SITORA

AI as the NHS Front Door

Safety, Accountability and Public Trust in AI Triage

A Sitora policy report for policymakers, NHS leaders, regulators and public-sector AI suppliers

Prepared by Sitora
July 2026
Draft for policy engagement and stakeholder circulation

Contents

1. Executive summary
2. Purpose and scope
3. Background: the NHS AI triage shift
4. Evidence base and current regulatory landscape
5. Risk analysis
6. The Sitora NHS AI Trust Test
7. Policy recommendations
8. Implementation roadmap
9. Stakeholder map
10. Conclusion

Appendix A: Risk register

Appendix B: Assurance checklist

References

1. Executive summary

The NHS is accelerating the use of artificial intelligence across health services, including AI-enabled triage inside the NHS App to direct patients to the most appropriate NHS service. NHS England has described this as part of a wider technology, digital and data overhaul, funded by £10 billion over three years. [1]

Sitora supports responsible AI adoption in healthcare. AI can reduce administrative pressure, improve navigation, shorten queues and help patients reach appropriate services faster. However, AI triage is not a normal app feature. When a tool directs a patient towards a GP, pharmacy, A&E, self-care or delayed appointment, it becomes part of the clinical access pathway.

The core concern is not whether the NHS should use AI. The core concern is whether public healthcare AI is sufficiently safe, independently audited, clinically accountable, transparent, inclusive and continuously monitored before it is used at scale.

This report argues that the NHS needs a clear, public-facing AI triage assurance model. The public should not be asked to trust a black box at the front door of care. They should be able to trust that the system has been properly assessed, that high-risk and uncertain cases can escalate to a human, that patient data is protected, and that accountability is defined before harm occurs.

Sitora position in one line

AI can help the NHS, but AI triage must be clinically safe, independently audited, transparent, accessible and accountable before it becomes the front door to public healthcare.

Priority asks

Policy ask	What it means in practice
Mandatory AI impact assessments	Every NHS AI triage deployment should assess clinical risk, equality, accessibility, privacy, cyber risk, procurement risk and patient safety.
Independent audit before scale	AI triage should not rely only on vendor assurances. Independent review should test safety, bias, explainability, performance and failure modes.
Human escalation guarantee	Patients must have a clear route to human review where symptoms are serious, uncertain, worsening, complex, or where the patient disagrees with the recommendation.
Clear accountability	Government and NHS England should define responsibility between NHS bodies, clinicians, commissioners and suppliers before rollout.
Ongoing monitoring	AI triage should have post-deployment monitoring for incidents, near misses, patient complaints, model drift and demographic disparities.

2. Purpose and scope

This report has been prepared as a practical lobbying and policy document. It is intended for MPs, parliamentary committees, the Department of Health and Social Care, NHS England, regulators, patient organisations, health technology suppliers and public-sector decision makers.

The report focuses on AI triage in NHS patient access pathways, especially where AI helps direct patients through the NHS App or other digital entry points. It does not attempt to oppose NHS digital transformation. Instead, it sets out the safeguards needed for responsible adoption.

The report draws on official NHS, UK regulator, NICE, MHRA, ICO, CQC, WHO and peer-reviewed sources. It also uses recent reporting on the NHS App AI triage rollout as the trigger for policy engagement. News reporting is used only to describe the policy context; governance recommendations are based primarily on official and regulatory sources.

3. Background: the NHS AI triage shift

NHS England has announced that a new AI triage tool in the NHS App will help direct patients to the most appropriate NHS service. The same announcement references wider use of AI tools, including AI notetaking, as part of a major technology and data programme across England. [1]

The NHS App is increasingly being positioned as a central digital route into care. In its elective care reform plan, NHS England described the NHS App as the 'single, comprehensive and trusted digital front door to care for patients'. [2] This language matters because a front door controls access. If an AI tool influences access, then it must meet a higher governance threshold than an ordinary digital convenience feature.

