80% is Not Good Enough

How Reveal, powered by Castle Hall, has crossed the AI gap in Institutional Due Diligence

reveal

In July, a research unit at Massachusetts Institute of Technology published a study revealing a shocking 95% failure rate in corporate AI projects. The message is clear: putting AI into production is nowhere near as easy as the first "wow" demo makes it seem.

So why have 95% of AI projects to date "failed", or at least fallen short of expectations? We see two main reasons.

First, as a Harvard University team commenting on the MIT research observed:

> *What went wrong? While experimentation is good, without a connection to the true business opportunity, experiments inevitably fall short of hopes and expectations. It sounds obvious, but by framing AI as radical and disruptive, we often lose sight of the connection to the most fundamental objective of business:* ***to solve problems for customers.***

Corporate AI Projects

5%
Success Rate

95%
Failure Rate

Statistics from Massachusetts Institute of Technology

Put simply, a successful AI solution isn't going to emerge, phoenix-like, from a pile of prompts or by handing everyone CoPilot and saying "have at it." A business needs to identify a project, define it precisely, and work — hard — to test and iterate an AIpowered workflow that produces tangible results.

Second, AI plays on our emotions as business leaders. The first (minimum viable) demo in the boardroom often looks incredible. But as testing deepens, that first version turns out to be good – just not good enough. We call this the 80% plateau.

At Castle Hall Diligence, we've experienced these challenges first-hand while building

Reveal, our AI-powered operational due diligence platform. Our journey led to a clear lesson: real AI impact requires deep domain expertise, not just technology.

THE 80% PLATEAU

Understanding AI's Current Limitations

Like many organizations exploring AI, we discovered that current systems excel at producing "80% reports" — outputs that look polished but reveal serious gaps once they face real-world due diligence testing. These limitations show up in two critical ways.

01

Inconsistency

Running the same query multiple times yields different results, undermining the reliability due diligence demands. Run a tool 10 times and you'll see what we mean.

02

Accuracy

AI can aggregate vast information, but iterating models to be consistently accurate is hard. Accuracy here means both what AI includes in a report — and what it misses that should be there.

The "Golden Master" Challenge

Both inconsistency and accuracy issues boil down to a fundamental question: What does a perfect report look like?

At Ai4 in Las Vegas, a panelist put it simply:

"AI can't produce a 100% report if humans haven't defined what 100% looks like."

And of course, a tech team can't define that on its own. In operational due diligence — just like investment research, pharma, or legal reviews — the "golden master" isn't a static template. It's the accumulated knowledge from thousands of investigations, the ability to spot subtle patterns, and the instinct to know when something "doesn't quite add up."

BUILDING REVEAL

A Practitioner-Led Approach

We built **Reveal** on one core principle: technology must follow expertise — not the other way around.

We assembled a team of highly experienced ODD practitioners who work hand-in-hand with AI developers, creating a feedback loop that continuously refines Reveal's capabilities.

The Critical Role of Subject Matter Experts

Our practitioners are central to Reveal's accuracy:

Error Detection & Correction

Identifying both Type 1 errors (wrong info included) and Type 2 errors (critical info omitted), then teaching the AI how to avoid them..

Contextual Understanding

Explaining *why* something is wrong or missing, not just flagging it.

Quality Definition

Defining what a "complete, accurate" report looks like across different entity types.

Iterative Refinement

Every fix can affect something else. It's often "one step forward, 0.9 steps back." Human experts keep the model on track.

Technical Challenges **We've Overcome**

Through development and testing, we've tackled several core technical issues:

Entity Resolution

Distinguishing between multiple entities with the same name (e.g., "Apex") using sophisticated disambiguation logic.

Confirmation Bias Prevention

Ensuring the AI doesn't add "extra" info just to make reports seem better.

Source Hierarchy

Dynamic source selection tailored to entity type — drawing on premium sources like Financial Times and Bloomberg for major institutions, and alternative sources where coverage is sparse.

Misinformation Filtering

Training the system to spot and reject unreliable, AI generated "slop."

THE PATH FORWARD

Realistic AI Implementation

We built **Reveal** on one core principle: technology must follow expertise — not the other way around.

We assembled a team of highly experienced ODD practitioners who work hand-in-hand with AI developers, creating a feedback loop that continuously refines Reveal's capabilities.

And yes, humans make mistakes too — which means we're not chasing an impossible ideal. **We're raising the baseline.**

01

Start with expertise

Subject matter experts must lead, not just advise.

02

Invest in the feedback loop

Real-world refinement is essential.

03

Treat AI as a journey

AI isn't perfect — but Reveal is already operating beyond the 80% plateau. We're consistently achieving 93–95% accuracy, delivering what used to take 30 hours of research in 30 minutes or less.

CONCLUSION:

Beyond the Hype

The MIT findings echo our experience: real AI impact demands deep domain knowledge, realistic expectations, and relentless refinement.

Reveal, powered by Castle Hall, bridges the gap between current AI capabilities and the rigorous standards of institutional due diligence. It's not a magic solution that replaces human judgment. It's a powerful force multiplier — accelerating how experts identify and understand risk.

As we continue to develop Reveal, we remain committed to this practitioner-led approach. Technology serves expertise; it doesn't replace it. And in a field where missing a single risk can have major consequences, this balance isn't just smart — it's essential.