National Commission into the Regulation of AI in healthcare consultation

Call for Evidence Questions

Question 1: Which of the following best describes your view about the need to change the UK's framework for regulating AI in healthcare? (optional)

Significant reform: The current framework requires substantial changes

Question 2.1: To what extent do you agree or disagree that the current regulatory framework is sufficient in the following domains: Safety and Performance Standards (optional)

Strongly disagree

Question 2.2: To what extent do you agree or disagree that the current regulatory framework is sufficient in the following domains: Data Privacy and Data Governance (optional)

Disagree

Question 2.3: To what extent do you agree or disagree that the current regulatory framework is sufficient in the following domains: Transparency (optional)

Disagree

Question 2.4: To what extent do you agree or disagree that the current regulatory framework is sufficient in the following domains: Requirements for clinical evidence (optional)

Strongly disagree

Question 2.5: To what extent do you agree or disagree that the current regulatory framework is sufficient in the following domains: Post Market Surveillance (optional)

Strongly disagree

Question 3: How would you rate the current framework's impact on innovation? (optional)

Somewhat loose [lacks necessary controls]

Question 4: How might the UK's framework for regulation of AI in healthcare be improved to ensure the NHS has fast access to safe and effective AI health technology? (optional)

The regulation model needs to develop from an individual to a whole-system framework that covers prevention and population health AI, mandates for generation and transparency for real-world evidence (RWE), operationalises equity, and enables faster scale-up through stepwise authorisation and shared national infrastructure. The UK framework for regulation must also consider international interoperability, as AI innovation crosses jurisdictions, ensuring the approach is in line with the **EU AI Act risk classifications**, and **FDA Predetermined Change Control Plans (PCCPs)** for AI-enabled medical devices. This aligns with WHO call for harmonised, system-wide digital-health governance and strong data foundations (interoperability, data quality, standards) (WHO).

a. **Address gap in regulation of AI in prevention & population health**

Current regulations defines medical purpose at the *individual* clinical level (diagnosis, prevention, monitoring, treatment, clinical risk prediction tools), and does not extend to public-health and population-health digital tools. This creates a significant regulatory "blind spot" for (i) AI supporting population health decisions (for example surveillance, risk stratification, commissioning, predictive analytics) and (ii) direct-to-consumer AI (such as GenAI-enabled behavioural interventions with a focus on self management of health, wellbeing and prevention). **The MHRA should consider proportionate categories for prevention/D2C and population-health AI, co-developed with British Standards Institute, NICE, the Faculty of Public**

Health, and the public, to define expectations for safety, transparency, data governance and performance at system scale (aligning with WHO's call for harmonised, system-wide digital-health governance) (ref WHO).

b. Address Real World Evidence (RWE) gap including effectiveness, population and system benefits, and equity.

Regulatory frameworks have an important role to encourage innovation which supports population benefits. Yet, there is no mandated requirement for real world testing or simulation of pilots before deployment; and once deployed diversity of the health and care digital infrastructure, data quality, inter-operability of data, populations served means evidence is not generalisable across health and care systems in the UK. Evidence in procurement-usable formats (such as benefits realisation plans, health equity impact assessments, interoperability profiles) and transparency on outcomes pre- and post-deployment across diverse NHS data and digital infrastructures across the UK is essential to operationalising AI in the care system.

The MHRA should consider stepwise authorisation across (a) RWE (including model performance (eg calibration, drift, interoperability), evaluation including system wide benefits (considering data linkage to capture benefits across whole systems), ROI, harms (including mitigation of clinical harms); and (b) population outcomes (including measurable health equity outcomes (differential performance across demographic groups).Alongside the use of The MHRA AI Airlock as a regulatory sandbox to test tools in "less-than-ideal" NHS data environments, to verify clinical, operational, equity and population outcomes before wide rollout – this should extend to ensuring causal validity for AI informing commissioning/resource decisions.
Encourage NHS organisations to **involve public-health** expertise in procurement and evaluation to assess population risks/benefits and organisational impacts, drawing on their expertise in evaluation (Cresswell 2024, Cresswell 2025, Scott 2025), health equity assessments, co-production and working with communities and trust, data analytics and outcome measurements at a population and whole health and care system level.

c. Operationalise equity and public-health ethics.

