain focus and prevent hallucination. For example, Patient agents are strictly prevented from initiating conversations with other agents, as open-ended patient dialogue frequently leads to irrelevant conversational topics. Conversely, medical staff agents are mandated to initiate a conversation if they are walking to another agent for a direct interaction. However, to keep agent actions well-defined and to avoid unrealistic mid-task branching, we treat an en route movement as an atomic action in the current simulator. This means staff would not initiate new conversations while traveling. We acknowledge that real ED work is frequently interrupted. We will leave the modeling, allowing interruption to future work.

These specialized behavioral controls, including movement patterns, interaction protocols, and communication constraints, are engineered directly into the agent specifications. This decision was made after observing that relying solely on the LLM for these intricate, role-specific behaviors resulted in inconsistent and often undesirable agent actions. By explicitly defining these specialized behaviors, we guarantee the necessary fidelity and predictability required for a robust and realistic ED simulation.

2.9 Front-end interface

To improve usability, we implemented a simple front-end webpage that supports controlling our EDSim simulator directly through the UI, shown in Figure 7. The interface

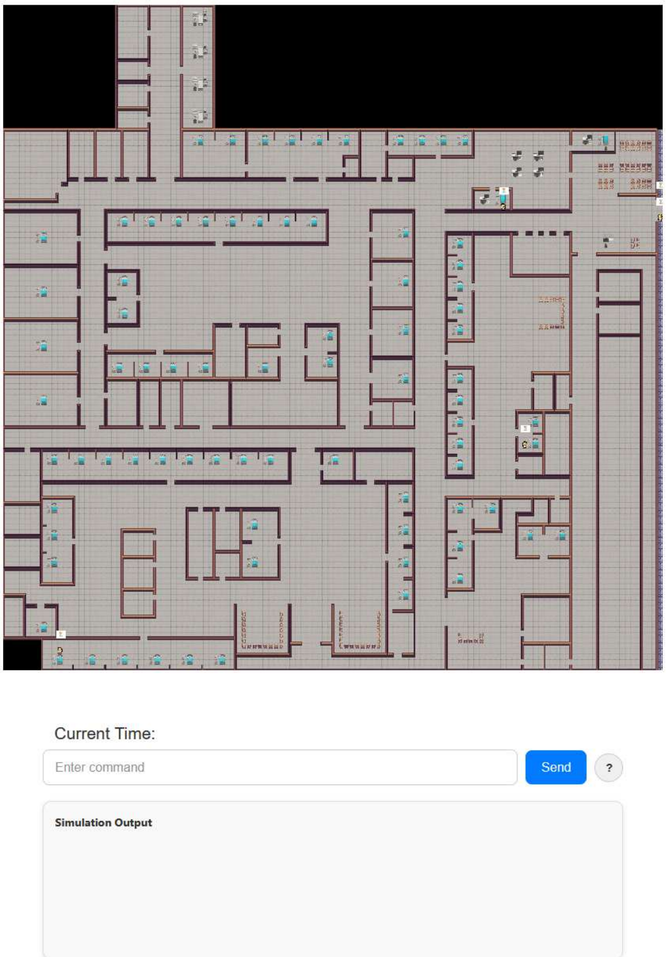

Fig. 7 Front-end interface of EDSim.

includes a command input box that allows users to enter commands through the webpage, making the simulator more accessible for testing and demonstrations.

In addition, we have a simulation configuration page for setting up a new run. The details are shown in Figure 8. This page centralizes the key simulation parameters, including staffing levels, patient-rate modifiers, and timing settings. Users can quickly define a scenario and start the simulation without editing code or scripts.

To support experimentation across a range of ED configurations, the simulation exposes a comprehensive set of user-configurable parameters. These parameters govern staffing levels, patient arrival rates, clinical workflow timings, patient behavioural rules, spatial movement characteristics, and hospital admission dynamics. Users can modify these values to model different operational scenarios, such as surge conditions,

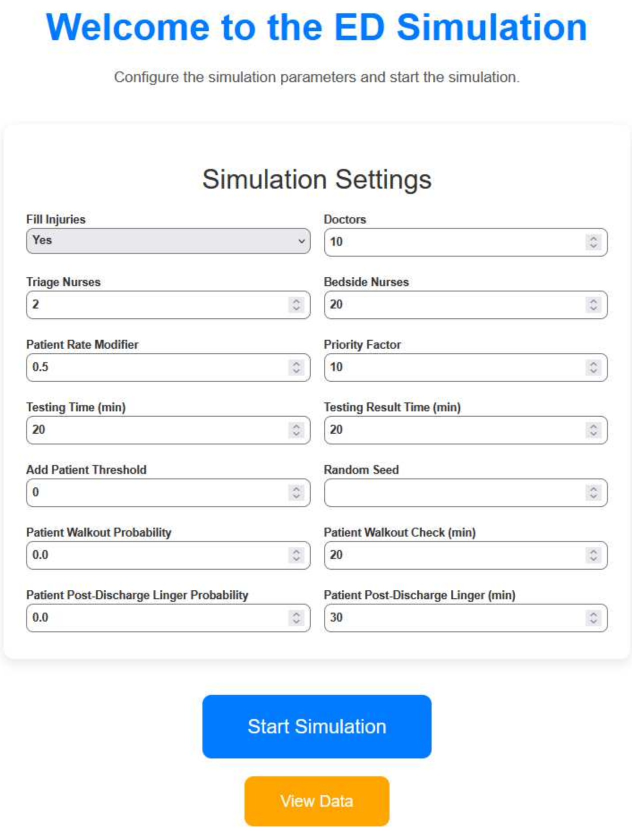

Fig. 8 Simulation configuration page for EDsim.

reduced staffing, or varying acuity distributions. Table 3 summarizes the full set of configurable parameters and their descriptions.

We provide a live web-based dashboard for real-time monitoring of the simulation state, shown in Figure 9. The dashboard displays summary status cards showing the current simulation time, step count, number of patients currently in the ED, and the number of completed patient visits. Below these, interactive charts visualize the distribution of patients across workflow states, zone occupancy relative to bed capacity, queue lengths for triage, bedside nurse, pager, and physician queues, nurse utilization broken down by activity status, and per-physician patient load. A tabular view at the bottom reports stage-level timing breakdowns for each completed patient, enabling direct inspection of individual wait times and durations across the care pathway.

