



TRUSTWORTHY AI IN HEALTH: GOVERNING A DOUBLE-EDGED SWORD

Introducing a Policy Brief of the Global Health Hub Germany Working Group on AI in Health

Rachel Firestone and Allison Colbert on behalf of the Working Group on AI in Health

INTRODUCTION

- Artificial Intelligence (AI)’s ability to process data and detect patterns beyond human capacity is transforming healthcare and public health.
 - Associated risks are not fully understood and can cause serious harm – particularly in high-stakes clinical and public health contexts.
- The EU AI Act provides the first horizontal legal framework for AI, but does not by itself ensure equitable benefit distribution or meaningful public communication - posing a gap between legal compliance and trustworthy deployment.
 - Failures in oversight directly affect access, patient and public safety, and public trust.

AI IN HEALTHCARE INCREASINGLY REPRESENTS A DOUBLE-EDGED SWORD

- The same technologies that can expand access and improve outcomes can also amplify harm at scale when governance mechanisms are weak or absent.
- This brief provides evidence-based recommendations to strengthen transparency, oversight, and accountability in health AI governance**
→ Particularly as Germany operationalizes the EU AI Act for health.

CASE STUDIES

AI IN MAMMOGRAPHY SCREENING IN GERMANY

14 million women receive mammography screening per year through a double-reading system involving 2-3 radiologists.

AI System

- An image-classification system classifies images (normal/suspicious).
- Radiologists review their own read and the AI output before the final diagnosis.

Drawing on evidence from two health AI deployments in high-income clinical contexts



AI DEPLOYMENT IN MENTAL HEALTH IN THE UK

A national program provides structured treatment for anxiety and depression across 200+ services nation-wide.

AI System

- A rule-based chatbot-LLM interacts with users through a text-based interface.
- The AI conducts patient intake and referral assessments before humans are involved.

WHERE THIS LEAVES US

Key Governance Lessons

Access and efficiency gains are promising

Transparency and explainability remain limited

Independent evaluation and post-deployment accountability infrastructure is insufficient

Critical Patterns

Evidence generation remains heavily dependent on AI developers

Deployment occurs faster than governance infrastructure development

EU AI Act compliance does not guarantee ongoing equitable, trustworthy performance

Why This Matters Now

Patient safety and equity: Poorly governed AI systems can amplify errors, biases, and unequal outcomes at scale, particularly among vulnerable populations

Public trust: Trust in health AI depends not only on technical capacity but also on long-term transparency, accountability, and meaningful patient engagement

Competitiveness and collaboration: Trustworthy governance can become a strategic advantage that strengthens public confidence, supports adoption, and reinforces leadership

TEASER - FULL POLICY BRIEF TO BE PUBLISHED ON JULY 8TH, 2026

Presenting targeted, time-sensitive actions for trustworthy health AI governance