The current regulatory approach is for tools to defined intended population group – the emphasis needs to encourage development of AI across population groups. The MHRA should mandate equity-oriented evidence: subgroup performance, bias/fairness monitoring, fairness metrics (e.g., equalized odds, calibration across subgroups, demographic parity), and measurable health-equity outcomes that should be mandated as part of a risk assessment. The MHRA should include explicit consideration of public health ethical principles (justification, proportionality, equity/fairness, transparency/accountability, and public acceptability) – to consider does the expected public good of using AI outweigh any limits placed on autonomy, and is that balance justified, proportionate and fair?

Question 5: How should the regulatory framework manage post-market surveillance for AI health technologies? (optional)

Post-market surveillance for AI must shift from a narrow, clinical risk based model to a whole-system, population-level, equity-oriented surveillance framework. Traditional mechanisms, including manufacturer vigilance reporting and the Yellow Card system, are designed to capture individual clinical harms. But there is a lack of awareness of this system, specifically for raising concerns of AI for prevention, outside of healthcare systems. AI also introduces additional risks, such as population-level, system-level and equity-related risks (and potentially benefits), that are not visible through these conventional routes. Surveillance for AI must be continuous, active, and system-wide. It should integrate equity, population-health ethics, prevention, and public legitimacy. Strengthened post market surveillance should consider the following;

a. **Expand beyond clinical safety to include population-health benefits and prevention of harms**

The post market surveillance is currently focused on clinical harms. Whilst the DHSC *Data-Driven Health and Care Technology* includes best practice for bias testing and monitoring, and real-world conditions this is not **embedded in regulation**. This creates a gap given the potential for AI to amplify inequities, mis-allocate resources, or influence population-level outcomes.

We recommend expanding them to require:

- drift and recalibration monitoring in real world data and infrastructure
- bias and equity surveillance (including surveillance of use outside of intended population, or populations which were absent in the training data)
- population or system level impact (benefits and harms)
- prevention (of harms and poor health).
  The post market surveillance will need to be tailored to different forms of AI (eg continuously learning AI: drift detection, revalidation requirements; Foundation or agentic models: hallucinations, unsafe reasoning patterns, inappropriate autonomy, and emergent behaviours).

Emerging evidence on conversational and agentic AI systems supports the potential of these tools in prevention, but also the need for caution for when considering the integration within prevention-focused public health strategies. Recent evidence demonstrates that agentic AI can provide personalised guidance and improve engagement in ways that are scalable across populations (Artificial intelligence for public health can harness data for healthier populations | Nature Health). However, these systems may produce inaccurate and unsafe advice, which may be disproportionate for particular groups (depending on language, culture, or help-seeking behaviour) - underscoring the need for rigorous safety, validation, and oversight when deploying agentic AI for prevention (Large language models in global health | Nature Health). This extend to the need for a dedicated approach to address the unique challenges of foundational models for both clinical and public use (eg prompt injection, hallucination, emergent capabilities, fine-tuning drift).

A route to strengthening the approach may be through the existing DCB0129 (manufacturer) and DCB0160 (deployer) requirements (NHS England Digital). But there remains consideration of safety in direct to consumer tools or those operating at a population level which would fall outside of the current DCB0129 and DCB0160 requirements.

b. Enable efficient information sharing on post market surveillance between NHS organisations, manufacturers and regulators.

The MHRA should consider co-developing across stakeholders, a model which can facilitate cross-NHS learning to accelerate safe adoption, such as standardised reporting formats and accessible platforms. Also aligning with international

regulations (EU AI Act and FDA) to reduces inequities resulting from fragmentation and shares learning around safety and effectiveness.
From a public-health perspective, it would be helpful to consider how the MHRA will operate where the NHS acts as both developer and deployer. For example, clarify the assurances for AI tools developed directly by the NHS (eg proposed AI prompt tools within the Federated Data Platform, or generative AI tools outlined in the NHS 10-Year Plan), in line with external products.

