

Value Investing Platform — Functional Specification v1.0

Status: Design baseline. Written market-agnostically; UK is the first instantiated market, Japan the intended second.

Scope: Single user, single portfolio. No multi-user, permissions, or sharing.

Purpose of this document: Define *what* the system does and *why*, at a level that constrains implementation without dictating it. Technical building blocks are named as categories only.

0. Founding premises

Every design decision below descends from four conclusions reached before this spec was written. If a future change contradicts one of these, the premise should be revisited explicitly rather than eroded silently.

1. **Edge is structural, not analytical.** Analytical capability is commoditising. What persists is access to places institutional capital cannot go: too small, too illiquid, too foreign, too unaggregated. The system's job is to point attention there.
2. **Numbers from a database, judgment from the model, decision from the human.** Each layer has a fixed role. Numeric data is never sourced from an LLM. Judgment is never outsourced entirely. The final call is never automated.
3. **The forward decision record replaces backtesting.** A concentrated, multi-year, judgment-driven strategy produces too few independent decisions for historical validation, and its inputs cannot be reconstructed point-in-time. The journal is the validation mechanism.
4. **Approximately right beats precisely wrong.** If a thesis requires a model to work out, it is too complicated. Coarse ranges, decomposed scores, no composite ranking.

1. System overview

Seven functional components, in process order. Components 1–4 narrow; 5–6 verify and decide; 7 records and reviews.

# Component	Nature	POC substitute
1 Market & Data Source Registry	Reference data, human-curated	Markdown/Excel in project folder

# Component	Nature	POC substitute
2 Philosophy & KPI Directory	Reference data, versioned rules	Excel + markdown
3 Investment Compass	Scored filter, two levels	Excel worksheet
4 Screener	Mechanical, deterministic	Excel + manual data pull
5 Knowledge Base	Document store + extraction	Project folder (PDFs) + chat context
6 Review Cockpit	Human + AI, structured workflow	This conversation thread
7 Decision Journal & Watchlist	Append-only record	Excel, one row per entry

Deferred (named, not designed): position sizing, portfolio monitoring, sell discipline, cycle temperature dial. See §9.

2. Component 1 — Market & Data Source Registry

Purpose. Answer, per market: where does data come from, how good is it, and can an agent reach it?

Functional requirements.

- One record per market (country/exchange). Records are additive; markets are never deleted, only deprioritised.
- Per market, a list of data sources with: type (regulator filing archive, exchange feed, commercial screener, company IR site), coverage (which universe, which years), format (structured/XBRL, HTML, PDF), cost, and access method.
- **Agent-access annotation** is a first-class field, not a footnote: authentication requirements, rate limits, robots/ToS position, whether automated retrieval is permitted, and known failure modes. A source a human can browse but an agent cannot use must be marked as such.
- **Data quality flag** per source, on a coarse scale, with the specific known defects noted (stale share counts, mistranslated line items, missing delisted entities).
- Explicit note on **survivorship**: does this source retain delisted/acquired entities?

Design note. This registry is the practical expression of the "unaggregated, not unavailable" rule. A market whose filings exist but are unaggregated is a target. A market whose information does not exist is out of scope.

3. Component 2 — Philosophy & KPI Directory

Purpose. Convert the value-investing literature into a small set of computable metrics and a small set of checklists. It is a rules artefact, not a research library.

Hard constraint. The directory's output is roughly **15 computable metrics** and **3–4 checklists**. It is not a repository of everything great investors have said. Scope creep here is the single most likely way this project stalls.

Functional requirements.

- **Metric definitions**, each with: formula, required input fields, unit, and the source line items it depends on. Metrics must be computable from filing data without judgment.
- **Anchor definitions for valuation.** Three anchors are required, not one: (a) the company's own long-run history, (b) current peer multiples, (c) **M&A / control transaction comparables**. The third is routinely absent from commercial screeners and is central to markets where takeover is the gap-closure mechanism.
- **Normalisation conventions**, notably: balance-sheet items expressed in **days**, measured against the company's own multi-year history rather than a sector table.
- **Versioning.** Thresholds and definitions change. Every metric definition carries a version, and every journal entry records which version it was evaluated under. Without this, historical decisions become uninterpretable.

Initial metric set (starting point, not canon). Cheapness: EV/EBIT, P/B, FCF yield, net-cash-to-market-cap. Quality: gross profitability, ROE/ROIC, F-score components, gross margin stability. Safety: net debt/EBITDA, interest cover, cash conversion (FCF vs. reported earnings), accruals ratio. Alignment: share count trend, shareholder yield, insider transactions. Liquidity: ADV.

4. Component 3 — Investment Compass

Purpose. Direct attention to the less-trodden ground. Two distinct levels, deliberately separated because they run at different cadences and on different data.

