KINETIC ALPHA RESEARCH · AI INVEST

AI Tool Diligence

A plugin for evaluating AI-driven investment tools, the strategies they generate, and the portfolios they actually produce.

What this is

AI Tool Diligence turns the five-part Kinetic Alpha research series on AI in wealth management into something repeatable. The scoring dimensions, the autonomy taxonomy, the control-surface analysis and the deflation mathematics all come from that work; the plugin operationalizes them as six skills you run inside Claude. It is a diligence instrument, not an advice engine: it classifies structure and disclosure risk, and it states plainly what cannot be known from public information.

It runs as a Claude Code / Cowork plugin — it is not a standalone application. Installation and use are covered on the last page.

The six skills

`screen-offer`	**Fast triage of an advertised offer** Paste an ad, a URL, or a Telegram invite. Routes on the custody question first, runs the registration lookups that actually matter for that class of product, scores tiered red flags, and returns a risk verdict with what is unknowable stated plainly. Target: five minutes.
`evaluate-platform`	**Documentary teardown of a named platform** Pulls Form ADV, Form CRS, BrokerCheck, terms of service and disclosure pages; places the product on a five-level autonomy ladder; scores six dimensions of documented investor protection; and tests marketing claims against filings.
`deflate-claim`	**Stress-test a performance number** Applies the deflated Sharpe ratio, minimum backtest length and break-even trial count, then works a presentation-defect checklist. Extends into implementability when the user has an actual strategy rather than a statistic.
`evaluate-portfolio`	**Decompose an actual book** Upload holdings as CSV or paste a position list; the engine regresses every holding on a factor set, rebuilds the covariance, and returns factor exposure, variance contributions, correlation clusters, concentration and independent bets. A position is not an atom — it expands into weighted exposures to named factors, and the value is discovering that positions which look independent are the same bet.

`monitor-portfolio`	**Re-run the decomposition on a cadence** Keyed to the strategy's rebalancing frequency; reports what drifted against both the original anchor and the prior run. Alerts are tiered and deterministic. Where the tool made a specific commitment — a volatility target, a rebalancing frequency — its own promise becomes the threshold.
`review-registry`	**Resolve past verdicts and recalibrate** Revisits past evaluations, marks each falsifiable prediction confirmed or refuted, and proposes concrete changes to the scoring. This is the skill that makes the others improve — the calibration record is the only honest basis for trusting the verdicts.

The measurement that justifies the portfolio engine

AI tools produce thematically coherent baskets. A tool asked for "AI infrastructure exposure" returns names spanning technology, industrials and utilities: diversified on a sector pie chart, one bet in reality. Standard diversification measures cannot see this.

On synthetic data with known structure, a single-factor bet and a genuinely diversified control both returned an effective position count of 8.0, and effective N by risk — the measure most systems stop at — returned 8.0 against 7.52. In a drift test where a diversified book was rebalanced into one theme, effective N by risk **rose**, from 7.56 to 7.60, while the book went from 5.74 independent bets to 2.23. What separated them was independent bets (1.36 vs 5.88) and the largest single risk source (95% of portfolio risk vs 44%). Those two numbers lead every portfolio output, and both are computed from the trace-normalised eigen-spectrum of the risk-scaled position matrix.

The feedback cycle

Every evaluation appends a record carrying a verdict and at least one falsifiable prediction with a review date. review-registry later checks what happened, marks each prediction confirmed or refuted, and proposes changes to the reference files — a flag's tier, a rubric anchor, a weight. Scoring changes are never silent: each is logged with the evidence that justified it. The point is the refutations. An agent that only logs its wins has learned nothing.

The calibration that matters most

Absence from registration databases is the default and expected state for most advertised investing products. Newsletters, screeners and general-audience signal services are lawfully outside the adviser perimeter. Absence becomes damning only alongside custody or discretion — so every skill routes on that question first.

What it will not do

It will not tell you a product is safe, recommend an investment, or estimate returns. It will not recommend trades, allocations or securities from a decomposition — that output describes exposure over a stated window; it does not prescribe or forecast. It will not claim to have determined whether a product's "AI" is real; no external observer can settle that from public information, and where that is the question the honest answer is "unknowable, and here is what would have to be true."

Installing and running it

The download is a **.plugin** bundle — a Claude Code / Cowork plugin. It runs inside Claude; there is no separate app to launch.

1 · Install the bundle in Claude Code or Cowork via your plugin settings (or `/plugin` in Claude Code), pointing at the downloaded ai-tool-diligence.plugin file.

2 · The six skills then trigger in natural language — "screen this offer," "evaluate PortfolioPilot," "deflate this Sharpe ratio," "decompose my portfolio," "monitor it monthly," "review the registry."

3 · The portfolio engine needs daily adjusted price history, which the caller supplies — the two Python scripts (decompose.py, drift.py) are deterministic and take no network access. Requires Python 3 with numpy.

Structure: skills/ (six) · references/ (lookup-playbook, red-flags, scoring-rubric, deflation-math, presentation-defects, registry-schema, house-standard, portfolio-schema, risk-decomposition, monitoring-protocol, market-data) · scripts/ (decompose.py, drift.py, factors-singlename.json).

Companion research: the five-part AI in Wealth Management series at kineticalpha.com/ai-invest — the autonomy ladder, reading the filings, where the return shows up, the enforcement layer, and validation. MIT-licensed. Not investment, legal or compliance advice.