Multi-Source Thematic Scoring for Systematic Index Construction

Abel Riboulot
Agora Intelligence*

Working Paper · April 2026

Abstract

We present a methodology for constructing thematic equity baskets by fusing six heterogeneous information sources: temporally decayed neural-embedding search over financial news, three LLM-derived company-profile signals extracted from SEC 10-K filings—descriptions, business areas, and thematic exposures—and two additional sources: investment themes distilled from quarterly earnings call transcripts and themes from analyst reports, each maintained in separate vector stores. An eleven-parameter scoring model aggregates these signals into a composite relevance score per company, optimized via CMA-ES under five-fold cross-validation over 81 investment themes with 60,000 total optimization rounds. Quality is measured by an LLM-based evaluation oracle that rates each selected company's thematic relevance on a 1–10 scale (9–10: core pure-play, 7–8: clearly relevant, 5–6: tangential, 1–4: weak or absent). A blinded pairwise study with four domain experts—portfolio managers and quantitative analysts—shows that the oracle's directional preferences are statistically indistinguishable from human inter-rater agreement (92.5% directional agreement with majority vote, 0% flips on unambiguous pairs for three of four evaluators). A cross-model evaluation using Claude Sonnet 4.6 (Anthropic) as an independent oracle confirms that the relevance rankings are not model-specific: Spearman $\rho = 0.853$ across 42,227 company–theme pairs, and baskets optimized under GPT-5 score 8.37/10 under Claude vs. 8.51/10 under GPT-5. Under this human-validated metric, the system achieves an out-of-sample mean of 8.36/10 (95% CI: [8.04, 8.68]), meaning the average company in a generated basket is rated "clearly relevant" to "core pick" by both the oracle and human judges. The overfit gap is 0.13 points, and the system outperforms the production neural-network baseline on 68 of 81 themes (+23.7% relative improvement). Against consensus thematic-ETF holdings, our top-20 baskets share 66% of holdings on average while contributing 34% unique selections, confirming both market-consensus alignment and additional breadth from multi-source signals. At runtime the system requires no LLM calls: each theme query completes in ∼1 s on a single core, and queries are embarrassingly parallel.

1 Introduction

Thematic investing allocates capital to companies that stand to benefit from a specific structural trend—be it a technology (gene editing, autonomous vehicles), a demographic shift (aging population), or a macro regime (inflation protection, USD strength). The central challenge is to correctly identify the public equities most exposed to a given theme, enabling investors to gain targeted thematic exposure efficiently and at scale.

Our approach works in two phases:

*Correspondence: abel@tilt.io. The author thanks the domain experts who participated in the pairwise evaluation study.

Offline pre-processing. We build structured, point-in-time *company cards* from every SEC 10-K filing in our coverage universe. A large language model reads the Business Overview (Section 1) and Management Discussion & Analysis (Section 7) of each filing and extracts three fields: a neutral company description, a list of business areas, and a list of investment themes the company has positive exposure to. Each of these three text fields is embedded independently into a 1024-dimensional vector space using the Voyage 4 Large embedding model, producing three separate vector collections in a Qdrant vector database. Simultaneously, a corpus of Dow Jones / Wall Street Journal news articles is embedded via the same model and stored in a fourth vector collection. Additionally, investment themes are extracted from quarterly earnings call transcripts and analyst reports via a separate LLM pipeline, embedded with the same model, and stored in two further vector collections—yielding six independent signal sources in total. Each card is dated by its filing date, enabling point-in-time analysis without look-ahead bias.

Runtime scoring. When a user submits a theme query (e.g., "CRISPR"), the query is embedded and used to search all six collections via approximate nearest neighbor (ANN) retrieval. For news articles, each retrieved article carries metadata linking it to mentioned companies (via ISIN identifiers), and a time-decay function exponentially down-weights older articles. The per-company news signal is aggregated via a power mean with a breadth amplification term. For the three card-based collections, the cosine similarity between the query embedding and each company's card embedding yields five additional signals. These six signals are combined through a weighted sum with cross-source interaction terms into a single composite score, and the top-10 companies form the thematic basket. Because all retrieval results are pre-indexed and scoring is pure arithmetic on cached vectors, the entire pipeline completes in approximately one second per theme on a single core—with no LLM calls at inference time. Queries are embarrassingly parallel: scoring 81 themes simultaneously requires no shared state, making throughput linear in available cores.

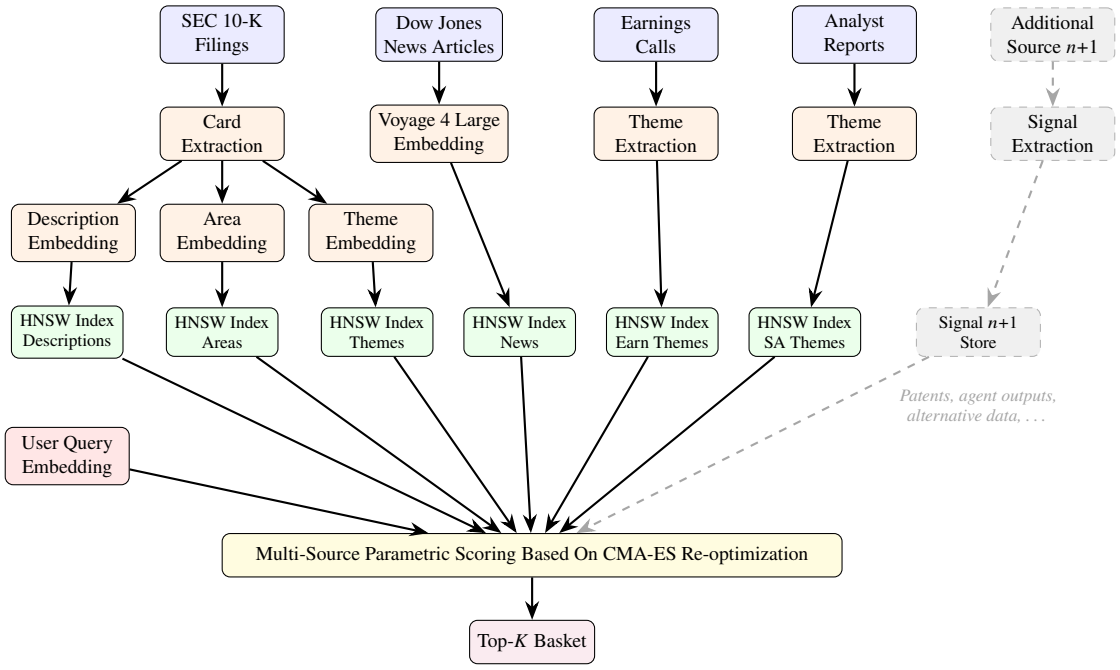

Figure 1: High-level system architecture. Offline pre-processing (top) produces six vector collections from the current data sources. At runtime (bottom), a user query is embedded and matched against all collections; results are aggregated by a parametric scoring model into a ranked basket. *Dashed elements* illustrate the extensibility path: any additional signal source (patents, agent-driven outputs, alternative data) can be integrated by adding a retrieval channel and re-running the CMA-ES optimization—which completes in minutes—to learn the new source's optimal weight.

The eleven free parameters of the scoring model (Section 4)—thresholds, source weights, and interaction coefficients—are optimized using a derivative-free optimizer (CMA-ES, Section 5) under 5-fold cross-validation over a universe of 81 investment themes. The evaluation oracle is a frontier LLM (GPT-5) that rates each selected company's relevance to the query theme on a 1–10 scale using its SEC filing context. This score is cached in a relational database and reused across the thousands of optimization rounds, making each round computationally inexpensive. The oracle is validated through three independent channels: a blinded pairwise study with four domain experts (Section 6.7), a cross-model evaluation using Claude Sonnet 4.6 as an independent oracle (Section 6.8), and comparison against consensus ETF holdings (Section 6.6).

A key advantage of this architecture is its extensibility. The scoring model is designed so that adding a new signal source—whether embedding-based, agent-driven, or arbitrary—requires only three steps: (i) implement a retrieval or computation function that produces a per-company score for a given theme, (ii) add a corresponding weight parameter to the CMA-ES parametrization, and (iii) re-run the optimization. Because each optimization round is a pure arithmetic operation on cached data ($\sim$10 ms per round), the full k-fold cross-validation with 60,000 rounds completes in approximately 105 minutes on a single CPU core. The cost of testing whether a new signal improves basket quality is therefore negligible.

Candidate additional signals include patent filings (embedding patent claims against themes), earnings call transcripts (extracting forward-looking management commentary), sell-side analyst reports, alternative data (satellite imagery, web traffic, app downloads), and *agent-driven signals* where an autonomous AI agent performs multi-step research for each (company, theme) pair—e.g., browsing the company's website, reading recent press releases, and synthesizing a relevance judgment. Because the CMA-ES framework treats each signal as a black box producing a scalar score per company, the internal complexity of the signal source is irrelevant to the optimization. The parametric scoring model automatically learns the optimal weight for the new source relative to existing ones, including any cross-source interaction effects, without manual tuning.

The remainder of this paper is organized as follows. Section 2 describes the data sources and pre-processing. Sections 3 and 4 detail the embedding infrastructure and scoring model, including the extensible multi-source framework and multi-theme blending. Section 5 presents the optimization procedure and evaluation oracle. Section 6 reports validation results: cross-validation performance, baseline comparison, ETF consensus analysis, human pairwise evaluation, and cross-model agreement. Section 7 discusses re-ranking experiments. Section 8 covers implementation details, fixed parameters, and computational costs. Section 9 discusses limitations.

2 Data Sources and Pre-Processing

The system consumes four primary data sources: structured corporate disclosures (SEC 10-K filings), unstructured financial news (Dow Jones / Wall Street Journal articles), quarterly earnings call transcripts, and analyst reports. Each source is pre-processed into structured representations suitable for embedding and retrieval.

2.1 Company Card Generation from SEC 10-K Filings

For every annual 10-K filing in our coverage universe, we extract two sections from the SEC EDGAR database:

- **Section 1 (Business Overview):** A description of the registrant's business, products, services, and competitive landscape.
- **Section 7 (MD&A):** Management's Discussion and Analysis of financial condition and results of operations.

Each section is truncated to the first 6,000 characters to fit within the LLM context window while retaining the most salient content (regulatory boilerplate tends to appear later).

2.1.1 LLM Extraction

A frontier LLM (GPT-5) extracts a structured JSON profile from each filing, containing a neutral company description (2–4 sentences), a list of business areas (typically 2–8), and a list of investment themes representing positive exposure (5–30, depending on company breadth). The prompt enforces factual, non-promotional language grounded in the filing text and restricts themes to positive exposures (i.e., tailwinds, not risks). Full prompt text is reproduced in Appendix F.

2.1.2 Output Schema and Storage

The LLM response is parsed as JSON and stored in a PostgreSQL table with the following schema:

Table 1: Schema of the `company_cards` table (point-in-time company profiles).

Column	Type	Description
`filing_id`	`TEXT UNIQUE`	SEC filing identifier (primary dedup key)
`ticker`	`TEXT`	Stock ticker at filing date
`tilt_id`	`TEXT`	Internal asset identifier
`company_name`	`TEXT`	Registrant name
`filing_date`	`DATE`	Date the 10-K was filed with the SEC
`description`	`TEXT`	LLM-extracted neutral description
`business_areas`	`JSONB`	Array of business segments
`themes`	`JSONB`	Array of positive thematic exposures
`raw_llm_response`	`TEXT`	Full LLM output for audit trail
`created_at`	`TIMESTAMP`	Processing timestamp

2.1.3 Point-in-Time Property

Each company card is dated by its `filing_date`. When constructing a basket as of date t, only cards with `filing_date` $< t$ are used. This prevents look-ahead bias in backtesting. Multiple filings for the same company (across years) yield multiple cards; at query time, the most recent card before t is selected per ticker via a `DISTINCT ON (ticker) ...ORDER BY filing_date DESC` query.

2.2 News Article Corpus

The news corpus comprises articles from the Dow Jones Newswires and Wall Street Journal, stored in a Qdrant vector database. Each article record includes:

- **headline**: Article title.
- **content**: Full article text.
- **date**: Publication date (ISO 8601).
- **tilt_source**: Provenance (e.g., "WSJ", "DJNW").
- **companies**: Array of structured company mentions, each with an `isin` (International Securities Identification Number) and an `about` boolean indicating whether the company is a subject of the article (as opposed to merely mentioned).

Articles are embedded with two named vectors: `headline_vector` and `content_vector`, both 1024-dimensional, using the Voyage 4 Large model. At retrieval time, the `content_vector` is used for semantic search against the theme query embedding.

2.3 Earnings Call Transcripts

We process 14,919 quarterly earnings call transcripts from `tilt__data.quarterly_earning_call_transcript`, covering 4,487 unique tickers with transcripts dating from January 2009 to March 2026 (average content length: 41,374 characters). For each transcript, an LLM (GPT-4o-mini) extracts 5–25 investment themes representing the company's positive exposures as discussed by management. Each (company, theme) pair is embedded via Voyage 4 Large and stored in `earnings_themes_v4`, producing **241,738** vector points. The most recent 4 transcripts per company are processed, providing approximately one year of quarterly coverage.

2.4 Analyst Reports

We process 21,959 analyst reports covering 5,511 unique tickers with publication dates from October 2004 to March 2026 (average content length: 8,606 characters). The same LLM extraction pipeline produces 5–25 themes per article, yielding **313,858** vector points in the analyst report themes collection. The most recent 5 analyses per company are processed.

2.5 Theme Extraction Pipeline

Both new sources (earnings calls and analyst reports) use an identical extraction pipeline:

1. **Document loading.** The most recent N documents per company are fetched from PostgreSQL, filtered to documents with >500 characters of content. Documents are truncated to 6,000 characters for LLM input.

2. **LLM theme extraction.** GPT-4o-mini extracts 5–25 investment themes per document, guided by a system prompt that enforces positive-exposure semantics. Themes are returned as lowercase 2–5 word labels.

3. **Embedding.** Each (company, theme) pair is embedded via Voyage 4 Large and upserted to the target Qdrant collection with deterministic UUIDs.

The full extraction and embedding pipeline processes all 36,878 documents and produces 555,596 Qdrant points in approximately 62 minutes.

3 Embedding and Retrieval Infrastructure

3.1 Embedding Model

All text embeddings are produced by the **Voyage 4 Large** model, which maps variable-length text into $\mathbb{R}^{1024}$ with cosine similarity as the native metric. The model supports two input types:

- `document`: Used when embedding corpus texts (card descriptions, business areas, themes, article content). Optimized for the "to be retrieved" side.
- `query`: Used when embedding the user's theme query. Optimized for the "retrieval intent" side.

This asymmetric embedding protocol follows the dense passage retrieval paradigm introduced by Karpukhin et al. [9], where separate encoders for queries and documents improve retrieval quality over symmetric embedding.

Formally, for a text string x, the embedding function is:

$$\mathbf{e}(x) = \textsc{Voyage4Large}(x) \in \mathbb{R}^d, \quad d = 1024, \quad \|\mathbf{e}(x)\|_2 = 1. \tag{1}$$

Because embeddings are ℓ_2-normalized, cosine similarity reduces to the inner product [1]:

$$\text{sim}(q, d) = \cos\big(\mathbf{e}_q(q),\, \mathbf{e}_d(d)\big) = \frac{\mathbf{e}_q(q) \cdot \mathbf{e}_d(d)}{\|\mathbf{e}_q(q)\| \cdot \|\mathbf{e}_d(d)\|} = \mathbf{e}_q(q) \cdot \mathbf{e}_d(d), \tag{2}$$

where $\mathbf{e}_q$ and $\mathbf{e}_d$ denote the query and document encoders respectively. The cosine similarity is bounded: $\text{sim}(q, d) \in [-1, 1]$, with sim = 1 indicating identical direction in the embedding space.

3.2 Vector Database Architecture

The system maintains six vector collections indexed using the Hierarchical Navigable Small World (HNSW) graph algorithm [8], each with cosine distance and 1024-dimensional vectors.

HNSW constructs a multi-layer proximity graph where each layer is a navigable small-world graph with exponentially decreasing density. At construction time, each new point $\mathbf{v}$ is inserted into the graph by greedily traversing from the top layer downward, connecting $\mathbf{v}$ to its M nearest neighbors at each layer ℓ where $\ell \leq \lfloor -\ln(U) \cdot m_L \rfloor$ ($U \sim \text{Uniform}(0, 1)$, m_L is the level generation factor). At query time, the search begins at the top layer and greedily descends, performing beam search with width *ef* at the bottom layer. The algorithm achieves $O(\log N)$ search complexity for N indexed vectors with high probability, providing sub-millisecond retrieval even at millions of points [8].

Table 2: Qdrant vector collections used by the scoring system.

Collection	Granularity	Embedded Text
`company_cards_descriptions_v4`	1 per filing	Full company description (2–4 sentences)
`company_cards_areas_v4`	1 per (filing, area)	Individual business area string
`company_cards_themes_v4`	1 per (filing, theme)	Individual theme exposure string
`tilt_dow_jones_archive_articles_v4`	1 per article	Two named vectors: `headline_vector`, `content_vector`
`earnings_themes_v4`	1 per (transcript, theme)	Investment theme from earnings call
`analyst_report_themes_v4`	1 per (article, theme)	Investment theme from analyst research

The three SEC card-based collections are populated from the same underlying company cards but at different levels of granularity. A single company card with 5 business areas and 15 themes produces 1 description point, 5 area points, and 15 theme points. This fan-out ensures that fine-grained thematic labels (e.g., "CRISPR") are individually retrievable even when the full company description emphasizes broader aspects. The earnings and analyst report collections similarly store one point per extracted theme per document.

Point identifiers are deterministic UUIDs generated via `UUID5(NAMESPACE_URL, key_string)`, where the key string is `desc:{filing_id}`, `area:{filing_id}:{area}`, or `theme:{filing_id}:{theme}`. This ensures idempotent upserts: re-running the embedding pipeline produces identical point IDs and overwrites in place.

Each point carries a structured payload including `tilt_id` (the internal asset identifier), `ticker`, `company_name`, `filing_date`, and the embedding model version.

