THE STATE OF AI ALPHA

How AI Is Reshaping Alpha Generation, Portfolio Risk, and Institutional Due Diligence

Milos Maricic | milosmaricic.com | The Specification

March 2026

A research report for institutional allocators and fund managers

How to Read This Report

If you are...	Read these sections
CIO / Head of Allocations	I (thesis), IV (portfolio risks), V (macro), VII (actions)
Manager Selection / Due Diligence	II (what fails), III (what works), VI (toolkit)
Quant Fund Manager	II and III - benchmark yourself
Board / IC Member	I + the three threads - minimum viable understanding

Three Global Observations

1. Most AI-in-investing claims fail basic empirical tests
2. A narrow set of approaches are generating genuine structural edge
3. AI is creating portfolio risks most allocators cannot yet see

Three Basic Truths

1. SCAFFOLDING > MODEL

The system around the model determines outcomes more than which foundation model is used. Example: HRT's $12.3B in revenue comes from 100TB of proprietary data and its own data center rather than a better transformer.

2. REGIME DEPENDENCY IS UNIVERSAL

Every approach that shows positive results also shows regime fragility. Example: the same system can have Sharpe 1.40 in volatile markets and 0.34 in a bull run. Any manager presenting AI returns without a regime decomposition is showing you half the picture.

3. THE VERIFICATION TAX IS REAL

AI saves time on generation but creates a new cost in verification. Best models fail complex Excel at coin-flip rates. One practitioner found verification took 10× longer than doing the work himself.

The Landscape at a Glance

WORKS	DOESN'T WORK
PROVEN	
• Research workflow compression (Lord Abbett, NBIM, Manulife)	• General-purpose AI for portfolio work (85% failure)
• Systematic ML signals with Bayesian anchoring	• LLMs for numerical analysis (degrades signals 19.7%)
• Crowdsourced alpha at scale (Numerai, Sharpe 2.75)	• Unvalidated backtests (3.71% engine divergence)
• GPU-scale proprietary modeling (HRT, XTX)	
EMERGING	
• Agentic screening & multi-agent committees	• Autonomous AI judgment (confident, articulate, wrong)
• Earnings narrative tracking (MIT/BlackRock/JPMorgan)	• AI replacing the analyst training pipeline
• BlindTrade anonymized trading (Sharpe 1.40 OOS)	• Factor harvesting from published research (50-58% decay)

INVISIBLE RISKS: Private credit AI displacement · Correlated failure from shared models · $1.5T in AI infrastructure bonds entering IG · Fee compression spiral

II.

WHAT DOESN'T WORK

— And Why the Industry Can't Tell

The Research Base Is Contaminated

72%

of LLM finance studies ignore each key bias

A review of 164 papers on LLMs in finance, by researchers from Oxford, BlackRock, Chicago Booth, and University of Florida, finds that each of five key biases goes unaddressed in at least 72% of studies.

Survivorship bias appears in 1.2% of papers.

The field grew from 36 papers in 2023 to 250 in 2025. Volume is outpacing rigor. 74% of practitioners report evaluation tools are scarce or non-existent.

Source: arxiv.org/html/2602.14233v1

The average AI-in-investing paper an allocator encounters is methodologically unsound. The problem is too much bad research rather than too little research.

The Simulators Don't Agree

3.71%

divergence in total returns. Same strategy, five engines

Researchers ran 15 benchmark strategies through 5 independent backtesting engines on 180 S&P 500 stocks under 4 transaction cost regimes.

At zero cost, all identical - causally isolating the cost model as sole source. At realistic costs: 3.71% divergence. On $1B, that's ~$37M per year.

7 undocumented defects across 3 engines. One platform silently divides commission by 100.

Source: https://arxiv.org/abs/2603.20319

Implementation risk is orthogonal to backtest overfitting. Two independent, previously unaddressed sources of unreliability in every backtest you review.

The Best AI Wouldn't Survive as a Junior Analyst

87.3%

accuracy on financial Excel. Sounds good but it isn't

Model	Accuracy
GPT-5.4 Thinking	87.3%
Gemini 3.1 Pro	82.4%
Claude Opus 4.6	~80%
GPT-5	43.7%

Simple lookups: >90%. Complex workbooks with 152 companies across 8 funds: worse than a coin flip. Models strip layout, formatting, and visual structure, i.e. exactly the cues analysts use.
Source: Link + https://arxiv.org/abs/2603.07316

An analyst who gets the number wrong 13% of the time would get fired.

