dakota marketplace

DAKOTA INSIGHTS · THE AI STRATEGY MISTAKE

You're Debating the Wrong Thing. It's a Data Problem, Not a Model Problem.

Every investment firm is racing to pick an AI model. Almost none of them are asking the question that actually determines whether their AI strategy succeeds or fails.

30+	69K+	$3.6T	300+
YEARS OF PRACTITIONER DATA	PRIVATE MARKET TRANSACTIONS	IN TRANSACTION VALUE TRACKED	DAILY DEAL SIGNALS

GC

Gui Costin
Founder & CEO, Dakota

DAKOTA MARKETPLACE · PRIVATE MARKETS INTELLIGENCE

12 Min Read

THE MISDIAGNOSIS

Every Firm Is Racing to Pick a Model. The Model Is Already a Commodity.

The board conversation sounds the same at every large investment firm right now: *"We need an AI strategy. Do we use OpenAI? Do we go with Microsoft Copilot? Should we build on Claude?"* These are reasonable questions. They are also, largely, the wrong ones. The model is not where your AI strategy will be won or lost.

OpenAI, Anthropic, Google, and Microsoft are engaged in one of the most intense technology races in history. They are spending tens of billions of dollars to out-train each other. The result — which most investment firms haven't fully absorbed yet — is that the underlying AI models are becoming commodities at a remarkable pace.

GPT-4 was a revelation eighteen months ago. Today it's table stakes. Claude, Gemini, and a growing field of open-source models are competitive at nearly every task. This commoditization will continue. The intelligence layer — the raw reasoning capability of the model — is not where differentiation will live.

So if the model is a commodity, what isn't? What is the actual differentiator that will determine whether your AI strategy produces a genuine competitive advantage — or just an expensive experiment?

> *The model is the engine. But an engine without fuel doesn't move. In private markets, the fuel is proprietary LP and GP data — and that data is not a commodity. It never will be.*

THE REAL PROBLEM

An AI Model Is Only as Good as the Data It Has Access To.

A powerful language model connected to weak, incomplete, or inaccurate data produces confidently wrong answers. In private markets — where every LP relationship, every allocation decision, every fundraising insight has real financial consequence — that's not an acceptable outcome. It's a liability.

Consider what a portfolio manager or fundraising professional actually needs to know. Who are the active allocators in infrastructure right now? Which public pensions have capacity for a first-time manager? What is this GP's actual fund performance history, and who are the right LPs to introduce them to? These questions require data that is accurate, current, relationship-level, and private-markets-specific. No general-purpose model has it. No internet scraper can build it. It has to be built the hard way — by practitioners, over years, through direct market participation. That is an entirely different class of asset than the model itself.

THE STACK

Where the Actual Value Lives in an AI Strategy.

To understand why the model selection debate misses the point, it helps to look at how an AI system actually works in practice — and which layers of that system are defensible.

LAYER 4 · The AI Model — COMMODITY ↓

OpenAI, Claude, Gemini, Copilot — the raw reasoning engine. Extraordinarily capable. Rapidly commoditizing.

LAYER 3 · The Connection Protocol (MCP) — INFRASTRUCTURE

The Model Context Protocol — the standard that lets AI agents connect to live data sources dynamically, in real time, in natural language.

LAYER 2 · The Data — Dakota Marketplace — THE MOAT ↑

30+ years of practitioner-built LP/GP intelligence. Relationship-level accuracy. Daily refreshed. Irreplaceable by any scraper or aggregator.

LAYER 1 · Your Workflow — THE GOAL

Fundraising, LP coverage, due diligence, deal sourcing, executive search — the actual work your firm needs to accelerate.

The model — Layer 4 — is where most firms are focused. The data — Layer 2 — is where the actual competitive moat lives. The connection between them, the MCP server at Layer 3, is what makes the data natively usable by any AI model your firm chooses to run.

Here is what this means practically: it doesn't matter whether your firm runs Claude or GPT-4 or Gemini. If your AI can't touch accurate, relationship-level LP and GP data, your strategy is built on a foundation that will produce mediocre results regardless of how sophisticated the model is.