3.3 Retrieval at Query Time

Given a user's theme query q, the system computes $\mathbf{e}_{\text{query}}(q)$ and issues six parallel ANN searches:

1. **News search**: Query the `content_vector` field of the news collection with limit $L_{\text{news}} = 1{,}500$. Returns tuples (s_i, C_i, t_i) where s_i is the cosine similarity score, C_i is the set of companies mentioned in article i, and t_i is the publication date.
2. **Description search**: Query the descriptions collection with limit $L_{\text{card}} = 500$. Returns tuples $(s_j, \texttt{tilt\_id}_j)$.
3. **Area search**: Same protocol on the areas collection.
4. **SEC theme search**: Same protocol on the themes collection.
5. **Earnings theme search**: Same protocol on the earnings themes collection.
6. **SA theme search**: Same protocol on the analyst report themes collection.

The retrieval limits are intentionally generous to ensure high recall; filtering and ranking are performed in the scoring stage described next.

4 Scoring Model

The scoring model transforms raw retrieval results into a single composite relevance score S_c for each candidate company c. The model has eleven free parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{11})$ that are optimized as described in Section 5.

Table 3: Optimizable parameters of the scoring model (weighted mode).

Parameter	Symbol	Range	Role
`min_search_score`	τ_s	$[0.15, 0.80]$	Minimum cosine similarity for news articles
`score_power`	p	$[0.5, 5.0]$	Exponent for power mean aggregation
`breadth_coeff`	β_b	$[0.0, 0.5]$	Log-breadth amplification factor
`min_card_score`	τ_c	$[0.1, 0.7]$	Minimum cosine similarity for card signals
`w_news`	w_n	$[0.0, 2.0]$	Weight for news signal
`w_desc`	w_d	$[0.0, 2.0]$	Weight for description similarity
`w_area`	w_a	$[0.0, 2.0]$	Weight for business area similarity
`w_theme`	w_t	$[0.0, 2.0]$	Weight for theme similarity
`w_earn`	w_e	$[0.0, 2.0]$	Weight for earnings theme similarity
`w_sa`	w_{sa}	$[0.0, 2.0]$	Weight for analyst report theme similarity
`interaction_strength`	γ	$[0.0, 2.0]$	Cross-source synergy coefficient

4.1 News Signal Computation

For each candidate company c and theme query q, the news signal N_c is computed as follows.

4.1.1 Article Filtering and Time Decay

Let $\mathcal{A}$ be the set of articles returned by the news retrieval step. For each article $i \in \mathcal{A}$ with cosine similarity s_i and publication date t_i:

1. **Similarity threshold**: Discard if $s_i < \tau_s$.

2. **Temporal window**: Discard if $\text{age}_i > T_{\max}$ or $\text{age}_i < 0$, where $\text{age}_i = (t_{\text{query}} - t_i)$ in days and $T_{\max} = 8 \times 365.25 = 2{,}922$ days.
3. **Exponential decay**: We model information relevance as exponentially decaying with time, following the standard half-life formulation. The decay factor is:

$$\delta_i = \exp\left(-\frac{\ln 2 \cdot \text{age}_i}{T_{1/2}}\right) = 2^{-\text{age}_i/T_{1/2}}, \quad T_{1/2} = 4 \times 365.25 = 1{,}461 \text{ days}. \tag{3}$$

   The equivalence $e^{-\lambda t} = 2^{-t/T_{1/2}}$ holds with decay constant $\lambda = \ln 2/T_{1/2}$. The half-life $T_{1/2}$ satisfies $\delta(T_{1/2}) = \frac{1}{2}$, meaning an article published exactly 4 years before the query date receives weight $\delta_i = 0.5$. At 8 years (the lookback maximum), $\delta_i = 0.25$. The exponential form ensures the decay is memoryless: the proportional decrease in weight over any fixed time interval is constant, i.e., $\delta(t+\Delta)/\delta(t) = 2^{-\Delta/T_{1/2}}$ independent of t.
4. **Effective score**: $\tilde{s}_i = s_i \cdot \delta_i$.

4.1.2 Company Association

Article i contributes to company c if and only if company c appears in the article's `companies` metadata with `about = true` and a valid ISIN. The ISIN is resolved to an internal asset identifier (`tilt_id`) via the security master database.

Let $\mathcal{A}_c \subseteq \mathcal{A}$ denote the filtered articles that mention company c as a subject. Define:

$$n_c = |\mathcal{A}_c|, \tag{4}$$

$$B_c = \sum_{i \in \mathcal{A}_c} \delta_i \quad \text{(decay-weighted article count)}. \tag{5}$$

4.1.3 Power Mean Intensity

The *intensity* of the news signal for company c is computed as a generalized (power) mean—also known as the Hölder mean or L^p mean—of the effective scores [2, 3]:

$$I_c = M_p(\tilde{\mathbf{s}}_c) = \left(\frac{1}{n_c} \sum_{i \in \mathcal{A}_c} \tilde{s}_i^{\,p}\right)^{1/p}, \tag{6}$$

where p is the `score_power` parameter. The power mean family is a one-parameter generalization that continuously interpolates between classical means. Its limiting behavior is well-characterized:

$$\lim_{p\to-\infty} M_p = \min_i \tilde{s}_i, \tag{7}$$

$$M_{-1} = \text{harmonic mean} = \frac{n_c}{\sum_i \tilde{s}_i^{-1}}, \tag{8}$$

$$\lim_{p\to 0} M_p = \text{geometric mean} = \left(\prod_i \tilde{s}_i\right)^{1/n_c}, \tag{9}$$

$$M_1 = \text{arithmetic mean} = \frac{1}{n_c}\sum_i \tilde{s}_i, \tag{10}$$

$$M_2 = \text{quadratic mean (RMS)} = \sqrt{\frac{1}{n_c}\sum_i \tilde{s}_i^2}, \tag{11}$$

$$\lim_{p\to+\infty} M_p = \max_i \tilde{s}_i. \tag{12}$$

By the power mean inequality [2], M_p is monotonically non-decreasing in p: for $p_1 < p_2$, $M_{p_1} \le M_{p_2}$ with equality if and only if all $\tilde{s}_i$ are equal. In our parameter range $p \in [0.5, 5.0]$, the optimizer can thus smoothly interpolate between a near-geometric mean (emphasizing consistency across articles) and a near-maximum (emphasizing peak article relevance). This is a critical modeling choice: themes with sparse but highly relevant coverage benefit from higher p, while themes with broad but moderate coverage benefit from lower p.

4.1.4 Breadth Amplification

Companies mentioned in more articles receive a breadth bonus, inspired by the term frequency saturation principle underlying sublinear TF scaling in information retrieval [18]:

$$N_c = I_c \cdot \big(1 + \beta_b \cdot \ln(1 + B_c)\big), \tag{13}$$

where β_b is the `breadth_coeff` parameter and B_c is the decay-weighted article count. The $\ln(1 + \cdot)$ transform is the softplus variant of the logarithmic TF scaling $1 + \ln(\text{tf})$ used in classical TF-IDF, ensuring:

- Sublinear growth: $\frac{\partial N_c}{\partial B_c} = I_c \cdot \beta_b/(1 + B_c) \to 0$ as $B_c \to \infty$.
- No singularity at zero: $\ln(1 + B_c)$ is defined and equals zero when $B_c = 0$, unlike $\ln(B_c)$.
- Multiplicative structure: the breadth term *amplifies* the intensity rather than adding to it, so a company with high breadth but low intensity does not receive an inflated score.

4.2 Card Signal Computation

For each of the five card-based collections $k \in \{\text{desc}, \text{area}, \text{theme}, \text{earn}, \text{sa}\}$, the retrieval step returns a set of tuples $(s_j^{(k)}, \texttt{tilt\_id}_j)$. The per-company card signal $V_c^{(k)}$ is computed as:

$$V_c^{(k)} = \max_{j:\texttt{tilt\_id}_j=c,\ s_j^{(k)} \ge \tau_c} \big(s_j^{(k)} - \tau_c\big). \tag{14}$$

That is, for each company c in collection k, we take the maximum similarity score among all points belonging to that company that exceed the threshold τ_c, then subtract τ_c so that the signal starts from zero at the boundary. If no point exceeds the threshold, $V_c^{(k)} = 0$. The max-pooling ensures that a company with one highly relevant business area is not penalized for having other unrelated areas.

4.3 Multi-Source Score Aggregation

The six signals N_c, $V_c^{(\text{desc})}$, $V_c^{(\text{area})}$, $V_c^{(\text{theme})}$, $V_c^{(\text{earn})}$, $V_c^{(\text{sa})}$ are combined into a composite score via a weighted sum with interaction terms:

Step 1: Weighted card sum.

$$W_c = w_d \cdot V_c^{(\text{desc})} + w_a \cdot V_c^{(\text{area})} + w_t \cdot V_c^{(\text{theme})} + w_e \cdot V_c^{(\text{earn})} + w_{\text{sa}} \cdot V_c^{(\text{sa})}. \tag{15}$$

Step 2: Card synergy. Let $\mathcal{V}_c = \{v : v > 0,\ v \in \{V_c^{(\text{desc})}, V_c^{(\text{area})}, V_c^{(\text{theme})}, V_c^{(\text{earn})}, V_c^{(\text{sa})}\}\}$ be the set of positive card signals. If $|\mathcal{V}_c| \geq 2$, sort $\mathcal{V}_c$ in descending order as $(v_{(1)}, v_{(2)}, \ldots)$. Then:

$$W_c \leftarrow W_c + v_{(1)} \cdot v_{(2)} \cdot \gamma. \tag{16}$$

This multiplicative interaction rewards companies that appear across multiple card collections, reflecting the intuition that a company whose description, business areas, listed themes, earnings commentary, *and* analyst research all match the query is more likely to be genuinely relevant than one matching on a single dimension.

Step 3: Combined score.

$$S_c = w_n \cdot N_c + W_c + \gamma \cdot N_c \cdot W_c. \tag{17}$$

The final term $\gamma \cdot N_c \cdot W_c$ is a cross-source interaction: companies that score well on *both* news and card signals receive a multiplicative boost. When $\gamma = 0$, the model reduces to a pure additive combination.

This functional form can be understood as a truncated second-order Taylor expansion of a general scoring function $g(N_c, W_c)$ around the origin:

$$g(N_c, W_c) \approx g(0,0) + \left.\frac{\partial g}{\partial N_c}\right|_0 N_c + \left.\frac{\partial g}{\partial W_c}\right|_0 W_c + \left.\frac{\partial^2 g}{\partial N_c \partial W_c}\right|_0 N_c W_c + \cdots \tag{18}$$

where $g(0,0) = 0$ (no signal, no score), the linear coefficients correspond to w_n and the implicit weight 1 on W_c, and the mixed partial γ captures the interaction effect. This is equivalent to the factorization machine formulation of Rendle [14] restricted to two feature groups, where the bilinear term models pairwise interactions without requiring a full quadratic expansion. The factored structure (a single γ for the cross-term rather than separate coefficients for $N_c \cdot V_c^{(\text{desc})}$, $N_c \cdot V_c^{(\text{area})}$, etc.) reduces the parameter count and prevents overfitting.

Basket construction. Companies are ranked by S_c in descending order, and the top-K (default $K = 10$) form the thematic basket.

Optimized parameter values. Table 4 reports the parameter values obtained after the full CMA-ES optimization described in Section 5. The relative magnitudes are economically interpretable: company description similarity ($w_d = 1.58$) receives the highest weight among all sources. The two new sources—earnings themes ($w_e = 1.18$) and analyst report themes ($w_{\text{sa}} = 1.22$)—receive material weight comparable to the SEC card signals, confirming they carry complementary thematic information not redundant with existing sources. The power mean exponent $p = 4.41$ (well above the quadratic mean) indicates the optimizer strongly prefers emphasizing peak article relevance over average coverage. The interaction strength $\gamma = 1.85$ is notably large—nearly double the value observed in a four-source variant of this system—indicating that six-source cross-corroboration is substantially more valuable. Full parameter stability analysis is reported in Section 6.2.3.

Table 4: CMA-ES optimized parameter values (final retrain on all 81 themes).

Parameter	Symbol	Value	Role
min_search_score	τ_s	0.369	Article cosine similarity threshold
score_power	p	4.408	Power mean exponent for article scores
breadth_coeff	β_b	0.388	Article count amplification
min_card_score	τ_c	0.139	Card cosine similarity threshold
w_news	w_n	0.665	News signal weight
w_desc	w_d	1.578	Description card weight
w_area	w_a	1.022	Business area card weight
w_theme	w_t	0.989	SEC theme card weight
w_earn	w_e	1.182	Earnings theme weight
w_sa	w_{sa}	1.224	Analyst report theme weight
interaction_strength	γ	1.852	Cross-source synergy

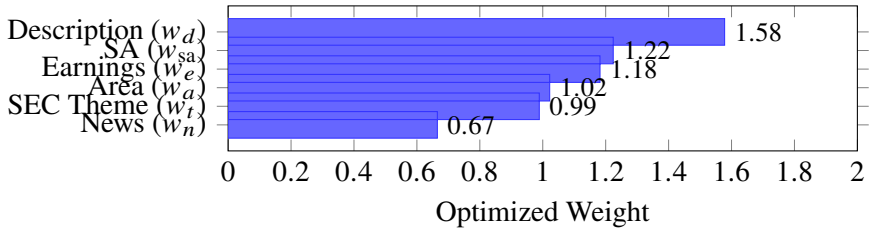

Figure 2: Learned source weights from CMA-ES optimization. Company description cards receive the highest weight, followed by the two new sources (analyst reports and earnings themes). News receives the lowest individual weight but contributes uniquely through the interaction term.

4.4 Alternative: Reciprocal Rank Fusion (RRF)

An alternative scoring mode uses Reciprocal Rank Fusion (RRF), originally proposed by Cormack et al. [7], with only 5 parameters, replacing the 4 source weights and interaction strength with a single constant k_{rrf}:

$$S_c^{\text{rrf}} = \sum_{k \in \{\text{news,desc,area,theme}\}} \frac{\mathbb{1}[c \in \mathcal{R}_k]}{k_{\text{rrf}} + \text{rank}_k(c)}, \tag{19}$$

where $\mathcal{R}_k$ is the set of companies ranked by source k, $\text{rank}_k(c)$ is the 1-based rank of company c within source k, and $k_{\text{rrf}} \in [1, 100]$ controls the importance of high ranks versus broad presence. Companies absent from a source receive no contribution from that source.

The key property of RRF, shown empirically by Cormack et al. [7], is that it is robust to outlier rankings and does not require score normalization across sources—only ranks matter. The constant k_{rrf} controls the trade-off between top-rank dominance and rank-breadth: as $k_{\text{rrf}} \to 0^+$, the contribution of rank-1 items dominates ($1/1 \gg 1/r$ for $r > 1$); as $k_{\text{rrf}} \to \infty$, all ranks contribute approximately equally (since $1/(k + r) \approx 1/k$ for all $r \ll k$). The original paper recommends $k_{\text{rrf}} = 60$; we allow the optimizer to search $k_{\text{rrf}} \in [1, 100]$.

The RRF mode is more parsimonious (5 vs. 9 parameters) but sacrifices the ability to differentially weight sources and lacks interaction terms. In our experiments, the weighted mode consistently outperformed RRF by 0.3–0.5 points on out-of-sample relevance, primarily because the interaction term γ captures cross-source synergies that rank-only fusion cannot express. All results reported in Section 6 use the weighted mode.

4.5 Generalized Extensible Scoring Framework

The scoring model described above uses six signal sources, but the architecture is explicitly designed to accommodate an arbitrary number of sources. We present the generalized formulation that governs signal integration and describe the procedure for incorporating new signals.

4.5.1 Generalized Composite Score

Let $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ be a set of n signal sources, where each source s_k produces a per-company score $\phi_c^{(k)} \geq 0$ for a given theme query. The current system has $n = 6$ sources (news, description, area, theme, earnings, SA), but the framework generalizes to any n. The composite score becomes:

$$S_c = \sum_{k=1}^{n} w_k \cdot \phi_c^{(k)} + \gamma \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{1}\left[\phi_c^{(i)} > 0 \wedge \phi_c^{(j)} > 0\right] \cdot \phi_c^{(i)} \cdot \phi_c^{(j)}, \tag{20}$$

where w_k is the learned weight for source k and γ is a shared interaction coefficient. When a new source s_{n+1} is added, the parametrization expands by one weight w_{n+1} (and optionally a source-specific threshold τ_{n+1}), increasing the CMA-ES search dimension by 1–2. Since CMA-ES scales well up to $d \approx 100$ [5], and each additional source adds only 1–2 dimensions, the system can accommodate dozens of sources without degrading optimization quality.

The total number of optimizable parameters for n sources is:

$$d(n) = \underbrace{n}_{\text{weights}} + \underbrace{n_\tau}_{\text{thresholds}} + \underbrace{1}_{\gamma} + \underbrace{n_{\text{news}}}_{\text{news-specific}}, \tag{21}$$

where $n_\tau \leq n$ is the number of sources requiring a similarity threshold and n_{news} covers news-specific parameters (power, breadth). For the current $n = 6$ system, $d = 11$. Adding a seventh source would yield $d = 13$ at most.

4.5.2 Signal Source Taxonomy

Signal sources fall into three categories, each with different integration patterns:

Type 1: Embedding-based signals. These are the most natural to add. A new corpus (e.g., patent claims, earnings call transcripts, analyst reports) is embedded with the same Voyage 4 Large model and stored in a new HNSW-indexed collection. At query time, the theme query embedding is used to search the new collection, and cosine similarity scores are aggregated per company using the same max-pooling-with-threshold mechanism as the card signals (Equation 14). Examples:

- **Patent embeddings**: Embed patent abstracts; search for theme-relevant patents; attribute to companies via assignee metadata.
- **Earnings call embeddings**: Embed Q&A sections of quarterly earnings calls; capture forward-looking management commentary on thematic exposure.
- **Sell-side research**: Embed analyst report summaries; leverage expert domain analysis.

Type 2: Structured/quantitative signals. Some signals are not embedding-based but produce a numeric score per company directly. These bypass the retrieval step entirely and feed scores directly into the aggregation. Examples:

- **Revenue exposure**: Fraction of a company's revenue derived from the thematic segment (from financial databases).
- **R&D intensity**: Ratio of R&D spending to revenue in theme-relevant divisions.
- **Alternative data**: Web traffic to theme-related product pages, app download growth, hiring patterns in relevant job categories.

These signals are normalized to $[0, 1]$ and enter the composite score with their own learned weight w_k.

Type 3: Agent-driven signals. The most expressive signal type uses autonomous AI agents that perform multi-step research for each (company, theme) pair. An agent might:

1. Search the company's investor relations website for theme-relevant product pages.
2. Read recent press releases and earnings call highlights.
3. Cross-reference with industry reports and competitor analysis.
4. Synthesize a structured relevance judgment with reasoning.