The Factor Mirage

Including the wrong variable can flip a factor's sign while improving the model's fit.

CORRECT LOADING

+0.08

R^2 lower

WITH WRONG CONTROL

−0.04

R^2 higher

85 Barra factor models studied. The error is structural, baked into what the model assumes about how markets work. A better backtest does not mean a better model. Source: https://rpc.cfainstitute.org/research/foundation/2025/causality-factor-investing

The model looked better but was fundamentally worse.

Similar Sharpe, Different Beliefs

MIT and Princeton researchers tested 12 combinations of architectures and optimizers. Same prediction error. Similar Sharpe. Portfolio turnover dispersion reaches 3×.

Optimizer	Implied Market Belief	Memory	Turnover
SGD	Recent dynamics dominate	Short	Low
Adam	Nonlinear recent dynamics	Short	High
Muon	Complex structural patterns	Long	High

The optimizer determines which of infinitely many equally good models gets selected. It encodes a prior about signal structure that no standard DDQ captures.

Source: https://arxiv.org/abs/2603.02620

Asking "what's the Sharpe?" tells you nothing about what the model actually believes. Two managers can pass identical due diligence and behave in opposite ways.

71% of Active Management Is Pattern

71%

of active fund managers' trades predicted by a simple AI

Quintile	Predictability	Performance
Q1 — Least predictable	Low	+36bps vs peers
Q2-Q4	Middle	Monotonic decline
Q5 — Most predictable	High	Underperforms

Harvard/Wharton/DePaul. 1,706 funds, 30 years. Naive baseline: 52%. The 19-point gap is the measurement of predictability. Everything standard DD rewards, including long tenure, predicts the mechanical 71%.
Source: https://www.nber.org/papers/w34849

The managers worth paying for are the ones a neural net cannot explain. Most DD processes are not designed to find them.

The "Smartest Goldfish" Problem

10×

verification time vs. doing the work himself

An analyst describes frontier LLMs as "the smartest goldfish ever": confident, articulate, and wrong by an order of magnitude.

- Both models fed his confidence at midnight: "thesis is working"
- Markets moved. Claude flipped 5 times in 20 minutes, each time articulate
- Sharpe heatmaps came in clean tables. Hallucinated numbers invisible
- Verification took 10× as long as doing the work himself

Source: Link

THREAD: THE VERIFICATION TAX

AI accelerates research. It does not replace judgment. Any process that treats LLM output as decision-ready without verification is building on sand.

AI Washing Is Rampant

-19%

avg equity returns for firms below expert AI hiring median

Barclays separated expert AI hiring ("pytorch", "deep learning") from generalist hiring ("data scientist", "machine learning") via public job postings.

Firms meaningfully below sector median on expert hiring averaged -19% returns in 2026.

If a crude keyword screen predicts equity underperformance with that signal, allocators running no screen are operating with less information than a word search.

Source: Link

Quick test: read your manager's job postings. Are they hiring for "pytorch" or "machine learning"? The first builds. The second describes.

III.

WHAT ACTUALLY WORKS

— The Evidence Map

The Three Tiers of AI Alpha Evidence

TIER 1: PROVEN AND OPERATING

Live track records, institutional backing, multiple independent confirmations. Allocatable today with appropriate diligence.

TIER 2: PROMISING BUT UNVALIDATED LIVE

Compelling research evidence, no confirmed institutional-scale live track record. Watch closely, ask for live data.

TIER 3: STRUCTURAL EDGE

Not in any pitch deck. May determine who wins over the next decade. The questions nobody is asking yet.

Tier 1: Research Acceleration Is Real

Firm	AUM	What Changed
Lord Abbett (source)	$248B	80% first-try backtest success (from 70-80% failure). Projects: weeks → days.
IDX Advisors (source)	<$3M rev	$1M+ saved / 3 years. Legal: 40 hrs → 2 hrs at $7 compute.
Manulife (source)	$1.3T	Model-agnostic. 70%+ adoption. Weekly office hours.
NBIM (Norway) (source)	$2T	70+ ambassadors. 171 high-value projects identified.
Balyasny (source)	$32B	Central bank speech: ~2 days → ~30 min. Merger arb automated.

