NHS

AI Disclosure Series

18 Critical Breaches in AI Workforce, Triage, Ambient Voice & Medical Data Sovereignty

A systematic examination of governance failures, liability gaps, and accountability breaches in NHS AI deployment.

Series: 18 of 18 Disclosures

Disclosure Breach Index

Breach 1 — Public Criminal-Threat Governance and Staff Liability Displacement

NHS AI Disclosure Series

Do not place workers at the visible endpoint of system-level liability.

Series: 1 of 18

1 The System Owner's Duties

- Procurement
- Permissions
- Audit design
- Training
- Clinical safety
- Vendor governance
- Escalation pathways
- Lawful access design

2 The AI-Mediated Environment

Staff using AI triage
↓
Ambient voice technology
↓
Workflow tools
↓
Cloud systems
↓
Model-generated outputs

Workers operate inside systems they did not design, procure, or control.

3 The Contradiction

Deploy AI nationally -> threaten staff with dismissal or prison for misuse

- Ambiguity over who processed what
- Who verified what
- What the model generated
- What entered the record
- What was retained

4 Visible Endpoint Consequences

- Disciplinary exposure
- Employment liability
- Tribunal reliance
- Protected-disclosure detriment risk
- Stress and verification burden
- Reputational harm

NHS England · ICBs · GP practices · Vendors · Cloud providers · Staff

Responsibility must be distributed — not displaced.

This is not ordinary confidentiality enforcement.
You cannot deploy the system and displace the liability.

Breach 2 – Mass Named-User AI Licence Exposure and Audit-Trail Liability

NHS AI Disclosure Series

Series: 2 of 18

A personal login can become a personal liability trail when the system architecture is not fullydisclosed.

1 What the Licence Creates

- Named staff account
- Unique login / user ID
- Role permissions
- Prompt and output history
- Timestamps and access logs
- Usage analytics / telemetry
- Organisational audit link
- Retention and export pathway

2 The Audit-Trail Environment

Named user account
↓
AI prompts
↓
Model outputs
↓
Telemetry / logging
↓
Vendor or SaaS platform
↓
Organisational audit trail

The worker becomes identifiable across the system chain.

3 The Contradiction

Buy mass named-user AI licences -> convert worker activity into traceable audit material -> later rely on the same trail for liability

- Who actually authored the text?
- What was AI-generated?
- What was merely suggested?
- What was retained by the vendor?
- What was exported into HR, governance, or legal use?

4 Visible Endpoint Consequences

- Disciplinary exposure
- Performance monitoring
- Tribunal evidence risk
- Protected-disclosure detriment risk
- False attribution risk
- Reputational harm

NHS England ····· ICBs / trusts / GP practices ····· Vendors ····· Cloud providers ····· Staff accounts

Audit responsibility must be governed – not silently personalised.

A named-user AI licence is not just access. It is an identity-linked audit architecture.

This is not ordinary software licensing.

When the licence is personal, the liability trail becomes personal.

Breach 3 — AI Triage Processing Sovereignty and Undisclosed Decision Engine

NHS AI Disclosure Series

Patient access must not be routed through a decision layer whose processing engine is not visibly disclosed.

Series: 3 of 18

1 What the Triage System Receives

- Patient symptoms
- Questionnaire answers
- Demographic details
- Prior-care or access context
- Urgency indicators
- Appointment or routing request

2 The Hidden Processing Chain

Patient-access input
↓
NHS app / portal / GP interface
↓
Triage application layer
↓
Rules / model / inference engine
↓
Cloud or SaaS processing environment
↓
Output score / routing recommendation
↓
Logging / telemetry / support-access pathway

Patient access may be machine-classified before the engine is publicly understood.

3 The Contradiction

Deploy AI triage nationally -> route patients through an undisclosed decision engine

- What model or rules decide urgency?
- Where does inference execute?
- Who validated under-triage risk?
- What is logged or retained?
- Where is the human-review point?
- What crosses vendor or cloud boundaries?

4 Visible Endpoint Consequences

- Urgency classification
- GP / pharmacy / A&E / self-care redirection
- Appointment prioritisation
- Delayed or denied access risk
- Equality and digital-exclusion exposure
- Clinical safety and liability consequences

NHS England ····· ICBs / trusts ····· GP practices ····· Vendors ····· Cloud providers ····· Patients

Patient-access routing responsibility must be disclosed — not hidden inside the engine.

A triage engine is not just software. It is a healthcare-access decision layer.

This is not ordinary digital intake.

If the engine is undisclosed, the access decision is not transparently governed.

Breach 4 – Data Execution, Modelling Location, and Cross-Border Data-Bleed Non-Disclosure

NHS AI Disclosure Series

Storage claims do not disclose where NHS data is executed, inferred, logged, analysed, or exposed.

