POINT OF VIEW

Pega Customer Decision Hub (CDH) vs. AI-Native Custom Build

When to build and when to buy in the age of agentic AI

Authors:

David Steuer, MD
Global Pega Practice Lead
Accenture

Rob Walker, GM
1:1 Customer Engagement
Pega

Executive Summary

The buy-vs-build question for AI-driven decisioning is not about capability alone; it is about the hidden costs of compliance, scale, and governed change.

Some discussions of "AI-native decisioning" still center on using a large language model as the decisioning engine itself, generating customer recommendations in real time at the moment of interaction. This is not a close call at enterprise decision volumes: LLM-orchestrated execution carries latency and non-determinism. This is incompatible with real-time performance requirements and regulatory audit obligations-while also running at a materially higher cost than deterministic decisioning.

That said, this topic keeps resurfacing in executive conversations. Some early programs (including at large enterprises) initially explored real-time LLM Next-Best-Action (NBA)inferencing before quickly reversing course for the practical reasons above. The core issue is simple: language models will always provide an answer—whether there is sufficient evidence to justify that answer or not. Conversely, purpose-built enterprise decisioning requires statistical rigor and traceable evidence in the moment.

The question that does warrant rigorous analysis, and where organizations are actively considering consequential commitments,-is whether modern AI development tooling, agentic frameworks, and LLM-assisted engineering can compress the build timeline for a custom decisioning platform to the point where it becomes a credible long-term alternative to CDH. That is where the genuine debate sits, and it is what this document examines.

A custom AI-native build can be "good enough" in some cases where mature models, strong in-house engineering, lower compliance needs, and modest personalization ambition are sufficient. But the central point is not that custom can "match" CDH; it is that reproducing CDH's enterprise-grade capabilities becomes difficult and cost-prohibitive as organizations scale across models, channels, customer personas, volumes, and regulatory jurisdictions.

This document outlines considerations for the buy vs. build decision in the agentic era—where design-time acceleration is real, but runtime guarantees, governed change, and compliance outcomes remain the decisive factors.

Contents

1. The debate reframed

Recent developments by Anthropic, Cursor, and OpenAI have redefined the boundaries for the buy vs. build decision. Companies using or planning to use NBA solutions are rethinking their approach. A large global tech company with a sophisticated engineering and modeling team built a "custom NBA" (CNBA) implementation—using Databricks, Kafka, Redis, and a boutique "near-real-time" (NRT) sentiment model. To date, this is delivering measurable retention results at scale with over 1,200 rules, 1,000 actions across 30K agents.

The emergence of AI coding agents (Claude, Copilot, Cursor) has genuinely compressed design-time complexity. GUIs get built faster. Rules get iterated more quickly. What once required months of bespoke development now takes sprints.

But the question is not whether a custom build can work. It is whether the total cost of ownership—across compliance, scale, governance, and agentic evolution—makes it the right choice for your organization.

In practice, the build phase rarely surfaces the core challenge. It emerges at the transition from prototype to production, when governance, regulatory compliance, operational monitoring, and enterprise-scale change management become non-negotiable requirements. The sections that follow examine where the substantive gaps lie and what they cost to close.

A final reframing is essential in the agentic era: Even if LLMs can write code faster, that does not change the nature of the Build-Your-Own approach. CDH (and other platforms) have always competed against manual coding. Faster code generation does not eliminate the same hard problems that appear at scale-governed runtime execution, auditability, regulatory explainability, and enterprise change control.

1 The Institute of Internal Auditors (IIA): Artificial Intelligence Auditing Framework (September 13, 2024): https://www.theiia.org/en/content/tools/professional/2023/the-iias-updated-ai-auditing-framework/

2. Five Foundational Capabilities

We use the term “foundational” deliberately. At a small scale, surface-level capability can support a prototype or narrowly scoped production use cases. At enterprise scale, sustained load requires a foundation: governed decisioning, repeatable compliance outcomes, and reliable runtime guarantees. The five capabilities examined in this section are foundational to enterprise customer decisioning. These are not features to be bolted on after the system is running, but the load-bearing layer that determines whether it can run safely at scale.

