

Robo Agents For All

A Decentralized Data Marketplace for LLMs & GWMs

RAFA Research, San Francisco, 10/2025

“Combined with emerging information markets, crypto anarchy will create a liquid market for any and all material which can be put into words and pictures.”

– Timothy C. May, [The Crypto Anarchist manifesto](#)

The Current State of Proprietary Data and AI	3
Vast Amounts of Data Are Locked in Silos and Underutilized	3
Proprietary Data as a Strategic Commercial Asset	5
Implications of Data for AI Development	7
The Economic Opportunity: Data Marketplaces and the Future Data Economy	9
How RAFA accelerates LLMs and GWMs	13
What gets monetized? The Insight Product Classes	15
Market potential	15
Why a decentralized marketplace unlocks this TAM and shifts market share from centralized marketplaces	17
Segment opportunities & Examples	18
Pricing & packaging patterns that work	19
Key Takeaways	19

The Current State of Proprietary Data and AI

Proprietary Data is The Critical Asset that Will Shape AI's Next Decade

Proprietary data resides in the vast troves of information locked within companies and organizations. It provides them a source of operational and competitive advantages. But, it cannot be directly monetized and can only be utilized internally. Proprietary data is increasingly recognized as a critical commercial asset for the coming decade. A recent observation by a Microsoft VP noted that *over 95% of the world's commercially valuable data remains locked in silos*, rather than freely accessible on the internet. This implies that only a single-digit percentage of data (mostly public web data) is currently fueling today's cutting edge AI models, while the overwhelming majority is sequestered behind firewalls and organizational boundaries. We believe that proprietary data will remain a key asset for businesses and a driving force for AI development in the next ten years. We examine how most data is underutilized and "trapped" in silos, the strategic value of unlocking this data, the implications for AI progress, and the economic opportunity emerging through data marketplaces. The goal is to illustrate why **unlocking proprietary data – through sharing or trading – could unleash enormous value**.

Vast Amounts of Data Are Locked in Silos and Underutilized

The fact that ~95% of valuable data is locked away has profound implications for AI development. Today's most advanced AI systems are largely trained on public datasets – for example, open internet text, publicly available images, etc. That public data represents only the tip of the iceberg of all data out there. As a result, there is a limit to what current AI models can learn, especially about specialized or proprietary contexts. If the status quo persists (data remaining siloed), we may see a bifurcation in AI development: a handful of big companies with massive proprietary datasets (e.g. proprietary search logs, private user data) will continue to dominate certain AI capabilities, while smaller players struggle to keep up. In other

words, AI advances could be unevenly distributed, with those controlling unique data vaults maintaining an edge. Multiple studies across industries reveal that the **vast majority of data collected is never actually used or analyzed**, reinforcing the notion that valuable data is “locked away.” For example, an engineering/construction industry analysis found “*over 95% of all data¹ captured goes unused*” in that sector. Similarly in healthcare, as much as *95%² of hospital-generated data goes unused* for any analytic or clinical value. These striking figures are not outliers – they align with broader enterprise trends. Forrester Research estimates that **60% to 73% of all enterprise data³ is never analyzed for insights**. In fact, McKinsey reported that only about *1% of all data collected* has ever been analyzed or used in any fashion. In other words, well over 90% of valuable data remains siloed and untapped across the world.

One consequence of siloed, underutilized data is a massive economic opportunity cost. IDC estimates that **data silos cost the global economy around \$3.1 trillion every year⁴ in lost productivity and revenue**. When data is fragmented across departments or held in proprietary silos, organizations struggle with incomplete information, duplicated efforts, and missed insights – all of which hurt the bottom line. Gartner likewise found that data inefficiencies (often a result of siloed and inaccessible data) cost companies on average \$12.9 million annually⁵. These

1

<https://www.forconstructionpros.com/business/press-release/21031884/fmi-corp-study-95-of-all-data-captured-goes-unused-in-the-ec-industry#:~:text=Some%20key%20findings%20include%3A>