3 Scenario-based Simulation Results

In this section, we present the simulation results for two scenarios. We start with the standard scenario comparing key statistical data from simulations with aggregate

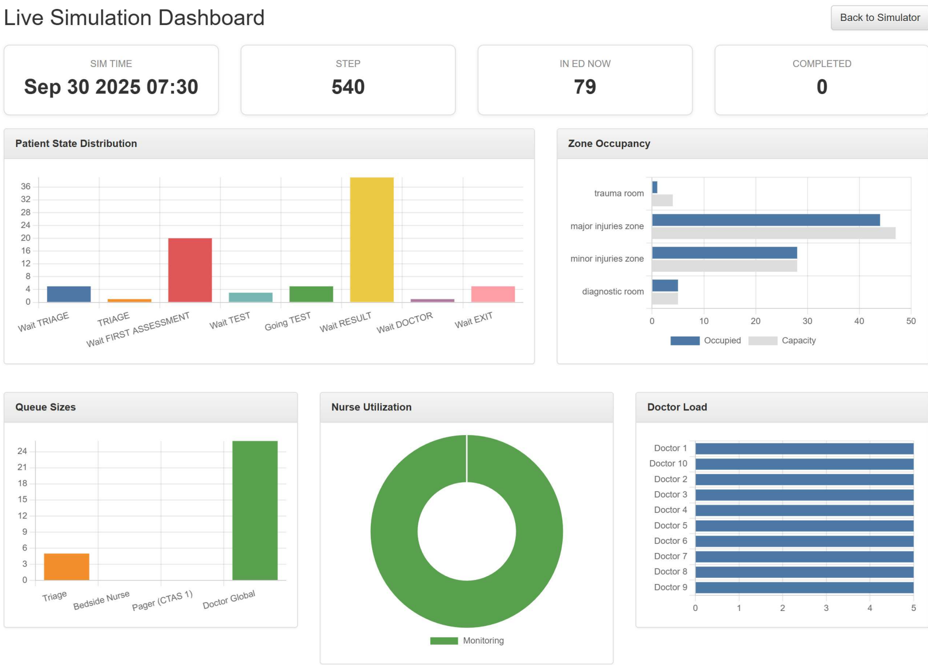

Fig. 9 Simulation dashboard for EDSim.

historical data of a major Level-1 Trauma Center. Then, we show the simulation results with patient volume surge, reduced number of nurses, additional physicians, diagnostic imaging improvements, and discharge-to-leave improvements.

3.1 Experimental Setup

This subsection describes the environment in which the experiments were conducted, including the simulator's architecture, configuration parameters, and the characteristics of the AI agents.

3.1.1 Baseline settings

The ED simulator was built from Python version 3.9.12 for the backend simulation logic and Phaser for the frontend visualization, and a Django web framework. Using the historical aggregate data, we configured EDSim baseline settings to replicate a hospital's typical daily patient load and acuity mix. Table 4 shows the configuration used for resource capacities and staff processing times to mimic the ED.

Table 5 demonstrates the distributions of CTAS used for arriving patients. Key parameters such as hourly arrival rates, distributions of CTAS triage categories, staffing levels of physicians and nurses, and service time distributions were calibrated to reflect this historical aggregate data.

Table 3 Configurable simulation parameters defined in the scenario metadata file.

Parameter	Description
General Settings	
Start Date	The calendar date on which the simulation begins.
Start Time	The starting time of the simulation (date and time).
Seconds per Step	The number of real-world seconds each simulation tick represents.
Staffing	
Physicians	Number of physician agents initialized at simulation start.
Triage Nurses	Number of triage nurse agents initialized at simulation start.
Bedside Nurses	Number of bedside nurse agents initialized at simulation start.
Patient Arrival	
Arrival Rate Modifier	Scaling factor applied to the hourly arrival rate distribution; values less than 1.0 reduce the arrival intensity.
Surge Baseline Rate	Baseline arrival rate during surge scenarios.
Initial Conditions	
Bed Fill Fraction	Fraction of available ED beds to pre-fill with patients at startup.
Waiting Room Preload	Number of patients placed in the waiting room at initialization.
Departure Window	Maximum time offset assigned to pre-loaded patients' departure timers, distributing their exits across the simulation.
Clinical Workflow	
Priority Factor	Multiplier applied to CTAS scores when computing queue priority; higher values increase the separation between acuity levels.
Max Patients per Physician	Maximum number of active patients assigned for a single physician.
Testing Time	Time in minutes for a patient to travel to and complete a test.
Result Wait Time	Time in minutes for diagnostic test results to become available.
Testing Probability	Per-CTAS probability that a patient requires diagnostic testing.
Diagnostic Room Capacity	Maximum number of patients that can occupy the diagnostic testing area simultaneously.
Patient Behaviour	
Walkout Probability	Probability that a waiting patient leaves without being seen.
Linger Probability	Probability that a discharged patient lingers in the ED.
Linger Duration	Duration a lingering patient remains in the ED after discharge.
Hospital Admission & Boarding	
Simulate Admission	Yes or no for simulating admission and boarding.
Admission Probability	Per-CTAS probability that a patient is admitted to hospital following disposition.
Min Boarding Duration	Minimum duration for admitted patients awaiting transfer.
Max Boarding Duration	Maximum duration for admitted patients awaiting transfer.
Spatial Movement	
Travel Speed	Agent walking speed in meters per second.

3.1.2 AI agents configuration

The simulation incorporates four AI agent roles: Patients, Physicians, Triage Nurses, and Bedside Nurses. Users can edit each agent's configuration and adjust the number of agents for each simulation run. We provide options to start the ED with patients in all beds and injury zones to reflect full capacity. Additional patients were dynamically introduced throughout the simulation based on real-world arrival data to ensure realistic workload patterns.

We use *gpt-4o-mini* and *text-embedding-3-small*. The state machine provides a structured framework for agent behavior, while the LLMs enable more nuanced and

Table 4 Resource and agent capacities for EDSim.

Input variable	Capacity	Process time (minutes)
Triage nurse	up to 5	
Assessment		LLM conversation
Discharge		10
Bedside nurse	up to 25	
Diagnostic escort to test		20
rest between tasks		15
Discharge		10
Physician	up to 15	
Treatment (high priority)		Consultation via chat
Treatment (low priority)		Consultation via chat
Consultation		4.5
post-visit rest		15
Diagnostic imaging room	up to 4 tables	20
Beds - high acuity zone	up to 47	
Beds - low acuity zone	up to 28	
Beds - trauma room	up to 4	

All inputs were based on the data collected from a major Level-1 Trauma Center in Canada.

Table 5 Proportion of total visits by CTAS score.

CTAS	Proportion (%)
1	5.9
2	39.3
3	41.5
4	10.9
5	2.4

context-aware decision-making within predefined states. The specific implementation for each agent role is detailed below:

Patients. Patient agents are controlled by a state machine, which determines their progression through the ED. Their conversational interactions and personality generation are controlled by an LLM. Progression to the next state is triggered by successful communication with medical staff. Their role is to enter the ED and request treatment from the staff, adhering to their instructions. Initial symptoms are assigned based on historical aggregate distributions. In addition, they have the opportunity to leave the ED by making an LLM call. LLM makes the decision based on state, personality, time spent in the state, and condition. Personality is assigned as unstable or impatient, unstable being that they may leave randomly, and impatient, that they may leave after a long time.

Physician. Physician agents are largely controlled by the LLM. They possess an assessment queue to identify and respond to patients located in the injury zones and are scripted to initiate conversations with them. They make an LLM call to pick a Patient based on their preference. The Physician in this simulation has a preference to pick the next one up in the queue. Their goal is to speak with patients, assign necessary tests, and determine discharge status.

