Simulation of an AI–assisted workflow for emergency department triage: a comparative efficiency study

Simulación de un flujo de trabajo asistido por inteligencia artificial en el triaje de un servicio de urgencias: estudio comparativo de eficiencia

Alejandro Martín-Quirós[1], Milagros Jaen Cañadas[2], Paula Miralles Atencia[3], Víctor Gómez Carrillo[1], Laura Ávila Huertas[4], Francisco Miralles Linares[1]

Emergency departments (EDs) face growing care pressure due to increasing patient volumes, limited resources, and the need to provide rapid and accurate attention. Diagnostic delays are a key factor in ED overcrowding and are associated with increased morbidity and mortality.[1,2] Traditionally, diagnostic tests are ordered only after the initial medical evaluation, which can generate delays—especially in overcrowded settings.[3] Triage allows clinicians to stratify patients and assign priority, but it does not activate diagnostic test ordering. With advances in artificial intelligence (AI), particularly large language models (LLMs) such as ChatGPT (OpenAI), there is increasing interest in their application as clinical support tools.[4,5] These models can process free-text input and generate relevant recommendations,[6] although their use in real clinical environments remains limited.

This study assessed the ability of ChatGPT (GPT-4, March 2024 version) to suggest diagnostic tests based on triage notes, and whether these suggestions matched real medical decisions. It also simulated the operational impact of an alternative workflow in which diagnostic tests are ordered before medical evaluation (Circuit 2) vs the traditional workflow (physician before test, Circuit 1). The analysis explored how this intervention could reduce care times according to hospital type (public or private) and saturation level (low, medium, high).

We conducted a retrospective observational study using triage data from an academic private tertiary referral center in Spain. The protocol was approved by *Hospital Costa del Sol* (Málaga, Spain) Ethics Committee. A total of 100 consecutive adult patients who presented to the ED with a free-text triage narrative were included. Minors, patients without medical evaluation, and cases with incomplete data were excluded. A total of 84 patients were analyzed. For each triage note, ChatGPT was asked to suggest the most appropriate diagnostic test, restricted to common ED tests (blood or urine analysis, ECG, radiography, ultrasound, CT). Actual medical decisions were drawn from health records, and 3 independent physicians determined the main diagnostic test ordered in each case. Concordance was established when the model's recommendation matched the medical test. In discordant cases, a second round was performed by adding the full medical anamnesis to the model prompt to determine whether concordance improved.

Two diagnostic pathways were simulated for each patient: Circuit 1 (triage –> physician –> test) and Circuit 2 (triage –> test –> physician). Estimated times for each step (physician evaluation, diagnostic tests) were obtained from the literature and internal reports (Tables 1 and 2)[7-13] and were adapted to each hospital type and level of saturation (Table 3). Time differences between circuits were calculated, and paired t-tests were used for comparison, along with ANOVA and linear regression to analyze predictors of time savings (Supplementary data: https://revistaemergencias.org/extra/5061.pdf).

Most of the 84 patients analyzed were categorized as level IV under the Manchester Triage System (58.3%), followed by levels III, II, and V. In 58 cases (69.0%), the diagnostic test suggested by ChatGPT based on triage coincided with the medical order, yielding a kappa index of 0.41 with a 95% confidence interval (95% CI) of 0.37–0.44.

Table 1. Estimated waiting times from triage to medical evaluation

Hospital type	Saturation level	Estimated waiting time (minutes)
Public	Low	20–30
	Medium	50
	Very high	> 90 (may exceed 2 hours)
Private	Low	10–15
	Medium	20
	Very high	30

Estimated waiting times from triage to medical evaluation in Spanish hospital emergency departments, stratified by hospital type and saturation level. Times are based on official public reports and internal hospital audits.[7-14]

Table 2. Estimated times for diagnostic testing

	Hospital type	Estimated time (minutes)
Plain radiography	Public	30 (up to 60 if highly saturated)
	Private	15–30
Blood tests (CBC, coagulation, chemistry)	Public	35
	Private	30–45
Electrocardiogram	Public	< 10
	Private	5–10
Urine dipstick	Public	5
	Private	5
Vital signs	Public	2–5
	Private	2–5
Urgent CT scan	Public	30 (up to +60 if highly saturated)
	Private	20–30
Ultrasound	Public	60
	Private	30–45

Estimated times for performing common diagnostic tests in Spanish hospital emergency departments according to hospital type.

