Policy Brief on Cybersecurity in Healthcare

George Karavokiros (BYTE), Sotiris Athanasopoulos (Hygeia), Elias Dakos (ATC), Artemis Voulkidis (Synelixis), Thomas Selegny (RESALLIENCE)

September 2025

ATLANTIS

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement No. 101073909.

ATLANTIS LSP#2
Policy Brief on Cybersecurity in Healthcare

Executive Summary

Europe's health sector remains a high criticality, mid maturity domain. ENISA's latest NIS360 assessment places healthcare at the upper end of moderate maturity yet still in the risk zone, with larger organisations driving most of the improvement and many smaller providers struggling with basic hygiene and legacy systems.

Key Challenges

- **AI-driven disinformation** undermining trust in vaccines, treatments, and emergency health measures.
- **Cyberattacks on hospital IT systems** disrupting clinical workflows (EHR, imaging, labs), delaying treatments and risking patient safety.
- **Fragmented cross-border coordination** slowing collective responses to health-related disinformation and cyber incidents.
- **Limited operational capacity** to deliver rapid, targeted corrections at scale.
- **Operational impacts on hospitals** (ED crowding, elective cancellations, staff redeployment) triggered by coordinated online narratives.

This brief builds on clinical and operational lessons from the ATLANTIS Large Scale Pilot 2 (LSP#2) at Hygeia Hospital, Athens and results in prioritized and actor specific policy recommendations, on three strategic priorities:

1. **Accelerate automated counter-narratives**. Build AI assisted drafting and dissemination so trustworthy information can be produced and deployed quickly, with humans maintaining oversight.
2. **Deliver in-channel corrections**. Place authoritative messages inside the same communities and platforms and through hospital-controlled channels (patient portals, appointment SMS, on-site signage) so the right audiences actually see them.
3. **Strengthen cross-border coordination**. Enable authorities and partners to share a common operating picture and act together through agreed protocols, joint tooling and regular exercises.

Introduction

According to ENISA's report on cybersecurity maturity and criticality assessment of NIS2 sectors (ENISA, 2025), health is assessed at the upper end of *moderate* maturity and sits mid-table overall, as shown in Figure 1. Health sits below leaders such as electricity, telecoms and banking and is classed in the "risk zone". These are the sectors that rank comparatively lower than others in terms of maturity but have a criticality score that is higher than their maturity score.