Triage Nurse. The Triage Nurse agent is scripted to remain within the Triage Room, where incoming patients arrive for their triage assessment. Their goal in the conversation, generated by the LLM, is to provide a Canadian Triage and Acuity Scale (CTAS) score and assign them to either the high acuity zone, the low acuity zone, or the trauma room. The LLM dynamically generates these conversations to yield appropriate CTAS according to the CTAS guidelines.

Bedside Nurse. Bedside Nurse agents have triggers to respond to patients who are ready to be transferred to their assigned room or to the diagnostic room for testing. The Bedside Nurse will retrieve the next patient in the queue, engage in conversation, and facilitate their movement to the next room. For testing, they will bring the patient to the diagnostic room for testing, wait 20 minutes for the testing to be completed, and then escort them back to their previous room.

3.2 Baseline scenario validation using historical data

We ran multiple simulation replications under the baseline settings to produce aggregate performance metrics for a normal day. The detailed per-patient results can be found in the supplementary. The simulation outputs were compared against the historical aggregate ED metrics stratified by triage acuity to assess fidelity, shown in Table 6. This experiment served as an internal validation, assessing whether the LLM-driven simulation reproduces the aggregate flow patterns observed in an ED. We found that the baseline EDSim closely reproduced the observed patient flow patterns. In particular, the model's wait time distributions aligned well with historical aggregate data across all triage levels, with mean and median values falling within expected ranges across acuity levels. These results, summarized in our Table of time-interval statistics, demonstrate that EDSim's baseline configuration successfully mirrors real-world ED performance. This validation builds confidence that the LLM-driven agent-based approach can achieve macro-level accuracy in addition to plausible micro-level interactions.

We report ED time-interval descriptive statistics stratified by CTAS for both aggregate historical data and the simulation; simulation intervals were derived by aggregating per-patient stage durations into the corresponding time intervals, and we summarize heavy-tailed waiting times using medians [IQR], means (SD), observed real-world min-max, and simulation percentiles (p10/p50/p90/p95) to enable direct comparison.

3.3 Patient volume surge scenario

After establishing our baseline, we examined a patient influx scenario to explore how the ED behaves under an extreme influx of patients. In this scenario, we sweep the

Table 6 Arrival to PIA (stage1) by CTAS (minutes): aggregate historical throughout vs simulation.

CTAS	Historical med [IQR]	Historical mean (SD)	Historical min–max	Sim n	Sim p10/p50/p90/p95	Sim mean (SD)
Stage 1: Arrival to PIA						
1	8 [11]	12.7 (21.4)	0–440	10	18.4/127.5/215.9/218.7	125.9 (70.8)
2	111 [201]	150.0 (139.0)	0–1384	79	58.9/175.1/404.9/507.3	215.6 (150.5)
3	203 [209]	218.0 (148.0)	0–1019	84	69.4/197.4/436.0/525.9	231.4 (176.3)
4	212 [195]	230.0 (143.0)	0–829	11	96.6/236.7/543.5/549.5	270.1 (175.1)
5	195 [200]	221.0 (151.0)	1–870	3	151.2/195.2/623.6/677.1	355.4 (326.2)
Stage 2: PIA to Disposition						
CTAS	Historical med [IQR]	Historical mean (SD)	Historical min–max	Sim n	Sim p10/p50/p90/p95	Sim mean (SD)
1	243 [319]	296.0 (248.0)	0–1411	10	96.6/195.0/427.6/599.0	254.1 (208.9)
2	248 [250]	296.0 (220.0)	1–1440	79	73.1/206.8/633.3/892.2	289.3 (263.2)
3	185 [227]	233.0 (199.0)	0–1430	84	50.1/161.2/552.3/754.5	245.2 (226.6)
4	97 [148]	146.0 (150.0)	0–1363	11	147.0/326.8/611.5/766.7	364.0 (240.7)
5	69 [108]	113.0 (132.0)	1–1240	3	123.5/204.0/245.5/250.6	187.7 (77.5)
Stage 3: Disposition to Leave						
CTAS	Historical med [IQR]	Historical mean (SD)	Historical min–max	Sim n	Sim p10/p50/p90/p95	Sim mean (SD)
1	79 [288]	313.0 (760.0)	0–22916	10	3.0/75.7/196.5/270.7	90.6 (107.6)
2	0 [100]	201.0 (725.0)	0–21881	79	1.2/3.7/108.2/324.1	61.8 (201.1)
3	0 [4]	116.0 (535.0)	0–21014	84	1.1/3.1/195.4/415.9	57.8 (129.4)
4	0 [0]	50.7 (341.0)	0–15939	11	1.1/1.7/3.8/129.7	25.1 (76.4)
5	0 [0]	29.5 (460.0)	0–18891	3	0.3/1.3/2.6/2.7	1.4 (1.4)

Table 7 Stress test vs baseline: overall percent change in mean stage times (all CTAS).

Metric (min)	Baseline mean (SD)	Stress mean (SD)	Abs diff	% change
ArrivaltoPIA	223.3 (165.4)	390.1 (314.9)	166.7	74.7%
PIAtoDispo	270.3 (241.5)	460.8 (354.3)	190.4	70.4%
DispotoLeave	58.4 (159.7)	117.2 (253.7)	58.8	100.7%
Total ED	552.0 (323.7)	968.0 (490.4)	416.0	75.4%

overall arrival rate by modifying the arrival-rate parameter. We hold staffing constant and report where performance "tips", which refers to a sharp rise in waiting-room time, IN BED saturation, and LWBS spikes, using aggregated daily metrics and patient-level timeline distributions.

As shown in Table 8, in this patient volume surge experiment, the overall arrival rate was increased by 60% to mimic a sudden surge while holding staffing and other resources constant.

In addition, we adjusted post-disposition settings to reflect slower bed turnover, where many patients were made linger in the ED after discharge by increasing the post-discharge occupancy time and its probability. This scenario simulated a worst-case overcrowding situation in which new patients flood in faster than usual, and discharged patients do not vacate beds promptly.

The upstream stages of patient flow exhibited substantial degradation under influx conditions (Table 8, Table 7). The mean arrival-to-PIA time increased by 74.7%, from 223.3 minutes (SD = 165.4) to 390.1 minutes (SD = 314.9), and the mean PIA-to-disposition time increased by 70.4%, from 270.3 minutes (SD = 241.5) to 460.8 minutes (SD = 354.3). Notably, the standard deviations nearly doubled across both stages, indicating that patient experiences became far more variable under stress. Lower-acuity patients were especially affected, often waiting several hours before assessment as higher-acuity cases consumed the limited physician capacity. As physicians managed a larger active queue, follow-up tasks and reassessments were deferred, further extending the treatment phase.

The disposition-to-departure stage showed the largest proportional degradation of any stage, with mean time increasing by 100.7%, from 58.4 minutes (SD = 159.7) to 117.2 minutes (SD = 253.7). As bed turnover slowed and in-bed occupancy remained elevated, a subset of patients experienced greatly prolonged post-disposition boarding. Although the median disposition-to-leave time remained low in both conditions, reflecting that the majority of patients still departed quickly, the mean reveals that tail-end patients were stuck substantially longer under stress. The growth in standard deviation (from 159.7 to 253.7 minutes) confirms that the heavy tail of the distribution expanded significantly, making the mean and standard deviation essential for capturing the true impact of overcrowding on this stage. Table 7 illustrates how these stage-level delays compound at the system level: mean total ED length of stay increased by 75.4%, from 552.0 minutes (SD = 323.7) to 968.0

Table 8 Baseline vs Influx simulation stage durations (minutes) by CTAS (influx trimmed to match baseline CTAS 1–4; CTAS 5 kept complete).

