Journal of Medical Imaging and Interventional Radiology (2026) 13:25
https://doi.org/10.1007/s44326-026-00114-0

ORIGINAL RESEARCH

AI as a clinical decision support system: accuracy of an open GPT-model in emergency practice

Jérémie Arthaud[1] · Harriet C. Thoeny[1,2,3] · Vincent Grek[3] · Lucien Widmer[1,2]

Received: 2 April 2026 / Accepted: 13 July 2026
© The Author(s) 2026

Abstract

Background Up to 30% of imaging examinations are deemed inappropriate, leading to unnecessary radiation exposure, emergency department (ED) overcrowding, and rising healthcare costs. While clinical decision support systems (CDSSs) such as ESR iGuide aim to improve imaging appropriateness, their clinical impact and user adoption remain limited. Large language models (LLMs) may offer a faster, more accessible alternative for radiologic decision-making.

Purpose To evaluate the performance of OpenAccessGPT in reducing inappropriate radiological examinations requested from an ED by comparing its recommendations with ESR iGuide and assessing appropriateness and time efficiency.

Materials and methods This retrospective, single-center, semi-quantitative cohort study was conducted at a tertiary hospital in Switzerland. A total of 201 consecutive ED radiology requests (March 2023–April 2024) met inclusion criteria (50 CT, 50 US, 50 X-rays, 51 MRI). Four hypotheses were tested: (1) OpenAccessGPT's ability to reduce imaging requests; (2) concordance with ESR iGuide; (3) time efficiency; and (4) the rate of inappropriate examinations in clinical practice when confronted to ESR iGuide. Statistical analyses included proportion tests and paired t-tests ($p < .05$).

Results Mean patient age was 55.3 years (SD = 20.9); 45% were female. Eleven cases were excluded due to unavailable ESR iGuide scenarios. OpenAccessGPT advised against imaging in 6% of cases (12/201), with 33.3% agreement with ESR iGuide (5/12). Overall concordance was 78.9% (150/190), below the 90% threshold ($\chi^2 = 25.79$, $p < .001$). GPT was significantly faster (21.6 s vs. 101 s; $\Delta = 79.6$ s; t[189] = 12.8; $p < .001$). Institutional practice aligned with ESR iGuide in 82.6% of cases ($\chi^2 = 11.46$, $p < .001$).

Conclusion OpenAccessGPT showed only limited ability to reduce inappropriate imaging. Despite substantial faster response times, its clinical reliability remained insufficient for standalone use.

Keywords Clinical decision support systems · Large language models · Generative pre-trained transformer · ESR iGuide · Emergency department

✉ Lucien Widmer
lucien.widmer@h-fr.ch

Jérémie Arthaud
jeremie.arthaud@h-fr.ch

Harriet C. Thoeny
harriet.thoeny@h-fr.ch

Vincent Grek
vincent.grek@gmail.com

[1] Department of Radiology, University Training and Research Hospital of Fribourg (HFR), Chemin des Pensionnats 2-6, 1700 Fribourg, Switzerland

[2] Department of Medical and Surgical Specialties, Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland

[3] Department of Urology, Inselspital, University Hospital Bern, Bern, Switzerland

Abbreviations

ACR	American College of Radiology
ACR AC	American College of Radiology Appropriateness Criteria
AI	Artificial intelligence
CDSSs	Clinical decision support systems
ED	Emergency department
ENT	Ear, nose, and throat
ESR	European Society of Radiology
LLMs	Large language models
MSK	Musculoskeletal

Published online: 23 July 2026

Inroductiont

Clinical decision-making in emergency departments is a complex, multifactorial process that integrates rapid diagnostic reasoning with evidence-based practice. This cognitive framework remains susceptible to systematic biases, contributing to persistently high rates of inappropriate imaging requests [1]. Contemporary studies indicate that 20–50% of radiological examinations ordered in emergency settings fail to meet established appropriateness criteria, raising concerns over patient radiation exposure, optimal resource allocation, escalating healthcare costs and greenhouse gas emissions [2–9] .