WHAT THE EXPERTS SAY

The Industry Has Been Saying This for Years. Now It's Decisive.

> *In the last decade, the biggest shift in AI was a shift to deep learning. I think it's quite possible that in this decade the biggest shift will be to data-centric AI. With the maturity of today's neural network architectures, for a lot of practical applications the bottleneck will be whether we can efficiently get the data we need.*

Andrew Ng

Co-Founder of Google Brain · Former Chief Scientist at Baidu · Founder of Coursera & Landing AI — IEEE Spectrum

> *Competitive advantage in AI goes not so much to those with data but those with a data engine: iterated data acquisition, re-training, evaluation, deployment, telemetry. And whoever can spin it fastest.*

Andrej Karpathy

Former Director of AI at Tesla · Founding Member of OpenAI — X (formerly Twitter)

> *Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.*

Clive Humby

British Mathematician & Data Science Pioneer · Architect of Tesco's Clubcard — ANA, 2006

The thesis is not new. What's new is that natural-language AI has finally made the data layer the binding constraint — and exposed which firms have spent the last three decades building it, and which have spent the last three quarters scrambling to look like they did.

THE COMPARISON

What "Model First" vs. "Data First" Actually Looks Like.

QUESTION	MODEL-FIRST APPROACH	DATA-FIRST APPROACH
Who are the right LP prospects for this fund?	Generic answer from public sources. No relationship context.	Curated list from 30+ years of practitioner data with contact-level detail.
What is this GP's real track record?	Press release data and aggregator summaries. Often incomplete.	Verified fund-level performance data from Dakota's transaction database.
Which pensions have capacity right now?	Stale public filings. No signal on current mandate or allocation pace.	Daily-refreshed commitment data with allocation pace and contact intelligence.
Who is the right person to call?	LinkedIn scrape. No accuracy guarantee. No relationship mapping.	Dakota's relationship graph — practitioner-verified contacts, updated continuously.
Competitive differentiation over time	None. Every firm running the same model gets the same quality answers.	Compounding. The data moat grows as Dakota's coverage deepens.

THE TAKEAWAY

The right side of this table is not theoretical. It is how Dakota's MCP-connected dataset actually performs when an AI agent queries it in plain English. The left side is what every firm using a generic LLM on top of public or scraped data is shipping today — and calling an AI strategy.

THE SECOND PROBLEM NO ONE IS TALKING ABOUT

Even If You Solve the Data Problem, You Still Have to Solve Completeness.

Here is the part of the conversation that almost no one in the AI strategy discussion is having: even if you accept that data is the real issue — not the model — you have to reckon with the fact that most LP and GP databases in the market are not just imperfect. They are massively, structurally incomplete.

The vast majority of databases in this space were built to serve one channel, in one region, for one type of institution. Typically: RIAs and family offices in the United States. Full stop. Everything else — other countries, other institution types, other channels — is either missing, thinly covered, or updated so infrequently that the data is functionally stale.

That is a fatal problem for any serious AI strategy in private markets. An AI model is only as complete as the universe it can see. If your data layer doesn't include entire continents, entire institutional categories, or entire channels of capital, your AI doesn't know what it doesn't know — and it will produce answers with the same confident tone whether it has complete coverage or a 40% sample.

> *An AI that can only see part of the market isn't an intelligence advantage. It's a liability — because it doesn't know what it's missing, and neither do you.*

What the Completeness Gap Costs You in Practice

Imagine running an AI-powered LP targeting exercise for a new infrastructure fund. Your model is excellent. But your database only covers U.S. public pensions, U.S. family offices, and a handful of U.K. endowments.

You've just excluded: every sovereign wealth fund in the Middle East and Asia, every Canadian pension fund (among the world's most sophisticated infrastructure allocators), every Nordic institution (some of the most active LP communities globally), every German and Swiss bank running private markets mandates, every Latin American family office, and the entire global insurance company universe.

Your AI didn't fail. Your data failed your AI. The model produced the best answer it could with an incomplete universe — and you never knew what you missed.

THE COMPLETENESS STANDARD

What True Global Coverage Means — And Why It's Hard to Build.

