

Viewpoint

# AI Triage in Primary Care: Building Safer and More Equitable Real-World Evidence

Aymn Alamoudi<sup>1,2</sup>, MSc; Evangelos Kontopantelis<sup>3,4</sup>, PhD; Salwa S Zghebi<sup>1</sup>, PhD; Benjamin Brown<sup>1</sup>, PhD

<sup>1</sup>Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, England, United Kingdom

<sup>2</sup>Department of Public Health, School of Nursing and Health Sciences, Jazan University, Jazan, Jazan Region, Saudi Arabia

<sup>3</sup>Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, England, United Kingdom

<sup>4</sup>Division of Family Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

**Corresponding Author:**

Aymn Alamoudi, MSc

Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health

University of Manchester

Williamson Building, 5th Floor

Oxford Road

Manchester, England, M13 9PL

United Kingdom

Phone: 44 161 306 6000

Email: [aymn.alamoudi@postgrad.manchester.ac.uk](mailto:aymn.alamoudi@postgrad.manchester.ac.uk)

## Abstract

Artificial intelligence triage in general practice is developing rapidly within the primary care digital transformation, promising efficiency gains and safety standardization in overwhelmed primary care systems. However, current evidence is drawn from retrospective validations, emergency settings, or vignettes, with scant evaluation of real-world outcomes and almost no equity-stratified safety data, despite known disparities across age, ethnicity, language, and deprivation. From a sociotechnical standpoint, which considers the fit between people, tasks, technology, and organizational context, risks arise not only from algorithmic bias and undertriage but also from human factors, workflow misalignment, governance gaps, and inadequate postdeployment monitoring. We argue that ensuring artificial intelligence triage is safe and equitable requires real-world evaluations in primary care settings, equity-focused performance reporting using theoretically informed frameworks, and rigorous postmarket surveillance. Without these, deployment may widen existing health inequalities rather than moderate them.

(*J Med Internet Res* 2026;28:e88396) doi: [10.2196/88396](https://doi.org/10.2196/88396)

**KEYWORDS**

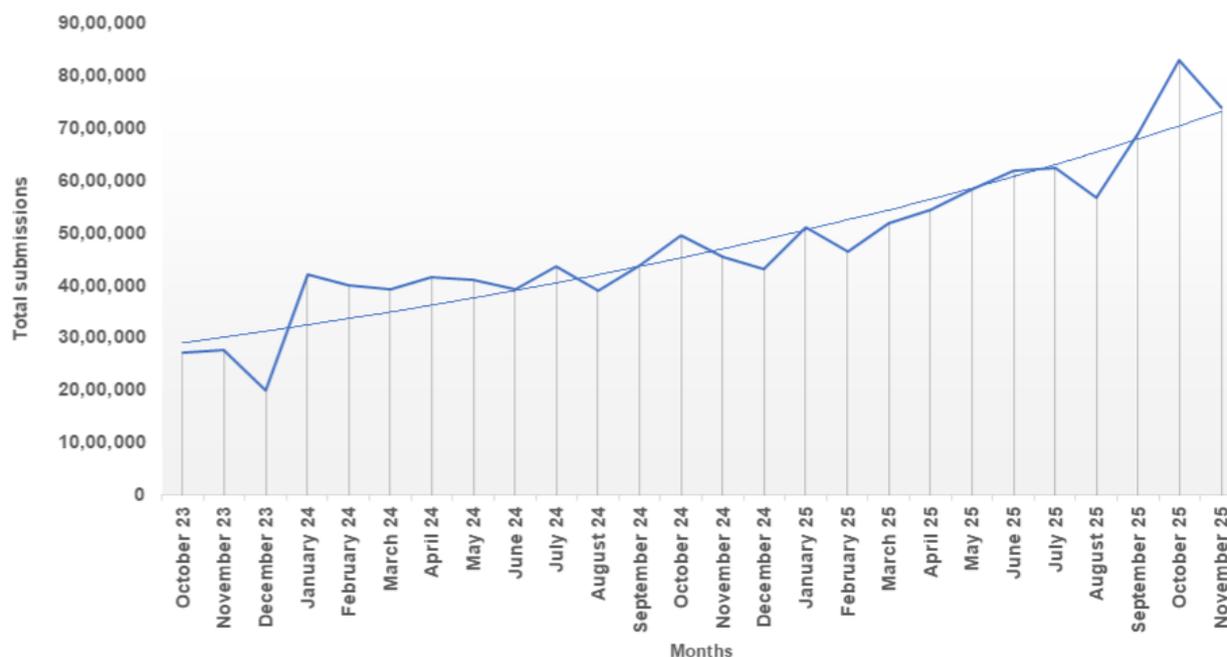
artificial intelligence; AI; triage; primary care; general practice; patient safety; health equity; real-world evidence; clinical decision support

## Introduction

Globally, primary care faces sustained growth in demand, increased patient complexity, and a workforce whose full-time equivalent growth has not matched demand, resulting in persistent access pressures and delays [1,2]. The COVID-19 pandemic accelerated the adoption of remote and digital access, including online consultations and reinforced strategic commitments to “digital front door” models within health systems, such as the National Health Service [3-5]. Online

consultation submissions in England rose from approximately 2.7 million in October 2023 to a peak of 8.3 million in October 2025, as seen in [Figure 1](#), highlighting the rapid and sustained growth of digital entry points into general practice (GP) [6]. Evidence also suggests that digital access is not equity neutral. A systematic review of inequalities in remote GP consultations found differential use by sociodemographic characteristics, with internet-based consultations more frequently used by younger, more affluent, and more educated groups, and noted that the impact of these inequalities on clinical outcomes remains uncertain [7].

**Figure 1.** Growth in monthly online consultation submissions in England. Data source: National Health Service England release [6]. Information from NHS England, licenced under the current version of the Open Government Licence.



In this context, artificial intelligence (AI)-enabled triage combines structured questions, red-flag pathways, and machine learning (ML) risk stratification with electronic health record (EHR) integration and clinician oversight to route patients more efficiently and potentially improve safety [8-10]. In this viewpoint, we distinguish between 3 related but conceptually distinct system types: symptom checkers, clinical decision support systems (CDSSs), and AI triage. Symptom checkers are patient-facing digital tools that provide health advice or triage recommendations directly to users, often without clinician oversight, and have been widely evaluated in consumer and emergency contexts [8]. CDSSs are clinician-facing tools embedded within clinical workflows or EHRs that support decision-making through alerts, risk scores, or guideline-based recommendations [9,10]. AI-enabled triage refers to digital systems that collect patient-reported information and generate urgency or routing recommendations (eg, self-care, routine review, urgent GP assessment, or emergency referral), with or without clinician oversight [11]. These systems may be embedded within online consultation platforms, patient-facing symptom checkers, or CDSS. Importantly, not all online consultation systems are AI enabled and not all AI-enabled triage systems function as stand-alone symptom checkers.