Emergency departments in Western healthcare systems face mounting pressure from rising patient volumes, leading to operational strain and delays in care delivery [10–13]. Alongside this growth, CT scan utilization has surged—Germany reported a 40% increase over seven years, and Switzerland observed a 34% rise between 2008 and 2018 [14–16]. This escalation reflects both increasing demand and advances in imaging technology, which, while expanding diagnostic capabilities, have also produced more complex and time-intensive interpretations. As a result, radiologists have become a critical bottleneck in emergency workflows [16].

To address appropriateness concerns, clinical decision support systems (CDSSs) have been introduced. The American College of Radiology Appropriateness Criteria (ACR AC), developed in 1993, offer evidence-based guidelines for over 3,700 clinical scenarios [17, 18]. The European Society of Radiology (ESR) subsequently launched ESR iGuide, a web-based platform that provides structured recommendations via a similar appropriateness scoring system [19, 20].

However, the real-world impact of CDSSs has been modest. A study involving 941 clinicians and 22,279 imaging requests reported a small improvement in appropriateness (77% to 80%) when best practice alerts were triggered using ACR criteria [21]. Another randomized trial with 3,511 providers showed a 6% reduction in high-cost imaging utilization [22]. A recent German multicenter study found ESR iGuide implementation did not significantly reduce inappropriate examinations, with ~6% remaining inappropriate post-deployment. Persistent barriers to adoption include workflow disruption and low perceived usefulness [23].

In recent years, artificial intelligence has emerged as a transformative force in medicine. Large language models (LLMs) have demonstrated promising performance in modality selection, achieving high concordance with expert radiologists in controlled evaluations [24]. Yet, proprietary models raise concerns about patient data privacy and environmental sustainability [25–28].

This study aimed to determine whether an open-access, privacy-preserving LLM can improve imaging appropriateness in emergency setting relative to an established clinical decision support system [29].

Materials and methods

Study design

This single-center, retrospective, comparative, semi-quantitative cohort study was conducted at the University Training and Research Hospital of Fribourg (HFR), Switzerland. The protocol was approved by the Cantonal Ethical Committee of "Canton de Vaud" (BASEC-ID 2023-00572) and informed consent was obtained from all participants. The following hypotheses were tested: (1) OpenAccessGPT's ability to reduce imaging requests by at least 10% as measured by a one-sided test of proportion; (2) whether GPT gave different recommendations compared to ESR iGuide as indexed by significantly less than 90% of concordance, measured by a one-sided test of proportion; (3) whether GPT was significantly faster than ESR iGuide as measured by a paired one-sided t-test; and finally (4) whether our institutional practice differed from ESR iGuide, either significantly above or below our 90% concordance threshold.

Radiological examinations requests and population eligibility

Radiological requests were retrieved using consecutive chronological sampling, with deliberate balancing to include equal proportions of US, X-rays, CT scans, and MRI. Requests were eligible for inclusion if they originated from the emergency department, contained sufficient clinical information, and were covered by an institutional data use agreement for research purposes. Requests were excluded if they involved multiple imaging modalities requested simultaneously for the same clinical indication in a single patient.

Index GPT system and reference clinical decision support system

The performance of a GPT model was evaluated against that of a validated CDSS. The GPT model was OpenAccessGPT, an open-access large language model (GPT-3.5) queried via text prompt to simulate imaging recommendation generation. The validated CDSS used was ESR iGuide, a web-based platform developed by the European Society of Radiology that provides guideline-based imaging recommendations with an appropriateness score.

Data acquisition

Data were extracted from the Radiology Information System (RIS), including patient age, sex, examination type, full clinical information, degree of emergency, and examination date. Emergency was classified according to the four levels of the Swiss Triage Scale.

Data processing

To allow standardized evaluation, the original clinical data extracted as free text from the electronic radiological requests were translated from the local language French into English using DeepL traduction software, then reviewed for linguistic accuracy by a single investigator (J.A.) [30]. Both OpenAccessGPT and ESR iGuide were queried by the same investigator between March and April 2025, ensuring temporal consistency in system versions.

