A KernelCentric AI Architecture for Emergency Triage: From Healthcare Knowledge Management to Simulated Dispatch Scenarios*

Demetrio Alessandro Trunfio[1], Ferdinando Potenza[1], and Anna Mezzapesa[1].

Abstract— **Healthcare is one of the domains where organizational complexity, cognitive workload, and regulatory constraints tend to slow down the adoption of new digital technologies. This work introduces a Kernel AI as a dorsal layer that orchestrates healthcare knowledge management, AI agents, and clinical workflows. The proposed architecture is instantiated on the endtoend emergency triage process, from the initial call to the dispatch center to patient arrival in the emergency department, with the goal of supporting patients, families and frontline staff (dispatchers, nurses, physicians, ambulance crews) and reducing errors related to fatigue and cognitive overload. Building on this conceptual model, the paper outlines a simulated emergency dispatch and territorial environment as the basis for future experimental evaluation of Kernel AI in realistic multicall, multiunit scenarios involving ground ambulances and helicopter emergency medical services (HEMS). This workin progress is part of a masters thesis ongoing and focuses on the conceptual design and triage workflow scenario, leaving implementation and quantitative validation to future work.**

I. Introduction

Healthcare organizations operate under high cognitive workload, tight regulatory constraints and complex, multiactor workflows, which makes them significantly slower than other sectors in absorbing the accelerating pace of digital innovation [1]. At the same time, the rise of Generative AI and Large Language Models (LLMs) is transforming how knowledge can be accessed, synthesized and delivered to clinicians and managers, opening new opportunities but also raising concerns about safety, reliability and integration into existing clinical pathways [2], [3].

In this context, knowledge management (KM) in healthcare has evolved from a focus on information systems and explicit documentation toward an ecosystem perspective, where digital technologies and AI systems become active nodes in the organizational knowledge infrastructure [1]. However, current AI deployments in healthcare often remain siloed tools or isolated decision support modules, loosely coupled with clinical workflows and with limited consideration of cognitive constraints and workload at the frontline [2], [3]. This gap is particularly evident in emergency care, where dispatch centers and emergency departments must process high volumes of heterogeneous information under time pressure, with significant risk of error, fatigue and burnout for staff [5], [6].

*This work was not supported by any organization.

[1]Department of Engineering, LUM Giuseppe Degennaro University, 70010 Casamassima (BA), Italy. {d.trunfio, f.potenza, a.mezzapesa}@studenti.lum.it

This workinprogress paper proposes a Kernel AI as a dorsal layer that orchestrates healthcare knowledge management, AI agents and clinical workflows in a unified architecture [4], [7], [8]. The proposed Kernel AI is instantiated on the endtoend emergency triage process, from the initial call to the dispatch center to patient arrival in the emergency department, with the goal of supporting patients, families and frontline staff (dispatchers, nurses, physicians, emergency medical services (EMS) crews, including ground ambulances and helicopter emergency medical service (HEMS) units) in choosing appropriate care pathways and reducing errors related to fatigue and cognitive overload [5], [6]. Building on this conceptual model, the paper outlines an experimental roadmap based on a simulated emergency dispatch and territorial environment, designed to evaluate the impact of the Kernel AI in realistic multicall, multivehicle scenarios [4], [7].

Existing work on healthcare knowledge management, Generative AI, multiagent systems and clinical BPM has not yet introduced a kernelcentric architecture that treats the orchestration layer itself as a firstclass Kernel AI engine for coordinating agents, workflows and knowledge resources in emergency care; this workinprogress paper represents a first step in that direction, using emergency triage as a testbed for the proposed research question.

The contributions of this paper are threefold. First, it defines a kernelcentric architecture that positions AI as an orchestration layer for knowledge, agents and workflows in healthcare emergency settings [4], [7], [8]. Second, it tailors this architecture to the emergency triage process by mapping the roles of different agents and knowledge resources along the pathway from telephone triage to inhospital triage [5], [6]. Third, it sketches a simulationbased evaluation strategy grounded in realistic dispatch and territorial scenarios, which will be used in future work to quantitatively assess the impact of the Kernel AI on process performance and cognitive load [4], [7].

