



# Automated Trading AI Agent

---

Product Strategy and Implementation

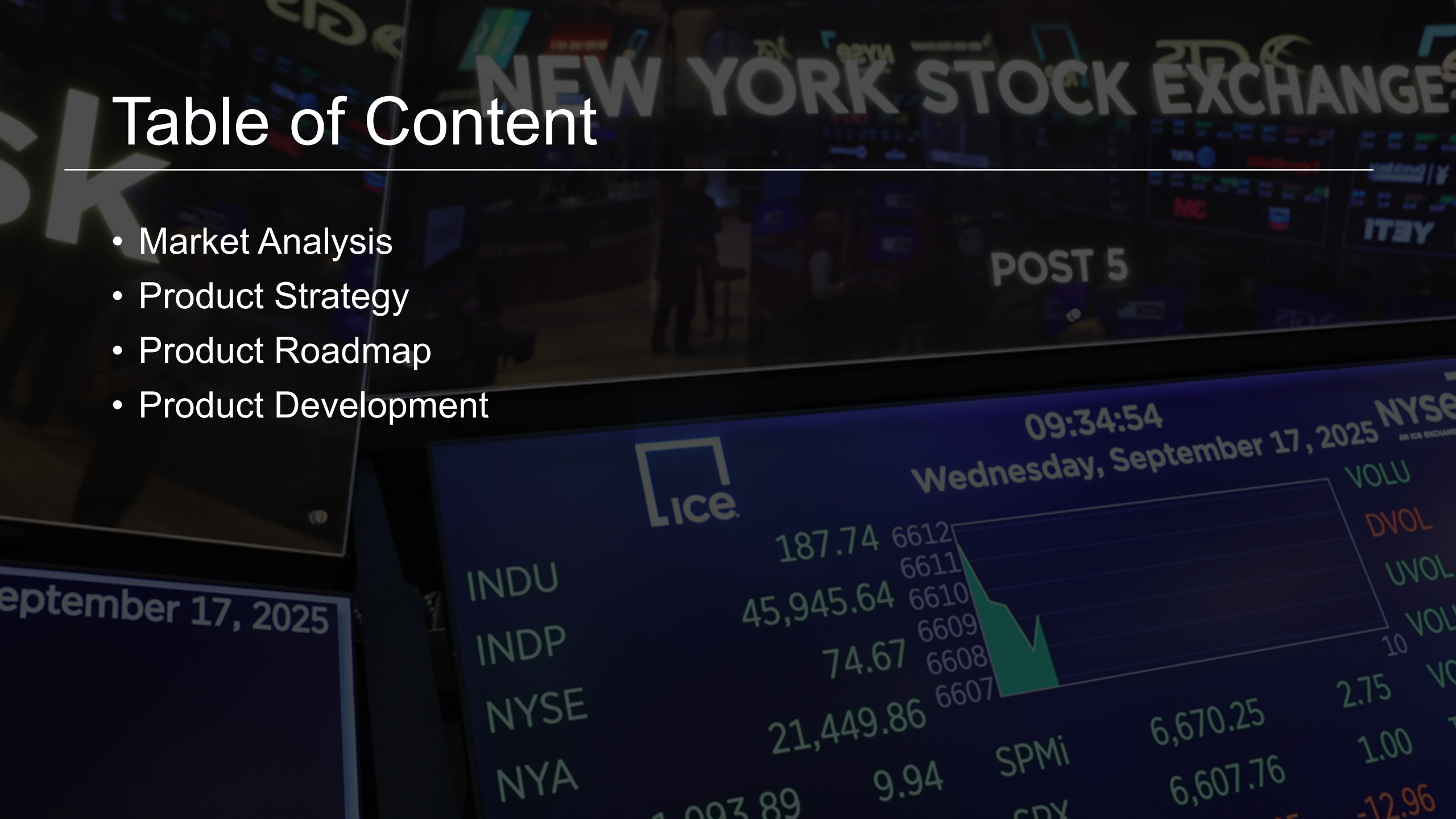
Eva Liu - [wenyuliuemail@163.com](mailto:wenyuliuemail@163.com)



# Table of Content

---

- Market Analysis
- Product Strategy
- Product Roadmap
- Product Development







NEW YORK STOCK

POST 5

# Market Analysis

ber 17, 2025



09:34:5

Wednesday, S

INDU

187.74 6612  
45,945.64 6611  
6610

INDP

74.67 6609  
6608  
6607

NYSE

21,449.86

NYA

1,093.89

9.94

SPMi

UTIL

06 38

0.13

SPX

CDM



# Market Analysis: Large Broker-Dealer Industry in Wealth Management

## Market Overview

Large broker-dealers (B/Ds) in wealth management—primarily the traditional wirehouses (Morgan Stanley Wealth Management, Merrill Lynch/Bank of America, UBS Wealth Management, Wells Fargo Advisors) plus scaled national/regional and independent B/D platforms (e.g., LPL Financial, Ameriprise, Raymond James, Edward Jones)—form a core channel of the U.S. (and global) wealth management industry.

They combine brokerage execution, advisory (fee-based and commission), banking/lending, research, and holistic planning for mass-affluent through ultra-high-net-worth (UHNW) clients.

The industry sits within a broader ecosystem of ~\$60–130+ trillion global wealth AUM – Assets Under Management (varying by definition and year), with U.S. advised assets and relationships continuing to expand amid demographic shifts, market growth, and technology disruption.

## Upstream

Suppliers, Inputs,  
and Enablers

- Asset managers / product manufacturers (BlackRock, Vanguard, Capital Group, alternatives providers);
- Custodians and clearing (Pershing, National Financial, DTCC);
- Data/research providers (Bloomberg, FactSet, Morningstar);
- Technology platforms (CRM, portfolio management, trading systems, AI/analytics vendors, TAMPs);
- Capital markets / investment banking arms; regulators (SEC, FINRA, CFTC, state securities regulators);
- Talent pipelines (universities, CFA/CFP programs, recruiting firms);
- Insurance carriers and structured product issuers;
- Fintech/WealthTech infrastructure.

## Downstream

Clients,  
Distribution, End  
Users

- Mass affluent, high-net-worth (HNW), ultra-high-net-worth (UHNW) individuals and families;
- Family offices;
- Institutional/endowment clients (selectively);
- Registered investment advisors (RIAs) and hybrid platforms that use B/D custody or clearing;
- Corporate retirement / stock-plan participants;
- Next-generation heirs (wealth transfer);
- Digital self-directed channels (e.g., E-TRADE within Morgan Stanley);
- Banks' private banking arms.

# Rising AI Trend towards Wealth Management in 2035

## 1. Invest in agentic AI: From automation to orchestration

CEOs must see beyond the automation of AI as a platform of semi-autonomous agents that collaborate with advisors, client salespeople, service team members, back-office staff, and each other to execute multistep workflows. Across industries, initial use cases of AI agents to automate limited tasks have been shown to deliver up to 3 to 5 percent improvements in productivity annually. Teams of AI agents that executed complex workflows not only resulted in productivity improvements but also could increase growth by more than 10 percent.