For OpenAccessGPT, a standardized prompt was used: "[patient age] + [patient gender] + [clinical indication from the radiology request] + 'According to current international guidelines, is there a recommended radiological examination? If yes, which one is recommended?'" When contrast use was not specified but potentially required, the additional query "with or without contrast injection?" was submitted. The time required from prompt introduction to recommendations from both OpenAccessGPT and ESR iGuide was recorded.

Only the first response from OpenAccessGPT and the "usually appropriate" (appropriateness rating 7–9) recommendation from ESR iGuide were recorded. All outputs were stored in a structured database, with screenshots archived on a secure institutional FileCloud server in compliance with data protection regulations [31].

Classification of Appropriateness.

For analysis, recommendations were dichotomized as appropriate or inappropriate. For the OpenAccessGPT vs. ESR iGuide comparison, the recommendations queried from both systems were considered as concordant only if modality, body part, and contrast phase matched exactly.

Statistical analysis

All statistical tests were performed with R version 4.3.1 with the base functions if not stated otherwise. Continuous variables were reported as mean ± SD and range; categorical variables as counts and percentages. Statistical significance was set at $p < .05$.

For the first and second hypotheses, a one-sided test of given proportion was performed without correction with 90% as the target value. For the third hypothesis, a Welch's one-sided paired t-test was performed between the "prompt to response" time of the two conditions (GPT vs. ESR iGuide), with a Cohen's *d* computed with the R package DescTools as an effect size. For the fourth hypothesis, the same analysis as the first and second hypotheses was performed, but two-sided as we did not state which condition (i.e., institutional practice vs. ESR iGuide) would drive the difference.

Sample size

A conservative sample size of 175 radiological requests was determined with a-priori power analyses to ensure the highest standard of statistical power (i.e., 95%) for our four hypotheses, as detailed in Table 1. We increased this sample size to 200 during data acquisition to account for potential exclusions before performing the analyses.

Due to the absence of concordant scenarios in ESR iGuide, 11 clinical cases could not be processed by this system and had to be excluded, leaving 190 patients for analysis across hypotheses 2, 3 and 4. These 11 cases included six cases of testicular pain, a suspected fracture of the nose, a case of worsened ear swelling despite antibiotic treatment, an inflammatory syndrome of unknown origin, a case of quadriceps pain without trauma and a painful subcutaneous lipoma.

Results

Radiological examinations requests and population characteristics

Two-hundred-and-one consecutive radiological examinations requested by the emergency department between March 2023 and April 2024 meeting the inclusion criteria were selected. The mean patient age was 55.3 years (SD = 20.9 years), with females representing 45% of the cohort (90/201). The average degree of emergency according to the Swiss Triage Scale was 2.9 (SD = 0.89). The final dataset comprised 50 CT scans, 50 US, 50 X-rays, and 51 MRI, as detailed in the flowchart (Fig. 1). Table 2 provides a summary of the main findings.

Among the CT scans, 18 were related to neuroimaging or ear-nose-throat (ENT) assessments, 12 to urogenital and abdominal conditions, 11 to cardio-thoracic concerns, 3 to musculo-skeletal (MSK) cases, 3 were whole-body scans for polytrauma, and 2 involved cross-specialty indications. Among the US examinations, 31 addressed abdominal pathologies, 10 MSK issues, and 9 urogenital complaints. The 50 X-rays comprised 23 MSK and 27 chest studies. Regarding MRI, 12 examinations focused on neuroimaging

Table 1 Description of hypotheses and power analyses

	Hypothesis	Description	Power analysis
Primary endpoint	Reduction of imaging request based on OpenAccessGPT recommendations	Number of radiological examinations actively discouraged by OpenAccessGPT, in comparison to those that were ultimately performed in routine clinical practice.	A-priori power analysis for a one-sided test of one proportion indicated that a sample size of 120 was needed to achieve a power of 95% for a small effect size (Cohen's h = 0.3).
Secondary endpoints	Appropriateness of OpenAccessGPT recommendations when compared to ESR iGuide	Proportion of imaging recommendations generated by OpenAccessGPT that were classified as "usually appropriate" according to ESR iGuide criteria.	Same as first hypothesis.
	Time saving with OpenAccessGPT	Time to obtain a recommendation using OpenAccessGPT compared to ESR iGuide.	A-priori power analysis computed with G*Power for a paired one-sided t-test indicated that a sample size of 175 was needed to achieve a power of 95% for a difference of 10 s between OpenAccessGPT and ESR iGuide with a conservatively estimated pooled standard deviation of 40 s.
	Appropriateness of institutional practices compared to ESR iGuide recommendations	Number of mismatch examination between practice and ESR iGuide recommendations.	A-priori power analysis for a two-sided test of one proportion indicated that a sample size of 144 was needed to achieve a power of 95% for a small effect size (Cohen's $h = 0.3$).