4.1 Market level (quarterly cadence, slow-moving)

Scored dimensions per market:

- Rule of law and enforcement quality
- Minority-shareholder treatment
- **Presence of a gap-closure mechanism** — activism, takeover activity, regulatory pressure on undervaluation, index reform. A market that is cheap with no mechanism is a value trap at national scale.
- Data obtainability (from Component 1)
- Platform accessibility (broker reach)
- Aggregate cheapness — **ranked last, deliberately**

Explicit warning encoded in the design: country-level cheapness is often sector composition or governance discount in disguise, and cheapness without a closure mechanism can persist for decades before ending badly.

4.2 Company level (every screen run)

Deterministic fringe filters: market capitalisation ceiling, average daily volume band, analyst coverage count, index membership (exclusion is a positive signal), free float.

Design note. The company-level filter encodes the core premise: we want candidates that are *physically unusable* by large funds. This is the filter that must not be relaxed for convenience.

5. Component 4 — Screener

Purpose. Reduce the market universe to a reading queue. Mechanical, deterministic, and deliberately loose.

Functional requirements.

- Runs one screen definition across all active markets simultaneously. **No per-market quotas.** If output is 30 candidates in one market and 3 in another, that distribution is information.
- Hard filters from the Compass (company level) plus 3–4 hard metric filters. Remaining metrics used for **ordering only**.
- **No composite score.** Output carries decomposed dimension scores (cheapness / quality / safety / alignment) side by side. Ranking orders the reading queue and does nothing else.

- Output stored with a **run timestamp and the screen definition version used**, so the candidate set is reproducible and comparable across runs.
- **Every screen number is provisional.** Screen output is a candidate list, never a fact. Any number used in a decision must be verified against a primary document (enforced in Component 6).
- **Industry clustering** on output. Industry mental models are expensive to build and cheap to reuse; presenting candidates grouped by industry means the second and third name in a sector cost a fraction of the first.

Second entry path. Special situations (spin-offs, post-bankruptcy equities, forced index selling, announced strategic reviews, insider-buying clusters) will never surface on a fundamental screen. An **event feed** is a parallel input to the same downstream pipeline, not a separate process.

6. Component 5 — Knowledge Base

Purpose. Hold primary documents and their extracted summaries. Lazy, tiered, permanently cached.

Three-tier fetch policy — bulk ingestion of a market's filings is explicitly rejected as the fastest way to kill the project:

Tier	Trigger	Fetch	Volume
1	Screen run	No documents. Structured data only.	~50 names
2	Entering kill pass	Latest annual + 2 recent interims	~30 names
3	Survived kill pass	5–10 years of filings	~10 names

Functional requirements.

- Blob storage for raw documents (PDF/HTML), never overwritten.
- For each document: source URL, retrieval timestamp, fiscal period covered, entity ID, document type.
- **Extraction layer:** AI-generated standardised one-pager per document — financials in a consistent shape, plus flagged items (auditor change, related-party transactions, dilution events, going-concern language, customer concentration). Stored alongside the raw document as a semi-structured record, never in place of it.

- The extraction is a *reading aid*. The raw document remains the authority for any number entering a decision.

Entity model. Keyed on **ISIN**, not ticker — tickers collide and change across venues. Every stored figure carries currency, fiscal-period end, and source document ID from the first record onward. Retrofitting either is materially painful.

7. Component 6 — Review Cockpit

The heart of the system, and the only component where the human is irreplaceable. Five passes in fixed order.

Pass A — Industry model (per industry cluster, reusable)

Build or retrieve a mechanistic model of how the industry works: capital cycle, supply/demand dynamics, cost structure, disruptors, typical margin and working-capital behaviour.

- This is AI-led. It is *scaffolding*, not edge — the macro layer of any industry is well-analysed and freely available, and no advantage is claimed there.
- **Verification requirement:** the model must be triangulated against checkable sources — competitor filings, trade data, industry press. With a zero baseline, fluent-and-wrong is indistinguishable from fluent-and-right unless checked.
- Cached and versioned per industry. Cost amortises across all candidates in the cluster.
- **Preference rule:** if a mispricing is only visible through a sophisticated industry model, it is outside the circle of competence regardless of who built the model.

Pass B — Kill pass (inversion)

Explicitly seeking reasons to **discard**, not reasons to buy. Target: eliminate roughly 70% of Tier-2 candidates. Budget ~20 minutes per company using the extracted one-pager, verified against the filing.

Runs the **Rule Repository** (§7.1).

Also asks the two structural questions:

- **"Why did this company survive?"** A sub-scale survivor in a scale industry must have something structural — a niche, a captive customer, a regional advantage. If it cannot be answered from the filings, the company is not understood.
Caution: "family-controlled and nobody forced consolidation" is persistence, not a moat.

- **"Has the entity changed materially through M&A?"** If so, as-reported history may be fiction and a pro-forma reconstruction is required before any historical comparison is valid.

