MACH ALLIANCE

eBook

Evaluating AI Vendors for Composable Commerce Environments

2025

Contents

MACH ALLIANCE

Introduction

As AI-native capabilities become integral to composable commerce, business leaders and decision-making delegates are tasked with making decisions about systems that think, learn, and evolve, often with little transparency and high operational impact. As an expert in due diligence on vendor selection as well as investments in technology, we have found that experience in making decisions in the selection process is mission-critical. However, with AI being a "new era", there are many areas that even experts don't yet have experience in. A key question in evaluation is no longer "*Can this AI do something that solves a problem?*" but rather "*Can I trust this AI to work the way my business needs it to, under real-world conditions?*" (deploying a hybrid human and SaaS approach).

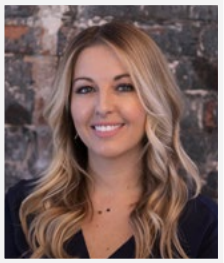
This document outlines a focused set of critical questions, considerations, and operational insights that leaders can use when evaluating AI vendors for composable ecosystems, especially when those tools play a role in customer experience, automation, or decision-making. Drawing from our collective experience across enterprise technology, retail, and composable commerce implementations, we've identified the most critical evaluation criteria that separate truly enterprise-ready AI solutions from those that merely promise compatibility.

This represents our initial framework for AI vendor evaluation in composable environments, with plans to expand these guidelines based on community feedback and evolving market needs.

Contributors

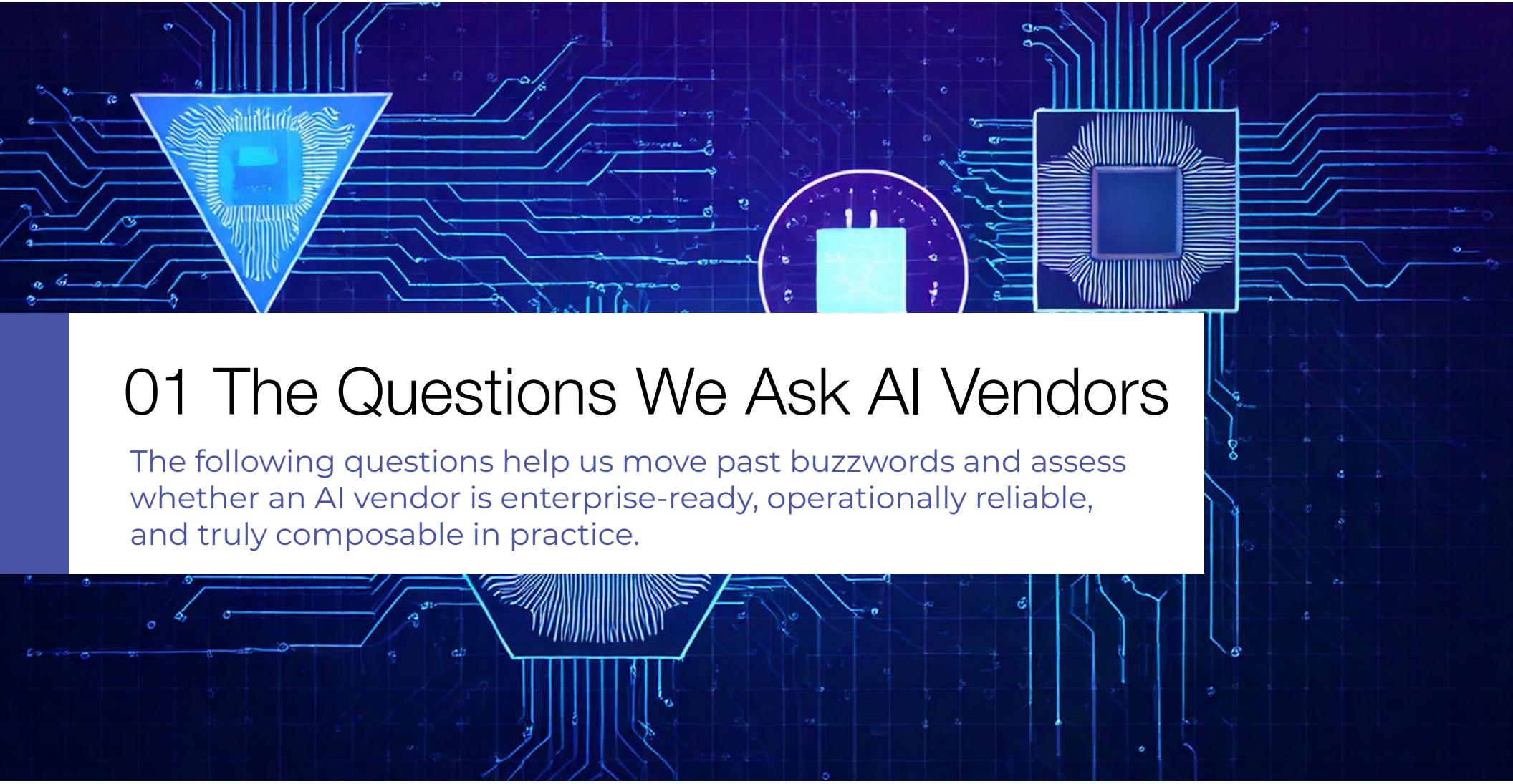
This framework has been developed by MACH Alliance Ambassadors, practitioners and technology leaders working across enterprise commerce, to help brands evaluate AI vendors with the same rigor they apply to their own decisions.