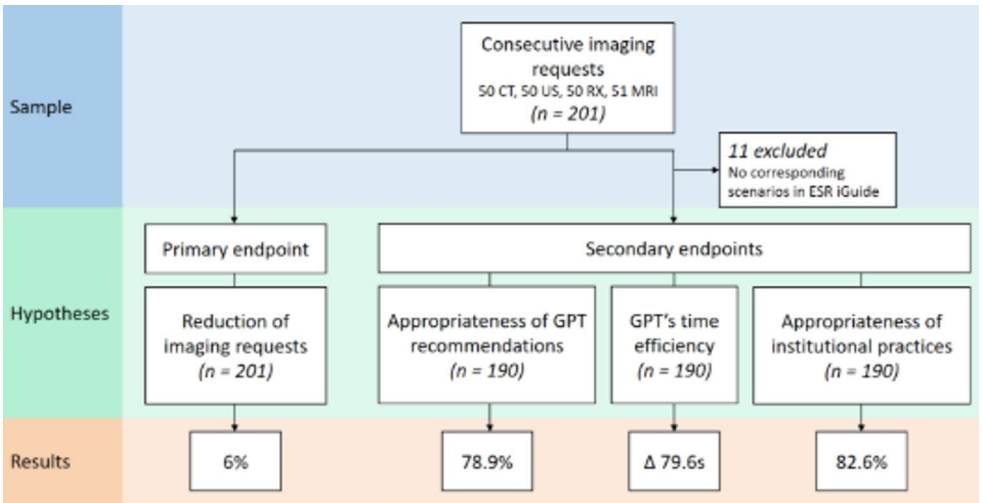


Fig. 1 Flowchart of radiological requests through the study

Table 2 Summary of main findings

Endpoint	Metric	Result	Predefined subjective threshold	*p*-value	Conclusion
Reduction of imaging request	% cases GPT advised against imaging	6.0% (12/201)	≥10% reduction	0.972	No significant reduction
GPT - ESR iGuide concordance	% agreement	78.9% (150/190)	≥90%	<0.001	Significantly below threshold
Time efficiency	Mean response time difference	GPT 79.6 s faster (21.6 s vs. 101 s)	GPT faster	<0.001	GPT significantly faster
Institutional practice – ESR iGuide concordance	% agreement	82.6% (157/190)	≥90%	<0.001	Significantly below threshold

Table 3 Details of disagreement between OpenAccessGPT and ESR iGuide recommendations

Parameter	Description
Imaging modalities recommended by OpenAccessGPT for the 40 discordant cases	8 CT, 7 US, 12 X-rays, 5 MRI, 8 cases with actively not recommended examinations.
Type of disagreement	57.5% (23/40) involved modality mismatches; 7.5% (3/40) involved mismatches in anatomical region.
CT scans recommended by OpenAccessGPT	50% (4/8) were downgraded to US by ESR iGuide.
X-rays recommended by OpenAccessGPT	66% (8/12) were upgraded by ESR iGuide to either CT or MRI (evenly distributed).
MRI recommended by OpenAccessGPT	80% (4/5) were not recommended by ESR iGuide, mostly related to low back pain without red flags.
No imaging recommended	OpenAccessGPT: 20% (8/40); ESR iGuide: 15% (6/40); no overlapping cases.

In total, this study found that 21.1% (40/190) of imaging requests resulted in discordant recommendations between OpenAccessGPT and ESR iGuide.

and 39 on MSK. Contrast enhancement was administered in 37 out of 50 CT scans and in 12 out of 51 MRI scans.