II. Background and Motivation

A. Knowledge Management and GenAI in Healthcare

Healthcare systems have progressively moved from paper-based records and fragmented information systems toward digitally mediated knowledge infrastructures, where clinical, organizational and relational knowledge coexist in highly regulated environments [1]. Within this evolution, knowledge management (KM) has shifted from a narrow focus on repositories and document management to a broader concern with how digital tools support evidencebased practice, reduce inequalities in access to knowledge and enable learning health systems [1]. The emergence of Generative AI and LLMs further amplifies this shift by providing new capabilities for summarization, retrievalaugmented reasoning and contextaware recommendation, but also introduces new risks related to hallucinations, bias, lack of transparency and difficulties in translating promising models into safe, sustainable clinical implementations [2], [3].

B. Emergency Triage and Workflow Complexity

Emergency departments and dispatch centers represent some of the most complex nodes of the healthcare system, characterized by unpredictable demand, timecritical decisions and multiprofessional coordination [5]. Modern triage systems structure the initial assessment into standardized acuity levels and aim to optimize resource allocation, yet empirical studies show that telephone triage and subsequent emergency department attendance remain prone to variability and potential mismatches between perceived urgency and actual clinical need [5], [6]. This setting creates a challenging combination of high cognitive workload, fragmented information flows between prehospital services and hospitals, and vulnerability to errors driven by fatigue and overload, which makes it an ideal stress test for any AIenhanced knowledge and workflow architecture [5], [6].

C. Toward KernelCentric Orchestration Architectures

In parallel, research on multiagent and hybrid intelligence systems in healthcare has begun to explore how autonomous or semiautonomous agents can support complex planning and coordination tasks, for example in infrastructure planning and resource allocation under constraints [4]. At the process level, clinical pathways and care processes have increasingly been modelled using workflow and business process management formalisms, such as BPMN, to enable execution, monitoring and continuous improvement of clinical workflows [7]. Beyond individual tools, these strands of work point toward the need for kernellike orchestration architectures, where a stable core coordinates processes, agents and knowledge resources while allowing peripheral components to evolve over time [4], [7], [8]. The Kernel AI proposed in this paper builds on these ideas and adapts them to the specific context of emergency triage, with the additional objective of explicitly accounting for cognitive load and knowledge management requirements in the design of the orchestration layer. Fig. 1 provides a highlevel view of the proposed Kernel AI architecture, highlighting the interaction between healthcare knowledge management, AI agents and clinical workflows in the context of emergency triage [4], [7], [8].

III. Kernel AI Architecture for Emergency Triage

The proposed Kernel AI architecture is organized into four main layers that connect emergency triage workflows with knowledge resources and AI agents (Fig. 1) [4], [7], [8]. At the bottom, an infrastructure layer provides connectivity to existing healthcare information systems, emergency dispatch platforms and data sources, including electronic health records, triage logs and operational metrics [1], [7]. Above this, the kernel orchestration layer acts as a stable core responsible for managing workflows, coordinating agents, logging decisions and enforcing governance and compliance rules [4], [8].

On top of the kernel, an intelligent services layer hosts a configurable ecosystem of AI agents and knowledge services, such as retrievalaugmented LLMs, triage support agents, dispatch recommendation agents and data quality monitors [2], [4]. These agents interact with a healthcare knowledge layer that integrates clinical guidelines, triage protocols, historical cases and local organizational rules, curated as part of the broader knowledge management effort [1], [7]. At the interaction layer, the outputs of the Kernel AI are exposed to human users through rolespecific interfaces for dispatchers, nurses, physicians and ambulance crews, designed to support decisionmaking without increasing cognitive load [5], [6].

A. Kernel orchestration layer

The kernel orchestration layer provides a stable core that manages the lifecycle of emergency triage workflows, from the initial telephone contact to inhospital triage and handover [5], [7]. It coordinates the execution of process models, such as those expressed in business process modelling notations for clinical pathways, and routes events and messages between prehospital and inhospital components [7], [8]. The kernel is also responsible for logging decisions, maintaining an auditable trail of recommendations and actions, and enforcing organizational and regulatory constraints that govern how AIgenerated suggestions can be used in clinical practice [2], [3]. In addition, the Kernel AI acts as a guardrail layer that is hierarchically stronger than individual agents, enforcing a set of hard governance rules that must hold across all recommendations. Conflicts between agents (for example, a telephone triage agent suggesting low urgency while a dispatch recommendation agent proposes HEMS activation) are handled through kernellevel arbitration policies that combine rulebased checks, priority schemes and consistency constraints derived from clinical protocols and organizational policies. When clinically critical disagreements are detected, the case is escalated to human operators through explicit alerts and explainable summaries of the conflicting recommendations and the rationale for the safe option proposed by the kernel. By separating orchestration and governance concerns from the implementation of individual agents and services, the kernel layer supports incremental evolution of the overall architecture without disrupting frontline workflows [4], [8].