Recent reporting has also described AI triage integration into the NHS App allowing large numbers of patients to book appointments directly with clinicians. [3] Separately, reporting on digital triage in A&E settings has highlighted potential benefits around waiting-time reduction, but also concerns around accessibility and clear guidance for redirected patients. [4]

The pressure on NHS access is real. GP appointment demand, A&E crowding, staffing pressures and administrative workload all create strong reasons to explore technology. However, the response cannot be to scale clinical access automation without transparent safeguards.

Triage decisions influence urgency and routing. A poor recommendation can create delay, false reassurance or inappropriate redirection. A safe governance model must therefore treat AI triage as a patient-safety issue, an equality issue, a data-protection issue and an accountability issue at the same time.

4. Evidence base and current regulatory landscape

4.1 NHS digital clinical safety standards

NHS England's digital clinical safety assurance guidance explains that the hazard log is a formal document under DCB0129 and DCB0160. It records hazards attributable to a product, potential clinical impacts, mitigations, controls and final assessment of risk. [5]

NHS England has commenced a review of digital clinical safety standards DCB0129 and DCB0160 to ensure they remain up to date and aligned with advances in healthcare technology and clinical practice. [6] This review is directly relevant to AI because probabilistic and adaptive systems create risks that conventional software standards may not capture fully.

4.2 MHRA: software and AI as a medical device

The MHRA states that software and AI play an essential part in health and social care and that many such products are regulated as medical devices or in vitro diagnostic medical devices in the UK. [7] For AI triage, classification depends on intended purpose and clinical function. If the tool influences patient routing or clinical decision-making, its regulatory status should be assessed explicitly rather than assumed.

4.3 NICE evidence standards

NICE provides an Evidence Standards Framework for digital health technologies. NHS England's Accelerated Access Collaborative has incorporated this framework into its work on digital health technologies. [8] AI triage systems should be mapped to appropriate evidence tiers, with stronger evidence expected where technology has direct consequences for patient access or clinical risk.

4.4 ICO and data protection

The ICO's AI and data protection guidance explains best practice for data-protection-compliant AI and how data protection law applies to AI systems processing personal data. [9] Health triage involves particularly sensitive information, including symptoms, medical history, risk factors and potentially protected characteristics. Any NHS AI triage deployment must therefore be designed around fairness, transparency, data minimisation and accountability.

4.5 CQC role and expectations

The CQC has set out its role and expectations in relation to AI in health and social care, acknowledging both benefits and risks and connecting AI use to safe, equitable and person-centred care. [10] This is important because AI may affect regulated care quality even when the software is supplied by a third party.

4.6 WHO ethics and governance

The World Health Organization's guidance on ethics and governance of AI for health identifies ethical challenges and risks, and sets out consensus principles to ensure AI works for the public benefit. [11] WHO's later guidance on large multi-modal models reinforces the need for governance by governments, technology companies and healthcare providers. [12]

4.7 AI and Digital Regulations Service

The AI and Digital Regulations Service, associated with NICE and other UK health bodies, is intended to help the NHS and wider care system identify, pilot and roll out impactful AI and digital technologies. [13] Its existence supports the principle that health AI adoption requires coordinated regulatory guidance rather than ad hoc procurement decisions.

4.8 UK pro-innovation AI regulation

The UK government's pro-innovation AI regulation framework is built around principles including safety, transparency, fairness, accountability and contestability. [14] Sitora's position is aligned with responsible innovation: the NHS should move quickly where benefits are clear, but should not treat public healthcare AI as exempt from rigorous assurance.