Each one is formed through years of accumulated engineering investment, operational iteration, and regulatory certification; and each one carries the defining quality of a true foundation: invisible when things are working, and non-negotiable when they are not. Pega CDH has reached production maturity in all five dimensions. The question for any organization considering a custom build is whether they are planning to reproduce them independently and what that requires.

1 Regulatory Explainability

2 Real-Time Decisioning

3 Always-On Learning at Scale

4 Operational Decisioning Intelligence

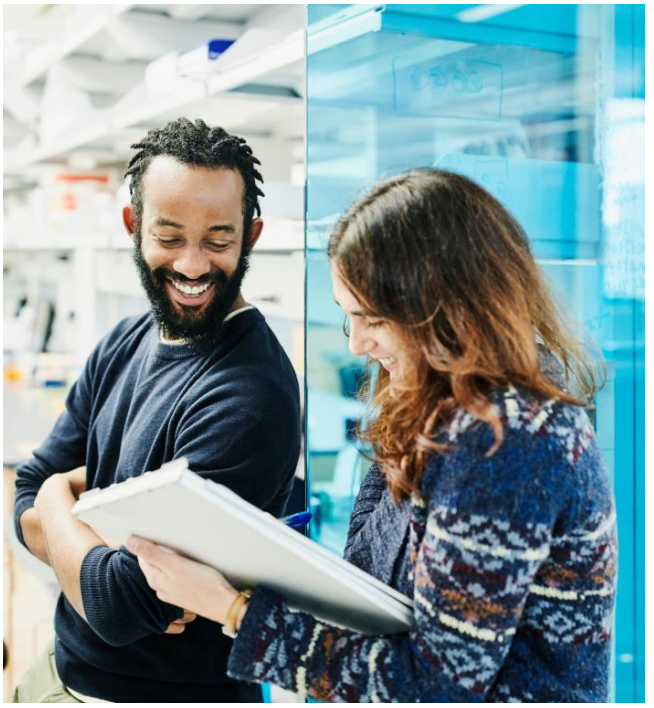
5 Change Governance

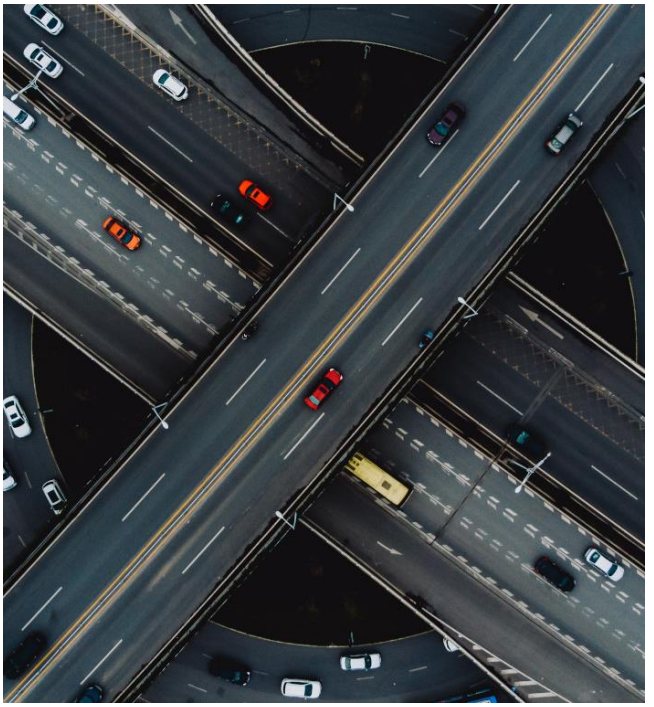
Capability I – Regulatory Explainability

SHAP (Shapley Additive Explanations) values are a valuable data science tool, but regulatory and audit expectations often extend beyond model-level explanations. Under GDPR Article 22, individuals have the right not to be subject to certain decisions based solely on automated processing that produce legal or similarly significant effects, and where exceptions apply, controllers must provide safeguards including the right to obtain human intervention, to express a point of view, and to contest the decision. In parallel, organizations operating under regimes such as HIPAA and the CCPA must protect sensitive data and honor defined individual rights and notices regarding how personal information is used and disclosed.

CDH provides a time-stamped audit trail of which strategy was active, which customer context was evaluated, which eligibility rules were applied, why a specific action was selected over alternatives, and which governance checkpoint approved that strategy's deployment.

