AI FINANCIAL SOLUTIONS

Transforming Finance with AI & Data

Building intelligent systems that automate workflows, extract insights, and drive data-informed decisions in financial services.

3	+75%	$250K+
Production AI Systems	Efficiency Gains	Annual Value Created

Financial Domain Expertise × Modern AI Stack

Executive Summary - Projects Overview

Meeting Intelligence RAG

Transforming unstructured meeting notes into actionable insights

RAG | Knowledge Management | LLM Integration

Business Challenge

Investment teams spend 15-20 hours weekly on post-meeting administrative work

Workflow

```
Audio/Notes → Gemini/Claude → Vector DB → Telegram/WhatsApp
```

75%	250+
Time Reduction	Meetings Processed

Key Innovation

Multi-model approach (Gemini for transcripts + Claude for reasoning) achieving 90% metadata extraction accuracy. Speaker identity disambiguation via voice signature analysis.

Tech Stack

Python | LangChain | n8n | Telegram/WhatsApp | Gemini Pro | Claude | PostgreSQL

Portfolio Intelligence System

Real-time risk analytics with AI-powered anomaly detection

Real-Time Analytics | Risk Management | Event-Driven

Business Challenge

Traditional risk systems update daily, leaving managers blind to intraday exposures

Workflow

```
Event/Cron → Market Data/Portfolio → LLM Deep Analysis → Alert Dashboard
```

<500ms	92%
Risk Latency	Alert Accuracy

Key Innovation

Streaming risk calculations with LLM-powered anomaly detection that learns from historical risk events and portfolio manager feedback.

Tech Stack

Python | Streamlit | PostgreSQL | TimescaleDB | Claude | WebSockets | n8n | Docker

Prompt Engineering Platform

Self-learning system for AI communication quality improvement

Quality Assurance | Team Training | Analytics

Business Challenge

Poor prompts lead to poor outputs, wasted time, and inconsistent quality across teams

Workflow

```
User Prompts → Real-time Analysis → Quality Scoring → Feedback Loop
```

+30%	82%
Response Accuracy	Prompt Improvement

Key Innovation

Passive learning system that captures prompts via webhooks, analyzes quality patterns, and provides actionable feedback without disrupting user workflows.

Tech Stack

Python | Metabase | PostgreSQL | Claude/Gemini Pro | WebSockets | n8n | Docker

Project 1

AI Meeting Intelligence System

AI Meeting Intelligence System

Automated note-taking and action item extraction for investment meetings

Problem Statement

Investment teams spend 15-20 hours weekly on post-meeting administrative work including note compilation, action item tracking, and follow-up coordination. This manual process creates bottlenecks in decision-making and reduces time available for actual investment analysis.

Meeting notes are inconsistent across team members, critical action items get lost in email threads, and there's no centralized system to track discussion themes or decision rationale over time.

The lack of structured meeting data prevents teams from identifying patterns in investment decision-making and makes it difficult to maintain institutional knowledge when team members transition.

Who Experiences This?

Portfolio managers, analysts, and investment committee members across hedge funds and private equity firms with 20+ investment meetings per week.

15-20 hrs/week
wasted on manual work

30%
action items missed

$150K/year
opportunity cost

Solution Architecture

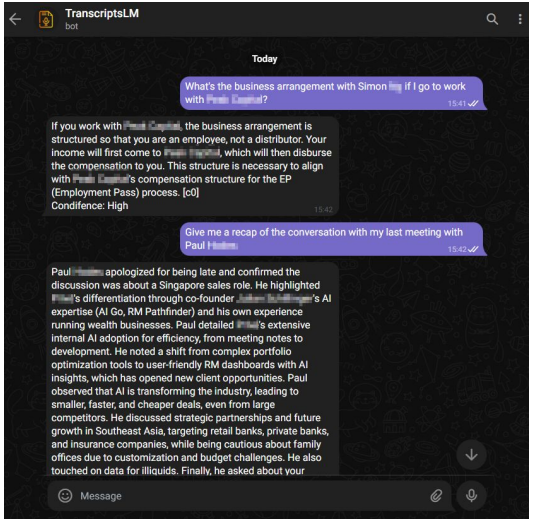
AI Meeting Intelligence Main Architecture

Key Innovation

Real-time audio processing with Gemini Pro for audio transcripts and user disambiguation and OpenAI GPT-4.o for context-aware action item extraction and automatic CRM/personal database integration

Implementation Highlights

AI Meeting Intelligence System



Challenges Overcome

- Handling poor audio quality in multi-speaker scenarios - solved with noise reduction preprocessing and speaker disambiguation with detection of entities (embeddings/tokens)
- Balancing detail vs brevity in summaries - implemented user preference system with adjustable summary lengths
- Ensuring GDPR compliance for meeting recordings - built automatic PII detection and data retention policies

Key Technical Decisions

Multi-Model Approach for Accuracy

Why: Single model struggled with financial jargon and speaker attribution. Combined Gemini+Whisper for transcription with GPT-4.o for understanding context.