True completeness in private markets data means coverage across every major geography, every institutional category, and every relevant channel. That is not a checkbox exercise. It took decades to build, and it cannot be replicated by scraping public sources or aggregating third-party feeds.

THE COMPLETE LP UNIVERSE — DAKOTA'S COVERAGE STANDARD

- »» Public pension funds — all 50 U.S. states, U.K., Europe
- »» Sovereign wealth funds — globally, across the Middle East, Asia, Nordics, and beyond
- »» Family offices — U.S., Latin America, Europe, Middle East, Asia-Pacific
- »» Endowments and foundations — globally
- »» Insurance companies — U.S. and international
- »» Banks and private wealth platforms — German, Nordic, Swiss, U.K. institutions
- »» Broker-dealers — across all U.S. channels, retail and institutional
- »» Funds of funds — global

GEOGRAPHIC REACH

United States	**United Kingdom**	**Germany & DACH**	**Nordic Region**
Middle East & GCC	**Latin America**	**Canada**	**Asia-Pacific**
Continental Europe	**Rest of World**		

Geographic completeness is only half the equation. Within each of those jurisdictions and institution types, every single relevant person must be tracked. That means the CIO, the portfolio manager running private equity, the analyst conducting due diligence, and the investment committee member with ultimate allocation authority. Not just the institution — every decision-maker inside it.

And that contact data has to be accurate in real time. The private markets industry experiences roughly 25% annual turnover in investment personnel. That means in any given year, one in four of the contacts in your database has changed roles, left for a different institution, or retired. A database that is not refreshed continuously is not just outdated — it is actively misleading your AI and your team. The wrong contact at the right institution is worse than no contact at all. It wastes relationship capital and signals a lack of seriousness to the LP community.

THE GP SIDE OF THE EQUATION

LP Data Is Only Half the Picture. The GP Universe Matters Just as Much.

Every discussion of private markets data that focuses exclusively on the LP side is leaving an enormous amount of value on the table. The GP dataset — managers, their funds, fund performance, portfolio companies, and the private company market broadly — is the other half of the intelligence equation, and it is just as difficult to build correctly.

01 All GPs & All Funds

Complete coverage of general partners across every strategy — buyout, growth equity, venture, real assets, credit, infrastructure. Every fund they have ever raised, with vintage, size, and strategy detail.

02 Fund Performance

Verified performance data for those funds — IRR, MOIC, DPI, TVPI — at the fund level. Not press release data. Practitioner-sourced, verified returns that an LP due diligence team can actually act on.

03 Portfolio & Private Companies

Every portfolio company across the GP universe, plus non-portfolio private companies — with C-suite executives tracked at each. A complete map of the private company executive universe that no public source can replicate.

04 2,000+ Transactions Monthly

Over two thousand private company market transactions tracked every single month — M&A, growth investments, buyouts, recapitalizations. The most current signal available on where capital is moving in private markets.

This is what separates a data infrastructure business from a data vendor. A vendor sells you a slice of the market — a list of names, a database of fund sizes, a directory of contacts. An infrastructure business gives your AI a complete, continuously refreshed picture of the entire private markets ecosystem: every LP, every GP, every fund, every portfolio company, every executive, every transaction. That is the universe your AI needs to see to give you answers you can actually act on.

Most databases in this market are not even close. They were built for one use case, in one region, for one type of buyer. They are snapshots, not engines. And a snapshot — however accurate in the moment it was taken — is the wrong foundation for an AI strategy that needs to operate in real time, across a global market that never stops moving.

THE THIRD PROBLEM — THE MOST UNDERESTIMATED

Complete, Global Data Still Fails You If It Isn't Updated Every Day.

There is a third dimension to this problem that almost no one in the AI strategy conversation is accounting for, and it may be the most operationally decisive of all. Private markets data is not static. It is not even close to static. It is one of the most dynamic data environments in all of institutional finance — updating not quarterly, not weekly, but every single day at extraordinary volume and velocity.

A database that was complete and accurate six months ago is a different database today. In private markets, that gap is not a nuisance. It is the difference between actionable intelligence and expensive noise. If your AI is reasoning from stale data, it is not just giving you outdated answers — it is generating confident, well-structured, completely wrong answers.

