VIKRAM JHA - DEFENSIBLE AI
CONFORMANCE CHECKLIST

UTILIZATION-REVIEW AI EVIDENCE CHECKLIST

Version 1.0 - 4 August 2026
https://vikramjha.work/artifacts/utilization-review-ai-evidence-checklist

Eleven statutory duties for AI in utilization review, and the evidence each one requires you to be able to produce.

WHY IT EXISTS

Maryland defines artificial intelligence, for utilization review, as a system that varies in its level of autonomy and produces outputs that can influence physical or virtual environments. That is a description of an agent - written into state insurance law in May 2025, months before the April 2026 revised US interagency model risk guidance placed generative and agentic AI expressly outside its scope. The controls for agentic decisioning in health insurance are being specified by state insurance regulators, not by federal model-risk supervision, and they bind the vendor conducting the review as directly as the payer. Each duty below is a yes or no a system can either evidence or cannot.

WHAT IT CONTAINS

1. Scope test - the three questions that determine whether the duties bind you, including the two that reach vendors
2. The eleven duties of 15-10B-05.1(C), each with the evidence required and a test you can run
3. The ablation test for group-data reliance - the only version that produces evidence rather than assertion
4. The prohibition in (D), treated as an entitlement question rather than a process one
5. Reporting and production obligations, and the architectural precondition each one carries
6. Why the AI-involvement flag must be captured at decision time and cannot be reconstructed from logs
7. Closure rules - including why evidence produced under examination is not evidence

THE ELEVEN DUTIES

Subsection (C) requires an entity in scope to ensure each of these. The test column is the part that matters: a duty you cannot test is a duty you are asserting.

(C)(1)

What the statute requires: Determinations rest on the enrollee's own clinical history, circumstances or record.
The test: List every enrollee-specific input the tool saw for one determination. Access to the chart is not use of the chart.

(C)(2)
What the statute requires: Determinations are not based solely on a group dataset.
The test: Run the determination with and without the enrollee-specific inputs. If the output does not change, group data is deciding.

(C)(3)
What the statute requires: Criteria and guidelines comply with the requirements of the title.
The test: For a determination made six months ago, produce the criteria version then in force.

(C)(4)
What the statute requires: The tool does not replace the provider's role under 15-10B-07.
The test: Report override rate and median reviewer dwell time. A role exercised in seconds at volume is a role in name.

(C)(5)
What the statute requires: Use does not result in unfair discrimination prohibited by law.
The test: Produce the latest disparity analysis and what changed because of it. Excluding protected attributes is not a finding.

(C)(6)
What the statute requires: The tool is fairly and equitably applied, including per HHS guidance.
The test: Is it applied to some populations and not others, and is that difference justified or just rollout sequencing?

(C)(7)
What the statute requires: The tool is open to inspection for audit or compliance review by the Commissioner.
The test: Request the inspection package with five days' notice. Did it exist, or was it assembled?

(C)(8)
What the statute requires: Written policies in the filed utilization plan state how the tool is used and what oversight is provided.
The test: Read the filed description to an engineer who built it. Divergence between filing and behaviour is the finding.

(C)(9)
What the statute requires: Performance, use and outcomes reviewed and revised at least quarterly, to maximise accuracy and reliability.
The test: Does the cadence survive a mid-quarter model version change made on the provider's schedule, not yours?

(C)(10)
What the statute requires: Patient data is not used beyond its intended and stated purpose, consistent with HIPAA.
The test: Query the retrieval layer under a purpose the data was not collected for. Does it refuse, or rely on restraint?

(C)(11)
What the statute requires: The tool does not directly or indirectly cause harm to an enrollee.
The test: Name the monitoring that would catch a systematic delay. Indirect harm does not show in determination accuracy.

(D)
What the statute requires: The tool may not deny, delay or modify health care services.
The test: Try to make it issue an adverse determination without a provider decision. If only process stops it, process is the control.

REPORTING AND PRODUCTION - AND WHAT EACH PRESUMES ABOUT YOUR ARCHITECTURE
QUARTERLY REPORT STATING, PER ADVERSE DECISION, WHETHER AN AI, ALGORITHM OR OTHER SOFTWARE TOOL WAS USED
Source: 15-10A-06, field added 2025
Architectural precondition: The flag is per decision and must be captured at decision time. "The tool was in the pipeline" is a different claim from "the tool was used in making this decision."

CAUSAL ACCOUNT WHERE ADVERSE DECISIONS FOR A SERVICE TYPE RISE 10% IN A YEAR OR 25% OVER THREE
Source: 15-10A-06(a)(2), added 2026
Architectural precondition: Pre- and post-deployment comparability has to be designed in. It cannot be assembled once an examination is open.

PRODUCTION OF ALL DOCUMENTS RELATED TO AN ADVERSE DECISION, INCLUDING THOSE HELD BY A PRIVATE REVIEW AGENT ACTING FOR THE CARRIER
Source: 15-10B-21, added 2026
Architectural precondition: Production reaches through the vendor. Contractual audit rights are necessary and not sufficient - the records must exist in producible form.

HOW TO USE IT
1. Select one material historical adverse determination where a tool was involved - not a synthetic case, and not the best-documented one.
2. Work the eleven duties in order, linking the record that proves each. Do not paste explanations into the evidence column.
3. Mark anything missing, indirect or unreconstructible as a gap, with an owner and a date.

4. Have the reconstruction performed by someone outside the team that built the system. The failure mode is familiarity, not competence.
5. If a vendor conducts any part of the review, run the scope and production sections against them, not only against yourself.

Built from public standards and general practice. No client information appears anywhere in this document. Regulatory instruments move - check anything cited here against its primary source on the day you rely on it.