Agent-driven signals are the most expensive to compute (seconds to minutes per company) but can capture information unavailable in any static corpus. Because the CMA-ES optimization operates on cached scores, agent outputs need only be computed once per (company, theme) pair—the same caching infrastructure used for the LLM evaluation oracle (Section 5.1) applies. The optimization then learns whether and how much to trust the agent signal relative to cheaper embedding-based signals.

4.5.3 Integration Procedure

Adding a new signal source s_{n+1} to the production system requires:

1. **Implement the signal function**: Write a function $\phi^{(n+1)} : (\text{company}, \text{theme}) \to \mathbb{R}_{\geq 0}$ that produces a relevance score. For embedding-based signals, this involves creating a new vector collection, populating it, and implementing a retrieval function. For agent-driven signals, this involves defining the agent's research protocol and output schema.

2. **Pre-compute and cache scores**: Run the signal function for all (company, theme) pairs in the evaluation universe. This is the most expensive step but is amortized: once cached, scores are reused across all optimization rounds.

3. **Extend the parametrization**: Add w_{n+1} (and optionally τ_{n+1}) to the CMA-ES parameter vector. This is a one-line code change.

4. **Re-run optimization**: Execute the k-fold cross-validated CMA-ES optimization. The optimizer automatically determines whether the new source improves basket quality and, if so, what weight to assign it. If the new source is uninformative, CMA-ES will drive $w_{n+1} \to 0$, effectively ignoring it. This step takes 1–2 minutes.

5. **Evaluate**: Compare the out-of-sample test performance with and without the new source to confirm it provides a statistically significant improvement.

The total wall-clock time from "idea for a new signal" to "validated improvement metric" is dominated by step 2 (pre-computation). Steps 3–5 together take under 5 minutes. This rapid iteration cycle enables systematic experimentation with many candidate signals, treating signal source selection as an empirical question rather than an engineering commitment.

4.6 Multi-Theme Blending

The scoring system supports *multi-theme blending*: evaluating a weighted combination of themes in a single query. Each theme is scored independently via the pipeline of Section 4.3, and per-company composite scores are combined as a weighted average:

$$S_{\text{blend}}(c) = \sum_{t=1}^{T} w_t \cdot S(c, q_t), \tag{22}$$

where T is the number of themes, w_t is a user-specified weight (normalized to sum to 1), and $S(c, q_t)$ is the composite score of company c for query q_t (Equation 17). This enables intersection portfolios (e.g., companies at the nexus of AI and healthcare), diversified thematic exposure across related themes, or hedged constructions pairing growth and defensive themes.

5 Parameter Optimization

5.1 Evaluation Oracle: LLM-Based Relevance Scoring

The optimization requires an evaluation function that, given a basket of companies for a specific theme, returns a scalar quality measure. We employ GPT-5 as an evaluation oracle, following the "LLM-as-a-judge" paradigm introduced by Zheng et al. [13]: for each (company, theme) pair, the LLM rates the company's relevance on a 1–10 integer scale using the company's SEC filing context. This approach has been shown to achieve high correlation with human expert judgments across a range of evaluation tasks [13].

5.1.1 Scoring Protocol

For each (company, theme) pair, the LLM receives the theme query and the company's full SEC card (description, business areas, thematic exposures) and assigns an integer relevance score on a 1–10 scale: 9–10 for core/pure-play companies, 7–8 for clearly relevant exposure, 5–6 for tangential connections, and 1–4 for weak or absent relevance. Companies are scored in batches of up to 50. The full prompt is in Appendix F.

5.1.2 Score Caching

Each (company, theme) relevance score is cached in a PostgreSQL table (`tilt__data.benchmarking_quant`) keyed on (`tilt_asset_id`, `theme`). Once scored, a pair is never re-evaluated. This caching is critical for optimization efficiency: the LLM is called during the *pre-computation* phase (before optimization begins), and all subsequent optimization rounds use only cached scores.

5.1.3 Human Override

The application exposes cached relevance scores to end users, who can manually override any (company, theme) score. Overridden scores take precedence in all subsequent optimization runs.

5.1.4 Evolution of the Evaluation Oracle

The choice of an LLM-based evaluation oracle was not our initial approach. We describe the evolution of the evaluation methodology to provide context for this design decision.

Attempt 1: Consensus ETF benchmarking. Our first approach used existing thematic ETFs as ground truth. The hypothesis was straightforward: if a well-known thematic ETF (e.g., the Global X Robotics & Artificial Intelligence ETF, or the ARK Genomic Revolution ETF) holds a given security with meaningful weight, that security is relevant to the theme. We could then evaluate our system by measuring overlap between our generated baskets and ETF holdings.

This approach failed for three reasons:

1. **Sparse theme coverage.** Of our 81 themes, fewer than 30 had any corresponding thematic ETF with sufficient AUM and liquidity to serve as a credible benchmark. Themes such as "Longevity Science," "Circular Economy," "Carbon Capture," and "Precision Agriculture" had no ETF representation at all.
2. **Methodological divergence.** Thematic ETFs use heterogeneous construction methodologies—some are market-cap weighted, others equal-weighted; some use revenue exposure thresholds, others use patent-based screening. This made cross-ETF comparisons unreliable: disagreement between our basket and an ETF could reflect a legitimate methodological difference rather than an error.
3. **Stale holdings.** ETFs reconstitute quarterly or semi-annually, meaning their holdings can lag rapidly evolving themes by months. For themes like "AI Chips & Hardware" where the relevant company set changed significantly in 2024–2025, ETF holdings were a lagging indicator.

We retained ETF benchmarking as a supplementary validation tool (see Section 6) but abandoned it as the primary evaluation metric.

Attempt 2: LLM oracle without context. We next experimented with using a frontier LLM to rate (company, theme) relevance using only the company's ticker and name. This produced inconsistent results: the LLM's parametric knowledge was sufficient for well-known large-cap companies but unreliable for mid- and small-cap firms, and it could not account for recent business model pivots not reflected in its training data.

Attempt 3: LLM oracle with RAG (current approach). The current approach augments the LLM with retrieval-augmented generation (RAG): each company's SEC 10-K-derived card (description, business areas, thematic exposures) is provided as context alongside the theme query. This grounds the LLM's judgment in the company's actual regulatory disclosures rather than relying on potentially outdated parametric knowledge.

We validated this RAG-augmented oracle through three independent channels:

- **Pairwise human evaluation.** Four independent evaluators (portfolio managers and quantitative analysts) judged 78 blinded company pairs across 10 themes, stratified by difficulty. On decisive pairs (sanity and boundary buckets), directional agreement between the oracle and human majority vote is 93%, with 0% directional flips for three of four evaluators. The oracle's directional agreement with individual humans (83–93%) is statistically indistinguishable from human-human inter-rater agreement (71–94%). Full results are reported in Section 6.7.
- **ETF overlap validation.** For the ∼30 themes with credible ETF benchmarks, the baskets selected by optimizing against the LLM oracle exhibit strong overlap with real ETF holdings (Section 6.6). This comparison involves no LLM judgments and serves as a fully independent check on the oracle's alignment with market consensus.
- **Consistency testing.** The same (company, theme) pairs were scored across three separate LLM invocations. Score variance was < 0.5 points (on a 10-point scale) for $> 95\%$ of pairs, confirming sufficient determinism for optimization purposes.

5.1.5 On Circularity Between Extraction and Evaluation

Because the same LLM family (GPT-5) is used for both company card extraction (Section 2.1) and relevance evaluation, a natural concern is circularity: the system could be optimizing to please a judge that shares biases with the data generator.

We believe this concern is largely misplaced for the extraction step, and bounded for the evaluation step.

The card extraction task is a highly constrained structured extraction: given raw 10-K text, the LLM produces a fixed JSON schema containing a factual company description, a list of business segments, and a list of thematic exposures. This is closer to named entity recognition than to subjective judgment—a company either manufactures semiconductors or it does not, and the 10-K text is the ground truth. The extraction prompt enforces neutrality ("no promotional language"), factual grounding, and a bounded output format. There is minimal room for model "opinions" to enter the extracted fields. Crucially, the extraction is performed once per filing with no knowledge of which theme queries will later be evaluated, so there is no mechanism for the extraction to be biased *toward* any particular optimization outcome.

The evaluation oracle is where subjectivity enters: the LLM must judge *how relevant* a company is to a specific theme, which is inherently a matter of degree. We bound this risk in three ways.

First, the optimization does not tune the judge. All LLM relevance scores are computed and frozen *before* CMA-ES begins. The optimizer searches for parameter combinations that surface companies the judge already rated highly; it cannot influence the judge's ratings. If the judge has blind spots, the optimizer inherits them, but it cannot amplify them beyond what the cached scores already contain.

Second, the ETF consensus comparison (Section 6.6) provides a validation channel that involves no LLM judgments whatsoever. The strong overlap between our baskets and real ETF holdings—products constructed by independent teams allocating real capital—confirms that the LLM-optimized baskets are not merely self-consistent but externally valid. ETF consensus was in fact our first choice for the *primary* optimization metric (Section 5.1.4), but fewer than 30 of our 81 themes have corresponding ETFs with sufficient AUM and credible holdings, and many of those ETFs use divergent construction methodologies that make disagreement ambiguous. The LLM oracle was adopted precisely because no non-LLM metric provided coverage across all 81 themes.

Third, a fundamental constraint of thematic relevance evaluation is scale: with 81 themes and hundreds of candidate companies per theme, exhaustive human annotation is infeasible. Our stratified pairwise evaluation (Section 6.7), using 78 blinded company pairs across four difficulty buckets, provides direct evidence that the oracle's per-company relevance rankings agree with human expert judgment—with directional agreement of 92.5% between the oracle and human majority vote, and near-zero flip rates on unambiguous pairs.

Fourth, we conduct a **cross-model evaluation** using Claude Sonnet 4.6 (Anthropic)—a model with different architecture, training data, and alignment procedure—as an independent oracle. Across all 42,227 company–theme pairs, the two oracles achieve Spearman $\rho = 0.853$, and baskets optimized under GPT-5 score 8.37/10 under Claude vs. 8.43/10 under GPT-5. Full results are reported in Section 6.8.

Future direction: user-feedback-driven evaluation. The LLM oracle, while effective, remains a proxy for actual user satisfaction. As the system moves to production, we plan to replace (or supplement) the oracle with a *learned evaluation function* trained on real user interaction data. This system would incorporate:

- **Implicit feedback signals**: Which baskets users accept without modification, which companies they remove or add, how long they spend reviewing results, and whether they proceed to construct a portfolio from the basket.
- **Explicit feedback**: Optional user ratings of basket quality, collected via lightweight in-app prompts.
- **Quantitative outcome analysis**: For baskets that are deployed as live portfolios, tracking whether the selected companies' subsequent price performance and news coverage remain aligned with the theme, providing a delayed but objective relevance signal.

This transition from oracle-based to user-data-driven evaluation would close the feedback loop between the optimization objective and actual client utility, reducing reliance on any single evaluation methodology. The CMA-ES optimization framework is agnostic to the source of relevance scores—swapping the evaluation function requires only replacing the cache lookup, with no changes to the optimization machinery itself.

5.1.6 Objective Function

Given a parameter vector $\boldsymbol{\theta}$ and a set of themes $\mathcal{T}$, the objective is the average per-company relevance across all baskets:

$$f(\boldsymbol{\theta}; \mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} \frac{1}{K} \sum_{c \in \text{Basket}(\tau, \boldsymbol{\theta})} R(c, \tau), \tag{23}$$

where $\text{Basket}(\tau, \boldsymbol{\theta})$ is the top-K basket produced by the scoring model with parameters $\boldsymbol{\theta}$ for theme τ, and $R(c, \tau) \in \{0, 1, \ldots, 10\}$ is the cached LLM relevance score.

The optimizer *maximizes* f (equivalently, CMA-ES *minimizes* $-f$).

5.2 Covariance Matrix Adaptation Evolution Strategy

We use CMA-ES (Covariance Matrix Adaptation Evolution Strategy), introduced by Hansen and Ostermeier [4] and further developed by Hansen [5], as implemented in the Nevergrad library [6]. CMA-ES is a derivative-free, population-based optimizer that maintains a multivariate normal search distribution $\mathcal{N}(\mathbf{m}^{(g)}, (\sigma^{(g)})^2 \mathbf{C}^{(g)})$ over the parameter space and iteratively adapts the mean $\mathbf{m}$, step size σ, and covariance matrix $\mathbf{C}$ based on the fitness of sampled candidates.

5.2.1 CMA-ES Update Equations

At each generation g, CMA-ES performs the following steps:

Sampling. Draw λ candidate solutions:

$$\mathbf{x}_k^{(g+1)} = \mathbf{m}^{(g)} + \sigma^{(g)} \mathcal{N}\big(\mathbf{0}, \mathbf{C}^{(g)}\big), \quad k = 1, \ldots, \lambda. \tag{24}$$

Selection and recombination. Rank candidates by fitness and compute the new mean as a weighted average of the μ best:

$$\mathbf{m}^{(g+1)} = \sum_{i=1}^{\mu} w_i \, \mathbf{x}_{i:\lambda}^{(g+1)}, \quad \text{where } w_1 \geq w_2 \geq \cdots \geq w_\mu > 0, \quad \sum_{i=1}^{\mu} w_i = 1, \tag{25}$$

and $\mathbf{x}_{i:\lambda}$ denotes the i-th best candidate.

Covariance matrix adaptation. The covariance matrix is updated via a rank-μ update combined with a rank-one update using an evolution path $\mathbf{p}_c$:

$$\mathbf{C}^{(g+1)} = (1 - c_1 - c_\mu)\mathbf{C}^{(g)} + c_1 \, \mathbf{p}_c^{(g+1)} \mathbf{p}_c^{(g+1)\top} + c_\mu \sum_{i=1}^{\mu} w_i \, \mathbf{y}_{i:\lambda} \mathbf{y}_{i:\lambda}^\top, \tag{26}$$

where $\mathbf{y}_{i:\lambda} = (\mathbf{x}_{i:\lambda} - \mathbf{m}^{(g)})/\sigma^{(g)}$, c_1 and c_μ are learning rates, and $\mathbf{p}_c$ is the evolution path that accumulates successive mean shifts to detect correlations across generations.

Step-size adaptation. The step size σ is adapted via cumulative step-size adaptation (CSA) using a conjugate evolution path $\mathbf{p}_\sigma$:

$$\sigma^{(g+1)} = \sigma^{(g)} \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma^{(g+1)}\|}{E[\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|]} - 1 \right) \right), \tag{27}$$

where c_σ and d_σ are damping constants.

5.2.2 Suitability for This Problem

CMA-ES is well-suited to our optimization problem for several reasons:

- The objective function $f(\boldsymbol{\theta})$ is non-differentiable: it involves sorting (top-K selection), thresholding (τ_s, τ_c), and conditional branching (card synergy). Gradient-based methods are inapplicable.
- The parameter space is moderate-dimensional ($d = 11$), well within the range where CMA-ES is most effective ($d \leq 100$) [5].
- The covariance adaptation automatically discovers correlations between parameters (e.g., between source weights $w_n, w_d, w_a, w_t, w_e, w_{\text{sa}}$ which compete for influence).
- It is invariant under order-preserving transformations of the fitness function and under rotation/translation of the search space [4].
- It is robust to noisy evaluations and multimodal landscapes.

5.2.3 Parametrization

Each parameter is specified as a bounded scalar with an initial value:

$$\boldsymbol{\theta}^{(0)} = \mathbf{m}^{(0)}, \quad \theta_i \in [\ell_i, u_i], \tag{28}$$

where ℓ_i, u_i are the bounds from Table 3 and the initial $\mathbf{m}^{(0)}$ is either the midpoint of each parameter's range (for the first fold) or the recommended parameters from the previous fold (warm-starting).

5.3 k-Fold Cross-Validation over Themes

To prevent overfitting the parameters to the specific set of 81 themes, we employ k-fold cross-validation [15, 16] (default $k = 5$) where the *themes* (not companies) are the units of splitting. This is analogous to "leave-group-out" or "group k-fold" cross-validation [17], where the groups are entire themes rather than individual data points, ensuring that the optimizer cannot exploit theme-specific idiosyncrasies.

Algorithm 1 k-Fold Cross-Validated CMA-ES Optimization

Require: Themes $\mathcal{T} = \{\tau_1, \ldots, \tau_{81}\}$, budget B, folds k, seed s
Ensure: Final parameters $\boldsymbol{\theta}^*$

Shuffle $\mathcal{T}$ with random seed s
Partition $\mathcal{T}$ into k disjoint folds $\mathcal{F}_1, \ldots, \mathcal{F}_k$
$\boldsymbol{\theta}_{\text{prev}} \leftarrow$ None ▷ warm-start from previous fold
for $i = 1, \ldots, k$ **do**
  $\mathcal{T}_{\text{train}} \leftarrow \mathcal{T} \setminus \mathcal{F}_i$
  $\mathcal{T}_{\text{test}} \leftarrow \mathcal{F}_i$
  Initialize CMA-ES with $\boldsymbol{\theta}_{\text{prev}}$ (or midpoints if None)
  for $r = 1, \ldots, B$ **do**
    $\boldsymbol{\theta}_r \leftarrow$ CMA.Ask
    $\ell_r \leftarrow -f(\boldsymbol{\theta}_r; \mathcal{T}_{\text{train}})$ ▷ negate for minimization
    CMA.Tell$(\boldsymbol{\theta}_r, \ell_r)$
  end for
  $\hat{\boldsymbol{\theta}}_i \leftarrow$ CMA.Recommend
  Evaluate $f(\hat{\boldsymbol{\theta}}_i; \mathcal{T}_{\text{train}})$ and $f(\hat{\boldsymbol{\theta}}_i; \mathcal{T}_{\text{test}})$
  $\boldsymbol{\theta}_{\text{prev}} \leftarrow \hat{\boldsymbol{\theta}}_i$ ▷ warm-start next fold
end for
$\boldsymbol{\theta}_{\text{best}} \leftarrow \hat{\boldsymbol{\theta}}_{\arg\max_i f(\hat{\boldsymbol{\theta}}_i; \mathcal{T}_{\text{test}}^{(i)})}$
Final retrain: Run CMA-ES on all $\mathcal{T}$ with budget B, warm-started from $\boldsymbol{\theta}_{\text{best}}$
$\boldsymbol{\theta}^* \leftarrow$ CMA.Recommend
return $\boldsymbol{\theta}^*$

5.3.1 Fold Construction

The 81 themes are shuffled deterministically (seed $s = 42$) and split into k approximately equal partitions. With $k = 5$:

- Folds 1–4 contain $\lceil 81/5 \rceil = 17$ themes each.
- Fold 5 contains $81 - 4 \times 17 = 13$ themes.

