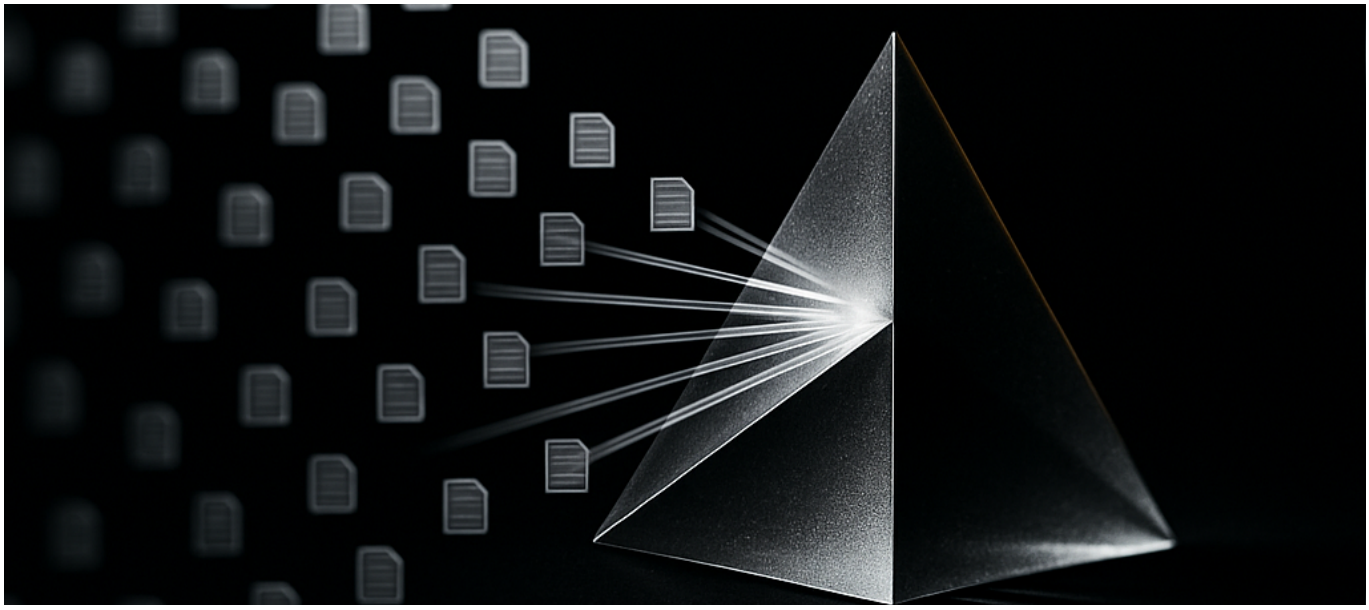


FDA Launches Elsa AI Tool Agency-Wide on February 5—Generative AI Now Reads Adverse Events, Compares Labels, and Generates Database Code for 18,000+ Employees

The FDA just deployed generative AI to 18,000 employees without buying a single SaaS license from an AI vendor. That sound you hear is a thousand enterprise sales teams rewriting their government pitch decks.

The News: FDA Goes All-In on Internal AI

On February 5, 2025, the U.S. Food and Drug Administration launched Elsa, an internally developed generative AI tool, to every employee across the agency. This isn't a pilot

program. This isn't a limited rollout to a single division. This is 18,000+ federal employees—from drug reviewers to administrative staff—with access to generative AI starting now.

Elsa handles three primary use cases: summarizing adverse event reports, comparing labels between drug products, and generating code for nonclinical databases. The first two directly touch drug safety and approval workflows. The third accelerates internal data infrastructure work that previously required specialized technical staff.

This marks the first federal agency deployment of enterprise generative AI at this scale. Other agencies have experimented with AI tools in isolated pockets. The FDA just made generative AI a standard tool for everyone, everywhere, all at once.

The timing matters. Just seven months after Sam Altman testified before Congress about AI regulation, a key regulatory body now uses the technology it may eventually oversee. The FDA didn't wait for perfect policy frameworks or industry consensus. They shipped.

Why This Matters: The Government Build-vs-Buy Decision Just Changed

Enterprise AI vendors have spent the past two years positioning themselves for the government market. Microsoft, Google, Amazon, and a parade of startups have built FedRAMP-compliant offerings, hired government sales teams, and lobbied for procurement budgets. The FDA's move suggests that strategy may be fundamentally flawed.