## 2. Reinvest AI technology dividends into innovation

Automation and AI will continue to expand capacity and reduce costs across middle and back offices. But efficiency alone is not a strategy. The real differentiator will be how firms redeploy productivity gains to fund “change-the-business” operations to unlock innovation in services, experiences, and business models. Leaders should have a clear vision of which competitive archetype their firm will pursue, shift from isolated technology projects to domain-level changes, and track realized savings.

## 3. Lead an AI-fluent, human-centered organization

The integration of AI is an organizational transformation. While it has been critical for business leaders to understand technology applications in the digital age, AI is making it more important than ever. CEOs must champion AI fluency across the enterprise, ensuring that leaders and teams understand how to use, govern, and collaborate with intelligent systems responsibly. AI fluency must include strategic literacy: understanding where AI can create value, what new business models it enables, and how to balance innovation with risk.

## 4. AI-enabled digital wealth managers

A novel fourth archetype of digital-first, AI-native firms would redefine advice delivery for the next generation of investors. Combining autonomous AI agents with human oversight, these firms would deliver personalized, low-cost advice at scale, expanding access to financial guidance. As the industry confronts a growing shortage of advisor talent—driving human-led services to become more expensive and exclusive—the best advisors are likely to migrate further upmarket.



# Competitive Landscape: AI Agents / Platforms for Thematic Automated Trading & Portfolio Construction

|                 | Application Features                                                                                                                                                                                                                                                                                                                          | Functionality                                                                                                                                                               | Use Case                                                                                                                                                                   | Managed Wealth Scale                                                                                                                                                                                                |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Magnifi (TIFIN) | <ul style="list-style-type: none"> <li>Natural-language semantic search across ETFs, stocks, mutual funds, SMAs;</li> <li>thematic discovery;</li> <li>AI chatbot for research &amp; explanations;</li> <li>linked multi-brokerage account analysis;</li> <li>themed managed portfolios;</li> <li>portfolio comparison tools</li> </ul>       | AI-powered investment discovery & research; conversational queries (e.g., “AI infrastructure ETFs with low expense ratios”); trade execution support; personalized insights | Advisors quickly surface and compare thematic ideas (AI, clean energy, defense, etc.), build client-ready recommendations, and explain rationale without complex screeners | Thousands of investors with >\$5B in linked assets (earlier figures cited hundreds of millions; platform focuses more on discovery than discretionary AUM)                                                          |
| Vise            | <ul style="list-style-type: none"> <li>AI-driven hyper-personalized portfolio construction;</li> <li>automated rebalancing &amp; tax-loss harvesting;</li> <li>long/short capabilities;</li> <li>real-time market commentary &amp; client explanations;</li> <li>multi-asset support include thematic sleeves;</li> </ul>                     | Build, manage, and explain customized portfolios at scale; opportunistic automation while keeping advisors in control; AI insights for client conversations                 | Advisors create thematic or goal-based model portfolios quickly, manage them with continuous optimization, and generate explainable client narratives                      | Platform assets >\$100B (as of mid-2026 reporting); discretionary AUM in strategies reported in the multi-billion range (earlier figures ~\$6B+ discretionary / tens of billions platform)                          |
| Investnet       | <ul style="list-style-type: none"> <li>Insights AI (agentic next-best-action engine);</li> <li>Generative BI (natural-language queries on data);</li> <li>decision intelligence layer; workflow automation;</li> <li>portfolio analytics &amp; rebalancing tools;</li> <li>integration across planning, investments, and reporting</li> </ul> | Surfaces personalized opportunities (including thematic tilts); automates insights and recommendations; embeds AI into advisor workflows for faster decisioning             | Advisors at scale receive proactive alerts and construct/adjust portfolios with AI-supported thematic or factor overlays while maintaining governance                      | Platform assets ~\$7 trillion+; widely used across the industry; It reflects the total client assets flowing through its technology, reporting, trading, UMA/SMA programs, model marketplaces, and related services |

# SWOT Analysis: Targeting Large Broker-Dealer Financial Advisors

## Strengths

- **Thematic specialization + agentic capability:** Ability to focus narrowly on theme-driven discovery, construction, supervised execution, and narrative generation can differentiate from broader platforms that treat themes as one of many features.
- **Advisor productivity leverage:** Large B/D advisors face high client loads and time pressure. An agent that reduces research-to-recommendation cycle time while generating explainable outputs.
- **Integration potential with existing scale players:** Platforms such as Envestnet (~\$6.5–7T platform assets, 110k+ advisors) and Aladdin already serve the target segment.

## Weaknesses

- **Regulatory and supervision constraints:** Fully autonomous trading agents face significant hurdles under fiduciary/suitability standards, best-execution rules, and firm-level supervision requirements.
- **Limited brand trust and distribution at launch:** Large B/D advisors and home offices prefer proven, integrated platforms. A new entrant lacks the existing relationships, compliance track record, and enterprise sales muscle of incumbents.
- **Data and integration barriers:** Access to real-time firm-specific client data, proprietary research, and trading rails is restricted. Without deep partnerships, the agent risks operating in a silo with incomplete context.

## Opportunities

- **Thematic investing growth:** Continued investor interest in megatrends (AI, energy transition, geopolitics, infrastructure) creates demand for specialized, dynamic exposure tools beyond static ETFs.
- **Hybrid / supervised automation models:** Opportunity to position as a “co-pilot” agent that proposes trades or rebalances within firm-approved guardrails, lowering the barrier to adoption inside wirehouses.
- **AI capability acceleration:** Rapid advances in agentic architectures, reasoning, and multimodal models can enable more sophisticated multi-step workflows.

## Threats

- **Incumbent platform dominance:** Envestnet, BlackRock Aladdin, and major wirehouse internal builds already control distribution, data, and advisor workflows at massive scale. They can rapidly add competing agentic features.
- **Build-vs-buy by large B/Ds:** Morgan Stanley and peers are actively deploying GenAI tools (e.g., Aladdin Auto Commentary). They may prefer internal development or exclusive partnerships over third-party agents.
- **Regulatory tightening:** Heightened scrutiny of AI in investment advice, algorithmic trading, and fiduciary processes could slow adoption or impose costly compliance requirements.



NEW YORK STOCK

POST 5

# Product Strategy

ber 17, 2025



09:34:5

Wednesday, S

INDU

187.74 6612

INDP

45,945.64 6611

NYSE

74.67 6610

NYA

21,449.86 6609

UTIL

1,093.89 6608

9.94

SPMi

0.13

SPX

06 38

CDM



# Product Vision: Theme-to-Execution AI Agent platform for Financial Advisors

## Positioning Statement

For **financial advisors at large wirehouse/broker-dealer teams** who need to respond to **client-driven theme requests** (AI, clean energy, defense) with specific, timely trades, the **Automated Trading Agent** is a **theme-to-execution platform** that translates a theme into compliant, executable trade baskets across client portfolios in minutes — unlike fragmented research tools and OMS platforms that leave advisors stuck between hours of manual screening and missed market windows.