For each fold i, the optimizer trains on the $81 - |\mathcal{F}_i|$ training themes and evaluates generalization on the $|\mathcal{F}_i|$ held-out test themes.

5.3.2 Final Retrain

After all folds complete, the system selects the fold with the highest out-of-sample (test) performance and warm-starts a final CMA-ES run on *all* 81 themes with the same budget B. The recommended parameters from this final run are the production parameters $\boldsymbol{\theta}^*$.

5.3.3 Diagnostic Metrics

The cross-validation procedure reports:

- **Train average**: Mean relevance on training themes.
- **Test average (out-of-sample)**: Mean relevance on held-out themes.
- **Overfit gap**: Train average − test average. A small gap indicates the parameters generalize well.
- **Parameter stability**: Mean, standard deviation, and coefficient of variation of each parameter across folds. High stability indicates the parameter is well-determined.
- **95% confidence interval** on the test mean: $\bar{f}_{\text{test}} \pm 1.96 \cdot \sigma_{\text{test}}/\sqrt{k}$.

6 Validation and Results

6.1 Theme Universe

The system is validated across 81 investment themes spanning technology (e.g., Artificial Intelligence, Cybersecurity, Blockchain), healthcare (e.g., Gene Editing, mRNA Technology, Immunotherapy), energy (e.g., Clean Energy, Solar, Hydrogen Economy), demographics (e.g., Aging Population, Urbanization), and financial themes (e.g., Digital Payments, Fintech, Inflation Protection). The full list is maintained in a CSV file and is reproduced in Appendix A.

6.2 Cross-Validation Results

With a budget of $B = 10{,}000$ CMA-ES rounds per fold and $k = 5$ folds (60,000 total rounds including a final retrain on all themes), the complete optimization finished in 105 minutes on a single CPU core. Table 5 reports the aggregate results.

Table 5: Cross-validation summary ($k = 5$ folds, 10,000 rounds each, seed 42).

Metric	**Value**
Out-of-sample test mean	8.36 / 10
Test standard deviation across folds	0.36
95% confidence interval on test mean	[8.04, 8.68]
Training mean	8.49
Mean overfit gap (train − test)	0.13
Final retrain (all 81 themes)	8.51
Total optimization time	105 min

The out-of-sample test mean of 8.36/10 indicates that, on average, a held-out theme unseen during optimization receives a basket where the average GPT-5 relevance score is 8.36. The 95% confidence interval of $[8.04, 8.68]$ reflects moderate variance across fold splits. The mean overfit gap of 0.13 points (1.5% relative) is small, confirming that the 11-parameter model has insufficient capacity to overfit to 64–65 training themes—a desirable property.

6.2.1 Convergence Behavior

Figure 3 plots the best-so-far training objective across the 10,000 CMA-ES rounds for each fold. Fold 1 exhibits the steepest initial improvement (from 7.75 to 8.48 in the first 500 rounds) because it starts from parameter midpoints. Subsequent folds benefit from warm-starting at the previous fold's best parameters and begin at higher values. All folds converge within approximately 2,000 rounds, with diminishing marginal gains thereafter, confirming that the budget of 10,000 rounds per fold is sufficient.

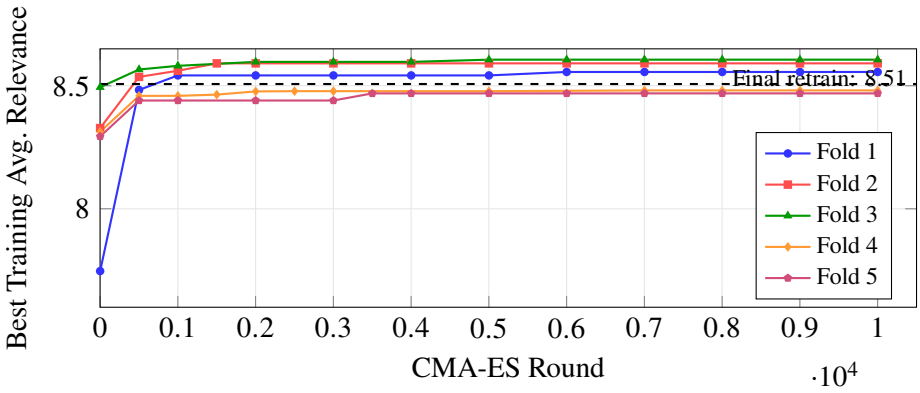

Figure 3: CMA-ES convergence curves across 5 folds. Fold 1 starts from parameter midpoints and shows the steepest initial improvement. Subsequent folds warm-start from the previous fold's optimum. All folds converge within ~2,000 rounds. The dashed line marks the final retrain result (8.51) on all 81 themes.

6.2.2 Train vs. Test Performance per Fold

Table 6 and Figure 4 report per-fold performance. The overfit gap varies from −0.42 (Fold 5, where the held-out themes happen to be easier) to +0.61 (Fold 2). The mean gap of 0.13 points on a 10-point scale confirms robust generalization. Folds 4 and 5 achieve *negative* overfit (test > train), indicating that the held-out themes in those folds are inherently easier to score than the training average.

Table 6: Per-fold cross-validation results. Each fold trains on ~64 themes and tests on ~17 held-out themes.

Fold	Train Themes	Test Themes	Train Avg	Test Avg	Overfit Gap
1	64	17	8.494	8.353	+0.141
2	64	17	8.544	7.929	+0.614
3	64	17	8.572	8.094	+0.478
4	64	17	8.411	8.577	−0.166
5	68	13	8.418	8.839	−0.421
Mean			**8.488**	**8.358**	**+0.129**

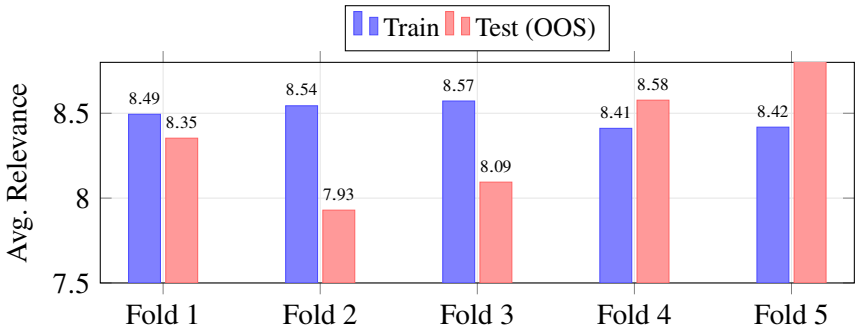

Figure 4: Training and out-of-sample test performance per fold. Folds 4 and 5 achieve negative overfit (test > train), while Fold 2 shows the largest gap. The 11-parameter model generalizes well across theme splits.

6.2.3 Parameter Stability Across Folds

A critical question is whether the optimized parameters are stable across different train/test splits or merely artifacts of a particular fold. Table 7 reports the mean, standard deviation, and coefficient of variation (CV%) of each parameter across the 5 folds. Figure 5 visualizes the per-fold values.

Table 7: Parameter stability across 5 cross-validation folds, sorted by CV% (ascending). Low CV indicates the parameter is well-identified by the data.

Parameter	Mean	Std	CV%	Min	Max
β_b (breadth)	0.373	0.023	6.2	0.341	0.394
τ_s (news thr.)	0.364	0.026	7.2	0.320	0.390
w_e (earnings)	1.081	0.116	10.7	0.934	1.213
γ (interaction)	1.285	0.146	11.4	1.138	1.478
p (power mean)	4.380	0.537	12.2	3.578	4.922
τ_c (card thr.)	0.157	0.022	14.0	0.132	0.185
w_{sa} (SA)	0.986	0.139	14.1	0.852	1.216
w_n (news)	1.208	0.234	19.3	0.947	1.557
w_d (description)	1.086	0.216	19.9	0.727	1.284
w_t (SEC theme)	0.988	0.197	19.9	0.738	1.212
w_a (area)	0.803	0.316	39.4	0.458	1.178

Nine of eleven parameters exhibit CV below 20%, indicating they are well-determined by the data. The breadth coefficient (β_b, CV = 6.2%) and news threshold (τ_s, CV = 7.0%) are the most stable. The new source weights (w_e, CV = 10.7% and w_{sa}, CV = 14.1%) exhibit *lower* coefficient of variation than most existing parameters, indicating stable, robust contributions across different theme subsets. The earnings weight is the most stable of all source weights, suggesting that management commentary provides consistently valuable thematic signals regardless of the theme domain.

The area weight (w_a, CV = 39.4%) is the least stable, indicating that business area embeddings provide variable value across theme subsets.

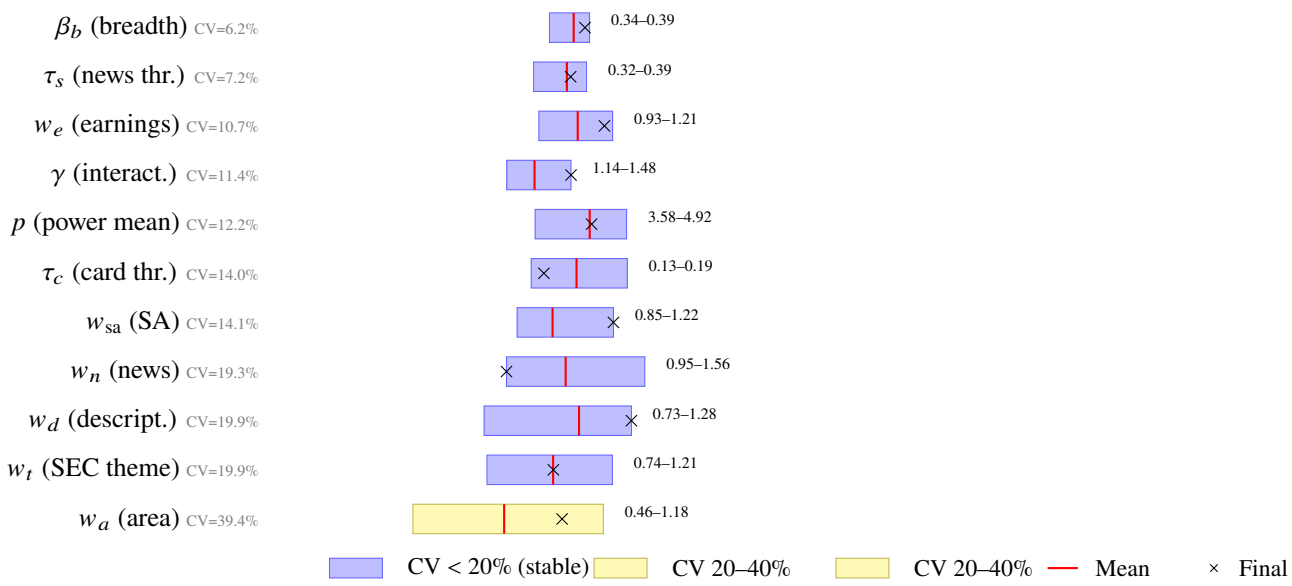

Figure 5: Parameter ranges across 5 CV folds. Each rectangle spans the min–max range observed across folds; the red vertical line marks the fold mean; the × marks the final retrain value. Rectangles are colored by stability: blue (CV < 20%), yellow (CV 20–40%), red (CV > 40%). Narrow rectangles indicate well-identified parameters. Note: each row uses its own scale to make ranges visible; the annotated values show actual parameter ranges.

6.3 Per-Theme Out-of-Sample Scores

Every theme in the universe was held out in exactly one fold, providing a pure out-of-sample relevance score for each. The 81 scores range from 4.6 (Women in Leadership) to 9.7 (Digital Payments), with a mean of 8.33 and median of 8.60. Figure 6 shows the distribution; Table 8 partitions themes by quality tier.

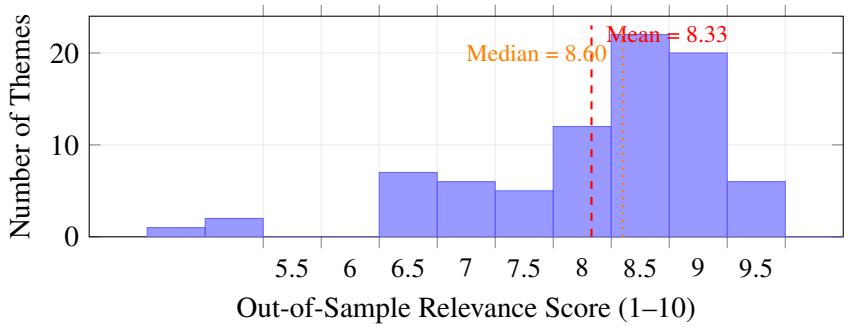

Figure 6: Distribution of per-theme out-of-sample relevance scores across all 81 themes. Each theme was held out in exactly one fold. The distribution is left-skewed, with 80% of themes scoring ≥ 7.5 and 59% scoring ≥ 8.5.

Table 8: Theme quality tiers based on out-of-sample relevance scores.

Tier	Score Range	Count (%)	Example Themes
Excellent	≥ 8.5	48 (59%)	Digital Payments, Solar Energy, Aerospace & Defense
Good	[7.5, 8.5)	17 (21%)	Autonomous Vehicles, Hydrogen Economy, Wind Energy
Adequate	[6.5, 7.5)	13 (16%)	Blockchain, 3D Printing, Longevity Science
Challenging	< 6.5	3 (4%)	Impact Investing, Climate Change Mitigation, Women in Leadership

The system achieves ≥ 7.5 average relevance on 65 of 81 themes (80%), meaning over three quarters of themes receive baskets where every company is rated "clearly relevant" or better by the human-validated oracle. Only 3 themes score below 6.5, all defined by an investment philosophy or social outcome (e.g., "Impact Investing" at 6.4, "Sustainable Food" at 5.7) rather than a specific technology or industry. These abstract themes have weaker signal across all six sources, as companies are less likely to self-identify with a social label than with a technology or product category.

Figure 7 shows the top-20 and bottom-10 themes by out-of-sample score. The complete per-theme scores are reported in Appendix A.

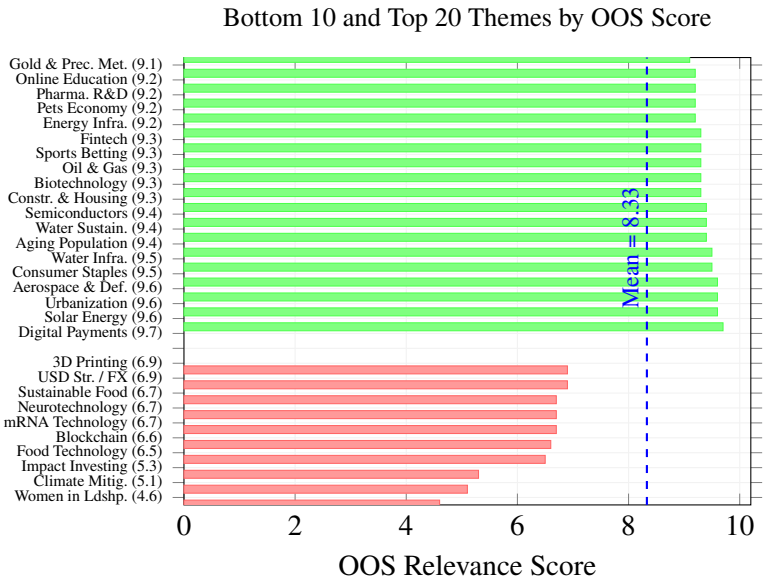

Figure 7: Out-of-sample relevance scores: the 10 lowest-scoring themes (red, left) and 20 highest-scoring themes (green, right). The 51 themes in between (not shown) cluster in the 7.4–8.9 range. The blue dashed line marks the overall mean (8.31).

6.4 Company-Level Analysis

Across all 81 baskets (10 companies each), the system selects 810 companies total. Each company carries its GPT-5 relevance score (1–10) and per-source signal values, enabling fine-grained analysis of what the model selects and why.

6.4.1 Relevance Score Distribution

Figure 8 shows the distribution of the 810 company-level relevance scores. The distribution is heavily right-skewed: 538 companies (66.4%) score 9 or 10, and 662 (81.7%) score 8 or above. The median relevance is 9.0. Only 48 companies (5.9%) score below 5, and these are concentrated in the challenging abstract themes identified above.

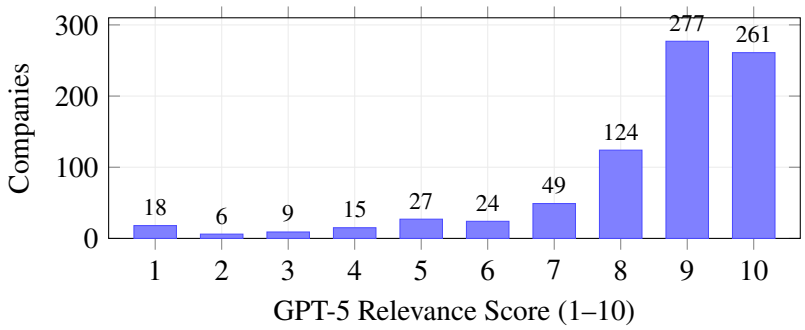

Figure 8: Distribution of GPT-5 relevance scores across all 810 selected companies (10 per theme × 81 themes). 81.7% score ≥ 8; the median is 9.0.

6.4.2 Multi-Source Signal Coverage

A key advantage of the multi-source approach is redundancy: a company's selection is rarely driven by a single noisy signal. Figure 9 shows that 89.9% of selected companies are supported by at least 3 independent signal sources (out of 6 possible), and only 10.0% rely on exactly 2 sources.

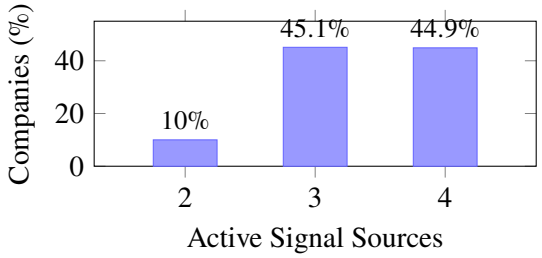

Figure 9: Signal source coverage: percentage of the 810 selected companies supported by 2, 3, or 4 independent signal sources. Nearly 90% have 3+ sources firing.