Question 6: Which statement best reflects your view on the current legal framework for establishing liability in healthcare AI tools? (optional)

Gaps exist: existing laws work for most cases but leave uncertainty in some scenarios

Question 7: How could manufacturers of AI health technologies, healthcare provider organisations, healthcare professionals, and other parties best share responsibility for ensuring AI is used safely and responsibly? (optional)

Ensuring the safe and responsible use of AI in health and care requires a shared, clearly defined set of responsibilities across the entire AI lifecycle. No single actor can meaningfully assure safety and responsible use, in isolation. A coordinated and transparent approach is essential to maintain public trust, prevent harms, and ensure that AI delivers population-level benefit and supports equity.

There are two specific considerations from a public health perspective:

A. AI can alter population outcomes *independently of individual clinical accuracy*. Responsibility must therefore be shared across manufacturers, providers, and regulators to monitor individual, system, and population-level impacts throughout deployment and scale-up (Panteli 2025)
B. Equity risks must be jointly owned across manufacturers, regulators and providers. Bias and inequity are core risks to population benefits, and attributable to socio-technical factors including digital infrastructure, data quality, model design, implementation choices, digital exclusion (for both users and health and care professionals – spanning access to digital tools, digital skills, trust, and user needs) (WHO 2022; Davies 2021 ; Mahl, 2025;

Panteli 2025; WHO 2026 – in draft). These risks cannot be mitigated by any single party acting alone.

C. Manufacturers should be responsible for transparent performance claims, ongoing model monitoring, and support for real-world evaluation across diverse settings. Healthcare providers and professionals should ensure safe local implementation, appropriate use, workforce training, and escalation of harms. Regulators and commissioners should set proportionate, enforceable expectations for real-world evidence, equity impact assessment, and interoperability, using procurement and commissioning as key levers for safety and accountability.

Public and patient engagement should be a shared responsibility, supporting trust, acceptability, and equitable adoption. **The MHRA should work with the Faculty of Public Health and other partners to define a shared accountability framework, to ensure safe and responsible AI use to the benefit of populations.**

Question 8: In the event of an adverse patient outcome where an adverse patient outcome involved an AI tool, where do you think liability should lie? (optional)

Liability for adverse patient outcomes involving AI should be understood as shared and context-dependent, rather than a simple either/or between the AI and Clinician. A potential approach is to consider who was best positioned to prevent the harm rather than attributing blame solely to clinicians or developers.
Clinicians retain professional responsibility for applying judgment, but their liability should be mitigated where reliance on an AI system was reasonable, especially if the tool was validated, embedded into workflows, used appropriately. Liability could be shared if:

- The AI's reasoning or confidence level was poorly explained
- The interface made the correct recommendation easy to miss or misunderstand
- Institutional policies pressured clinicians to ignore or discount AI outputs

Developers should be accountable for model safety, transparency, and ongoing performance, including undisclosed biases or drift, while deploying organisations

carry responsibility for governance, training, and safe integration to avoid automation bias.

From a public-health perspective, liability frameworks must also address population-level harms.

Consideration of the role of MHRA to act on notifications of harm are needed. Should the MHRA have explicit powers or expectations to recall, pause or constrain the use of AI where harms, or inequities are detected, rather than simply monitoring them. Overall, a lifecycle, system-level approach is needed in which responsibility aligns with control, foreseeability, and the ability to manage risk across developers, deployers, clinicians, and regulators. As AI becomes more autonomous and embedded in standards of care, liability is likely to shift gradually away from individual clinicians and toward system-level and manufacturer accountability, much like aviation or pharmaceuticals.