Building this governance layer on top of a custom stack is not easily implemented even with advanced AI tools. It is a multi-year compliance certification effort that no custom build inherits for free. No custom build starts with this on day one. CDH already carries a decade of regulatory certification across multiple jurisdictions.





Capability II – Real-Time Decisioning (and what "real-time" actually means)

The custom CNBA stack in this example achieved near-real-time re-decisioning using a clever combination of Kafka event streaming, a custom-built near-real time sentiment model operating over a 2-minute call transcript window, and Redis caching. It works. At least for a narrow use case (retention in an assisted channel), performance benchmarks show it is technically comparable on a like-for-like comparison.

But the question is to what degree can a custom solution deliver real time personalization at scale:

- The custom implementation caches a re-decisioned result triggered by an async event. When the front end requests the re-decisioned action, it retrieves from cache. That is fast cache retrieval, not live arbitration.
- Pega CDH arbitrates simultaneously across all eligible actions and a multiple of treatments in a single <100ms pass, meaning it simultaneously applies propensity models (in real-time), business constraints, ethical guardrails, and channel suitability rules. No pre-computation. No cache dependency.
- At 1,000 actions, the custom approach is feasible. At 10,000 actions across 65M customers — profile of a large global bank—the delta between cached re-decisioning and live arbitration becomes operationally significant.

At that scale, cached re-decisioning is not a performance trade-off—it is an architectural ceiling. A custom BYOB stack must pre-compute, pre-cache, and pre-narrow the decision space before the customer arrives. That is the opposite of arbitration. It means the system has already decided what it can consider before it knows what the customer needs.

Custom builds can achieve impressive near-real-time performance at defined scale. The capability of Pega's CDH is the breadth of simultaneous arbitration inside that window at scale, not speed alone for a limited set of actions.

The commercial implication becomes visible at scale. When arbitrating across 100's of simultaneous actions against live customer context at enterprise volume, the precision of a genuinely real-time arbitrated decision compounds over time in ways that pre-computed recommendations cannot replicate. Organizations that have closed this gap consistently report that adaptive models refine their signals with each interaction, creating a performance advantage that widens as volumes grow.

Capability III – Always-On Learning at Scale

Naive Bayes and Gradient Boosting are open source. The critical component is not the algorithm. It is the automated orchestration of the full adaptive learning lifecycle—across every channel, in real time, at enterprise scale—without manual retraining cycles.

CDH closes the loop automatically: interaction → signal → propensity update → next interaction across any channel, continuously. At a bank with 65M customers and 5B interactions per month, that operational maturity is a multi-year build, not a few sprints.

The adaptive modeling framework also drives cross-channel orchestration consistency. Because CDH maintains a single decisioning brain across all inbound and outbound channels, each customer interaction informs the next one, regardless of where it occurs. That contextual continuity is not a design principle but rather an architectural outcome. When a customer engages via mobile, that signal updates the model estate. When they arrive on the web or speak to the contact center shortly after, CDH's decision reflects what was learned. Channel-specific decisioning systems cannot replicate this without a centralized arbitration authority. CDH is that decision authority.

At a smaller scale (retention-focused, single channel, defined signal set), the custom adaptive loop is achievable. Business results supported that on a custom solution. The gap is not a maturity failure. It is a different category of problem instead.

The consequence is cumulative. Each additional channel, product category, or market segment multiplies the operational complexity of the adaptive learning estate. CDH absorbs that complexity automatically. A custom build requires it to be re-engineered at each expansion, at a cost that compounds in proportion to the ambition of the program.

Capability IV – Operational Decisioning Intelligence

The three capabilities above address what CDH decides and why those decisions are trustworthy. There is a fourth capability that custom builds most consistently lack: continuous visibility into whether those decisions are actually working. Action Analysis and Decision Health provide operational visibility into the decision funnel and the distribution of actions available for each decision, helping teams identify when decisions have too few (or no) actions to compete effectively and where filtering occurs.