Series: 4 of 18

1 What the System Actually Involves

- Patient symptoms and clinical data
- Patient speech and consultation content
- Staff prompts and workflow inputs
- Clinical notes and summaries
- Telemetry, logs, and audit events
- Support-access and analytics pathways

2 The Hidden Processing Chain

SaaS application layer
↓
Cloud environment
↓
Model / inference engine
↓
Telemetry / logging
↓
Support-access route
↓
Sub-processors
↓
Backup / disaster recovery
↓
Model-improvement pathway

Storage location does not reveal the execution truth.

3 The Contradiction

Say the data is stored in the UK -> but do not disclose where it is executed, inferred, logged, debugged, retained, or exposed

- Where does inference occur?
- What crosses borders?
- Who can access support logs?
- What telemetry is retained?
- Where is the human-review point?
- Which vendor or cloud layers operate the chain?

4 Visible Endpoint Consequences

- Cross-border data-bleed risk
- Opaque processing sovereignty
- Clinical-record formation uncertainty
- Staff audit and liability exposure
- Vendor / cloud dependency risk
- Public accountability failure

NHS England ..... ICBs / trusts / GP practices ..... Vendors ..... Cloud providers ..... Patients & staff

Execution responsibility must be disclosed — not hidden behind storage claims.

A storage statement is not an execution map.

This is not ordinary data hosting. If the execution chain is undisclosed, the legal processing truth is undisclosed.

Breach 5 – AI Triage Output Traceability, Accuracy, and Human-Review Failure

NHS AI Disclosure Series

If triage output affects care routing, the source, logic, and human review must be visibly traceable.

Series: 5 of 18

1 What the Triage Output Produces

- Urgency score or classification
- GP / pharmacy / A&E / self-care routing
- Appointment prioritisation
- Advice or next-step recommendation
- Escalation or clinician-review trigger
- Record or audit entry

2 The Traceability Chain

Patient-access input
↓
Rules / model / inference engine
↓
Triage output or score
↓
Human-review checkpoint
↓
Routing / appointment outcome
↓
Audit trail / retained record

If the output cannot be traced back through the chain, accountability collapses.

3 The Contradiction

Use AI triage output to shape patient access -> but do not visibly prove how the output was produced, checked, corrected, or overruled

- What generated the output?
- What evidence supported it?
- Who reviewed or approved it?
- Can the output be challenged or corrected?
- Was under-triage tested?
- Is the reasoning visible to staff or patients?

4 Visible Endpoint Consequences

- Misrouted patient access
- Delayed or denied care risk
- False reassurance or false urgency
- Clinical safety exposure
- Equality and transparency concerns
- Liability and complaint escalation

NHS England · ICBs / trusts · GP practices · Vendors · Cloud providers · Patients

Triage-output accountability must be visible from source to decision.

A triage output is not just a suggestion. It is a patient-access decision trace.

This is not ordinary digital intake. If the output is not traceable, the access decision is not safely governed.

Breach 6 – Voice / Speech AI Accent Validation Failure and Unsafe National Scaling from Sussex Pilot Evidence

NHS AI Disclosure Series

Speech recognition is not intelligence. If accent and voice validation are not visibly proven, national scaling is unsafe.

Series: 6 of 18

1 What the Speech System Receives

- Patient speech
- Clinician speech
- Accent and dialect variation
- Background noise / interruptions
- Multi-speaker overlap
- Clinical shorthand and medication names
- Symptom descriptions
- Timing and negation phrases

2 The Risk Chain

Spoken consultation / care conversation
↓
Speech recognition / transcription
↓
AI interpretation / summarisation
↓
Draft note / structured output
↓
Patient record entry
↓
Care decision / routing / advice

If the first capture is wrong, the full chain can carry the error forward.

3 The Contradiction

Use ambient voice and speech AI nationally -> but do not visibly prove accent, dialect, noise, disability, and language validation at national scale

- Was Sussex pilot evidence enough for national scaling?
- Were accent and dialect variations tested?
- How are negation errors handled?
- How are drug-name confusions handled?
- Where is the human verification point?
- What enters the patient record if the transcript is wrong?

Example: NO CHEST PAIN -> CHEST PAIN

4 Visible Endpoint Consequences

- Misheard or distorted symptoms
- False reassurance or false urgency
- Unsafe clinical record formation
- Inequality and language-bias exposure
- Staff verification burden
- Clinical safety and liability risk

NHS England · ICBs / trusts · GP practices · Vendors · Cloud providers · Patients & staff

Speech-pathway accountability must be proven from voice capture to record entry.

A speech system is not just dictation. It is a clinical meaning-capture layer.

This is not ordinary voice software. One misheard word can rewrite the patient.

Breach 7 – Ambient Voice Technology and AI-Mediated Clinical Record Formation

NHS AI Disclosure Series

Series: 7 of 18

When AI sits between the spoken consultation and the medical record, record formation itself becomes a governance issue.