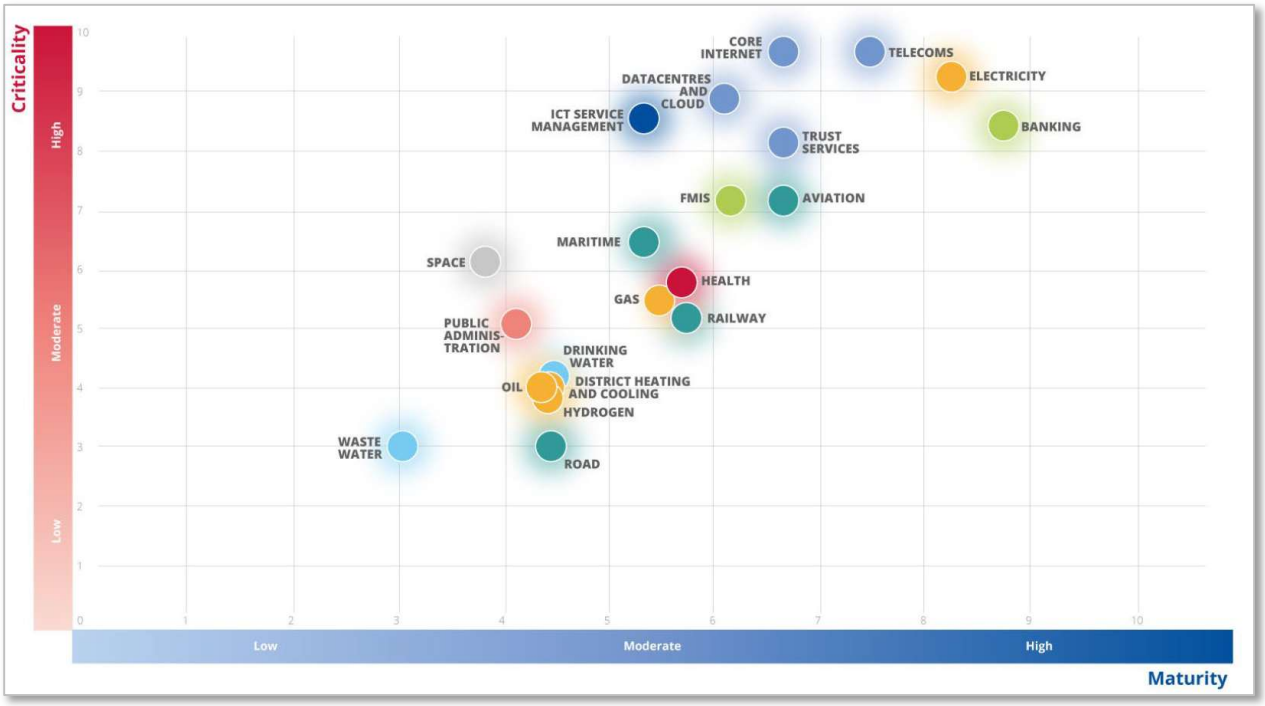

Figure 1: Cybersecurity maturity and criticality overview by sector according to (ENISA, 2025).

The **ATLANTIS LSP#2** exercise at Hygeia Hospital tested realistic cyber-physical scenarios, including:

- Manipulation of **electronic health records (EHRs)** to alter diagnoses or treatment plans.
- **DDoS attacks** on eHealth platforms, disrupting telemedicine and appointment systems.
- **Misinformation surges** during a public health crisis, eroding trust in official guidance.

- **Coordinated rumours** triggering cancellations or spikes in demand for specific services (e.g., oncology, paediatrics), disrupting scheduling and bed management.

LSP#2 at Hygeia moved beyond generic infodemic guidance by exercising an end to end response inside a working hospital, pairing cyber physical incidents (EHR manipulation and DDoS on eHealth and appointment systems) with a surge of health misinformation to test joint situational awareness and human supervised in channel corrections. These scenarios revealed three urgent needs:

1. **Automated counter-narratives** to contain harmful misinformation before it spreads widely.
2. **In-channel corrections** via both social platforms and hospital systems (EHR portals, SMS reminders, call-centre scripts) to reach affected audiences quickly and effectively.
3. **Cross-border coordination** for a unified, timely response to both cyber and information threats

The following sections set out concrete measures, roles and implementation steps informed by LSP#2.

Promoting Automated Counter-Narratives and In-Channel Corrections

Current Status

Disinformation has long been used by state and non-state actors to influence opinion, erode trust in institutions, and advance political or economic aims. Two developments have greatly expanded its reach and sophistication: the global spread of social media and the rise of Artificial Intelligence (AI) tools for content creation. Platforms such as Facebook and X use algorithms that amplify provocative or misleading narratives, exploiting filter bubbles and microtargeting to reach millions instantly. At the same time, AI-driven text, image, and video synthesis enables the creation of convincing but fabricated testimonials, citations, and deepfakes at scale, lowering barriers for large-scale influence operations and complicating detection, attribution, and policy responses.

The COVID-19 pandemic exposed the healthcare sector's acute vulnerability to such campaigns. The surge of mixed-quality information during crises (termed an Infodemic) overwhelms the public's ability to identify reliable guidance, posing serious health risks. In 2020, the World Health Organization (WHO) formally recognised this phenomenon (Posetti & Bontcheva, 2020). At EU level, the European Commission and High

Representative issued a Joint Communication on COVID-19 disinformation, a factsheet of immediate measures, and later guidance to strengthen the Code of Practice on Disinformation (European Commission and High Representative, 2020; European Commission, 2021). Research warns that fast-spreading misinformation undermines health behaviours, trust in healthcare and governance, and can destabilise societal cohesion and democracy (Zielinski, 2021; Borges do Nascimento et al., 2022). The ease of AI-enabled content generation and rapid social media dissemination heightens these risks.