Reduction of imaging request based on openaccessgpt recommendations

OpenAccessGPT recommended performing radiological examination in 94% (189/201) of cases for which an examination was actually performed at our institution. This means that 6% (12/201) of the requests were assessed by OpenAccessGPT as not requiring an examination. This falls short of our predefined clinical significance threshold of less than 90% of advised cases ($\chi^2 = 3.63$, $p = .972$), indicating that GPT does not significantly reduce radiology examinations. These 12 cases included 3 CT, 1 US, 4 X-rays and 4 MRI studies. OpenAccessGPT recommended further clinical evaluation or preliminary investigations before considering imaging in 33.3% (4/12) of cases. In a similar proportion (33.3%, 4/12), the clinical scenario involved low back pain without red flags (red flags included: recent trauma, patient age below 20 years or above 50 years, features suggestive of inflammatory back pain, suspected infection, systemic deterioration, or neurological deficits). No specific observations were identified for the remaining 4 out of 12 cases. ESR iGuide supported OpenAccessGPT's decision against imaging in 5 of the 12 cases (41.7%).

Appropriateness of OpenAccessGPT recommendations when compared to ESR iGuide

Among the 190 cases assessed by both OpenAccessGPT and ESR iGuide, complete agreement was achieved in 78.9% (150/190) of recommendations, significantly below the hypothesized threshold of 90% ($\chi^2 = 25.79$, $p < .001$). There were 21.1% (40/190) discordant cases. Primary source of disagreement was a mismatch in imaging modality, accounting for 57.5% (23/40). This was followed by mismatches in examinations that were actively not recommended, accounting for 35% (14/40). The remaining 7.5% (3/40) of cases involved mismatches in the anatomical region recommended for assessment. Table 3 provides further details of the mismatches.

Time saving with OpenAccessGPT

The time needed to obtain a recommendation from OpenAccessGPT was under one minute in all cases, with a mean interval from prompt query to response of 21.6 s (SD = 4.65 s, range 13–45 s). OpenAccessGPT mean response time was 79.6 s faster than ESR iGuide, for which time requested to obtain a recommendation was less consistent (mean = 101 s, SD = 86.4 s, range = 18–542 s, Cohen's $d = 1.3$; $t_{[189]} = 12.84$, $p < .001$).

Appropriateness of Institutional Practice Compared to ESR iGuide Recommendations

Among the 190 radiological examinations evaluated using ESR iGuide, agreement between recommendations and actual institutional practice at our institution was observed in 82.6% (157/190) of cases. This concordance rate falls significantly below the expected 90% threshold ($\chi^2 = 11.46$, $p < .001$). Of the 33 discordant cases, ESR iGuide advised against any imaging in 36% (12/33) of cases, with 58% (7/12) of these involving patients presenting with low back pain. Additionally, in 21% (7/33) of discordant cases, ESR iGuide recommended performing a CT examination, whereas US was instead performed in clinical practice. Two recurrent indications were identified among these cases: suspected appendicitis in 28,5% of cases (2/7), and diffuse abdominal pain in patients with a history of abdominal surgery 28,5% of cases (2/7). No specific pattern was identified in the remaining cases.

Discussion

This study assessed whether OpenAccessGPT, a privacy-focused open-access GPT-3.5–based model, could reduce radiological requests in an emergency department, while maintaining recommendation appropriateness and efficiency compared with ESR iGuide, an established CDSS. We also evaluated the proportion of inappropriate imaging in our institution based on ESR iGuide standards.

OpenAccessGPT advised against imaging in 6% of cases, lower than our predefined clinical significance threshold ($\chi^2 = 3.63$, $p = .972$). It achieved 78.9% concordance with ESR iGuide, below the hypothesized threshold of 90% ($\chi^2 = 25.79$, $p < .001$), but generated recommendations 79.6 s faster on average than ESR iGuide ($p < .001$).

