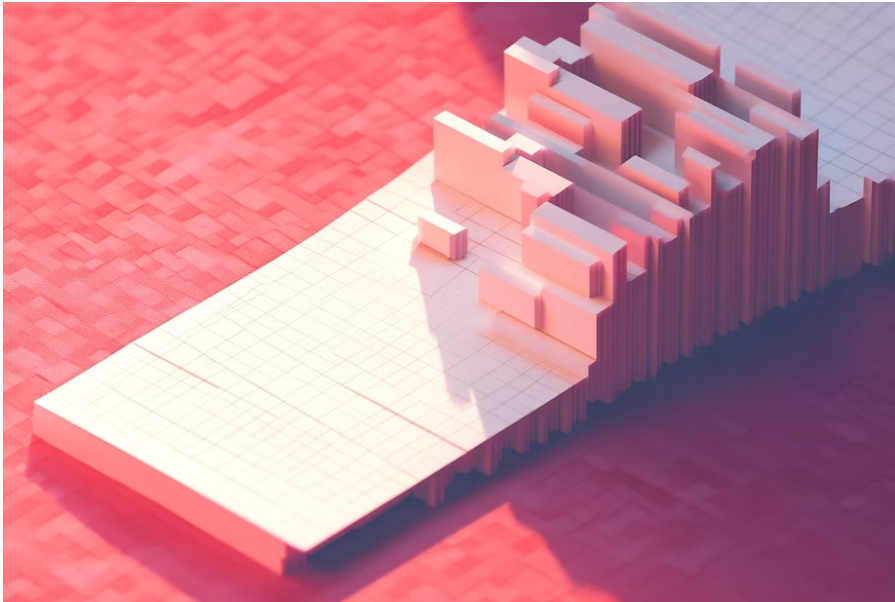# Why Your AI Projects are Built on a House of Cards

September 3, 2025

This week's latest from Project Flux

We're in the middle of an AI gold rush, with businesses pouring billions into artificial intelligence, hoping to strike it rich with unprecedented efficiency and innovation.

Yet, for all the investment and hype, a staggering number of projects are quietly imploding. A recent report from S&P Global highlights a deeply concerning trend: the failure rate of AI projects is not only high but *rising* [1]. This isn't just a minor setback; it's a systemic crisis rooted in a fundamental misunderstanding of how we measure success. We are, in essence, building complex, expensive, and potentially world-changing systems on a foundation of flawed and inadequate metrics.

A groundbreaking [MIT report](#) has peeled back the curtain on this uncomfortable reality, revealing that our current methods for evaluating AI are not just failing but are actively misleading us [2]. The traditional benchmarks and leaderboards we rely on to gauge AI performance are often disconnected from real-world applications, creating a dangerous gap between what we *think* our AI can do and what it *actually* delivers. This isn't just an academic debate; it's a critical business issue with profound implications for project delivery professionals. As the people responsible for turning AI's promise into reality, we are on the front lines of this measurement crisis, and it's our projects that pay the price for these flawed evaluations.

> "We are flying blind, making multi-million-pound decisions based on metrics that are little more than vanity numbers."

This isn't about a lack of effort or investment. It's about a lack of a shared, meaningful language for what

'good' looks like in AI. We celebrate models that top the charts on abstract benchmarks, but we fail to ask the hard questions about whether those benchmarks have any bearing on the messy, complex, and unpredictable environments our projects operate in. The result is a cycle of hype, investment, and disappointment that is eroding trust in AI and costing businesses dearly. We are, in effect, trying to build a skyscraper on foundations of sand, and the cracks are already beginning to show.

The Industry Blind Spot: Our Obsession with Leaderboards

The core of the problem lies in our industry-wide obsession with standardised benchmarks and leaderboards. While these tools provide a sense of order and progress, they have become a dangerous crutch, encouraging a narrow, reductionist view of AI capabilities. The arXiv paper "On the Six Contending Paradigms of AI Evaluation" brilliantly deconstructs this issue, identifying six distinct evaluation paradigms that are often conflated or ignored [3]. These paradigms range from task-specific performance metrics to more holistic assessments of robustness, fairness, and real-world impact. The trouble is, we've become fixated on the former, while largely neglecting the latter.

This narrow focus creates a veneer of scientific rigour, but it's a dangerously misleading one. We see a model climb the ranks of a popular leaderboard and assume it's 'better' in a general sense, but we fail to scrutinise what that leaderboard is actually measuring. Is it testing for the specific capabilities our project needs? Is it evaluating the model's performance in a context that resembles our own? More often than not, the answer is no.

This blind spot is exacerbated by the complexities of measuring AI ROI. As a Devoteam article points out, the value of AI is often indirect and difficult to quantify with traditional financial metrics [4]. It might manifest in improved customer satisfaction, enhanced decision-making, or reduced operational risk – all of which are notoriously tricky to measure. By focusing on easily quantifiable but ultimately superficial metrics, we are missing the bigger picture and failing to capture the true value (or lack thereof) of our AI investments. This isn't just a technical issue; it's a strategic one. As project delivery professionals, we need to be able to articulate the value of our work in a language that the business understands, and our current measurement toolkit is simply not up to the task.