B. Agent ecosystem and intelligent services

The intelligent services layer hosts an ecosystem of AI agents that encapsulate specific decisionsupport and coordination functions along the emergency triage pathway [2], [4]. Examples include agents that assist telephone triage by suggesting questions and preliminary risk levels, dispatch

Figure 1: KERNEL AI ARCHITECTURE FOR EMERGENCY TRIAGE

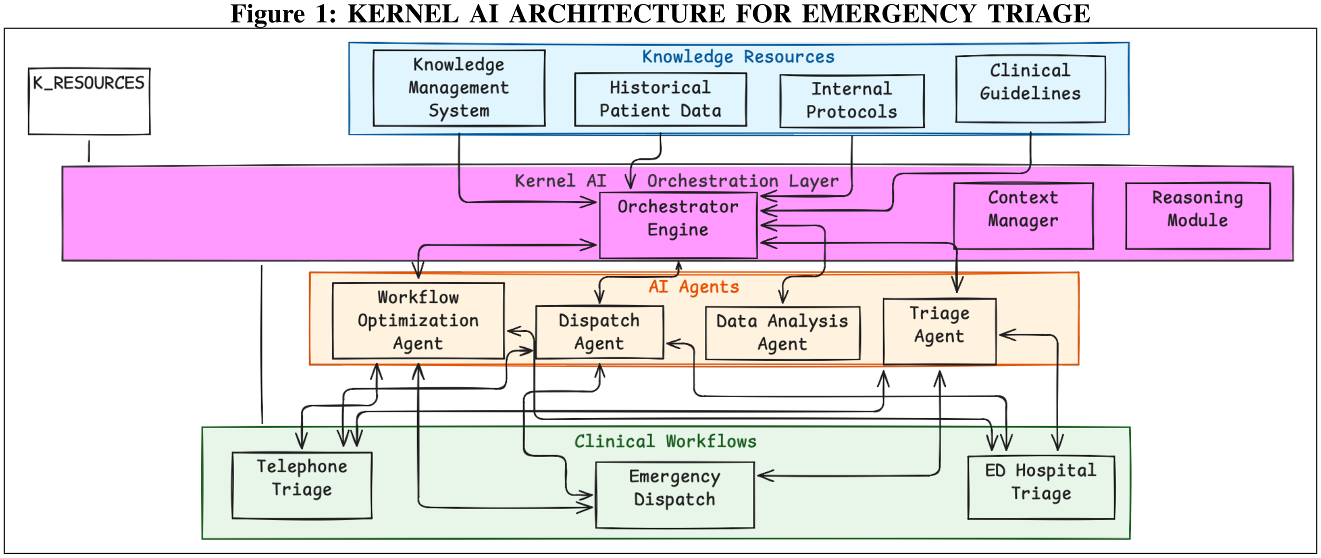

Fig. 1. Highlevel Kernel AI architecture for emergency triage, showing the orchestration layer between knowledge resources, AI agents and clinical workflows.

recommendation agents that propose emergency medical services (EMS) unit types, including ground ambulances and HEMS units, and appropriate destination facilities, and triage support agents that help emergency department staff assign acuity levels and prioritize patients [5], [6]. These agents can use retrievalaugmented LLMs to ground their outputs in uptodate clinical guidelines and local protocols, while the kernel orchestrates their interactions and ensures that their suggestions remain aligned with the overall workflow and governance rules [2], [3], [7].

C. Healthcare knowledge layer and interaction layer

The healthcare knowledge layer aggregates and structures the knowledge needed for emergency triage, including evidencebased clinical guidelines, triage scales, organizational policies and historical data from past cases [1], [5]. This layer provides a domainspecific context for AI agents through indexing, retrieval and representation mechanisms that allow LLMbased services to access relevant knowledge while minimizing unsupported generations [2], [3]. At the interaction layer, the outputs of the Kernel AI are delivered through user interfaces tailored to the roles and information needs of dispatchers, nurses, physicians and ambulance crews, emphasizing concise, explainable recommendations that can be accepted, modified or overridden by human operators [5], [6]. This design aims to support decisionmaking and reduce cognitive load without replacing professional judgement.