2

<https://www.healthcarefinancenews.com/news/most-data-generated-not-used-its-fullest-potential#:~:text=CHICAGO%20,according%20to%20healthcare%20IT%20experts>

3

<https://fatman.fi/blog-mckinsey-says-only-1-of-generated-data-has-ever-been-analyzed-where-do-you-stand/#:~:text=%2A%20Machine.off%2C%20nifty>

4

<https://www.dataversity.net/articles/breaking-down-data-silos-for-digital-transformation-success/#:~:text=In%20the%20race%20to%20become.one%20of%20the%20top%20five>

5

<https://nektar.ai/5-ways-siloed-data-is-burning-your-revenue/#:~:text=This%20is%20the%20case%20of.siloed%20data>

numbers highlight that *the value is there in the data, but it remains largely “trapped”*. Unlocking even a fraction of siloed data could recover billions or trillions in value through better decisions, new products, and efficiency gains.

It’s clear that organizations are aware of this problem. An overwhelming 90% of businesses cite data silos as a major obstacle to growth⁵. The persistence of siloed data is often due to historical organizational structures, separate IT systems, and cultural resistance to sharing information. Within companies, different teams frequently store data in isolated systems that “don’t talk to each other,” meaning *most of the data an organization has can’t be easily combined for insights*⁵. And on a larger scale, proprietary datasets remain confined within individual organizations due to privacy, competition, or lack of mechanisms for exchange. The end result is the same: **only a tiny sliver of the world’s data (likely <5%) is readily accessible for training AI models or informing decisions**, while everything else remains locked down.

Proprietary Data as a Strategic Commercial Asset

Despite the challenges in utilizing it, proprietary data is enormously valuable – many call it the “**new oil**” or “**new gold**” of the modern economy. In fact, data has been described as *“the new raw material of business, an economic input almost on a par with capital and labor.”*⁶ Companies and governments today collect more information than ever, and that information – from customer behavior logs to supply chain metrics – has tangible monetary value. As one foresight analysis put it, *“data becomes a currency with a value and a price”* in the digital age. And like any currency or asset, data can be bought, sold, or traded – which is why experts argue it “**requires a marketplace where anything that is information is represented.**”

Crucially, proprietary datasets are **unique and often high-quality**, making them a potential competitive moat for companies that own them. A rival can copy your

6

<https://www.futureagenda.org/foresights/value-of-data/#:~:text=It%20has%20never%20been%20easier,T his%20is%20gold%20dust%20for>

software features, but they cannot easily copy years' worth of accumulated user data or domain-specific records that your organization holds. This exclusivity is why *proprietary data is seen as a new economic moat* in many industries. For example, large tech firms have amassed proprietary datasets (search query data, social network data, e-commerce purchase data, etc.) that give them a formidable advantage in developing AI models and targeted services. These data troves are *not available to the public or competitors*, making them a key commercial asset.

Industry observers note that while state-of-the-art AI models (LLMs) are often trained on public web data, the private data companies⁷ hold can provide “critical nuance” and specialized knowledge that generic models lack.

In domains like healthcare, finance, or enterprise software, insiders acknowledge that *domain-specific proprietary data is indispensable for high-performing AI*. As one healthcare AI firm's clinical officer explained, only a tool built with real **behavioral health-specific data** could capture the subtle nuances of therapy conversations – a general-purpose model trained on broad internet text simply wouldn't have those insights. In short, **unique data from specialized domains remains crucial** for achieving superior AI results in those fields.