### Industry proven core value position

10% – 25%

**Improved Conversation rate**

AI-powered lead scoring

5 – 15 p.p.

**NPS Improvement**

Better client experience

25% – 30%

**Capacity Unlock Potential**

Planning, portfolio management

+15% – 20%

**Revenue per advisor**

Cross sell, retention, capacity

## Target Persona

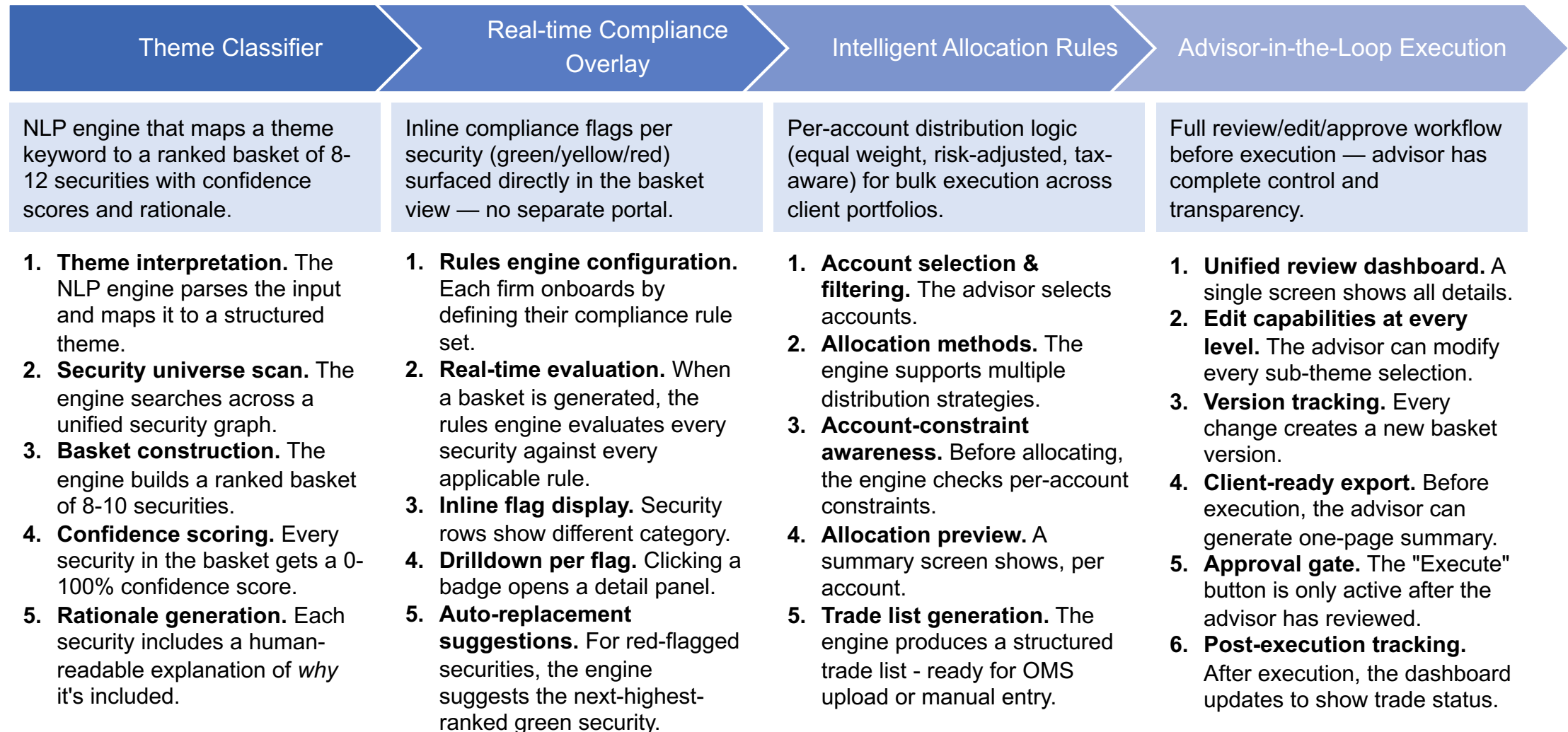
### Financial Advisor at a Large Broker-Dealer

- Manages 50-150 client portfolios with varying mandates and risk profiles
- Clients increasingly ask for thematic exposure but can't name specific securities
- Operates under strict firm compliance: approved lists, suitability, concentration limits
- Spends 2-4 hours per theme researching across Morningstar/YCharts/Bloomberg
- Then manually enters trades account-by-account in the OMS

### Current Pain Points

1. **Selection Paralysis:** Research tools are built for analyzing individual securities, not mapping concepts to baskets
2. **Compliance Gauntlet:** Compliance is in a separate system with no real-time feedback, forcing the advisor to invest hours building a basket without knowing which securities will pass
3. **Execution timing:** Theme opportunities are time-sensitive, but manual execution misses the window
4. **The Window Closes:** The end-to-end process — research, compliance, execution — was designed by different vendors for different users with no one optimizing for speed

# Product Mission: Accelerate Value Generation speed with Compliant Decision Making





# Measurable Objectives: Define the Winning points towards Platform Success

## Early Stage based on MVP

| Metric                       | Definition                                                                     | Target                      | Proves                                 |
|------------------------------|--------------------------------------------------------------------------------|-----------------------------|----------------------------------------|
| Active advisor seats         | Unique advisors generating $\geq 1$ theme basket per week                      | 50 seats                    | Product-market fit                     |
| Teams penetrated             | Unique advisor teams (not just individual advisors) with $\geq 3$ active seats | 10 teams                    | Team-level stickiness, network effect  |
| Seat expansion rate          | % of teams that grow from initial seats to more seats within 6 months          | $\geq 40\%$ of teams expand | Land-and-expand is happening           |
| Advisor-to-advisor referrals | New seats originating from existing advisor invitations (not sales-driven)     | $\geq 25\%$ of new seats    | Organic growth, product selling itself |
| Weekly active rate           | % of licensed seats that log in and generate at least one basket per week      | $\geq 60\%$                 | Seats aren't shelfware                 |
| Theme diversity              | Number of unique themes searched across the platform                           | $\geq 50$ distinct themes   | Advisors are exploring broadly         |