What We're Missing: The Human Element in a World of Machines

In our relentless pursuit of technical perfection, we are overlooking the most critical factor in any project's success: the human element. We are so focused on optimising algorithms and chasing leaderboard scores that we are failing to account for the complex, often irrational, and deeply human contexts in which our AI systems will be deployed. The RAND Corporation's report on AI project failures provides a sobering reminder of this reality, highlighting that the root causes of failure are often not technical but organisational and cultural [5].

We are failing to ask the right questions at the outset of our projects. We are not adequately defining what success looks like in human terms. We are not engaging with the end-users to understand their needs,

their workflows, and their fears. We are not building the necessary feedback loops to capture the nuanced, qualitative data that tells us whether our AI is actually helping or hindering. We are, in short, treating AI as a purely technical problem, when it is, in fact, a deeply socio-technical one.

> "We're engineering solutions in a vacuum, and then we're surprised when they shatter on contact with the real world."

This is where the role of the project delivery professional becomes absolutely critical. We are the bridge between the technical teams building the AI and the business units that will use it. We are the ones who can translate between the arcane language of machine learning and the practical realities of the shop floor. We are the ones who can ensure that the human element is not just an afterthought but a central consideration in every stage of the project lifecycle. To do this, we need to expand our own skillsets, moving beyond traditional project management methodologies to embrace a more holistic, human-centric approach. We need to become adept at user research, stakeholder management, and change management. We need to become champions for the user, ensuring that their voices are heard and their needs are met. And we need to be brave enough to challenge the status quo, to push back against the tyranny of meaningless metrics, and to demand a more nuanced and meaningful conversation about what it truly means for an AI project to succeed.

What We Can Actually Do About It: A New Manifesto for AI Project Delivery

This isn't a time for despair; it's a time for action. As project delivery professionals, we are uniquely positioned to lead the charge in redefining how we measure and deliver AI projects. Here's a four-point manifesto for a new, more effective approach:

1. Redefine Success Beyond the Leaderboard: We must move beyond a myopic focus on standardised benchmarks and develop a richer, more nuanced understanding of what success looks like for each individual project. This means working closely with stakeholders to define clear, measurable, and meaningful business outcomes from the outset. It means developing custom evaluation metrics that are tailored to the specific context of our projects. And it means embracing a more qualitative approach to measurement, capturing user feedback, and assessing the real-world impact of our AI systems.

2. Champion a Human-Centric Approach: We must put the human element back at the heart of our projects. This means conducting thorough user research to understand the needs, workflows, and pain points of the people who will be using our AI systems. It means co-designing solutions with end-users, ensuring that our AI is not just technically proficient but also intuitive, usable, and trustworthy. And it means investing in change management, providing the training, support, and communication that people need to adapt to new ways of working.

3. Build a Culture of Critical Evaluation: We must foster a culture of healthy scepticism and critical evaluation within our project teams. This means encouraging open and honest conversations about the limitations of our AI systems. It means creating safe spaces for people to raise concerns and challenge assumptions. And it means rewarding people for identifying and addressing problems, not just for hitting

arbitrary targets. We need to move away from a culture of "AI solutionism" and towards a more realistic and responsible approach to innovation.

4. Demand Better from Our Tools and Partners: We must demand more from the vendors and partners who provide our AI tools and platforms. This means asking tough questions about their evaluation methodologies. It means pushing for greater transparency in how their models are trained and tested. And it means refusing to be seduced by impressive-sounding but ultimately meaningless marketing claims. We need to be informed and discerning consumers of AI technology, not just passive recipients.

This is not just about improving our project success rates; it's about shaping the future of AI in a way that is responsible, ethical, and genuinely beneficial to humanity. As project delivery professionals, we have a critical role to play in this endeavour. Let's seize this opportunity to lead the way.

Call-to-action: Don't let your projects become another statistic. Subscribe to Project Flux for the insights and strategies you need to navigate the complexities of AI delivery and build a future that works.

References

[1] S&P Global. (2025, August 22). *AI Project Failure Rates on the Rise*. CIO Dive. Retrieved from https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/

[2] MIT Computer Science and Artificial Intelligence Laboratory. (2025, February 26). *MIT Report on AI Measurement Failures* [Video]. YouTube. https://youtu.be/ly6YKz9UfQ4?feature=shared

[3] Arvind, V., et al. (2025). *On the Six Contending Paradigms of AI Evaluation*. arXiv. https://arxiv.org/html/2502.15620v1

[4] Devoteam. (n.d.). *The Complexities of Measuring AI ROI*. Devoteam. Retrieved from https://www.devoteam.com/expert-view/the-complexities-of-measuring-ai-roi/

[5] RAND Corporation. (2023). *The Role of Organizational and Cultural Factors in AI Project Failures*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2680-1.html