1 What the Ambient System Captures

- Patient speech
- Clinician speech
- Consultation context
- Background interruptions / overlap
- Medication names and symptoms
- Questions, answers, and timing
- Nuance, hesitation, and clarification
- Spoken content intended for notes

2 The Record-Formation Chain

Spoken consultation / care conversation
↓
Ambient capture
↓
Speech recognition / transcription
↓
AI summarisation / structuring
↓
Draft note / letter / template content
↓
Human review / edit point
↓
Patient record entry
↓
Downstream care / admin / evidence use

Once AI sits between speech and record entry, the note is no longer a direct human capture.

3 The Contradiction

Deploy ambient voice nationally -> allow AI-generated notes, summaries, and structured content to enter the patient record -> but do not visibly prove what was heard, what was transcribed, what was inferred, what was omitted, who verified it, or what finally entered the record

- What exactly did the system capture?
- What did the AI add, remove, or smooth?
- Who verified the final wording?
- What happens if context is lost?
- Can the patient challenge the note?
- Which version becomes the official record?

4 Visible Endpoint Consequences

- Distorted or incomplete clinical record
- Lost nuance, safeguarding, or symptom context
- Medication / history misunderstanding
- Staff verification burden
- Patient complaint and evidential dispute risk
- Clinical safety and liability exposure

NHS England · ICBs / trusts · GP practices · Vendors · Cloud providers · Patients & staff

Record-formation accountability must be proven from consultation speech to final record entry.

An ambient voice tool is not just a note taker. It is a clinical record-formation layer.

This is not ordinary dictation. If AI forms the note, the record chain must be provable.

Breach 8 – AI-Generated Clinical Notes, Summaries, and Structured Content without Visible Proof

NHS AI Disclosure Series

When AI creates clinical notes and summaries, the official record must show how the content was generated, verified, and approved.

Series: 8 of 18

1 What the AI System Produces

- AI-generated clinical notes
- AI summaries of consultation
- Structured data extraction and formatting
- Problem lists and care plans
- Referral letters and reports
- Follow-up and task recommendations
- Population and risk stratification content

2 The AI Note-Generation Chain

Consultation audio / ambient capture
↓
Transcription
↓
AI understanding / interpretation
↓
AI summarisation / note generation
↓
Structuring / coding / labelling
↓
Draft note or summary output
↓
Human review interface
↓
Edits / additions / deletions
↓
Approval / sign-off
↓
Record entry into patient record

If any step is hidden or unproven, the content in the record is not trustworthy.

3 The Contradiction

Use AI to write the note -> but do not visibly prove how the note was generated, checked, edited, verified, or approved

- What exact data was used?
- What did the AI include, infer, or omit?
- What prompts or models shaped it?
- Who reviewed and edited it?
- What changed between draft and final?
- Was it clinically safe and accurate?
- Can the patient challenge the note?
- Which version is the official record?

4 Visible Endpoint Consequences

- Inaccurate or misleading clinical record
- Missing context, nuance, or patient meaning
- Wrong medication or history recorded
- Staff verification burden and overload
- Patient complaint and evidential dispute risk
- Clinical safety and liability exposure
- Systemic data quality collapse

NHS England · ICBs / trusts · GP practices · Vendors · Cloud providers · Patients & staff

Note-generation accountability must be visible from data-in to final record-out.

AI can help draft notes — but it must never silently write the official truth.

This is not ordinary automation. If AI forms the clinical note, the record must be fully provable, reviewable, and challengeable.

Breach 9 – Missing Audit Trails, Version History, and Immutable Log Non-Disclosure

NHS AI Disclosure Series

If the chain of actions cannot be traced, the decision cannot be trusted.

Series: 9 of 18

1 What the System Does

- Captures and processes patient data
- AI models infer, predict or generate outputs
- Humans review, edit or approve
- Outputs are written to the patient record
- Changes, updates and deletions occur over time
- Logs, metadata and system events are generated

2 The Hidden Trace Chain

User / system action
↓
Data accessed or imported
↓
AI model / prompt / parameters
↓
Inference / generation output
↓
Human review / edits
↓
Approval / sign-off
↓
Record entry / update
↓
Subsequent changes or deletions
↓
Log events / metadata captured

This entire chain must be traceable, versioned and tamper-evident to prove what really happened.

3 The Contradiction

Use AI outputs and record changes -> but do not visibly disclose the complete, immutable audit trail or version history

- What data was accessed?
- Which model and version was used?
- What prompt or parameters were set?
- What did the AI actually output?
- What was edited, by whom, and why?
- What was approved or overruled?
- What changed after entry – and when?
- What was deleted – and by whom?
- Can the chain be independently verified?
- Can the patient review the full history?

4 Visible Endpoint Consequences

- Inability to prove what really happened
- Lost context and missing history
- Undetected errors, bias or fabricated content
- Undisclosed deletions or backdating
- No patient ability to challenge the record
- Legal risk to fairness, accountability and safety
- Systemic loss of trust and data integrity

NHS England · ICBs / trusts · GP practices · Vendors · Cloud providers · Patients & staff

Audit-trail accounttability must be end-to-end, timestamped, versioned and tamper-evident.

If the trail is hidden, the truth is hidden.

This is not ordinary IT logging. A patient record is a legal and clinical truth.
If the audit trail is undisclosed, the truth of the record is not provable.