This viewpoint advances 3 linked arguments. First, we argue that triage in primary care is a safety-critical and equity-sensitive function, such that errors or delays can produce serious harm and unequal outcomes. Second, we show that the current evidence base for AI-supported triage is dominated by emergency department (ED), vignette, and retrospective studies, with little real-world or equity-stratified evaluation in GP. Third, we argue that AI triage operates as a sociotechnical system shaped by human behavior, workflows, and governance, meaning that algorithmic accuracy alone cannot guarantee safety

or fairness. This viewpoint aims to outline a practical agenda for evaluating and governing AI-enabled triage in GP that integrates real-world safety outcomes, equity-stratified performance reporting, and sociotechnical implementation and surveillance. The intended audience includes GP clinicians and practice leaders, digital health and AI developers, evaluators and implementation scientists, and policymakers and regulators responsible for deployment and monitoring. Our contribution is to consolidate a practical, real-world evaluation and governance agenda for AI triage in GP that integrates sociotechnical safety (workflow and human factors), equity-stratified performance reporting (including an example of fairness), and postdeployment surveillance.

### Triage in Primary Care Is Safety Critical and Equity Sensitive

In health care, triage refers to the systematic process of assessing patient urgency and risk to determine the appropriate level, timing, and pathway of care [12,13]. In primary care, triage does not establish a diagnosis but prioritizes patients for self-care, routine review, urgent GP assessment, or emergency referral based on presenting symptoms, clinical risk, and service capacity [14,15]. This function is safety critical because misclassification can lead to delayed diagnosis, inappropriate self-management, or unnecessary escalation [8-10]. AI-enabled triage promises standardization and auditability but introduces novel patient-safety risks, such as automation bias, algorithmic mistriage, and digital exclusion, particularly for socially disadvantaged groups [8,16,17]. Moreover, model performance (eg, sensitivity, specificity, calibration, and error rates) may vary by age, ethnicity, language, or limited digital access, unless these dimensions are intentionally tested and monitored [18-22].

The absence of equity evaluation may pose a significant risk. If models are calibrated primarily on majority language, younger, or White-majority cohorts, AI triage may systematically de-escalate or deprioritize patients whose symptom descriptions diverge due to cultural or linguistic factors. Coupled with cognitive and automation biases in human users, the most vulnerable groups risk unsafe disposition, such as self-care advice when urgent assessment is indicated.

The “equity blind spot” in AI triage is not merely a technical glitch; it reflects broader systemic oversight. To operationalize safe, equitable AI, embedding framework-informed stratification, for example, PROGRESS-Plus, fairness metrics, and multidimensional performance reporting into every stage of model development, validation, and deployment is needed. Without this, AI triage may reinforce and even amplify existing health care disparities if deployed without adequate safeguards.

## What the Evidence Shows and What It Misses

### What Current Studies Show

Controlled evaluations of AI triage report high technical performance, with area under the receiver operating characteristic (singular) curve values typically between 0.82 and 0.94 and sensitivities often exceeding 0.75. However, these studies are predominantly retrospective, vignette based, or conducted in EDs and hospital settings. They rarely reflect routine workflows in GP [23-26].

Recently, Abualruz et al [12] reviewed 22 studies on AI-supported triage, most of which were carried out in emergency, acute, or hospital-based settings. Only a small subset examined outpatient or primary-care use. As a result, the current literature tells us little about how AI triage performs in everyday GP or how it affects patient safety in real workflows. Table 1 presents the distribution of published AI-supported triage studies by clinical setting.

**Table 1.** Distribution of published artificial intelligence–supported triage studies by clinical setting (N=22).

Clinical setting	Study type	Studies, n (%)
Emergency department or hospital	Real patient data	19 (86)
Primary care	Clinical vignettes or qualitative studies	3 (14)
Primary care	Real patient data	0 (0)

Most studies have been conducted in emergency or hospital settings using real patient data (19/22, 86%). In total, 14% (3/22) of the studies relied on clinical vignettes or qualitative interviews, and none (0/22, 0%) evaluated AI-supported triage using real-world patient data in routine GP.

### Why This Evidence Is Not Sufficient for Safe and Equitable GP Deployment

Equity reporting is also sparse. Few studies disaggregate performance by age, ethnicity, language, or socioeconomic status [27-30]. Intersectional analyses, for example, age×ethnicity or ethnicity×deprivation, are almost absent. Subgroup calibration, false-negative rates, and false-positive rates (FPRs) are rarely reported.

Study design further limits interpretability. Vignette-based and retrospective analyses do not capture real-world pressures, such as workload variation, free-text symptom input, clinician overrides, or case-mix drift [31]. Prospective designs, such as controlled interrupted time series or cluster-randomized trials, are almost never used in GP settings [1,3]. Without these designs, safety effects cannot be attributed reliably to AI deployment.

Postdeployment monitoring is also underdeveloped (Table 2). Few studies report ongoing calibration checks, subgroup performance dashboards, or systematic incident reporting aligned to the World Health Organization (WHO) International Classification for Patient Safety [32,33]. As a result, health systems lack visibility into how AI triage safety changes over time.

**Table 2.** What the evidence on artificial intelligence triage shows and what it misses.

Domain	What studies typically show	What is usually missing
Accuracy	High area under the receiver operating characteristic curve (≈0.82-0.94) and reasonable sensitivity in retrospective and vignette-based studies	Performance under real general practitioner workload, with free-text input, comorbidity, and clinician override
Setting	Predominantly emergency departments, acute care, or simulated cases	Routine general practice, community clinics, and longitudinal follow-up
Safety outcomes	Agreement with clinicians or reference standards	Delayed diagnosis, avoidable emergency use, or patient harm
Equity	Rare or absent subgroup reporting	Performance by age, ethnicity, language, deprivation, or intersectional groups
Monitoring	One-off validation at model development	Postdeployment drift, subgroup miscalibration, and incident tracking

Consistent with this, an ED-focused scoping review found limited demographic breakdowns and no multidimensional analyses, leaving equity implications unclear [27]. Within UK GP settings, stratified data on undertriage or performance by deprivation or ethnicity are rare [16,29], and subgroup calibration metrics or true-positive rate (TPR) and FPR reporting are notably absent. A recent international review found variable triage accuracy, poor calibration reporting, and limited deployment-level evaluation, reinforcing that the evidence gap is global [34].

## ***Why AI Triage Is a Sociotechnical System***

### **Overview**

AI-enabled triage systems are not isolated algorithms; they operate within complex care delivery systems where human factors, workflows, and trust dynamics profoundly shape safety. A sociotechnical system perspective, exemplified by the systems engineering initiative for patient safety framework, analyzes how people, tasks, tools, and organizational structures interact to influence patient safety.

By contrast, implementation frameworks focus on organizational readiness, technology adoption, and sustainability [35,36]. Together, these complementary approaches emphasize that deploying an accurate ML model alone does not guarantee safe outcomes; safety depends on both workflow integration and organizational adoption.

Human factors and trust are central. Health care professionals and patients must interpret AI-generated recommendations within their cognitive, ethical, and emotional contexts. A recent qualitative work on AI-based triage in Swedish primary care underlines how trust emerges from lived experience, transparency, and perceived reliability. Both patients' and professionals' trust is contingent on real-world usability and clear decision roles and not just model accuracy [31].