Gap - Underdeveloped Automation

AI-enabled monitoring platforms now scan social media and online news in real time, using natural language processing and machine learning to flag suspect content and alert authorised users, an approach also used in the ATLANTIS project. Yet producing effective health counter-narratives, such as patient-facing explanations, corrections to unsafe home remedies, vaccine/treatment clarifications and service-status updates, remains largely manual, slowing responses during fast-moving crises. Corrections often fail to reach the same audiences exposed to the falsehood, especially when misinformation circulates within closed communities. The practice of "in-channel correction," delivering rebuttals in the same spaces where false narratives spread, is still underdeveloped.

While malicious actors increasingly deploy bots, algorithms, and Large Language Models (LLMs) to generate and disseminate false content at scale, defensive automation lags behind. Fully autonomous AI debunking remains rare, with current systems requiring substantial human input, making them "less powerful to tackle disinformation than [AI is] to create it" (Aouati et al., 2024). This asymmetry leaves defenders outpaced.

Automated counter-narratives also pose risks, i.e., false positives, suppression of legitimate criticism, bias against smaller languages, adversarial manipulation, and privacy or accountability concerns. To address censorship concerns, automation should be done under transparent criteria and public guidance, prioritising additive corrections such as labels and replies rather than removal. Health messaging should be separate from platform enforcement, supported by human in the loop review for high impact cases and independent oversight that includes smaller Member States and civil society. Activation should be proportionate and time limited, with auditable logs, regular reporting and an available appeal route.

Policy Need

The EU should lead efforts to develop AI-assisted counter-narrative systems, focusing on: (a) **Content generation**, using AI to produce accurate, compelling rebuttals or myth-busting posts as soon as threats are detected; and (b) **Content dissemination**, automating delivery into the same channels where falsehoods spread, including delivery

through hospital-controlled channels (patient portals, appointment SMS, IVR/call-centre scripts) for patients already engaged in care pathways. For example, an AI system overseen by human fact-checkers could generate verified corrections to a trending health hoax and circulate them as replies, comments, or banners. Under the DSA's risk-mitigation provisions, platforms should be encouraged or required to flag viral health misinformation and enable in-platform corrections, such as pinning fact-checks or notifying users who shared false content, while enabling public-health authorities and hospitals to 'whitelist' authoritative health messaging for rapid elevation during incidents. The steps in the proposed response flow are shown in Figure 2.