Pass C — Anomaly bucket classification (forced choice, required field)

Every survivor is classified into exactly one bucket. Free text is not permitted; the discipline is in being unable to leave it vague.

Bucket	Meaning	Requirement to proceed
Unlooked-at	No coverage, no story, illiquid	A named catalyst, or explicit acceptance of indefinite patience
Misunderstood	Market sees it and is wrong	A defensible variant view, evidenced
Cheap for reason	Decline, governance abuse, trapped cash	Discard

This classification *is* the thesis. It determines whether a gap-closure mechanism exists at all.

Pass D — Independent assessment

A separate assessment thread, given **facts only — not the thesis**. Financials, filing excerpts, ownership structure, price. Asked to produce its own view and, separately, to build the strongest case that this is a value trap.

Known limitation, recorded deliberately: a fresh context removes conversational contamination (accumulated framing, commitment to a prior position, drift toward agreement) but *not* correlated model error. Same weights, same priors, same systematic blind spots. This yields an independent reading, not an independent reasoner. Where practical, use a different model family. Do not treat agreement between passes as confirmation.

Pass E — Human inversion and decision

The human attacks the assessment. Default outcome is **discard**; the burden of proof sits on the buy case.

Kill criterion rule — protects against rhetorical drift: a candidate is killed only when an objection is (a) reduced to a falsifiable claim, (b) checkable against a primary document, and (c) verified. A model conceding under persistent pressure is not evidence. *"The niche might not be durable"* is a worry. *"Segment note shows largest customer at 40% of revenue, acquired by a competitor last year"* is a kill.

Structured judgment prompts (the Marks layer) applied to survivors:

- What does the consensus believe about this company, and what is the evidence for that belief? (Price-target distribution, consensus estimates vs. observed run-rate, coverage count — evidence, not assertion.)
- Why is it wrong?
- What am I paying versus conservative value?
- What would kill this thesis?
- **What am I explicitly giving no credit for?** (Upside deliberately marked at zero.)
- **What are the limits of my competence here?** An honest disclosure field; also the too-hard-pile trigger.

7.1 Rule Repository

A growing checklist applied during Pass B. Modelled on Pabrai's practice of encoding others' documented mistakes.

Structure per rule:

- Rule statement
- **Provenance** — the specific documented failure it derives from. Generic prudence is weak; *"Buffett/Dexter Shoe: foreign low-cost competition destroyed a moat believed durable"* is applicable because it describes what the failure looked like from the inside.
- Type: deterministic (computable) or judgment (AI-assessed)
- **Output is three-way: pass / fail / cannot-determine.** Not a numeric score. "Could not establish who controls this company" is a finding, not a gap.
- Evidence field and confidence level for judgment rules
- Version and date added

Judgment rules are legitimate. Management track record, shareholder alignment, integrity signals, scandal history — these are assessed non-deterministically and partially subjectively, exactly as a human analyst would, and are no less admissible for it. The requirement is that the assessment carries its evidence.

Growth policy. Start at roughly 20–30 genuinely fatal rules (leverage, dilution history, related-party transactions, cash conversion, auditor changes, control structure, pension deficit, customer concentration). Grow primarily from **your own errors**, recorded via the journal review loop. A borrowed 200-item checklist that is not understood is theatre, and 200 rules × 50 candidates is an unusable workload.

8. Component 7 — Decision Journal & Watchlist

The record that replaces backtesting. Everything upstream exists to produce entries here.

8.1 Entry schema (frozen at decision time)

Modelled on the VIC header block, which places conclusions before evidence and forces commitment:

- Upside %, **with the derivation method named**
- Downside %, **with the derivation method named** — external, non-self-serving anchors preferred (e.g. the most bullish sell-side target, or liquidation value)
- Duration / expected time to resolution
- Catalyst
- Key risks
- Reference statistics: price, 52-week range, market cap, net leverage, **ADV**
- Anomaly bucket (from Pass C)
- **Scenario ladder** — 3–4 discrete cases with per-share values, replacing any point estimate
- **Pre-committed kill criteria** — specific, primary-document-checkable conditions that would end the position
- Circle-of-competence disclosure
- **Input snapshot** — the screen figures, metric versions, and rule-repository version in force at decision time
- The independently generated bear case (Pass D output), stored in full

8.2 Immutability and review

- **Append-only.** The original entry is never edited. Each quarterly review is a dated entry appended beneath it.
- Reviews are scored **against the pre-committed kill criteria**, not as fresh evaluations. Fresh looks drift — the thesis quietly gets rewritten to fit the price.
- Review structure: which criteria still hold, which broke, what is new that was not anticipated, verdict (hold / add / trim / exit).
- The diff between entry one and entry eight is the readable record of whether judgment was right, wrong, or moving the goalposts.