- **Danielle Diliberti** - CEO, Sommsation



- **Dylan Valade** - Management Consultant, Cuesta Partners

- **Krishna Gangu** - Distinguished Architect, JCPenney



01 The Questions We Ask AI Vendors

The following questions help us move past buzzwords and assess whether an AI vendor is enterprise-ready, operationally reliable, and truly composable in practice.

1. Composable Architecture and Integration

What to ask:

- Is your product fully API-based? Do you support multi-tenant or modular deployment?
- How do you handle real-time vs. batch interactions?
- Can we test integrations in staging environments with production-level functionality?
- What technologies and partners are used to develop, run, and monitor your system?
- How much of this solution did you acquire versus build? Is it running on your existing infrastructure or did you start fresh?
- What is your strategy to support interoperability protocols? MCP (Model Context Protocol), Agent2Agent, or any others that enable vendor independence?
- How do you decide when to use open source or proprietary technologies in your services? How does this impact our ability to avoid vendor lock-in?

Why it matters:
Even the most promising AI breaks down if it can't integrate cleanly into your stack or scale across your user base. API limitations, lack of environment parity, and reliance on fragile batch processes are early red flags that can undermine composable architecture benefits.

Practical insight:
One company found a vendor's "real-time" analytics offering was actually a daily batch export. By the time sales teams got the insights, the moment had passed.

2. Data, Training, and Model Behavior

What to ask:

- What is the origin of your training data? Is it proprietary, synthetic, or third-party?
- Can we fine-tune the model with our domain-specific data without compromising data sovereignty?
- How frequently are models retrained, and can we trigger retraining programmatically?
- Are audit logs or decision explanations available through standard interfaces?
- What skills and mindset does my team need to maximize the value of our data and processes using your service?

Why it matters:
AI should reflect *your business context*, not a generalized dataset. Without visibility into how the model learns and evolves, you can't troubleshoot or adapt it to your needs. Data protection and bias are also areas of concern, especially when data flows between multiple vendors in your stack.

Practical insight:
In one case, a company needed AI-generated product descriptions for high-end goods. The model, trained on general e-commerce data, introduced tone inconsistencies that undercut brand equity.

