Data foundations checklist: the 5 pillars for AI-ready data

Assess your team's readiness to deliver AI at scale — and uncover your biggest opportunities to improve.

Why data foundations matter

Gartner predicts that 60% of AI projects unsupported by AI-ready data will be abandoned by the end of 2026. Most fail before the model is even trained. The culprit? Weak data foundations.

Even the most sophisticated AI tools will underdeliver without one critical element: a data foundation your teams can trust. High-performing AI starts here - not with models or algorithms.

At Emergn, we help global enterprises build the data, strategy, and capability to turn AI ambition into measurable business results. Strong data foundations give your AI a competitive edge by helping you:

- Ensure AI outputs are accurate, relevant, and trusted.
- Reduce rework and delays caused by data issues.
- Unlock cross-functional collaboration and faster delivery.

For more information about Emergn, please contact us or visit emergn.com
UK +44 808 189 2043 US +1 617 482 0830 Email info@emergn.com

How to use this guide:

This checklist is designed to help you quickly evaluate how prepared your organization's data is to support AI initiatives at scale.

For each section:

1. Answer the questions.
2. Self-assess your current state:
   - Red: Significant gaps; urgent focus required.
   - Amber: Some capability; targeted improvement needed.
   - Green: Strong foundation; maintain and improve.
3. Review the guidance to understand why it matters, the risks if ignored, and what to do next. Teams often find value in completing this assessment together. Use it to spark discussion, surface alignment gaps, and identify the next best moves.

Pillar #1: Data quality

QUESTIONS:	YOUR ASSESSMENT:		
Is the data accurate, complete, and timely?	Red ☐	Amber ☐	Green ☐
Are definitions standardized and applied consistently across systems and teams?	Red ☐	Amber ☐	Green ☐

Why this matters:

High-quality data is the lifeblood of effective AI. If the data is incomplete, out-of-date, or defined differently across teams, AI will generate misleading outputs—fast. Product-led organizations treat data quality as a product feature, not an IT afterthought.

Broader Implications:

- Low-quality data undermines trust in AI outputs.
- Teams waste time reconciling conflicting information.
- Higher AI failure rates and wasted investment.

What 'good' looks like:

Data is >95% accurate and complete, refreshed on a regular cadence (daily/weekly depending on need). Definitions are standardized and consistently applied across teams and systems.

Pillar #2: Ownership and accountability

QUESTIONS:	YOUR ASSESSMENT:		
Do all critical datasets have clearly assigned owners?	Red ☐	Amber ☐	Green ☐
Are business and technical owners collaborating on data standards and definitions?	Red ☐	Amber ☐	Green ☐

Why this matters:

Without clear ownership, no one is accountable for ensuring data is accurate, up-to-date, and relevant. This leads to inconsistent definitions, duplicate data sources, and mistrust in AI outputs. In product-led organizations, ownership ensures data decisions are tied to customer and business outcomes, not just technical feasibility.

Broader Implications:

- Data initiatives stall without a decision-maker.
- Poor governance leads to compliance risks and costly rework.
- AI models trained on ownerless data are more prone to bias and errors.

What 'good' looks like:

Every critical dataset has a named owner. Business and technical leaders share responsibility for data standards and definitions, ensuring decisions are tied to customer and business outcomes, not just technical feasibility.

Pillar #3: Continuous data improvement

QUESTIONS:	YOUR ASSESSMENT:		
Are there ongoing processes to monitor and improve data quality?	Red ☐	Amber ☐	Green ☐
Are product teams empowered to fix upstream issues as part of daily work?	Red ☐	Amber ☐	Green ☐

Why this matters:

Data readiness is not a one-time project; it's an ongoing practice. Without a continuous improvement loop, even high-quality data will degrade over time hurting AI performance and agility. Product-led organizations embed data improvement into daily workflows so that quality scales with capability.

Broader Implications:

- Declining data quality over time leads to AI performance drops.
- Increased operational costs from rework and patch fixes.
- Loss of competitive advantage as agility declines.

What 'good' looks like:

Data quality is continuously monitored with KPIs for accuracy, timeliness, and completeness. Product teams address upstream issues as part of daily work, not ad-hoc projects.

Pillar #4: Context and metadata

QUESTIONS:	YOUR ASSESSMENT:		
Is data accompanied by clear definitions, lineage, and usage rules?	Red ☐	Amber ☐	Green ☐
Do teams understand the business meaning behind the numbers?	Red ☐	Amber ☐	Green ☐

Why this matters:

AI can process data, but without human-provided context, results can be technically correct yet strategically irrelevant. Context ensures that outputs are interpreted accurately, decisions are informed, and compliance risks are minimized. In product-led organizations, context turns raw data into actionable insights that directly inform decisions.

Broader Implications:

- Misinterpretation of AI outputs leading to poor strategic choices.
- Loss of stakeholder trust in data and AI initiatives.
- Higher risk of regulatory non-compliance.

What 'good' looks like:

Critical datasets include metadata (lineage, definitions, usage rules). Teams share a common understanding of what the data represents and how it informs decisions.

Pillar #5: Accessibility and integration

QUESTIONS:	YOUR ASSESSMENT:		
Can authorized teams easily access the data they need without bottlenecks?	Red ☐	Amber ☐	Green ☐
Do teams understand the business meaning behind the numbers?	Red ☐	Amber ☐	Green ☐

Why this matters:

Clean, complete data is useless if it's locked in silos. Inaccessible data slows delivery and limits AI's ability to combine insights across domains. In product-led organizations, accessibility and integration enable teams to move from idea to validated value faster.

Broader Implications:

- Slower time-to-market for AI-powered features.
- Increased costs from duplicated storage and effort.
- Missed opportunities for cross-functional innovation.

What 'good' looks like:

Authorized teams access data in minutes, not days. 80–90% of critical systems are integrated, allowing data to flow seamlessly across the enterprise.

Your AI needs many elements to succeed, but the foundation must be right: data your teams can trust and use with confidence. The stronger that foundation, the faster you can deliver AI at scale, with outputs that are accurate, trusted, and tied to business outcomes.

Email Emergn at
thrive@emergn.com
to schedule a **20-minute AI Data Readiness Review** with us.

More Reds or Ambers than Greens? That's a signal your AI foundations need work, and a chance to fix the issues that slow delivery, frustrate teams, and erode trust in outputs.

We'll help you:

- **Pinpoint** your top 2–3 most urgent data gaps.
- **Sequence** improvements for fastest impact.
- **Uncover** hidden risks before they derail your AI investment.

Don't just build AI. Build the data foundation that gives you lasting advantage

For more information about Emergn, please contact us or visit emergn.com
UK +44 808 189 2043 US +1 617 482 0830 Email info@emergn.com