Similar issues are emerging in teledentistry and dental triage, where AI-enabled chatbots and triage systems are used to prioritize pain, infection, trauma, or urgent referral pathways [37]. Early work includes prototype "intelligent dental triage systems" and evaluations of AI chatbots for dental queries, but the same core risks apply: safety-critical undertriage, unequal performance for patients with language barriers or limited digital access, and workflow integration challenges in busy dental practices [38]. AI tools in dental assessment and smile analysis, such as Dynasmile, a video-based AI smile analysis platform in orthodontics, illustrate the expanding role of AI beyond workflow triage into diagnostic and aesthetic decision support in oral care [39].

This cross-domain comparison reinforces our central claim: AI triage should be evaluated as a sociotechnical intervention with equity-stratified safety reporting and postdeployment monitoring, regardless of clinical specialty.

### **Explainable AI to Support Calibrated Trust and Reduce Automation Bias**

In safety-critical triage, explanations should aim to support calibrated trust not persuasion. Evidence from clinical

decision-support research suggests that well-designed explanations can improve clinician understanding and trust calibration, while poorly designed explanations can increase overreliance and automation bias [40,41].

In practice, explainable AI for GP triage should be workflow integrated and low burden, including the following: (1) a short list of the main drivers for escalation (eg, red-flag symptoms and abnormal risk profile), (2) uncertainty indicators or confidence bands, (3) an "override required" prompt for high-risk edge cases, and (4) safety-netting text that is consistent with the triage rationale. Explanation stability is also important; near-identical inputs should not produce inconsistent rationales, as this undermines trust and may increase unsafe deference [40,42].

These requirements for explanation design reinforce the importance of sociotechnical fit described subsequently.

### **What Real-World Deployments Show**

AI triage does not operate as a stand-alone algorithm. In GP, it is embedded in symptom checkers, online consultations, patient apps, and EHR-based decision-support tools. These systems influence how patients describe symptoms, how clinicians prioritize work, and how care is delivered.

Evidence from multiple countries shows potential efficiency gains. In Iceland, an ML triage model for respiratory symptoms improved previsit risk stratification in community clinics [43]. Similarly, Brazil's primary care referral triage system demonstrated improved appropriateness of specialist referrals [44]. Studies from Sweden and Italy reported improved workflow transparency but persistent concerns about trust, usability, and clinician acceptance [31,33].

However, most deployments remain early stage, small scale, or limited to specific pathways. Systematic reviews of symptom checkers across Europe, the United States, Spain, Canada, and Asia report wide variability in triage accuracy and frequent mismatches between algorithmic and clinician assessments, particularly for complex multimorbidity and non-native-language users [8,14,34,45].

### **Why Workflow, Trust, and Integration Matter**

Qualitative evidence highlights that sociotechnical fit is critical. Steerling et al [31] found that both health care professionals and patients require alignment with clinical judgment, transparency, and oversight before trusting AI-based triage. Similarly, Siira et al [46] identified 3 interacting barriers in Swedish primary care: (1) professional skepticism or resistance and trust, (2) organizational readiness and digital maturity, and (3) technical limitations and poor EHR integration. Successful sites mitigated these by hands-on leadership and staff training, "superuser" networks, and iterative codevelopment with vendors. Even where efficiency gains were perceived, unresolved integration gaps and complex case-mix sustained workload and safety concerns.

Evidence from the UK primary care e-visits (14 practices; 16 staff, 24 patients; 2020-2021) identified 7 concrete AI use cases—workflow routing, directing, prioritization, postsubmission adaptive questioning, writing assistance,

self-help information, and autobooking—and found acceptability hinged on clinical oversight, timely responses, and ongoing evaluation. Perceived upsides were workload relief and faster help, while risks were depersonalization and mistriage, if poorly implemented [47].

### What This Means for Safety

These findings show that AI triage safety depends on how tools interact with people, workflows, and organizational routines. This aligns with sociotechnical safety theory, particularly the systems engineering initiative for patient safety framework, which emphasizes the fit between tasks, technology, and organizational context [32,48,49].

AI also offers enhanced structured data capture, natural language processing (NLP)-enabled symptom interpretation, EHR-integrated safety netting, and auditable decision trails [50]. NLP-enabled symptom interpretation can, for example, recognize “feeling pressure in chest” as equivalent to angina, supporting safer triage for patients who do not use standard terminology [26,45]. However, these benefits are offset by risks. Algorithmic mistriage, automation bias, and poor integration can delay escalation or overload urgent pathways [8,13].

Without governance mechanisms, such as version control, audit logs, safety dashboards, and periodic revalidation, performance may degrade over time as populations, language, and risk profiles change [32]. Therefore, sociotechnical alignment is not optional; it appears essential for safe AI-enabled triage.

### The Equity Blind Spot

Although AI-enabled triage systems are promoted as equitable tools for managing primary care demand, the current evidence reveals a persistent equity blind spot, driven by

underrepresentation, limited fairness measurement, and neglect of equity-focused monitoring [51].

Equity in digital health demands more than equal treatment; it requires fair opportunity to achieve safe, good outcomes, especially when baseline disadvantages exist. Equity reporting should use structured tools. The PROGRESS-Plus framework, developed by the Cochrane-Campbell equity methods group, was designed to systematically identify and report on equity-relevant factors in health research [17]. It extends the original PROGRESS (place of residence, race and ethnicity, occupation, gender, religion, education, socioeconomic status, and social capital) acronym with “plus” dimensions, including age, disability, and language [17]. The framework was created to help researchers illuminate disparities that might otherwise be masked in aggregated data and has since been widely applied in clinical trials, systematic reviews, and digital health evaluations.

Alternative frameworks, such as the health equity impact assessment tool, as seen in [Textbox 1](#), are used prospectively to anticipate equity impacts before interventions are deployed [52]. The SIITHIA (Strengthening the Integration of Intersectionality Theory in Health Inequality Analysis) checklist provides structured criteria for identifying inequities in digital health [53]. More recently, the digital health equity framework extends this approach to digital interventions and multidimensional analyses [54]. In practice, these can be operationalized by setting thresholds (eg, sensitivity gaps  $\leq 5\%$  between groups), with breaches triggering model review and corrective action. Complementing this, algorithmic fairness metrics, such as equal opportunity (equal TPRs), equalized odds (matching both TPR and FPR), and calibration integrity, are critical for measuring subgroup performance and detecting systematic bias. Without these frameworks, inequities may go undetected beneath aggregated performance.

**Textbox 1.** Frameworks for evaluating safety and equity in artificial intelligence triage.

#### Safety and sociotechnical performance

- Systems engineering initiative for patient safety describes how people, tasks, tools, and organizational context interact to shape patient safety [48].
- World Health Organization International Classification for Patient Safety provides a standardized taxonomy for reporting and classifying safety incidents [32,49,55].

#### Implementation and adoption

- Consolidated Framework for Implementation Research assesses organizational readiness and barriers to and facilitators of implementation.
- Nonadoption, abandonment, scale-up, spread, and sustainability [35,36] framework evaluates the complexity and long-term viability of digital health technologies.
- Reach, effectiveness, adoption, implementation, and maintenance framework evaluates long-term reach, effectiveness, adoption, and maintenance [56].
- Human, organization, and technology-fit framework examines alignment between human, organizational, and technical factors [57].

