Beyond more data: Eval-driven intelligence as the next enterprise AI frontier

June 28, 2026

Enterprise AI programs need to move beyond data-heavy training and toward workflow-based evals: measurable, governed systems that learn how work actually gets done.

Many enterprise AI leaders could tell you how much data they have prepared for model training. Fewer could clearly say whether the model understands the decisions, exceptions, controls and escalation logic that shape work inside the business.

This is becoming a defining gap in enterprise AI. Many organizations still approach model training with a familiar assumption—collect more data, clean more data, label more data, and then train on historical patterns. If the data is already well-structured, consistent, and complete, this can work. But in most enterprises, data is fragmented, inconsistent, and deeply tied to evolving processes—making the pursuit of "perfect data" a never-ending effort.

A claims decision, procurement exception, customer service resolution or compliance review is not just a data point. It is the result of policy interpretation, risk judgment, missing information, business context and human expertise. Historical records may show what decision was made, but not always why it was made.

This is why enterprise AI needs a new operating model. The next phase will not be won by feeding more data into larger models. It will be won by translating workflow context into measurable learning systems.

Cognizant AI Data Services: Workflow, evals and continuous improvement

Think of enterprise AI training in three connected phases to see how the model changes:

Workflow grounding: AI training needs to begin with the work itself. That means mapping the task, the users involved, the systems accessed, the documents reviewed, the decisions made, the exceptions encountered and the points where human judgment is required.

The objective is not to create a generic process map. It is to expose the operating logic that makes the workflow perform well. What information matters? What policy applies? What signals indicate risk? When should the AI act, ask for clarification or escalate to a human?

Eval-driven design: Once the workflow is understood, the next step is to define what "good" looks like. This is where evaluations become the control plane for AI performance. Evals are not generic accuracy tests. They are business-specific measures of whether the AI system is doing the right work in the right way. They can test whether a recommendation follows policy, whether a high-risk case is escalated, whether the explanation cites the right evidence, whether the response is complete and whether the model knows when not to act.

This matters because enterprise AI does not fail only when it is wrong. It fails when it is wrong in ways that matter: approving what should be escalated, producing an unsupported recommendation, exposing sensitive information, creating rework or eroding trust.

Scenario-led optimization: With the workflow and evals in place, training data becomes more targeted. Instead of depending only on large volumes of raw customer data, teams can create representative scenarios that reflect the real operating environment.

These scenarios can include standard cases, edge cases, incomplete cases, conflicting information, policy exceptions, adversarial prompts, escalation moments and known failure modes. They can be built from anonymized examples, expert judgment, synthetic data or controlled samples.

The model can then be trained or optimized inside a simulated task environment. It can read documents, retrieve information, apply rules, call tools, produce structured outputs and receive feedback against the evals. The result is not just a model that predicts a historical pattern. It is a model that practices the work safely before being exposed to live operations.

Why this changes the economics of enterprise AI

An eval-driven approach changes the adoption curve because it reduces several sources of friction that slow enterprise AI programs.

First, it reduces raw-data dependency. Large historical data dumps are no longer the only route to improvement. Organizations can begin with workflow definitions, expert rubrics, curated examples and controlled scenarios.

Second, it lowers sensitive-data exposure. Where appropriate, scenarios can use masked, anonymized, synthetic or policy-derived examples. This does not remove the need for governance, but it can reduce the need to move large volumes of sensitive customer information into training pipelines.

Third, it helps organizations make progress before every data-quality issue is solved. Most enterprises do not have perfectly clean, fully labeled and fully governed data estates. AI Training Data Services allows useful progress to begin while broader data maturity continues.

Fourth, it improves edge-case readiness. Traditional datasets often underrepresent rare events

But in enterprise workflows, rare events can carry the highest regulatory, financial or reputational risk. Scenario design allows teams to deliberately test these cases before production.

Finally, it creates a faster learning loop. Every failure can become a new eval. Every exception can become a new scenario. Every production insight can become a new training signal.

Consider claims decision support

A claims team may want AI to help adjusters review documents, identify missing evidence, assess policy fit and recommend next steps.

The traditional approach would start by collecting historical claims, cleaning records, labeling outcomes and training a model to predict approval, rejection or escalation. But those records may not explain why decisions were made. Some decisions may reflect outdated policy. Some may depend on undocumented conversations. Some edge cases may be too rare to influence the model. Some records may contain sensitive information that should not be used in raw form.

Evals-based approach starts differently. The team maps the claims workflow, defines the documents reviewed, identifies the facts required, clarifies the policies that apply and agrees where human escalation is mandatory.

The team then defines evals: high-risk claims must escalate, missing documentation must be flagged, policy-violating approvals must be avoided, recommendations must cite supporting evidence and uncertain cases must be routed to a human.

Next, representative scenarios are created. These include simple claims, incomplete claims, suspicious claims, ambiguous claims and claims with conflicting documents. The model is tested in a simulated claims environment, reviewed by experts and first deployed in shadow mode before it influences live decisions.

The outcome is not simply a model that predicts claims outcomes. It is an AI system that understands the workflow, respects risk boundaries and improves with the business.

How to move toward eval-driven intelligence

To begin the shift, enterprise leaders should take the following steps:

Start with workflows, not datasets. Identify the high-value workflows where AI can improve speed, quality, control or customer experience. Map the real decision points, exceptions and risks before deciding what data is needed.

Build the eval suite early. Define what good looks like before deployment. Include accuracy, completeness, compliance, escalation behavior, explanation quality, evidence use, privacy controls and human handoff rules.

Use scenarios to train for reality. Create representative and edge-case scenarios that reflect how work actually happens. Use anonymized, synthetic or expert-authored cases where raw data is sensitive, fragmented or incomplete.

Next steps in enterprise AI training

Enterprise AI is moving from access to capability. Foundation models are becoming widely available. The differentiator is whether organizations can make those models work inside their own workflows, decisions and controls.

Traditional training begins with data and hopes the model learns the work. AI Training Data Services begins with the work and teaches the model how to perform it.

Organizations that operationalize this shift will build AI systems that are safer, more measurable and more business ready. They will reduce raw-data dependency, improve governance and create continuous learning loops that convert every exception into improvement.

The next step is evaluation-driven enterprise AI—because in complex workflows, intelligence is not just in the data, but in understanding how good work actually gets done.

Cognizant brings this to life through proven expertise in AI Model Training & Data Service, underpinned by an evals-driven framework and a suite of proprietary platforms and accelerators. This enables rapid, scenario-based training and evaluation, reducing risk of failure while significantly improving performance and productivity for model training.

If you're looking to move from experimentation to production-grade AI, this is where to start.

Author

Dr. Venkatesh Upadrista
Global Head of AI Offerings—IOA
Drvenkatesh.upadrista@cognizant.com

Follow

Dr Venkatesh Upadrista leads AI offerings for Cognizant's Intuitive Operations and Automation (IOA) service line, driving end-to-end growth—from sales and solutioning to delivery and client engagement for AI-led transformations.

With over two decades of leadership experience, he is an active AI researcher and innovator, having authored multiple books and published several research papers in AI and blockchain. He also leads advanced research in Artificial general intelligence at a leading academic institution in the UK, focusing on next-generation areas such as world models, meta-learning, and causal reasoning—advancing the development of human-like intelligence systems for real-world enterprise impact.