## Market Winning Stage with full Functionalities

| Metric                         | Definition                                                                                              | Target                | Proves                                                |
|--------------------------------|---------------------------------------------------------------------------------------------------------|-----------------------|-------------------------------------------------------|
| AUM influenced                 | Total dollar value of baskets generated and approved (whether executed or not)                          | \$500M                | Product is effectively increasing the profit          |
| Time saved per basket          | Manual baseline (est. 2-4 hours) vs. agent-assisted time — measured via advisor quarterly surveys       | $\geq 85\%$ reduction | Core promise: hours $\rightarrow$ minutes             |
| Compliance first-pass rate     | % of generated securities that pass all compliance rules on first evaluation (before any advisor edits) | $\geq 85\%$           | Classifier understands the firm's universe            |
| Basket-to-execution conversion | % of generated and approved baskets that reach OMS export or trade execution                            | $\geq 60\%$           | Baskets aren't just exploration — they lead to action |
| Annual recurring revenue       | Total ARR from seat licenses + AUM-based fees                                                           | \$80M                 | Business viability                                    |

# Value Proposition: Theme-based, Compliant, and Ready-to-Execute portfolio solution in minutes

## Value Proposition

---

For **large broker-dealer financial advisors** who need to respond quickly and compliantly to client interest in thematic investing, this AI Agent delivers a **purpose-built, advisor-controlled workflow** that turns a theme keyword into a **ready-to-execute, multi-account portfolio** in minutes — without sacrificing oversight or regulatory discipline.

### Competitive Differentiation

- **Versus Magnifi / TIFIN:** Stronger on compliance visibility and multi-account bulk execution; more focused on actionable trading rather than pure discovery.
- **Versus Vise:** More specialized for rapid thematic idea generation and explicit compliance flagging, with a lighter, advisor-controlled workflow rather than full discretionary portfolio management.
- **Versus Investnet / Aladdin:** Complements rather than replaces these platforms. Acts as a fast, thematic “front-end” agent that can feed into existing enterprise systems while solving the specific pain of theme-to-compliant-basket conversion.
- **Versus internal wirehouse tools:** Offers a more specialized, modern NLP-driven theme classifier and true multi-account allocation logic that many internal systems still lack in a unified view.

## Feature Differentiation

---

### O1 – Theme Classifier

Dramatically accelerates research. Advisors move from client request (“I want AI infrastructure exposure”) to a transparent, ranked shortlist in seconds instead of hours of manual screening.

### O2 – Real-time Compliance Overlay

Eliminates the friction of switching to separate compliance portals or model lists. Advisors instantly see which names are approved, restricted, or require extra review — reducing compliance risk and speeding decision-making.

### O3 – Smart Allocation Rules

Enables true scale. Advisors can apply the same thematic idea across dozens or hundreds of accounts while respecting individual risk profiles and tax situations, without building each portfolio manually.

### O4 – Advisor-in-the-loop Execution

Preserves the advisor’s professional judgment and fiduciary responsibility. The Agent proposes; the advisor decides. This design satisfies both regulatory expectations and advisor preference for control.





