ISSUE 5 · 2026-06-26 · FDA AI-ENABLED DEVICE CYBER TESTING

Your AI-enabled medical device is about to get a deficiency for cyber testing

Issue 5: If your device has AI in it, FDA already asked for more cyber testing than you are doing

KEY TAKEAWAYS

- In 2025 FDA cleared 295 AI/ML-enabled devices, roughly 97% of them through 510(k), and most almost certainly did not perform the cybersecurity testing FDA's January 7, 2025 draft guidance asks for.
- Section XII of that draft (starting on page 34) changes the scope of 'cybersecurity testing' the moment a trained model is involved: a different attack surface, tested at a different time, on different infrastructure.
- The 2023 final guidance everyone internalized aims all its testing (threat model, SBOM, security risk assessment, architecture views, vuln scanning, fuzzing, pentest under a Secure Product Development Framework) at the finished, shipping device, but a trained model breaks that boundary because most AI-specific risk lives in the pipeline that never ships.
- Section XII names seven AI-specific threat categories: data poisoning, model inversion and stealing, model evasion, data leakage, overfitting, model bias (including embedded backdoors), and performance drift.
- Only model evasion is primarily a runtime, on-device threat that a standard penetration test covers well; the other six live upstream in data ingestion, training and evaluation environments, inference endpoints, model artifacts, and the update/retraining mechanism.
- A submission resting on a pentest of the deployed model has evidence for roughly one of the seven categories, leaving the other six untested on infrastructure the pentest scope never includes.

> **"Section XII is telling you the device was never the whole system."**

The baseline most people are testing to

The September 2023 final guidance is stable and well understood, with mature activities and predictable deficiency letters, but all of its testing is aimed at the finished device and its interfaces. That boundary is right for a conventional device because the attack surface is the

deployed software, but a trained model is only the last stage of a pipeline where most AI-specific risk lives.

What FDA actually named

Section XII lists seven AI-specific cyber threats as examples: data poisoning, model inversion and stealing, model evasion, data leakage, overfitting, model bias including embedded backdoors, and performance drift. Reading each against where it actually materializes exposes the testing problem.

Six of seven are upstream of the device

Only model evasion is primarily an on-device runtime threat a standard pentest touches. Data poisoning hits the ingestion and preprocessing pipeline, inversion and stealing target the inference endpoint and model artifact, data leakage targets the training store and inference path, overfitting and bias are created during training, and performance drift is a lifecycle threat against the data distribution and retraining mechanism, so a device pentest evidences roughly one of the seven.

What Section XII asks you to submit

Beyond the 2023 baseline it wants AI-specific deltas folded into the risk management report, threat model, risk assessment, and labeling, plus at minimum fuzz and penetration testing appropriate to the model's risks, a Security Use Case View, and described controls for data vulnerability and leakage. The practical read is traceable test cases mapped to each of the seven categories spanning runtime inference and the upstream pipeline.

Why the timing is the expensive part

Conventional premarket cyber testing happens late because the device is done and the surface is fixed, but six of these seven threats were set during training and must be tested in the training environment while the data still exists. If the gap surfaces late the remediation is a retrain, not a patch, which carries schedule and budget risk of a different order at the worst possible point.

One honest caveat

The AI-enabled guidance is still draft, the comment period closed April 2025, and public deficiency feedback has not surfaced because these products are only now reaching submission in volume. The claim is not settled enforcement, only that the requirement is written down, the testing it describes differs fundamentally from today's practice, and the cost of being wrong is asymmetric.

FDA cleared 295 AI/ML-enabled devices in 2025; roughly 97% went through 510(k).

AI-enabled device software functions draft guidance posted January 7, 2025; cybersecurity requirements sit in Section XII starting on page 34.

Seven named AI-specific threat categories; 6 of 7 are upstream of the device and only model evasion is a runtime/on-device threat.

Baseline document: 'Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions,' final September 2023.

Comment period on the AI draft closed April 2025; still not finalized as of late June 2026.

Section XII minimum asks: fuzz (malformed input) testing, penetration testing appropriate to model risks, a Security Use Case View, and data controls (access controls, encryption, anonymization/de-identification).

About ELTON. ELTON is the leading AI vulnerability discovery and cybersecurity verification platform for medical devices. ELTON builds a digital twin of each device, continuously identifies new vulnerabilities across the full stack, and uses AI to determine which are actually exploitable, with every finding and every dismissal evidenced for regulatory review. ELTON meets FDA premarket (524B) and postmarket cybersecurity vulnerability testing and management requirements. Read the newsletter and talk to us at **eltoncyber.com**.