Beyond ROI: Are We Using the Wrong Metric in Measuring AI Success?

Author: David Gallacher
Industry Fellow / Faculty Director, Professional Programs

A Response to MIT's "95% AI Failure" Study

MIT's recent report claiming that 95% of generative AI projects fail to deliver measurable return on investment has sent shockwaves through the business community. The study, "The GenAI Divide: State of AI in Business 2025," found that despite $30-40 billion in enterprise investment, 95% of organizations studied are seeing zero return on their AI initiatives. While this headline grabbing statistic has dominated media coverage, it reveals a more fundamental question: Are we measuring AI success with the wrong metrics entirely?

The obsession with traditional ROI in AI implementations reflects the same flawed thinking that has plagued every major technological transformation. The expectation that a new technology can accrue immediate financial benefits. When email was introduced, companies didn't immediately see their profits increases. When the internet emerged, organizations didn't abandon it because quarterly earnings didn't immediately spike. Yet somehow, with AI, we've reverted to expecting immediate, measurable financial returns within months of pilot programs being launched.

The Fundamental Measurement Problem

MIT's study defined success narrowly as "deployment beyond pilot phase with measurable KPIs" and "ROI impact measured six months post pilot." This approach ignores the reality of how transformational technologies create value and reflects a profound misunderstanding of AI's current role in organizational transformation.

The problem isn't that AI doesn't work—it's that we're applying industrial-era metrics to a cognitive-era transformation. Traditional ROI calculations work well for capital

equipment purchases or process optimizations, but they fundamentally misunderstand how AI creates value in modern knowledge work.

Alternative Metrics That Actually Matter

Return on Efficiency (ROE) Instead of focusing solely on revenue increases, organizations should measure time savings and productivity gains. When a marketing team reduces content creation time from hours to minutes, or when legal teams accelerate contract review by 60%, the value isn't immediately visible in quarterly earnings but represents significant efficiency improvements that compound over time.

Quality of Work Enhancement AI often improves the quality of outputs rather than just speed. Better analysis, more comprehensive research, enhanced creativity, and reduced errors represent real value that traditional ROI calculations miss entirely.

Employee Satisfaction and Retention Generic tools like ChatGPT excel for individuals because of their flexibility, and our research shows personal AI productivity tools demonstrate the highest success rates. When employees can eliminate mundane tasks and focus on higher-value work, job satisfaction increases—a benefit that shows up in retention metrics, not immediate profit calculations.

Workforce Capability Expansion AI enables individuals to perform tasks previously requiring specialized skills or team coordination. A single employee can now conduct market research that previously required a consulting firm or create marketing materials that once needed an agency. This capability expansion represents strategic value that transcends quarterly financial metrics.

Understanding the AI Adoption Spectrum

Our research at UC Berkeley's SCET AI Commons initiative has identified why the MIT study's findings, while alarming, misses the bigger picture. We've developed a framework that recognizes five distinct use cases for AI implementation:

Personal Productivity AI shows the highest success rates precisely because individuals can experiment, iterate, and measure value in terms of personal efficiency rather than corporate ROI. The MIT study found that startups, which often don't have

such entrenched business processes to begin with, are much more likely to find genAI can deliver ROI. This aligns with our observation that personal and small-team implementations consistently outperform enterprise-wide initiatives.

Team-Level AI represents perhaps the most promising near-term opportunity but requires measurement frameworks that capture collaborative efficiency and team capability enhancement rather than direct profit impact.

Enterprise-Wide AI, where most of the MIT study's failures likely occurred, demands the most sophisticated change management approaches and the longest measurement timeframes. Expecting six-month ROI from enterprise transformation is like expecting immediate profits from other long term transformation implementations.

The Learning Gap vs. The Measurement Gap

MIT researchers exposed a "learning gap"—people and organizations simply did not understand how to use the AI tools properly or how to design workflows that could capture the benefits of AI while minimizing downside risks. But there's an equally critical measurement gap: organizations don't understand how to quantify AI's value creation in knowledge work environments.

Consider these scenarios that traditional ROI calculations miss:

- A customer service team that handles 20% more complex inquiries without additional staff because AI handles routine questions

- Engineers who can explore more design alternatives because AI accelerates prototyping

- Analysts who provide more comprehensive reports because AI expands their research capabilities

- Managers who make better decisions because AI enhances their access to relevant information

Each represents real business value, but none shows up as immediate profit increases in quarterly reports.