THE PRIVATE MARKETS DATA VELOCITY PROBLEM

300+ DAILY	New private company transactions tracked every single day — M&A, growth investments, buyouts, recapitalizations across the private markets universe. *Miss a day, miss the signal on where capital is moving before anyone else knows.*
100+ WEEKLY	New investments posted by public pension funds — commitment disclosures that reveal who is actively deploying capital, into what strategies, at what pace. *A week-old pension commitment feed is already behind in a fast-moving fundraise.*
Continuous WEEKLY–MONTHLY	Family offices and RIAs making new investments, updating mandates, and shifting allocation priorities — individually smaller in size but enormous in aggregate signal. *The most opaque segment of the LP universe and the fastest-moving. Real-time tracking is not optional.*
13F EVERY 45 DAYS	SEC 13F filings, ETF and fund-of-funds disclosures — updated on a 45-day cycle, giving a near-real-time window into the public-facing holdings of institutional investors. *Institutional mandates shift every quarter. A database refreshed annually misses three full cycles.*
25% CONTINUOUSLY	Annual contact-level turnover across the investment industry. One in four decision-makers — the CIOs, due diligence leads, portfolio managers — changes roles, firms, or exits the market every year. *A static contact database is not just stale — it actively destroys relationship capital when your team calls the wrong person.*

THE THREE-PART STANDARD

Stack the Signals Together. The Picture Becomes Clear.

Stack these signals together and the picture becomes clear: in a single quarter, the private markets landscape sees tens of thousands of new transactions, thousands of new LP commitment disclosures, continuous mandate shifts across the family office and RIA universe, multiple rounds of 13F updates, and significant turnover in the very contacts your team needs to reach. The database that served you well at the start of the year looks materially different by the end of it.

> *The question is not whether your data was accurate when you licensed it. The question is whether it is accurate right now, at the moment your AI is using it to generate an answer you are about to act on.*

This is the full three-part standard that any investment firm serious about AI must hold its data provider to. Not one of these conditions, not two — all three, simultaneously, without compromise:

I

The Right Data Layer

Proprietary, practitioner-built LP and GP intelligence — not scraped, not aggregated, not reverse-engineered from public filings. Built by people who were in the market.

II

Complete Global Coverage

Every institution type, every jurisdiction, every channel — from U.S. public pensions to Nordic sovereign wealth funds to Latin American family offices to Middle Eastern banks. No blind spots.

III

Dynamic, Daily Refresh

300+ transactions daily. 100+ pension commitments weekly. Continuous contact verification. 13F updates every 45 days. The data that powers your AI has to move as fast as the market does.

Any database that cannot meet all three conditions is not a foundation for an AI strategy. It is a liability dressed up as one. And right now, in the private markets data landscape, there is exactly one provider built to that standard.

But there is a fourth condition — one that most firms won't discover until they've already made a very expensive mistake.

THE FOURTH STANDARD — THE MOST DANGEROUS MISTAKE

Commingling Similar Datasets Will Destroy Your AI Strategy.

Investment firms building AI strategies on private markets data almost universally start with the same instinct: *"We'll use the best of each provider. PitchBook for GPs, Dakota for LPs, a third source for RIAs, a fourth for sovereign wealth funds."* It sounds reasonable. It sounds like diversification. It is, in practice, one of the most damaging things you can do to the AI system you're building.

Here is the fundamental problem. When you feed an AI model two different records of the same entity — the same RIA, the same family office, the same public pension, the same sovereign wealth fund — from two different data providers, you are not giving it more information. You are giving it a contradiction. And AI systems handle contradictions very badly.