IV. Illustrative Triage Workflow and Simulation Roadmap

A. Emergency triage workflow from call to emergency department

The illustrative scenario covers the endtoend emergency triage workflow, from the initial telephone call to the dispatch center to patient arrival in the emergency department (ED) [5], [6]. In the first phase, calltakers collect key information about symptoms, context and patient characteristics, often under stress and with incomplete descriptions, and assign a preliminary urgency level using structured triage protocols [5]. In the second phase, dispatch decisions are made regarding the type and number of emergency medical services (EMS) units to be mobilized, including ground ambulances and helicopter emergency medical services (HEMS) units when required by distance, geography or severity [6]. During the prehospital phase, EMS crews monitor the patient, communicate updates and may transmit vital signs or prealerts to the receiving ED. Finally, on arrival at the hospital, ED triage nurses perform an inperson assessment and assign a formal acuity level, integrating prehospital information with clinical examination [5].

Fig. 2 illustrates a simplified sequence diagram of the interactions between the dispatch center, the Kernel AI, emergency medical services (EMS) units and the emergency department (ED) along the triage pathway.

Within this workflow, the Kernel AI can support different actors through rolespecific agents coordinated by the orchestration layer. Telephone triage agents assist calltakers by suggesting clarifying questions and preliminary risk stratification; dispatch recommendation agents propose EMS unit types and destination facilities consistent with protocols and resource availability; triage support agents help ED staff to prioritize patients based on integrated prehospital and inhospital data [2], [4], [7]. Across all phases, the kernel maintains an endtoend view of the case, logs decisions and provides a shared knowledge context that aims to reduce variability and errors related to fatigue and cognitive overload [1], [5].