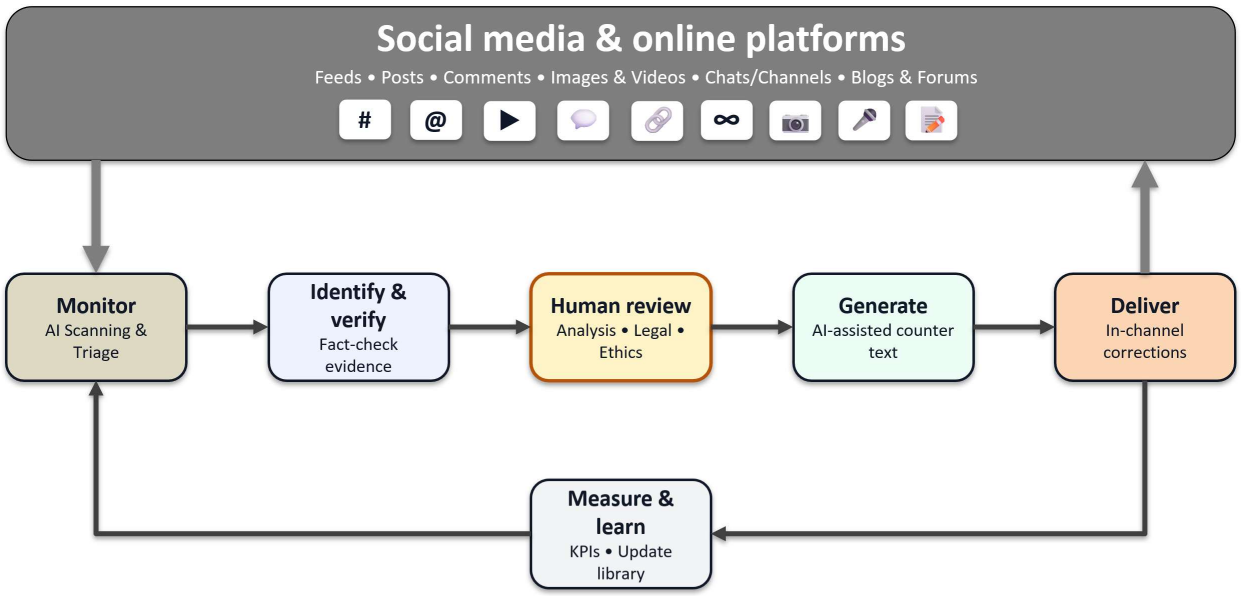

Figure 2: Proposed health-misinformation response flow

Trustworthiness is essential: AI models must ensure accuracy, fairness, and robustness, supported by transparent governance, auditability, and continuous monitoring. Initiatives like the THEMIS 5.0 project can embed bias detection and mitigation throughout system design. The **EU AI Act** and the **Ethics Guidelines for Trustworthy AI** (European Commission, 2024) reinforce these principles through legal and ethical requirements to minimise bias and promote diversity.

Automation can also power **counter-narrative libraries**, pre-crafted messages on recurring disinformation themes, adaptable and deployable within minutes, aided by machine translation for multilingual reach. Authorities could also use **cell broadcast alerts** to deliver urgent health messages directly to citizens, bypassing platforms, particularly valuable when moderation weakens, as seen in a major platform's 2023 withdrawal from the EU's voluntary Code of Practice (Deutsche Welle, 2023; Reuters, 2023). Protocols should define when such emergency messaging is justified, e.g. in cases of dangerous medical misinformation.

EU funding programmes (**Horizon Europe**, **Digital Europe**) should prioritise counter-disinformation innovation, building on ATLANTIS and LSP#2 tooling (situational awareness, disinformation campaign monitoring tools etc.). Practical solutions, such as dashboards for rapid cross-platform debunking and browser extensions showing official rebuttals to exposed users, can shorten the feedback loop, ensuring accurate information is injected quickly into the same streams as the falsehood.

Improving Cross-Border Collaboration and Coordinated Responses

Current Status

Health disinformation often crosses borders as patients travel, care pathways link across Member States and clinical narratives spread across languages, yet responses remain largely national. Within the EU, Member States maintain their own, unevenly resourced structures. Collaborative mechanisms exist, such as the **Rapid Alert System (RAS)** launched in 2019 for information exchange (EEAS, 2024), the **EU Hybrid Fusion Cell**, and **EEAS** initiatives like the **East StratCom Task Force**. The EU also cooperates internationally via the **G7 Rapid Response Mechanism** and **NATO StratCom**.

Despite these measures, coordination is still ad hoc and insufficient for the scale of the threat. In healthcare, a false narrative about a disease or treatment can quickly move from one country's online sphere to another's. A 2024 global policy review identified a key gap: the "absence of a standardised, global framework for addressing cross-border misinformation," resulting in uneven enforcement and siloed national action (Pangotra, 2024). In the EU, this means one state may counter a health hoax swiftly while a neighbour remains unaware, enabling exploitation of coordination gaps (Sessa et al., 2024). Following COVID-19, the **Council of the EU** urged stronger collective resilience and coordinated counter-disinformation efforts across Europe (Aouati et al., 2024).

