AI Might Not Be Failing You - You Might Be Failing AI

Why So Many AI Pilots Miss the Mark and How to Measure What Actually Matters

By Justin Amendola and Alex Ruiz
In partnership with ARTXMACHINA, LLC

Section 1:

The Problem Isn't AI - It's How We're Measuring It

The "AI Failure" Narrative

Headlines have grown increasingly skeptical about AI. Many describe AI pilots as stalled, and questions about ROI are popping up everywhere. Executives are left wondering if they've put too much hype behind AI too quickly.

But hold on. This narrative doesn't match what's actually going on inside organizations. According to McKinsey's State of AI (March 2025), companies are adopting AI at an impressive rate across many functions, with marketing leading the way in the use of generative AI. Teams are experimenting, and many tools are being deployed, with investments still going strong.

The disconnect stems less from what AI does than from how organizations choose to evaluate it.

"The real AI gap isn't adoption, it's measurement. Most organizations are experimenting with AI. Almost none are instrumenting it like a business process."

Just using AI doesn't guarantee success. HubSpot's AI Trends for Marketers (2025) shows that while AI usage has doubled year over year, only a small percentage of teams can link that usage to real business outcomes.

AI often gets layered onto disjointed processes without proper tracking. When results disappoint, AI gets blamed instead of the systems around it.

Moreover, the lack of governance and readiness makes it tougher to see measurable impacts. McKinsey emphasizes that without a clear operational model, data discipline, and workflow clarity, the bottom-line impact remains just out of reach.

In short, AI isn't failing; our methods of measurement are.

The Real Issue: Misaligned Measurement

Most AI initiatives get evaluated using metrics that sound good but mean little. Consider:

- Time saved.
- Content volume.
- Drafts generated.
- Adoption rates.

These are standalone metrics that often miss the point. They don't answer the real questions:

- Is the work better?
- Is it more reliable?
- Did it empower people to focus on higher-value tasks?
- Did it change outcomes, not just outputs?

Success lies in measuring **workflow performance** rather than superficial tool capabilities.

As the **AXM Operational Metrics Framework** advises:

"Measure the workflow, not just the tool."

Focus on tracking how well processes work - think operational efficiency, quality, brand consistency, and employee experience - as signals of real transformation. These should link directly to lagging indicators such as conversion lifts, cost savings, or increased capacity.

"Everyone counts AI usage.
Almost nobody measures whether it works."

AI shines brightest when it changes how we work, enhances our judgment and creativity, and is treated like any other business process in terms of measurement.

From Tool ROI to Workflow ROI

Instead of asking, **“Did this AI tool work?”** rethink it to, **“Did this AI-enabled process outperform the old one?”**

What’s needed here is a solid testing framework that prioritizes causality over anecdotes. Organizations often struggle to justify investments based on subjective impressions like saying, “this feels faster” or “people seem to like it.”

> “The most important AI question isn’t ‘Did this tool work?’ - it’s ‘Did this AI-enabled process outperform the old one, consistently?”

To change that perspective, we should approach AI measurement like testing for product-market fit:

1. **Start small.**
2. **Measure rigorously.**
3. **Scale what has proven effective.**

This iterative mind-set distinguishes organizations that merely talk about AI from those that genuinely benefit from it.

Section 2:

How to Measure AI's Impact - and What (*and Where*) to Measure

Defining What You're Testing and Why

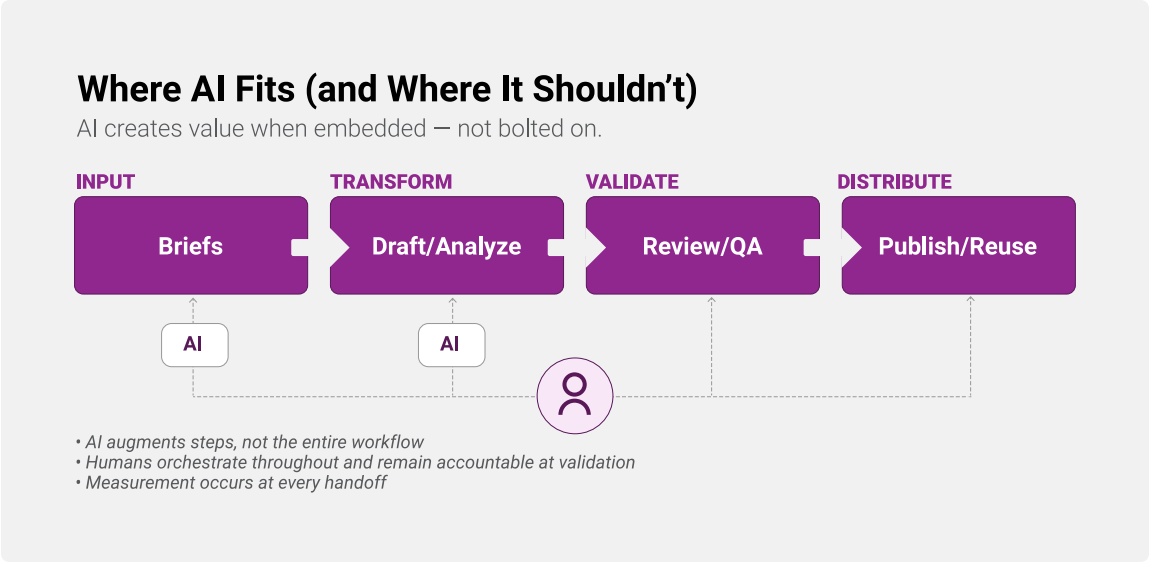
It's crucial to clearly define the focus of your tests. Many pilots tend to start with vague goals, such as "Let's try ChatGPT for marketing." Yet, without specifying whether you're focusing on content generation, imagery, data analysis, or workflow automation, you risk missing the mark.

AI encompasses various capabilities that each require distinct success criteria. AI isn't one thing. The chatbot writing your emails needs different success metrics than the algorithm routing your support tickets. Most pilots fail because they measure content generation like workflow automation.

Each of these affects workflows differently, requiring tailored measurement strategies in different contexts. For example, AI used in marketing needs a different evaluation approach than AI employed in sales or HR.

From "AI Use Case" to "AI-Enabled Workflow"

AI doesn't create value in isolation; it thrives when embedded within existing workflows. So, how can we re-think AI experimentation? By breaking down workflows into four clear stages



1. **Input Creation:** This involves gathering briefs, data, requirements, or any source material that kicks off the process.
2. **Transformation:** Here, we draft, analyze, synthesize, or generate various content pieces.
3. **Validation:** This step includes reviews, compliance checks, brand assessments, and approvals.
4. **Distribution and Reuse:** Finally, it's about publishing, localizing, versioning, and archiving the materials.

AI doesn't replace all four stages. Instead, it often enhances specific steps, particularly those that might be hidden from leadership but are frustrating for employees.

Consider this scenario: marketing campaign content development

In a traditional workflow:

- Briefs arrive in inconsistent formats.
- Drafts are created from scratch.
- Review cycles tend to run longer than they should.
- Localization happens manually, often under deadline pressure.

But with AI, this process can transform:

- Intake briefs are summarized into clean, structured formats.
- First drafts are generated using clearly defined brand prompts.
- Tone, terminology, and compliance issues are flagged automatically.
- Localized versions are produced and reviewed side by side.

If we only measure success by the amount of "content generated", we miss the bigger picture. The real value surfaces as faster turnaround times, fewer revisions, and higher approval rates on the first pass. The lesson here? Start by defining the workflow, then figure out how AI enhances it.

Start Where the Work Hurts

Many AI projects begin with executive pain points: "We need more content, faster," or "We have to do more with fewer people." While these concerns are valid, a more fruitful approach focuses on employee pain points: the repetitive, manual, or frustrating tasks that stall progress every day.

> "AI delivers the fastest ROI where work is repetitive, error-prone, and emotionally draining not where it's most visible to executives."

When organizations tackle real friction points, they uncover use cases that are simpler to measure, quicker to benchmark, and more likely to expand. According to McKinsey's AI in the Workplace (2025), the organizations that achieve the clearest ROI link AI adoption to specific workflow challenges instead of lofty strategic goal.

Guiding Principle: Start where the work hurts, not where the PowerPoint slide suggests.

Mapping Friction to Measurable Hypotheses

To truly harness employee pain points, you need to turn them into testable hypotheses. For instance, let's say a creative operations team reports spending excessive time reformatting briefs and hunting down missing information. An AI-assisted intake process can streamline those briefs, highlight missing elements, and route requests more efficiently.