A SINGLE RIA · TWO DATA SOURCES · WHAT THE AI SEES

DATA PROVIDER A	
AUM	**$2.4B**
Head of Alts	**Jennifer Walsh**
PE Allocation	**12%**
Open to New GPs	**Yes**
Last Updated	**Q2 2024**

DATA PROVIDER B	
AUM	**$3.1B**
Head of Alts	**Michael Torres**
PE Allocation	**8%**
Open to New GPs	**No**
Last Updated	**Q4 2023**

WHAT THE AI HAS TO DO WITH THE CONFLICT

- »» **Pick one arbitrarily** — the AI selects one record with no reliable basis for the choice. 50% chance of using the wrong data, presented with full confidence.
- »» **Average the conflict** — the AI blends the two records ($2.75B AUM, ~10% PE allocation) — producing a number accurate to neither source and reflecting no real-world entity.
- »» **Hedge the answer** — the AI surfaces the contradiction and refuses to answer definitively. Your team gets "sources conflict" where they needed action.
- »» **Silently choose wrong** — the most dangerous outcome. The AI produces a clean, confident answer using stale or incorrect data — and no one knows. Your team calls Michael Torres. He left 18 months ago.

THE COMMINGLING PROBLEM, BY ENTITY TYPE

It's Not an Edge Case.
It Happens Everywhere.

This is not a theoretical concern. It is what happens every single time an AI system encounters the same entity described differently by two sources. Multiply that across 5,000 RIAs, 3,000 family offices, 500 public pension funds, 200 sovereign wealth funds, and 50,000 GP fund records — and you have not built an AI strategy. You have built a sophisticated confusion engine that produces authoritative-sounding answers based on a data foundation that is internally at war with itself.

> *Feeding an AI two conflicting records of the same entity doesn't give it more information. It gives it an impossible problem to solve — and it will solve it incorrectly, confidently, and invisibly.*

RIAS

Two providers will have different AUM figures, different contact lists, different allocation data — often from different reporting periods. The AI cannot know which is current. The 25% annual contact turnover means at least one source is almost certainly wrong on personnel.

FAMILY OFFICES

Among the most opaque entities in private markets. Two providers tracking the same family office will have gathered their data through entirely different methods — one from public filings, one from direct outreach — producing fundamentally different pictures of the same institution.

PUBLIC PENSIONS

Commitment timing, allocation pacing, and investment committee contacts shift constantly. A provider pulling from public disclosures and a provider with direct sourcing will show different states of the same fund — at different points in time — with no flag to the AI that they disagree.

SOVEREIGN WEALTH FUNDS

Global SWFs are notoriously opaque. Coverage varies dramatically by provider. Commingling two incomplete pictures of an SWF doesn't produce a complete picture — it produces two partial, potentially contradictory records that the AI will treat as a single confused entity.

GP & FUND RECORDS

Fund performance figures — IRR, MOIC, vintage — will differ across providers based on calculation methodology, reporting date, and data sourcing. An AI reasoning across two sets of performance numbers for the same fund will produce analysis accurate to neither.

THE ONLY CORRECT ANSWER

There Is No Technical Solution. There Is Only a Single Source of Truth.

There is no technical solution to conflicting source data. Deduplication logic, entity resolution algorithms, and confidence scoring can paper over the problem — but they cannot solve it. The AI still has to choose between two records that describe different realities for the same entity.

The only answer is a single source of truth. **Pick one data provider.** Evaluate it against the full four-part standard — proprietary and practitioner-built, complete across all geographies and channels, dynamically refreshed daily, and a single coherent universe with no internal contradictions. Then connect it to your AI via MCP and build from there.

The firms that try to stitch together multiple data providers will spend months discovering why it doesn't work. The firms that start with a single source of truth will be compounding their advantage while everyone else is debugging their data pipeline.

WHY THE DATA DECISION IS DIFFERENT

This is what makes the data selection decision so consequential — and so different from every other vendor decision an investment firm makes. You are not choosing a tool. You are choosing the *epistemological foundation* of your AI system.

Everything the AI knows, believes, and concludes about the private markets universe will flow from that choice. Get it right, and the system compounds. Get it wrong, and no amount of prompt engineering, model selection, or workflow optimization will save you.

THE COMPLETE STANDARD — ALL FOUR, NON-NEGOTIABLE

I

Proprietary & Practitioner-Built

Not scraped. Not aggregated. Built by people who were in the market, making calls, verifying relationships.

II

Complete Global Coverage

Every institution type, every jurisdiction, every channel. No blind spots in the LP or GP universe.