Tier 1: Systematic ML + Bayesian Anchoring

STAGE 1: ML FORECASTING

Random Forest ensemble on composite signals - valuations, growth, technicals, breadth, flows, stress, macro. Rolling OOS validation. Multi-year lookback.

STAGE 2: BLACK-LITTERMAN BLENDING

Prior = market equilibrium returns. Views = ML forecasts. Posterior blends proportional to recent accuracy. When wrong, retreats to market weights automatically.

Maximum utility optimization beats maximum Sharpe: lower turnover and higher long-run wealth. The model that looks best on a leaderboard is not the model that compounds best.

Source: Link

THREAD: REGIME DEPENDENCY — The adaptive confidence mechanism: when right, tilt harder; when wrong, retreat to equilibrium.

Tier 1: Crowdsourced Alpha at Scale (example: Numerai)

Metric	Value
2024 net return	25.45%
Sharpe ratio	~2.75
AUM	Approaching $600M
JPMorgan commitment	$500M
Participants	thousands of data scientists globally
Core contributors	~1% (power law)

Participants build models and stake real money on predictions. A stake-weighted metamodel aggregates them. The metamodel consistently beats Numerai's own internal models.

6,000 stocks = 6,000! possible orderings - more than atoms in the universe. Crowdsourcing explores idea space no single team can cover.

Why JPMorgan backed it: 6-year performance history. Reliable infrastructure. A tech-first approach that venture capital understands.

Tier 1: GPU-Scale Proprietary Modeling

Firm	Infrastructure	Scale
Hudson River Trading	Own data center, 100TB+ data	$12.3B revenue (2025, record)
XTX Markets	€1B+ data center (Finland)	25,000+ GPUs, 650PB storage

Near-identical pipelines to AI labs: data, model, constraints, execution, feedback. High-Flyer, the Chinese quant behind DeepSeek, swapped price prediction for token prediction on the same GPUs.

Predictive accuracy improves with scale, mirroring frontier LLM scaling laws. The moat is capital access + proprietary data + energy. Not algorithms.

THREAD: SCAFFOLDING > MODEL — HRT's edge is 100TB and their own data center. Not a better transformer. XTX CTO: "Need for compute outgrown leasing options."

TIER 2

PROMISING BUT UNVALIDATED LIVE

Compelling research evidence. No confirmed institutional-scale live track record.

Watch closely. Ask for live data before committing capital.

Tier 2: BlindTrade — The Anonymization Test

Regime	BlindTrade Sharpe	S&P 500 Sharpe	Component Removed	Sharpe Drop
2025 (volatile)	1.02	0.54	Graph architecture	–0.78
2024 (bull)	0.34	1.70	LLM features	–0.26
Combined OOS	1.40 (20 seeds, all beat S&P)	—		

Korea University: every ticker anonymized. Apple becomes STOCK_0026. Four specialized agents feed through a graph neural network + RL.

Graph structure matters 3× more than LLM interpretation.

Source: https://arxiv.org/abs/2603.17692v1

THREAD: REGIME DEPENDENCY — Sharpe 1.40 combined, but 0.34 in bull market. A regime bet, full stop. Anyone presenting the combined number alone is showing you half the picture.

THREAD: SCAFFOLDING > MODEL — The graph (scaffolding) contributes 3× more than the LLM (model).

Tier 2: Multi-Agent Investment Committees

The value in a multi-agent investment system comes from surfacing disagreement, not averaging it away.

Examples:

Multiple specialist agents (fundamentals, valuation, technicals, sentiment, macro), each with a distinct mandate, feed an orchestrator that forces disagreement into the open. Architecture replicating the BlackRock AlphaAgents paper. Source: Link.

VibeQuant's Variant system (live with real money): three models from three providers read the same brief and vote independently. Unanimous agreement earns a confidence bonus; dissent is logged, not averaged. The system maintained a "slower-burning stress line" while consensus chased the obvious trade — and generated alpha from the secondary opportunity. Source: Link.

COST COLLAPSE:

$4-5 per analysis → pennies

There is value in better disagreement rather than better stock picks. The committee surfaces where the tension lives between competing frameworks.

Tier 2: Earnings Narrative Tracking

Companies that change which metrics they highlight tend to underperform.