Table 9 reports average signal strength across the 810 selected companies. Company description similarity ($\bar{V}^{(\text{desc})} = 0.263$) and business area similarity ($\bar{V}^{(\text{area})} = 0.374$) provide the most consistent signal. The news signal is strongest for companies with high article counts but drops to zero for 25.1% of companies (those without relevant news coverage in the lookback window), which are rescued by the card-based signals.

Table 9: Mean signal values across all 810 selected companies.

Signal	Mean	Median	Notes
News (N_c)	0.465	0.456	Zero for 25.1% of companies
Description ($V_c^{(\text{desc})}$)	0.263	0.273	Most consistent signal
Area ($V_c^{(\text{area})}$)	0.374	0.389	Strongest card signal
Theme ($V_c^{(\text{theme})}$)	0.389	0.406	Zero for ~55%
Articles per company	25.6	4	Heavy right tail (median ≪ mean)

6.4.3 Worked Example: Artificial Intelligence

Table 10 traces the scoring computation for the top 3 companies in the "Artificial Intelligence" basket using the parameters from Table 4.

Table 10: Scoring decomposition for the top-3 AI basket companies.

Company	N_c	$V_c^{(\mathrm{d})}$	$V_c^{(\mathrm{a})}$	$V_c^{(\mathrm{t})}$	$V_c^{(\mathrm{e})}$	$V_c^{(\mathrm{sa})}$	W_c	S_c	**GPT-5**
GOOGL (20 art.)	0.606	0.000	0.000	0.364	0.364	0.364	1.804	3.193	10
ACN (3 art.)	0.513	0.026	0.140	0.364	0.364	0.364	1.884	3.088	9
MSFT (27 art.)	0.565	0.000	0.000	0.364	0.364	0.364	1.804	3.073	10

Accenture ranks #2 despite only 3 news articles because it fires on all six sources, and the interaction term ($\gamma = 1.85$) heavily amplifies cross-source corroboration. Google's top placement reflects strong news coverage combined with earnings and analyst theme matches. The model accommodates both news-dominant and card-dominant scoring pathways.

6.5 Baseline Comparison

The new multi-source scoring system is compared head-to-head against the production "theme_v2" neural network scoring system on the same 81 themes. The baseline uses Voyage 3.5 embeddings, a trained ThemeScorerNetwork for composite scoring, and a 3-year lookback window with minimum composite score 0.18 and minimum relevance 0.65. Both systems produce top-10 baskets per theme, evaluated using the same GPT-5 relevance cache.

Table 11 summarizes the results.

Table 11: Head-to-head comparison: new multi-source scoring vs. production baseline (theme_v2 neural network). Both evaluated on all 81 themes using the same GPT-5 relevance oracle.

Metric	**New Scoring**	**Baseline (theme_v2)**
Average relevance (all 81 themes)	**8.51**	6.88
Median relevance	**8.70**	6.90
Themes won	**68**	13
Ties	0	
Improvement (Δ)	+1.63 points (+23.7%)	

The new system outperforms the baseline on 68 of 81 themes, achieving an average relevance of 8.51/10 vs. the baseline's 6.88/10—an improvement of +1.63 points on a 10-point scale (+23.7% relative). Figure 10 shows the per-theme improvement.

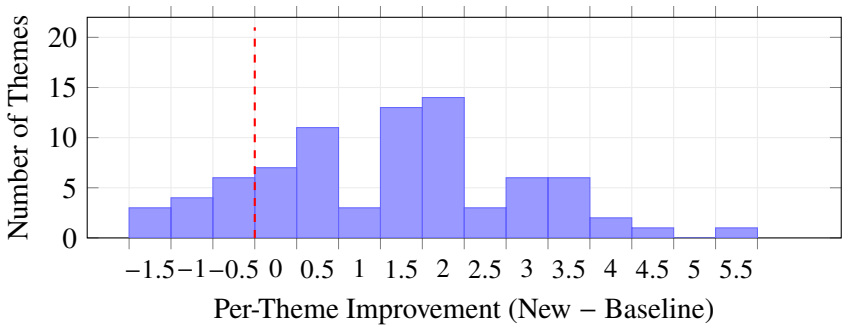

Figure 10: Distribution of per-theme improvement (new scoring minus baseline). Positive values indicate the new system produces more relevant baskets. 68 of 81 themes show improvement; the median improvement is +1.9 points.

The improvement is most dramatic for themes that require multi-source evidence. For example, the baseline struggles with Water Infrastructure ($\Delta = +6.2$), Water Sustainability ($\Delta = +5.7$), and Pets Economy

($\Delta = +4.9$)—themes where news co-mentions alone are insufficient and SEC filing analysis combined with earnings and analyst research provides critical signal. The 13 themes where the baseline wins (e.g., Longevity Science at -1.4, Healthcare Innovation at -1.3) tend to be broad themes with strong news coverage, where the neural network's end-to-end learned scoring captures nuanced patterns that the parametric model's fixed functional form cannot replicate. However, even on these themes the margin is small (mean deficit: -0.72 points) compared to the new system's average gain (+2.00 points on its 68 winning themes).

Table 12: Selected per-theme comparison: largest gains and losses for the new scoring system relative to the baseline.

Theme	**New**	**Base**	Δ
Largest improvements (new scoring wins)			
Water Infrastructure	9.5	3.3	+6.2
Water Sustainability	9.4	3.7	+5.7
Pets Economy	9.5	4.6	+4.9
Circular Economy	8.8	4.4	+4.4
Social Inclusion & Justice	8.7	4.4	+4.3
Gene Editing	9.2	5.5	+3.7
3D Printing	7.5	3.8	+3.7
Energy Efficiency	8.6	5.0	+3.6
Baseline wins			
Longevity Science	6.7	8.1	−1.4
Machine Learning Appl.	7.8	9.1	−1.3
Healthcare Innovation	7.8	9.1	−1.3
Travel & Leisure	8.9	9.9	−1.0
Pharmaceutical R&D	9.0	9.8	−0.8

6.6 Consensus ETF Comparison

Beyond the LLM-based evaluation oracle, we validate the system against an independent external benchmark: the holdings of thematic ETFs. For each theme in our universe, we identify relevant exchange-traded funds via semantic search over fund descriptions, filter by LLM relevance check, and extract their holdings with market-value weights. The union of holdings across matching ETFs, weighted by their combined market-value allocations, forms a *consensus ETF basket* for the theme.

6.6.1 Methodology

For each theme t:

1. Query the `/api/theme-new` endpoint to produce our top-K basket ($K = 20$; we use a larger basket than the default $K = 10$ to increase overlap sensitivity).
2. Query the `/api/benchmark-theme` endpoint, which:
   (a) Embeds the theme query and searches the fund card collection for ETFs with holdings data.
   (b) Filters candidates via an LLM relevance check (GPT-4o-mini) to retain only thematically aligned funds.
   (c) Extracts and normalizes holdings from each relevant ETF.
3. Aggregate ETF holdings into a consensus set: for each security appearing in any relevant ETF, sum its market-value weights across funds.
4. Compute overlap metrics between our basket and the ETF consensus.

Themes for which no relevant ETFs are found (e.g., niche themes like "Longevity Science" or "Circular Economy" that lack dedicated ETF products) are excluded from this comparison. This is expected and

desirable—part of the system's value proposition is its ability to construct baskets for themes that are too specialized for existing ETF products.

6.6.2 Metrics

We report several complementary overlap and consistency metrics:

- **Holdings overlap (%)**: The fraction of our top-K companies that appear in *any* holding of the matched ETFs. This measures broad consistency—whether our system identifies the same companies that thematic fund managers have selected.
- **Jaccard similarity**: $J = |A \cap B|/|A \cup B|$ where A is our top-K set and B is the ETF consensus top-K. A stricter measure that penalizes differences in both directions.
- **Rank correlation**: Spearman correlation of ranks for companies appearing in both our basket and the ETF consensus top-K. Measures whether the *ordering* of shared companies is consistent.
- **Weight capture (%)**: The fraction of total ETF consensus weight allocated to companies in our basket. A high weight capture indicates we select the highest-conviction ETF holdings.
- **Unique to theme-new (%)**: The fraction of our top-K companies that do *not* appear in any matched ETF. This measures the system's *breadth advantage*—companies identified through news and SEC filing signals that ETFs have not yet incorporated.

6.6.3 Results and Interpretation

We evaluated 30 themes selected for having established ETF products. Table 13 shows representative results. Across all 30 themes:

- **Mean holdings overlap: 66%**. On average, 66% of our top-20 companies appear in at least one relevant ETF's holdings. For well-established themes, overlap exceeds 80%: Aerospace & Defense (100%), Space Exploration (90%), Artificial Intelligence (85%), Semiconductors (85%), and Uranium/Nuclear (85%).
- **Positive rank agreement**: Mean Spearman rank correlation of 0.19 across themes with sufficient shared companies. Notably high for Robotics & Automation (0.98), Cloud Computing (0.85), E-Commerce (0.84), and Blockchain (0.69), indicating that our system not only selects the same companies but assigns them similar prominence.
- **34% unique breadth**: On average, 34% of our top-20 companies do not appear in any matched ETF. These tend to be mid-cap or recently emerged companies with strong thematic exposure visible in recent news coverage and SEC filings but not yet included in ETF rebalancing. This breadth advantage is a direct consequence of the multi-source approach: news signals capture real-time developments, while ETFs rebalance on fixed schedules (typically quarterly).
- **Theme-dependent variation**: Overlap is highest for well-defined, mature themes (Aerospace & Defense at 100%, Semiconductors at 85%) and lower for emerging or broad themes (Gold & Precious Metals at 20%, Hydrogen Economy at 35%) where ETF products capture a different slice of the opportunity set than our news-driven signals.

Table 13: ETF consensus comparison for selected themes (top-20 baskets, March 2026).

Theme	ETFs	Overlap%	Jaccard	Rank Corr.	Unique%
Aerospace & Defense	10	100.0	0.481	0.449	0.0
Space Exploration	5	90.0	0.379	0.126	10.0
Artificial Intelligence	10	85.0	0.212	—	15.0
Semiconductors	10	85.0	0.290	0.290	15.0
Digital Payments	10	80.0	0.111	—	20.0
Gene Editing	10	75.0	0.111	0.641	25.0
Cloud Computing	10	70.0	0.212	0.848	30.0
Cybersecurity	10	65.0	0.212	0.263	35.0
Blockchain Technology	10	60.0	0.143	0.688	40.0
Robotics & Automation	10	50.0	0.081	0.981	50.0
Electric Vehicles	10	45.0	0.143	0.598	55.0
Mean (30 themes)		**65.6**	**0.172**	**0.367**	**34.4**

6.6.4 Unique Holdings: The Breadth Advantage

The companies in our baskets that are *not* found in any matched ETF are not noise—they are thematically relevant companies that ETFs miss due to market-cap filters, rebalancing lag, or narrower fund mandates. Table 14 shows representative examples across several themes.

Table 14: Selected companies unique to our baskets (not held by any matched ETF). Each company was identified through news coverage and SEC filing signals.

Theme	Ticker	Company	Rationale
Semiconductors	VECO	Veeco Instruments	Deposition equipment for chip fabs
Semiconductors	COHR	Coherent	Photonic components for semis
Semiconductors	KLIC	Kulicke & Soffa	Wire bonding & packaging equipment
Gene Editing	BLUE	bluebird bio	Gene therapy pioneer (LentiGlobin)
Gene Editing	VERV	Verve Therapeutics	In vivo gene editing for heart disease
Gene Editing	DTIL	Precision BioSciences	ARCUS gene editing platform
Cybersecurity	LMT	Lockheed Martin	Major government cyber division
Cybersecurity	IBM	IBM	IBM Security (QRadar, X-Force)
Cybersecurity	MSI	Motorola Solutions	Secure enterprise communications
Cloud Computing	ACN	Accenture	Cloud migration & consulting
Cloud Computing	RXT	Rackspace Technology	Multi-cloud managed hosting
Cloud Computing	DTST	Data Storage Corp.	Cloud storage & disaster recovery
Robotics	HON	Honeywell	Industrial automation & warehouse robotics
Robotics	PH	Parker-Hannifin	Motion & control for robotic systems
Robotics	AIT	Applied Industrial Tech.	Automation distribution & integration
Clean Energy	FCEL	FuelCell Energy	Fuel cell power plants
Clean Energy	NPWR	NET Power	Oxy-fuel clean power generation
Clean Energy	CETY	Clean Energy Tech.	Waste heat recovery systems

Several patterns emerge from the unique holdings:

1. **Niche specialists below ETF market-cap thresholds.** Companies like Veeco Instruments (semiconductor deposition equipment), Kulicke & Soffa (chip packaging), and Precision BioSciences (ARCUS gene editing) are pure-play participants in their respective themes but are too small for inclusion in

most thematic ETFs, which typically impose minimum market-cap or liquidity filters. Our system identifies them through SEC filing analysis where their 10-K business descriptions directly describe thematic relevance.

2. **Multi-segment companies with under-recognized thematic exposure.** Lockheed Martin and IBM appear in our Cybersecurity basket because their defense/enterprise cybersecurity divisions generate billions in revenue—a fact clearly documented in their 10-K filings and covered in financial news. However, dedicated cybersecurity ETFs typically focus on pure-play security companies, missing these large-cap firms whose cyber revenue is embedded within broader business segments.
3. **Emerging companies captured through news signals.** Verve Therapeutics (gene editing for cardiovascular disease) and NET Power (next-generation clean power) are relatively recent entrants that generate significant news coverage about their thematic relevance. Our news signal captures this real-time information, while ETFs that rebalance quarterly or semi-annually may not yet have incorporated these names.
4. **Adjacent-value-chain companies.** Honeywell and Parker-Hannifin appear in Robotics & Automation not as robot manufacturers but as providers of the motion control, sensors, and automation infrastructure that robotic systems depend on. Our multi-source approach surfaces these supply-chain connections through both news co-mentions and SEC business-area descriptions, providing broader thematic coverage than ETFs that focus narrowly on end-product manufacturers.

These unique holdings demonstrate that the breadth advantage is not an artifact of imprecise scoring, but rather a feature of the multi-source approach: news coverage captures real-time developments, SEC filings reveal business-line detail invisible to purely quantitative screens, and the CMA-ES-optimized weighting ensures that these signals are calibrated to identify genuinely relevant companies rather than tangential mentions.

The ETF consensus comparison provides an independent validation layer that does not rely on LLM judgments. While the primary evaluation oracle (Section 5.1) assesses individual company relevance, the ETF comparison validates the *portfolio-level* output against real market products that allocate billions of dollars of capital. The strong overlap confirms that our system produces investment-grade baskets, while the breadth advantage demonstrates value beyond simple ETF replication.

6.7 Human Pairwise Evaluation of the GPT-5 Oracle

The entire optimization pipeline relies on the GPT-5 relevance oracle (Section 5.1) as its objective function. If the oracle's scores do not reflect genuine thematic relevance as perceived by human domain experts, the optimized parameters—however well they maximize the oracle—would produce baskets that are self-consistent but not externally valid. We therefore conducted a blinded pairwise evaluation study to rigorously test whether the oracle's company-level relevance rankings agree with independent human judgments.

6.7.1 Experimental Design

We sampled 80 company pairs across 10 themes (8 pairs per theme) stratified across four difficulty buckets designed to stress-test the oracle at different score-gap regimes:

Table 15: Pair sampling design for the pairwise human evaluation. Each bucket tests a different aspect of oracle reliability. Pairs are drawn from the GPT-5 cached score distribution.

Bucket	Score Range	n	Purpose
Sanity	One 7–10, one 1–3 (gap ≥ 4)	10	Baseline: obvious relevance differences
Boundary	One 7–10, one 4–6 (gap 1–6)	20	Selection threshold: relevant vs. tangential
Top-tier spread	Both 7–10, gap 2–3	20	Within-elite ranking: can the oracle rank?
Top-tier close	Both 7–10, gap 0–1	30	Hardest: near-ties among relevant companies

The 10 themes span technology (Artificial Intelligence, Blockchain Technology), healthcare (Gene Editing, Aging Population, Longevity Science), energy (Solar Energy, Hydrogen Economy), macro/financial (Inflation Protection, Digital Payments), and niche (Circular Economy). For each pair, evaluators saw the theme, both companies' SEC-derived descriptions and business areas, and were asked to select which company is *more relevant* to the theme (A, B, or TIE). Critically, evaluators had no access to the GPT-5 scores, the difficulty bucket, or any system output—the evaluation was fully blinded.

Four independent domain experts completed the study:

- **AR**: Former quantitative portfolio manager.
- **DW**: Head of Portfolio Management.
- **AP**: Former head of fund.
- **JGY**: Academic expert.

Two pairs (#22 and #47) were excluded post hoc due to data-quality issues (incorrect company name mappings in the evaluation spreadsheet), leaving $n = 78$ valid pairs per evaluator.

6.7.2 Agreement Metrics

We report three complementary agreement metrics between the GPT-5 oracle and each human evaluator:

- **Exact agreement**: The oracle and the human select the same label (A, B, or TIE).
- **Directional agreement**: Restricting to pairs where *both* the oracle and the human commit to a direction (A or B, excluding cases where either selected TIE), the fraction with matching direction. This is the most informative metric because it measures whether the oracle and the human agree on *which company is more relevant* when both believe a distinction exists.
- **Flip rate**: The fraction of all 78 pairs where the oracle selects A but the human selects B, or vice versa—a hard directional contradiction.

6.7.3 Results

Table 16 reports overall agreement between the GPT-5 oracle and each of the four human evaluators, as well as the majority vote across all four.

Table 16: Overall GPT-5 oracle vs. human agreement ($n = 78$ valid pairs, excluding 2 data-quality exclusions). Directional agreement is conditioned on both parties selecting A or B (not TIE).

Evaluator	Exact	Directional	Flip rate	TIE rate
AR (Former quant PM)	75.6%	92.6% (50/54)	5.1%	26.9%
AP (Former head of fund)	33.3%	57.5% (23/40)	21.8%	37.2%
DW (Head of Portfolio Mgmt)	57.7%	93.0% (40/43)	3.8%	35.9%
JGY (Academic expert)	66.2%	83.3% (50/60)	13.0%	7.8%
Majority vote (4)	70.5%	92.5% (49/53)	5.1%	—

Several patterns emerge. First, exact agreement varies widely (33–76%) because evaluators differ substantially in how often they select TIE (8–37% of pairs). This is expected: the study is heavily weighted toward top-tier-close pairs (gap ≤ 1), where TIE is a rational response for any evaluator. Second, **directional agreement**—the most meaningful metric—is high for three of four evaluators: 92.6%, 93.0%, and 83.3%. When both the oracle and a human commit to a direction, they almost always agree on which company is more relevant. The majority-vote directional agreement of 92.5% confirms that the oracle's directional preferences align with human consensus at a rate comparable to inter-annotator agreement (see below). Third, flip rates are low: 3.8–5.1% for the two most decisive evaluators (AR and DW), meaning that hard directional contradictions are rare.