3. Transparency and Explainability

What to ask:

- Do predictions come with confidence scores, traceable logic, or source references accessible through your platform? Can these be configured per use case?
- How do you handle edge cases or ambiguous scenarios?
- Can we view a changelog of model updates through your system?
- Are rollback options available, and can they be triggered when needed?
- What is the funding source for your AI services? Is it profitable yet, and how does this affect long term stability?
- Is there anything you would never use your AI for, any clear no-go zones that align with ethical business practices?

Why it matters:
When AI behavior shifts, operational teams need to *understand and explain it*, especially in customer-facing or regulated environments. Transparency becomes even more critical when AI decisions impact multiple downstream systems.

Practical insight:
A personalization engine reprioritized homepage items. Conversion rates dropped, but there were no logs or explanations to identify the cause. The fix took days, and eroded internal confidence in the AI system.

4. Rollout, Updates, and Change Management

What to ask:

- How are model updates or new features deployed? Do they follow established deployment practices?
- Are changes communicated in advance with impact assessments?
- Do you offer feature flags, A/B testing, or staged rollout options that work with existing operational workflows?
- Can releases align with our product or business cycle rather than forcing us to adapt to your schedule?
- Assuming successful adoption, how can we maintain control over costs? Can you rate limit calls by user group or implement usage-based pricing tiers?
- Our developers have made app changes that inadvertently result in thousands or millions of unexpected API calls in production where a fast rollback is not feasible. Can production services be gracefully disabled to avoid runaway costs from implementation mistakes?
- Can we use our own API keys and contracts for services in your application? (OpenAI, AWS, Google, Azure…)
- Are your existing teams now supporting both your AI and existing products or did you hire separately for the AI effort? Have you redirected resources from your current offerings to fund and build AI? Are you still delivering feature updates to your existing products on schedule?
- What needs to be true for us to be successful using your service? What does our team need to know? Do we need to hire or reassign any roles?

Why it matters:
Even strong AI can cause disruption if rolled out poorly. Sudden model changes can break workflows, confuse users, or trigger unintended downstream effects across your technology stack. A strong vendor will help streamline training and education for teams while respecting your operational processes.

Practical insight:
An AI tool that powered pricing suggestions released a silent update mid-quarter. It recalibrated discounts without approval, throwing off margin forecasts for a retail chain.

5. Support, Documentation, and Escalation

What to ask:

- Do we have a dedicated account or customer success contact with technical expertise?
- What are your support hours and response time commitments for different severity levels?
- Are self-service diagnostics, admin tools, or developer sandboxes available with comprehensive documentation?
- How quickly can engineering triage a model-specific issue, and do you have escalation paths to AI specialists?

Why it matters:
AI issues are harder to debug than typical SaaS errors. Without clear support and escalation paths, small issues can escalate into operational crises that affect your entire integrated system.

Practical insight:
One AI customer service assistant misrouted high-value clients due to a configuration error. Without direct escalation access, resolution took three days, well beyond brand SLA expectations.

6. Governance, Bias, and Brand Alignment

What to ask:

- What safeguards do you use to detect and mitigate bias in model outputs?
- Can model tone, behavior, or logic be adjusted to match our brand or ethics through configuration rather than custom development?
- Do you support human-in-the-loop override capabilities that can be integrated into our workflows?
- Can we implement and enforce specific content guidelines or suppression rules?
- How do you handle regulatory compliance requirements that may vary by geography or industry?

Why it matters:

AI reflects the data it sees, but *your brand reflects what customers experience*. If the AI contradicts your voice, values, or tone, it does more harm than good. In integrated environments, these issues can propagate across multiple touchpoints and channels.

Practical insight:

A generative tool produced marketing copy for a premium product line. Phrases like “cheap thrill” and “boozy fun” appeared, undermining the luxury positioning across all marketing channels.

02 How We Apply This in Real Business Contexts

These aren't theoretical concerns. They're decisions we've made under pressure, in market, and with real customers. Below are example scenarios that highlight how these questions play out across industries.