Evidence map

Source area	Relevance to NHS AI triage	Implication
NHS digital clinical safety	Clinical safety case, hazard log, DCB0129/DCB0160	AI triage needs explicit hazard analysis and clinical risk controls.
MHRA	Software/AI may be regulated as medical device depending on intended purpose	Regulatory classification must be documented before procurement and rollout.
NICE ESF	Evidence standards for digital health technologies	Higher-risk triage tools need stronger evidence of effectiveness and safety.
ICO	AI and personal data; fairness and transparency	Patient-facing AI must explain data use and meet UK GDPR obligations.
CQC	Safe, equitable and person-centred care	AI use should be assessed as part of quality and safety, not treated as separate IT.
WHO	Ethics and governance for AI in health	Public benefit, human oversight and accountability should be central design requirements.

5. Risk analysis

5.1 Mis-triage

The most direct risk is that a patient is directed to the wrong care pathway. A patient needing urgent care may be advised to use a less urgent route. A patient needing GP assessment may be reassured at home. The risk is amplified because patients may interpret an NHS App recommendation as authoritative.

5.2 False reassurance and delayed care

Triage systems must be designed to recognise uncertainty. A safe system should not force a confident recommendation where the symptom pattern is ambiguous or high-risk. It should escalate uncertain cases rather than creating false reassurance.

5.3 Bias and unequal performance

AI systems may perform differently across age, sex, ethnicity, disability, language, health literacy and socio-economic groups. Patient descriptions of pain, symptoms and urgency vary. Any deployment should test for demographic disparities and should not assume that average performance is enough.

5.4 Digital exclusion

A digital-first pathway may disadvantage elderly patients, disabled patients, those without smartphones, people with low English proficiency, and people with low digital confidence. A safe model needs parallel routes and not just digital design improvements.

5.5 Data protection and confidentiality

AI triage may process special category health data and potentially sensitive contextual data. Governance must address data minimisation, lawful basis, transparency, access controls, supplier processing, retention, secondary use and audit trails.

5.6 Accountability gap

If AI contributes to harm, responsibility can become unclear. The NHS body, GP practice, clinician, commissioner, app operator and supplier may each point to a different part of the chain. Accountability should be defined in advance, not litigated after harm.

5.7 Automation bias

Patients and staff may over-trust an AI-generated recommendation because it appears objective or official. Training, warnings and escalation routes should be designed to prevent over-reliance.

5.8 Model drift and post-deployment change

AI performance can change over time due to population changes, pathway changes, supplier updates, coding changes or model drift. Safety approval at launch is not enough. Continuous monitoring is required.

6. The Sitora NHS AI Trust Test

Sitora recommends that any AI triage system used in NHS patient access pathways should pass a six-point public trust test before national scale-up.

1. Clinical safety validation

Evidence across urgent, ambiguous, complex and routine scenarios; testing against real-world symptom descriptions; documented false negative and false positive risks.

2. Independent AI audit

External review of safety, bias, accessibility, explainability, cyber risk, data protection and model governance, not only supplier self-assessment.

3. Human escalation route

Clear ability for patients and staff to escalate to a human where symptoms are serious, worsening, complex, uncertain, or the patient disagrees.

4. Defined accountability

Named accountable owners across NHS body, clinical safety officer, procurement lead, supplier and incident-response process.

5. Patient transparency

Plain-English notification when AI is used, what role it plays, what it cannot do, and how to challenge or seek human review.

6. Ongoing monitoring

Live monitoring of incidents, near misses, complaints, demographic disparities, model drift, supplier changes and clinical outcomes.

How the Trust Test fits existing standards

The Sitora Trust Test is not proposed as a replacement for existing NHS, NICE, MHRA, ICO or CQC frameworks. It is a public-facing assurance layer designed to translate those requirements into practical questions that policymakers, NHS boards and patients can understand.

7. Policy recommendations

Recommendation 1: Require a published AI impact assessment

Every AI triage deployment should produce a public summary covering clinical risk, equality impact, accessibility, data protection, cyber risk, human oversight and patient communication.

Recommendation 2: Mandate independent audit before scale-up

Before national rollout, AI triage tools should undergo independent assurance. Audit findings should be summarised publicly, with commercially sensitive material redacted only where necessary.

