



GUIDE



Build vs. Buy

The Hidden Costs of a DIY Data Security Solution

For organizations serious about AI, data classification isn't optional infrastructure - it's the foundation everything else is built on. Here's what to consider before you decide.



Every AI initiative your organization launches depends on one thing: knowing what data you have, where it lives, who can access it, why it matters to the business, and the risk it creates. Without that foundation, AI systems operate on blind trust. Data classification is no longer a back-office compliance task — it's a strategic prerequisite. The question isn't whether to invest in it. The question is whether to build it yourself or partner with a purpose-built platform.

01

Time to Value

Can your AI roadmap afford a 2–4 year wait?

Building a classification engine internally is an engineering milestone measured in years, not quarters. For organizations that have committed to AI as a strategic priority, that timeline isn't just slow — it's a compounding disadvantage.

2–4 years

to meaningful coverage when building internally

Days

to operational visibility with a purpose-built platform

Build Internally	Buy (Cyera)
AI initiatives stalled pending classification readiness	Operational in days to weeks
Multi-year gap before trusted data foundation exists	AI pipelines grounded in trusted classification from day one
Competitors with purpose-built platforms gain years of advantage	Accelerate your AI roadmap without rebuilding the foundation
10–15 specialist engineers required to sustain it	Proven scalability and continuous coverage updates
Relies on static rules, fixed labels, or generic LLMs — limited to predefined or generic data types and prone to false positives	Uses adaptive, AI-native models that identify proprietary, business-specific, and previously unknown data without manual rule-writing

If data classification is required to unlock the full value of AI, then the build timeline is not just an engineering consideration — it is a business risk.



Your best engineers deserve better problems

The question isn't whether your team can build a classification engine — they almost certainly can. The real question is whether they should, and what they won't be building while they do. Classification infrastructure requires continuous maintenance as environments evolve, new sources emerge, and regulations shift. It absorbs the capacity of the people you need building agents, applications, and AI-driven products.

A realistic internal classification build requires 10–15 engineers across data engineering, ML, cloud infrastructure, security, and backend development — absorbing the same team you need building agents, workflows, and AI-driven products.

Build Internally	Buy (Cyera)
Engineering hours: 10–15 specialists across data engineering, ML, cloud infrastructure, security, and backend development to build and sustain it	Engineering hours redirected to agents, workflows, and AI-driven products that define your competitive edge
Maintenance burden: continuous re-tuning as environments evolve, new data sources emerge, and regulations shift	Maintenance burden absorbed by the platform — continuous coverage updates without pulling engineers off product work
Opportunity cost: every hour spent on classification infrastructure is an hour not spent on differentiated product work	Opportunity cost recovered — engineering capacity stays focused on what only your team can build

The most valuable thing Cyera delivers may not be the platform itself — it's the engineering hours it gives back. Let your team focus on what they want to build: agents, workflows, and AI-driven capabilities that define your competitive edge.

Building internally means owning every gap and failure

When you build and operate your own classification capability, you take full ownership of every outcome it produces — including the failures. This liability compounds across four dimensions.



Regulatory Compliance

GDPR, HIPAA, CCPA, and emerging AI regulations require demonstrable, continuously updated coverage. Internal builds that lag behind shifting requirements create direct enforcement exposure.



Classification Accuracy

High false-negative rates mean sensitive data goes undetected — over-shared, under-protected, or fed into AI systems without controls. Homegrown approaches built on pattern-matching or generic LLMs are especially prone to missing proprietary, unstructured, and business-specific data — exactly the data regulators and adversaries care most about. Your organization bears full accountability for those gaps.



Speed of Change

New data sources and AI pipelines are introduced continuously. Internal teams that can't keep pace create gaps — and regulators don't make allowances for development backlogs.



Key Person Dependency

Internally built systems concentrate knowledge in a small number of engineers. Attrition or competing priorities can degrade the capability entirely, with no external escalation path.