# Product Roadmap

---

ber 17, 2025

NEW YORK STOCK

POST 5



09:34:5

Wednesday, S

INDU

187.74 6612

INDP

45,945.64 6611

NYSE

74.67 6610

NYA

21,449.86 6609

UTIL

1,093.89 6608

9.94

SPMi

0.13

SPX

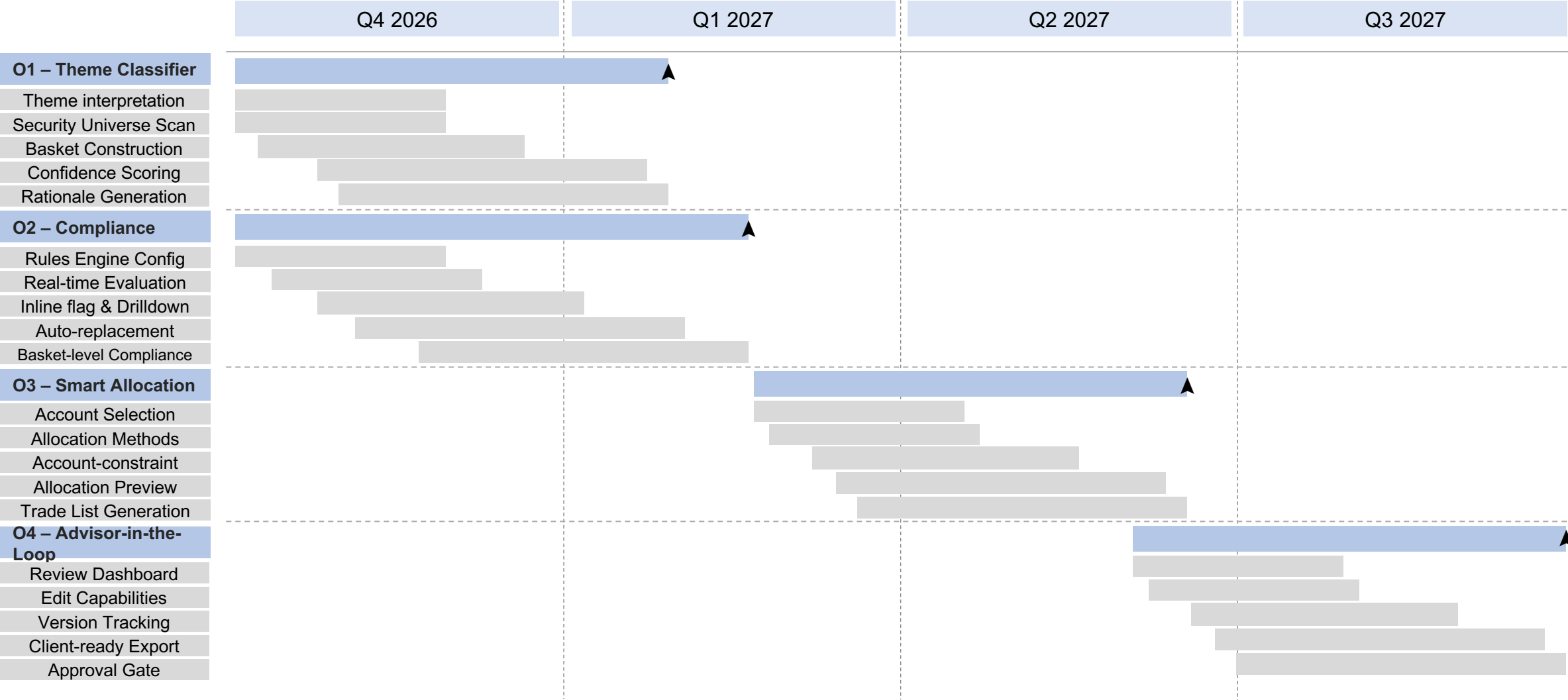
06 38

CDM

# Product Roadmap based on Feature Epic and Priority Allocation

|                     | Application Features                                                                                                                                                                                                                                          | Configurability                                                                                                                                                                                                                                                                                             | Priority | Estimated Timeline         |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------|
| Theme Classifier    | <ul style="list-style-type: none"><li>▪ Theme interpretation &amp; disambiguation;</li><li>▪ Security universe scan;</li><li>▪ Basket construction;</li><li>▪ Confidence scoring;</li><li>▪ Rationale generation;</li></ul>                                   | <ul style="list-style-type: none"><li>▪ Filter by security type;</li><li>▪ Set min market cap and liquidity threshold;</li><li>▪ Exclude specific sectors or tickers per advisor preference;</li><li>▪ Bias toward or away from the firm's internal research ratings;</li></ul>                             | High     | 3-6 months; (6-12 sprints) |
| Compliance Overlay  | <ul style="list-style-type: none"><li>▪ Rules engine configuration;</li><li>▪ Real-time evaluation;</li><li>▪ Inline flag display;</li><li>▪ Drilldown per flag;</li><li>▪ Auto-replacement suggestions;</li><li>▪ Basket-level compliance posture;</li></ul> | <ul style="list-style-type: none"><li>▪ Rule schema is per-firm, not hard-coded;</li><li>▪ Rules have effective dates;</li><li>▪ Yellow warnings can require or bypass attestation per firm policy;</li><li>▪ Full audit log of every compliance — who saw what, when, and what action they took;</li></ul> | High     | 3-6 months; (6-12 sprints) |
| Smart Allocation    | <ul style="list-style-type: none"><li>▪ Account selection &amp; filtering;</li><li>▪ Allocation methods;</li><li>▪ Account-constraint awareness;</li><li>▪ Allocation preview;</li><li>▪ Trade list generation;</li></ul>                                     | <ul style="list-style-type: none"><li>▪ Allocation method policy;</li><li>▪ Account constraint thresholds;</li><li>▪ Rebalancing policy;</li><li>▪ Saved account groups;</li><li>▪ Default allocation preferences;</li></ul>                                                                                | Mid-High | 2-4 months; (4-8 sprints)  |
| Advisor-in-the-Loop | <ul style="list-style-type: none"><li>▪ Unified review dashboard;</li><li>▪ Edit capabilities at every level;</li><li>▪ Version tracking;</li><li>▪ Client-ready export;</li><li>▪ Approval gate;</li><li>▪ Post-execution tracking;</li></ul>                | <ul style="list-style-type: none"><li>▪ Approval workflow rules. The firm defines what requires attestation;</li><li>▪ Audit trail requirements. Configurable by firm policy;</li><li>▪ Client communication templates;</li><li>▪ OMS export format</li></ul>                                               | Mid      | 2-4 months; (4-8 sprints)  |

# Project Plan with Explicit timeline for Feature Implementation







# Product Development

ber 17, 2025

# NEW YORK STOCK

POST 5



09:34:5  
Wednesday, S

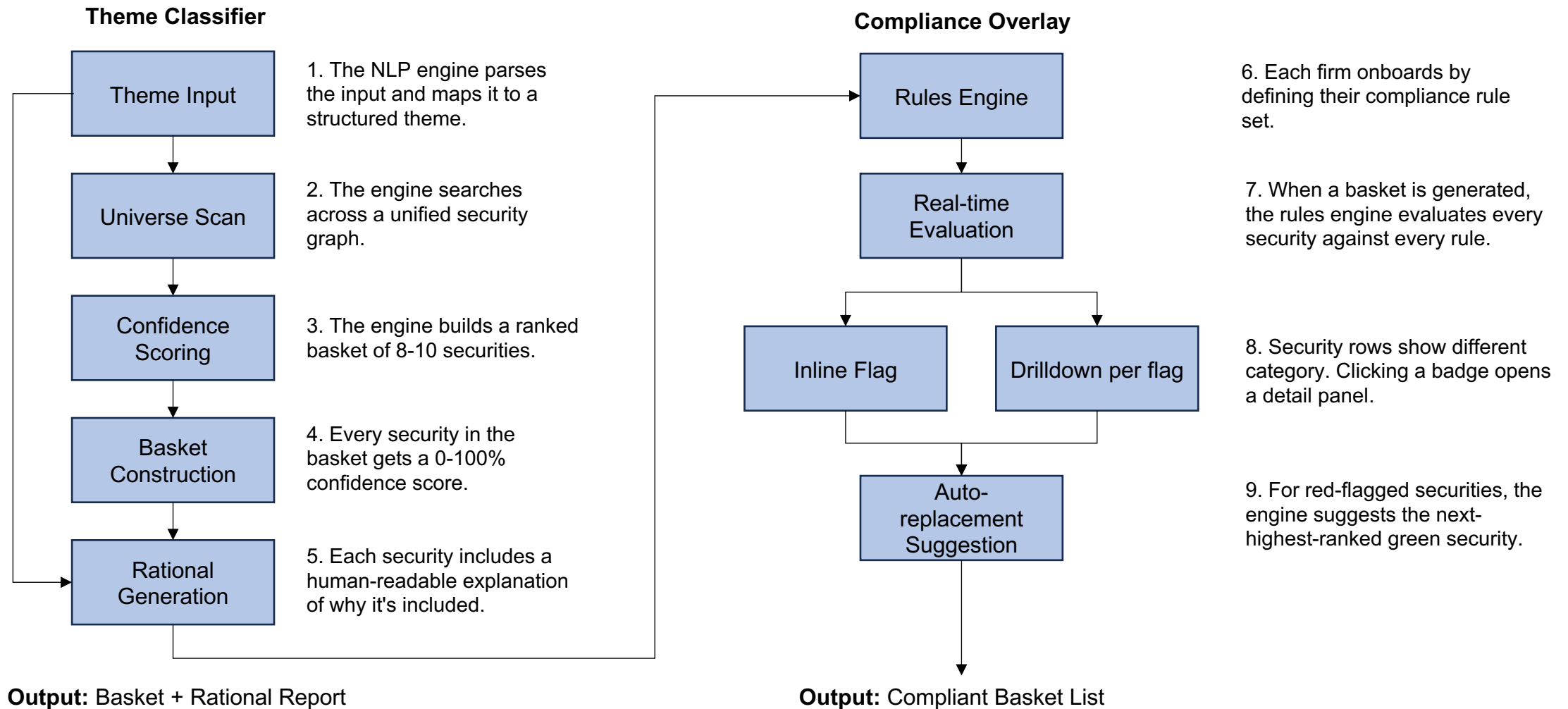
INDU  
INDP  
NYSE  
NYA  
UTIL

187.74 6612  
45,945.64 6611  
74.67 6610  
21,449.86 6609  
1,093.89 6608  
06 38 6607