Your hypothesis could be straightforward:

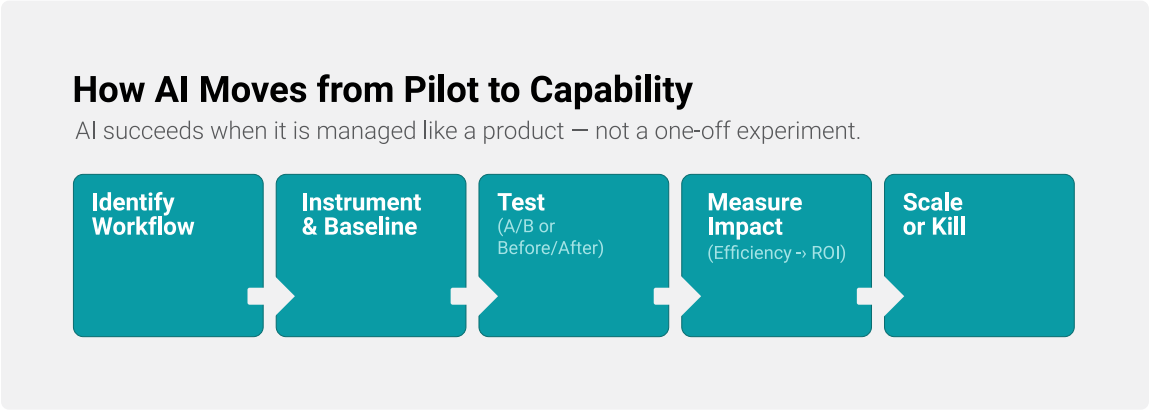
- Turnaround time will decrease.
- Error rates tied to incomplete briefs will drop.
- The "fulfilling work percentage" will increase.

Measurement should then zero in on the timing from intake to draft, the causes of revisions, and shifts in employee sentiment, converting frustration into actionable evidence.

Organizations that skip this step often find that while AI "feels helpful," it doesn't drive meaningful metrics. It's not about low adoption rates; it's about a lack of clearly defined, testable hypotheses.

Building a Measurement Framework

Once you've defined a use case, your measurement strategy should follow a structured approach:



- **Workflow Selection**: Choose distinct, repeatable processes to focus on.
- **Baseline Capture:** Gather pre-AI performance data across time, cost, and accuracy.
- **Study Design:** Use observational methods for speed and experimental designs for causality.
- **Instrumentation:** Record timestamps, AI modes, and reviewer notes.
- **Metric Hierarchy:** Prioritize metrics that cover efficiency, quality, employee experience, and overall value.
- **Governance:** Establish a review cadence and rigor thresholds.

Study Design in Practice

Take a marketing operations team tasked with testing AI-assisted email campaign development.

The baseline shows a 12-day turnaround, numerous revisions, and a high error rate. After introducing AI for first-draft generation and brand review, compare pre- and post-AI performance over identical time periods. Quick improvements may appear but keep an eye on confounding factors.

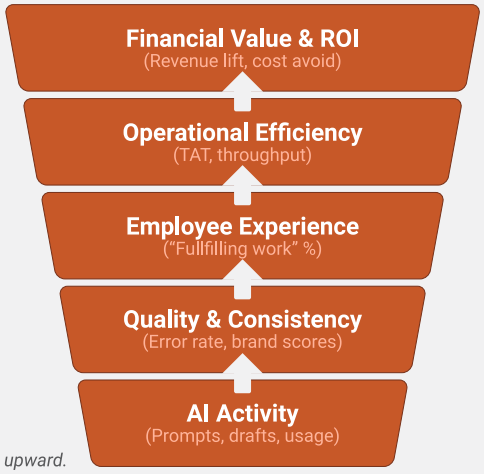
Use a true experimental design where campaigns randomly receive either AI assistance or stick with traditional methods. The performance differences like faster launch times and fewer revisions, can then be directly attributed to AI involvement. Executives care more about results than how the technology works; the key is whether the AI process outperforms the old way.

What to Measure

Effective AI measurement encompasses five critical dimensions:

From AI Activity to Business Impact

Measuring only the bottom rung guarantees disappointment.



AI value compounds upward.

1. **Operational Efficiency:** Look at turnaround times and throughput variance.
2. **Quality and Brand Consistency:** Monitor error rates and brand adherence.
3. **Employee Experience:** Gauge the "fulfilling work percentage" through surveys.
4. **Value and ROI:** Investigate net revenue opportunities created by reinvested capacity.
5. **Meta-Metrics:** Assess AI usage coverage and data completeness.