8.3 Error profile

Periodic aggregation across closed decisions to identify **clustered** errors — systematic over-trust of cash-rich balance sheets, under-weighting of dilution, mistaking cyclical peaks for stable earnings. Clustered errors feed back as new rules in §7.1. This loop is the system's only genuine learning mechanism.

8.4 Watchlist

For the most common outcome: *good company, wrong price*. Holds the thesis, a target entry price, and a re-check trigger. Prevents rediscovering the same company from scratch in two years, and means that when a market dislocates, the target list already exists.

8.5 Coverage ledger

What was screened, when, under which definitions, and with what outcome. Prevents re-reading the same company every quarter and makes the funnel's attrition visible.

9. Deferred components

Named so the journal schema accommodates them; not designed here.

- **Position sizing** — inputs: scenario ladder, conviction, liquidity (ADV), correlation to existing holdings.
 - **Portfolio monitoring** — trigger re-check on new filings, price moves through thresholds, or breach of a pre-committed kill criterion.
 - **Sell discipline** — thesis broken / fair value reached / better opportunity available. Pre-committed, not decided in the moment.
 - **Cycle temperature dial** — a market-level aggressiveness setting (credit spreads, IPO volume, covenant quality, retail participation, narrative temperature) answering one quarterly question: deploy eagerly, normally, or hoard cash. A screen on *when*, not *what*.
-

10. Anti-scope — deliberately not built

Recorded with reasons so these are not rediscovered expensively.

Excluded	Reason
Backtesting engine	A concentrated multi-year strategy yields ~8–15 independent decisions per decade — no statistical power. Judgment inputs

Excluded	Reason
	cannot be reconstructed point-in-time. Documented anomaly literature supplies the priors; the journal supplies the validation.
LLM as numeric data source	No knowledge timestamps, no as-reported guarantee, hallucination risk, not reproducible. Disqualifying for anything entering a decision. LLMs read; databases and filings supply numbers.
Composite scoring	Precision theatre. A single blended number invites trust it has not earned. Dimensions stay decomposed and ordinal.
Pattern-mining loop over historical filings	Look-ahead contamination is a property of model weights and cannot be prompted away. Post-cutoff filings restore ignorance but only at real time. Unbounded hypothesis space makes multiple-testing correction uncomputable.
Fine-tuning on historical outcomes	Noisy labels, contaminated data, and it hides the learned rules where they cannot be inspected. Explicit written rules give the same compression with visibility.
Technical/price-based signals	Crowded by better-resourced participants; no structural edge available.
Execution and order management	Broker handles it. No advantage to rebuilding.
Real-time price feeds	Multi-year horizons do not require intraday data.
Global coverage at launch	Point-in-time and aggregation problems are unsolved outside a handful of markets. One market, done properly.
Macro / country-timing bets	Fully visible to everyone with better information. Structural edge is bottom-up only.

Standing rule against over-engineering. The thin version — a candidate table, a manually run screen, one text file per thesis — already captures most of the value. Every component beyond that must earn its place by removing work actually being done, not work imagined. The most likely failure of this project is a beautiful system that is never used.

11. POC configuration

Target-state block	POC substitute
Cloud-hosted application	None. Manual process.
HTML user interface	A dedicated conversation thread
Relational store (entities, metrics, screen runs)	Excel workbooks in the project folder
Document/object store (extractions, assessments)	Markdown files in the project folder
Blob store (filings)	Project folder — PDFs
Scheduled screen runs	Manual, on demand
Event feed	Manual monitoring

First market: United Kingdom. Rationale: filings in English and in a legal framework the user can read critically; assessable governance regime; confirmed gap-closure through observed takeover activity at premiums; no currency layer; structurally neglected AIM/small-cap segment. Japan scores marginally better on raw statistics and has a stronger institutional catalyst, but the user's ability to read documents critically outweighs the difference at proof-of-concept stage.

Known UK-specific traps to encode as rules: pension deficits exceeding market capitalisation; AIM liquidity adequate for entry and punishing on exit.

POC success criterion. Not "did it find a winner" — that is unknowable for years. The test is: **does the process produce a decision the user would actually act on, with a written thesis they could defend and pre-committed criteria they could be held to?** If the surviving candidates feel unactionable, the fault is upstream and better found before anything is built.

12. Open questions

1. Metric coverage and quality for UK small caps at the target size band — resolvable only by attempting the first screen.
2. Whether the Pass D independent assessment produces materially different conclusions in practice, or merely restates Pass B. If the latter, the pass needs redesign or removal.

3. Initial screen thresholds — deliberately unset. To be calibrated against first-run output volume rather than chosen in the abstract.
4. Industry cluster granularity — how narrowly to define an industry before the model stops being reusable.
5. Whether the event-feed path warrants its own kill-pass variant, since special situations fail conventional quality screens by construction.