

The True Cost of the Validation Burden

Executive Summary

AI only creates value when its outputs can be trusted. The human effort required to validate those outputs—the validation burden—has become one of the most significant and overlooked costs in clinical AI adoption. This piece defines the burden, explains what drives it, and outlines why understanding it is essential for evaluating AI's true operational impact and total cost of ownership.

Clinical AI is becoming deeply embedded in documentation, care coordination, screening programs, specialty workflows, and quality initiatives. Across these use cases, one universal truth sits beneath every technological advance:

No matter what AI is designed to do, its outputs must be trusted before they can be used in clinical care.

That trust doesn't come from algorithms alone—it comes from **validation**. And because validation requires ongoing human review, oversight, and judgment, it carries a cost that many organizations dramatically underestimate. As health systems scale their AI initiatives, they are discovering that the most significant expense is often not the software itself—it's the *validation burden* that comes with it.

Why Validation Matters in Clinical AI

Validation isn't a one-time technical check—it's a continuous clinical workflow.

Because AI recommendations directly influence clinical decisions, validation also functions as a safety checkpoint—ensuring the information is accurate, complete, and appropriate before it affects patient care. Whenever an AI tool generates a recommendation, synthesizes clinical information, or flags an action item, trained staff must confirm that the output is accurate, clinically relevant, and safe to act upon. They must compare the AI's interpretation against the underlying patient data and ensure that the information reflects true clinical context. As these review steps accumulate across hundreds or thousands of outputs, many organizations encounter what becomes a growing *validation burden*—the hidden operational load created when humans must verify AI-generated information at scale.

Different AI technologies generate different types of outputs—some highly structured, some probabilistic, some context-dependent. Because of these variations, the amount of human oversight required can differ significantly from one AI system to another. Some outputs require only light review, while others must be examined in detail. The underlying architecture affects the *volume and complexity* of validation, which directly influences operational burden.

This is why different AI systems create very different operational workloads—and why many health systems underestimate the true cost of validation.

The Hidden Costs Most Buyers Miss

The cost of validation rarely appears on a budget proposal, but it drives a significant portion of the total cost of ownership for clinical AI.

Human review consumes:



Time

Often several minutes per output



Staffing

Additional navigators, clinicians, or quality personnel



Budget

Salary and operational expenses that scale with volume

Across thousands or tens of thousands of outputs, the burden grows rapidly. In many cases, the ongoing labor cost of validation exceeds the cost of the AI software itself. Yet these costs often remain invisible until an organization attempts to operationalize AI at scale.



How the Validation Burden Affects Care Delivery

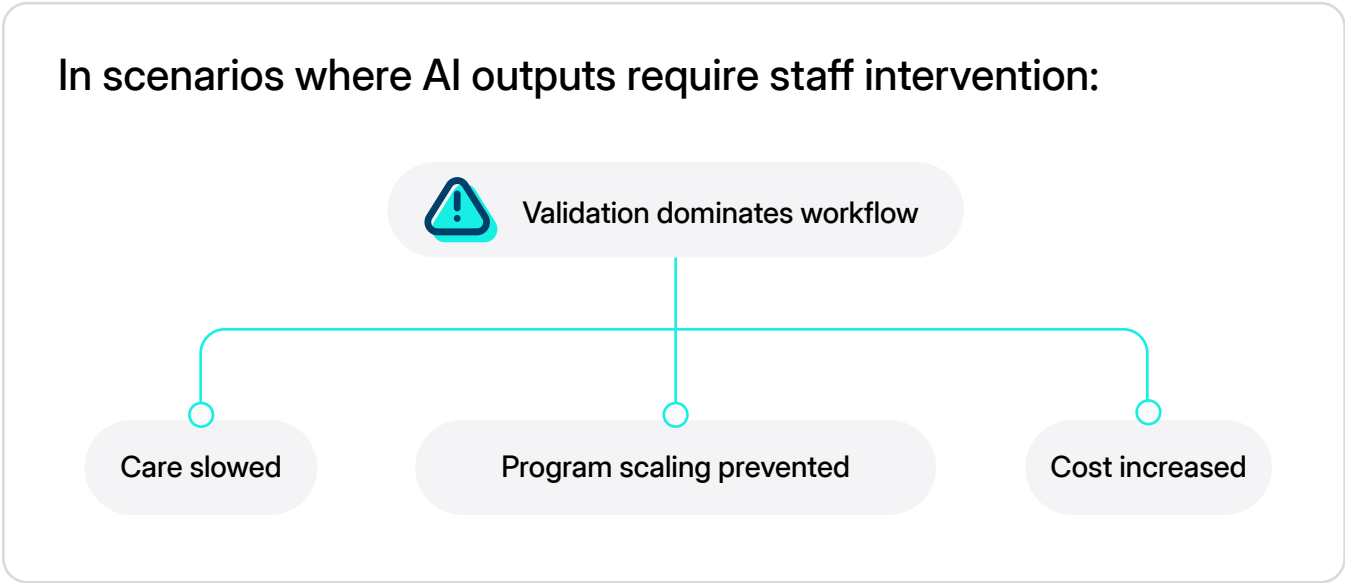
A heavy validation load doesn't just strain budgets—it slows care.