#### Equity and fairness

- PROGRESS-Plus identifies social stratifiers, such as age, ethnicity, language, deprivation, and disability [17].
- Health Equity Impact Assessment tool evaluates potential equity impacts before deployment [52].
- The SIITHIA (Strengthening the Integration of Intersectionality Theory in Health Inequality Analysis) checklist and the digital health equity framework support intersectional and digital-specific equity analysis [53,54].

The frameworks presented in [Textbox 1](#) allow AI triage to be evaluated across safety, implementation, and equity dimensions. Combining them enables a more comprehensive assessment than any single lens can provide.

[Textbox 2](#) provides an illustrative example of how to report equal opportunity (true positive rate; sensitivity) across intersectional PROGRESS-Plus strata.

**Textbox 2.** Illustrative example: reporting equal opportunity across intersectional strata.

- Equal opportunity requires similar true positive rates (TPRs; sensitivity) across groups for individuals who truly need urgent care.
- Step 1 includes defining the safety-critical outcome ( $Y=1$ ), for example, “urgent same-day clinical assessment required” based on a reference standard (eg, clinician adjudication, emergency department attendance within 24 to 48 hours, or diagnosis of a time-critical condition).
- Step 2 involves computing TPR within each subgroup.
- $TPR \text{ (sensitivity)} = \text{true positive (TP)} / (\text{TP} + \text{false negative [FN]})$
- Step 3 involves reporting TPR across PROGRESS-Plus strata and intersectional strata, for example, age group×ethnicity or ethnicity×deprivation quintile. An example of reporting is as follows:
  - White, least deprived:  $TP=180; FN=20 \rightarrow TPR=0.90$
  - White, most deprived:  $TP=150; FN=30 \rightarrow TPR=0.83$
  - Minority ethnicity, least deprived:  $TP=70; FN=20 \rightarrow TPR=0.78$
  - Minority ethnicity, most deprived:  $TP=55; FN=25 \rightarrow TPR=0.69$
- Step 3a involves quantifying uncertainty in subgroup estimates. For each subgroup, TPRs should be reported with measures of uncertainty (eg, 95% CIs or standard errors), particularly where subgroup sample sizes are small. This enables assessment of whether observed differences are robust or compatible with random variation.
- Step 4 summarizes the disparity. This approach makes equity risks visible that would be hidden in overall performance metrics.
- Absolute gap (maximum TPR – minimum TPR) =  $0.90 - 0.69 = 0.21$
- Flag threshold: investigate if the gap is greater than 0.05 or if any subgroup TPR is less than 0.80
- These figures are illustrative. In practice, equity assessments should consider statistical uncertainty (eg, CI overlap and subgroup sample size) alongside point estimates. Empirical data on equity-stratified artificial intelligence triage performance in primary care remain limited.

Several dimensions of bias have been documented or are anticipated in an AI-driven triage system, as mentioned subsequently.

The first dimension is age. Older adults are frequently underrepresented in development datasets, increasing the risk of atypical presentations or those with limited digital literacy [58,59].

The second dimension is language and ethnicity. NLP models are extremely sensitive to linguistic variation, dialects, multilingual input, or limited English proficiency. However, model evaluations rarely account for these, threatening safe triage in diverse populations [60].

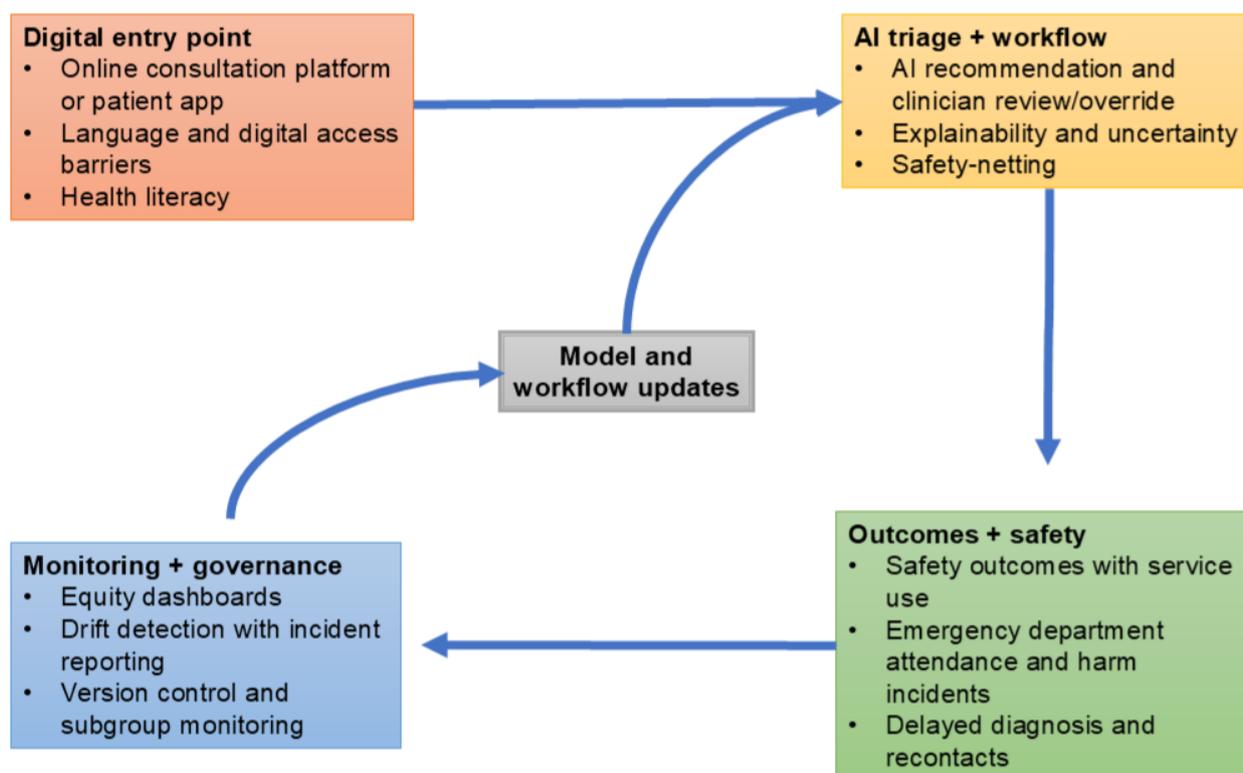
Furthermore, broader AI research (outside primary care) shows that racial and ethnic biases in algorithmic systems persist [61]. For instance, algorithms that relied on health care cost as a proxy for illness systematically undertriaged Black patients due to unequal access-driven cost differences [60]. Similarly, AI in imaging frequently underdiagnoses emergent pathology in marginalized groups. Black women have shown significantly higher underdiagnosis rates in medical imaging models.

Telephone or digital triage evaluations suggest that low-income individuals, ethnic minority groups, and displaced patients experience worse outcomes, though quantitative data remain sparse [62].