A custom-built decisioning stack generates telemetry. CDH generates decisioning intelligence. Action Analysis provides per-stage funnel visibility to identify precisely where actions are being suppressed. ADM Health Check automatically flags broken feedback loops where model performance has degraded to chance level. Explainability Extract provides a time-stamped, human-readable record of why each action was selected, satisfying both internal review and regulatory audit. Building equivalent diagnostic capability independently is a multi-year engineering commitment. CDH delivers it as a baseline.

Capability V – Change Governance

MLOps and CI/CD pipelines manage the technical lifecycle of a model. Changing a credit card upsell strategy in a regulated institution is not purely a technical event. It requires compliance review, legal sign-off, model validation, bias testing, and staged rollout—a multi-stakeholder governance workflow that MLOps does not deliver out of the box.

CDH's deliberate separation of design-time and runtime is by design. GenAI can assist at design time (Blueprint agents generating strategy hypotheses in natural language). Statistical AI governs runtime execution where precision, consistency, and auditability are non-negotiable. This is exactly what a Chief Risk Officer wants to hear.

A custom build might use a Claude-built UI, demonstrating that AI accelerates design-time iteration. The question is: when that strategy goes live, does the change governance infrastructure satisfy your risk committee?

Organizations that have built change governance capability independently find it becomes a permanent engineering program rather than a project. Each regulatory update, model change, or strategy revision triggers the full multi-stakeholder cycle. CDH converts that ongoing commitment into a configurable workflow: what takes months to coordinate through a custom change process takes days through CDH's design-time governance tools.

3. The Agentic Dimension

The emergence of agentic AI reshapes this debate in a specific way. The comparison is no longer GUI vs. GUI. It is:

- Custom build: AI agents that help developers iterate on code and rules faster at design time.
- Pega CDH: AI agents with full access to a purpose-built decisioning control plane, (including Scenario Planner, Value Finder, Action Performance Tracker, NBA Designer, and Ops Manager as agentic tools).

CDH's agentic control plane means that the governed infrastructure is itself agentically accessible. Any organization's agents, including externally built ones, can connect to and orchestrate CDH. This is not a walled garden; it is a governed foundation that agentic workflows can operate on top of.

Agentic AI has genuinely accelerated the design-time phase of a decisioning build. Faster prototypes, generated rule logic, and scaffolded interfaces are real advantages that both paths benefit from. The asymmetry is at runtime. Governed real-time arbitration across live customer context, engagement policy management, regulatory explainability at the moment of decision, and adaptive model orchestration at enterprise scale are operational architecture problems. They do not get easier as the system grows, and faster code generation does not make them smaller. CDH's agentic layer addresses both dimensions: GenAI tools that accelerate strategy development at design time, and a proven governed runtime that any agent can orchestrate rather than rebuild from scratch.

Gartner predicts that over 40% of agentic AI projects will be canceled by the end of 2027, citing escalating costs, unclear business value, or inadequate risk controls. Related research on enterprise GenAI adoption also reports that only a small minority of integrated pilots reach production, reinforcing the importance of governed runtime infrastructure, monitoring, and accountability when moving beyond demos.

Two agentic-era shifts: content advantage has leveled, and agentic operations reduce the historic BAU complexity tax, making day-to-day change faster without removing controls.

4. A Framework for the Decision

Rather than a binary recommendation, evaluate against Scale & Performance Requirements; Regulatory & Compliance Obligations; Agentic Architecture Readiness; and Commercial Viability. The first three are threshold requirements, not preferences.

If these are not important for your business objectives and you have mature models, a strong in-house engineering team, dedicated program capacity, and a strong proclivity for DIY, then a custom solution may be good enough.