Breach 10 – Secondary Use, Function Creep, and Purpose-Binding Failure

NHS AI Disclosure Series

Series: 10 of 18

Data collected for one reason does not stay for one reason.

1 What the System Does

- Collects NHS data for clinical care
- Expands use to research, planning, training, or AI
- Links data across services, time, and datasets
- Creates new models, insights, or predictions
- Uses outputs for decisions or sharing beyond care
- Retains data indefinitely for "future benefits"

2 The Function Creep Chain

Data collected for care
↓
Stored in central systems
↓
Tagged for "improvement"
↓
Reused for research / AI training
↓
Linked with other datasets
↓
New purposes emerge
↓
Policy / commercial use
↓
Indefinite retention / reuse

Purpose drift happens silently unless the record proves purpose-boundary enforcement.

3 The Contradiction

Use data beyond the original care purpose -> but do not visibly prove lawful purpose change, fresh basis, and proportionality each time

- What was the original purpose?
- What is the new purpose?
- What lawful basis applies now?
- Was the patient told?
- Was secondary use necessary?
- Was proportionality tested?
- Could the purpose be achieved without patient data?
- Who benefits from the new use?
- Can the patient opt out?
- When will the data stop being used?
- How is function creep prevented?

4 Visible Endpoint Consequences

- Loss of patient control and autonomy
- Profiling, surveillance and decision harm
- Unlawful or unethical use of sensitive data
- Commercial exploitation without consent
- Erosion of trust in the NHS
- Legal, regulatory and reputational liability
- Data kept forever for unknown futures

NHS England · ICBs / trusts · GP practices · Vendors / AI developers · Researchers / partners · Policy makers

Purpose-boundary accountability must be visible at every reuse decision.

Data is not a commodity. It was given for a reason – not every reason.

This is not ordinary data reuse. If purpose changes are not lawful, proportionate, and transparent – the use is not permitted.

Breach 11 – Combined AI System Interactions and Compound Risk Failure

NHS AI Disclosure Series

When multiple AI systems interact, small failures compound into unsafe outcomes.

Series: 11 of 18

1 What the System Does

- Multiple AI systems operate across the patient pathway
- AI outputs feed other AI systems
- Decisions are made across linked models and workflows
- Data, predictions and actions amplify each other
- Small errors become compound harm
- No one sees the full system picture

2 The Interaction Chain

AI system A (triage / prediction)
↓
AI system B (risk score / routing)
↓
AI system C (documentation / summarisation)
↓
AI system D (decision support / alerts)
↓
AI system E (coding / billing / reporting)
↓
Integrated workflow / automation
↓
Action taken on patient
↓
Compound outcome

The whole system must be visible, testable, accountable and safe – not just the parts.

3 The Contradiction

Use multiple AI systems in combination -> but do not visibly prove the compound risk, emergent behaviour, or system-wide safety

- How do the systems interact in reality?
- What failures amplify other failures?
- What are the emergent risks?
- What was the system tested as a whole?
- Who owns the compounded outcome?
- Can any one system override the chain?
- What happens when data drifts?
- How is the full chain monitored?
- Can the patient or clinician challenge compound decisions?
- What is the worst-case outcome?

4 Visible Endpoint Consequences

- Wrong patient, wrong care, wrong pathway
- Harm multiplied across linked systems
- No single point of accountability
- Invisible risks and black-box decisions
- Legal, regulatory and professional liability
- Loss of trust and unsafe care delivery

NHS England — ICBs / trusts — GP & clinical teams — Vendors / AI developers — Cloud providers — Patients & public

System-level safety requires system-level visibility, testing, and governance.

It is not enough for each AI system to be safe in isolation.
The chain, the interactions, and the compound behaviours must be safe too.

This is not ordinary IT integration. It is a system-of-systems risk.
If the combination is not proven safe, *the* outcome is not safely governed.

Breach 12 – Consent, Notice, and Patient Knowledge Mismatch

NHS AI Disclosure Series

A patient cannot consent to what they are not told.

Series: 12 of 18

1 What the System Does

- Collects and processes patient data
- Uses AI for decisions, predictions, or automation
- Shares or links data across systems
- Uses data for research, training, or development
- Reuses data for new purposes
- Keeps changes, updates and profiling over time

2 The Consent & Notice Chain

Information is provided to patient
↓
Patient understands the information
↓
Patient is given real choice
↓
Patient gives informed consent
↓
Consent is recorded and linked
↓
Any change in use is reviewed
↓
Consent is updated or renewed
↓
Consent remains valid and current

Consent is not a checkbox.
It is an ongoing relationship of trust and information.

3 The Contradiction

Use patient data in ways beyond what was explained or consented to -> but do not visibly prove valid, informed, and ongoing consent

- What did the patient know at the time?
- Was the purpose clearly explained?
- Was the use changed later?
- Was consent re-checked or updated?
- Can the patient withdraw consent easily?
- Is consent linked to specific uses?
- Is consent valid for profiling or AI?
- Was the patient told about data sharing?
- Can the patient see how their data is used?
- Is there proof of understanding?
- What happens if the patient says no?
- How is consent kept up to date?