Table 3. Simulated times according to hospital type, saturation level, and circuit

	Saturation level degree of saturation	Circuit	Mean (SD) (minutes)	Median (Min–Max) (minutes)
Public hospitals	Low saturation	Circuit 1	48.0 (24.0)	35.0 (25.0-85.0)
		Circuit 2	35.7 (12.3)	25.0 (25.0-60.0)
	Medium saturation	Circuit 1	73.0 (24.0)	60.0 (50.0-110.0)
		Circuit 2	50.9 (2.9)	50.0 (50.0-60.0)
	High saturation	Circuit 1	143.0 (24.0)	130.0 (120.0-180.0)
		Circuit 2	120.0 (0.0)	120.0. (120.0-120.0)
Private hospitals	Low saturation	Circuit 1	25.0 (13.2)	20.0 (12.5-50.0)
		Circuit 2	18.7 (7.8)	12.5 (12.5-37.5)
	Medium saturation	Circuit 1	32.5 (13.2)	27.5 (20.0-57.5)
		Circuit 2	22.7 (5.0)	20.0 (20.0-37.5)
	High saturation	Circuit 1	42.5 (13.2)	37.5 (30.0-67.5)
		Circuit 2	30.7 (2.2)	30.0 (30.0-37.5)

Estimated total times (in minutes) for Circuits 1 and 2, stratified by hospital type (public vs private) and emergency department saturation level (low, medium, high).
SD: standard deviation.

Among the 26 discordant cases, after adding the anamnesis, the model modified its recommendation in 14 cases (53.8%); 9 of these aligned with the medical test. Thus, overall concordance increased to 67 of 84 cases (79.8%), resulting in a Cohen's kappa of 0.44 (95% CI, 0.41–0.47). Table 3 illustrates the triage level according to the Manchester system, the diagnostic test requested by the AI model, and the degree of concordance with the physician's request—both at the initial triage stage and after including anamnesis in the discordant cases.

Simulation demonstrated that Circuit 2 was systematically faster—or at least equal—vs Circuit 1 in all scenarios. The average total time was 53.6 minutes in Circuit 1 vs 39.2 minutes in Circuit 2, with mean savings of 14.4 minutes per patient (SD, 8.1). The greatest savings were observed in highly saturated public hospitals (> 22 minutes), and the smallest in low-demand private hospitals (6.3 minutes). Nevertheless, all scenarios showed statistically significant differences ($P < .001$). Multivariate regression confirmed that both hospital type and saturation level were significant predictors of time saved. Public hospitals were associated with an additional savings of 9.8 minutes vs private hospitals ($\beta = 9.8$; $P < .001$). For every additional saturation level, an average of 4.0 minutes was saved ($\beta = 4.0$; $P < .001$). The model explained 12.1% of the variance in time saved ($R^2 = 0.121$).

These findings suggest that incorporating AI-generated diagnostic recommendations at triage, along with workflows that enable early activation of tests, can significantly reduce care times. The greatest benefits occur in high-pressure settings, such as saturated public hospitals. Unlike former studies that used structured tools or predefined protocols, this study relied on real free-text inputs, better representing the complexity of the clinical environment.[14,15] Additionally, the results showed that symptoms recorded at triage often do not reflect the true clinical problem, generating discordance. When the LLM re-evaluated cases with the added clinical anamnesis, concordance improved. However, no trend was observed toward requesting more or fewer tests, nor were specific differences found related to reason for consultation or triage level. This was likely related to the sample size in both conditions.