III

Dynamic Daily Refresh

300+ transactions daily. 100+ pension commitments weekly. Continuous contact verification. Moving as fast as the market.

IV

A Single Source of Truth

One coherent, internally consistent universe. No commingled records, no conflicting data, no invisible contradictions.

THE FIFTH STANDARD — THE ONE THAT MAKES EVERYTHING ELSE MATTER

The Goal Is Continuous Improvement. Standards I–IV Make It Possible.

Here is the thing about the first four standards that most firms miss until it is too late. They are not ends in themselves. They are **preconditions** for the only outcome that actually matters in an AI strategy: a model that gets smarter every single day, compounds its intelligence over time, and becomes an increasingly powerful competitive weapon the longer you run it.

That is the goal. Not a model that answers questions adequately. Not a chatbot that surfaces names from a database. A continuously improving intelligence system that knows more about the private markets landscape this quarter than it did last quarter — because it has been trained on better data, more data, and more current data.

That goal — continuous training, compounding intelligence — is what the first four standards are protecting. And here is the brutal truth: **if you fail any one of them, the fifth doesn't just underperform. It collapses entirely.**

HOW EACH FAILED STANDARD DESTROYS CONTINUOUS TRAINING

STANDARD I VIOLATED · DATA ISN'T PRACTITIONER-BUILT

The model gets faster at being wrong.

You are training on scraped approximations and public-source guesses. Every training cycle makes it more confident in information that was never accurate to begin with.

↓

STANDARD II VIOLATED · COVERAGE IS INCOMPLETE

The model develops blind spots it doesn't know it has.

Trained on a partial universe, it learns to reason about the 40% of the market you covered. Every cycle deepens its ignorance of what it can't see.

↓

STANDARD III VIOLATED · DATA ISN'T DYNAMICALLY REFRESHED

The model becomes an expert on a market that no longer exists.

Each training cycle compounds the staleness. Your AI gets sophisticated about a six-month-old version of reality.

↓

STANDARD IV VIOLATED · DATA IS COMMINGLED

The model embeds internal contradictions more deeply with every cycle.

It learns that the same RIA has two different AUM figures, two different contact lists. It gets better at navigating conflicts it should never have encountered.

↓

STANDARD V · THE OUTCOME

Continuous improvement becomes impossible.

You cannot train your way to intelligence on a foundation of inaccurate, incomplete, stale, conflicting data. The model compounds its errors with increasing sophistication and decreasing transparency. You've built a system that gets confidently worse over time.

THE COMPOUNDING INTELLIGENCE FLYWHEEL

What Continuous Training on Clean Data Actually Builds.

> *Training a model on bad data doesn't produce a bad model that you can see is bad. It produces a sophisticated model that has learned, deeply and durably, to be wrong — and that presents its errors with the same fluency and confidence as its correct answers.*

The inverse — what happens when all four standards are met — is the actual prize. A model trained continuously on proprietary, practitioner-built data, covering the entire global LP and GP universe, refreshed daily, from a single coherent source of truth, does something remarkable over time. It compounds.

Every new pension commitment tracked, every contact updated, every transaction recorded, every fund performance figure verified — feeds back into a model that gets incrementally smarter. The gap between that model and a competitor's model widens every quarter. The institutional knowledge it accumulates becomes genuinely irreplaceable. The employees using it stop thinking of it as a tool and start thinking of it as a colleague — one that has processed more market intelligence than any human team could absorb in a decade.

THE COMPLETE FIVE-STANDARD FRAMEWORK — ALL NON-NEGOTIABLE

1. **Clean, complete, daily-refreshed data enters the system**
   →
2. **Model trains on accurate, current, internally consistent intelligence**
   →
3. **Employees get better answers, faster — and act on them with confidence**
   →
4. **Better decisions generate better outcomes — which informs even smarter future queries**
   ↑
5. **The model compounds. The gap over competitors widens. The advantage becomes structural.**

This is what every investment firm says they want when they talk about an AI strategy. And this is exactly what becomes unreachable the moment they cut corners on any of the four foundational standards. The fifth standard is not a separate requirement. It is the reason the other four exist.