4 Visible Endpoint Consequences

- Loss of patient autonomy and choice
- Lack of transparency and trust
- Hidden uses and function creep
- Legal and regulatory non-compliance
- Ethical failure and public harm
- Erosion of trust in the NHS
- Reputational and systemic damage

NHS England ---- ICBs / trusts ---- GP practices ---- Vendors / AI developers ---- Cloud providers ---- Patients & public

Consent must be clear, informed, specific, revocable, and visible at every use.

Trust is earned through truth, clarity, and choice – not through silence.

This is not ordinary IT permission. This is a patient's right to know and choose.

Without full notice and valid consent, the use *is not lawful or ethical.*

NHS
AI Disclosure Series

Breach 13

Cybersecurity Blind Spot

Authorised Use as Exposure Pathway

1 Authorised NHS User

Doctor / Admin / Staff with access

2 Patient Record / NHS System

EHR / Clinical System / Databases

3 AI Layer / Integration

AI Model / Plugin / Middleware

- Cloud Processing
- Vendor Support Access
- Telemetry / Logs
- Sub-processors / APIs

Exposure can occur without a hack

Traditional cyber view

Focuses only on external attack

- Risk is not only unauthorised intrusion.
- Normal permitted use can still move data outward.
- Cloud, support, logging and integrations widen the exposure chain.
- Patients may be exposed without any visible breach event.

Real exposure path

Permitted access + hidden downstream pathways

If access is authorised but the pathway is not tightly controlled, exposure is still exposure.

NHS
AI Disclosure Series

Breach 14

AI Hallucination, Verification Burden & Liability Transfer

1 Clinical input / source record

Consultation notes, history, tests, context

2 AI draft output

Examples: summary, discharge note, triage output, draft letter

3 Human review under pressure

High workload, time pressure, competing demands

4 Final decision / signed output

Approved note, record entry, communication sent

False output can be accepted as true

Vendor claim

- Saves time
- Reduces admin burden
- Supports accuracy

Operational reality

- Output may hallucinate or smooth over missing facts
- Verification burden returns to staff
- Time pressure encourages rubber-stamping
- Liability remains with the human user / organisation

Core contradiction

The system is sold as time-saving, but safety depends on time-consuming verification.

- **Generative output is probabilistic, not guaranteed fact.**
- **Every unsafe draft must be caught before sign-off.**
- **If full verification is required, the workload is not removed.**
- **Risk is transferred downstream to staff and patients.**

Liability shift

The AI supplier warns of errors; the clinician and NHS carry the consequence.

If it must be constantly verified, the burden is not removed — it is relocated.

BREACH 15 VENDOR, SUB-PROCESSOR, CLOUD, TELEMETRY AND SUPPORT-ACCESS CHAINS NOT VISIBLY MAPPED

Unseen supply chains create unaccountable data routes and uncontrolled risk.

NHS AI DISCLOSURE

THE ISSUE

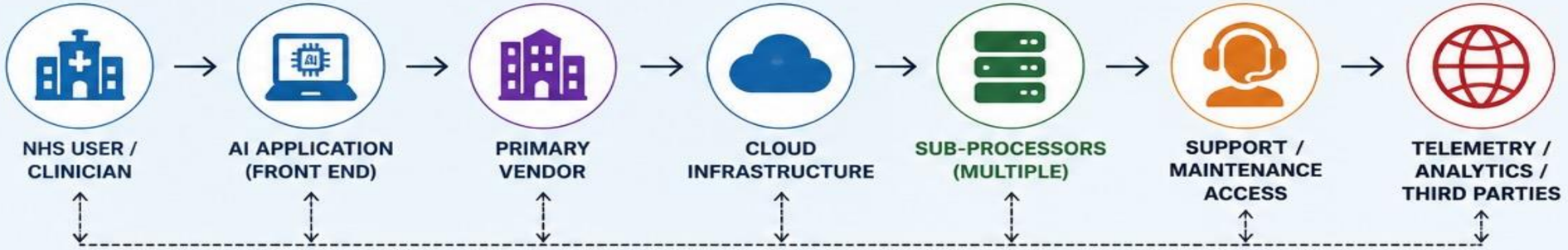
The full chain of handling for NHS AI systems — including vendors, sub-processors, cloud providers, telemetry collectors and remote support services — is not publicly mapped or disclosed.

This creates blind spots where personal data may be accessed, copied, processed or stored outside the UK and outside direct NHS control.