Recommendation 3: Create a human review guarantee

Patients should never be trapped in a fully automated route. The NHS should guarantee human escalation for high-risk, uncertain, vulnerable or disputed cases.

Recommendation 4: Define accountability in procurement contracts

NHS supplier contracts should include audit access, incident cooperation, performance obligations, liability provisions, change-control requirements and post-deployment monitoring duties.

Recommendation 5: Publish a patient-facing AI notice

Patients should be told when AI is being used, what it is used for, whether it is diagnostic or routing-only, what data it uses and how to challenge or escalate.

Recommendation 6: Establish AI incident reporting

The NHS should create a national reporting route for AI-related patient safety incidents and near misses, including triage misdirection and delayed care.

Recommendation 7: Test with vulnerable and excluded groups

Assurance should include older adults, disabled patients, people with limited English, low digital literacy, complex long-term conditions and minority communities.

Recommendation 8: Require board-level AI governance

NHS organisations deploying AI triage should have named executive ownership, clinical safety ownership and information governance ownership.

8. Implementation roadmap

Phase	Timeline	Action	Owner
Phase 1	0-3 months	Publish national assurance requirements for AI triage; define accountable owner and minimum audit scope.	DHSC / NHS England
Phase 2	3-6 months	Run independent audits of pilots; test vulnerable groups; publish	NHS England / suppliers /

		patient-facing notices.	auditors
Phase 3	6-12 months	Introduce incident reporting, board reporting and national monitoring metrics.	NHS England / regulators
Phase 4	12-24 months	Scale only systems that pass safety, fairness, accessibility and accountability thresholds.	NHS England / Integrated Care Boards

9. Stakeholder map

Patients and carers

Need understandable routes, human escalation, privacy protection and the right to challenge unsafe or unsuitable outcomes.

GP practices and frontline clinicians

Need clear clinical accountability, workflow safety, training and protection from being blamed for supplier or system failures.

NHS England and DHSC

Need national assurance, public trust, procurement standards and evidence of safe productivity gains.

Regulators

Need visibility of AI use, incident data, safety cases, equality effects and supplier accountability.

Suppliers

Need clear expectations for evidence, audit access, monitoring, change control and patient-facing transparency.

Parliament and public bodies

Need assurance that innovation is not outpacing safeguards in public healthcare.

10. Conclusion

AI can support the NHS, but it should not be allowed to quietly become the gatekeeper to care without stronger public assurance.

The right position is not anti-AI. It is pro-trust, pro-safety and pro-accountability. Healthcare AI should earn public confidence through evidence, audit and clear governance.

The NHS does not just need AI that works. It needs AI patients can trust.

Appendix A: AI triage risk register

Risk	Potential harm	Likelihood	Impact	Suggested controls
Mis-triage	Delayed care, worsening condition, emergency escalation	Medium	High	Clinical validation; red flag rules; human escalation; incident monitoring
False reassurance	Patient delays seeking urgent help	Medium	High	Uncertainty thresholds; safety-net advice; prominent emergency warnings
Bias	Unequal access or unsafe outcomes for groups	Medium	High	Demographic testing; equality impact assessment; continuous disparity monitoring
Digital exclusion	Vulnerable patients locked out of care	High	High	Non-digital alternatives; accessibility testing; assisted digital support
Privacy failure	Loss of trust, legal breach, sensitive data exposure	Medium	High	DPIA; data minimisation; access controls; supplier contracts; audit logs
Accountability gap	Delayed redress, clinician scapegoating, public mistrust	Medium	High	Contractual liability; named owners; clear incident process
Automation bias	Over-reliance by patient or staff	Medium	Medium	Training; explainability; warnings; human override
Model drift	Performance declines after launch	Medium	High	Post-deployment monitoring; change control; revalidation