Figure 2: EMERGENCY TRIAGE SEQUENCE DIAGRAM

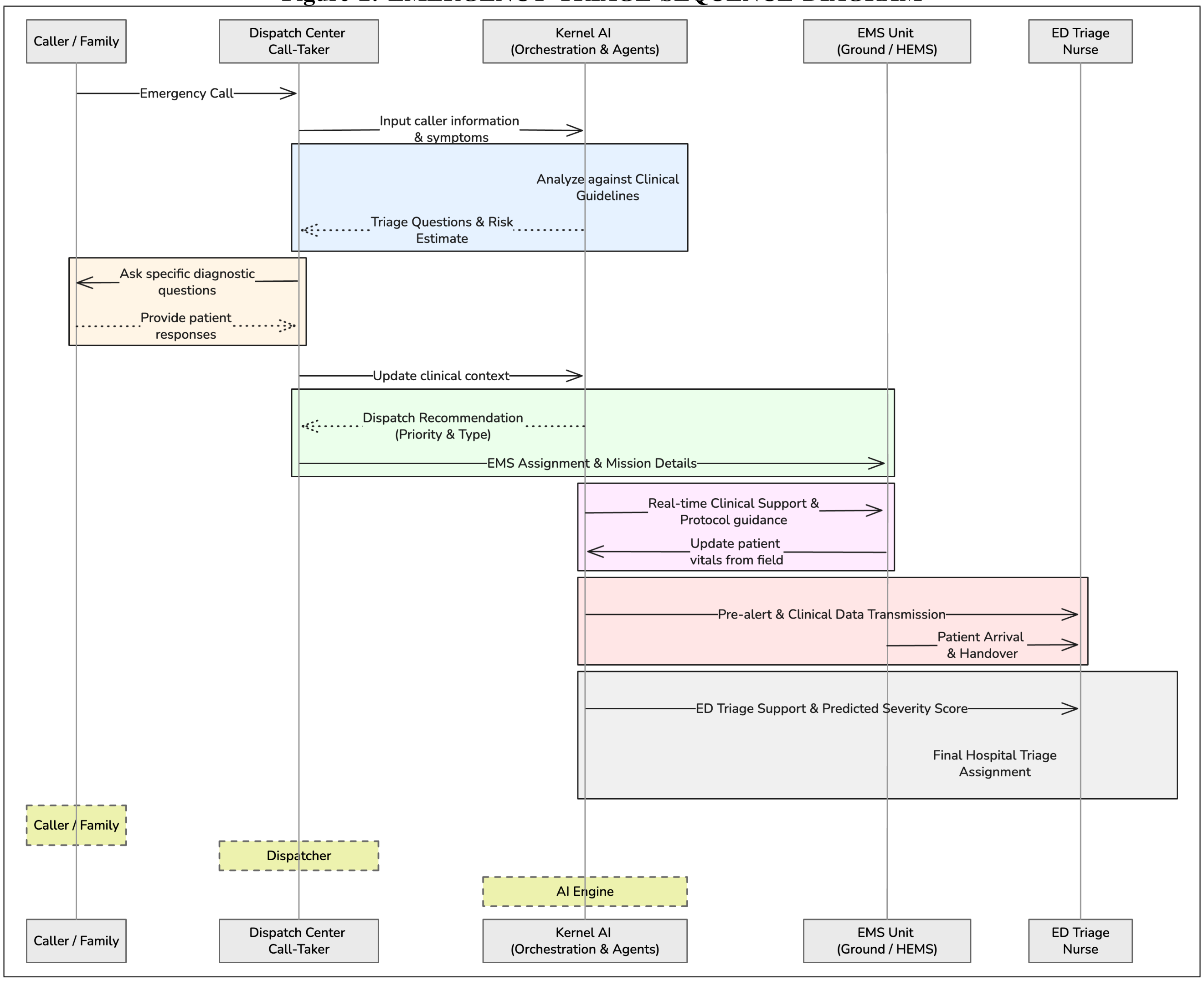

Fig. 2. Simplified sequence diagram of the emergency triage process from call to ED arrival, including interactions with the Kernel AI.

B. Simulationbased evaluation roadmap

To evaluate the proposed architecture without interfering with realworld emergency operations, the work envisions a simulationbased roadmap grounded in realistic dispatch and territorial scenarios [4], [7]. The idea is to configure a simulated emergency dispatch center and a model of a selected area of the Bari province, with a given density of calls, EMS bases, ground ambulances and potential HEMS activation patterns. This environment will be used to generate multicall, multiunit scenarios in which baseline workflows (without the Kernel AI) are compared to workflows augmented by the kernelcentric architecture. Key outcome measures will include process indicators such as time from call to EMS dispatch and time to ED arrival, as well as quality indicators related to triage accuracy and alignment with clinical protocols [5], [6].

In this roadmap, the simulation environment provides controllable yet realistic conditions to test how different configurations of agents and knowledge services within the Kernel AI affect systemlevel performance. The workinprogress status of this study implies that the present paper focuses on the conceptual design of the architecture and the definition of the triage workflow and simulation plan, while future work will implement the prototype, calibrate the simulation and conduct quantitative experiments based on the proposed metrics [4], [7], [8].

Beyond textual information, the envisioned Kernel AI also targets richer, multimodal inputs in future iterations. In particular, the agent ecosystem is expected to exploit speech-based features (e.g., prosody, background noise, overlapping voices) and, when clinically and ethically appropriate, optional video streams to better contextualize emergency calls and anticipate the likely scenario on scene [2], [5]. In this perspective, the Kernel AI would not only assist calltakers in structuring questions, but could also support proactive preparation of emergency medical services (EMS) crews by providing synthesized, realtime summaries and visual cues derived from callerprovided audio and video, subject to strict governance and privacy constraints [3], [6]. The working hypothesis is that such kernelcentric, multimodal orchestration can reduce timetointervention, improve triage consistency and enhance readiness of EMS teams in high-stress situations, a claim that will need to be quantitatively tested in the simulation environment described above [4], [7]. In line with the available dispatch centre data from the Bari province, the simulation will first establish baseline values for timetodispatch, timetoED and triage alignment with protocols, and then test kernelcentric configurations against working hypotheses of moderate but clinically meaningful improvements (for example, singledigit percentage gains in timeliness and consistency) rather than fixed numerical targets.