Single-axis analysis (age or ethnicity) is insufficient. Intersecting vulnerabilities (eg, older adults from minority ethnic backgrounds with language barriers) can compound risk and increase mistriage. However, only a limited number of studies disaggregate safety performance by intersectional subgroups, leaving some of the most disadvantaged populations effectively invisible in assessments. Such analyses often lack power; therefore, findings should be treated as exploratory unless supported by large, multisite cohorts.

## *What Must Change: a Research and Policy Agenda*

In this viewpoint paper, we argue that AI triage stands at a crossroad; it has the potential to improve safety and access in primary care but requires real-world evaluation, equity-focused monitoring, and sociotechnical governance. [Figure 2](#) summarizes the proposed real-world evaluation and governance loop for AI triage in GP. Digital entry points feed into AI-supported triage and clinician workflow, producing safety and service-use outcomes that inform monitoring and governance (including equity dashboards, drift detection, and incident reporting). Governance outputs drive model and workflow updates, enabling continuous improvement.

**Figure 2.** Conceptual loop for evaluating and governing artificial intelligence (AI)-enabled triage in general practice as a sociotechnical intervention.

To shift AI triage from a hypothetical promise to an equitable, safe reality in GP and primary care settings, we propose 5 interrelated priorities, as mentioned subsequently.

The first priority is real-world evaluations in primary care (prospective or retrospective). Current evidence is dominated by vignette experiments or ED contexts. Prospective, real-world evaluations—such as controlled interrupted time series or (cluster) randomized controlled trials (RCTs)—that assess patient safety outcomes (eg, delayed diagnoses and avoidable emergency use), workflow effects, and override behaviors in GP practice are urgently needed. Where randomization is infeasible, controlled interrupted time series with matched controls and prespecified safety outcomes can provide strong quasi-experimental evidence.

The second priority is equity-stratified performance reporting. AI triage systems should be evaluated through the equity lenses, such as PROGRESS-Plus and other fairness metrics. This means disaggregating performance (TPR, FPR, and calibration) by age, ethnicity, language, deprivation, and their multidimensional aspects to identify and mitigate differential risks. Without such reporting, disparities will remain hidden.

The third priority is causal evaluation designs. Observational signals are helpful, but causality demands rigorous designs. Interrupted time series around AI deployment and RCTs, where feasible (to attribute safety effects more definitively to AI interventions), and propensity score methods, instrumental variables, or target trial emulation, where RCTs are infeasible, should be conducted.

The fourth priority is postmarket surveillance and governance infrastructure. Effective governance is not optional.

Organizations should adopt frameworks such as people, process, technology, and operations to ensure structured oversight across personnel, process, technology, and operations. AI triage requires continuous monitoring, with statistically principled detection of drift, subgroup miscalibration, and emerging hazards [63].

The fifth priority is human-AI collaboration and implementation research. Research should shift from algorithm-centric evaluation to sociotechnical integration. Studies should examine how clinicians interpret, override, or trust AI suggestions; how AI supports (rather than disrupts) workflow; and how organizational culture shapes safe AI adoption. Mixed methods research combining qualitative insights with quantitative safety metrics will be critical.

Prospective evaluations in GP are challenging but feasible. Cluster-randomized trials and interrupted time series require careful handling of contamination between clinicians, cointerventions during rollout, seasonal and demand shocks, and variation in practice digital maturity. Outcome measurement also depends on data linkage (eg, EHR, urgent care, ED attendance, and diagnostic follow-up), and governance processes can slow implementation. Although such designs remain underused for AI-enabled triage, they are well-established in evaluating complex service interventions in primary care and are methodologically appropriate for this context [64,65].

## Conclusions

AI triage offers potential for improving primary care efficiency, safety, and consistency, but current evidence leaves critical gaps. Without intersectional, real-world safety evaluations,

implementation is not just uncertain; it may be ethically risky and may inadvertently magnify existing health inequities. Over the coming years, this field must commit to responsible, equity-focused, system-aware evidence generation. That means embedding AI evaluation within the disordered realities of practice, governance mechanisms that ensure fairness and transparency, and human-AI systems that augment care rather than add workload. Operationally, this requires three commitments: (1) prospective, real-world evaluations in GP;

(2) equity-stratified performance reporting guided by frameworks, such as PROGRESS-Plus; and (3) rigorous postmarket surveillance with drift and subgroup monitoring and WHO International Classification for Patient Safety-aligned incident reporting. Without these, deployment risks amplify inequities rather than reducing them. With these commitments, AI triage can better deliver on its potential of safer, more equitable primary care.

## Acknowledgments

Generative artificial intelligence (ChatGPT, version 5.2; OpenAI) was used to support language editing for clarity and readability. All content was reviewed and revised by the authors, who take full responsibility for the final manuscript.

## Funding

This work received no specific funding. EK is partly funded by the National Institute for Health and Care Research HealthTech Research Centre in Emergency and Acute Care (grant NIHR205301) and the Manchester British Heart Foundation Centre for Research Excellence (grant RE/24/130017).

## Conflicts of Interest

None declared.

## References

1. Sidaway-Lee K, Pereira Gray SD, Khan N, Abraham L, Evans P. GP continuity: the keystone of general practice. *InnovAiT*. Apr 24, 2024;17(7):313-320. [doi: [10.1177/17557380241246742](https://doi.org/10.1177/17557380241246742)]
2. Hobbs FD, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. *Lancet*. Jun 04, 2016;387(10035):2323-2330. [FREE Full text] [doi: [10.1016/S0140-6736\(16\)00620-6](https://doi.org/10.1016/S0140-6736(16)00620-6)] [Medline: [27059888](https://pubmed.ncbi.nlm.nih.gov/27059888/)]
3. The NHS long term plan. NHS England. Jan 2019. URL: <https://www.longtermplan.nhs.uk/> [accessed 2025-05-25]
4. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for COVID-19. *BMJ*. Mar 12, 2020;368:m998. [doi: [10.1136/bmj.m998](https://doi.org/10.1136/bmj.m998)] [Medline: [32165352](https://pubmed.ncbi.nlm.nih.gov/32165352/)]
5. Reidy C, Papoutsi C, Kc S, Gudgin B, Laverty AA, Greaves F, et al. Qualitative evaluation of the implementation and national roll-out of the NHS App in England. *BMC Med*. Jan 21, 2025;23(1):20. [FREE Full text] [doi: [10.1186/s12916-024-03842-w](https://doi.org/10.1186/s12916-024-03842-w)] [Medline: [39838384](https://pubmed.ncbi.nlm.nih.gov/39838384/)]
6. Submissions via online consultation systems in general practice. NHS England. URL: <https://digital.nhs.uk/data-and-information/publications/statistical/submissions-via-online-consultation-systems-in-general-practice> [accessed 2026-01-24]
7. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. *BJGP Open*. Jun 2021;5(3):040. [FREE Full text] [doi: [10.3399/BJGPO.2021.0040](https://doi.org/10.3399/BJGPO.2021.0040)] [Medline: [33712502](https://pubmed.ncbi.nlm.nih.gov/33712502/)]
8. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ Open*. Aug 01, 2019;9(8):e027743. [FREE Full text] [doi: [10.1136/bmjopen-2018-027743](https://doi.org/10.1136/bmjopen-2018-027743)] [Medline: [31375610](https://pubmed.ncbi.nlm.nih.gov/31375610/)]
9. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc*. 2003;10(6):523-530. [FREE Full text] [doi: [10.1197/jamia.M1370](https://doi.org/10.1197/jamia.M1370)] [Medline: [12925543](https://pubmed.ncbi.nlm.nih.gov/12925543/)]
10. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. Apr 02, 2005;330(7494):765. [FREE Full text] [doi: [10.1136/bmj.38398.500764.8F](https://doi.org/10.1136/bmj.38398.500764.8F)] [Medline: [15767266](https://pubmed.ncbi.nlm.nih.gov/15767266/)]
11. Technologies T. Technologies to address wait times in the emergency department. *Health Technologies*, National Library of Medicine. Jul 2025;EN0058. [FREE Full text] [Medline: [40857378](https://pubmed.ncbi.nlm.nih.gov/40857378/)]
12. Abualruz H, Yasin I, Abu Sabra MA, Abunab HY, Azayzeh R, Zubidi Y, et al. The role of artificial intelligence in enhancing triage decisions in healthcare settings: a systematic review. *Appl Nurs Res*. Dec 2025;86:152024. [doi: [10.1016/j.apnr.2025.152024](https://doi.org/10.1016/j.apnr.2025.152024)] [Medline: [41330654](https://pubmed.ncbi.nlm.nih.gov/41330654/)]
13. Peta D, Day A, Lugari WS, Gorman V, Ahayalimudin N, Pajo VM. Triage: a global perspective. *J Emerg Nurs*. Nov 2023;49(6):814-825. [doi: [10.1016/j.jen.2023.08.004](https://doi.org/10.1016/j.jen.2023.08.004)] [Medline: [37925222](https://pubmed.ncbi.nlm.nih.gov/37925222/)]