Appendix B: Assurance checklist for NHS AI triage

- Has the intended purpose been defined clearly, including whether the system is diagnostic, routing-only or decision-support?
- Has MHRA medical device status been assessed and documented?
- Has a DCB0129/DCB0160 clinical safety case and hazard log been completed where applicable?
- Has a data protection impact assessment been completed and approved?
- Has the system been tested for vulnerable groups, language variation and digital exclusion?
- Has independent audit been completed before scale-up?
- Does the patient receive a plain-English notice that AI is being used?
- Is there a clear route to human review or escalation?
- Are incident and near-miss reporting routes defined?
- Are supplier update, model change and revalidation processes contractually controlled?
- Are board-level accountable owners named?
- Are post-deployment metrics reviewed regularly and publicly summarised where appropriate?

References

[1] NHS England (2026). NHS accelerates artificial intelligence rollout to cut waiting times and improve care for millions. https://www.england.nhs.uk/2026/07/nhs-accelerates-artificial-intelligence-rollout-to-cut-waiting-times-and-improve-care-for-millions/

[2] NHS England (2025). Reforming elective care for patients. https://www.england.nhs.uk/long-read/reforming-elective-care-for-patients/

[3] Digital Health (2026). AI tool allows a million patients to book appointments via NHS App. https://www.digitalhealth.net/2026/04/ai-tool-allows-a-million-patients-to-book-appointments-via-nhs-app/

[4] The Guardian (2026). A&E patients with non-urgent ailments may be told to come back later under NHS plans. https://www.theguardian.com/society/2026/jun/10/patients-accident-emergency-non-urgent-ailments-nhs-plans-overcrowding

[5] NHS England. Digital clinical safety assurance. https://www.england.nhs.uk/long-read/digital-clinical-safety-assurance/

[6] NHS England / NHS Digital. Review of digital clinical safety standards: DCB0129 and DCB0160. https://digital.nhs.uk/data-and-information/information-standards/governance/latest-activity/standards-and-collections/review-of-digital-clinical-safety-standards-dcb0129-and-dcb0160

[7] MHRA / GOV.UK. Medical devices: software and artificial intelligence (AI). https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device

[8] NICE. Evidence standards framework for digital health technologies. https://www.nice.org.uk/what-nice-does/digital-health/evidence-standards-framework-esf-for-digital-health-technologies

[9] ICO. Guidance on AI and data protection. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/

[10] Care Quality Commission (2026). Artificial intelligence in health and social care: CQC role, expectations and plans. https://www.cqc.org.uk/about-us/transparency/artificial-intelligence-health-social-care-cqcs-role-expectations-plans

[11] World Health Organization (2021). Ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200

[12] World Health Organization (2025). Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. https://www.who.int/publications/i/item/9789240084759

[13] NICE. Artificial intelligence and digital regulations service. https://www.nice.org.uk/what-nice-does/digital-health/artificial-intelligence-ai-and-digital-regulations-service

[14] Department for Science, Innovation and Technology (2023). A pro-innovation approach to AI regulation. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

[15] NHS England. Artificial intelligence and machine learning. https://www.england.nhs.uk/long-read/artificial-intelligence-ai-and-machine-learning/

[16] NHS England / NHS Digital. AI knowledge repository. https://digital.nhs.uk/services/ai-knowledge-repository

[17] Health Research Authority. Artificial Intelligence and Digital Regulations Service. https://www.hra.nhs.uk/planning-and-improving-research/research-planning/how-were-supporting-data-driven-technology/artificial-intelligence-and-digital-regulations-service/

[18] Unsworth et al. (2021). The NICE Evidence Standards Framework for digital health and care technologies. npj Digital Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC8236783/

[19] Morley et al. (2021). Toward a framework for evaluating the safety, acceptability and efficacy of AI systems for health. https://arxiv.org/abs/2104.06910

[20] Alattal et al. (2025). Integrating Explainable AI in Medical Devices: Technical, Clinical and Regulatory Insights and Recommendations. https://arxiv.org/abs/2505.06620