It's worth noting that there is an ongoing debate about data “moats.” Some venture analysts argue that the emergence of foundation models (trained on web-scale data) has *narrowed the gap* between companies with proprietary data and those without. Modern AI techniques like transfer learning and synthetic data can, to some extent, compensate for not having a gigantic private dataset. However, even these analysts concede that **in high-stakes or highly specialized applications, proprietary data can still tip the scales⁷**. Moreover, having exclusive data is one of the few levers companies can pull to differentiate their AI in an era when algorithms rapidly diffuse. Thus, looking ahead a decade, we conclude that proprietary data will *continue to be viewed as a key asset. It is the primary asset that businesses will monetize, guard, and leverage* for competitive advantage. Corporate leaders are already moving in this direction: in one survey, 95% of organizations said they were introducing new KPIs and strategies to measure and maximize the value of

7

<https://review.insignia.vc/2025/03/10/ai-moat/#:~:text=This%20isn%E2%80%99t%20to%20say%20data,s pecific%20training>

their private data⁸. Data is becoming a core part of business value, often appearing on balance sheets or as a driver of M&A decisions.

Implications of Data Exchange for AI Development

The fact that ~95% of valuable data is locked away has profound implications for AI development. **Today's most advanced AI systems are largely trained on public datasets** – for example, open internet text, publicly available images, etc. That public data represents only the tip of the iceberg of all data out there. As a result, there is a *limit to what current AI models can learn*, especially about specialized or proprietary contexts. If the status quo persists (data remaining siloed), we may see a bifurcation in AI development: a handful of big companies with massive proprietary datasets (e.g. proprietary search logs, private user data) will continue to dominate certain AI capabilities, while smaller players struggle to keep up. In other words, **AI advances could be unevenly distributed**, with those controlling unique data vaults maintaining an edge.

On the other hand, if mechanisms emerge to **unlock and share proprietary data in a privacy-preserving, secure way, the next decade of AI progress could accelerate dramatically**. Imagine if even a portion of the currently siloed 95% of data becomes available (through marketplaces or collaboration) – this would vastly expand the training materials and knowledge base for AI models. *Generative AI's quality and accuracy would leap forward* in many domains, since models could learn from rich troves of enterprise data (customer interactions, scientific experiments, manufacturing data, etc.) that are currently off-limits. Many high-value AI applications – like medical diagnosis, industrial optimization, or enterprise analytics – require exactly those kinds of proprietary datasets to train on. Unlocking them would enable AI to tackle problems and industries that have so far been under-served due to data scarcity. **Analysts emphasize that foundation models pre-trained on public data still need domain-specific fine-tuning to perform expertly in fields like healthcare or finance⁷**, so access to proprietary data will be the key to truly expert AI systems in those areas.

There is also an important **democratization angle**. If data remains siloed, AI power stays concentrated with data-rich organizations. But if a *data marketplace* or sharing framework allows data-poor companies (or researchers) to access datasets they don't have, it levels the playing field. **More actors could train competitive models**, spurring innovation across the board. We are already seeing early signs of this: forward-looking enterprises are beginning to augment their internal data with external data sources to improve their analytics and AI – a trend that is “propelling the growth” of data marketplaces⁹. Essentially, organizations realize they can't rely only on their own info; by tapping into outside datasets (from partners or marketplace vendors), they can greatly enhance model performance and insights. Over the next decade, this *networked data sharing* could become a norm, fueling a virtuous cycle: better data access leads to better AI models, which create more economic value, incentivizing further data sharing.

Of course, accessing proprietary data at scale raises concerns – privacy, security, and intellectual property need to be managed carefully. But the development of privacy-preserving technologies (like federated learning, differential privacy) and clear data licensing frameworks are making it possible to utilize sensitive data without exposing or misusing it. If those governance pieces fall into place, **the unlocked 95% of data could herald a golden age for AI**, where innovations aren't limited by data hunger. The bottom line is that *proprietary data will remain a linchpin for AI – either as a guarded treasure driving asymmetric advantage, or as a traded commodity fueling broad-based AI advancement*. In both scenarios, its importance only grows as we push against the frontiers of what AI can do.