Stage 1: Arrival to PIA

CTAS	Base n	Base p10/p50/p90/p95	Base mean (SD)	Influx n	Influx p10/p50/p90/p95	Influx mean (SD)
1	10	18.4/127.5/215.9/218.7	125.9 (70.8)	10	26.0/237.9/725.5/814.7	319.2 (283.2)
2	79	58.9/175.1/404.9/507.3	215.6 (150.5)	53	78.2/300.2/773.2/937.8	383.2 (301.3)
3	84	69.4/197.4/436.0/525.9	231.4 (176.3)	60	92.8/328.0/898.4/1188.9	426.3 (343.1)
4	11	96.6/236.7/543.5/549.5	270.1 (175.1)	13	145.6/188.3/661.9/808.9	306.4 (265.6)
5	3	151.2/195.2/623.6/677.1	355.4 (326.2)	3	186.3/194.0/660.6/718.9	385.2 (339.5)

Stage 2: PIA to Disposition

CTAS	Base n	Base p10/p50/p90/p95	Base mean (SD)	Influx n	Influx p10/p50/p90/p95	Influx mean (SD)
1	10	96.6/195.0/427.6/599.0	254.1 (208.9)	10	194.4/402.6/886.2/1034.0	493.0 (319.0)
2	79	73.1/206.8/633.3/892.2	289.3 (263.2)	53	142.0/381.8/808.1/1001.9	453.9 (371.9)
3	84	50.1/161.2/552.3/754.5	245.2 (226.6)	60	116.6/388.7/889.6/939.1	459.6 (338.1)
4	11	147.0/326.8/611.5/766.7	364.0 (240.7)	13	125.8/302.0/1004.8/1136.2	501.4 (427.8)
5	3	123.5/204.0/245.5/250.6	187.7 (77.5)	3	93.7/265.8/568.4/606.2	320.2 (300.4)

Stage 3: Disposition to Leave

CTAS	Base n	Base p10/p50/p90/p95	Base mean (SD)	Influx n	Influx p10/p50/p90/p95	Influx mean (SD)
1	10	3.0/75.7/196.5/270.7	90.6 (107.6)	10	1.5/3.7/382.1/683.2	132.7 (314.9)
2	79	1.2/3.7/108.2/324.1	61.8 (201.1)	53	1.3/3.7/359.6/637.8	101.4 (229.1)
3	84	1.1/3.1/195.4/415.9	57.8 (129.4)	60	1.0/2.1/471.5/524.6	130.9 (270.7)
4	11	1.1/1.7/3.8/129.7	25.1 (76.4)	13	1.1/4.0/255.8/546.9	133.2 (269.2)
5	3	0.3/1.3/2.6/2.7	1.4 (1.4)	3	1.3/1.8/2.7/2.8	2.0 (0.9)

minutes (SD = 490.4). Overall, this stress test demonstrates how exceeding ED operating capacity induces a nonlinear degradation under increased demand, with both the central tendency and variability of wait times increasing across all three stages. This scenario underscores the importance of surge planning and highlights how our simulator can identify the thresholds at which an ED transitions from stable operation to an overwhelmed state.

Overall, this stress test demonstrates how exceeding ED operating capacity induces a nonlinear degradation under increased demand, with both the central tendency and variability of wait times increasing across all three stages. This scenario underscores the importance of surge planning and highlights how our simulator can identify the thresholds at which an ED transitions from stable operation to an overwhelmed state.

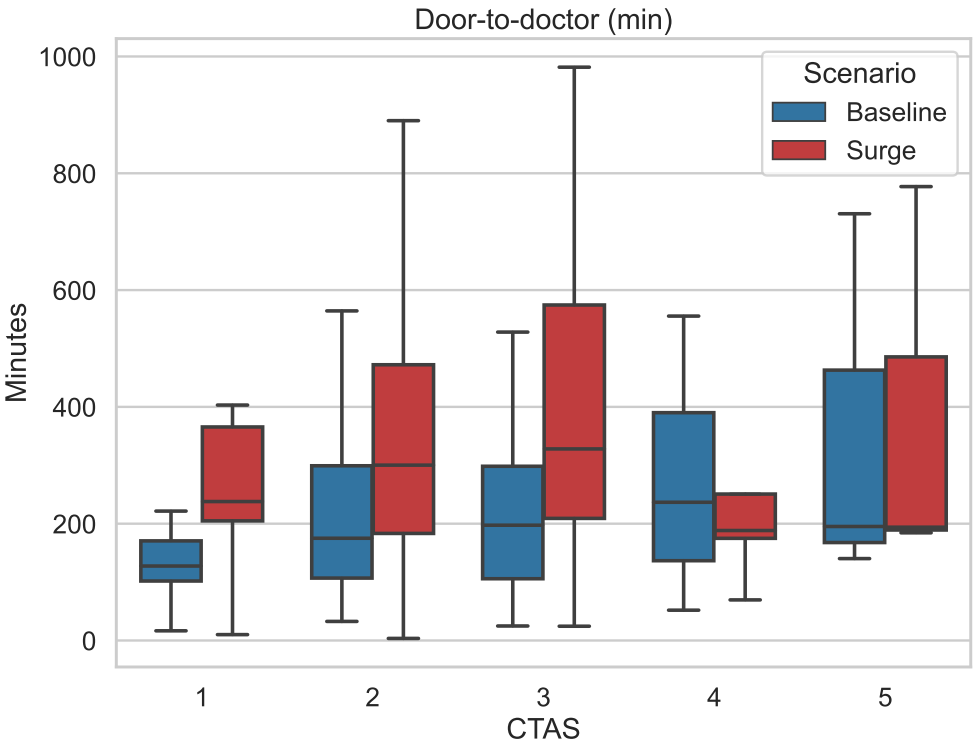

Fig. 10 Distribution of door-to-doctor times (minutes) by CTAS level under baseline (n = 187) and surge (n = 139) conditions. Boxes represent the interquartile range (Q1–Q3), horizontal lines indicate the median, and whiskers extend to 1.5× IQR. Under surge conditions, CTAS 1–3 patients experienced substantially longer and more variable wait times before physician assessment, whereas CTAS 4–5 showed less pronounced changes.