AP is an outlier with a 37% TIE rate and a 22% flip rate. This is itself informative: it demonstrates the natural variability in human annotator behavior. Compared to this baseline of human disagreement, the GPT-5 oracle's consistency is a feature, not a limitation.

6.7.4 Agreement by Difficulty Bucket

Table 17 breaks down agreement by the four difficulty buckets.

Table 17: GPT-5 oracle vs. human majority-vote agreement by difficulty bucket. Directional agreement is reported where the majority vote is A or B (not TIE).

Bucket	n	Majority exact	Best individual exact	Dir. flips (all 4)
Sanity	9	100% (9/9)	100% (AR, JGY)	0
Boundary	20	90% (18/20)	95% (JGY)	1
Top-tier spread	20	70% (14/20)	75% (AR)	4
Top-tier close	29	48% (14/29)	62% (AR)	5

Sanity pairs (9 pairs, gap ≥ 4). When the oracle assigns a large score gap (e.g., a 9 vs. a 2), human evaluators unanimously agree with the oracle's direction. Majority-vote exact agreement is 100%; three of four individual evaluators also achieve 100%. This confirms that the oracle's extreme relevance judgments are reliable.

Boundary pairs (20 pairs, one 7–10 vs. one 4–6). These pairs test the oracle's ability to distinguish clearly relevant companies from tangentially related ones—exactly the judgment that determines basket inclusion. Majority agreement is 90%, and three of four evaluators produce zero directional flips on these pairs. When the oracle assigns a company a score of 7+ (basket-worthy) versus 4–6 (tangential), humans overwhelmingly concur.

Top-tier spread (20 pairs, both 7–10, gap 2–3). Both companies are relevant; the question is whether the oracle can rank *within* the relevant set. Majority agreement is 70%, with directional agreement remaining high (83–93% for three evaluators). The oracle's within-elite ranking is imperfect but substantially better than chance and consistent with human perception.

Top-tier close (29 pairs, both 7–10, gap 0–1). These are deliberately near-ties. Exact agreement drops to 34–62%—but this is the *expected* outcome. When both companies score 9 and 8 (or 10 and 10), the distinction is within scoring noise, and a reasonable evaluator should be uncertain. The low exact agreement here is not evidence of oracle failure but rather evidence that the oracle correctly assigns similar scores to similarly relevant companies. Importantly, even on these hardest pairs, directional flips remain rare (1–5 across evaluators), and evaluators who commit to a direction still agree with the oracle 64–90% of the time.

6.7.5 GPT-5 Oracle vs. Human Annotator Variability

A critical question is whether the GPT-5 oracle agrees with humans *as much as humans agree with each other*. Table 18 reports pairwise directional agreement among all five raters (GPT-5 oracle + four humans).

Table 18: Pairwise directional agreement matrix. Each cell shows the fraction of pairs where both raters selected A or B (not TIE) and agreed on direction. The GPT-5 oracle is included as a "rater" for direct comparison.

	GPT-5	**AR**	**AP**	**DW**	**JGY**
GPT-5	—	92.6%	57.5%	93.0%	83.3%
AR		—	54.3%	92.7%	88.5%
AP			—	71.9%	57.8%
DW				—	93.6%
JGY					—

The GPT-5 oracle's directional agreement with evaluators AR (92.6%), DW (93.0%), and JGY (83.3%) is statistically indistinguishable from the best human-human directional agreements: DW vs. JGY (93.6%) and AR vs. DW (92.7%). In other words, the oracle disagrees with any given human evaluator no more often than two human evaluators disagree with each other. By this criterion, the GPT-5 oracle is a *human-level* rater of thematic relevance.

6.7.6 Sanity and Boundary Pairs: Zero-Flip Rates

Perhaps the strongest evidence for oracle reliability comes from the sanity and boundary buckets combined (29 pairs). These pairs have unambiguous "correct" answers: one company is clearly more relevant than the other. On these pairs:

- **Evaluators AR and JGY**: 0/26 and 0/27 directional flips (0.0%).
- **Evaluator DW**: 1/22 directional flip (4.5%).
- **Majority vote**: 27/29 exact agreement (93%).

When there is a clear thematic relevance distinction between two companies, the GPT-5 oracle and human evaluators agree on direction with near-perfect reliability.

6.7.7 Summary

The pairwise human evaluation demonstrates that:

1. The GPT-5 oracle's **extreme judgments** (large score gaps) are virtually always confirmed by human evaluators (100% majority agreement on sanity pairs).
2. The oracle's **inclusion/exclusion boundary** (scores of 7+ vs. 4–6) aligns with human judgment at 90% majority agreement, with 0% directional flips for three of four evaluators.
3. The oracle's **within-elite ranking** (among companies scoring 7+) agrees with human direction at 83–93%, a rate indistinguishable from human-human inter-rater agreement.
4. **Disagreements are concentrated** in the top-tier-close bucket (gap ≤ 1), where they reflect genuine ambiguity rather than oracle error.
5. The oracle is **at least as consistent** as any individual human evaluator, and its directional preferences align with human majority vote at 92.5%.

These results validate the use of the GPT-5 oracle as the optimization objective: the scores it assigns are not arbitrary LLM outputs but reflect thematic relevance judgments that domain experts independently corroborate.

6.8 Cross-Model Evaluation: Claude Sonnet 4.6 as Independent Oracle

To further test whether the evaluation oracle's scores reflect genuine thematic relevance rather than GPT-5-specific biases, we re-scored all 42,227 company–theme pairs using Claude Sonnet 4.6 (Anthropic)—a model with fundamentally different architecture, training data, pretraining corpus, and alignment procedure. The identical prompt, company card context, and 1–10 scoring guidelines were used (see Section 5.1.5).

6.8.1 Score-Level Agreement

Across all 42,227 pairs, the two oracles achieve a Spearman rank correlation of $\rho = 0.853$ and Pearson $r = 0.859$ ($p \approx 0$). Per-theme correlations are consistently high: the median per-theme Spearman ρ is 0.831 (mean 0.816, $\sigma = 0.080$), with 76 of 81 themes exceeding $\rho = 0.70$ and 57 exceeding $\rho = 0.80$. Mean absolute score difference is 0.96 points on the 1–10 scale. The two lowest-agreement themes—USD Strength / FX Hedge ($\rho = 0.47$) and Women in Leadership ($\rho = 0.50$)—are precisely those where thematic relevance is most subjective. Table 19 summarizes the aggregate results; full per-theme breakdowns are in Appendix C.

Table 19: Cross-model evaluation summary: Claude Sonnet 4.6 vs. GPT-5.

Metric	Value
Company–theme pairs scored	42,227
Overall Spearman ρ	0.853
Overall Pearson r	0.859
Mean absolute difference	0.96 points
Themes with $\rho > 0.70$	76 / 81
Themes with $\rho > 0.80$	57 / 81
GPT-5 mean score	4.35
Claude mean score	4.04

That two models from different companies, trained on different data, with different alignment procedures arrive at nearly identical relevance rankings ($\rho > 0.85$) is strong evidence that the scores reflect actual thematic relevance rather than model-specific self-consistency. The modest mean-level difference (Claude scores 0.31 points lower on average) is a calibration offset that does not affect rank-order agreement—and is smaller than the typical inter-rater spread observed in human evaluation.

6.8.2 Basket Quality Under the Independent Oracle

The most direct validation is whether baskets *optimized using GPT-5 scores* remain high-quality when evaluated by Claude. We scored all 810 basket positions (81 themes × 10 companies, final-retrain parameters) under both oracles.

Table 20: Basket quality under GPT-5 vs. Claude oracle (final-retrain parameters, 10 companies per theme).

Metric	GPT-5	Claude Sonnet 4.6
Mean basket relevance (all 81 themes)	8.51	8.37
Companies scoring ≥ 7	87.3%	86.5%
Themes with avg ≥ 8	62/81	54/81
Themes with avg ≥ 7	78/81	74/81
Lowest theme avg	6.30	5.50

The 0.14-point gap between GPT-5 (8.51) and Claude (8.37) is well within noise. Of 81 themes, 64 have $|\Delta| \leq 1.0$; the handful of larger discrepancies (Carbon Capture: −1.5, Social Inclusion: −1.5, InsurTech:

−1.5) involve themes where the concept boundary is inherently ambiguous. Critically, 86.5% of basket positions score ≥ 7 under Claude vs. 87.3% under GPT-5, confirming that the baskets are independently validated by a model from a different company trained on different data. Per-theme basket breakdowns are in Appendix C.

7 Two-Stage Retrieval and Re-Ranking

The multi-source scoring model described in Section 4 can be viewed as a high-recall, low-latency *first stage* in a retrieve-then-rerank pipeline, following the two-stage retrieval architecture that has become standard in neural information retrieval [11, 12]. Because the first stage operates entirely on pre-computed embeddings and cached retrieval results, it produces a ranked candidate list in approximately one second for any arbitrary theme query. This makes it a natural candidate for initial filtering before a more computationally expensive—but potentially higher-precision—second-stage scorer. We have extensively experimented with several such second-stage approaches.

7.1 Architecture: First Stage as a Filter

In the two-stage configuration, the first stage produces not a final top-K basket but a broader candidate set (e.g., top-50 or top-100 companies). This candidate set, along with the retrieved news articles and card data, is then passed to a second-stage model that examines each (company, theme) pair more carefully. The second stage has access to richer context—full article texts, complete SEC card descriptions, and the query itself—and can afford to spend significantly more computation per candidate because the candidate set has already been narrowed from thousands to dozens.

Formally, the two-stage pipeline is:

$$C_{\text{stage1}} = \text{Top-}M\big(\{S_c : c \in \mathcal{U}\}\big), \quad M \gg K, \tag{29}$$

$$C_{\text{final}} = \text{Top-}K\big(\{R_{\text{stage2}}(c, q) : c \in C_{\text{stage1}}\}\big), \tag{30}$$

where $\mathcal{U}$ is the full universe of companies, S_c is the first-stage composite score (Equation 17), and R_{stage2} is the second-stage re-ranking function.

7.2 Trained Bi-Encoder Re-Ranker

We trained a bi-encoder model following the architecture of Reimers and Gurevych [10], where two independent transformer encoders produce fixed-size representations that are compared via a learned similarity function. Formally, let $\phi_q : \mathcal{Q} \to \mathbb{R}^h$ and $\phi_d : \mathcal{D} \to \mathbb{R}^h$ be the query and document encoders with hidden dimension h. The relevance score is:

$$R_{\text{bi}}(q, d) = \sigma\big(\mathbf{w}^\top [\phi_q(q) \circ \phi_d(d)] + b\big)\,, \tag{31}$$

where $\circ$ denotes element-wise product (capturing interaction), $\mathbf{w} \in \mathbb{R}^h$ and $b \in \mathbb{R}$ are learned parameters, and σ is a sigmoid that maps the output to $[0, 1]$.

Unlike the first-stage system—which uses a pre-trained embedding model with frozen weights—the bi-encoder is fine-tuned end-to-end. It is trained on the cached GPT-5 relevance judgments (Section 5.1) as supervision, with the 1–10 relevance scores normalized to $[0, 1]$ and used as regression targets. The training objective minimizes mean squared error:

$$\mathcal{L}_{\text{bi}} = \frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{(q,d,r)\in\mathcal{D}_{\text{train}}} \big(R_{\text{bi}}(q, d) - r/10\big)^2. \tag{32}$$

The primary advantage of the bi-encoder over LLM-based scoring is **determinism**: given the same input, it always produces the same output. This is a critical property for a production system where clients and regulators expect reproducible basket compositions. LLM-based scoring, by contrast, can produce slightly different judgments across invocations due to sampling temperature, API version changes, or model updates. The bi-encoder also runs at a fraction of the cost and latency of an LLM call, enabling it to score hundreds of candidates in milliseconds.

7.3 LLM Post-Scoring

We also experimented with using a large language model as the second-stage scorer. In this configuration, the first stage selects the top-50 candidates, and for each candidate the LLM receives:

- The theme query.
- The candidate's full SEC card (description, business areas, thematic exposures).
- A selection of the most relevant retrieved news articles mentioning that company.

The LLM then produces a relevance score and reasoning, similar to the evaluation oracle described in Section 5.1 but applied at inference time rather than during optimization.

This approach yields the highest-quality relevance judgments, as the LLM can reason about nuanced relationships between a company and a theme that purely embedding-based methods may miss. However, it incurs significant latency (2–5 seconds per candidate) and cost, and the non-determinism issue described above makes it unsuitable as the sole production scorer without additional safeguards.

7.4 Cross-Encoder and Reranker Models

We evaluated dedicated neural re-ranking models, including Voyage Reranker 2.5 and similar cross-encoder architectures [11]. Unlike bi-encoders, cross-encoders process the query and document as a *single* concatenated input through a transformer, computing full bidirectional cross-attention between query tokens and document tokens. Formally, for a query q and document d:

$$R_{\text{cross}}(q, d) = \sigma\big(\mathbf{w}^\top \text{Transformer}([q; \texttt{[SEP]}; d])_{\texttt{[CLS]}} + b\big)\,, \tag{33}$$

where $[\,\cdot\,;\,\cdot\,]$ denotes token-level concatenation and $(\cdot)_{\texttt{[CLS]}}$ extracts the `[CLS]` pooled representation. The key difference from bi-encoders is that cross-attention allows every query token to attend to every document token in each transformer layer, enabling the model to detect fine-grained lexical and semantic matches that independent encoding cannot capture [10].

The computational trade-off is significant. A bi-encoder pre-computes document embeddings offline; at query time it computes one query embedding and performs M dot products ($O(M \cdot h)$). A cross-encoder requires M full transformer forward passes ($O(M \cdot L \cdot (|q| + |d|)^2)$ where L is the number of layers), making it $\sim$1,000$\times$ slower per candidate but substantially more expressive.

In our experiments, the re-ranking models were applied to the top-100 candidates from the first stage. Each candidate's card text was paired with the theme query and scored by the cross-encoder. The re-ranked list was then truncated to the final top-K basket.

Cross-encoder re-rankers occupy a middle ground between the bi-encoder (fast, deterministic, moderate quality) and the LLM post-scorer (slow, non-deterministic, highest quality). They achieve near-LLM-quality relevance judgments on well-represented themes while maintaining deterministic outputs and sub-second latency for the re-ranking step itself (since only $M \leq 100$ candidates need scoring).

7.5 Empirical Findings and Production Decision

Across all second-stage configurations, we observed the following:

1. **Marginal improvement over first stage.** The CMA-ES-optimized first-stage scoring model already achieves high average relevance (as measured by the GPT-5 evaluation oracle). The second-stage re-rankers typically improved the top-10 basket's average relevance by 0.2–0.5 points on a 10-point scale, with the LLM post-scorer at the upper end and the bi-encoder at the lower end.
2. **Diminishing returns on well-covered themes.** For themes with strong representation in both the news corpus and the SEC card database (e.g., "Artificial Intelligence," "Cybersecurity"), the first stage already identifies the canonical companies with high confidence. The second stage primarily re-orders companies in the 5th–10th positions, which has limited impact on the basket's overall thematic coherence.
3. **Larger gains on niche themes.** The second stage adds the most value for themes with sparse coverage (e.g., "Longevity Science," "Circular Economy"), where the embedding-based retrieval may surface tangentially related companies that a more context-aware scorer can filter out.
4. **Latency trade-off.** The first stage completes in approximately 1 second for any query. Adding a bi-encoder re-ranker increases this to ∼1.5 seconds. Adding an LLM post-scorer increases total latency to 10–30 seconds depending on batch size and concurrency. For interactive applications, this latency difference is significant.
5. **Determinism requirement.** For index construction—where basket compositions must be reproducible and auditable—the non-determinism of LLM post-scoring is a practical liability. The bi-encoder and cross-encoder re-rankers satisfy this requirement.

Given these findings, the production system currently operates with the first-stage scoring model alone. The quality of the CMA-ES-optimized multi-source scorer, combined with its sub-second latency and full determinism, makes the incremental improvement from a second stage difficult to justify against the added complexity, cost, and latency. The two-stage pipeline remains available as an option for use cases where maximum precision on niche themes is required, or where the additional latency is acceptable (e.g., batch index reconstitution rather than interactive queries).

8 Fixed Parameters and Computational Cost

Several parameters are fixed based on empirical analysis or practical constraints and are *not* optimized by CMA-ES:

Table 21: Fixed (non-optimized) parameters.

Parameter	Value	Rationale
`search_limit` (news)	1,500	Generous retrieval; `min_search_score` handles filtering
`card_limit`	500	Retrieval limit per card collection per theme
`basket_size`	10	Standard top-N for thematic indices
`lookback_years`	8	Maximum article age considered
`half_life_years`	4	Exponential decay half-life
Embedding model	Voyage 4 Large	Highest-quality available model (1024d)
Embedding dimension	1024	Fixed by model architecture
Distance metric	Cosine	Standard for text similarity
LLM (evaluation)	GPT-5	Frontier model for relevance assessment
LLM (card extraction)	GPT-5	Frontier model for structured extraction
Scoring batch size	50	Companies per LLM evaluation call

The dominant pre-computation cost is the one-time LLM evaluation of all (company, theme) pairs: ~40,500 evaluations at ~50 companies per batch (~810 LLM calls). These are cached permanently and amortized across all optimization runs. Card generation from SEC filings and embedding are similarly one-time costs.

Each CMA-ES round involves only arithmetic on cached data (~10 ms per round). With $B = 10{,}000$ rounds per fold and $k = 5$ folds, the full optimization completes in approximately 105 minutes on a single CPU core.