Decision Dimension	Pega CDH	AI-Native Custom Build
Scale & Performance Requirements	Scale and Performance Requirements	Scale and Performance Requirements Scale and Performance Requirements
Scale: 65M+ customers / 5B+ interactions/month	✅ Purpose-built	⚠️ Significant re-engineering required
Real-time multi-action arbitration (<100ms)	✅ Native	⚠️ Achievable, but cache ≠ live arbitration
1000's of self-learning (adaptive) models	✅ Native	❌ Custom MLOps overhead
Regulatory & Compliance Obligations		
Regulatory audit trail (GDPR, CCPA, HIPAA)	✅ Built-in, decade-certified	⚠️ Must build governance layer from scratch
Explainability for regulators (not just SHAP)	✅ Time-stamped strategy audit	⚠️ Requires significant compliance build
Change governance with multi-stakeholder sign-off	✅ Design/runtime separation	⚠️ Supported by custom agents but requires infrastructure & context engine build to prevent agent drift and support operational change and security
Agentic Architecture Readiness		
Agentic control plane (Scenario Planner, NBA Designer…)	✅ Available now	❌ Not available off-shelf
Full lifecycle workflow/case management integration	✅ Available now	❌ Not available off-shelf
Central location for all predictive modeling and analytics	✅ Available now	❌ Not available off-shelf
Commercial Viability		
Time to first value	✅ Fast prototype	✅ Fast prototype
Flexibility / bespoke logic	⚠️ Within platform model	✅ Full control
Cost at small scale (single use case)	⚠️ Platform licensing overhead	✅ Lower initial investment

5. Conclusion – Completing the Agentic Stack

The central finding of this analysis is not that custom builds fail. It is that CDH does not compete with agentic AI, it completes it. AI agents accelerate design-time iteration, generate strategy hypotheses, and compress development cycles. CDH provides the governed decisioning control plane—real-time arbitration, regulatory explainability, adaptive learning, and enterprise change governance—that makes those agentic builds viable, auditable, and scalable over the long term. Without that control plane, organizations face the prospect of rebuilding the same foundational capabilities independently, at a compounding cost, every time they expand to a new channel, jurisdiction, or product line.

The most durable architecture is not pure buy or pure build; it is a deliberate coexistence. Custom AI and agentic tooling handle the differentiated logic, proprietary models, and rapid experimentation where in-house teams add the most value. CDH serves as the decisioning authority underneath: arbitrating and orchestrating actions consistently across channels, enforcing governance at runtime, and closing the adaptive learning loop at enterprise scale. That division of labor lets organizations move fast at the edges without sacrificing the governed foundation at the center.

Pega and Accenture bring both the evidence and the analytical tools to help organizations pressure-test this decision against their own reality where customer volumes, regulatory exposure, model maturity, engineering capacity, and commercial ambition can vary widely. The framework in this document is a starting point. The next step is a structured assessment that maps your organization's specific decisioning landscape to the capabilities that matter most and identifies where CDH, custom build, or a combination of both delivers the strongest outcome.

Glossary

Acronym	Definition
CDH	Customer Decision Hub. Pega's centralized AI decisioning engine that arbitrates across actions, treatments, and channels in real time.
NBA	Next Best Action. The output of the decisioning process — the optimal action, offer, or treatment to present to a customer at a given moment, balancing business goals with customer relevance.
CNBA	Custom Next Best Action. A bespoke, internally built stack designed to replicate NBA-style decisioning outside of a platform like Pega CDH — the "BYOB" (Build Your Own Brain) approach the paper is evaluating.
NRT	Near Real-Time. A decisioning or data processing mode that operates with a short but non-zero lag (typically seconds to a few minutes), often achieved through event streaming and caching rather than true live arbitration.
ADM	Adaptive Decision Manager. Pega's embedded machine learning component within CDH that continuously updates propensity models based on customer interactions without requiring offline retraining cycles.
MLOps	Machine Learning Operations. The engineering discipline covering the deployment, monitoring, versioning, and retraining of ML models in production — in the paper's context, the ongoing operational burden a custom stack places on internal teams.
CI/CD	Continuous Integration / Continuous Delivery. A software engineering practice for automating the build, testing, and deployment of code changes — relevant here because a custom CNBA stack requires full CI/CD pipelines that a platform like CDH abstracts away.
Ops Manager	Pega's Operations Manager, a tool within the Pega platform for monitoring, managing, and scaling deployed decisioning and workflow processes in production environments.

Thank you

About Pega

Pega delivers innovative software that crushes business complexity. From maximizing customer lifetime value to streamlining service to boosting efficiency, we help the world's leading brands solve problems fast and transform for tomorrow. Pega clients make better decisions and get work done with real-time AI and intelligent automation. And, since 1983, we've built our scalable architecture and low-code platform to stay ahead of rapid change. Our solutions save people time, so our clients' employees and customers can get back to what matters most.
For more information, visit us at **WWW.PEGA.COM**