The FDA built Elsa internally rather than licensing an enterprise AI product. This decision carries enormous implications for the estimated $30 billion federal AI market that vendors have been salivating over.

Consider what the FDA avoided by building in-house: per-seat licensing fees that would scale to 18,000 users, vendor lock-in on a rapidly evolving technology, data sovereignty concerns about sensitive health information flowing through third-party systems, and procurement timelines that stretch into years. Government software procurement typically moves at geological speeds. The FDA compressed that timeline dramatically by keeping development internal.

The winners here are obvious: agencies with technical talent willing to follow the FDA's lead, and infrastructure providers (cloud, compute) who still get the underlying spend. The losers are enterprise AI SaaS vendors who assumed government would be their largest customer segment.

> The FDA just demonstrated that building AI tools is faster than buying them through government procurement. Every agency CIO noticed.

There's a second-order effect worth watching. The FDA reviews drugs, devices, and food safety for the entire U.S. market. Any efficiency gains in their review process directly affect how quickly new treatments reach patients. If Elsa shaves even 10% off adverse event review time, that compounds across thousands of drug safety reviews annually. The downstream impact on pharmaceutical timelines could be substantial.

Technical Deep Dive: What Elsa Actually Does

The public details on Elsa's architecture remain limited, but we can infer significant technical decisions from its stated capabilities.

Adverse Event Summarization

Adverse event reports—the forms submitted when someone experiences a negative reaction to a drug—are notoriously dense. A single serious adverse event report can run 20+ pages, mixing medical terminology, patient narratives, and clinical assessments. The FDA receives over 2 million of these reports annually through the FAERS (FDA Adverse Event Reporting System) database.

Summarizing these reports requires more than basic text compression. The model must identify clinically relevant details, distinguish between primary and secondary events, flag potential drug interactions, and maintain medical accuracy. A hallucination in this context isn't just embarrassing—it could affect drug safety decisions.

This suggests the FDA either fine-tuned a foundation model on medical text or implemented robust retrieval-augmented generation (RAG) with their existing adverse

event databases. Given the sensitivity, they likely added multiple verification layers before any AI-generated summary enters official review workflows.

Label Comparison

Drug labeling comparison sounds simple until you understand the complexity involved. Drug labels contain dosing information, contraindications, warnings, clinical trial data summaries, and legal language that varies between generic and brand-name products. When the FDA needs to compare labels—during generic drug approvals, label updates, or safety reviews—analysts currently read both documents manually and document differences.

An effective AI solution for label comparison needs to: parse structured and semi-structured document formats, identify semantically equivalent statements even when phrased differently, flag meaningful differences versus cosmetic changes, and track changes across label versions over time. This is a sophisticated document understanding problem that goes well beyond simple diff tools.

Database Code Generation

The code generation capability for nonclinical databases represents the most technically interesting use case. “Nonclinical databases” at the FDA typically means the massive data repositories holding animal study results, laboratory data, and pre-human trial information that supports drug applications.

These databases use a mix of legacy systems, proprietary formats, and modern data warehouses. Generating code to query, transform, or analyze this data requires understanding both the data schemas and the scientific context of what analysts need.

The FDA essentially built an internal code assistant trained (or prompted) with knowledge of their specific data infrastructure. This is precisely the kind of specialized application that general-purpose AI coding assistants like GitHub Copilot handle poorly. Copilot doesn’t know the FDA’s internal table structures or data conventions.

The technical insight here: domain-specific AI tools outperform general-purpose alternatives when you have proprietary data structures and specialized workflows.

The FDA apparently concluded that building a custom solution would outperform any vendor offering.

The Contrarian Take: What Everyone Gets Wrong About This

Most coverage of the FDA announcement focuses on the "AI in government" angle, celebrating modernization or warning about automation risks. Both framings miss the actual significance.

This Isn't About AI—It's About Technical Sovereignty

The FDA's decision to build rather than buy reflects a broader shift in how sophisticated organizations think about core technology capabilities. When AI touches your most sensitive workflows—in the FDA's case, drug safety and approval decisions—outsourcing that capability to a vendor creates unacceptable risks.

What happens when your AI vendor's model gets updated and suddenly performs differently? What happens when they raise prices, knowing you're locked in? What happens when they get acquired and your data access terms change? What happens when their system has an outage during a critical safety review?

The FDA answered these questions by taking control of the entire stack. This is the same logic driving major banks to build internal AI capabilities rather than relying solely on vendors. It's the same reasoning behind defense agencies insisting on on-premise deployments.