Our findings extend prior work by Rosen and Saban (2023), who reported an 87.6% concordance between ChatGPT-3 and ESR iGuide across 97 abdominal CT referrals [32]. Their narrow design, restricted to a single modality and indication, likely facilitated this higher performance. By contrast, our study included CT scans, MRI, US, and X-rays across a broad range of emergency scenarios, better reflecting clinical reality. This broader scope may explain our lower concordance (78.9%). Consistent with this, 5.5% of our cases were excluded due to absence of corresponding ESR iGuide scenarios, compared with 3% in Rosen and Saban's study, mainly reflecting indications such as testicular pain not yet covered by ESR iGuide. Together, these differences highlight the limitations of both LLM and guideline-based systems when applied beyond tightly controlled contexts.

Within the subset of cases where OpenAccessGPT recommended against imaging, concordance with ESR iGuide was only 33.3%, suggesting limited potential for reducing overuse, despite the small sample size. This is further reflected in the overall inappropriate imaging rate of 17.4% observed in our cohort (per ESR iGuide), lower than the 20–30% reported in prior studies, and pointing to OpenAccessGPT's relative underperformance in curbing unnecessary imaging.

Qualitative review showed that in about one third of discordant cases OpenAccessGPT cited insufficient clinical information. Another third involved MRI for acute low back pain without red flags, reflecting a well-documented driver of unnecessary imaging. These findings illustrate how pressures such as patient expectations or defensive medicine may drive requests beyond what CDSS or OpenAccessGPT alone can counteract.

This study demonstrated a moderate concordance between OpenAccessGPT and ESR iGuide recommendations (78.9%). More than half of mismatches were modality-related. Notably, both polytrauma cases in our sample fell into this category, with OpenAccessGPT suggesting more conservative imaging than ESR iGuide's whole-body CT scan recommendations. Current guidelines from the European Society of Emergency Radiology endorse whole-body CT scan as standard, reflecting its diagnostic value in critical trauma [33]. Such highly critical decisions, usually made by trauma teams, remain challenging for generalized algorithms, whether CDSS or LLM-based.

Regarding time efficiency, OpenAccessGPT clearly outperformed ESR iGuide, with a highly significant five-fold advantage. This may reflect the difference between OpenAccessGPT's natural language interface and ESR iGuide's structured, multi-field input. Usability challenges of CDSS interfaces have been reported previously and remain major barriers to adoption [23]. Although ESR iGuide response times decreased slightly with repeated use, indicating a learning curve, OpenAccessGPT maintained a substantial speed advantage. Newer LLMs supporting voice input may further amplify this benefit, especially if integrated into tailored clinical interfaces with pre-saved prompts to streamline workflows.

Several limitations should be acknowledged. First, the single-center retrospective design introduces selection bias, and limits generalizability. Because the analysis was restricted to examinations that were already performed, the study cannot support inference about actual decision avoidance or practice change. Nevertheless, consecutive inclusion and balanced modality sampling aimed to reflect real-world practice. Second, ESR iGuide's outputs depend on keyword selection, which affects reproducibility. This variability was preserved deliberately to replicate clinical practice. More fundamentally, comparing a probabilistic language model with a deterministic, rule-based system is conceptually imperfect, as the two operate under different assumptions. Third, this study evaluated a single, non-fine-tuned LLM (OpenAccessGPT, based on GPT-3.5, without real-time access to medical literature or clarified training data cutoff). This limits generalizability to other LLM platforms: newer, larger, or domain-specific fine-tuned models may perform differently, and our findings should be interpreted as reflecting this specific model configuration rather than LLM-based imaging decision support in general. Finally, exclusion of cases not represented in ESR iGuide reflects intrinsic coverage limitations of guideline-based systems themselves, and should be considered when interpreting whether observed differences reflect true performance gaps or merely ESR iGuide's scenario coverage.

Our results highlight both the potential and shortcomings of LLMs in radiological decision support. These findings suggest value not in standalone deployment, but in hybrid models that combine LLMs' efficiency with the evidence-based rigor of established CDSSs. Realizing this will

require prospective multicenter studies, real-world usability assessments, and integration of broader clinical data, alongside continued attention to data privacy and environmental sustainability as AI is deployed in healthcare. In summary, GPT-based systems are not yet viable as standalone clinical tools in this context, but represent a promising foundation for next-generation decision support, provided their development proceeds under rigorous validation and ethical oversight.

