

NeuralMesh™ AIDP in Financial Services

Continuously current, permission-aware AI knowledge for quantitative and high frequency trading, fraud detection, and banking services, all hosted on NeuralMesh AIDP

Challenges

- Regulatory and risk data locked out of AI systems
- Market and trading data goes stale before it can be queried
- Every quant and risk AI project rebuilds the same pipelines
- Compliance teams block AI rollouts over permission gaps
- Strict data residency rules limit what AI can touch
- GPU clusters sit idle while engineering teams hand-roll ETL
- Research velocity stalls while competitors iterate faster

Solution

AIDP watches NeuralMesh™ in real time, vectorizing trading data, research, and data records the moment they land, update, or change, with POSIX permissions enforced at query time

Benefits

- Trading data is always current
- Semantic search out of the box
- Regulatory-grade POSIX permission enforcement
- No pipelines to build or maintain
- Live monitoring and status dashboards in one user interface

AI-ready Financial Services data without manual burden

NeuralMesh™ AIDP is a software and hardware add-on to NeuralMesh™ that is deployed from the **WEKA App Store**. It transforms how financial services organizations operationalize AI factories, automatically handling ingestion, chunking, embedding, vectorization, and POSIX-enforced retrieval across trading data, risk models, regulatory filings, and financial records without custom engineering.

Rather than building and maintaining custom data pipelines for every quant model, risk system, or compliance project, AIDP automatically processes financial data during ingest, keeping it continuously AI-ready as market data arrives, regulatory filings update, or trading records change. AIDP delivers ready-to-consume AI outcomes such as semantic search over financial documents and real time data monitoring.

Financial services firms invest heavily in GPU infrastructure and hit the same wall: getting trading data, risk models, and compliance documents into AI systems reliably, securely, and at the speed markets demand. **AIDP removes that wall.**

Why choose WEKA's AI Data Platform?

AIDP is designed for easy deployment in financial services AI factories across quantitative research, risk analysis, fraud detection, regulatory intelligence, and trading infrastructure. AIDP extends NeuralMesh from a high-performance storage system into a complete solution that stores financial data at scale, keeps it current, enforces data access controls at query time, and delivers AI outcomes directly.

The problem AIDP solves for financial services organizations

Most financial services AI pilots fail to reach production because the data infrastructure underneath isn't built for continuous AI operations. For quantitative and high frequency trading workloads, risk systems, and fraud detection workflows to run cleanly, data must be parsed, chunked, embedded, and indexed.

As financial market data changes, the entire pipeline must stay in sync. **Without AIDP**, this means:

- **Per-project pipeline debt:** Every quant strategy, risk model, or fraud detection system requires a custom pipeline built from scratch, maintained indefinitely
- **Stale market and risk context:** When trading records, prices, or regulatory filings update in NeuralMesh, downstream AI systems don't automatically reflect those changes creating decisions made on stale data
- **Research velocity loss:** Every new strategy or model rebuilds the same ingestion and vectorization work. Time spent on infrastructure is time not spent finding alpha
- **Slow time-to-production:** Selecting, integrating, and validating an AI data stack delays the first AI outcome by months. Financial teams can't wait quarters to get to production
- **Compliance and security exposure:** Standard pipelines strip ACLs at ingest. In regulated financial environments, this blocks AI deployment until security teams can verify governance holds through every layer of the AI stack

AIDP offers two fundamental benefits

1. Continuous, automated data preparation simplifies workflows

AIDP watches WEKA NeuralMesh filesystems in real time. As market data arrives, trading records change, or regulatory filings update, AIDP automatically handles the full preparation pipeline without batch jobs or manual intervention. AIDP:

- Detects every file create, update, and delete via continuous monitoring without a batch scheduler, polling, or reindexing cycle
- Parses and chunks financial documents, research reports, market data feeds, and structured records as they arrive
- Embeds POSIX ACL, UID, and GID metadata at ingest. If a trader or analyst can't see the source file, they can't see the vector. No middleware, no compliance gap
- Keeps vector indexes synchronized with the live state of your financial data, including market moves, file deletions, and record updates propagate automatically

That continuous data readiness is what makes the second benefit possible: delivering AI outcomes directly to the financial professionals who need them, without additional engineering overhead.