The Economic Opportunity: Data Marketplaces and the Future Data Economy

For investors and businesses, the recognition of proprietary data's value is giving rise to a new **data economy**. In particular, **data marketplaces** are emerging as platforms to realize the economic potential of siloed data. A data marketplace is a platform where organizations can **buy, sell, or exchange data products** – essentially turning data into a traded asset. The vision is to create a secure, compliant environment where companies can *monetize their data assets* and others can acquire valuable datasets to power their models, all while preserving privacy and ownership rights. Major industry players are getting behind this idea. For instance, Snowflake's Data Marketplace allows companies to share live datasets cloud-to-cloud; AWS Data Exchange, Oracle Data Marketplace, and dedicated startups like Dawex and Narrative are all building infrastructure for data trading. These efforts reflect a growing market opportunity that didn't exist at scale a few years ago.

The total addressable market (TAM) for data marketplaces and data monetization is already measured in the billions and growing rapidly. Recent market research valued the global data **monetization** market (the broader practice of turning data into economic value) at about \$3.5 billion in 2024, projected to reach \$12–13 billion by 2032¹⁰. More specifically, the market for **data marketplace platforms** – the tools and services enabling data exchange – is expected to grow at ~25% CAGR in the coming years, **reaching around \$5–8 billion by 2030**¹¹. This high growth rate speaks to surging demand for solutions that can unlock siloed information. Importantly, this TAM spans virtually every sector: *financial services, healthcare, consumer goods, media, technology, the public sector, manufacturing*, and more are all beginning to participate in data marketplace activity¹¹. No single industry dominates the data exchange market, but some are leading the way. The **Banking/Financial Services (BFSI) sector currently holds the largest share** of the data monetization market – roughly 18–20% of the pie – thanks to its abundant streams of customer and transaction data¹¹. Finance firms have been early to recognize that their vast datasets on customer behavior, risk, and markets can be productized or utilized for AI models.

¹⁰

<https://www.fortunebusinessinsights.com/data-monetization-market-106480#:~:text=The%20global%20data%20monetization%20market,in%202024>

¹¹

<https://www.grandviewresearch.com/press-release/global-data-marketplace-market#:~:text=Data%20Marketplace%20Platform%20Market%20To,from%202025%20to%202030>

Close behind are sectors like **healthcare**, which is rich in data (electronic health records, clinical trial data) and now actively looking to capitalize on it, and **retail/consumer goods**, where e-commerce and loyalty data are highly valuable for analytics¹⁰. **Technology and media companies**, already data-savvy, are also significant contributors, both as providers and consumers in data marketplaces¹¹. Even the **public sector** is joining in, with government agencies exploring data sharing to spur innovation (open data initiatives, data collaboratives) – often with an eye on maintaining privacy while allowing commercial use of certain datasets.

Benefits: The economic benefits of breaking down data silos and enabling data exchange are multi-fold. First, **companies can create new revenue streams by monetizing data they historically kept internally**. Instead of letting data just sit in archives, organizations can package it (in anonymized, aggregated form) and sell it to others who need it. This was virtually impossible in the past due to legal and technical barriers, but data marketplace platforms now provide the trust and security layers to make it feasible. A cited example is SafeGraph (a provider of foot-traffic location data) which significantly expanded its customer base by listing datasets on Snowflake's marketplace, turning its niche data into a global product¹². Consultancies like Cognizant have similarly used marketplaces to **quickly source high-quality external data** for client projects, saving time and delivering better insights. These cases hint at the efficiency gains – instead of every company trying to collect or generate every data point, a marketplace allows them to *buy exactly what they need* from others who have it. This **reduces duplication of effort** in the economy and lets each organization focus on its strengths while still benefiting from data collected by others.