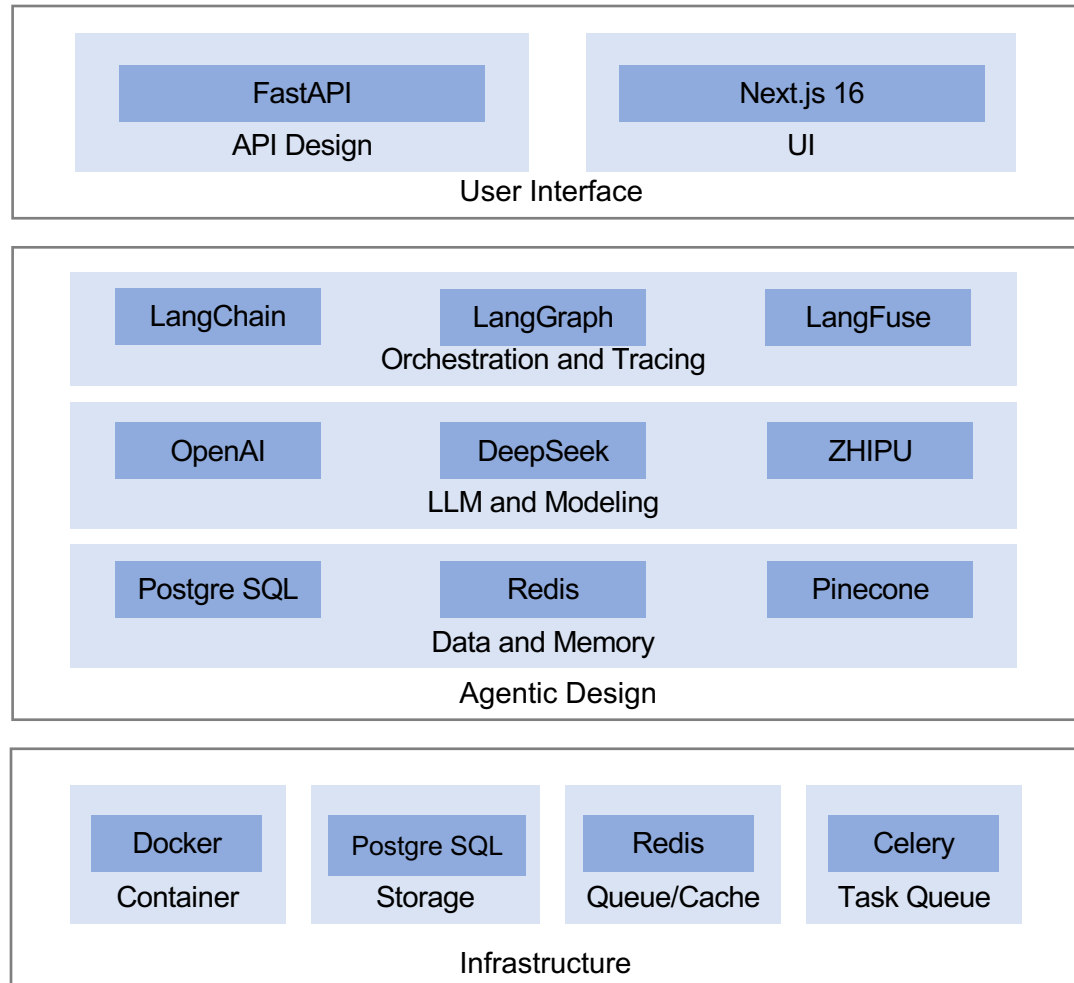
9.94 SPMi  
0.13 SPX  
CDM



# MVP Workflow: Verifiable Workflow to Implement Theme Selection and Compliance check



# Technical Stack: Tiered Technical Solution based on Application Features



## User Interface Layer:

- Entry points for end users, web/mobile apps, chat, IDE plugins
- Need infrastructure to parse, validate, and act on structured outputs, web/mobile apps, CLIs, Slack bots, browser extensions

## Agentic Design Layer:

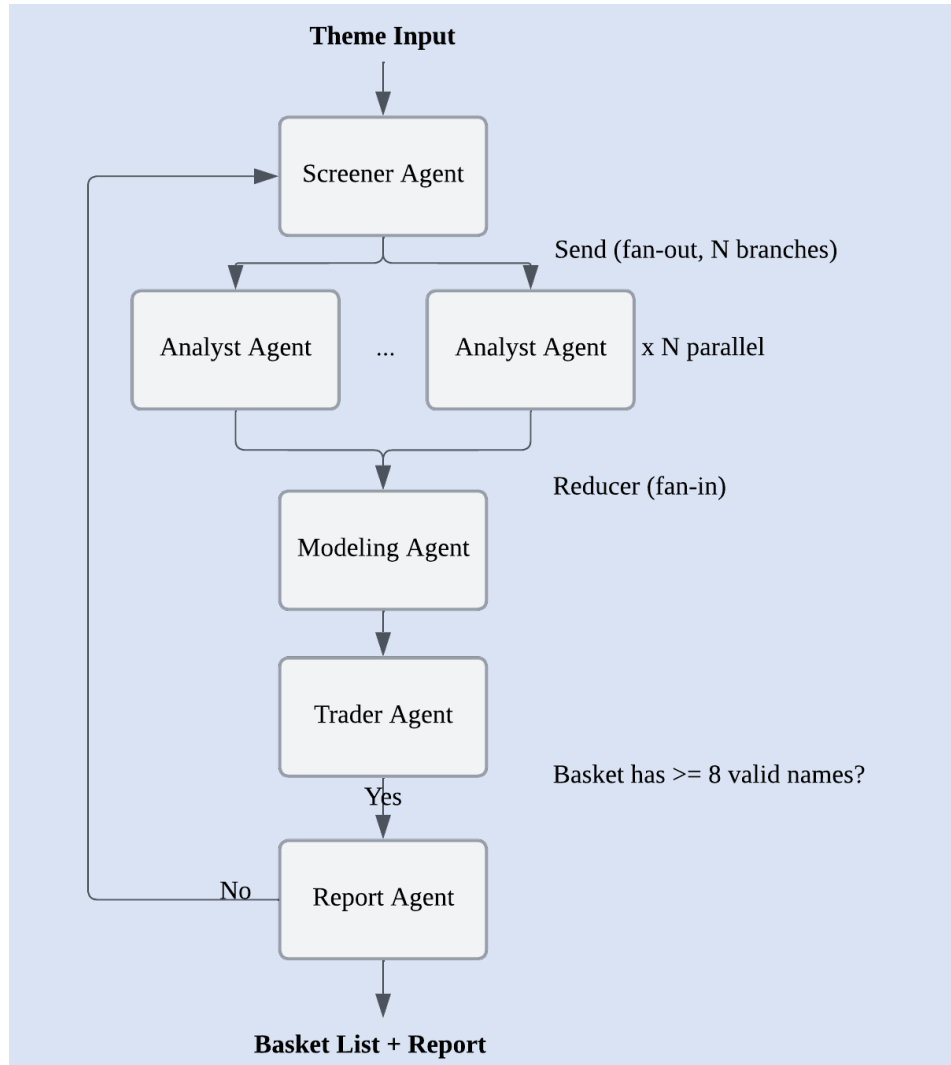
- **Data Management:** This is what makes a stateless API feel intelligent
- **Core intelligence:** Foundation models (LLMs like GPT-style)
- **Responsible for generation and research**
- **Orchestration:** “Glue logic” connecting models to real-world workflows

## Infrastructure Layer:

- Provides compute power, storage, and networking
- Optimized for large-scale AI workloads and parallel processing



# Theme Classifier: LangGraph based Multi-Agent Orchestration Implementation



## 1. Screener Agent

Candidate list (50-150 tickers)

Convert an investment theme into a bounded, verifiable list of candidate tickers.