Gap - Fragmented Response Framework

Despite the cross-border nature of disinformation, the EU's response remains fragmented along national lines. Use of the **RAS** has been limited, with "relatively few highly engaged EU member states" sharing information and "major differences" in threat perception undermining trust, allowing false health narratives to be debunked in one country yet spread unchecked in another (Pamment, 2020). The **European Court of Auditors** found in 2021 that Member States were not using the RAS to its full potential, noting it had "never issued alerts" or coordinated joint attribution as intended (European Court of Auditors, 2021). Since then, progress includes daily RAS updates during the Ukraine war

and new initiatives such as the **FIMI toolbox**, **ISAC**, and election-focused measures. Still, the core challenge is achieving consistent engagement and trust across all Member States.

Specialised bodies like the **EEAS StratCom Task Force**, **European Digital Media Observatory**, and **EU Hybrid Fusion Cell** address information manipulation but none functions as a 24/7 crisis coordination centre akin to the Emergency Response Coordination Centre for disasters. Current platforms focus on information exchange and strategic guidance, not real-time operational management of infodemics. The **Digital Services Act** (Regulation (EU) 2022/2065) imposes binding obligations on very large platforms to assess systemic risks (Art. 34) and mitigate them (Art. 35), including boosting authoritative health information. While this offers a legal framework for managing future infodemics, the EU still lacks a permanent operational hub to coordinate rapid, cross-border action.

Policy Need

Europe needs stronger cross-border and cross-sector coordination to counter health disinformation crises, integrating public-health (DG SANTE, HERA, ECDC) with cyber and platform responses. A model could be drawn from the **Emergency Response Coordination Centre (ERCC)**, which operates 24/7 for disaster relief. An equivalent **EU "Information Emergency" hub** could unite experts from affected Member States, EU bodies (DG ECHO, DG Santé, EEAS StratCom), and platforms to share real-time intelligence and deliver unified responses, joint public messages, pooled translation resources, and synchronised debunking across all national media. Speaking with one voice would limit opportunities for adversaries to exploit national divides.

The **European Commission** could establish cross-border protocols for health misinformation, leveraging the **ECDC**'s new communication network or **HERA** to coordinate and standardise measures. This might include regular multi-country drills and upgrading the **Rapid Alert System** with stronger analytics and faster dissemination so all partners can act within hours of a flagged rumour. The **European Civil Protection Mechanism** could also expand to cover "disinformation emergencies," recognising their potential societal disruption on par with natural disasters or epidemics.

Conclusions

Healthcare has improved its cybersecurity maturity but still sits in the risk zone where gaps translate into patient-safety risk, throughput loss and capacity constraints. Lessons from LSP#2 show that joint training, shared situational awareness, and systematic learning can raise readiness across the system. EU policymakers should:

1. **Automate and accelerate** trusted message delivery.
2. **Embed healthcare-specific protocols** into cross-border coordination.
3. **Measure and track** response effectiveness through clear KPIs.

Three priorities should guide actor-specific prioritized actions, as in Table 1. Use automation to draft and deliver corrective messages at speed with humans in the loop. Deliver in-channel corrections so trusted information reaches the same audiences as the false claims. Strengthen cross-border coordination to act on a common operating picture.

Table 1: Prioritized actions and responsible actors

Timeframe	Action	Responsible Actors
Short-term	• Pilot AI-assisted counter-narrative tools • Build sector-specific libraries • Conduct targeted hospital drills	EU Commission (funding), Member States, hospitals, platforms
Medium-term	• Standardize EU-wide protocols • Integrate hospitals into RAS workflows • Run cross-border exercises	EU Commission (DG SANTE and HERA, DG ECHO, DG CNECT), HERA, ECDC, health ministries, national patient-safety authorities
Long-term	• Establish Information Emergency Hub • Embed KPIs • Secure sustainable funding	EU Commission (DG ECHO/ERCC, DG SANTE and HERA, DG CNECT), ECDC, Member States, EEAS StratCom

The Digital Services Act provides a legal basis for risk assessment and mitigation by very large online platforms, but an operational gap remains. The EU still lacks a standing information emergency function to coordinate rapid responses. Targeted funding should support the development of practical tools and exercises and progress should be tracked through clear outcome metrics such as:

- **Time to detect misinformation.**
- **Time to correct.**
- **% of exposed audience reached** (incl. patients already on care pathways).
- **Engagement rate with counter-narratives**.
- **Clinical/operational recovery**:
    - time to restore critical services

- % missed appointments rescheduled within 72h
- ED inflow normalisation time.