Second, **unlocking data via marketplaces can dramatically improve ROI on AI and analytics initiatives**. When internal data is combined with external complementary data, organizations can discover insights that were previously unattainable. For example, a business might augment its customer purchase data with third-party demographic or geolocation datasets to better personalize its marketing. Studies indicate that using a richer variety of data translates into more accurate models and better decision-making, which in turn drives higher revenue

12

<https://nimbusintelligence.com/2023/03/snowflakes-data-marketplace-a-game-changer-for-data-monetization-with-sql/#:~:text=SafeGraph%20is%20one%20of%20them,value%20of%20its%20data%20assets>

and profitability. In fact, data-driven organizations are **19 times more likely to be profitable** and substantially more likely to acquire and retain customers, according to analytics surveys¹³. Data marketplaces make it easier to become data-driven, by removing the friction in finding and licensing useful datasets. Additionally, **marketplaces encourage data to be standardized and high-quality** (since data providers compete and are vetted), which addresses a common problem of messy data that companies struggle to use. Overall, by **enabling efficient data sharing, marketplaces help maximize the economic return on data as an asset** – both for the data seller (who monetizes something that was idle) and the buyer (who gains insights or capabilities they otherwise wouldn't have).

Finally, there is a broader societal and macroeconomic benefit. As more data gets used (responsibly), we can expect **productivity and innovation to rise**. Entire new services and industries can be built on exchanged data – for instance, smart city applications using shared traffic and IoT data, or precision medicine AI systems trained on pooled healthcare data. One report noted that *the added value from “Internet of Things” data alone could be \$1.9 trillion worldwide over a five-year period*¹⁴. Much of that value will be realized only if data flows to where it's needed, rather than remaining in isolated pockets. In the coming decade, **data marketplaces could play a pivotal role in unlocking these trillions in value by connecting data “haves” with “have-nots.”** As the Future Agenda project remarked, data is like gold dust for organizations and *“demands a marketplace”* to truly realize its worth¹⁴. By creating liquidity in the data asset class, we set the stage for a more data-rich economy where AI and analytics can drive growth across all sectors.

Proprietary data is poised to remain a critical commercial asset over the next decade, underpinning competitive advantage and AI innovation. The evidence is clear that *most* of the world's useful data (well over 90%) is currently locked away in silos, underutilized but imbued with tremendous latent value. This presents a stark choice and opportunity: organizations can either continue hoarding data

¹³

<https://nektar.ai/5-ways-siloed-data-is-burning-your-revenue/#:~:text=Enabling%20this%20data%20to%20flow,more%20likely%20to%20be%20profitable>

¹⁴

<https://www.futureagenda.org/foresights/value-of-data/#:~:text=By%202020%20people%20and%20connected,themselves%2C%20but%20from%20the%20understanding>

(gaining limited value from it), or they can embrace new models of data sharing and monetization to unlock value on an unprecedented scale. The rise of data marketplaces and exchanges suggests that the tide is turning toward the latter – treating data as a tradeable asset to fuel the broader AI ecosystem. Investors are understandably eyeing this space, as the market for technologies and platforms that enable data exchange is growing rapidly, driven by demand in nearly every industry. The **total addressable market spans billions of dollars and is expanding by double digits annually**¹⁵, reflecting the recognition that data has become “the new oil” for the digital economy.

For AI development, the implications of tapping into proprietary data stores are profound. It could overcome the data scarcity that limits today’s models, enabling more powerful and specialized AI applications – in effect, *supercharging AI progress* in domains from medicine to finance. It could also democratize AI, allowing startups and smaller players to access quality data and compete, rather than ceding ground to only the data-rich tech giants. The economic benefits, too, are compelling: less wasted data means recouping trillions in lost productivity, and new data-driven services mean fresh revenue streams and efficiencies across the board. As one analysis succinctly put it, *data is now an economic input as important as capital or labor*¹⁶ – and just as markets exist to allocate capital and labor, efficient markets for data are now forming to allocate this valuable resource.