These metrics should be analyzed collectively, not in isolation. A faster turnaround rate alongside rising error rates isn't a success story. True impact emerges as quicker work without compromising quality, paired with improved employee satisfaction and operational gains.

"AI impact doesn't show up as cost savings first.
It shows up as faster workflows, fewer errors,
and better employee experience - financial impact follows."

Once leading indicators stabilize, you can then venture into financial impact modeling. AI ROI rarely starts as cost savings; it often manifests first as increased capacity—and should be measured as such.

Section 3:

Where AI Delivers Measurable Workflow ROI

Realizing AI Value in Your Own Operations

Across various industries and organizations of all sizes, AI creates enduring value when woven into existing workflows. The focus should be on how work is changing, rather than just how much output is increasing.

EXAMPLE 1:

Marketing Video Production

Scaling creative output without sacrificing quality or burning out teams.

With the rising demand for video content across different channels, marketing teams face pressure to produce more assets, in more formats, and at greater speeds without compromising brand standards or overwhelming their creative staff.

Thanks to AI, teams can reduce their cycle times and lessen manual efforts, while still keeping creative judgment firmly in human hands.

How AI fits Into the Workflow:

- Drafting creative briefs and scripts to kickstart projects.
- Generating initial concept lists and rough storyboards.
- Producing multiple formats from a master video for different channels.

Humans remain crucial for narrative flow, tone, brand voice, and final quality.

Workflow ROI Ladder Outcomes:

- **Efficiency & Speed:** Shorter times from brief to first cut, quicker versioning.
- **Quality & Consistency:** Fewer revision rounds, higher first-passage approval rates.
- **Employee Experience:** Less repetitive work, leaving more time for creative choices.

What Leaders Measure:

- Time taken from brief to usable asset.
- Scores from brand consistency audits.
- Reduction in late-stage revisions adds to creative morale.

True business value won't always show as immediate cost savings; often, it's about **capacity creation** that allows teams to launch campaigns more swiftly and efficiently.

EXAMPLE 2:

Customer Support

Enhancing Speed, Consistency, and Quality of Escalations at Scale

In customer support, organizations constantly juggle speed, cost efficiency, and customer satisfaction all while navigating regulatory compliance. Here, AI plays a role in filtering and standardizing workflows, not replacing human agents.

How AI Fits Into the Workflow:

- AI helps identify common customer inquiries and directs them to appropriate self-service responses.
- It assists in developing standardized call scripts that align with brand and policy guidelines.

Humans then tackle the complex cases that require judgment, empathy, or escalation.

Workflow ROI Ladder Outcomes:

- **Efficiency & Speed:** Reduced number of routine calls reaching agents; faster resolutions for straight-forward cases.
- **Quality & Consistency:** More uniform customer experiences; lowered compliance risks.
- **Employee Experience:** Less repetitive work leads to higher confidence and lower stress among agents.

What Leaders Measure:

- Average handle time and first-contact resolution rates.
- Trends in script adherence and compliance errors.
- Agent satisfaction and signs of escalation fatigue.

As these indicators improve, they translate into financial benefits, driving better customer satisfaction, reduced churn, and optimized support capacity.

The Measurement Through-Line

In these examples, AI succeeds for a common reason: **measurement emphasizes workflow performance rather than merely tracking AI activity**. Organizations that keep an eye on:

- How long tasks take.
- How often work needs correction.
- How consistently work meets established standards.
- How employees feel about their tasks.

We can clearly observe AI's impact moving up the **Workflow ROI Ladder** - from efficiency gains to quality improvements, enhanced employee experiences, and ultimately fostering business value.

AI isn't solely about doing more work. Its true power lies in transforming how that work gets done and measurement provides the lens through which leaders can observe that change.

From Measurement to Meaning

When AI initiatives are rooted in genuine pain points, paired with clear hypotheses, and guided by disciplined baselines, AI evolves from merely a tool to a valuable capability.

"AI doesn't fail silently.
Organizations fail to listen correctly, and measurement is how AI tells you whether it's working."

As highlighted in the Marketing AI Institute's 2025 State of AI Report, the gap between hype and real ROI lies not in capability but in how we define and measure success.

The encouraging thought is: AI might not be failing you. You could simply be asking it to solve the wrong problems or perhaps measuring the wrong things when it succeeds.