KEY RISKS

- Unseen onward transfers of patient data
- Jurisdictional and legal enforcement exposure
- Remote access by support staff with broad privileges
- Telemetry and analytics creating secondary data profiles
- No contractual visibility on sub-processor behaviour
- Incident response gaps across multiple parties

TYPICAL AI SUPPLY CHAIN (EXAMPLE)



BLIND SPOTS: EACH LINK MAY INVOLVE UNDISCLOSED SUB-PROCESSORS, CROSS-BORDER TRANSFERS, OR REMOTE ACCESS BEYOND NHS VISIBILITY

EXAMPLES OF HIDDEN TOUCHPOINTS

- Cloud region replication and backup providers
- AI model training services and data enrichment partners
- Telemetry pipelines and product analytics tools
- Error logging and performance monitoring platforms
- Customer support portals and remote diagnostic tools
- Identity providers and authentication services
- Software update servers and content delivery networks
- Third-party libraries, plugins and open-source components

WHY THIS MATTERS

- Patients have no meaningful control or visibility over these chains.
- Contracts may permit onward sharing without explicit patient knowledge.
- Legal responsibility remains with the NHS even when processing is outsourced.
- Breach or misuse anywhere in the chain becomes an NHS breach.

REQUIRED ACTIONS

- [x] Full supplier and sub-processor registers must be published.
- [x] Data flow maps must show every processor, location and purpose.
- [x] Support and admin access must be least-privilege, logged and reviewed.
- [x] Telemetry must be minimised, justified and transparent.
- [x] Patients must be informed in clear language about all chains.
- [x] Regular independent audits of the entire processing ecosystem.

LEGAL & ETHICAL FRAMEWORK

UK GDPR (Art. 13, 14, 28) · Data Protection Act 2018 · NHS Constitution
Caldicott Principles · Common Law Duty of Confidentiality

TRUTHFARIAN PRINCIPLE

Truth exists only where every link in the chain is known, owned, minimised and accountable.

EQUILIBRIUM IMPACT

Unmapped chains increase Ownership-Load (Ω) and decrease Coherence (ΔC).

$$Eq(S) = \Sigma(\Delta C - \Delta\Omega) < 0 \Rightarrow \text{System in Imbalance}$$

BREACH 16

ACCOUNTABILITY GAP BETWEEN NHS ENGLAND, ICBs, GP PRACTICES, VENDORS, AND STAFF

NHS

Case Reference: NHS AI Audit Disclosure Series

When everyone touches the AI chain, no one can be allowed to disappeer from responsibility.

THE AI ACCOUNTABILITY CHAIN – MANY ACTORS, ONE PATIENT, NO CLEAR ACCOUNTABLE OWNER

1. NHS ENGLAND
Sets national policy, standards and digital strategy. Contracts frameworks. Delegates delivery.

2. ICBs
Commission services, approve tools, manage budgets, oversee risk – but vary widely.

3. GP PRACTICES
Deploy AI tools in clinical workflows. Rely on vendors, train staff, manage local risk.

4. VENDORS
Design, build and supply AI systems. Control models, updates and performance.

5. CLOUD / HOSTING PROVIDERS
Store and process NHS data. Provide infrastructure, uptime and security controls.

6. STAFF / USERS
Use AI outputs in real time. Make decisions. Carry liability without control over system.

THE PATIENT
The person who bears the risk, harm and consequence.

THE ACCOUNTABILITY GAP (THE BREACH)

This fragmented structure creates a systemic accountability void:

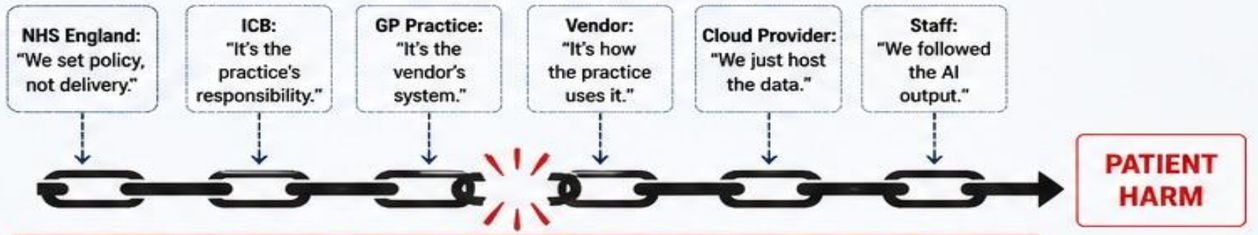
- No single organisation is clearly responsible for end-to-end AI safety.
- Contracts often disclaim liability between organisations and vendors.
- Regulators rely on self-reporting by the same organisations they regulate.
- When harm occurs, each actor points to another party.
- Patients and staff cannot identify who is legally accountable for the harm.

RESULT: RESPONSIBILITY IS FRAGMENTED BEFORE THE PUBLIC CAN EVEN IDENTIFY WHO IS ACCOUNTABLE.

EXAMPLES OF HARM WHERE ACCOUNTABILITY BREAKS DOWN

- AI gives wrong clinical suggestion
- AI misses red flag in data
- Patient data leaked via vendor system
- Bias causes unequal treatment
- Record fabricated or altered
- Staff rely on AI and are blamed

THE BROKEN CHAIN IN ACTION



NO ONE OWNED THE OUTCOME. EVERYONE CONTRIBUTED TO THE RISK.