The Discovery of a New Workforce

We've essentially discovered a new continent of AI workers—highly capable, efficient, and tireless digital employees who never need coffee breaks, never have bad days, and never call in sick. But like any new workforce integration, this requires time to optimize workflows, establish management practices, and measure appropriate outcomes.

The MIT study's focus on six-month ROI timeframes reflects the same thinking that would have deemed the internet a failure in 1995 because most corporate websites weren't generating immediate profits. Transformational technologies require transformational thinking about value measurement.

The Success Stories Hidden in Plain Sight

MIT found the biggest ROI in back-office automation—eliminating business process outsourcing, cutting external agency costs, and streamlining operations. These successes often appear as cost avoidance rather than revenue generation, making them invisible to traditional ROI calculations focused on top-line growth.

Companies aren't as quick to replace staff members who leave, especially in customer support and administrative roles. This represents significant value through avoided hiring costs, training expenses, and the productivity gains from AI handling routine tasks—but it doesn't register as "measurable ROI" in six-month pilot studies.

A Framework for Measuring AI Success

Organizations should adopt a multi-dimensional measurement framework that includes:

- **Efficiency Metrics**: Time saved, processes automated, tasks completed per employee.

- **Quality Metrics**: Error reduction, customer satisfaction improvements, decision accuracy

- **Capability Metrics**: New tasks enabled, skill amplification, creative output enhancement

- **Strategic Metrics**: Competitive advantage gained, market responsiveness improved, innovation acceleration

- **Human Metrics**: Employee satisfaction, retention rates, learning velocity

We Get What We Measure, and We Get What We Incent

The fundamental principle "we get what we measure" explains much of the disconnect revealed in MIT's study. Organizations measuring only short-term ROI will inevitably optimize for short-term financial returns, missing the efficiency gains and capability enhancements that represent AI's primary value creation in knowledge work.

Similarly, "we get what we incent" means that if executives are rewarded based on quarterly profit increases, they'll abandon AI initiatives that deliver tremendous efficiency gains but don't immediately boost the bottom line. When performance reviews and bonuses are tied to traditional financial metrics, employees and managers will naturally gravitate toward initiatives that move those specific needles, regardless of broader organizational value.

If organizations genuinely want AI to increase efficiency, reduce cognitive burden, and enhance human capability, they must establish metrics and incentive structures that reward these outcomes. This means:

- **Rewarding managers** for team productivity improvements and employee satisfaction scores, not just budget performance

- **Measuring and celebrating** time-to-completion reductions, quality improvements, and capability expansions

- **Incentivizing experimentation** with personal and team-level AI tools rather than penalizing departments that don't show immediate profit increases

- **Aligning executive compensation** with long-term transformation metrics rather than quarterly financial results

The companies succeeding with AI are those that have aligned their measurement and incentive systems with the actual value AI delivers. They measure what matters and reward the behaviors that drive AI adoption success.

The Path Forward

The MIT study's findings, while concerning, should not be interpreted as evidence that AI doesn't create value. Instead, they reveal that we're applying twentieth-century measurement frameworks and incentive structures to twenty-first-century technological transformation.

Organizations that succeed with AI will be those that:

1. **Start with personal and team-level implementations** where value is more immediately apparent and measurable

2. **Focus on efficiency and capability gains** rather than immediate profit increases

3. **Adopt longer measurement timeframes** appropriate for transformational change

4. **Develop new metrics** that capture knowledge work value creation

5. **Treat AI integration as organizational change management** rather than technology deployment

Conclusion

The 95% "failure" rate highlighted by MIT's study may in fact represent 95% of organizations measuring the wrong things at the wrong time with the wrong expectations. Just as the internet's value became clear over years, not quarters, AI's

transformational impact requires patience, proper measurement frameworks, and incentives, while understanding that efficiency gains often precede profit increases.

The question isn't whether AI creates value—our research clearly shows it does, particularly at the personal and team levels. The question is whether organizations will develop the measurement sophistication needed to recognize and optimize that value creation.

We're not experiencing an AI failure; we're experiencing a measurement failure. It's time to evolve our metrics to match the transformation we are truly experiencing.

This analysis builds on research from UC Berkeley's Sutardja Center for Entrepreneurship & Technology (SCET) AI Commons initiative and responds to "The GenAI Divide: State of AI in Business 2025" by MIT's NANDA initiative.