Backlogs delay downstream workflows, even when the AI is performing as intended. These delays affect care coordination, screening follow-up, quality reporting, and other critical clinical processes. Inconsistent human interpretation adds further variability and, in some cases, avoidable errors. These inefficiencies compound as AI usage expands into additional clinical domains.

Where the Validation Burden Becomes Especially Visible: Incidental Findings and Screening Programs

Although the validation burden applies across all clinical AI, it becomes particularly pronounced in programs that rely on detailed, structured clinical information—such as incidental findings management and screening programs like lung and breast.

These programs depend on AI to surface specific clinical characteristics that determine next steps, eligibility for follow-up, and the appropriate path of care. When AI outputs are incomplete, ambiguous, or inconsistently structured, staff must revisit the original clinical information to interpret the AI's output and gather the details required to act on it.



How AI Architecture Contributes to the Validation Burden

Different models generate outputs in different ways, which directly affects how much human validation they require.

Rule-based systems and many NLP approaches rely on pattern matching and may miss nuance or context, increasing the need for human oversight. Large Language Models (LLMs) produce flexible, narrative-style outputs that can vary from case to case and often require review to ensure accuracy and consistency. Computational Linguistics (CL) approaches generate structured, predictable outputs, which can reduce the amount of clarification and verification needed. These architectural differences shape the volume, predictability, and complexity of validation—and ultimately, the total operational effort required.

Validation burden is also heavily influenced by each model's precision and recall—how often it generates false positives or misses clinically relevant findings. Models that over-flag non-actionable items create large volumes of false positives that staff must review and dismiss. Models with lower recall require additional verification to ensure important findings are not overlooked. These error patterns differ across NLP systems, LLMs, and CL approaches, and they directly shape the human effort required to confirm that outputs are accurate and complete.

In incidental findings programs in particular, the burden is compounded by the need for AI to extract the clinical context required to determine whether follow-up is warranted and what the appropriate next step should be. These decisions depend on multiple attributes—such as the characteristics of the finding, relevant history, risk factors, comparisons to prior exams, and guideline-based thresholds. When AI extracts only partial information or interprets context inconsistently, staff must gather the missing details manually, adding substantial time and effort to the workflow.

How Workflow and Platform Design Add to the Validation Burden

The validation burden extends beyond the AI output itself. Placing a patient onto the correct care pathway and initiating appropriate follow-up requires clinical information that AI alone cannot provide—such as demographics, risk factors, smoking history, and other EHR-derived data. Staff must verify whether guideline-based follow-up is needed and determine the appropriate next step using clinical decision aids or evidence-based criteria.

These validation steps are part of the end-to-end workflow and contribute significantly to the total validation burden, regardless of the AI model used. How a platform or solution supports (or complicates) this process plays a major role in determining the true operational effort required.



Conclusion:

Understanding the Validation Burden

AI will play an increasingly central role in how care is delivered. But it can only reach its potential if its outputs can be trusted **without overwhelming clinical teams**. Validation is essential—yet the volume of validation required varies dramatically across AI systems and across the platforms that support them.

Ultimately, the true value of clinical AI isn't defined by its accuracy in isolation, but by how safely, efficiently, and reliably it can be put into practice.

The Validation Burden—the cumulative, often hidden cost of confirming that AI outputs are accurate, complete, and ready for clinical use—is one of the most overlooked aspects of AI adoption today.

Bringing this burden into full view is essential for building AI strategies that scale, protect clinical teams from operational overload, and deliver the level of trust that modern care demands.

As health systems expand their AI portfolios, the organizations that evaluate solutions through the lens of the Validation Burden—not just performance claims—will make more sustainable, defensible, and patient-centered investments.

About Eon

Eon is a healthcare technology company focused on supporting health systems in the identification and ongoing management of patients at risk of cancer and other life-threatening conditions. Powered by condition-specific clinical AI, Eon's longitudinal care management platform extracts incidental findings documented in radiology reports and helps ensure patients receive timely, guideline-based follow-up and remain in appropriate surveillance over time.

More than 70 health systems across over 1,200 facilities rely on Eon and its care management services to scale early detection programs, enable earlier diagnosis and treatment, and support sustained patient engagement—outcomes that also carry meaningful financial implications for health systems.