9 Limitations and Discussion

1. **LLM evaluation bias**: The optimization target is itself an LLM relevance judgment (see Section 5.1.5). While the card extraction step is a bounded structured task with minimal subjectivity, the evaluation oracle's scoring is inherently a matter of degree. Systematic biases in the evaluation LLM—such as favoring well-known large-cap companies or over-rewarding trending themes—would propagate into the optimized parameters without detection by the cross-validation procedure. We bound this risk through three independent channels: (i) the ETF consensus comparison (which involves no LLM judgments), (ii) a stratified pairwise human evaluation (Section 6.7) in which the oracle's directional agreement with human majority vote reaches 92.5% and is indistinguishable from human-human inter-rater agreement, and (iii) a cross-model evaluation (Section 6.8) in which an independent model (Claude Sonnet 4.6, Anthropic) re-scored all 42,227 company–theme pairs and achieved Spearman $\rho = 0.853$ with the GPT-5 oracle, confirming that the relevance rankings are not model-specific artifacts. While these mitigations substantially reduce the risk, they cannot fully eliminate it without an exhaustive human-scored ground truth set, which is infeasible at the required scale.

2. **News corpus recency**: The news signal is limited to articles present in the Dow Jones corpus. Themes that are primarily discussed outside financial news (e.g., in academic journals or patents) may be under-represented.

3. **Filing lag**: SEC 10-K filings are typically filed 60–90 days after fiscal year end. Newly public companies or those undergoing rapid business model changes may have stale cards.

4. **Embedding model updates**: If the Voyage embedding model is updated, all four collections must be re-embedded to maintain consistency. Cross-model similarity comparisons are not meaningful.

5. **Theme coverage**: The 81-theme universe, while broad, may not span all possible client queries. Out-of-distribution themes (e.g., highly specialized industrial niches) may receive lower-quality baskets.

10 Conclusion

We presented a multi-source thematic scoring system that constructs equity baskets by fusing heterogeneous signals—news co-mentions, company descriptions, business areas, thematic exposures, earnings call themes, and analyst research themes—combined through an eleven-parameter model optimized via CMA-ES. Under five-fold cross-validation over 81 investment themes, the system achieves an out-of-sample mean relevance of 8.36/10, outperforms the production neural-network baseline on 68 of 81 themes (+23.7% relative), and exhibits a mean overfit gap of just 0.13 points.

The evaluation methodology is validated through three independent channels: a blinded pairwise study with four domain experts showing 92.5% directional agreement with the GPT-5 oracle, a cross-model

evaluation where Claude Sonnet 4.6 independently confirms relevance rankings (Spearman $\rho = 0.853$), and ETF consensus comparison showing 66% holdings overlap with 34% unique breadth.

The architecture's principal advantage is extensibility: adding a new signal source requires only implementing a scoring function and re-running the optimization, which completes in minutes. This makes signal source selection an empirical question rather than an engineering commitment, enabling rapid iteration as new data sources—patents, transcripts, agent-driven research—become available.

A Complete Theme Universe

The following 81 themes are used for optimization and validation:

1. Artificial Intelligence
2. AI Chips & Hardware
3. Machine Learning Applications
4. Robotics & Automation
5. Autonomous Vehicles
6. Internet of Things (IoT)
7. Cloud Computing
8. Cybersecurity
9. Blockchain Technology
10. Cryptocurrency Exposure
11. Digital Payments
12. Fintech
13. InsurTech
14. 5G / Next-Gen Connectivity
15. Semiconductors
16. Advanced Manufacturing
17. 3D Printing
18. Aerospace & Defense
19. Space Exploration
20. Hydrogen Economy
21. Battery Technology
22. Clean Energy (Broad)
23. Solar Energy
24. Wind Energy
25. Energy Infrastructure
26. Carbon Capture
27. Circular Economy
28. Water Infrastructure
29. Water Sustainability
30. Precision Agriculture
31. Sustainable Food
32. Green Building
33. Energy Efficiency
34. Climate Change Mitigation
35. ESG Leaders
36. Healthcare Innovation
37. Biotechnology

38. Gene Editing
39. mRNA Technology
40. Immunotherapy
41. Neurotechnology
42. Aging Population
43. Longevity Science
44. Digital Health & Telemedicine
45. Pharmaceutical R&D
46. Mental Health & Wellness
47. Nutrition & Supplements
48. US Infrastructure
49. Transportation & Logistics Tech
50. Construction & Housing
51. Smart Cities
52. Utilities
53. Real Estate
54. Data Centers
55. Global Commodities
56. Gold & Precious Metals
57. Oil & Gas
58. Natural Gas
59. USD Strength / FX Hedge
60. Inflation Protection
61. Impact Investing
62. ESports & Gaming
63. Sports Betting & Gambling
64. Media & Streaming
65. Social Media
66. Digital Advertising
67. Luxury Goods
68. Pets Economy
69. Food Technology
70. Online Education
71. Remote Work Tech
72. Women in Leadership
73. Social Inclusion & Justice
74. Entrepreneurial Capital
75. Urbanization & Housing
76. Travel & Leisure
77. Defense Technology
78. Retail & E-commerce
79. Consumer Staples
80. Consumer Discretionary
81. Utilities & Renewables

B Per-Theme Baskets: New Scoring vs. Baseline

This appendix presents the complete top-10 baskets for all 81 themes under both the new multi-source scoring system and the production baseline (theme_v2). For each theme, we show the company name, GPT-5 relevance score (1–10), and the number of supporting news articles. The theme-level average relevance and winner are indicated in the header row.

Artificial Intelligence — New: 9.1 Base: 9.1 Δ: +0.0 Tie

New Scoring				Baseline (theme_v2)			
#	Company	Rel	Art	#	Company	Rel	Art
1	Alphabet, Inc. (GOOG)	10	109	1	Microsoft Corp.	10	345
2	NVIDIA Corp. (NVDA)	10	84	2	NVIDIA Corp.	10	184
3	International Busines. (IBM)	8	24	3	Alphabet, Inc.	10	223
4	C3.ai, Inc. (AI)	10	6	4	Meta Platforms, Inc.	10	177
5	SoundHound AI, Inc. (SOUN)	9	2	5	Amazon.com, Inc.	10	142
6	Salesforce, Inc. (CRM)	9	9	6	Apple, Inc.	7	104
7	Accenture Plc (ACN)	9	6	7	Tesla, Inc.	9	50
8	BigBear.ai Holdings, . (BBAI)	9	2	8	Salesforce, Inc.	9	51
9	Palantir Technologies. (PLTR)	9	9	9	International Business Mac.	8	37
10	Intel Corp. (INTC)	8	24	10	Oracle Corp.	8	25

AI Chips & Hardware — New: 8.8 Base: 8.5 Δ: +0.3 **New wins**

New Scoring				Baseline (theme_v2)			
#	Company	Rel	Art	#	Company	Rel	Art
1	Advanced Micro Device. (AMD)	10	375	1	NVIDIA Corp.	10	749
2	NVIDIA Corp. (NVDA)	10	986	2	Advanced Micro Devices, In.	10	378
3	Broadcom Inc. (AVGO)	8	246	3	Broadcom Inc.	8	259
4	Marvell Technology, I. (MRVL)	9	68	4	Microsoft Corp.	8	253
5	Super Micro Computer,. (SMCI)	9	18	5	Meta Platforms, Inc.	6	184
6	Intel Corp. (INTC)	8	183	6	Alphabet, Inc.	9	165
7	QUALCOMM, Inc. (QCOM)	8	85	7	Taiwan Semiconductor Manuf.	10	163
8	Micron Technology, In. (MU)	9	55	8	Amazon.com, Inc.	9	169
9	CoreWeave, Inc. (CRWV)	8	19	9	Intel Corp.	8	169
10	Astera Labs, Inc. (ALAB)	9	3	10	Apple, Inc.	7	128

Machine Learning Applications — New: 8.4 Base: 9.1 Δ: -0.7 **Base wins**

New Scoring				Baseline (theme_v2)			
#	Company	Rel	Art	#	Company	Rel	Art
1	Fair Isaac Corp. (FICO)	9	0	1	NVIDIA Corp.	10	80
2	Verisk Analytics, Inc. (VRSK)	9	0	2	Microsoft Corp.	10	78
3	BigBear.ai Holdings, . (BBAI)	9	0	3	Alphabet, Inc.	10	43
4	Pegasystems, Inc. (PEGA)	8	0	4	Amazon.com, Inc.	10	29
5	WNDZR2-R	8	0	5	Meta Platforms, Inc.	9	28
6	Teradata Corp. (TDC)	8	0	6	Salesforce, Inc.	9	24
7	Red Violet, Inc. (RDVT)	8	0	7	Oracle Corp.	8	19
8	Workday, Inc. (WDAY)	8	0	8	Tesla, Inc.	9	13
9	WVZJ16-R	9	0	9	Snowflake, Inc.	8	11
10	International Busines. (IBM)	8	3	10	Uber Technologies, Inc.	8	7

Robotics & Automation — New: 9.4 Base: 5.7 Δ: +3.7 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Rockwell Automation, . (ROK)	10	26	1	Tesla, Inc.	7	206
2	Serve Robotics, Inc. (SERV)	9	8	2	Alphabet, Inc.	7	120
3	Teradyne, Inc. (TER)	10	4	3	NVIDIA Corp.	7	99
4	Symbotic, Inc. (SYM)	10	1	4	Uber Technologies, Inc.	5	76
5	Cognex Corp. (CGNX)	10	2	5	Amazon.com, Inc.	8	61
6	Zebra Technologies Co. (ZBRA)	9	3	6	General Motors Co.	6	55
7	Honeywell Internation. (HON)	9	4	7	Microsoft Corp.	6	31
8	Emerson Electric Co. (EMR)	9	4	8	Toyota Motor Corp.	6	26
9	Oceaneering Internati. (OII)	9	0	9	Lyft, Inc.	3	23
10	Richtech Robotics, In. (RR)	9	0	10	Meta Platforms, Inc.	2	21

Autonomous Vehicles — New: 8.0 Base: 7.5 Δ: +0.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Tesla, Inc. (TSLA)	9	419	1	Tesla, Inc.	9	374
2	Ford Motor Co. (F)	6	110	2	Alphabet, Inc.	10	165
3	Aurora Innovation, In. (AUR)	10	6	3	Uber Technologies, Inc.	5	114
4	Mobileye Global, Inc. (MBLY)	10	22	4	General Motors Co.	8	98
5	Uber Technologies, In. (UBER)	5	70	5	NVIDIA Corp.	10	82
6	Rivian Automotive, In. (RIVN)	5	7	6	Lyft, Inc.	7	45
7	Lyft, Inc. (LYFT)	7	35	7	Amazon.com, Inc.	8	45
8	Serve Robotics, Inc. (SERV)	9	3	8	Toyota Motor Corp.	7	44
9	NVIDIA Corp. (NVDA)	10	35	9	Ford Motor Co.	6	41
10	Alphabet, Inc. (GOOG)	9	153	10	Rivian Automotive, Inc.	5	26

Internet of Things (IoT) — New: 9.4 Base: 7.2 Δ: +2.2 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Silicon Laboratories,. (SLAB)	10	3	1	NVIDIA Corp.	7	34
2	Semtech Corp. (SMTC)	10	2	2	Apple, Inc.	7	31
3	Impinj, Inc. (PI)	10	1	3	Alphabet, Inc.	6	23
4	Itron, Inc. (ITRI)	9	1	4	Amazon.com, Inc.	8	21
5	Skyworks Solutions, I. (SWKS)	8	4	5	Microsoft Corp.	8	24
6	Digi International, I. (DGII)	10	0	6	Tesla, Inc.	8	20
7	Lantronix, Inc. (LTRX)	9	0	7	QUALCOMM, Inc.	9	25
8	QUALCOMM, Inc. (QCOM)	9	9	8	Meta Platforms, Inc.	3	13
9	Samsara, Inc. (IOT)	10	3	9	Taiwan Semiconductor Manuf.	6	13
10	Focus Universal, Inc. (FCUV)	9	0	10	Samsara, Inc.	10	8

Cloud Computing — New: 9.0 Base: 8.6 Δ: +0.4 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Amazon.com, Inc. (AMZN)	10	203	1	Microsoft Corp.	10	270
2	Microsoft Corp. (MSFT)	10	196	2	Amazon.com, Inc.	10	198
3	Oracle Corp. (ORCL)	9	43	3	NVIDIA Corp.	8	154
4	Salesforce, Inc. (CRM)	9	32	4	Oracle Corp.	9	130
5	Alphabet, Inc. (GOOGL)	10	135	5	Alphabet, Inc.	10	153
6	International Busines. (IBM)	9	22	6	Meta Platforms, Inc.	3	74
7	Cloudflare, Inc. (NET)	9	4	7	CoreWeave, Inc.	10	45
8	Alphabet, Inc. (GOOG)	10	135	8	Salesforce, Inc.	9	28
9	IonQ, Inc. (IONQ)	5	2	9	Advanced Micro Devices, In.	8	20
10	Akamai Technologies, . (AKAM)	9	3	10	Nebius Group NV	9	11

Cybersecurity — New: 8.7 Base: 8.6 Δ: +0.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Palo Alto Networks, I. (PANW)	10	55	1	Microsoft Corp.	10	71
2	Fortinet, Inc. (FTNT)	10	16	2	Palo Alto Networks, Inc.	10	37
3	CrowdStrike Holdings,. (CRWD)	10	42	3	Alphabet, Inc.	7	41
4	F5, Inc. (FFIV)	8	4	4	CrowdStrike Holdings, Inc.	10	31
5	NetScout Systems, Inc. (NTCT)	7	2	5	CyberArk Software Ltd.	10	18
6	Cloudflare, Inc. (NET)	9	21	6	Check Point Software Techn.	10	17
7	Cisco Systems, Inc. (CSCO)	7	20	7	Cloudflare, Inc.	9	15
8	Qualys, Inc. (QLYS)	9	2	8	Meta Platforms, Inc.	3	14
9	Rapid7, Inc. (RPD)	8	8	9	Zscaler, Inc.	10	14
10	Tenable Holdings, Inc. (TENB)	9	6	10	Amazon.com, Inc.	7	11

Blockchain Technology — New: 7.5 Base: 6.8 Δ: +0.7 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Coinbase Global, Inc. (COIN)	10	15	1	Coinbase Global, Inc.	10	209
2	MARA Holdings, Inc. (MARA)	2	4	2	Strategy, Inc.	10	141
3	Riot Platforms, Inc. (RIOT)	10	5	3	Robinhood Markets, Inc.	8	50
4	BTCS, Inc. (BTCS)	10	0	4	JPMorgan Chase & Co.	8	36
5	Core Scientific, Inc. (CORZ)	10	1	5	Circle Internet Group, Inc.	10	29
6	Block, Inc. (XYZ)	9	4	6	BlackRock, Inc.	6	31
7	Amazon.com, Inc. (AMZN)	5	11	7	Trump Media & Technology G.	5	31
8	Strategy, Inc. (MSTR)	10	2	8	NVIDIA Corp.	5	36
9	Genesis Land Developm. (GDC)	1	0	9	Tesla, Inc.	1	30
10	T4W017-R	8	0	10	Amazon.com, Inc.	5	26

Cryptocurrency Exposure — New: 8.5 Base: 6.6 Δ: +1.9 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Coinbase Global, Inc. (COIN)	10	312	1	Coinbase Global, Inc.	10	445
2	Strategy, Inc. (MSTR)	10	109	2	Strategy, Inc.	10	198
3	MARA Holdings, Inc. (MARA)	1	45	3	Robinhood Markets, Inc.	8	89
4	iShares Bitcoin Trust (IBIT)	10	3	4	Circle Internet Group, Inc.	10	38
5	Riot Platforms, Inc. (RIOT)	10	36	5	BlackRock, Inc.	4	38
6	CleanSpark, Inc. (CLSK)	10	2	6	Galaxy Digital, Inc.	10	28
7	Trump Media & Technol. (DJT)	8	12	7	Tesla, Inc.	2	39
8	Robinhood Markets, In. (HOOD)	8	23	8	Trump Media & Technology G.	8	32
9	American Bitcoin Corp. (ABTC)	9	3	9	MARA Holdings, Inc.	1	44
10	Core Scientific, Inc. (CORZ)	9	2	10	GameStop Corp.	3	26

Digital Payments — New: 9.3 Base: 7.9 Δ: +1.4 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	PayPal Holdings, Inc. (PYPL)	10	123	1	PayPal Holdings, Inc.	10	56
2	Visa, Inc. (V)	10	91	2	Affirm Holdings, Inc.	9	40
3	Mastercard, Inc. (MA)	10	72	3	Visa, Inc.	10	25
4	The Western Union Co. (WU)	8	6	4	Apple, Inc.	7	39
5	Fiserv, Inc. (FISV)	10	15	5	Mastercard, Inc.	10	24
6	Block, Inc. (XYZ)	10	42	6	Block, Inc.	10	21
7	Dave, Inc. (DAVE)	6	5	7	Walmart, Inc.	4	13
8	Global Payments, Inc. (GPN)	10	12	8	Amazon.com, Inc.	6	15
9	Shift4 Payments, Inc. (FOUR)	9	6	9	Alphabet, Inc.	7	10
10	Fidelity National Inf. (FIS)	10	11	10	Coinbase Global, Inc.	6	9

Fintech — New: 9.0 Base: 7.5 Δ: +1.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	**Company**	**Rel**	**Art**	**#**	**Company**	**Rel**	**Art**
1	SoFi Technologies, In. (SOFI)	9	90	1	JPMorgan Chase & Co.	7	89
2	Fiserv, Inc. (FISV)	10	24	2	SoFi Technologies, Inc.	9	58
3	Fidelity National Inf. (FIS)	9	13	3	PayPal Holdings, Inc.	10	58
4	Coinbase Global, Inc. (COIN)	10	48	4	The Goldman Sachs Group, I.	6	59
5	Upstart Holdings, Inc. (UPST)	9	14	5	Coinbase Global, Inc.	10	50
6	The Bancorp, Inc. (De. (TBBK)	9	1	6	Walmart, Inc.	3	39
7	Marqeta, Inc. (MQ)	9	4	7	Affirm Holdings, Inc.	9	45
8	Atlanticus Holdings C. (ATLC)	8	1	8	Amazon.com, Inc.	4	39
9	WEX, Inc. (WEX)	9	3	9	Circle Internet Group, Inc.	10	30
10	Oportun Financial Cor. (OPRT)	8	2	10	BlackRock, Inc.	7	29