When stratified by CTAS level, the three boxplot comparisons reveal that surge-induced degradation is pervasive but not uniform across acuity groups, shown in

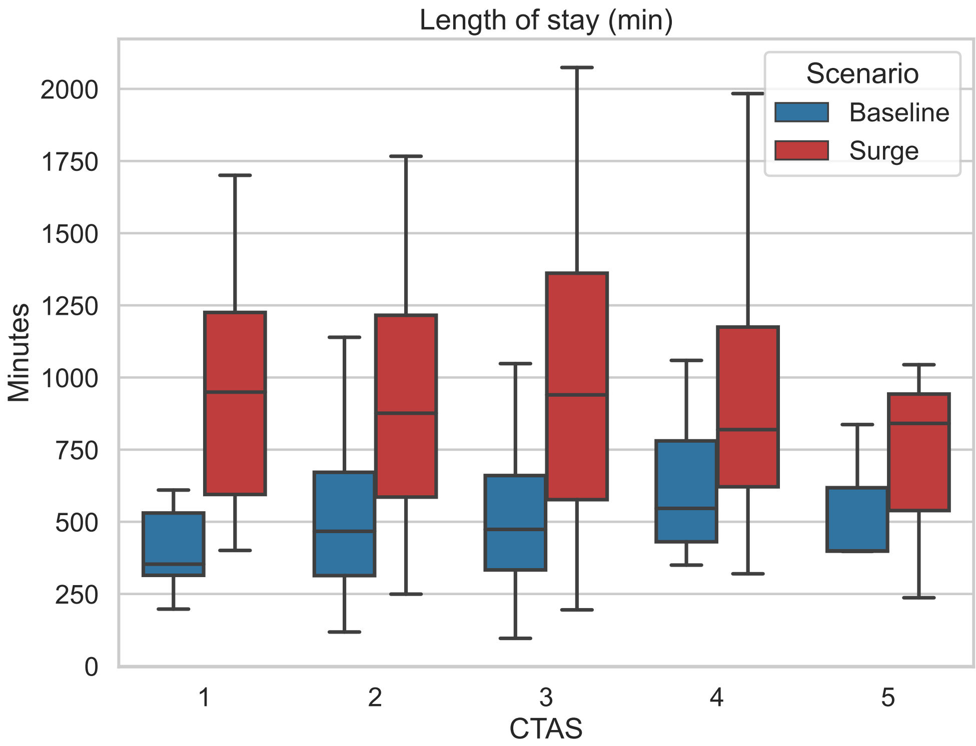

Fig. 11 Distribution of total ED length of stay (minutes) by CTAS level under baseline and surge conditions. All CTAS levels showed a marked upward shift in both central tendency and dispersion under surge, with the median length of stay approximately doubling across most acuity groups. CTAS 1 and CTAS 3 patients exhibited the widest surge distributions, with upper whiskers exceeding 1700 and 2000 minutes, respectively, reflecting extreme tail-end delays.

Figure. 10, Figure. 11, and Figure. 12. Door-to-doctor time showed the largest proportional median increase for CTAS 1 patients (127.5 to 237.9 minutes, an 87% increase), followed by CTAS 3 (197.4 to 328.0 minutes, +66%) and CTAS 2 (175.1 to 300.2 minutes, +71%), indicating that even the highest-acuity patients were not insulated from access delays during overcrowding. Length of stay degradation was universal: median values approximately doubled for every CTAS level, rising from a range of 353–546 minutes at baseline to 819-949 minutes under surge, confirming that the bottleneck effect propagates through the entire care pathway regardless of patient acuity. Treatment time distributions likewise shifted upward under surge, but the most striking feature is the explosion in variability, particularly for CTAS 4, whose surge interquartile range expanded from 283 minutes (165-448) to 836 minutes (151–987), suggesting that lower-acuity patients are disproportionately subject to intermittent deprioritization when physician workload is elevated. Across all three metrics, the widening of interquartile ranges and whisker spans under surge underscores that overcrowding does not merely increase average delays; it makes the patient experience far less predictable.

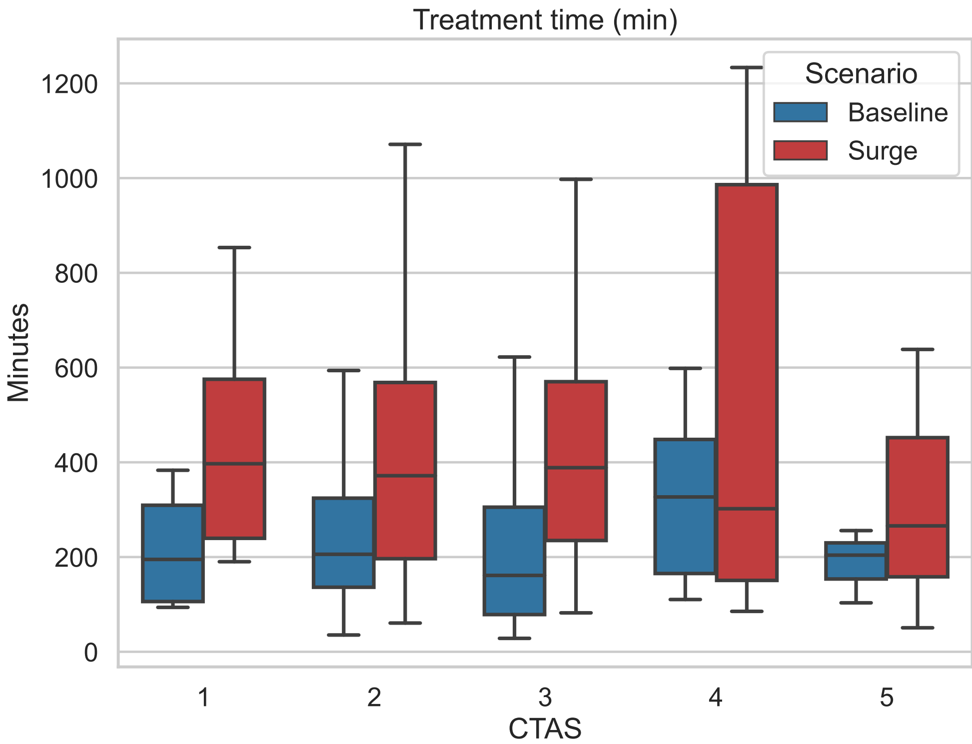

Fig. 12 Distribution of treatment time (minutes) by CTAS level under baseline and surge conditions. Treatment time captures the interval from physician initial assessment to disposition. Under surge, treatment times increased across most CTAS levels, with notably greater variability; CTAS 4 patients showed the widest surge interquartile range, suggesting that a subset of lower-acuity patients experienced substantially prolonged treatment phases when physician workload was elevated.

A counter-intuitive finding is that CTAS 4–5 door-to-doctor medians remained stable or even decreased under surge (CTAS 4: 236.7 to 188.3 minutes); however, given the small sample sizes (n = 11 and n = 13 for CTAS 4, n = 3 each for CTAS 5), this pattern likely reflects stochastic variation rather than a systematic effect and should be interpreted with caution.

4 Discussion

Our contribution is unique from existing traditional commercial simulators [37–39, 51] and other LLM agent frameworks [29, 32, 34, 52] in several key ways. First, we employ a *hybrid cognitive architecture* that combines deterministic clinical protocols with LLM reasoning. This ensures that agents strictly adhere to medical standards. In Canada, this refers to the Canadian Emergency Department Triage and Acuity Scale (CTAS) Guidelines [53]. Meanwhile, EDSim retains the flexibility to handle ambiguous scenarios and exhibit human-like reasoning. Second, our model is data-calibrated, parameterized using aggregated ED flow data to ensure that the foundational metrics, such as arrival patterns, acuity distributions, and service times, reflect actual hospital operations rather than theoretical assumptions. Third, we incorporate an editable full replica of a local hospital emergency department in Canada to enforce realistic spatial constraints on agent movement and interaction.