> The organizations that will define AI adoption aren't buying AI products—they're building AI capabilities.

The 18,000 Number Is Both Impressive and Misleading

Headlines trumpet "18,000+ employees" with AI access, and that number is real. But access

doesn't mean active use. Every enterprise software deployment knows the difference between licenses deployed and features actually used.

The more interesting number will emerge in 6-12 months: how many employees use Elsa weekly? Which use cases drive the most adoption? Where do employees ignore the AI and stick with manual processes? The FDA's deployment is a starting gun, not a finish line.

The agency will inevitably face the same adoption challenges every enterprise AI deployment encounters: prompt engineering is a skill most employees lack, trust in AI outputs varies wildly across personality types, and workflow integration matters more than tool quality. An AI tool that requires context-switching away from existing systems will see lower adoption than one embedded in daily workflows.

The Underrated Story: FDA Has Technical Talent

Building and deploying an agency-wide AI tool requires significant technical capability. Product managers, ML engineers, infrastructure specialists, security experts, and deployment teams all had to execute on aggressive timelines.

The FDA has historically struggled to compete with private sector salaries for top technical talent. That they pulled this off suggests one of two things: either they've improved their technical recruiting significantly, or they partnered with contractors who delivered effectively (a rare outcome in government IT).

Either way, the FDA's technical organization deserves more attention than it typically receives. They just shipped a complex AI product to 18,000 users. Most startups can't claim that.

Practical Implications: What This Means For Your Organization

If you're a CTO, senior engineer, or technical founder, the FDA's deployment offers several actionable insights.

Reconsider Build vs. Buy for Core AI Workflows

The conventional wisdom says "buy AI tools, don't build them"—the models are too expensive to train, the infrastructure too complex, the talent too scarce. The FDA's deployment challenges this assumption for specific use cases.

If your AI application involves: proprietary data that vendors can't access, specialized domain knowledge beyond general models, workflows touching regulatory or compliance functions, or cost structures where per-seat licensing doesn't scale—then building internally deserves serious evaluation.

The calculation has changed. Foundation models are available through APIs. RAG architectures provide customization without fine-tuning costs. Infrastructure providers offer managed services that reduce operational burden. Building a specialized AI application on top of existing foundation models is dramatically easier than building one from scratch.

Document Understanding is the Killer App

Two of Elsa's three use cases involve document understanding: summarizing adverse event reports and comparing labels. This aligns with emerging patterns across enterprise AI adoption.

Most enterprises are drowning in documents: contracts, reports, correspondence, technical documentation, regulatory filings. The humans who understand these documents are expensive and time-constrained. AI that can accurately parse, summarize, and compare documents provides immediate ROI.

If you're prioritizing AI initiatives, document understanding workflows deserve top placement. The technology has matured enough for production use, the value proposition is clear, and the evaluation criteria (accuracy, completeness) are measurable.

Start With Summarization, Not Generation

Note what Elsa doesn't do: it doesn't write new regulatory submissions, generate clinical trial protocols, or draft policy guidance. The FDA wisely constrained Elsa to tasks where AI

supports human decision-making rather than replacing it.

Summarization is the safest entry point for AI in high-stakes domains. The human reviewer still makes the decision; they just receive pre-processed information. If the AI makes an error, the human catches it. If the AI misses something, the original documents remain available.

Organizations deploying AI in regulated or high-stakes environments should follow this pattern. Start with summarization and analysis tasks that augment human judgment. Graduate to more autonomous capabilities only after building trust through demonstrated reliability.

Specific Technical Recommendations

For teams considering similar deployments:

- **Invest in evaluation infrastructure** before deployment. The FDA surely has robust testing for Elsa's outputs against known-good summaries and comparisons. You need the same.
- **Build retrieval layers specific to your domain**. General-purpose RAG won't understand your data structures, terminology, or implicit relationships. Custom retrieval significantly improves output quality.
- **Plan for model updates**. The foundation model underneath your application will change. Abstract your integration layers to accommodate model swaps without full rewrites.
- **Implement usage analytics from day one**. You cannot improve what you don't measure. Track which features get used, which queries fail, and where users abandon the tool.
- **Design for progressive disclosure**. Not every user needs every capability. Start simple, expand access to advanced features based on demonstrated competence.

The Vendor Landscape: Who Wins and Loses

The FDA's build decision reshapes competitive dynamics in enterprise AI.