V. Conclusion and Future Work

This work-in-progress paper has introduced a kernel-centric AI architecture for emergency triage, positioning a Kernel AI as a dorsal layer that orchestrates healthcare knowledge management, AI agents and clinical workflows from the initial emergency call to emergency department arrival [1], [4], [7], [8]. By grounding the proposal in the specific context of telephone triage, emergency medical services (EMS) dispatch and in-hospital triage, the paper outlines how role-specific agents and a shared knowledge layer can support frontline staff while addressing cognitive workload and variability in decision-making [5], [6]. Building on this basis, the work contributes by (i) introducing the Kernel AI as a first-class orchestration engine for coordinating agents, workflows and knowledge resources in emergency care, (ii) instantiating this kernel-centric architecture on the end-to-end emergency triage pathway, and (iii) defining a simulation-based evaluation roadmap in realistic multi-call, multi-unit scenarios as a first step toward quantitative assessment of its impact on process performance, triage quality and cognitive load.

The paper also sketches a simulation-based evaluation roadmap, based on a realistic model of a dispatch center and territorial configuration in the Bari province, to compare baseline workflows with those augmented by the Kernel AI in multi-call, multi-unit scenarios [4], [7]. The working hypotheses are that kernel-centric orchestration, potentially extended to multimodal inputs such as speech and video, can reduce time-to-intervention, improve triage consistency and enhance EMS team readiness under high workload, in line with prior evidence on structured triage and decision support in emergency care [2], [5], [6]. With respect to these future multimodal extensions, the use of audio and optional video streams in emergency calls is conceived within a layered governance framework that combines the right to life and access to emergency care with GDPR-compliant legal bases for emergency processing of health data and the high-risk AI requirements of the EU AI Act, distinguishing between strictly necessary real-time processing during life-threatening situations and consent-based storage or secondary use once the acute phase has passed. In line with the human-oversight requirements for high-risk AI systems, the Kernel AI is therefore conceived not as an autonomous decision-maker but as a guard-rail orchestration layer that keeps clinicians in the loop, with all critical recommendations remaining subject to human review, override and accountability, an aspect that will be investigated more deeply in the next phases of the research and prototyping work [3]. Beyond this work-in-progress paper, the research roadmap foresees three main steps: completing the detailed Kernel AI design for emergency triage, implementing a prototype integrated with the simulated dispatch and territorial environment, and conducting quantitative experiments on process performance, triage quality and indicators of cognitive load within the current academic year [2], [3], [7], [8].

Acnowledgement

The research has been shaped by earlier experiences in academic writing, including the acceptance of several extended abstracts and the submission of a coauthored paper with fellow students and faculty members to an IEEE conference, which provided a first exposure to the expectations and rigor of the control, decision and information technologies community. The author gratefully acknowledges the guidance and feedback of his supervisors and the teaching staff of the Department of Engineering at LUM University, who provided the conceptual and methodological foundations that underpin the present work.

References

[1] W. Vesperi, M. Ventura, C. L. Cristofaro, and A. M. Melina, *Digital evolution and emerging inequalities in healthcare: A scoping review through the lens of knowledge management*, *BMC Health Services Research*, vol. 25, no. 1, p. 1371, 2025.

[2] S. Reddy, *Generative AI in healthcare: An implementation science informed translational path on application, integration and governance*, *Implementation Science*, vol. 19, no. 27, pp. 112, 2024.

[3] J. C. L. Ong *et al.*, *Regulatory science innovation for generative AI and Large Language Models in health and medicine: A global call for action*, *arXiv preprint* arXiv:2502.07794, 2025.

[4] Y. Zhang *et al.*, *MedBuild AI: An agent-based hybrid intelligence framework for reshaping agency in healthcare infrastructure planning through generative design for medical architecture*, *arXiv preprint* arXiv:2511.11587, 2025.

[5] M. Christ, B. Grossmann, A. Winter, and M. Bingisser, *Modern triage in the emergency department*, *Deutsches Ärzteblatt International*, vol. 107, no. 50, pp. 892898, 2010.

[6] A. Gibson, S. Randall, S. R. Gould, G. K. C. Kelly, P. Y. Robinson, and D. S. Eagar, *Emergency department attendance after telephone triage: A population-based data linkage study*, *Health Services Research*, vol. 53, no. 2, pp. 11371162, 2018.

[7] *N. Rasooli, F. Jolai, M. M. Sepehri, and A. Tehranian, BPM application in clinical process improvement: A womens hospital case study*, *Business Process Management Journal*, vol. 30, no. 3, pp. 9861011, 2024.

[8] Z. Hou and Z. Yu, *Research of the workflow management system based on microkernel*, *Journal of Theoretical and Applied Information Technology*, vol. 47, no. 1, pp. 324332, 2013.