In our simulation results, these time intervals provide insight into ED performance under various conditions. Under normal conditions, the Arrival-to-PIA interval is relatively short, indicating that most patients are being seen in a timely manner. Dispo-to-Leave is especially informative for detecting output delays: a prolonged Dispo-to-Leave time often means admitted patients are boarding in the ED or discharged patients are delayed in departure. Our surge experiment quantifies this effect: the mean Dispo-to-Leave time doubled (+100.7%), and its standard deviation grew from 159.7 to 253.7 minutes, revealing that a tail of patients experienced extreme post-disposition boarding even though the majority still departed quickly. This heavy-tailed behaviour makes the mean and standard deviation essential complements to the median, which alone would understate the overcrowding impact on this stage. Even when physician assessment and treatment are efficient, downstream output constraints can lead to prolonged boarding after the decision to admit, stalling patient flow and amplifying congestion across the department. This finding is consistent with real-world ED crowding scenarios, where boarding admitted patients after the decision to admit is made is identified as a primary driver of prolonged ED stays and crowding [54]. By examining these three intervals, the simulation allows us to diagnose whether a surge in total ED length of stay is primarily stemming from delays before seeing a physician, during treatment, or after disposition. Being aware of the source of major delays would allow ED managers to make informed decisions addressing the most pressing bottlenecks with precious resources.

Both baseline and surge/influx experiments should be interpreted as a successful proof of concept: while our simulator already reproduces the expected nonlinear degradation under overload, we have not yet modeled the full scope of ED complexity (e.g., the complete patient mix, provider roles, and procedure/test pathways), which will be incorporated in future iterations.

Triage priority and initial patient queue. Our agentic simulation represents how patients are prioritized and queued at various stages (triage, bed assignment, and physician assessment) in a manner that mirrors real ED triage principles. In the real ED, the Canadian Triage and Acuity Scale (CTAS) classifies patients from Level 1 (most critical) to Level 5 (least urgent), and patients are generally seen in order of urgency rather than strictly by arrival time. CTAS Level 1 cases (resuscitation) are considered life-threatening and should be seen by a physician immediately, essentially 100% of the time, while CTAS Level 2 (emergent) patients are ideally seen within 15 minutes in about 95% of cases. Our simulation incorporates these principles by assigning higher priority to higher-acuity patients, but it also realistically models how operational constraints can still lead to some waiting even for high-acuity cases under certain conditions.

In our simulation, we model the pre-triage stage as a first-in-first-out (FIFO) queue based on arrival time when triage capacity is saturated. Before formal triage classification, patients wait in the order they enter the system. CTAS-1 is treated as a special case: patients flagged as CTAS-1 are immediately escalated, bypassing the FIFO queue. For example, a CTAS-1 patient gets immediate bed assignment and physician attention, ahead of anyone else waiting. For CTAS-2 to CTAS-5, patients remain in the FIFO pre-triage queue until assessed by the triage nurse. Our model's handling of CTAS-1 ensures those patients are attended to as quickly as possible, given resource availability. For patients CTAS-2 to CTAS-5, the simulation uses a priority-weighted queueing system. Rather than a simple FIFO queue, each patient's priority score is determined by their CTAS level and adjusted dynamically over time. This is equivalent to an accumulating priority queue approach [55], where waiting time gradually increases a patient's priority if they are not yet seen. In the simulation, a CTAS-2 patient will generally be prioritized over CTAS-3/4/5 patients for assignment to an available bed or physician. However, our simulation also ensures fairness so that as lower-acuity patients wait, their effective priority increases, allowing them to eventually be seen. We use this approach to balance urgency and wait time. Thus, while a newly arrived CTAS-2 patient will jump ahead of their CTAS-4 peers, a CTAS-4 who has been waiting for a long time will gradually rise in priority. This dynamic system is more reflective of real ED operations than a rigid, hard-coded rule, since real triage nurses and charge staff continuously reassess waiting patients' conditions and how long they've been waiting.

In the simulation, our physician assignment logic similarly takes into account CTAS-based priority but with some flexibility. Physicians pick the next patient from an assessment queue that is weighted by acuity, yet if a physician is dedicated to a certain zone, they will see the next sickest patient waiting in their zone. This means the prioritization is mostly by acuity, but operational factors can override a strict global priority order. The net effect in the simulation is that CTAS-1 patients are virtually always seen first, CTAS-2 patients are very likely to be seen next, and so on, but with occasional out-of-order selections caused by spatial/resource constraints. Notably, implementing this accumulating priority queue discipline improved the fairness and performance of the model; similar approaches have been shown to outperform simple static triage-priority rules in managing ED patient flow.

Reflection. Our results demonstrate that EDSim can simulate the workflows of an emergency department, achieving both macro-level accuracy and micro-level realism. The close match between the simulation's baseline outputs and the historical aggregate data of a major Level 1 Trauma Center is a crucial validation, as it indicates that our agentic simulation, when calibrated with real statistics, can reliably reproduce the complex dynamics of patient flow in the ED. Notably, the simulator not only captured aggregated metrics like median wait times but also the variability and heavy-tailed delays characteristic of ED operations. This suggests that the stochastic behaviors and queuing phenomena in the real system are being correctly simulated. Such fidelity is a prerequisite for using a simulator to experiment with operational changes. If the baseline were inaccurate, conclusions drawn from any what-if scenarios would be questionable.

The strong alignment with reality gives confidence that interventions tested in the simulation could translate into meaningful insights for the actual ED. The EDSim framework lays out the foundation for deploying customized interventions based on each ED's resource constraints and immediate needs. Moreover, we observed that the agentic roles in EDSim engaged in realistic interactions and decision-making processes. For example, patient agents left without being seen when waits became intolerable, and virtual providers communicated and reprioritized tasks in ways that mirror real clinical workflow. These qualitative aspects, while harder to quantitatively validate, add credibility and richness to the simulation, potentially uncovering issues such as communication delays or cognitive overload that existing simulators might miss.

Limitations. While EDSim demonstrates strong alignment with historical aggregate ED metrics across CTAS levels 2 through 5, CTAS 1 patients present a notable limitation. In practice, CTAS-1 cases often trigger an immediate response with a physician at the bedside within minutes. The historical medians are on the order of 8 minutes. However, our simulated CTAS-1 throughput times can be longer than expected due to inherent overhead in the current architecture: each interaction can require LLM API calls for agent decision-making and conversation generation, and the movement system requires agents to physically traverse the environment tile-by-tile. These computational and spatial costs impose a practical floor on how quickly any patient can be processed, regardless of priority. In a real ED, a CTAS-1 patient may bypass nearly all waiting and movement through direct-to-resus workflows and task preemption, but our proof-of-concept does not yet model true interrupt-driven, multi-provider resuscitation dynamics, nor can it compress LLM inference and pathfinding below their runtime constraints. As a result, matching the very short CTAS-1 door-to-physician median is difficult without adding more explicit resuscitation pathways, which we plan to incorporate in future work to better reflect resuscitation realities.

