Reddit Options Trader (ROT)

Overview

ROT is a real-time market signal system that ingests Reddit discussion, detects momentum, extracts tradable tickers, enriches them with market data, and produces structured trade ideas.

This project was intentionally designed **infrastructure-first**. Models are used only after data identity, timing, and noise constraints are solved.

The system answers one question:

> *Which Reddit discussions are becoming tradable signals right now, and why?*

Not:

> *What should I buy or sell?*

Design Philosophy

1. Infrastructure Before Intelligence

Most trading AI projects fail because they treat the internet as clean data. Reddit is adversarial by default.

Before any ML or reasoning:

- Deduplication
- Identity tracking
- Time-aware deltas
- Symbol validation
- Graceful API failure handling

If these are not solved first, models amplify noise.

System Architecture

Ingestion (PRAW)

Why: Reddit's API structure already encodes social dynamics (score, comments, timing).

Key decisions:

- Use `rising` / `hot` listings (momentum surfaces)
- Snapshot posts repeatedly over time
- Treat each snapshot as a state transition, not a datapoint

Avoided:

- Raw comment text modeling early
- Full historical crawls (low signal-to-noise)

Deduplication & Identity

Reddit reposts, crossposts, stickies, and daily megathreads can poison momentum.

Solution:

- Persistent `seen_posts` state
- Keyed by subreddit + post id
- Deduplication happens before trend scoring

Without this, velocity metrics are meaningless.

Trend Detection (Momentum Engine)

Metric:

- Score delta / time
- Comment delta / time

Why:
Raw counts lag. Deltas expose acceleration.

Threshold-based detection:

- No ML
- Fully explainable
- Deterministic

This makes downstream reasoning auditable.

Ranking

Two parallel rankings:

1. Global Momentum Ranking

Ranks *all* trending discussions.

Purpose:

- Observe macro narrative flow
- Debug momentum math
- Validate ingestion health

2. Ticker-Aware Ranking

Ranks only discussions with **validated tradable symbols**.

This is where real value begins.

Ticker Extraction

Problem: Capital letters ≠ tickers.

Solutions:

- Regex with preference for `$TICKER`
- Alias mapping (SPX → ^GSPC, TSMC → TSM)
- Hard exclusion lists (AI, USD, NATO, etc.)

Design choice:
Extraction before validation. Validation later.

Symbol Validation

Why:
Reddit invents symbols constantly.

Validator responsibilities:

- Equity vs index vs ETF
- Handle delisted / unsupported symbols
- Graceful degradation (never crash pipeline)

Validation errors are signals themselves.

Market Enrichment

Tool: Yahoo Finance

Cached data:

- Price
- Volume
- Earnings calendar
- Basic metadata

Design choice:

- Best-effort enrichment
- Missing data does not block event creation

Markets fail quietly. Systems must not.

Event Construction

An event is:

- A trend
- With validated tickers
- With evidence
- With market context

This creates a stable interface for reasoning layers.

Credibility Scoring

Lightweight heuristic score based on:

- Trend strength

- Market context
- Evidence quality

Not predictive.
Used for ranking and filtering.

Reasoning Layer (DeepSeek / LLM)

Role:

- Summarize context
- Explain why signal matters
- Generate structured packets

Explicitly not used for:

- Price prediction
- Trade execution

LLMs operate downstream, never upstream.

Trade Builder

Consumes reasoning packets + events.

Outputs:

- Hypothesis-style trade ideas
- Time horizon suggestions
- Risk framing

No automation. Human-in-the-loop by design.

Logging & Observability

Everything writes to JSONL:

- Snapshots
- Trend candidates
- Top signals

- Events
- Reasoning
- Trade ideas

This enables:

- Offline analysis
- Replay
- Auditing

What This Project Is Not

- Not a prediction engine
- Not an auto-trader
- Not a backtest toy
- Not ML-first

It is a **signal infrastructure layer**.

Expansion Roadmap

Short-Term

- Historical replay mode
- Comment-level sentiment deltas
- Per-subreddit baselines
- Better ETF/index resolution

Medium-Term

- Cross-platform ingestion (X, Discord)
- Regime-aware scoring
- Latency-aware weighting

Long-Term

- Market impact modeling
- Strategy plug-in layer
- Real-time execution adapters

Final Note

The core insight:

> Alpha systems are infrastructure problems disguised as ML problems.

This project was built to prove that point.