Acting now will close operational gaps, protect patient safety, and strengthen resilience before the next information-driven health crisis. Public trust in health institutions and information is the ultimate goal.

References

Aouati, O., Freguglia, P., Heiss, R., Patras, S., Pavlou, P., Pelsy, F. & Truc, M. (2024, July) *How to reduce the impact of disinformation on Europeans' health* (Study for the Subcommittee on Public Health, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament; PE 754.205). Luxembourg: European Parliament.

Borges do Nascimento, I.J., Pizarro, A.B., Almeida, J.M., Azzopardi Muscat, N., Gonçalves, M.A., Björklund, M. & Novillo Ortiz, D. (2022) 'Infodemics and health misinformation: a systematic review of reviews', *Bulletin of the World Health Organization*, 100(9), pp. 544–561. Available at: https://doi.org/10.2471/BLT.21.287654

Deutsche Welle (2023) 'EU: Twitter leaves voluntary pact on fighting disinformation'. Available at: https://www.dw.com/en/eu-twitter-leaves-voluntary-pact-on-fighting-disinformation/a-65751487

EEAS (2024) *European External Action Service*. Available at: https://eeas.europa.eu

European Commission (2021) *European Commission Guidance on Strengthening the Code of Practice on Disinformation*. COM(2021) 262 final, 26 May. CELEX: 52021DC0262.

European Commission (2024) *Digital Strategy*. Available at: https://digital-strategy.ec.europa.eu

European Commission and High Representative (2020) *Tackling COVID-19 disinformation: Getting the facts right*. JOIN(2020) 8 final, 10 June. CELEX: 52020JC0008.

European Court of Auditors (2021) *Special Report No 09/2021: Disinformation affecting the EU*. Luxembourg: Publications Office of the European Union.

European Union Agency for Cybersecurity (ENISA) (2025) *ENISA NIS360 – 2024: ENISA Cybersecurity Maturity & Criticality Assessment of NIS2 sectors*. ENISA. DOI: 10.2824/1378797. ISBN: 978-92-9204-687-3.

Menz, B., Modi, N., Sorich, M. & Hopkins, A.M. (2023) 'Health disinformation use case highlighting the urgent need for artificial intelligence vigilance – Weapons of mass disinformation', *JAMA Internal Medicine*, 184(1), pp. 92–96.

Pamment, J. (2020, July 15) *The EU's role in fighting disinformation: Taking back the initiative*. Carnegie Endowment for International Peace.

Pangotra, A. (2024, October 29) *Addressing cross-border misinformation: The need for international cooperation*. CyberPeace.

Posetti, J. & Bontcheva, K., United Nations Educational, Scientific and Cultural Organization (2020) *Disinfodemic: deciphering COVID-19 disinformation*. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000374416

Reuters (2023, May 26) 'Twitter cannot hide EU rules after exit code, EU's Breton says'. Available at: https://www.reuters.com/technology/twitter-cannot-hide-eu-rules-after-exit-code-eus-breton-says-2023-05-26

Sessa, M.G., Serrano, R.M., Romero Vicente, A., McNamee, J., Gentil, I. & Alaphilippe, A. (2024, October) *Countering disinformation: Issues and solutions for EU decision makers*. EU DisinfoLab.

Zielinski, C. (2021) 'Infodemics and infodemiology: a short history, a long future', *Revista Panamericana de Salud Pública*, 45, p. e40. Available at: https://doi.org/10.26633/RPSP.2021.40