5 Data Availability

The aggregated historical dataset used in this study is not publicly available due to privacy and confidentiality. To support reproducibility, we provide (i) an example dataset with calibrated noise added, and (ii) the detailed synthetic per-patient generated dataset used in our experiments. These files are available in the project GitHub repository at https://github.com/denoslab/EDSim.

6 Code Availability

The complete source code for this study is available at: https://github.com/denoslab/EDSim.

7 Acknowledgments

This work was supported by the New Frontiers in Research Fund (NFRF) Exploration Stream NFRF-E/00269-2024.

8 Author Contributions

J.W. conceptualized, designed, conducted the experiments, and wrote the original draft and final manuscript. Z.W. and A.B. developed the initial version of the simulation framework. H.L. led subsequent simulation development and refinement, later joined by D.V., O.O., and M.G. B.T. facilitated an emergency department field visit and provided physician insights that informed the workflow design and clinical realism of the simulation. M.H. coordinated and gathered the study data. S.C. provided emergency department domain input to guide clinical realism and workflow assumptions. T.C., Z.M., T.W., T.R., E.L., and J.Z. provided substantive feedback and domain-specific input throughout the study. S.D. and J.H. supervised the project and provided substantial input across study design, analysis, interpretation, and manuscript development. All authors critically revised the manuscript and approved the final version.

9 Competing Interests

The authors declare no competing interests.

References

[1] Varner, C.: Emergency departments are in crisis now and for the foreseeable future. CMAJ (2023)

[2] Health Information, C.I.: NACRS emergency department visits and lengths of stay. https://www.cihi.ca/en/nacrs-emergency-department-visits-and-lengths-of-stay. Accessed: 2025-07-27 (2025)

[3] Rader, T., Ritchie, L.: Emergency department overcrowding in canada: Multistakeholder dialogue. Canadian Journal of Health Technologies **3**(10) (2023)

[4] Slade, D., Scheeres, H., Manidis, M., Iedema, R., Dunston, R., Stein-Parbury, J., Matthiessen, C., Herke, M., McGregor, J.: Emergency communication: the discursive challenges facing emergency clinicians and patients in hospital emergency departments. Discourse & Communication **2**(3), 271–298 (2008)

[5] Flottemesch, T.J., Gordon, B.D., Jones, S.S.: Advanced statistics: developing a formal model of emergency department census and defining operational efficiency. Academic Emergency Medicine **14**(9), 799–809 (2007)

[6] Bragard, I., Dupuis, G., Fleet, R.: Quality of work life, burnout, and stress in emergency department physicians: a qualitative review. European Journal of Emergency Medicine **22**(4), 227–234 (2015)

[7] Hunsaker, S., Chen, H.-C., Maughan, D., Heaston, S.: Factors that influence the development of compassion fatigue, burnout, and compassion satisfaction

in emergency department nurses. Journal of nursing scholarship **47**(2), 186–194 (2015)

[8] Ríos-Risquez, M.I., García-Izquierdo, M.: Patient satisfaction, stress and burnout in nursing personnel in emergency departments: A cross-sectional study. International journal of nursing studies **59**, 60–67 (2016)

[9] Son, C., Sasangohar, F., Rao, A.H., Larsen, E.P., Neville, T.: Resilient performance of emergency department: Patterns, models and strategies. Safety Science **120**, 362–373 (2019)

[10] Gul, M., Guneri, A.F.: A comprehensive review of emergency department simulation applications for normal and disaster conditions. Computers & Industrial Engineering **83**, 327–344 (2015)

[11] Day, T.E., Al-Roubaie, A.R., Goldlust, E.J.: Decreased length of stay after addition of healthcare provider in emergency department triage: a comparison between computer-simulated and real-world interventions. Emergency Medicine Journal **30**(2), 134–138 (2013)

[12] Miller, M.J., Ferrin, D.M., Messer, M.G.: Fixing the emergency department: A transformational journey with edsim. In: Proceedings of the 2004 Winter Simulation Conference, 2004., vol. 2, pp. 1988–1993 (2004). IEEE

[13] Ahsan, K.B., Alam, M., Morel, D.G., Karim, M.: Emergency department resource optimisation for improved performance: a review. Journal of Industrial Engineering International **15**(Suppl 1), 253–266 (2019)

[14] De Santis, A., Giovannelli, T., Lucidi, S., Messedaglia, M., Roma, M.: A simulation-based optimization approach for the calibration of a discrete event simulation model of an emergency department. Annals of Operations Research **320**(2), 727–756 (2023)

[15] Hurwitz, J.E., Lee, J.A., Lopiano, K.K., McKinley, S.A., Keesling, J., Tyndall, J.A.: A flexible simulation platform to quantify and manage emergency department crowding. BMC medical informatics and decision making **14**(1), 50 (2014)

[16] Hawe, G.I., Coates, G., Wilson, D.T., Crouch, R.S.: Agent-based simulation for large-scale emergency response: A survey of usage and implementation. ACM Computing Surveys (CSUR) **45**(1), 1–51 (2012)

[17] Cabrera, E., Taboada, M., Iglesias, M.L., Epelde, F., Luque, E.: Optimization of healthcare emergency departments by agent-based simulation. Procedia computer science **4**, 1880–1889 (2011)

[18] Liu, Z., Rexachs, D., Epelde, F., Luque, E.: An agent-based model for quantitatively analyzing and predicting the complex behavior of emergency departments. Journal of computational science **21**, 11–23 (2017)

[19] Liu, Z., Rexachs, D., Epelde, F., Luque, E.: A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity. Computers & Industrial Engineering **103**, 300–309 (2017)

[20] Zou, C., Wang, J., Cheng, Y.: The impact of the variability of patient flow and service time on the efficiency of large-scale outpatient systems. IEEE Transactions on Computational Social Systems **10**(3), 1230–1240 (2022)

[21] Schiff, G.D.: System dynamics and dysfunctionalities: levers for overcoming emergency department overcrowding. Academic Emergency Medicine **18**(12), 1255–1261 (2011)

[22] Romano, E., Guizzi, G., Chiocca, D.: A decision support tool, implemented in a system dynamics model, to improve the effectiveness in the hospital emergency department. International Journal of Procurement Management **8**(1-2), 141–168 (2015)

[23] Wong, H.J., Wu, R.C., Caesar, M., Abrams, H., Morra, D.: Smoothing inpatient discharges decreases emergency department congestion: a system dynamics simulation model. Emergency Medicine Journal **27**(8), 593–598 (2010)

[24] McAvoy, S., Staib, A., Treston, G.: Can a system dynamics model of the emergency department show which levers reduce bottlenecks and delays to improve access to care? Systems Research and Behavioral Science **38**(1), 61–79 (2021)

[25] Zhang, X.: Application of discrete event simulation in health care: a systematic review. BMC health services research **18**(1), 687 (2018)