Acknowledgements During the preparation of this work the authors used large language models; ChatGPT and Perplexity; as well as DeepL Translator to assist with language editing, improving the clarity of the manuscript text. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Author contributions Jérémie ARTHAUD: conceptualization, methodology, investigation, writing-original draft. Harriet C. THOENY: conceptualization, methodology, writing-review and editing, supervision. Vincent GREK: writing-review and editing. Lucien WIDMER: conceptualization, methodology, data curation, writing-review and editing, supervision, project administration and funding acquisition.

Funding Open access funding provided by University of Fribourg. This study has received funding by the University Training and Research Hospital of Fribourg (HFR Research Grant: 2339). Open Access funding was provided by the same institution.

Data availability Data supporting the study findings are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Ethical approval and consent to participate This study was approved by the Cantonal Ethical Committee of "Canton de Vaud" (BASEC-ID 2023−00572).

Consent for publication Written informed consent was obtained from all participants in this study. There are no identifying characteristics of participants in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. De Alencastro L, Clair C, Locatelli I et al (2017) Raisonnement clinique: de la théorie à la pratique… et retour. Rev Med Suisse 562:986–989
2. Almen A, Leitz W, Richter S, Swedish Radiation Safety Authority S (Sweden) (2009) National Survey on Justification of CT-examinations in Sweden. Swedish Radiation Safety Authority, Stockholm. (Sweden)
3. Lehnert BE, Bree RL (2010) Analysis of appropriateness of outpatient CT and MRI referred from primary care clinics at an academic medical center: how critical is the need for improved decision support? J Am Coll Radiol 7:192–197. https://doi.org/10.1016/j.jacr.2009.11.010
4. Vilar-Palop J, Hernandez-Aguado I, Pastor-Valero M et al (2018) Appropriate use of medical imaging in two Spanish public hospitals: a cross-sectional analysis. BMJ Open 8:e019535. https://doi.org/10.1136/bmjopen-2017-019535
5. Ebdon-Jackson S, Frija G, European Society of Radiology (2021) Improving justification of medical exposures using ionising radiation: considerations and approaches from the European Society of Radiology. Insights into Imaging 12:2. https://doi.org/10.1186/s13244-020-00940-0
6. Thiel CL, Vigil-Garcia M, Nande S et al (2024) Environmental Life Cycle Assessment of a U.S. Hospital-based Radiology Practice. Radiology 313:e240398. https://doi.org/10.1148/radiol.240398
7. Cavanagh G, Schoen JH, Hanneman K et al (2025) Excess Greenhouse Gas Emissions Associated With Inappropriate Medical Imaging in the US Medicare Part B Population From 2017 to 2021. J Am Coll Radiol 22:621–629. https://doi.org/10.1016/j.jacr.2025.02.043
8. Vosshenrich J, Merkle EM, Heye T (2025) The carbon footprint of modern imaging. Curr Opin Urol 35:674–678. https://doi.org/10.1097/MOU.0000000000001337
9. Rosenkrantz AB, Cummings RW (2025) Utilization of Emergency Department Imaging from 2013 to 2023: A National Medicare Analysis. Radiology 316:e251395. https://doi.org/10.1148/radiol.251395
10. Pines JM, Hilton JA, Weber EJ et al (2011) International perspectives on emergency department crowding. Acad Emerg Med 18:1358–1370.https://doi.org/10.1111/j.1553-2712.2011.01235.x
11. Hugli OW, Potin M, Schreyer N, Yersin B (2006) Engorgement des centres d'urgences: une raison légitime de refuser l'accès aux patients non urgents? Rev Med Suisse 075:1836–1839
12. Consultations dans un service d'urgence en Suisse | OBSAN. (2013) https://www.obsan.admin.ch/fr/publications/2013-consultations-dans-un-service-durgence-en-suisse. Accessed 13 June 2025
13. Carron P-N, Sarasin F (2022) Redéfinir les urgences par ce qu'elles ne sont pas ? Rev Med Suisse 791:1479–1480
14. Samara ET, Aroua A, Bochud FO et al (2012) Exposure of the Swiss population by medical x-rays: 2008 review. Health Phys 102:263–270. https://doi.org/10.1097/hp.0b013e31823513ff
15. (PDF) Annual Exposure of the Swiss Population From Medical Imaging (2018) ResearchGate. https://doi.org/10.1093/rpd/ncab012
16. Mezrich J (2020) Radiology bottlenecks - Is there any utility of just in time inventory principles in the ED setting? Clin Imaging 66:54–56. https://doi.org/10.1016/j.clinimag.2020.04.035
17. Amis ES (2000) American College of Radiology standards, accreditation programs, and appropriateness criteria. AJR Am J Roentgenol 174:307–310. https://doi.org/10.2214/ajr.174.2.1740307