2. Ready-to-consume data generates better AI outcomes

Your quantitative or high frequency trading team should be building strategies and models, not babysitting ETL jobs. AIDP delivers AI outcomes directly to financial professionals and applications without requiring engineering teams to deploy or manage AI frameworks:

- **Market Signal Detection** tracks and displays when financial data lands, from earnings releases, analyst reports, regulatory filings, and market feeds, all surfaced the moment it arrives, with results limited to what the requesting user is authorized to see
- **Semantic Search** enables natural language queries over research reports, trading records, and risk documentation stored on NeuralMesh and accessed via browser UI or API, backed by NVIDIA NIMs, all managed by AIDP and supported by WEKA

How AIDP works: platform architecture

AIDP is a full-stack system of multiple integrated services that handles the complete journey from raw financial data to AI-ready outcomes in a secure, permissions-based environment. Key outcomes include:

Theme	Without AIDP	With AIDP
Data Freshness	AI projects use stale and outdated data when files change	Create, update, and delete events automatically propagate to the vector store
	Superseded filings, stale prices, and retracted records linger in the index	Superseded and deleted documents are purged from the index automatically
Compliance & Governance	AI responses ignore file-level permissions and governance procedures	POSIX ACL enforced from NeuralMesh through to every AI response
	Manual vector purge on record deletion creates compliance risk	File deletion propagates automatically to the vector index, honoring data retention and deletion rules natively
	Each RAG pipeline requires separate audit validation, multiplying the compliance burden across teams	One validated platform serves all teams, consolidating the audit and model-risk surface
Architecture & Maintenance	Custom data pipelines are built and maintained for every project	One platform continuously prepares all NeuralMesh data for AI
	IT teams manage AI frameworks directly	Semantic search and more are delivered out of the box
Visibility & Control	Siloed pipelines offer little or no operational visibility across pipelines	Live queue monitoring, status dashboards, and role-based governance in one UI

AIDP workloads in Financial Services

Quantitative research and strategy backtesting

Quantitative trading teams run agentic RAG and semantic search over research notes, backtest results, factor libraries, and market data to find and validate signals faster than competitors. Nearly all public market trading now runs on quantitative algorithms, so firms that test more strategies in less time compound a research velocity advantage. Those queries are only useful if the index reflects the latest data and only safe if each desk sees only what it is entitled to.

Because market data, prices, and research land continuously and batch reindexing lags, a strategy built on a stale factor or a superseded dataset rests on data that no longer holds. Legacy pipelines also strip permissions at ingest, so information barriers between desks depend on bolt-on middleware that risk and compliance teams may refuse to clear for production.

NeuralMesh AIDP solves this problem by:

- Reflecting new market data, prices, and research the moment they land, eliminating batch reindex gaps
- Purging superseded datasets and retracted results from the index automatically
- Enforcing desk-level file permissions (POSIX ACLs) at query time, keeping information barriers intact

With a research knowledge layer that NeuralMesh AIDP keeps current and permission-aware, quant teams test more strategies against the live state of the market and spend their time finding alpha, not maintaining indexes.

High frequency trading

HFT teams run large-scale scenario analysis and backtesting across years of tick data, order book history, and execution logs to tune signals and risk controls before they reach the market. Reducing wall clock time on these runs directly widens the set of scenarios a team can test. Agentic RAG over that history lets researchers compare current conditions to prior regimes and reuse past analysis instead of rerunning it.

In legacy approaches, backtest outputs and tick archives accumulate faster than teams can catalog them, locking prior runs in isolated directories where they are functionally invisible, while superseded datasets linger and pollute downstream queries. This forces researchers to repeat expensive runs that have already been conducted and rebuild the same search stack for every new model.

NeuralMesh AIDP solves this problem by:

- Indexing backtest outputs and tick datasets on landing, making them immediately searchable
- Purging superseded and retracted runs from search results automatically

- Enabling long-context queries across years of market history without GPU recompute, using WEKA Augmented Memory Grid

With NeuralMesh carrying the throughput and NeuralMesh AIDP governing the archive, HFT teams find and reuse every run they have executed instead of rebuilding the same search stack for each model.

Real-time sports and prediction markets

Prediction market and real-time sports platforms ingest live event feeds, odds movements, and historical outcomes, then run models that price and settle markets while events are still unfolding. Semantic search and agentic RAG over live and historical data let teams surface comparable past events and adjust models as conditions change, all while user volume spikes around each event.

In legacy environments, live feeds and model outputs arrive faster than batch pipelines can index them, so models query stale context at the exact moment odds are moving. Rebuilding a separate ingestion and search stack for every new market and scaling it for event-day spikes delays launches and leaves queries running against data that is already out of date.