WHY THIS IS A LEGAL & ETHICAL BREACH

- Violates the duty of accountability under UK GDPR Article 5(2)
- Breaches NHS AI Lab principles (Safety, Transparency, Accountability)
- Contravenes the NHS Constitution principle: "Accountable to patients" and the public for the services we provide"
- Creates unlawful risk of vicarious liability exposure for staff
- Undermines public trust and the rule of law

REAL-WORLD CONSEQUENCES

- Harm, error or bias in AI output
- Unlawful access or data misuse
- Unsafe triage or clinical delay
- False or incomplete records
- Staff blamed, disciplinaries, or struck off
- Patients left without remedy
- No organisation accepts responsibility

WHAT MUST HAPPEN

- Clear, documented accountability at each level
- Contracts that enforce shared, non-disclaimed liability
- Mandatory AI risk governance and audit trails
- Independent regulation with enforcement powers
- Transparency to patients and staff
- A single accountable owner for each system and outcome

PUBLIC INTEREST PRINCIPLE

Public services using AI must be accountable, transparent and answerable at every point in the chain. No more hiding behind layers.

ACCOUNTABILITY IS NOT OPTIONAL. IT IS THE LAW.

BREACH 17

LACK OF THIRD-PARTY AI ASSURANCE, PENETRATION TESTING & INDEPENDENT VALIDATION

NHS AI tools are deployed without independent assurance of safety, accuracy, security or bias.

NHS
AI DISCLOSURE SERIES
Case Reference: NHS AI Audit

THE ISSUE

NHS AI systems are procured and deployed without mandatory independent testing, red teaming, bias audits or clinical validation. Reliance on vendor self-assessment creates blind spots that can result in harmful outputs, unsafe recommendations and unlawful processing of patient data.

Without third-party assurance, the NHS cannot evidence that AI tools are safe, accurate, secure or fit for clinical purpose.

KEY RISKS

- Undetected bias and discriminatory outcomes
- Unsafe or incorrect clinical recommendations
- Security vulnerabilities remain untested
- Prompt injection and adversarial attacks
- Hallucinations and false information in records
- Model drift and performance degradation
- Non-compliance with UK GDPR and AI standards
- Patient harm and loss of public trust

WHERE THE ASSURANCE GAP EXISTS

NHS / ICB
Procures and approves AI tools without independent verification

→ ✕ →

VENDOR SELF-ASSESSMENT
Vendors validate their own systems (conflict of interest)

→ ✕ →

NO INDEPENDENT TESTING
No red teaming, penetration testing, or bias auditing

→ ✕ →

NO CLINICAL VALIDATION
No independent clinical safety validation or real-world testing

→ ✕ →

UNMONITORED DEPLOYMENT
AI used in practice without continuous independent oversight

→ ✕ →

PATIENT AT RISK
Harm, error or unlawful outcomes may occur

NO INDEPENDENT ASSURANCE = NO EVIDENCE OF SAFETY, ACCURACY OR COMPLIANCE
This is a systemic assurance failure and a breach of legal and ethical duties.

WHY INDEPENDENT ASSURANCE IS ESSENTIAL

- Detects bias, drift, hallucinations and safety risks
- Tests security against real-world threats
- Validates clinical safety and performance
- Confirms compliance with law and standards
- Provides evidence of due diligence
- Protects patients and staff
- Maintains public trust and accountability

LEGAL & REGULATORY STANDARDS

- UK GDPR Article 32 – Requires appropriate technical and organisational measures, including testing and evaluation.
- AI Regulation (EU) – Requires conformity assessment, robustness and accuracy testing for high-risk AI.
- NHS AI Lab Principles – Safety, Transparency, Accountability and Assurance.
- Common Law Duty of Care – Duty to ensure systems used in patient care are reasonably safe and fit for purpose.

REAL-WORLD CONSEQUENCES

- Harm to patients through unsafe or incorrect AI outputs
- Unlawful discrimination and inequality
- Data breaches and cyber exploitation
- Regulatory enforcement and fines (ICO, CQC)
- Litigation, compensation and reputational damage
- Loss of public trust in NHS services
- Staff liability exposure

REQUIRED ACTIONS

- Mandate third-party assurance for all NHS AI tools
- Independent red teaming and penetration testing
- Bias audits and fairness testing
- Clinical safety validation by independent experts
- Ongoing monitoring, logging and re-assurance
- Assurance reports must be public and auditable

ACCOUNTABILITY PRINCIPLE

If the NHS does not independently test and validate AI tools before and during deployment, it cannot claim it took reasonable steps to protect patients.

No assurance. No safety. No accountability.

TRUTHFARIAN EQUILIBRIUM IMPACT

Failure to assure creates hidden harm ($\Delta\Omega \uparrow$), while coherence and safety ($\Delta C \downarrow$) are not increased.

$Eq(S) = \Sigma(\Delta C - \Delta\Omega) < 0 \rightarrow$ System in Imbalance

Sustained imbalance = Institutional Failure.