In light of the supporting evidence, “>95% of the world’s commercially valuable data is locked in siloes” does emphatically hold up and underscores a key point for the next decade: **unlocking proprietary data is one of the biggest opportunities in technology and business**. Those who manage to liberate and leverage these data troves (while respecting privacy and security) stand to reap substantial rewards. Conversely, organizations that ignore their data assets or keep them walled off may fall behind in the age of AI. Proprietary data will not only remain a key asset; its importance will grow as companies and entire industries transform

¹⁵

<https://www.fortunebusinessinsights.com/data-monetization-market-106480#:~:text=The%20global%20data%20monetization%20market,in%202024>

¹⁶

<https://www.futureagenda.org/foresights/value-of-data/#:~:text=It%20has%20never%20been%20easier,T his%20is%20gold%20dust%20for>

data from a hidden, isolated resource into an open, shared driver of innovation and economic value.

How RAFA accelerates LLMs and GWMs

How models become true “knowledge experts” with marketplace data

LLMs (large language models).

- **From generic word predictors to expert guides:** Today’s frontier LLMs excel on public web data; they underperform on domain nuance (financial data that requires numerical and mathematical accuracy, legal contracts, patient notes, plant telemetry, private ledgers). Injecting first-party and partner data via RAG/fine-tuning moves them from generic chat to **decision-grade copilots** in [finance](#), health, supply chain, and industrial operations. The RAG tooling market itself is scaling quickly (est. ~\$11B by 2030), signaling enterprise demand for fact-grounded outputs¹⁷.
- **Agentic workflows:** LLMs orchestrate tools, query live sources with access controls, and produce **auditable recommendations** (“explain-your-work” traces). Enterprises are shifting from static RAG islands toward **agent-based workflow architectures** that query governed sources at runtime, preserving permissions and reducing data leakage risks—key for regulated sectors¹⁸.

GWMs / “world models.”

- **What they are:** World models build an internal representation of environments to **simulate** futures, enabling planning and “what-if” scenario evaluations that used to take tremendous amounts of expert time and

¹⁷

https://www.grandviewresearch.com/industry-analysis/retrieval-augmented-generation-rag-market-report?utm_source=chatgpt.com

¹⁸

https://www.techradar.com/pro/rag-is-dead-why-enterprises-are-shifting-to-agent-based-ai-architectures?utm_source=chatgpt.com

resources, not just text prediction. Think virtual sandboxes for reasoning, simulations and experimentation¹⁹.

- **Why it matters for insights:** With marketplace data (e.g., SKU flows, credit exposures, maintenance logs, electronic medical records), GWMs can run **counterfactuals and scenario planning on an unprecedented level**: “If we re-route inventory,” “If loan policy tightens 50 bps,” “If shift patterns change,” yielding quantified decision choices. Big labs position world models as a bridge to more general reasoning and physical AI such as industrial and consumer robotics. Recent announcements underscore their role in simulating complex real-world settings²⁰.

Other specialist models that monetize insights once data flows:

- **Time-series forecasters** for demand, risk, and pricing.
- **Graph models** for fraud, supply-chain, and knowledge graphs (entity linking across partners).
- **Multimodal models** (text+tables+images+sensor data) for quality control, claims, and field ops. Together with LLM agents and GWMs, these form a **stack** that ingests marketplace data and outputs revenue-relevant decisions.

¹⁹ https://arxiv.org/pdf/2507.05169?utm_source=chatgpt.com

²⁰

https://www.theguardian.com/technology/2025/aug/05/google-step-artificial-general-intelligence-deepmind-agi?utm_source=chatgpt.com

What gets monetized? The Insight Product Classes

1. **Financial Models:** As the [first implementation](#) of decentralized financial intelligence marketplace that gets monetized with very high value per insight, RAFA's decentralized AI driven investing platform serves as a proof of concept for the decentralized data marketplace envisioned in this whitepaper.
2. **Insight APIs** (per-query): Insurance and similar product underwriting scores, demand signals, churn/propensity, supplier risk etc can get priced by call or by seat, which is then recorded, is auditable and source-attributed, providing confidence and meeting compliance standards.
3. **Copilot subscriptions:** function-specific copilots ([Financial Planning & Analysis](#), Revenue Operations, Claims, Care) bundled with data entitlements.
4. **Simulation & planning sandboxes** (GWM-backed): scenario packs sold by industry (e.g., "holiday logistics stress test").
5. **Outcome-based contracts:** share of uplift/savings for optimization levers (pricing, routing, staffing).
6. **Data-plus-model bundles:** curated datasets with tuned models; revenue-share flows back to data providers through the marketplace.