THE MCP THESIS

The MCP Server Is the Bridge — And It Changes Everything.

The Model Context Protocol is an open standard, created by Anthropic, that defines how AI agents connect to external data in real time. It is the architectural shift that makes data-first AI strategies practical at enterprise scale.

Before MCP, connecting an AI model to a proprietary data source required custom API integrations — expensive, brittle, and slow to maintain. Every new model required a new integration. Every schema change could break the workflow. Firms with large data assets were effectively locked out of the AI-native world.

MCP changes that. When a data provider exposes its intelligence through an MCP server, any AI model with an authorized connection can query it dynamically — in natural language, in real time, without a developer in the loop. The AI figures out what to ask and how to ask it. The data provider doesn't need to know which model the firm is using. The firm doesn't need to rebuild their integration when they switch models.

WHAT THIS MEANS FOR YOUR FIRM

Dakota's MCP server means your AI — whatever model you choose to run — can query our LP and GP database directly. In natural language. In real time. With 30+ years of practitioner-built accuracy behind every answer.

You're not choosing Dakota over OpenAI or Claude. You're choosing Dakota *alongside* whichever model you run. We become the data layer. The model becomes interchangeable. And your AI strategy finally has a foundation that can actually deliver differentiated results.

That is what it means to be data-first.

30+	69K+	$3.6T	DAILY
YEARS OF PRACTITIONER INTELLIGENCE	PRIVATE MARKET TRANSACTIONS	IN TRANSACTION VALUE	REFRESH ACROSS EVERY LAYER

THE STRATEGIC QUESTIONS

Three Questions Every Investment Firm Should Be Asking Instead.

01 Can our AI actually access our most important data — or is it reasoning from what's publicly available?

Public data produces generic answers. Proprietary, practitioner-built data produces genuine edge. If your AI can't touch your best data, you don't have an AI strategy — you have an expensive search engine.

02 Is the data our AI runs on accurate enough to act on in private markets?

Confidently wrong is worse than uncertain. If your AI is producing LP prospect lists from scraped or aggregated data, your team will burn relationships acting on bad intelligence. Accuracy is not a nice-to-have — it is the whole game.

03 Is our data layer model-agnostic — or will we have to rebuild if we change models?

The model you choose today will not be the model you run in three years. The firms that win will be the ones whose data infrastructure works with any model, not the ones who bet everything on one provider.

THE BOTTOM LINE

The Firms That Get This Right Will Look Obvious in Retrospect.

Every decade or so, there is a technology transition that looks like a tooling decision from the outside and turns out to be a fundamental restructuring of competitive advantage. The move from on-premise to cloud looked like a hosting choice. It was actually a complete reimagining of how competitive moats are built in software.

The move from model-centric AI to data-centric AI is that kind of transition. The firms that recognize it early — the ones that stop asking "which model?" and start asking "what data?" — will compound advantages that become very difficult to close.

THE CONCLUSION

The Model Is a Commodity. The Data Is the Moat.

In private markets, the most valuable data in the world is the LP and GP intelligence that took 30 years to build by practitioners who were actually in the market, making calls, verifying relationships, tracking commitments. That data cannot be scraped. It cannot be replicated quickly. And it is now natively accessible to any AI model through Dakota's MCP server.

THE BOTTOM LINE

The model is a commodity. The data is the moat. The firms that understand that distinction, and build their AI strategy accordingly, will own the next decade of private markets intelligence. Everyone else will be running the same model as their competitors, on the same public data, wondering why their AI strategy isn't delivering differentiated results.

Ready to Build Your AI Strategy on the Right Foundation?

Dakota's MCP server connects 30+ years of practitioner-built LP and GP intelligence to whatever AI model your firm runs. **Model-agnostic. Daily refreshed. Built for private markets.**

30+	69K+	$3.6T	MCP
YEARS OF PRACTITIONER INTELLIGENCE	PRIVATE MARKET TRANSACTIONS	IN TRANSACTION VALUE	MODEL-AGNOSTIC CONNECTION

Connect 30+ years of practitioner intelligence to whatever AI model your firm runs.

› CONNECT WITH DAKOTA