18. Appropriateness Criteria https://www.acr.org/Clinical-Resources/Clinical-Tools-and-Reference/Appropriateness-Criteria. Accessed 12 June 2025
19. ESR ES of R EuroSafe Imaging Together - for patient safety. https://www.eurosafeimaging.org/information-for-referring-professionals-2/clinical-decision-support. Accessed 12 June 2025
20. European Society of Radiology (ESR) (2019) Methodology for ESR iGuide content. Insights into Imaging 10:32. https://doi.org/10.1186/s13244-019-0720-z
21. Doyle J, Abraham S, Feeney L et al (2019) Clinical decision support for high-cost imaging: A randomized clinical trial. PLoS ONE 14:e0213373. https://doi.org/10.1371/journal.pone.0213373
22. Palen TE, Sharpe RE, Shetterly SM, Steiner JF (2019) Randomized Clinical Trial of a Clinical Decision Support Tool for Improving the Appropriateness Scores for Ordering Imaging Studies in Primary and Specialty Care Ambulatory Clinics. AJR Am J Roentgenol 213:1015–1020. https://doi.org/10.2214/AJR.19.21511
23. Dijk SW, Wollny C, Barkhausen J et al (2025) Evaluation of a Clinical Decision Support System for Imaging Requests: A Cluster Randomized Clinical Trial. JAMA 333:1212–1221. https://doi.org/10.1001/jama.2024.27853
24. Gertz RJ, Bunck AC, Lennartz S et al (2023) GPT-4 for Automated Determination of Radiological Study and Protocol based on Radiology Request Forms: A Feasibility Study. Radiology 307:e230877. https://doi.org/10.1148/radiol.230877
25. The Italian Data Protection Authority halts ChatGPT's data processing operations. In: Clifford Chance. https://www.cliffordchance.com/content/cliffordchance/insights/resources/blogs/talking-tech/en/articles/2023/04/the-italian-data-protection-authority-halts-chatgpt-s-data-proce.html. Accessed 13 June 2025
26. L'équipe chargée de la sécurité d'une (2024)super IA chez OpenAI dissoute. In: rts.ch. https://www.rts.ch/info/sciences-tech/2024/article/l-equipe-chargee-de-la-securite-d-une-super-ia-chez-openai-dissoute-28507223.html. Accessed 13 June 2025
27. Towler L (2025) Search Engines vs AI: energy consumption compared. In: Kanoppi. https://kanoppi.co/search-engines-vs-ai-energy-consumption-compared/. Accessed 13 June 2025
28. Patterson D, Gonzalez J, Le Q et al (2021) Carbon Emissions and Large Neural Network Training
29. OpenAccessGPT (ed) (2023) https://www.openaccessgpt.org/privacy-first-serverless/. Accessed 13 June 2025
30. DeepL Translate The world's most accurate translator. https://www.deepl.com/translator. Accessed 19 Aug 2025
31. Sci-data https://sci-data.unifr.ch/ui/core/index.html#/. Accessed 19 Aug 2025
32. Rosen S, Saban M (2024) Evaluating the reliability of ChatGPT as a tool for imaging test referral: a comparative study with a clinical decision support system. Eur Radiol 34:2826–2837. https://doi.org/10.1007/s00330-023-10230-0
33. Wirth S, Hebebrand J, Basilico R et al (2020) European Society of Emergency Radiology: guideline on radiological polytrauma imaging and service (short version). Insights Imaging 11:135. https://doi.org/10.1186/s13244-020-00947-7

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.