Cognitive plausibility of LLMs as early clinical support tools is reinforced by the high AI-physician concordance (nearly 80%), surpassing studies limited to performance on medical examinations.[6] However, limitations must be considered, such as the use of data from a single private hospital and the simulated desing of the time analysis, which may differ from real-world practice. In conclusion, this pilot study shows that integrating diagnostic AI into triage—along with workflows that anticipate test ordering—can meaningfully enhance ED operational efficiency. The effects are especially pronounced in public hospitals and high-saturation contexts. Future prospective studies in real clinical environments will be necessary to validate these findings and explore their impact on patient-centered clinical outcomes.

References

1 Pines JM, Localio AR, Hollander JE, Baxt WG, Lee H, Phillips C, et al. The impact of emergency department crowding measures on time to antibiotics for patients with community-acquired pneumonia. Ann Emerg Med. 2007;50:510-6.
2 Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang LJ, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med. 2013;61:605-11.e6.
3 Schriger DL. Does the presence of diagnostic testing reduce ED length of stay? Ann Emerg Med. 2000;35:105-11.
4 OpenAI. GPT-4 Technical Report. 2023. (Consultado 29 Agosto 2025). Disponible en: https://openai.com/research/gpt-4
5 Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large Language Models Encode Clinical Knowledge. NPJ Digit Med. 2023;6:1-9.
6 Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education. PLOS Digit Health. 2023;2:e0000198.

7 Informe anual del Sistema Nacional de Salud, tiempos de espera en urgencias hospitalarias. Madrid: Ministerio de Sanidad, Consumo y Bienestar Social; 2022.
8 Fundación Instituto para el Desarrollo e Integración de la Sanidad (IDIS). Informe "Sanidad privada, aportando valor: Análisis de la situación 2023". Madrid: Fundación IDIS; 2023.
9 Sociedad Española de Medicina de Urgencias y Emergencias (SEMES). Recomendaciones para tiempos de actuación clínica en urgencias. Madrid: SEMES; 2021.
10 Sociedad Española de Radiología Médica (SERAM). Guía de práctica clínica en radiología urgente. Madrid: SERAM; 2022.
11 Grupo de trabajo de Laboratorio Clínico de SEMES. Manual de laboratorio en urgencias. Madrid: SEMES; 2021.
12 Organización de Consumidores y Usuarios (OCU). Encuesta nacional sobre tiempos de espera en urgencias públicas. Madrid: OCU; 2025.
13 Fundación Instituto para el Desarrollo e Integración de la Sanidad (IDIS). Estudio RESA 2023/2024: resultados en salud del sector privado. Madrid: Fundación IDIS; 2024.
14 Rao A. Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making. Radiology. 2023;308:e230236.
15 Hinson JS, Martinez DA, Cabral S, George K, Whalen M, Hansoti B, et al. Triage Performance in Emergency Medicine: A Systematic Review. Ann Emerg Med. 2019;74:140-52.

Author Affiliations: [1]Departamento de Medicina Interna, Hospital Vithas Xanit Internacional, Benalmádena, Málaga, Spain. [2]Departamento de Urgencias, Hospital Vithas Xanit Internacional, Benalmádena, Málaga, Spain. [3]Facultad de Medicina, Universidad de Navarra, Pamplona, Spain. [4]Facultad de Medicina, Universidad de Granada, Granada, Spain.
Email: a.martinquiros@gmail.com
Conflict of Interest Disclosures: None declared.
Author Contributions, Funding, and Ethical Responsibilities: The author has confirmed their authorship, the non-existence of external funding, and the maintenance of confidentiality and respect for patients' rights in the author's responsibilities document, publication agreement, and assignment of rights to EMERGENCIAS. The study was approved by the Ethics Committee of the Costa del Sol Hospital.
Article not commissioned by the Editorial Board and with external peer review.
Editor in Charge: Agustín Julián-Jiménez.
Corresponding Author: Alejandro Martín-Quirós. Departamento de Medicina Interna. Hospital Vithas Xanit Internacional. Av. de los Argonautas, s/n. 29630 Benalmádena, Málaga, Spain.