NeuralMesh AIDP solves this problem by:

- Reflecting live event data and odds movements in the index the moment they land
- Purging settled and voided markets from search results automatically, keeping queries clean
- Scaling search across live and historical data in one namespace without rebuilding pipelines per market

With NeuralMesh handling the live throughput and NeuralMesh AIDP keeping the index current, platform teams price and settle markets against the live state of each event instead of babysitting ingestion jobs.

Fraud detection

Fraud teams run models and agentic RAG over transaction streams, device and account histories, and prior case files to detect and respond to threats in seconds. Rapid access to transaction data is what lets a system flag and block a fraudulent pattern before it clears. Those detections are only accurate if the index reflects the latest transactions and only compliant if analysts see only the accounts they are authorized to review.

In legacy approaches, new transactions and case updates lag behind batch reindexing, so models score activity against a picture of the account that is already out of date. Standard pipelines also strip access controls at ingest, so account-level restrictions depend on separate middleware that security teams may refuse to clear for production use on sensitive customer data.

NeuralMesh AIDP solves this problem by:

- Reflecting new transactions and case updates in the index the moment they land
- Enforcing account and analyst file permissions (POSIX ACLs) at query time, blocking access to restricted records

- Purging closed and reversed cases from search results automatically

With a case and transaction knowledge layer that NeuralMesh AIDP keeps current and permission-aware, fraud teams act on the live state of every account and keep their attention on stopping threats, not maintaining indexes.

Payment analytics

Payment analytics teams run semantic search and agentic RAG across authorization data, settlement records, merchant profiles, and dispute histories to monitor flows, explain declines, and surface risk in near real time. These queries are only useful if the index reflects the latest payment activity and only safe if each team sees only the merchants and records it is entitled to.

In legacy approaches, payment records land continuously while batch reindexing lags, so analysis runs on settlement and dispute data that no longer matches the live ledger. Every new analytics program rebuilds the same ingestion and governance stack from scratch, delaying the first result and multiplying the compliance surface across teams.

NeuralMesh AIDP solves this problem by:

- Indexing authorization, settlement, and dispute records on landing, making them immediately searchable
- Enforcing merchant and team file permissions (POSIX ACLs) at query time, with no separate middleware
- Purging reversed and superseded records from the index automatically, keeping analysis clean

With one governed payment knowledge layer from NeuralMesh AIDP, analytics teams work from the live state of the ledger across programs instead of rebuilding the same stack for each one.

Banking solutions

Retail and commercial banking teams run agentic RAG and semantic search across customer records, product documentation, regulatory filings, and policy libraries to answer questions, check compliance, and support customers at scale. Demand swings with seasonal cycles and regulatory deadlines, and the models behind these answers grow more complex as banks consolidate data across lines of business.

In legacy approaches, policies are revised, products change, and filings are superseded, but downstream AI systems keep answering from documents that no longer apply, creating regulatory and customer risk. Each line of business also builds its own search and governance stack, so scaling for seasonal demand and audit requirements multiplies cost and validation work across the bank and pushes access control into middleware that security teams may never approve.

NeuralMesh AIDP solves this problem by:

- Purging withdrawn policies, superseded products, and stale filings from the index automatically

- Enforcing line-of-business file permissions (POSIX ACLs) at query time across shared data
- Scaling long-context queries for seasonal demand without GPU recompute, using WEKA Augmented Memory Grid

With one validated knowledge layer from NeuralMesh AIDP, banking teams answer from the current state of policy and product across every line of business instead of maintaining a separate stack for each.

Ready to get started?

Filesystem level permissions (POSIX-ACL)

Financial services AI programs stall at the access-control layer. AIDP resolves this natively, no custom middleware required.

- POSIX permissions are already the standard in regulated Linux environments
- AIDP enforces them at inference time, not as a post-retrieval filter
- Structurally, not by policy, entitlement-scoped queries can’t surface restricted vectors
- Audit lineage is native: every retrieval event is tied to the enforced permission set
- Result: the same access control model the CISO already manages governs the AI

AIDP turns “the CISO blocked AI” into “the CISO approved AI.”

0 hours	1	Auto
Deletion-to-searchable latency vs. batch purge window	Validated platform replaces per-team RAG pipeline silos	Record deletion honored natively; propagates to the vector index

Ready to bring WEKA AIDP to your Financial Services AI program?

Visit weka.io/contact to schedule an architecture briefing for your financial services AI program.

➔ See NeuralMesh AIDP in Action