[26] Wu, Z., Peng, R., Han, X., Zheng, S., Zhang, Y., Xiao, C.: Smart agent-based modeling: On the use of large language models in computer simulations. arXiv preprint arXiv:2311.06330 (2023)

[27] Park, J.S., Popowski, L., Cai, C., Morris, M.R., Liang, P., Bernstein, M.S.: Social simulacra: Creating populated prototypes for social computing systems. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, pp. 1–18 (2022)

[28] Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., Li, Y.: Large language models empowered agent-based modeling and simulation: A survey and perspectives. Humanities and Social Sciences Communications **11**(1), 1–24 (2024)

[29] Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. In: Proceedings of the

36th Annual Acm Symposium on User Interface Software and Technology, pp. 1–22 (2023)

[30] Almutairi, M., Chiang, C., Guo, H., Belcher, M., Banerjee, N., Milkowski, M., Volkova, S., Nguyen, D., Weninger, T., Yankoski, M., et al.: Simulating teams with llm agents: Interactive 2d environments for studying human-ai dynamics. arXiv preprint arXiv:2510.08242 (2025)

[31] Piao, J., Yan, Y., Zhang, J., Li, N., Yan, J., Lan, X., Lu, Z., Zheng, Z., Wang, J.Y., Zhou, D., et al.: Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. arXiv preprint arXiv:2502.08691 (2025)

[32] Li, J., Lai, Y., Li, W., Ren, J., Zhang, M., Kang, X., Wang, S., Li, P., Zhang, Y.-Q., Ma, W., et al.: Agent hospital: A simulacrum of hospital with evolvable medical agents. arXiv preprint arXiv:2405.02957 (2024)

[33] Teitge, B., Holodinsky, J., Wu, J., Lang, E., Drew, S.: Can human-aware agentic artificial intelligence transform emergency care workflows? Canadian Journal of Emergency Medicine (2026) https://doi.org/10.1007/s43678-026-01105-6

[34] Yao, Z., Yu, H.: A survey on llm-based multi-agent ai hospital. OSF (2025)

[35] Park, J.S.: generative agents. GitHub. Accessed 2026-02-04. https://github.com/joonspk-research/generative_agents

[36] Automation, R.: Arena Simulation Software. https://www.rockwellautomation.com/en-ca/products/software/arena-simulation.html. Accessed: 2026-01-29

[37] FlexSim Software Products, I.: Better Decisions in the Emergency Department. https://www.flexsim.com/healthcare/case-studies/better-decisions-emergency-department/. Accessed: 2026-01-28

[38] CreateASoft, I.: Emergency Room, Emergency Department Simulation. https://www.createasoft.com/emergency-room-simulation. Accessed: 2025-07-27 (2023)

[39] Simio: Digital Twin Simulation Software — Simio. https://www.simio.com/. Accessed: 2026-01-29

[40] Du, Z., LujieZheng, L., Hu, R., Xu, Y., Li, X., Sun, Y., Chen, W., Wu, J., Cai, H., Ying, H.: Llms can simulate standardized patients via agent coevolution. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 17278–17306 (2025)

[41] Horiuchi, D., Tatekawa, H., Shimono, T., Walston, S.L., Takita, H., Matsushita, S., Oura, T., Mitsuyama, Y., Miki, Y., Ueda, D.: Accuracy of chatgpt generated diagnosis from patient's medical history and imaging findings in neuroradiology

cases. Neuroradiology **66**(1), 73–79 (2024)

[42] Hirosawa, T., Kawamura, R., Harada, Y., Mizuta, K., Tokumasu, K., Kaji, Y., Suzuki, T., Shimizu, T.: Chatgpt-generated differential diagnosis lists for complex case–derived clinical vignettes: Diagnostic accuracy evaluation. JMIR Medical Informatics **11**, 48808 (2023)

[43] Masanneck, L., Schmidt, L., Seifert, A., Kölsche, T., Huntemann, N., Jansen, R., Mehsin, M., Bernhard, M., Meuth, S.G., Böhm, L., *et al.*: Triage performance across large language models, chatgpt, and untrained doctors in emergency medicine: comparative study. Journal of medical Internet research **26**, 53297 (2024)

[44] Zaboli, A., Brigo, F., Brigiari, G., Massar, M., Parodi, M., Pfeifer, N., Magnarelli, G., Turcato, G.: Chat-gpt in triage: Still far from surpassing human expertise-an observational study. The American Journal of Emergency Medicine **92**, 165–171 (2025)

[45] mapeditor: Tiled: Flexible level editor. https://github.com/mapeditor/tiled. Accessed: 2026-01-27 (2025)

[46] Roby, N., Smith, H., Hurdelbrink, J., Craig, S., Hawthorne, C., DuMontier, S., Kluesner, N.: Characteristics and retention of emergency department patients who left without being seen (lwbs). Internal and Emergency Medicine **17**(2), 551–558 (2022)

[47] Grant, L., Diagne, M., Aroutiunian, R., Hopkins, D., Bai, T., Kondrup, F., Clark, G.: Machine learning outperforms the canadian triage and acuity scale (ctas) in predicting need for early critical care. Canadian Journal of Emergency Medicine **27**(1), 43–52 (2025)

[48] Holodinsky, J.K., Sabir, L., McRae, A.D., Mason, S.M.: Busier emergency departments: same triage systems—can ai solve the "next-to-be-seen" problem? Canadian Journal of Emergency Medicine **27**(1), 5–6 (2025)

[49] Franc, J.M., Cheng, L., Hart, A., Hata, R., Hertelendy, A.: Repeatability, reproducibility, and diagnostic accuracy of a commercial large language model (chatgpt) to perform emergency department triage using the canadian triage and acuity scale. Canadian Journal of Emergency Medicine **26**(1), 40–46 (2024)

[50] Doudareva, E., Carter, M.: Using discrete event simulation to improve performance at two canadian emergency departments. In: 2021 Winter Simulation Conference (WSC), pp. 1–12 (2021). IEEE

[51] MOSIMTEC, L.: Emergency Department (ED) Simulation Model. https://mosimtec.com/emergency-department-ed-simulation-model/. Accessed: 2025-07-27 (2018)

[52] Almansoori, M., Kumar, K., Cholakkal, H.: Self-evolving multi-agent simulations for realistic clinical interactions. arXiv preprint arXiv:2503.22678 (2025)

[53] Hall, J.N., McCarron, J., Toarta, C., McLeod, S.L., Committee, C.N.A., Committee, N.T.: Canadian emergency department triage and acuity scale (ctas) guidelines 2025. Canadian Journal of Emergency Medicine, 1–4 (2025)

[54] Salehi, L., Phalpher, P., Valani, R., Meaney, C., Amin, Q., Ferrari, K., Mercuri, M.: Emergency department boarding: a descriptive analysis and measurement of impact on outcomes. Canadian Journal of Emergency Medicine **20**(6), 929–937 (2018)

[55] Cildoz, M., Mallor, F., Ibarra, A.: Analysing the ed patient flow management problem by using accumulating priority queues and simulation-based optimization. In: 2018 Winter Simulation Conference (WSC), pp. 2107–2118 (2018). IEEE

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- edsimfeb24supplementary.pdf