Market potential

Below, we summarize the value of “monetizable insights” by 2030. Below are credible, current market baselines that **become more fully addressable** once proprietary data exchanges are liquid and governed:

Component market (global)	2030 size / trajectory	Why it matters to insights
Financial Insights	~100B	Surfaces tacit, high value financial insights by monetizing them in a fair, verifiable marketplace
Conversational AI	~\$41B by 2030	Front-door for knowledge assistants & copilots ²¹ .
Decision Intelligence	~\$36B (GVR) to ~\$50B (M&M) by 2030	Converts data into decisions; where insight value is realized ²² .
AI Agents	~\$43–53B by 2030	Automates workflows; queries governed sources at runtime ²³ .

²¹

https://www.grandviewresearch.com/industry-analysis/conversational-ai-market-report?utm_source=chatgpt.com

²²

https://www.grandviewresearch.com/industry-analysis/decision-intelligence-market-report?utm_source=chatgpt.com

²³

https://finance.yahoo.com/news/ai-agent-market-forecast-reach-135600682.html?utm_source=chatgpt.com

Knowledge Mgmt Software	~\$32B by 2030 (other estimates higher)	Organizes enterprise know-how; foundation for retrieval ²⁴ .
RAG tooling/platforms	~\$11B by 2030 (multiple estimates)	Bridges proprietary data to LLMs with citations/grounding ¹⁷ .
Analytics-as-a -Service	~\$40–44B by 2030 (range across firms)	Packaged analytics delivery; natural channel for insight APIs ²⁵ .

Synthesis (insights TAM): Given heavy overlap, a conservative blended view yields **\$250–\$350B** in directly monetizable “insights” software/services by 2030, with upside if agentic/GWM simulation penetrates operations (pricing, routing, workforce, inventory). This sits within gen-AI’s broader economic impact (Per McKinsey, **\$2.6–\$4.4T annual value**), which becomes more attainable as proprietary data liquidity increases²⁶

²⁴

https://www.mordorintelligence.com/industry-reports/knowledge-management-software-market?utm_source=chatgpt.com

²⁵ <https://www.mordorintelligence.com/industry-reports/knowledge-management-software-market>

²⁶

<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>

Why a decentralized marketplace unlocks this TAM and shifts market share from centralized marketplaces

- **Access to unique signal means differentiated, verifiable accuracy.** Proprietary datasets add the “critical nuance” that foundation models lack. Domain-specific data demonstrably improves outcomes in Finance, healthcare and other verticals²⁷. Instead of relying on data “modulation”, “policy”, “ethics” and other risks that centralized providers like snowflake and Adobe have, a decentralized marketplace gives this access to anyone, verifiable openly, with cryptographic guarantees and not meatspace trust or contracts.
- **Governance & access control at query-time. A verifiable, trust minimized** Agentic architecture preserves permissions, satisfying security and —historically blockers to data sharing²⁸ via immutable on-chain contracts.
- **Distribution & liquidity for data products.** As with Snowflake/AWS-style exchanges, marketplaces reduce friction to **find/license/activate** external data, speeding cohort creation and model improvement. (This dynamic is already visible in data exchanges and is accelerating.)²⁹
- **Capital flows prove data’s cash value.** Large platforms now pay real money for proprietary datasets/tools to strengthen AI offerings (e.g., BlackRock buying Preqin to feed Aladdin’s analytics, Reddit, Brave etc monetizing their data sets etc serve as proofs of this accelerating trend)³⁰