## 2. Analyst Agent

Per-stock qualitative report + structured scores

Receive one ticker at a time from a pre-screened candidate universe and a theme definition. Then produce a grounded, source-cited analysis.

## 3. Modeling Agent

(a) Factor scoring; (b) Cross-sectional ranking → ranked list

Compute a transparent, reproducible composite score for every stock in the universe and output a ranked table.

## 4. Trader Agent

Basket construction: top 8-10 + weights + constraints

Take the ranked list from the Modeling Agent and construct an 8-10 name basket.

## 5. Report Agent

Rationale document

Write Investment Rationale based on theme definition, Analyst reports, Modeling Agent factor breakdowns.

# Multi-Agent Construction: Prompt Engineering for Role based System Messages

## Screener Agent

SystemMessage:

You are a Universe Screener for a thematic equity research desk. Your job is to convert an investment theme into a bounded, verifiable list of candidate tickers – not to judge or rank them.

Given a theme (e.g., "grid modernization," "GLP-1 supply chain," "onshoring of semiconductor capacity"), you will:

1. Decompose the theme into 3-6 sub-exposures (e.g. for "grid modernization": transmission equipment, smart meters, grid software, utilities capex beneficiaries, battery storage).
2. For each sub-exposure, call the `search\_holdings` and `search\_` tools to pull constituent tickers from relevant thematic ETFs, GICS sector sub-industries, and index membership – do not invent tickers from memory.
3. Deduplicate and output a candidate list of 50-150 US and major international listed equities/ETFs with: ticker, company name, GICS sub-industry, sub-exposure tag, market cap, and average daily dollar volume (for a later liquidity filter).
4. Flag (but do not exclude) any name with ADV under \$5M or market cap under \$300M – the Trader agent will apply the final liquidity screen.

Do not include an investment opinion. Do not rank. Output only the structured candidate table as JSON matching the provided schema. If a sub-exposure returns fewer than 5 candidates, say so explicitly rather than padding the list with loosely related names.

## Design Considerations

- **Role/persona framing** — each system prompt opens by naming a narrow role rather than one generalist agent doing everything. Single-responsibility prompts are more reliable than a mega-prompt.
- **Explicit procedure, not open-ended instruction** — most prompts give numbered steps ("1. Decompose the theme... 2. Call search\_holdings... 3. Output JSON") rather than a vague goal. This is a form of prompt-level chain-of-thought scaffolding that constrains how the model gets to an answer, not just what the answer should look like.
- **Grounding/anti-hallucination constraints** — repeated instructions like "cite a source for every claim," "output null rather than filling gaps," "do not invent tickers from memory." This is prompt engineering specifically aimed at the failure mode LLMs are worst at: confidently inventing facts.
- **Negative constraints as a design tool** — several prompts explicitly forbid something ("do not rank," "do not override the computed rank with your own opinion") specifically to stop an agent from drifting into a job that belongs to a different, more deterministic part of the pipeline.

# Modeling Agent: Specialized Algorithm for Stocks/ETFs Ranking and Selection

## Step A: Factor categories

| Factor       | Example metrics                                           |
|--------------|-----------------------------------------------------------|
| Thematic fit | Analyst agent's 1-5 relevance score, converted to numeric |
| Valuation    | P/E, EV/EBITDA, P/S relative to sector median             |
| Growth       | Revenue growth (TTM, fwd estimate), EPS revision trend    |
| Quality      | ROE, gross margin, debt/EBITDA, FCF conversion            |
| Momentum     | 3/6/12-month price return, RSI, distance from 52wk high   |
| Sentiment    | News/analyst sentiment score (LLM-derived, numeric)       |

## Step B: Normalize

For each factor  $f$ , the cross-sectional mean and standard deviation are computed over the stocks with non-missing values:

$$\mu_f = \frac{1}{|\mathcal{N}_f|} \sum_{i \in \mathcal{N}_f} x_{i,f} \quad \sigma_f = \sqrt{\frac{1}{|\mathcal{N}_f|} \sum_{i \in \mathcal{N}_f} (x_{i,f} - \mu_f)^2}$$

Then the z-score for stock  $i$  on factor  $f$  is:

$$z_{i,f} = \begin{cases} \frac{x_{i,f} - \mu_f}{\sigma_f} & \text{if } i \in \mathcal{N}_f \\ \text{undefined (NaN)} & \text{otherwise} \end{cases}$$

This is applied independently for every factor  $f \in \text{factor\_cols}$ , producing one new column  $z_{i,f}(\{\text{col}\_z\})$  per factor, the function loops over factors.

## Step C: Combine into a composite score

### Method 1: Weighted sum

Given a weight  $w_f$  for each factor (from default weights), the composite score for stock  $i$  is:

$$S_i = \sum_{f=1}^F w_f z_{i,f}$$

with the constraint that weights sum to 1.

### Method 2: Borda rank aggregation

Here you don't combine the z-scores directly — you first convert each factor into a rank, then sum the ranks.

$R_{i,f}$  = rank of  $z_{i,f}$  among  $\{z_{1,f}, \dots, z_{N,f}\}$ , descending

The Borda score is the sum of ranks across all factors:

$$B_i = \sum_{f=1}^F R_{i,f}$$

FinalRank <sub>$i$</sub>  = rank of  $B_i$  among  $\{B_1, \dots, B_N\}$ , ascending

## Step D: Rank

Given composite scores  $S_i$  for  $i = 1, \dots, N$  (from either Method 1 or Method 2 in Step C — I'll just call it  $S_i$  regardless of which produced it).

Let  $\sigma$  be the permutation of  $\{1, \dots, N\}$  that sorts stocks by descending composite score:

$$S_{\sigma(1)} \geq S_{\sigma(2)} \geq \dots \geq S_{\sigma(N)}$$

Then stock  $i$ 's rank is its position in that ordering:

$$\text{Rank}_i = k \quad \text{such that} \quad \sigma(k) = i$$

Equivalently, without needing to name the permutation explicitly, this is a counting formula — how many stocks strictly beat stock  $i$ , plus one:

$$\text{Rank}_i = 1 + |\{j \neq i : S_j > S_i\}|$$



# Evaluation: 3-Level Evaluation for Agent and System Functionality Validation

## Deterministic components (Modeling Agent, Trader Agent)

Unit-testable, standard software testing

- Because two of the five agents are deliberately deterministic, evaluation splits cleanly into two tracks.
- **Unit tests** on `compute_factor_scores`, `combine_scores`, `rank` and other functions against known input/output pairs (including edge cases: all-NaN factor column, tied composite scores, single-candidate universe).
- **Property tests**: composite score must be monotonic in each factor holding others fixed; rank must be a valid permutation of 1..N.
- **Trader constraint tests**: verify sector caps and liquidity floors are never violated in the output basket, across randomized ranked-list fixtures.

