



Executive AI Due-Diligence Scorecard

How to Use This Scorecard

This scorecard provides a structured approach to evaluate AI initiatives and investments. Each criterion should be scored based on its readiness and potential impact.

Scoring Ranges:

| | | | |
|----------------------------------|----------------------------------|--------------------------------------|------------------------------|
| 30-35 Enterprise-ready | 21-29 Viable with gaps | 15-20 Requires development | Below 15 High risk |
|----------------------------------|----------------------------------|--------------------------------------|------------------------------|

1. LLM / Tool Agnosticism

Score: 5

Runs multiple LLM providers natively; easy routing & switching; supports future models without rebuilds

Score: 4

Supports several providers; switching requires minor configuration; limited provider-specific constraints

Score: 3

Supports 2–3 models but with restrictions; switching requires developer work or context loss

Score: 2

Primarily locked into one provider; integrations available only via custom code or brittle plugins

Score: 1

Single-model lock-in with no ability to swap or add alternatives

2. Persistent Knowledge (Prompts, Threads, Outputs)

Score: 5

Saves prompts, chats, outputs, and notes automatically; searchable; organized by project; reusable

Score: 4

Can save and retrieve prompts/threads with tagging; some project structure

Score: 3

Allows manual saving or exporting; minimal organization; limited search

Score: 2

Chats disappear unless manually copy/pasted; no organizational structure

Score: 1

Cannot save or retrieve conversations or prompts in any structured way

3. Context Control & Recovery

Score: 5

Direct context editing + forking/versioning; can revert to earlier states instantly

Score: 4

Supports some editing of system prompts or memory; limited version history

Score: 3

Can reset context or start new threads; no editing or branching

Score: 2

Must manually copy/paste context to simulate control

Score: 1

One linear chat; model drift cannot be corrected

4. Integration with Existing Data & Systems

Score: 5

Secure connectors to docs, databases, and apps; natural-language querying; robust permissions

Score: 4

Supports key systems with authentication; some NL-to-data functionality

Score: 3

Connects to a few systems; primarily manual uploads or one-off APIs

Score: 2

Copy/paste workflows; inconsistent connectors; limited automation

Score: 1

Cannot access or work with existing internal data in any meaningful way

5. Cost Visibility & Guardrails

Score: 5

Real-time governance with provider-level usage tracking, per-project/user budgets, alerts, and optimization intelligence

Score: 4

Offers usage dashboards and some budget controls; limited alerts

Score: 3

Shows monthly or aggregate usage; no quotas or alerts

Score: 2

Only provider invoices; little granularity; teams blind to usage

Score: 1

No reporting or tracking; unclear who is using what and at what cost

6. Human-in-the-Loop Quality / Anti-Workslop

Score: 5

Formal quality process with role-based review, evaluation rubrics, required human oversight, and documented acceptance criteria

Score: 4

Consistent review structure in most workflows; optional rubrics

Score: 3

Teams review outputs manually; no formal standards

Score: 2

Occasional spot checks; inconsistent practices

Score: 1

Raw model outputs used without any human validation

7. Security, Governance & Portability

Score: 5

Enterprise SSO, granular RBAC, full audit logs, data export APIs, retention policies, VPC/private deployment options

Score: 4

Strong access controls and logging; data export available; some compliance certifications

Score: 3

Basic user roles and permissions; manual export processes; limited audit capability

Score: 2

Minimal access controls; data portability requires support tickets; unclear data policies

Score: 1

No governance features; data trapped in vendor system; no audit trail

Final Score

| Criteria | Score (1-5) |
|--|-------------|
| LLM / Tool Agnosticism | |
| Persistent Knowledge (Prompts, Threads, Outputs) | |
| Context Control & Recovery | |
| Integration with Existing Data & Systems | |
| Cost Visibility & Guardrails | |
| Human-in-the-Loop Quality / Anti-Workslop | |
| Security, Governance & Portability | |
| Total | |

Scoring Ranges:



- Dealbreaker Note:** Any criteria scoring 1-2 in critical areas should be considered a dealbreaker, regardless of the overall score, and warrants immediate attention or reconsideration of the initiative.