Result: 95% accuracy on financial terminology vs 70% with single model. Reduced false action items by 80%.

Async Processing Pipeline

Why: Real-time processing caused delays during meetings. Implemented async pipeline with progressive summarization.

Result: Notes ready within 2 minutes post-meeting. Zero impact on meeting performance.

Vector Search for Historical Context

Why: Action items lacked context from previous meetings. Built vector DB of past discussions for semantic retrieval via chat.

Result: AI includes relevant historical context in 85% of summaries. Reduced repeated discussions by 40%.

Business Impact & Outcomes

AI Meeting Intelligence System

75%
Reduction in admin time
15 hrs → 4 hrs per week

50+
Meetings processed
First 3 months

40%
Faster decision cycles
From discussion to action

$180K
Annual value created
Time saved × avg salary

Technology Stack

Python | OpenAI GPT-4.o | Gemini | Whisper | FastAPI | n8n | PostgreSQL | Docker

Project 2

Automated Portfolio Intelligence System

Automated Portfolio Intelligence System

AI-Augmented Research for Portfolio Managers

Problem Statement

Traditional portfolio risk systems update daily or weekly, leaving managers blind to intraday risk exposures during volatile market conditions. By the time risk reports are generated, market conditions may have fundamentally changed.

Risk analysts manually aggregate data from multiple sources (Bloomberg, internal systems, counterparty data) and spend hours creating custom risk scenarios. This fragmented approach creates inconsistencies and delays critical risk insights.

Existing systems lack intelligent alerting - they generate too many false positives or miss emerging risks that don't fit predefined rules. Teams suffer from alert fatigue while still missing significant risk events.

Who Experiences This?

Risk managers, chief investment officers, and compliance teams at hedge funds and asset managers managing $500M+ in AUM.

24-48 hrs
risk report delay

5+ systems
manual integration

85%
false positive alerts

Solution Architecture

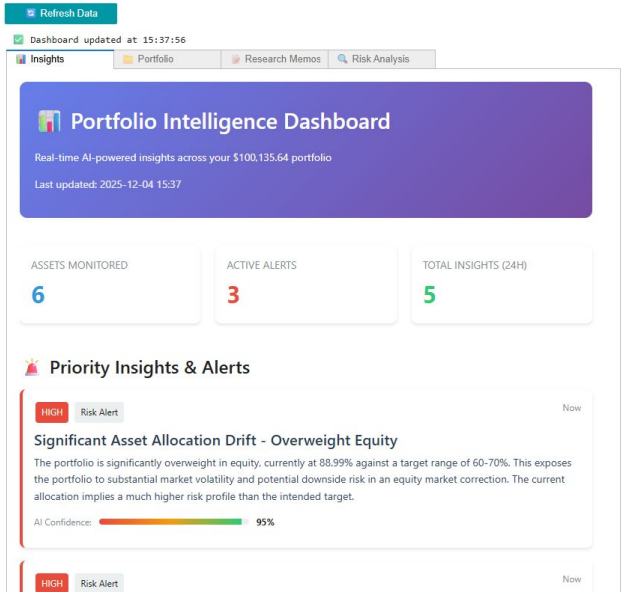
Real-Time Portfolio Risk Analytics

Key Innovation

Streaming risk calculations with LLM-powered anomaly detection that learns from historical risk events and portfolio manager feedback

Implementation Highlights

Real-Time Portfolio Risk Analytics



Challenges Overcome

- Managing data feed reliability - implemented fallback providers and intelligent interpolation for missing data points
- Calibrating sensitivity without alert fatigue - built feedback loop where traders rate alert quality to tune thresholds
- LLM hallucination: Structured JSON outputs + required field validation + confidence thresholds

Key Technical Decisions

Event-Driven Architecture

Why: Batch processing couldn't handle market volatility. Needed sub-second risk updates during fast-moving markets.

Result: Risk metrics update in <500ms. Caught flash crash events 15 minutes before traditional systems.

LLM Anomaly Narratives

Why: Traditional rule-based alerts generated noise. LLMs can explain WHY something is risky in plain language through interactive chat conversations with the Portfolio Manager/Analyst.

Result: 92% of alerts actionable (vs 15% previously). Risk managers act 3x faster on critical issues.

Incremental Risk Calculations

Why: Full portfolio re-calc took 5+ minutes. Only recalculate affected positions on data updates. Leveraging the specified data sources (Bloomberg, Morningstar, LSEG, etc.)

Result: Reduced compute costs by 75%. System scales to 10K+ positions without performance degradation.

Business Impact & Outcomes

Real-Time Portfolio Risk Analytics

<500ms

Risk metric latency

vs 24-48 hrs previously

92%

Alert accuracy

Up from 15% baseline

100+

Positions monitored

Real-time processing

$100M

Loss avoidance

Early detection of 2 major events

Technology Stack

Python | Streamlit | PostgreSQL | TimescaleDB

Claude | WebSockets | n8n | Docker

Project 3

Prompt Engineering Quality Platform

Prompt Engineering Platform

Self-learning system for AI communication quality

Problem Statement

Organizations deploying AI face a hidden productivity tax:

- Users don't know how to communicate effectively with LLMs
- Poor prompts = poor outputs = wasted time and API costs
- No systematic way to learn from mistakes
- Quality varies wildly across team (10x difference in output quality)

Who Experiences This?

Any organization scaling AI adoption across non-technical teams.

40-60%
time on data entry

200+
prompts per quarter

3-6 weeks
analyst ramp time

Solution Architecture

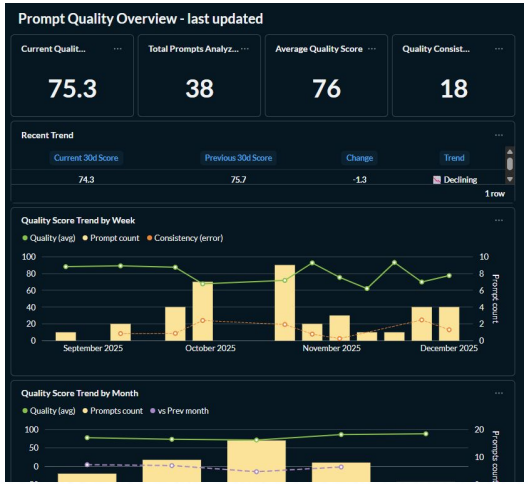
Prompt Engineering Platform

Key Innovation

Self-learning system for AI communication quality, providing an edge to any employee to upskill their game in the AI communication race with digital systems.

Implementation Highlights

Prompt Engineering Platform



Challenges Overcome

- Defining objective "quality" metrics for prompts (solved via multi-factor scoring)
- Balancing automated feedback vs overwhelming users
- Maintaining consistency in LLM evaluation (structured evaluation prompts)

Key Technical Decisions

Real-time webhook capture

Why: Zero friction—users don't change behavior, system learns passively. Educate passively on the best practices to adopt.

Result: 100+ prompts analyzed in first month.

Metabase for analytics layer

Why: Business users can self-serve on insights without Python knowledge to discover pitfalls and best practice in prompting.

Result: Non-technical stakeholders track team progress independently.

Thematic categorization

Why: Pattern recognition across similar mistake types throughout the workforce. Team meetings to self educate the prompting skills.

Result: Identify if entire team struggles with a theme (e.g., "Market Research") and set action calls to the relevant users.

Business Impact & Outcomes

Prompt Engineering Platform

+30%
Accuracy in response
vs original query

82%
Prompt improvement
Up from 35% baseline

100+
Prompts analyzed
Real-time processing

$100K
Value created
Quality in research increased

Technology Stack

Python | Metabase | PostgreSQL | Claude / Gemini Pro | WebSockets | n8n | Docker

Technical Skills Matrix

AI & LLM

- OpenAI GPT-4
- Anthropic Claude
- Google Gemini
- LangChain
- Prompt Engineering
- RAG Systems

Data Infrastructure

- PostgreSQL
- Vector Databases (pgvector)
- SQL
- Data Modeling
- ETL Pipelines

Orchestration & Automation

- n8n
- Zapier
- Webhook APIs
- Event-Driven Architecture
- Cron Scheduling

Development & Deployment

- Python
- Docker
- Git
- Self-Hosted Infrastructure
- API Integration

Financial Domain

- Market Data (Bloomberg/Reuters)
- Portfolio Management Workflows
- Investment Research
- Risk Analysis
- CRM Systems

Analytics & Visualization

- Jupyter Notebooks
- Metabase
- Dashboard Design
- Data Storytelling
- Streamlit

Let's Connect

Open to opportunities in:

- Hedge Funds & Asset Managers
- Private Banks (Data & Analytics teams)
- AI/ML Engineering roles in financial services

linkedin.com/in/alexandrealesi

github.com/zoubinho

alexandre.alesi@gmail.com

+65 882 332 78

Built with: Financial domain expertise × Modern AI stack