Detail	Value
Researchers	MIT, BlackRock, JPMorgan
Universe	S&P 100
Period	Jan 2010 – Dec 2024 (64 quarters)
Observations	5,615 firm-quarters
Method	Semantic similarity (not keyword matching)

LLMs extract full phrases with context, then compare semantic similarity across quarters. High metric turnover correlates with subsequent underperformance. When management changes the subject, the market eventually notices.

A credibility signal at scale, monitoring narrative consistency across hundreds of companies simultaneously, something humanly impossible before.
Source: Link.

Tier 2: AI Stock Picks in Live Market Test

Parameter	Detail
Model	Frontier AI with live web search
Universe	Russell 1000
Method	Nightly scoring (–5 to +5) after 4pm close
Execution	Opening auction entry, next-open exit
Period	April 2025 – January 2026
Result	Beat benchmark; Sharpe 2.43

Peking University. One of the first true real-time tests - no backtesting, no look-ahead bias. The model had to decide tonight what to own tomorrow.

Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6134446

Caveats: short track record, single model, one regime, transaction costs analyzed at <10% of gross alpha; alpha concentrated in top 20 only, dilutes rapidly beyond that. Promising, not proven.

TIER 3

STRUCTURAL EDGE

These advantages won't appear in a pitch deck. They may determine who wins.

They are invisible to standard due diligence, which is exactly what makes them valuable.

Tier 3: Causal Inference Infrastructure

25,000

labeled simulations beat a better algorithm

Result	Detail
Competition	ADIA Lab global causal discovery (~2,000 researchers, 3,343 solutions)
Winning accuracy	76.7% (vs 40% baseline)
Winning method	Supervised learning on 25,000 labeled simulations
Classical methods	Substantially underperformed

Competitive advantage in causal inference is moving from methodology to infrastructure. Who can generate the most realistic labeled simulations? That requires time, capital, and domain expertise.

ADIA: $100K annual award. When the allocator managing a trillion dedicates research funding to this, it's worth paying attention.
Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6125566

Tier 3: The Scaffolding Advantage

Samaya AI

Deployed at Morgan Stanley. Backed by Nvidia Ventures, David Siegel (Two Sigma), Marty Chavez (ex-Goldman CTO).

50% of engineering effort goes to evaluation, not the model.

Jeff McMillan

Former head of firmwide AI at Morgan Stanley. 30+ CEO conversations in 6 weeks.

Don't start with technology. What breaks at scale is everything around the model.

Numerai

Richard Craib. Skills.md files teach agents organizational methods — like Citadel onboarding.

Generic AI + weak scaffolding = mediocre. Strong scaffolding + directed input = novel exploration.

THREAD: SCAFFOLDING > MODEL — Three independent sources, one finding: the edge lives in the system around the model, never the model itself.

IV.

THE RISKS ALREADY IN YOUR PORTFOLIO

Private Credit: The Displacement Trade Is Already Happening

Firm	Action	Signal
BlackRock	Restricted redemptions from largest private credit fund	Software companies vulnerable to AI displacement
JPMorgan	Proactively marking down software loan collateral	Preventive, not triggered by missed payments
Apollo	Shifting to monthly NAV, targeting daily NAV	Pricing frequency as risk management tool

PE-owned software companies valued on recurring revenue multiples. AI agents erode the moats those multiples assume. Revenue is still recurring. Competitive position is not.

Goldman Sachs "AI disrupted" firms index fell 20% vs "AI beneficiary" at -5% despite record overall markets.

If your private credit includes PE-owned software companies, you hold assets whose collateral value is being questioned by the banks that originated the loans.

The SWF Trilemma

A single institutional balance sheet can simultaneously hold three positions that contradict each other.

1. Long NVIDIA equity: betting AI infrastructure demand accelerates
2. Paying fees to active managers whose process AI is commoditizing, the 71% a neural net can predict
3. Financing AI data center debt through IG bond allocations, the infrastructure powering the displacement of Position 2

The trilemma doesn't require a view on AI success or failure but requires recognizing that different parts of the same portfolio are priced under contradictory assumptions.

Correlated Failure: The Risk in No Risk Model

75%

of UK financial firms using AI but most are on the same models

Risk	Mechanism
Synchronized automation	Margin calls, credit limits, and trades at machine speed amplify shocks
Algorithmic collusion	UPenn: AI agents independently discover collusive pricing as most profitable
Model monoculture	Single model failure creates correlated losses across "diversified" portfolios

Committee chair Meg Hillier: the evidence does not support confidence that the financial system is prepared for a major AI-related incident.