Use Case: Customer Personalization Engine

- **Goal:** Improve product discovery with AI-based recommendations that integrate with existing commerce systems.
- **What works:** A vendor offering a sandbox with version-controlled APIs and detailed logic explainers
- **What fails:** Changing the recommendation logic without notice. Teams can't explain shifts in product performance or user paths.

Use Case: Generative Content for Product Pages

- **Goal:** Scale SEO content and product descriptions using AI
- **What works:** Allowing for editorial tone calibration and confidence scoring for fact-checking
- **What failed elsewhere:** One model hallucinated food pairings and used inconsistent terminology, requiring manual rewrites that negated efficiency gains

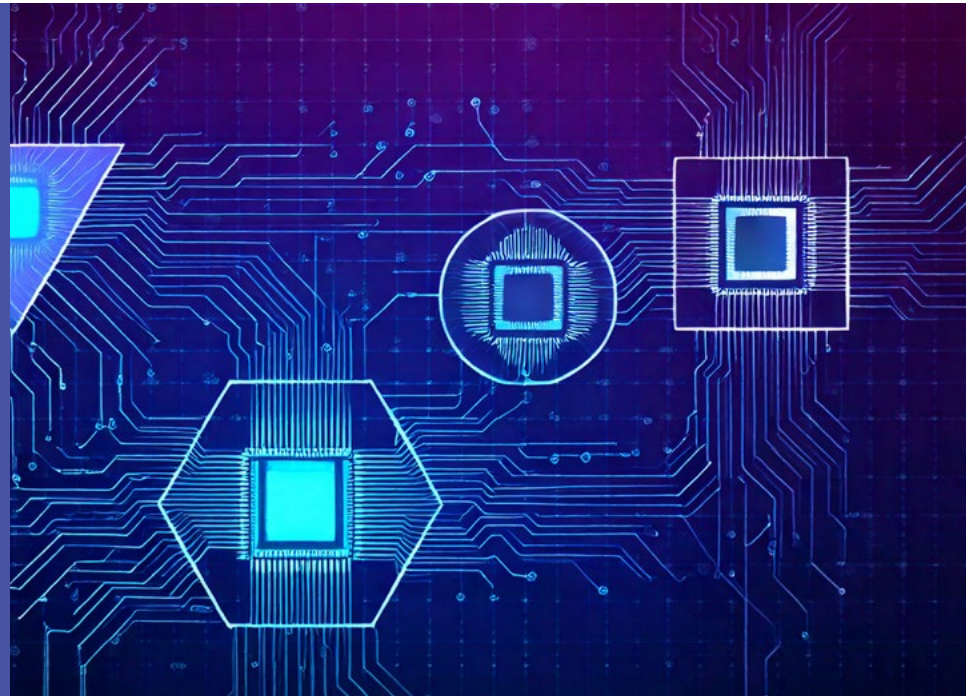
Use Case: Inventory Forecasting via AI Agents

- **Goal:** Use AI to detect anomalies and adjust demand planning
- **What works:** Transparent data pipeline mapping allowed for override logic and retraining triggers
- **What fails:** A "black box" vendor misread seasonal patterns and triggered unnecessary restock orders, costing thousands in overstock

Internal Dialogue That Guides Our Thinking

These are the conversations that happen internally when we evaluate AI systems. They may not show up on a scorecard - but they drive the decision:

- "Can my operating team explain what the AI is doing?"
- "If it changes tomorrow, how will we know?"
- "What happens when it's wrong - and who's accountable?"
- "Do we control the inputs, the outputs, or neither?"
- "Does this support how we work - or force us to work how *they* built it?"
- "How does this fit into our architecture without creating new silos?"
- "Can we replace this AI vendor without rebuilding our entire system?"