²⁷ <https://www.aicontio.com/2024/02/20/what-are-general-world-models>

²⁸ <https://www.techradar.com/pro/rag-is-dead-why-enterprises-are-shifting-to-agent-based-ai-architectures>

²⁹

<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/charting-a-path-to-the-data-and-ai-driven-enterprise-of-2030>

³⁰ <https://www.investopedia.com/blackrock-buys-private-markets-data-firm-preqin-8671884>

Segment opportunities & Examples

Financial services (largest buyer of data & analytics):

- Given [RAFA.AI](#)'s expertise in leveraging financial data to provide AI enabled workflows and productivity solutions to leading asset management firms in the US, we see many applications in this space with high value : LLM-agent copilots for running on chain capital allocation, KYC, portfolio commentary, client Q&A with real time data feeds; graph models for fraud; GWMs for liquidity and stress simulations. Monetization is done via a mix of seat licenses + outcome fees . The DI/Conversational AI stacks are already material here³¹.

Healthcare & life sciences:

- LLMs grounded on EHR/clinical notes can help with diagnostic accuracy and efficiency, treatment copilots, trial care-path simulations; multimodal models with imaging + notes. More importantly, as hospitals have longitudinal objective data on most used interventions and therapies, they can be utilized to improve the discovery and research efficiency of Pharma and Biotech. These use cases carry very high price-per-insight due to outcome value.

Industrial & logistics:

- Sensor/maintenance logs + world models = **digital twins** for throughput and downtime reduction; insights sold as simulation packs and ongoing optimizers.

Retail/CPG:

- Demand sensing (LLM+time series models), promo/price elasticity simulations, agent copilots for merchandising and supply planning, external

mobility and macro feeds bought via the marketplace are some of the use cases.

Pricing & packaging patterns that work

- **Tiered entitlements:** insight tiers map to data rights (internal-only → partner bundles → premium marketplace packs).
- **Hybrid metering:** base subscription + usage overage for API queries, token usage etc separate SKUs for “simulation hours” (GWMs).
- **Attribution & auditability:** per-answer citations back to data sources, with verifiable compute and proof of value (drives trust and rev-share).
- **Outcome SLAs:** optional gain-share contracts for pricing/route/loss-mitigation use cases.

Key Takeaways

RAFA’s decentralized data marketplace will enable revenue capture by:

- **Picks-and-shovels:** Agent orchestration, policy-aware connectors, lineage/citation layers, and secure query-time federation should see outsized demand as enterprises avoid centralizing sensitive data³² and single vendor/marketplace dependency due to policy and regime risks, as well as trust issues as have continued to occur on the leading AD platforms.
- **Vertical insight vendors:** Firms with privileged data access (or exclusive marketplace bundles) will command premium ARPU and lower churn.
- **Data rights as moat:** Entities aggregating scarce signals (transactions, events, telemetry) will price above commodity web corpora; M&A will target these rights (cf. Preqin)³³. These data rights are immutably recorded on the blockchain as a permanent timestamping and property rights guarantees.

³² <https://www.techradar.com/pro/rag-is-dead-why-enterprises-are-shifting-to-agent-based-ai-architectures>

³³

https://www.investopedia.com/blackrock-buys-private-markets-data-firm-preqin-8671884?utm_source=chatgpt.com

- **Simulation/native world-model plays:** As GWMs mature, **scenario-as-a-service** emerges in robotics, warehousing, and operations; watch for early revenue in planning tools³⁴.
- **Macro tailwinds:** Adoption of gen-AI is broad and rising fast in the enterprise; insights categories ride this curve. Given that the data moat companies started restricting access to data to model companies (see Apple's private computation policy, Google and Reddit's recent data restrictions on training oAI etc), data marketplaces will be the only way for monetizing data from commercially valuable data holders.