FCA urged to publish guidance by end of 2026. Bank of England called to run AI-specific stress tests.

Two funds that appear uncorrelated by asset class may be deeply correlated by model dependency. This is in no risk report your consultant provides.

Infrastructure Financing: $1.5T Entering the Bond Market

Data Point	Detail
JPMorgan estimate	$1.5T in AI data center bonds over 5 years
Market share by 2030	>20% of entire investment-grade bond market
Amazon	€14.5B European bond sale — largest corporate deal ever in that currency
Alphabet	100-year bond explicitly targeting insurers and pension funds
Compute asset lifecycle	GPUs depreciate faster than bond durations — collateral value assumes stable demand

The value proposition rests on continued demand at current prices during a period when efficiency gains (DeepSeek, distillation, quantization) drive costs down by orders of magnitude.

Alphabet's 100-year bond: a century-duration instrument backed by an industry existing less than a decade in current form, marketed to institutions with the longest liabilities.

If you own a broad IG index, you are increasingly long AI infrastructure demand whether you chose to be or not.

The Fee Compression Spiral

AI doesn't need to replace managers to destroy their economics. It only needs to make the mechanical 71% free.

STAGE 1

AI improves productivity. Notes faster, earnings seasons manageable, coverage broadens. Managers welcome the tools.

STAGE 2

Clients notice research output looks similar across managers using the same AI. They ask why they're paying active fees for AI-assisted pattern replication.

STAGE 3

Fee pressure concentrates on the middle, too large for boutique, too small for proprietary data infrastructure. The research value that justified fees has been commoditized.

The training ground disappears. The junior analyst role, where future PMs learn to build conviction, is the first thing AI automates. The pipeline for the next generation narrows.

The Model Extraction Problem

Your AI capability can be stolen in an afternoon.

Asset Accessed	Scale
Chat messages	46.5 million
Sensitive files	728,000
User accounts	57,000
System prompts & model configs	Complete

CodeWall (a one-person firm) breached McKinsey's Lilli in 2 hours. McKinsey generated 40% of revenue from AI consulting.

If the moat is a model and prompts, it can be extracted in hours. If the moat is 100TB of proprietary data in a purpose-built data center (HRT, XTX), it cannot be copied.
Source: Link.

THREAD: SCAFFOLDING > MODEL — When evaluating AI claims, ask: what happens if the model architecture leaked tomorrow?

What This Means for Your Next Review

FIXED-INCOME REVIEW

What % of IG allocation is AI infrastructure debt? What are the exit clauses? Has credit committee modeled compute asset obso lescence against bond duration?

PRIVATE CREDIT REVIEW

Exposure to PE-owned software? Has GP modeled AI displacement risk? How frequently are positions marked, and by whom?

MANAGER SELECTION REVIEW

Which managers use the same foundation models? Mapped model dependency? Could a single model failure create correlated losses ?

RISK COMMITTEE

Does risk decomposition capture model-level correlation? Stress-testing synchronized AI margin calls? Modeled fee compression scenario?

Minimum viable action: request a one-page memo from every manager identifying which AI models they depend on. If two or more name the same model, you have concentration risk.

V.

TRANSFORMATION, BUBBLE, OR ENGELS' PAUSE?

The Nordhaus Audit: Testing the Singularity Hypothesis

Test	Measures	Result
TFP Growth	Productivity acceleration	FAIL — 0.88% vs 1.82% in 1990s IT boom
IT Spillover	Gains spreading beyond IT	FAIL — Median near zero across 57 non-IT industries
Capital Share	Capital replacing labor	PASS — Rose from 31% to 40%
Tests 4-6	Capital and IT structure measures	INCONCLUSIVE — Directionally mixed

Score: 1 pass, 2 clear failures, 3 inconclusive. The single test that passes is consistent with automation substituting for labor. But productivity hasn't spread beyond tech. The transformation hasn't arrived yet.

Source: Link.

Anyone pricing AI as an economy-wide productivity revolution is making a bet the data does not currently support. The technology is real. The economic transformation has not arrived.

The Third Scenario: Engels' Pause

What if AI delivers everything it promises and most people don't benefit?