14. Tahernejad A, Sahebi A, Abadi AS, Safari M. Application of artificial intelligence in triage in emergencies and disasters: a systematic review. *BMC Public Health*. Nov 18, 2024;24(1):3203. [FREE Full text] [doi: [10.1186/s12889-024-20447-3](https://doi.org/10.1186/s12889-024-20447-3)] [Medline: [39558305](https://pubmed.ncbi.nlm.nih.gov/39558305/)]
15. Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf*. Feb 2015;24(2):103-110. [FREE Full text] [doi: [10.1136/bmjqs-2014-003675](https://doi.org/10.1136/bmjqs-2014-003675)] [Medline: [25589094](https://pubmed.ncbi.nlm.nih.gov/25589094/)]
16. Smart C, Newman C, Hartill L, Bunce S, McCormick J. Workload effects of online consultation implementation from a job-characteristics model perspective: a qualitative study. *BJGP Open*. Nov 21, 2022;7(1):BJGPO.2022.0024. [doi: [10.3399/bjgpo.2022.0024](https://doi.org/10.3399/bjgpo.2022.0024)]
17. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *J Clin Epidemiol*. Jan 2014;67(1):56-64. [doi: [10.1016/j.jclinepi.2013.08.005](https://doi.org/10.1016/j.jclinepi.2013.08.005)] [Medline: [24189091](https://pubmed.ncbi.nlm.nih.gov/24189091/)]
18. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. Jan 7, 2019;25(1):24-29. [doi: [10.1038/s41591-018-0316-z](https://doi.org/10.1038/s41591-018-0316-z)] [Medline: [30617335](https://pubmed.ncbi.nlm.nih.gov/30617335/)]
19. Barocas S, Hardt M, Narayanan A. *Fairness And Machine Learning: Limitations And Opportunities*. Cambridge, MA. The MIT Press; 2023.
20. Akter S, Dwivedi YK, Sajib S, Biswas K, Bandara RJ, Michael K. Algorithmic bias in machine learning-based marketing models. *J Bus Res*. May 2022;144:201-216. [doi: [10.1016/j.jbusres.2022.01.083](https://doi.org/10.1016/j.jbusres.2022.01.083)]
21. Jain A, Brooks JR, Alford CC, Chang CS, Mueller NM, Umscheid CA, et al. Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms. *JAMA Health Forum*. Jun 02, 2023;4(6):e231197. [FREE Full text] [doi: [10.1001/jamahealthforum.2023.1197](https://doi.org/10.1001/jamahealthforum.2023.1197)] [Medline: [37266959](https://pubmed.ncbi.nlm.nih.gov/37266959/)]
22. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci U S A*. Jun 09, 2020;117(23):12592-12594. [FREE Full text] [doi: [10.1073/pnas.1919012117](https://doi.org/10.1073/pnas.1919012117)] [Medline: [32457147](https://pubmed.ncbi.nlm.nih.gov/32457147/)]
23. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. Oct 2019;1(6):e271-e297. [FREE Full text] [doi: [10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)] [Medline: [33323251](https://pubmed.ncbi.nlm.nih.gov/33323251/)]
24. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. Mar 25, 2020;368:m689. [FREE Full text] [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)] [Medline: [32213531](https://pubmed.ncbi.nlm.nih.gov/32213531/)]
25. Delshad S, Dontaraju VS, Chengat V. Artificial intelligence-based application provides accurate medical triage advice when compared to consensus decisions of healthcare providers. *Cureus*. Aug 2021;13(8):e16956. [FREE Full text] [doi: [10.7759/cureus.16956](https://doi.org/10.7759/cureus.16956)] [Medline: [34405077](https://pubmed.ncbi.nlm.nih.gov/34405077/)]
26. Ivanov O, Wolf L, Brecher D, Lewis E, Masek K, Montgomery K, et al. Improving ED emergency severity index acuity assignment using machine learning and clinical natural language processing. *J Emerg Nurs*. Mar 2021;47(2):265-78.e7. [FREE Full text] [doi: [10.1016/j.jen.2020.11.001](https://doi.org/10.1016/j.jen.2020.11.001)] [Medline: [33358394](https://pubmed.ncbi.nlm.nih.gov/33358394/)]
27. Tyler S, Olis M, Aust N, Patel L, Simon L, Triantafyllidis C, et al. Use of artificial intelligence in triage in hospital emergency departments: a scoping review. *Cureus*. May 2024;16(5):e59906. [FREE Full text] [doi: [10.7759/cureus.59906](https://doi.org/10.7759/cureus.59906)] [Medline: [38854295](https://pubmed.ncbi.nlm.nih.gov/38854295/)]
28. Bernardi FA, Alves D, Crepaldi N, Yamada DB, Lima VC, Rijo R. Data quality in health research: integrative literature review. *J Med Internet Res*. Oct 31, 2023;25:e41446. [FREE Full text] [doi: [10.2196/41446](https://doi.org/10.2196/41446)] [Medline: [37906223](https://pubmed.ncbi.nlm.nih.gov/37906223/)]
29. Wise J. Electronic consultations offer few benefits for GP practices, says study. *BMJ*. Nov 06, 2017;j5141. [doi: [10.1136/bmj.j5141](https://doi.org/10.1136/bmj.j5141)]
30. Karran EL, Cashin AG, Barker T, Boyd MA, Chiarotto A, Dewidar O, et al. Using PROGRESS-plus to identify current approaches to the collection and reporting of equity-relevant data: a scoping review. *J Clin Epidemiol*. Nov 2023;163:70-78. [FREE Full text] [doi: [10.1016/j.jclinepi.2023.09.017](https://doi.org/10.1016/j.jclinepi.2023.09.017)] [Medline: [37802205](https://pubmed.ncbi.nlm.nih.gov/37802205/)]
31. Steerling E, Svedberg P, Nilsen P, Siira E, Nygren J. Influences on trust in the use of AI-based triage-an interview study with primary healthcare professionals and patients in Sweden. *Front Digit Health*. May 20, 2025;7:1565080. [FREE Full text] [doi: [10.3389/fdgth.2025.1565080](https://doi.org/10.3389/fdgth.2025.1565080)] [Medline: [40463579](https://pubmed.ncbi.nlm.nih.gov/40463579/)]
32. Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an international classification for patient safety: key concepts and terms. *Int J Qual Health Care*. Feb 2009;21(1):18-26. [FREE Full text] [doi: [10.1093/intqhc/mzn057](https://doi.org/10.1093/intqhc/mzn057)] [Medline: [19147597](https://pubmed.ncbi.nlm.nih.gov/19147597/)]
33. Mahlkecht A, Engl A, Piccoliori G, Wiedermann CJ. Supporting primary care through symptom checking artificial intelligence: a study of patient and physician attitudes in Italian general practice. *BMC Prim Care*. Sep 04, 2023;24(1):174. [FREE Full text] [doi: [10.1186/s12875-023-02143-0](https://doi.org/10.1186/s12875-023-02143-0)] [Medline: [37661285](https://pubmed.ncbi.nlm.nih.gov/37661285/)]
34. Riboli-Sasco E, El-Osta A, Alaa A, Webber I, Karki M, El Asmar ML, et al. Triage and diagnostic accuracy of online symptom checkers: systematic review. *J Med Internet Res*. Jun 02, 2023;25:e43803. [FREE Full text] [doi: [10.2196/43803](https://doi.org/10.2196/43803)] [Medline: [37266983](https://pubmed.ncbi.nlm.nih.gov/37266983/)]

35. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* Aug 07, 2009;4:50. [FREE Full text] [doi: [10.1186/1748-5908-4-50](https://doi.org/10.1186/1748-5908-4-50)] [Medline: [19664226](https://pubmed.ncbi.nlm.nih.gov/19664226/)]
36. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res.* Nov 01, 2017;19(11):e367. [FREE Full text] [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
37. Tuzlalı M, Baki N, Aral K, Aral CA, Bahçe E. Evaluating the performance of AI chatbots in responding to dental implant FAQs: a comparative study. *BMC Oral Health.* Oct 08, 2025;25(1):1548. [FREE Full text] [doi: [10.1186/s12903-025-06863-w](https://doi.org/10.1186/s12903-025-06863-w)] [Medline: [41063105](https://pubmed.ncbi.nlm.nih.gov/41063105/)]
38. Kaushik R, Rapaka R. A patient-centered perspectives and future directions in AI-powered teledentistry. *Discoveries (Craiova).* 2024;12(4):e199. [doi: [10.15190/d.2024.18](https://doi.org/10.15190/d.2024.18)] [Medline: [40109877](https://pubmed.ncbi.nlm.nih.gov/40109877/)]
39. Chen K, Qiu L, Xie X, Bai Y. Dynasmile: Video-based smile analysis software in orthodontics. *SoftwareX.* Feb 2025;29:102004. [doi: [10.1016/j.softx.2024.102004](https://doi.org/10.1016/j.softx.2024.102004)]
40. Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. *Healthcare (Basel).* Aug 29, 2025;13(17):2154. [FREE Full text] [doi: [10.3390/healthcare13172154](https://doi.org/10.3390/healthcare13172154)] [Medline: [40941506](https://pubmed.ncbi.nlm.nih.gov/40941506/)]
41. Abdelwanis M, Alarafati HK, Tammam MM, Simsekler MC. Exploring the risks of automation bias in healthcare artificial intelligence applications: a Bowtie analysis. *J Saf Sci Resil.* Dec 2024;5(4):460-469. [doi: [10.1016/j.jnlssr.2024.06.001](https://doi.org/10.1016/j.jnlssr.2024.06.001)]
42. Salimparsa M, Sedig K, Lizotte DJ, Abdullah SS, Chalabianloo N, Muanda FT. Explainable AI for clinical decision support systems: literature review, key gaps, and research synthesis. *Informatics.* Oct 28, 2025;12(4):119. [doi: [10.3390/informatics12040119](https://doi.org/10.3390/informatics12040119)]
43. Ellertsson S, Hlynsson HD, Loftsson H, Sigur Sson EL. Triaging patients with artificial intelligence for respiratory symptoms in primary care to improve patient outcomes: a retrospective diagnostic accuracy study. *Ann Fam Med.* 2023;21(3):240-248. [FREE Full text] [doi: [10.1370/afm.2970](https://doi.org/10.1370/afm.2970)] [Medline: [37217331](https://pubmed.ncbi.nlm.nih.gov/37217331/)]
44. Vergara PO, de Oliveira JC, Mattiello R, Montelongo A, Roman R, Katz N, et al. Accuracy of artificial intelligence for gatekeeping in referrals to specialized care. *JAMA Netw Open.* Jun 02, 2025;8(6):e2513285. [FREE Full text] [doi: [10.1001/jamanetworkopen.2025.13285](https://doi.org/10.1001/jamanetworkopen.2025.13285)] [Medline: [40459894](https://pubmed.ncbi.nlm.nih.gov/40459894/)]
45. Wallace W, Chan C, Chidambaram S, Hanna L, Acharya A, Daniels E, et al. Evaluating the diagnostic and triage performance of digital and online symptom checkers for the presentation of myocardial infarction: a retrospective cross-sectional study. *PLOS Digit Health.* Aug 2024;3(8):e0000558. [doi: [10.1371/journal.pdig.0000558](https://doi.org/10.1371/journal.pdig.0000558)] [Medline: [39102377](https://pubmed.ncbi.nlm.nih.gov/39102377/)]
46. Siira E, Tyskbo D, Nygren J. Healthcare leaders' experiences of implementing artificial intelligence for medical history-taking and triage in Swedish primary care: an interview study. *BMC Prim Care.* Jul 24, 2024;25(1):268. [doi: [10.1186/s12875-024-02516-z](https://doi.org/10.1186/s12875-024-02516-z)] [Medline: [39048973](https://pubmed.ncbi.nlm.nih.gov/39048973/)]
47. Moschogianis S, Darley S, Coulson T, Peek N, Cheraghi-Sohi S, Brown B. Seven opportunities for artificial intelligence in primary care electronic visits: qualitative study of staff and patient views. *Ann Fam Med.* May 27, 2025;23(3):214-222. [FREE Full text] [doi: [10.1370/afm.240292](https://doi.org/10.1370/afm.240292)] [Medline: [40425478](https://pubmed.ncbi.nlm.nih.gov/40425478/)]
48. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics.* 2013;56(11):1669-1686. [FREE Full text] [doi: [10.1080/00140139.2013.838643](https://doi.org/10.1080/00140139.2013.838643)] [Medline: [24088063](https://pubmed.ncbi.nlm.nih.gov/24088063/)]
49. Patient safety: making health care safer. World Health Organization (WHO). 2017. URL: <https://www.who.int/publications/i/item/WHO-HIS-SDS-2017.11> [accessed 2025-07-14]
50. Lobach D, Sanders GD, Bright TJ, Wong A, Dhurjati R, Bristow E, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep).* Apr 2012;(203):1-784. [Medline: [23126650](https://pubmed.ncbi.nlm.nih.gov/23126650/)]
51. Darley S, Coulson T, Peek N, Moschogianis S, van der Veer SN, Wong DC, et al. Understanding how the design and implementation of online consultations affect primary care quality: systematic review of evidence with recommendations for designers, providers, and researchers. *J Med Internet Res.* Oct 24, 2022;24(10):e37436. [FREE Full text] [doi: [10.2196/37436](https://doi.org/10.2196/37436)] [Medline: [36279172](https://pubmed.ncbi.nlm.nih.gov/36279172/)]
52. Olyaeemanesh A, Takian A, Mostafavi H, Mobinzadeh M, Bakhtiari A, Yaftian F, et al. Health Equity Impact Assessment (HEIA) reporting tool: developing a checklist for policymakers. *Int J Equity Health.* Nov 18, 2023;22(1):241. [FREE Full text] [doi: [10.1186/s12939-023-02031-0](https://doi.org/10.1186/s12939-023-02031-0)] [Medline: [37980523](https://pubmed.ncbi.nlm.nih.gov/37980523/)]
53. Government of Canada. How to integrate intersectionality theory in quantitative health equity analysis? A rapid review and checklist of promising practices. Public Health Agency of Canada. Ottawa, ON. URL: <https://www.canada.ca/en/public-health/services/publications/science-research-data/how-integrate-intersectionality-theory-quantitative-health-equity-analysis.html> [accessed 2026-01-24]
54. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. *NPJ Digit Med.* Aug 18, 2022;5(1):119. [FREE Full text] [doi: [10.1038/s41746-022-00663-0](https://doi.org/10.1038/s41746-022-00663-0)] [Medline: [35982146](https://pubmed.ncbi.nlm.nih.gov/35982146/)]
55. Carayon P. Sociotechnical systems approach to healthcare quality and patient safety. *Work.* 2012;41 Suppl 1(0 1):3850-3854. [FREE Full text] [doi: [10.3233/WOR-2012-0091-3850](https://doi.org/10.3233/WOR-2012-0091-3850)] [Medline: [22317309](https://pubmed.ncbi.nlm.nih.gov/22317309/)]

56. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. Sep 1999;89(9):1322-1327. [doi: [10.2105/ajph.89.9.1322](https://doi.org/10.2105/ajph.89.9.1322)] [Medline: [10474547](https://pubmed.ncbi.nlm.nih.gov/10474547/)]
57. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for health information systems: human, organization and technology-fit factors (HOT-fit). *Int J Med Inform*. Jun 2008;77(6):386-398. [doi: [10.1016/j.ijmedinf.2007.08.011](https://doi.org/10.1016/j.ijmedinf.2007.08.011)] [Medline: [17964851](https://pubmed.ncbi.nlm.nih.gov/17964851/)]
58. Chu CH, Donato-Woodger S, Khan SS, Nyrupe R, Leslie K, Lyn A, et al. Age-related bias and artificial intelligence: a scoping review. *Humanit Soc Sci Commun*. Aug 17, 2023;10(1):510. [doi: [10.1057/s41599-023-01999-y](https://doi.org/10.1057/s41599-023-01999-y)]
59. Shiwani T, Relton S, Evans R, Kale A, Heaven A, Clegg A, Ageing Data Research Collaborative (Geridata) AI group, et al. New Horizons in artificial intelligence in the healthcare of older people. *Age Ageing*. Dec 01, 2023;52(12):afad219. [FREE Full text] [doi: [10.1093/ageing/afad219](https://doi.org/10.1093/ageing/afad219)] [Medline: [38124256](https://pubmed.ncbi.nlm.nih.gov/38124256/)]
60. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. Oct 25, 2019;366(6464):447-453. [FREE Full text] [doi: [10.1126/science.aax2342](https://doi.org/10.1126/science.aax2342)] [Medline: [31649194](https://pubmed.ncbi.nlm.nih.gov/31649194/)]
61. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. Apr 04, 2019;380(14):1347-1358. [doi: [10.1056/nejmra1814259](https://doi.org/10.1056/nejmra1814259)]
62. Williams C, Shang D. Telehealth usage among low-income racial and ethnic minority populations during the COVID-19 pandemic: retrospective observational study. *J Med Internet Res*. May 12, 2023;25:e43604. [FREE Full text] [doi: [10.2196/43604](https://doi.org/10.2196/43604)] [Medline: [37171848](https://pubmed.ncbi.nlm.nih.gov/37171848/)]
63. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. *Bioengineering (Basel)*. Mar 29, 2024;11(4):337. [FREE Full text] [doi: [10.3390/bioengineering11040337](https://doi.org/10.3390/bioengineering11040337)] [Medline: [38671759](https://pubmed.ncbi.nlm.nih.gov/38671759/)]
64. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. Feb 01, 2017;46(1):348-355. [FREE Full text] [doi: [10.1093/ije/dyw098](https://doi.org/10.1093/ije/dyw098)] [Medline: [27283160](https://pubmed.ncbi.nlm.nih.gov/27283160/)]
65. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. Sep 29, 2008;337:a1655. [FREE Full text] [doi: [10.1136/bmj.a1655](https://doi.org/10.1136/bmj.a1655)] [Medline: [18824488](https://pubmed.ncbi.nlm.nih.gov/18824488/)]

## Abbreviations

**AI:** artificial intelligence  
**CDSS:** clinical decision support system  
**ED:** emergency department  
**EHR:** electronic health record  
**FPR:** false-positive rate  
**GP:** general practice  
**ML:** machine learning  
**NLP:** natural language processing  
**RCT:** randomized controlled trial  
**SIITHIA:** Strengthening the Integration of Intersectionality Theory in Health Inequality Analysis  
**TPR:** true-positive rate  
**WHO:** World Health Organization

*Edited by A Mavragani; submitted 24.Nov.2025; peer-reviewed by KMS Islam, M Kokash, J Grosser, K Chen; comments to author 07.Jan.2026; accepted 28.Jan.2026; published 04.Mar.2026*

*Please cite as:*

Alamoudi A, Kontopantelis E, Zghebi SS, Brown B  
*AI Triage in Primary Care: Building Safer and More Equitable Real-World Evidence*  
*J Med Internet Res* 2026;28:e88396  
URL: <https://www.jmir.org/2026/1/e88396>  
doi: [10.2196/88396](https://doi.org/10.2196/88396)  
PMID:

©Aymn Alamoudi, Evangelos Kontopantelis, Salwa S Zghebi, Benjamin Brown. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 04.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN

1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.