Certification Should Validate More Than Just the Model

We believe a MACH AI Certification should address:

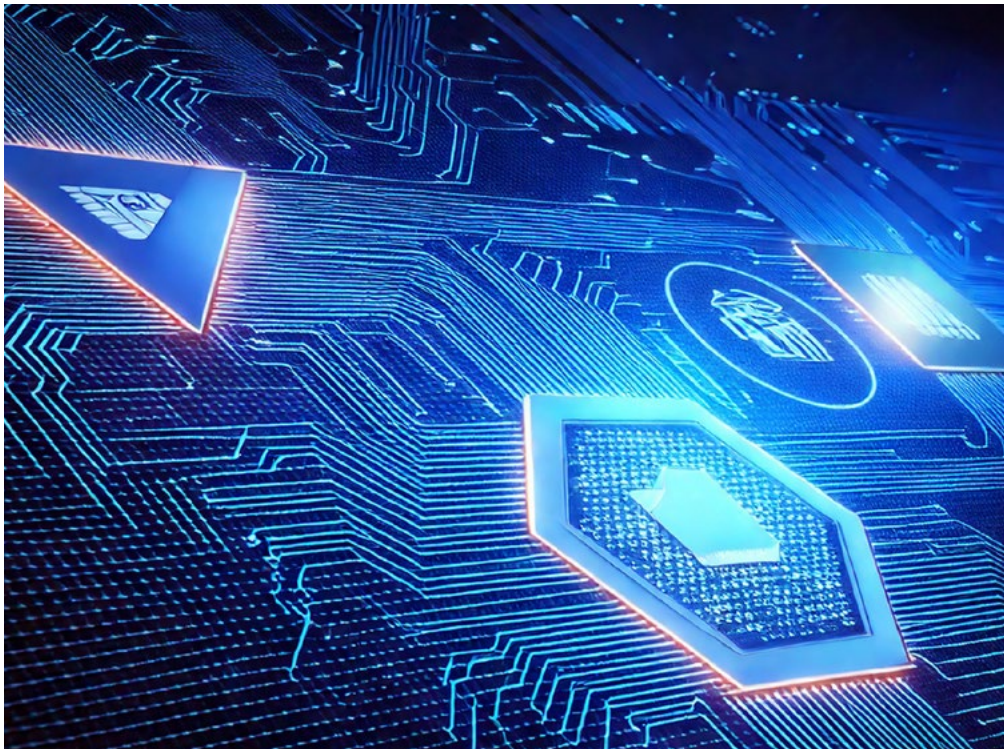
Area	Certification Implication
Transparency and Explainability	Vendors should provide visibility into AI logic, versioning, and update impacts through accessible interfaces and documentation
Operational Fit	Tools should adapt to client release cycles, escalation needs, and real-world workflows rather than forcing proprietary processes
Data Control and Customization	Clients should have agency over retraining, suppression rules, and fine-tuning inputs while maintaining data sovereignty
Support Maturity	Live support, SLAs, and self-service tools must be documented, auditable, and designed for technical teams
Ethical and Brand Alignment	Vendors must support configurable values, tone, fairness controls, and override capabilities for client-facing outputs
Integration & Independence	Solutions should avoid lock-in through standard protocols, portable data formats, and interoperable interfaces

Conclusion

Evaluating AI vendors for composable commerce requires looking beyond the model's capabilities to assess how well the entire solution fits into a modern, integrated architecture. The questions and frameworks outlined here help ensure that AI investments support rather than undermine the flexibility, scalability, and vendor independence that define successful composable strategies.

> The future of commerce is composable, and the AI that powers it must be as well.

Find out more? →



About The MACH Alliance

The MACH Alliance is a not-for-profit industry body advancing composable enterprise architecture through five core principles: Composable, Connected, Incremental, Open, and Autonomous. With over 100 global member companies, we provide certification, community, and guidance to help organizations adopt transformative technologies and future-proof their businesses. Learn more at machalliance.org, or on LinkedIn.