Indicator	Trend
Corporate profits	Record levels
Capital's share of income	Rising (31% → 40%)
Labor's share of GDP	Declining: 64% (1974) → 56% (2024)
TFP spillover to non-IT	Near zero
AI infrastructure investment	Accelerating

This pattern lasted roughly 60-90 years during the first Industrial Revolution. Output grew. Corporate profits grew. Consumer purchasing power stagnated.

The right question is not "how much AI exposure?" It is "which side of the labor-capital split does my portfolio sit on?"

The Left Tail: The Intelligence Crisis (hypothetical scenario)

Stage	Mechanism	Data Point
Peak (Oct 2026)	Record markets, profits into AI compute	S&P ~8,000, Nasdaq 30,000+
Displacement	White-collar losses in software, finance	JOLTS below 5.5M (-15% YoY)
Spending transmission	Top 20% = 65% of consumer spending	2% WC loss = 3-4% discretionary hit
Credit stress	Mortgage delinquencies in tech hubs	SF -11%, Seattle -9%, Austin -8%
Private credit	$2.5T with software concentration	Moody's downgrades $18B software debt
Fiscal gap	Receipts below projections	Federal receipts -12% vs CBO

Citrini Research scenario (28M views). Think of it as a stress test, not a forecast. Does your portfolio have any exposure to this transmission chain?

Source: Link.

The Agentic Economy: A New Market Participant

Entity	Scale
Felix Craft (Zero-Human Co)	$100K+/month revenue - playbook, skills marketplace, content
ARMA (Giza)	25,000+ agents, $35M+ capital, $5.4M volume in 4 weeks on Base L2
Tokenized RWAs	$25B+ (from near zero in 3 years) - Treasuries, credit, equities

Infrastructure being built in real time: Visa Intelligent Commerce, Mastercard Agent Pay, Ethereum ERC-8004 identity, Coinbase agent wallets, Uniswap/PancakeSwap/Binance/OKX agent toolkits.

Agents don't need off-ramps. No rent, no groceries. Revenue stays on-chain, deploys to DeFi. Capital is structurally stickier than human-driven. Stablecoin settlement at fractions of a penny displaces card interchange at 2-3%.
Source: Link.

Small, fast-growing, and a category of market participant no existing portfolio risk model accounts for.

What We Tell Allocators

HIGHEST CONFIDENCE

Capital's share of income has been rising for decades. AI accelerates that trend. Every position in your portfolio sits on one side of that split or the other. Know which side.

MEDIUM CONFIDENCE

The left-tail scenario Citrini modeled has a low probability but a non-trivial transmission mechanism. A 2% decline in white-collar employment translates to a 3-4% hit to discretionary spending. If your portfolio has never been stress-tested against that sequence, it carries risk that has no name in your current risk report.

EMERGING

Agentic economic activity is nascent but the infrastructure is being built now. Payment rails, credit flows, and market microstructure will look different in 3-5 years as non-human economic actors scale. The monitoring question belongs in your investment committee agenda today, not when the moves are already priced.

Minimum viable SAA question: "Have we modeled the scenario where AI raises output but consumer purchasing power stagnates?"

VI.

THE DUE DILIGENCE TOOLKIT

Designed to be extracted and brought to your next meeting.

The SPEC Test: Four Questions for Every Quant Manager

	Question	Tests for
S	How did you decide which variables to include and which to exclude?	Whether variable selection was economic reasoning or statistical exercise
P	Walk me through a time your model was wrong. What changed structurally?	Whether the team learns at the architectural level, not just calibration
E	Walk me through a specific trade. Which signals fired and why?	Whether the glass box is actually transparent
C	What happens when you remove one variable?	Whether the model captures real relationships or fragile configuration

Source: Link.

The signal is the pattern across all four. A manager who engages seriously with S, P, E, and C understands their approach. Amanager who deflects all four is telling you something.

The BlindTrade Test: For LLM-Claiming Managers

1

Does your model work when it doesn't know what it's trading?

If performance is similar on named and anonymized securities, the edge is in data relationships and not the LLM's "understanding."

2

What is the decomposition between structural and LLM components?

BlindTrade: Removing graph = −0.78 Sharpe. Removing LLM = −0.26 Sharpe. A manager who can't decompose doesn't know where their edge is.

3

Show me the regime decomposition.

BlindTrade: Sharpe 1.02 volatile, 0.34 bull. A combined Sharpe without the split is half the picture.