Question 9: Do you have any other evidence to contribute? You can submit written evidence in the comment box. Note: please confirm that you have the necessary permissions prior to sharing any documents in this way. (optional)

- Alagarajah *et al*. (2025) Bridging the Digital Divide: A Toolkit for Equity-Centred Regulation of Digital Mental Health https://www.regulatoryrapporteur.org/software/bridging-the-digital-divide-a-toolkit-for-equity-centred-regulation-of-digital-mental-health/1025.article
- Cresswell K, *et al* . (2025) A mixed methods formative evaluation of the United Kingdom National Health Service Artificial Intelligence Lab. npj Digital Medicine
- Cresswell K, *et al*. (2024) Evaluating Artificial Intelligence in Clinical Settings—Let Us Not Reinvent the Wheel. Journal of Medical Internet Research.
- Davies AR. *et al* Addressing the Digital Inverse Care Law in the Time of COVID-19: Potential for Digital Technology to Exacerbate or Mitigate Health Inequalities. Journal of Medical Internet Research 2021 https://www.jmir.org/2021/4/e21726/
- Faculty of Public Health (2025) *An opportunity to seize or a threat to mitigate? UK Public Health and AI Report*. London: FPH AI & Digital Public Health SIG.

Available at: https://www.fph.org.uk/policy-advocacy/special-interest-groups/artificial-intelligence-digital-public-health-sig/an-opportunity-to-seize-or-a-threat-to-mitigate-uk-public-health-specialists-views-on-artificial-intelligence-ai/ (Accessed: 14 January 2026).

- Fotheringham K. *et al* (2024) Accidental injustice: Healthcare AI legal responsibility must be prospectively planned prior to its adoption. Future Healthcare J https://www.sciencedirect.com/science/article/pii/S2514664524015716
- Gov.uk (2025) *AI Airlock Sandbox Pilot Programme Report*. Medicines and Healthcare products Regulatory Agency, 16 October. Available at: https://www.gov.uk/government/publications/ai-airlock-sandbox-pilot-programme-report (Accessed: 14 January 2026).
- Lawton T *et al* (2024) Clinicians risk becoming 'liability sinks' for artificial intelligence. Future Healthcare J https://pmc.ncbi.nlm.nih.gov/articles/PMC11025047/
- Leone de Castris A, Thomas C. (2025) AI governance around the world: United Kingdom. The Alan Turing Institute. (Accessed: 14 January 2026).
- Mahl D, *et al (*2025) Responsible artificial intelligence in public health: a Delphi study on risk communication, community engagement and infodemic management. BMJ Global Health https://gh.bmj.com/content/10/5/e018545
- Oskrochi Y (2025) Digital Health Technology Compliance With Clinical Safety Standards In the National Health Service in England: National Cross-Sectional Study. Journal of Medical Internet Research
- Panteli D, *et al* (2025) Artificial intelligence in public health: promises, challenges, and an agenda for policy makers and public health institutions. The Lancet Public Health https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667%2825%2900036-2/fulltext
- Scott I *et al* (2025). How can we promote greater adoption of AI in healthcare? | BMJ Quality & Safety
- Sudlow, C. (2024) *Uniting the UK's Health Data: A Huge Opportunity for Society (The Sudlow Review)*. Health Data Research UK. Available at: https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/ (Accessed: 14 January 2026).

- The King's Fund (2025) *AI in the NHS 2025: shaping the future of health care*. London: The King's Fund. Available at: https://www.health.org.uk/events/ai-in-the-nhs-2025 (Accessed: 14 January 2026).
- Woolley K, *et al* (2023) Mapping Inequities in Digital Health Technology Within the World Health Organization's European Region Using PROGRESS PLUS: Scoping Review. Journal of Medical Internet Research
- World Health Organisation, 2022 Equity within digital health technology within the WHO European Region: a scoping review. Ref WHO/EURO:2022-6810-46576-67595

Question 10: You can upload documents to be considered as part of this call for evidence. Note: please confirm that you have the necessary permissions prior to sharing any documents in this way. (optional)

No response