## LLM-driven components (Screener, Analyst, Report Agent)

Golden-set + groundedness checks

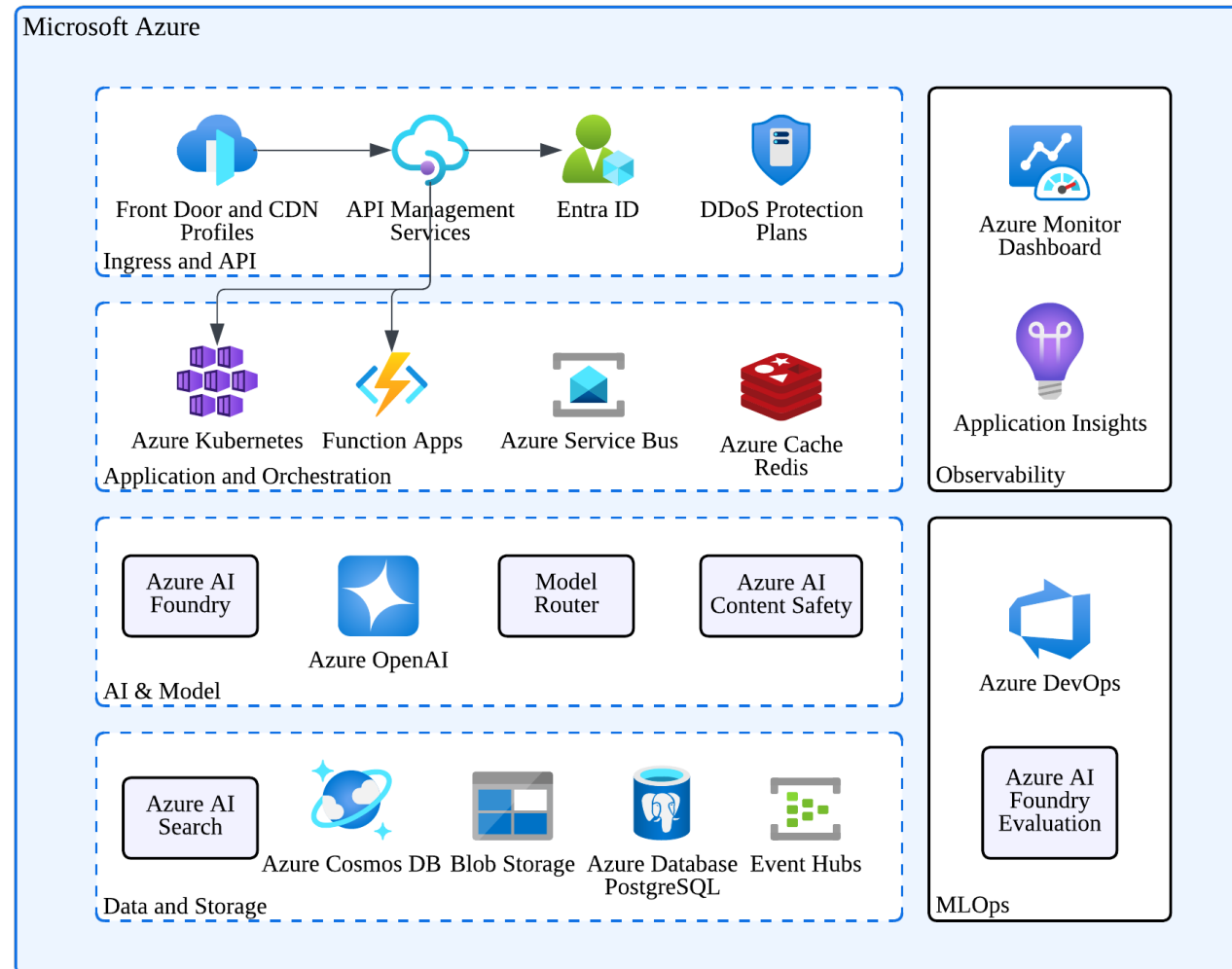
- **Candidate coverage golden set**: pull the top 10 holdings of each theme's mapped ETFs as the "known obvious names" a theme's Candidate Universe should generally include.
- **Basket plausibility check**: verify the final basket has some non-trivial overlap ( $\geq 2$ -3 tickers) with the union of top-15 holdings across a theme's mapped ETFs.
- **Groundedness check on Analyst Reports**: every `sources` entry must correspond to an actual tool-call result recorded in that run's Langfuse trace (catch hallucinated citations).
- **Groundedness check on the Rational Report**: regex-extract every number from `report.md`, build the allowed set from that run's analyst reports, flag any numbers that doesn't match rounding tolerance.

## System-level (The Theme Classifier system performance)

Overall Evaluation Strategy

- **Shadow-mode comparison**: whenever a theme's `config.factor_weights` change reuses that run's already-persisted analyst reports and `factor_panel` rows — no re-fetching, no LLM calls — and reruns only `compute_factor_scores` → `combine_scores(df, new_weights)` → `rank` → basket construction with the proposed weights.
- **Forward-looking performance tracking**: each run's basket gets its forward return tracked in `basket_performance` from date onward, benchmarked against the theme's own mapped reference ETF rather than a flat index or a fixed percentage target — this gives the existing alpha column real, theme-relative meaning.
- **Human review gate**: no theme's basket should be treated as client-facing output until at least one manual review pass.

# Infrastructure Deployment Implementation: Cloud solutions



1. Client Layer uses Azure Static Web Apps to serve SPAs globally from CDN edge nodes, with Azure Front Door providing anycast routing, WAF, and DDoS protection.
2. Ingress & API Gateway is anchored by Azure API Management (APIM), which is the critical hub — it provides token-aware rate limiting, per-tenant quotas, request transformation.
3. Application & Orchestration runs on Azure Kubernetes Service (AKS) hosting LangChain/Semantic Kernel microservices for agent controllers, and prompt builders.
4. AI & Model Layer centres on Azure OpenAI Service (GPT-4o, o1, o3-mini) — all data stays within Azure's boundary with no external API calls.
5. Data & Storage uses Azure AI Search as the primary vector store. Cosmos DB stores conversation history and agent state. Blob Storage is the document ingestion landing zone.
6. Observability & MLOps combines Azure Monitor + Application Insights for distributed tracing across all services with Prompt Flow + Langfuse for LLM-specific tracing, token cost tracking, and prompt versioning.
7. Azure DevOps / GitHub Actions manages CI/CD with Bicep/Terraform IaC.