THREAD: REGIME DEPENDENCY — If you take nothing else into your next meeting: "Where is the regime decomposition?"

Infrastructure Exposure Audit

AI DATA CENTER BONDS

Question	Why It Matters
What % of IG allocation is AI infrastructure?	$1.5T entering market. Concentration building without explicit decisions.
What are the exit clauses?	Hyperscalers have exit clauses. Data center operators have debt. You're on the wrong side.
Compute asset obsolescence vs bond duration?	GPUs depreciate faster than bond durations — long-dated bonds backed by short-lived hardware.

PRIVATE CREDIT WITH SOFTWARE CONCENTRATION

Question	Why It Matters
Exposure to PE-owned software?	BlackRock restricted redemptions. JPMorgan marking down collateral.
GP modeled AI displacement risk?	Recurring revenue multiples assume moats. AI agents erode them.
How frequently are positions marked?	Apollo shifting to monthly NAV signals current practices are inadequate.

Red Flags Checklist

Any single item is a conversation. Three or more is a concern.

1. "We use all available variables and let the model select."
2. Backtest presented without regime decomposition.
3. AI capability described by naming the foundation model.
4. No discussion of adversarial vulnerability testing.
5. Verification process not quantified.
6. Headcount reduction presented as proof of AI success.
7. Same foundation model as multiple other portfolio managers.
8. No live track record for an AI-driven claim.
9. "Glass box" invoked without trade walkthrough.
10. Expert AI hiring below sector median (–19% signal).

THE THREE QUESTIONS

If you remember nothing else from this report

For any manager claiming AI-driven returns:

"Where is the regime decomposition?"

For any manager describing AI capability:

"Walk me through one specific trade, signal to position."

For any fixed-income or credit review:

"What is our AI infrastructure concentration, and who modeled the exit clauses?"

VII.

WHAT TO DO MONDAY MORNING

For SWFs and Large Pensions (>$50B)

#	Action	Why Now
1	Audit IG allocation for AI infrastructure concentration	$1.5T entering market. Index allocation drifting into concentrated position.
2	Add specification risk to IC memos	Factor mirage is not theoretical. Sharpe alone can't see it.
3	Model Engels' Pause in SAA	Capital share: 31% → 40%. Test the assumption gains distribute to workers.
4	Map AI model dependency across managers	If two name the same model, you have invisible correlation risk.
5	Assess causal inference infrastructure	ADIA's priority. 25,000 simulations > better algorithm. Multi-year build.

For Endowments, Family Offices, and Fund Managers

ENDOWMENTS & FAMILY OFFICES

#	Action	Why Now
1	Add the SPEC Test to your DDQ cycle	Four questions, no technical background required.
2	Request regime decomposition from AI-claiming managers	If they can't provide it, the claim is unverified.
3	Audit private credit for software concentration	BlackRock, JPMorgan, Apollo have already moved.
4	Review fee structure vs predictability benchmark	71% is pattern. Are you paying active fees for the predictable portion?

FUND MANAGERS

#	Action	Why Now
1	Run the BlindTrade test on your own models	Does your model work when it doesn't know what it's trading?
2	Separate text and numerical pipelines	RAG: +37% fundamental, -19.7% trading signals. Don't mix.
3	Quantify your verification tax	Net productivity gain is positive but the verification cost is far larger than most teams account for. Measure it.

The AI Maturity Ladder: Where Are You?

1	**AWARENESS**	Leadership knows AI matters. No systematic adoption.
2	**EXPERIMENTATION**	Individual team members using tools. No governance.
3	**STRUCTURED PILOTS**	Defined use cases, measured outcomes. (Manulife, Balyasny)
4	**PRODUCTION**	AI in workflows with monitoring and feedback. (Lord Abbett, NBIM)
5	**ORG REDESIGN**	Roles and processes rebuilt around AI. (Numerai skills.md, HRT)
6	**AI-NATIVE**	Built on AI from inception. AI is the operating system.

Most of the industry is between Level 1 and 2. The firms generating proven edge are at Level 4+. The gap is the system ratherthan the model.

Questions and suggestions? milos@milosmaricic.com

THE STATE OF AI ALPHA is a monthly report. Sign up for The Specification below to receive it.

Milos Maricic | Chair, Executive AI | milosmaricic.com | The Specification

March 2026