Vizorfin

# AI Evaluation Framework for Fund Operations

Separating real capability
from marketing claims

**Darwin Palma**

Lead Data Engineer @ Vizorfin
vizorfin.com

# The Three Categories of AI

Every vendor in fund operations claims to be "AI-powered" in 2026. Most are using the term loosely. This framework helps you evaluate what AI can actually do — and where the real value is.

### ✅ Category 1: Pattern Recognition

The most mature and useful. Anomaly detection, reconciliation discrepancies, data quality drift. Measurable success criteria.

### ⚡ Category 2: Natural Language Processing

Querying operational data, extracting structured data from documents, summarizing regulatory updates. A productivity tool, not a decision-maker.

### ⚠️ Category 3: Predictive Analytics

The most hyped and least proven. Constrained by data quality, sample size, and inherent unpredictability. Useful for narrow operational tasks, not strategic decisions.

### ❌ Not AI: Marketing

Vague claims, no specific task, no metrics, no fallback. If a vendor can't explain what their AI does, treat the claim with skepticism.

💡 **Key principle:** AI built on poor data infrastructure amplifies problems rather than solving them. Fix the data foundation first. Then layer AI on top.

?

# The 5 Questions to Ask

When a vendor says their platform is "AI-powered," ask these questions. If they can't answer all five clearly, treat the claim with skepticism.

### 1 What specific task does the AI perform?

**Good answer:** "Our model detects anomalies in NAV calculations by comparing against historical ranges and administrator outputs."
**Bad answer:** "It uses AI to optimize fund operations."

> **Red flag:** If the answer is vague, the AI is probably marketing.

### 2 What data was it trained on?

**Good answer:** Specific datasets, time periods, and volume. Ideally, data similar to yours.
**Bad answer:** "Proprietary data" with no further detail.

> **Red flag:** If they can't answer, the model's behavior is opaque and unpredictable.

### 3 How do you measure accuracy?

**Good answer:** Precision, recall, F1 score — real metrics from real testing.
**Bad answer:** Anecdotes about how "accurate" it is in practice.

> **Red flag:** No quantitative metrics means no accountability.

### 4 What happens when the AI is wrong?

**Good answer:** Defined fallback process, human review step, exception handling, and clear documentation of limitations.
**Bad answer:** "It's very accurate, so this isn't an issue."

> **Red flag:** Every AI system fails sometimes. How a vendor handles failure reveals their maturity.

### 5 Can you turn it off?

**Good answer:** Yes — every AI feature can be disabled, overridden, or replaced with manual review.
**Bad answer:** "It's core to the platform."

**Red flag:** If you can't control the system, you don't own your operations.

# What AI Does Not Solve

To be direct about what AI cannot do in fund operations:

| CLAIM | REALITY |
|---|---|
| **AI replaces regulatory judgment** | No. Regulatory decisions require human expertise, context, and accountability. |
| **AI guarantees compliance outcomes** | No. Compliance depends on process, controls, and interpretation — not on technology alone. |
| **AI eliminates the need for human review** | No. AI can reduce workload. It cannot replace review. |
| **AI makes a bad data pipeline good** | No. It amplifies the problem. Garbage in, garbage out applies more than ever. |
| **AI is a substitute for data lineage** | No. Without lineage, AI decisions cannot be audited or defended. |

## Where AI Actually Helps

Based on my experience building these systems, AI delivers measurable value in:

1. **Data validation at scale** — flagging anomalies humans would miss
2. **Exception triage** — prioritizing which issues need attention first
3. **Document processing** — extracting structured data from unstructured sources
4. **Operational forecasting** — predicting workload for capacity planning
5. **Query interfaces** — letting non-technical users explore data safely

That's a narrower list than most vendors would give you. It's also more honest.

✅

# Your Evaluation Checklist

Use this checklist when evaluating any "AI-powered" fund operations platform.

## Before the Demo

1. Ask for the specific AI capabilities in writing. Not brochures — documentation.
2. Ask for accuracy metrics on those capabilities. Request the methodology.
3. Ask for references from clients with similar fund structures.

## During the Demo

1. Request a live demonstration of the AI making a real decision, not a scripted one.
2. Ask to see what happens when the AI is uncertain. What's the fallback?
3. Ask to turn the AI off and see the manual workflow. It should exist.
4. Ask how the AI handles edge cases — new fund structures, unusual transactions, regulatory changes.

## After the Demo

1. Ask for a pilot with your data before signing a long-term contract.
2. Define success metrics in advance. What would "the AI works" look like, measurably?
3. Plan for failure. What's the rollback? What's the manual fallback? Who owns incidents?

> 💡 **Final principle:** The best AI vendors are the ones that are precise about what their AI does and — more importantly — what it doesn't. If a vendor claims their AI "does everything," they're either misinformed or misleading. Either way, look elsewhere.

**Important:** This framework is for educational and informational purposes only. Vizorfin is a financial technology company and is not a registered investment adviser, broker-dealer, or financial services licensee in any jurisdiction, including the United States, United Kingdom, Australia, Canada, or New Zealand. Our software is a decision-support tool and should not be the sole basis for financial or compliance decisions. Individual results will vary. Consult with qualified financial, legal, and tax professionals regarding your specific situation.