TRUTHFARIAN DISCLOSURE: Patients have a right to assurance, evidence and accountability for every AI system used in their care.

www.truthfarianism.com

BREACH 18

NO CLINICAL ACCOUNTABILITY FOR AI DECISIONS MADE OUTSIDE THE CLINICIAN–PATIENT RELATIONSHIP

AI decisions that affect patient care are made without a responsible clinician in the loop. This is a breach of professional duty, patient safety law and constitutional principle.

NHS
AI DISCLOSURE SERIES

CASE REFERENCE: NHS AI AUDIT DISCLOSURE SERIES

THE ISSUE

AI systems make or influence clinical decisions (triage, prioritisation, diagnosis suggestions, risk scores, referrals, record creation) without a responsible, identifiable clinician being directly involved.

This breaks the foundational clinician–patient relationship and removes legal and professional accountability.

WHY THIS IS A BREACH

- Violates the GMC duty of personal responsibility (Good Medical Practice)
- Breaches the NHS Constitution principle: "Accountable to patients" for the care we provide"
- Contravenes patient safety law (duty of candour, safe systems of care)
- Creates unlawful risk of harm without accountable oversight
- Undermines Article 2 (Right to Life) and Article 8 (Right to Private Life) – HRA 1998

CORE BREACH STATEMENT

AI IS MAKING OR INFLUENCING CLINICAL DECISIONS WITHOUT A CLINICIAN OWNING THE DECISION.

If no clinician owns it, no one can be held accountable when harm occurs.

HOW AI DECISIONS ARE MADE OUTSIDE THE CLINICIAN–PATIENT RELATIONSHIP

AI SYSTEM
Processes data and generates outputs (e.g., triage score, risk flag, diagnosis suggestion)

SYSTEM WORKFLOW
Output is consumed by systems, dashboards or staff without mandatory clinician review

DECISION ACTIONED
AI output drives or influences clinical action (referral, treatment delay, record entry, etc.)

NO CLINICIAN OWNERSHIP
No identifiable clinician has personally reviewed, validated and owned the decision

PATIENT AT RISK
Harm, delay, error or inappropriate treatment may occur

NO CLINICIAN • NO OWNERSHIP • NO ACCOUNTABILITY • NO SAFETY
This is not innovation. This is abdication of professional duty.

REAL-WORLD CONSEQUENCES

- Mis-triage or delayed treatment
- False reassurance or missed red flags
- Inappropriate referrals or discharges
- Automated record creation errors
- No accountable clinician to answer when harm occurs
- Loss of public trust in NHS services
- Staff exposed to blame without control

EXAMPLES OF AI DECISIONS WITHOUT CLINICIAN OWNERSHIP

TRIAGE & PRIORITISATION
AI assigns scores that determine who is seen first – no clinician review.

DIAGNOSTIC SUGGESTIONS
AI suggests diagnoses which are accepted without clinician validation.

REFERRAL ROUTING
AI routes patients to services or rejects referrals automatically.

RECORD CREATION
AI drafts clinical notes or codes without clinician confirmation.

RISK STRATIFICATION
AI labels patients as 'low risk' – care is reduced or withdrawn.

LEGAL & PROFESSIONAL DUTIES

GMC Good Medical Practice
Doctors must take personal responsibility for all decisions in their care.

NHS Constitution
Patients have the right to care that is safe, accountable and overseen by qualified staff.

Duty of Candour
Requires openness when things go wrong.

Human Rights Act 1998
Protects the right to life (Art. 2) and private life (Art. 8).

Data Protection Act 2018 / UK GDPR
Decisions that materially affect individuals require appropriate human oversight (Art. 22, 5(1)(f)).

THE ACCOUNTABILITY GAP

NO ONE CAN BE HELD ACCOUNTABLE. THE PUBLIC LOSES. PATIENTS ARE PUT AT RISK.

WHAT MUST HAPPEN

- Every AI output that affects care must be reviewed and owned by a named clinician.
- Mandatory human-in-the-loop before action is taken.
- System design must enforce clinical accountability.
- Audit trails must show who reviewed, and what decision was made.
- No patient-facing AI decision without clinician validation.

ACCOUNTABILITY PRINCIPLE

If a decision affects a patient, a real clinician must own it.

No ownership. No safety. No accountability. No law.

TRUTHFARIAN EQUILIBRIUM TEST

$$Eq(S) = \Sigma(\Delta C - \Delta\Omega) \geq 0$$

When AI decisions occur without clinician ownership:

ΔC (coherence) ↓ | $\Delta\Omega$ (ethical load) ↑

$Eq(S) < 0$ → System in Imbalance (Untruthful)

Truthfarian Disclosure: Patients have a right to safe, accountable, human-supervised care. AI must serve care, not replace responsibility.

www.truthfarianism.com

NHS

The Truthfarian Principle

"Truth exists only where every link in the chain is known, owned, minimised and accountable."

Patients have a right to safe, accountable, human-supervised care.

AI must serve care, not replace responsibility.