Winners

Cloud infrastructure providers (AWS, Azure, GCP) still capture the underlying compute and storage spend. Whether an agency builds or buys AI applications, they need infrastructure. The FDA's deployment likely runs on one of these platforms, generating substantial revenue without the sales complexity of enterprise SaaS.

Foundation model API providers get a different kind of win. If the FDA built Elsa on top of a commercial foundation model accessed via API (rather than training their own), that provider captures usage-based revenue without the enterprise sales cycle. This is the picks-and-shovels position in the AI gold rush.

Systems integrators and government contractors with AI expertise become more valuable if the build approach spreads. Agencies will need help designing, building, and maintaining custom AI applications. The firms that can deliver will command premium rates.

Losers

Enterprise AI SaaS vendors targeting government face an existential challenge. Their pitch assumes agencies will prefer buying to building. If the FDA's success encourages other agencies to build internally, the addressable market shrinks dramatically.

AI startups planning government pivots need to rethink their strategy. The government market is slower and more complex than commercial sales. If that market also increasingly prefers internal builds, the expected revenue may never materialize.

Legacy government IT contractors without AI capabilities face displacement. If agencies prioritize AI deployment and decide to build, they'll work with partners who can deliver AI systems—not partners whose expertise is in maintaining decades-old infrastructure.

The Forward Look: Where This Goes in 6-12 Months

The FDA's deployment sets up several predictable next moves.

Other Agencies Will Follow

Government agencies benchmark against each other. The FDA's successful deployment gives cover to CIOs at other agencies who want to pursue similar initiatives. Expect announcements from HHS (FDA's parent department), CDC, and potentially DOD components within the next year.

The White House Office of Management and Budget will likely issue guidance encouraging (or requiring) agencies to evaluate AI tools for operational efficiency. The FDA's deployment provides the template.

FDA Will Expand Use Cases

Elsa's current capabilities are a starting point. The obvious next expansion areas include: automated initial screening of drug applications for completeness, faster identification of safety signals across adverse event databases, support for foreign language submissions (a growing portion of FDA workload), and assistance in drafting inspection reports and compliance letters.

Each expansion moves AI closer to core regulatory decisions. The FDA will proceed cautiously, but the direction of travel is clear.

The Vendor Response

AI vendors won't concede the government market without a fight. Expect several responses: aggressive pricing to make buy decisions more attractive than build, partnerships with government contractors who have existing agency relationships, open-source strategies that provide the technology while selling services around it, and intense lobbying for procurement policies that favor vendor solutions.

Microsoft has the strongest position here, given existing government cloud relationships. They can bundle AI capabilities with infrastructure agencies already use, reducing the case for standalone builds.

The Regulatory Implications

Here's the genuinely novel situation: the FDA now uses AI internally while also being the agency that oversees AI-powered medical devices and AI in drug development. They've moved from AI observer to AI operator.

This experience will inform their regulatory posture. Regulators who use technology daily develop more nuanced views than those who understand it only theoretically. The FDA's internal AI deployment may accelerate their ability to evaluate AI-based medical products—or it may create conflicts of interest that critics will highlight.

The Workforce Conversation

18,000 FDA employees now have access to AI that can perform portions of their jobs. The agency will face questions about staffing levels, skill requirements, and workforce development. Does AI mean fewer employees? Different employees? Employees doing different work?

These questions have no easy answers, but the FDA's deployment makes them concrete rather than theoretical. Union negotiations, hiring plans, and training budgets will all need to account for AI capabilities that didn't exist a year ago.

The Bigger Picture

The FDA's Elsa deployment represents something more significant than a single agency adopting a new tool. It marks the moment when AI stopped being something government studies and started being something government uses.

This transition happened faster than most predictions suggested. Two years ago, generative AI was a research curiosity. One year ago, government agencies were writing policy papers about AI risks. Today, the FDA has deployed AI to 18,000 employees for core mission functions.

The pace of change is accelerating, not slowing. Organizations that assumed they had years to develop AI strategies may have months. The gap between AI adopters and AI observers

will widen faster than anyone expected.

The FDA made a choice: they decided that waiting for perfect AI tools was more dangerous than deploying imperfect ones carefully. They decided that building internally gave them more control than buying externally. They decided that agency-wide access was better than restricted pilots.

These decisions will be validated or challenged by what happens next. But the FDA's willingness to make them—to ship AI to 18,000 employees and accept responsibility for the outcome—sets a standard that other organizations will be measured against.

The age of AI pilots is over. The FDA just showed what operational AI deployment looks like at scale, and everyone else is now playing catch-up.



WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting