Improving AI Decision Quality: Evaluation Framework for Financial Signal Systems

Automation of real-time predictive trading signals that send text/email alerts

AI-driven systems often generate outputs that appear correct in isolation, yet fail under real-world conditions due to weak reasoning, insufficient context, and poor risk calibration. This project was built to address that gap—focusing not on signal generation alone, but on decision quality under uncertainty.

The process began with a structured evaluation of widely used short-term trading indicators, drawing from those most commonly deployed by professional traders. These indicators were systematically backtested across a targeted universe of assets, including cryptocurrencies, commodities and leveraged ETFs. The objective was clear: develop a risk-adjusted, short-term options decision support system capable of identifying high-probability trade opportunities.

Through iterative testing, the model converged on a refined combination of Relative Strength Index (RSI), MACD, and volume dynamics, optimized for asymmetric trade horizons—approximately 15 days for upside positions and 3 days for downside moves. Initial architecture centered on a standalone application with continuous market scanning and real-time alerting. This evolved into a more efficient and scalable solution using custom Pine Script indicators within TradingView, paired with webhook-based alert delivery via Telegram.

Early results demonstrated strong predictive potential, but also revealed a critical flaw: an unacceptable frequency of false-positive signals. Addressing this required a shift from single-layer signal detection to multi-factor confirmation, introducing a medium-term trend component. The result was a composite oscillator that significantly improved signal quality and materially reduced risk exposure.

🔹 Evaluation Framework – Algo Attempt #1

A simple scoring system was created to evaluate the performance during the limited backtesting against the data for TQQQ and also the crypto currency XLM

Criteria	Description	Score (1–5)
Accuracy	Is the output factually correct?	3
Reasoning	Is the logic sound?	4

Consistency	Does it contradict itself?	5
Risk Awareness	Does it consider downside?	2
Signal Strength	Is this actionable or noise?	4
Total		18 / 25

🔹 Evaluation Framework – Combined Algo #2

The new algorithm is now more consistent and will now be paired with an oscillator indicator to improve accuracy

Criteria	Description	Score (1–5)
Accuracy	Is the output factually correct?	5
Reasoning	Is the logic sound?	5
Consistency	Does it contradict itself?	3
Risk Awareness	Does it consider downside?	4
Signal Strength	Is this actionable or noise?	5
Total		22 / 25

🔹 Applied Examples

Example:

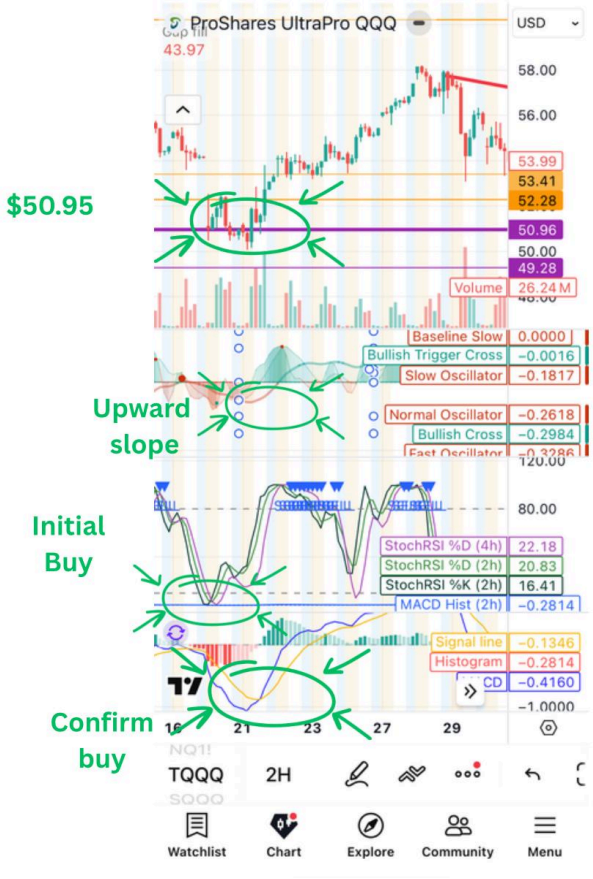
Four charts from top down:

1) Price and Volume 30 minute bars
2) Oscillator indicator
3) Custom indicator
4) MACD

In the picture below we have an example of a “false signal”; a buy signal before hitting another new low. The goal of risk reduction for short-term trades would have failed as an optimal buy,

so the AI found that a 2- hour time frame trigger is better than 30-minute bars for predicting multi-day, directional moves.

AI found the optimal set up below. By shifting to 2-hr bars, initial buy at 0 on the custom indicator, then, a confirm with a MACD cross upward from below,then finally affirmed with an upward sloping oscillator.



Input:

- Custom indicator touches: 0
- Oscillator moves from downward sloping to positive sloping with increasing volume
- MACD cross = MACD cross to the upside from the sell side of the indicator line on the 2-hour chart
- Sell indicator = when MACD histogram ticks down after the 4-hour RSI has crossed through 80 or higher

AI Output:

3 separate email alerts and duplicate text alerts through Telegram, backed up by webhooks within the TradingView platform:

Alert 1: Initial Buy = Custom indicator cross to upside after hitting zero on 2-hour chart

Alert 2: Buy Confirmation = MACD cross to the upside from the sell side of the indicator line on the 2-hour chart.

Alert 3: Sell = when MACD histogram ticks down after the 4-hour RSI has crossed through 80 or higher

---

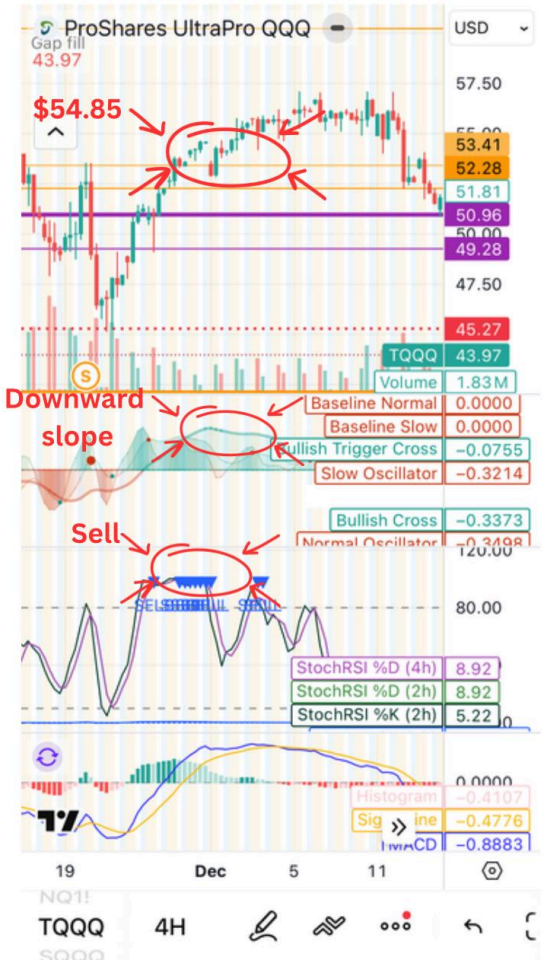
Your Evaluation:

- Buy Accuracy: 5
- Reasoning: 5
- Consistency: 4
- Risk Awareness: 4
- Signal strength: 5
- Final Score: 23 / 25

- Sell Accuracy: 1
- Reasoning: 2
- Risk Awareness: 5
- Reasoning: 4
- Consistency: 1
- Final Score: 12 / 25

Explanation:

- The short-term buy indicators worked very well at predicting profitable oversold conditions for buy opportunities. The addition of the Oscillator indicator improved overall prediction rates by affirming the buy was not a false signal.
- Overbought conditions were, however, completely overestimated; creating needless multiple sell signals which ultimately would forgo profit after a winning buy.
- Bearish momentum is not considered
- No downside risk analysis



The trade went from $50.95 to $54.85 in a 5-day time horizon with minimal risk. Multiple false sell signals. To hold longer would have gained 5% more in two more days; though with additional risk. A $2.95 profit in 5 days.

🔹 Section 4 — Failure Mode Analysis

Example Failure Types:

- Overconfident outputs on the sell side
- Ignoring conflicting signals (sideways trading, indecisiveness)
- Hallucinated reasoning

These failure modes mirror LLM hallucinations and ranking errors

---

🔹 Section 5 — Improvement Strategy

- Automate combining the two indicators avoiding human evaluation of the Oscillator indicator
- Penalize conflicting indicators and perform a thorough backtest after harnessing the historical data for the assets. Build a research agent that can perfect the indicators on a running basis after the daily close from the Alpaca API
- Require minimum confidence thresholds for the sell side indicator. Look at shorting strategies in addition to long-only buys, once perfected.