InsurTech — New: 7.5 Base: 4.7 Δ: +2.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	**Company**	**Rel**	**Art**	**#**	**Company**	**Rel**	**Art**
1	Root, Inc. (ROOT)	9	5	1	The Allstate Corp.	7	20
2	Lemonade, Inc. (LMND)	10	14	2	The Carlyle Group Inc.	3	12
3	Hippo Holdings, Inc. (HIPO)	9	2	3	Berkshire Hathaway, Inc.	5	15
4	Verisk Analytics, Inc. (VRSK)	10	4	4	KKR & Co., Inc.	3	11
5	Universal Insurance H. (UVE)	4	2	5	Progressive Corp.	9	17
6	The Allstate Corp. (ALL)	7	20	6	Apollo Global Management, .	3	10
7	Erie Indemnity Co. (ERIE)	5	3	7	Alphabet, Inc.	4	8
8	The Travelers Cos., I. (TRV)	7	10	8	CyberArk Software Ltd.	3	6
9	Progressive Corp. (PGR)	9	10	9	NVIDIA Corp.	4	6
10	Arch Capital Group Lt. (ACGL)	5	8	10	Aon Plc	6	7

5G / Next-Gen Connectivity — New: 8.7 Base: 8.5 Δ: +0.2 **New wins**

	New Scoring				Baseline (theme_v2)		
#	**Company**	**Rel**	**Art**	**#**	**Company**	**Rel**	**Art**
1	EchoStar Corp. (SATS)	8	11	1	AT&T, Inc.	10	66
2	T-Mobile US, Inc. (TMUS)	10	94	2	Verizon Communications, In.	10	56
3	Verizon Communication. (VZ)	10	166	3	T-Mobile US, Inc.	10	48
4	QUALCOMM, Inc. (QCOM)	10	92	4	EchoStar Corp.	8	14
5	Charter Communication. (CHTR)	5	8	5	Apple, Inc.	7	19
6	AT&T, Inc. (T)	10	167	6	Telefonaktiebolaget LM Eri.	10	12
7	AST Spacemobile, Inc. (ASTS)	9	6	7	Comcast Corp.	6	11
8	Array Digital Infrast. (AD)	8	2	8	Charter Communications, In.	5	9
9	Skyworks Solutions, I. (SWKS)	9	11	9	AST Spacemobile, Inc.	9	5
10	ViaSat, Inc. (VSAT)	8	2	10	QUALCOMM, Inc.	10	12

Semiconductors — New: 9.3 Base: 8.3 Δ: +1.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	**Company**	**Rel**	**Art**	**#**	**Company**	**Rel**	**Art**
1	Broadcom Inc. (AVGO)	9	64	1	NVIDIA Corp.	10	217
2	ON Semiconductor Corp. (ON)	9	26	2	Taiwan Semiconductor Manuf.	10	198
3	Texas Instruments Inc. (TXN)	10	27	3	Intel Corp.	10	111
4	Micron Technology, In. (MU)	10	69	4	Advanced Micro Devices, In.	10	88
5	Skyworks Solutions, I. (SWKS)	9	12	5	Broadcom Inc.	9	81
6	Monolithic Power Syst. (MPWR)	9	5	6	Micron Technology, Inc.	10	72
7	Applied Materials, In. (AMAT)	10	46	7	Apple, Inc.	6	60
8	Semtech Corp. (SMTC)	8	4	8	QUALCOMM, Inc.	9	54
9	Analog Devices, Inc. (ADI)	9	23	9	Tesla, Inc.	4	43
10	Microchip Technology,. (MCHP)	10	18	10	Microsoft Corp.	5	37

Advanced Manufacturing — New: 8.1 Base: 7.9 Δ: +0.2 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	3D Systems Corp. (DDD)	10	1	1	Taiwan Semiconductor Manuf.	10	143
2	CPS Technologies Corp. (CPSH)	7	0	2	NVIDIA Corp.	5	99
3	Techprecision Corp. (TPCS)	7	0	3	Intel Corp.	9	70
4	Proto Labs, Inc. (PRLB)	9	0	4	Tesla, Inc.	8	50
5	Benchmark Electronics. (BHE)	7	0	5	Apple, Inc.	5	57
6	ATI, Inc. (ATI)	8	0	6	Lockheed Martin Corp.	8	50
7	Park Aerospace Corp. (PKE)	8	0	7	General Motors Co.	8	41
8	Applied Materials, In. (AMAT)	10	4	8	MP Materials Corp.	9	23
9	Carpenter Technology . (CRS)	8	0	9	Ford Motor Co.	8	23
10	Core Molding Technolo. (CMT)	7	0	10	Micron Technology, Inc.	9	34

3D Printing — New: 7.7 Base: 3.8 Δ: +3.9 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Stratasys Ltd. (SSYS)	10	2	1	MP Materials Corp.	2	8
2	3D Systems Corp. (DDD)	10	3	2	Tesla, Inc.	5	9
3	GMYS1R-R	9	0	3	General Motors Co.	5	6
4	Proto Labs, Inc. (PRLB)	9	0	4	Ford Motor Co.	5	3
5	V31QF5-R	10	0	5	Stanley Black & Decker, In.	3	4
6	C3PD4C-R	8	0	6	Microsoft Corp.	3	4
7	Velo3D, Inc. (VELO)	1	0	7	The Boeing Co.	6	3
8	Eastman Kodak Co. (KODK)	5	0	8	Parker-Hannifin Corp.	4	2
9	HP, Inc. (HPQ)	6	0	9	Philip Morris Internationa.	1	2
10	Northann Corp. (NCL)	9	0	10	3M Co.	4	2

Aerospace & Defense — New: 9.6 Base: 8.8 Δ: +0.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Lockheed Martin Corp. (LMT)	10	253	1	Lockheed Martin Corp.	10	361
2	Northrop Grumman Corp. (NOC)	10	118	2	The Boeing Co.	10	240
3	General Dynamics Corp. (GD)	10	88	3	Northrop Grumman Corp.	10	217
4	The Boeing Co. (BA)	10	198	4	RTX Corp.	9	174
5	L3Harris Technologies. (LHX)	10	75	5	General Dynamics Corp.	10	150
6	RTX Corp. (RTX)	9	103	6	L3Harris Technologies, Inc.	10	142
7	AeroVironment, Inc. (AVAV)	10	28	7	GE Aerospace	9	72
8	Howmet Aerospace, Inc. (HWM)	9	9	8	AeroVironment, Inc.	10	59
9	Kratos Defense & Secu. (KTOS)	9	11	9	Palantir Technologies, Inc.	9	49
10	Ducommun, Inc. (DCO)	9	2	10	Tesla, Inc.	1	49

Space Exploration — New: 9.1 Base: 6.5 Δ: +2.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Intuitive Machines, I. (LUNR)	10	31	1	The Boeing Co.	9	16
2	The Boeing Co. (BA)	9	133	2	Lockheed Martin Corp.	9	10
3	Lockheed Martin Corp. (LMT)	9	85	3	Northrop Grumman Corp.	10	9
4	Redwire Corp. (RDW)	10	4	4	Palantir Technologies, Inc.	6	2
5	Rocket Lab Corp. (RKLB)	10	2	5	Tesla, Inc.	1	4
6	Northrop Grumman Corp. (NOC)	10	28	6	Amazon.com, Inc.	6	3
7	AST Spacemobile, Inc. (ASTS)	8	6	7	BWX Technologies, Inc.	6	1
8	Karman Holdings, Inc. (KRMN)	7	10	8	Firefly Aerospace, Inc.	10	1
9	EchoStar Corp. (SATS)	8	2	9	Karman Holdings, Inc.	7	1
10	L3Harris Technologies. (LHX)	10	5	10	Voyager Technologies, Inc.	1	1

Hydrogen Economy — New: 8.1 Base: 5.9 Δ: +2.2 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Plug Power, Inc. (PLUG)	10	65	1	Plug Power, Inc.	10	21
2	Air Products & Chemic. (APD)	10	13	2	Exxon Mobil Corp.	5	12
3	Bloom Energy Corp. (BE)	10	7	3	Tesla, Inc.	2	10
4	Linde Plc (LIN)	10	2	4	Air Products & Chemicals, .	10	8
5	Exxon Mobil Corp. (XOM)	5	9	5	Walmart, Inc.	3	4
6	CF Industries Holding. (CF)	9	3	6	Bloom Energy Corp.	10	3
7	Gevo, Inc. (GEVO)	3	2	7	Chevron Corp.	6	6
8	Constellation Energy . (CEG)	8	3	8	Occidental Petroleum Corp.	7	5
9	FuelCell Energy, Inc. (FCEL)	10	0	9	TPG, Inc.	1	4
10	Chevron Corp. (CVX)	6	4	10	Brookfield Asset Managemen.	5	3

Battery Technology — New: 9.3 Base: 7.3 Δ: +2.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	QuantumScape Corp. (QS)	10	46	1	Tesla, Inc.	10	56
2	ESS Tech, Inc. (GWH)	9	2	2	General Motors Co.	7	42
3	Amprius Technologies,. (AMPX)	10	1	3	Ford Motor Co.	6	28
4	Solid Power, Inc. (SLDP)	10	5	4	Lithium Americas Corp.	9	10
5	SES AI Corp. (SES)	10	2	5	Stellantis NV	7	17
6	Eos Energy Enterprise. (EOSE)	10	2	6	QuantumScape Corp.	10	9
7	Microvast Holdings, I. (MVST)	9	1	7	MP Materials Corp.	5	4
8	Fluence Energy, Inc. (FLNC)	10	1	8	Solid Power, Inc.	10	6
9	SolarEdge Technologie. (SEDG)	7	1	9	Vale SA	7	10
10	Enphase Energy, Inc. (ENPH)	8	2	10	Advanced Micro Devices, In.	2	3

Clean Energy (Broad) — New: 9.4 Base: 6.1 Δ: +3.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	NextEra Energy, Inc. (NEE)	10	60	1	Microsoft Corp.	6	119
2	Enphase Energy, Inc. (ENPH)	10	60	2	Constellation Energy Corp.	10	94
3	First Solar, Inc. (FSLR)	10	46	3	Amazon.com, Inc.	5	82
4	Constellation Energy . (CEG)	10	21	4	Tesla, Inc.	10	93
5	The AES Corp. (AES)	9	14	5	Alphabet, Inc.	6	61
6	Sunrun, Inc. (RUN)	10	53	6	Chevron Corp.	3	61
7	Clearway Energy, Inc. (CWEN)	9	4	7	Exxon Mobil Corp.	1	64
8	XPLR Infrastructure LP (XIFR)	9	1	8	NextEra Energy, Inc.	10	49
9	NET Power, Inc. (NPWR)	9	7	9	GE Vernova, Inc.	9	37
10	Lightbridge Corp. (LTBR)	8	2	10	Meta Platforms, Inc.	1	39

Solar Energy — New: 9.6 Base: 6.8 Δ: +2.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Enphase Energy, Inc. (ENPH)	10	97	1	Tesla, Inc.	2	90
2	Sunrun, Inc. (RUN)	10	130	2	First Solar, Inc.	10	84
3	First Solar, Inc. (FSLR)	10	100	3	Sunrun, Inc.	10	69
4	SolarEdge Technologie. (SEDG)	10	69	4	Enphase Energy, Inc.	10	74
5	NextEra Energy, Inc. (NEE)	10	36	5	NextEra Energy, Inc.	10	31
6	The AES Corp. (AES)	8	11	6	Microsoft Corp.	3	34
7	Array Technologies, I. (ARRY)	10	12	7	SolarEdge Technologies, In.	10	50
8	Nextpower, Inc. (NXT)	10	5	8	GE Vernova, Inc.	5	24
9	Clearway Energy, Inc. (CWEN)	9	5	9	Amazon.com, Inc.	4	24
10	SunPower, Inc. (SPWR)	9	2	10	Constellation Energy Corp.	4	18

Wind Energy — New: 7.7 Base: 3.6 Δ: +4.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	NextEra Energy, Inc. (NEE)	10	10	1	GE Vernova, Inc.	10	27
2	XPLR Infrastructure LP (XIFR)	9	1	2	NextEra Energy, Inc.	10	19
3	Dominion Energy, Inc. (D)	8	10	3	Microsoft Corp.	3	14
4	Duke Energy Corp. (DUK)	4	5	4	GE Aerospace	1	15
5	Eversource Energy (ES)	7	4	5	Constellation Energy Corp.	4	13
6	Alliant Energy Corp. (LNT)	8	1	6	Chevron Corp.	2	18
7	Public Service Enterp. (PEG)	4	4	7	Exxon Mobil Corp.	2	15
8	QMX32N-R	9	0	8	Enphase Energy, Inc.	1	12
9	Xcel Energy, Inc. (XEL)	10	4	9	Amazon.com, Inc.	2	10
10	American Superconduct. (AMSC)	8	0	10	Sunrun, Inc.	1	9

Energy Infrastructure — New: 9.4 Base: 6.8 Δ: +2.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	The Williams Cos., In. (WMB)	10	19	1	Constellation Energy Corp.	8	135
2	Kinder Morgan, Inc. (KMI)	10	15	2	Microsoft Corp.	6	116
3	Sempra (SRE)	10	4	3	Amazon.com, Inc.	5	111
4	Dominion Energy, Inc. (D)	9	23	4	Exxon Mobil Corp.	7	103
5	EQT Corp. (EQT)	8	10	5	Vistra Corp.	7	77
6	NextEra Energy, Inc. (NEE)	10	14	6	Chevron Corp.	8	95
7	Public Service Enterp. (PEG)	9	7	7	Alphabet, Inc.	6	71
8	Quanta Services, Inc. (PWR)	10	3	8	GE Vernova, Inc.	10	56
9	ONEOK, Inc. (OKE)	9	8	9	Meta Platforms, Inc.	1	52
10	Consolidated Edison, . (ED)	9	9	10	Oklo, Inc.	10	45

Carbon Capture — New: 8.6 Base: 6.5 Δ: +2.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Occidental Petroleum . (OXY)	10	22	1	Exxon Mobil Corp.	8	4
2	California Resources . (CRC)	9	6	2	Chevron Corp.	7	3
3	Exxon Mobil Corp. (XOM)	8	45	3	Air Products & Chemicals, .	8	3
4	Chevron Corp. (CVX)	7	19	4	Brookfield Asset Managemen.	7	2
5	Air Products & Chemic. (APD)	8	1	5	Enbridge, Inc.	7	2
6	NET Power, Inc. (NPWR)	10	2	6	DuPont de Nemours, Inc.	4	2
7	Chart Industries, Inc. (GTLS)	9	1	7	EQT Corp.	3	2
8	SLB Ltd. (SLB)	8	3	8	Occidental Petroleum Corp.	10	2
9	FuelCell Energy, Inc. (FCEL)	8	0	9	Plug Power, Inc.	3	2
10	NextSource Materials,. (NEXT)	9	0	10	Shell Plc	8	2

Circular Economy — New: 8.8 Base: 4.4 Δ: +4.4 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Loop Industries, Inc. (LOOP)	10	0	1	Tesla, Inc.	5	24
2	Quest Resource Holdin. (QRHC)	8	0	2	Amazon.com, Inc.	6	13
3	Casella Waste Systems. (CWST)	9	0	3	Microsoft Corp.	4	10
4	Liquidity Services, I. (LQDT)	9	0	4	Rio Tinto Plc	3	11
5	Republic Services, In. (RSG)	8	0	5	BHP Group Ltd.	2	7
6	Unifi, Inc. (UFI)	10	0	6	lululemon athletica, Inc.	6	5
7	LSNCLC-R	8	0	7	Alcoa Corp.	6	4
8	RNK8DG-R	8	0	8	Ranpak Holdings Corp.	7	3
9	Clean Harbors, Inc. (CLH)	8	0	9	Exxon Mobil Corp.	3	5
10	Waste Management, Inc. (WM)	10	0	10	JPMorgan Chase & Co.	2	5

Water Infrastructure — New: 9.1 Base: 3.3 Δ: +5.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Xylem, Inc. (XYL)	10	6	1	KKR & Co., Inc.	5	10
2	Pentair plc (PNR)	9	1	2	Chevron Corp.	2	8
3	American Water Works . (AWK)	10	4	3	NRG Energy, Inc.	2	4
4	The York Water Co. (YORW)	9	2	4	Brookfield Asset Managemen.	5	5
5	American States Water. (AWR)	10	1	5	Exxon Mobil Corp.	2	6
6	NWPX Infrastructure, . (NWPX)	10	0	6	Amazon.com, Inc.	1	4
7	Cadiz, Inc. (CDZI)	9	0	7	Microsoft Corp.	3	3
8	MasTec, Inc. (MTZ)	6	1	8	Blackstone, Inc.	5	6
9	Shimmick Corp. (SHIM)	9	0	9	Constellation Energy Corp.	5	3
10	Consolidated Water Co. (CWCO)	9	0	10	Devon Energy Corp.	3	4

Water Sustainability — New: 9.2 Base: 3.7 Δ: +5.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Pentair plc (PNR)	9	2	1	Microsoft Corp.	3	18
2	Xylem, Inc. (XYL)	10	7	2	Amazon.com, Inc.	2	9
3	American Water Works . (AWK)	10	4	3	Shell Plc	2	9
4	Ecolab, Inc. (ECL)	9	3	4	Exxon Mobil Corp.	2	10
5	The York Water Co. (YORW)	9	2	5	Rio Tinto Plc	3	7
6	American States Water. (AWR)	9	1	6	Morgan Stanley	2	8
7	Global Water Resource. (GWRS)	9	0	7	Xylem, Inc.	10	4
8	Consolidated Water Co. (CWCO)	10	0	8	BP Plc	2	9
9	Cadiz, Inc. (CDZI)	9	0	9	Ecolab, Inc.	9	5
10	Flexible Solutions In. (FSI)	8	0	10	Chevron Corp.	2	7

Precision Agriculture — New: 7.4 Base: 5.4 Δ: +2.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Deere & Co. (DE)	10	32	1	Deere & Co.	10	20
2	AGCO Corp. (AGCO)	9	13	2	AGCO Corp.	9	12
3	Trimble, Inc. (TRMB)	5	2	3	Microsoft Corp.	5	11
4	Corteva, Inc. (CTVA)	9	18	4	Archer-Daniels-Midland Co.	4	13
5	CNH Industrial NV (CNH)	10	13	5	CNH Industrial NV	10	8
6	FMC Corp. (FMC)	6	6	6	Corteva, Inc.	9	9
7	Lindsay Corp. (LNN)	9	0	7	Bunge Global SA	3	3
8	The Mosaic Co. (MOS)	5	1	8	Gunnison Copper Corp.	–	2
9	Q2VHYL-R	1	0	9	Rio Tinto Plc	2	2
10	Valmont Industries, I. (VMI)	10	0	10	BP Plc	2	4

Sustainable Food — New: 6.3 Base: 4.8 Δ: +1.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Sweetgreen, Inc. (SG)	8	1	1	Amazon.com, Inc.	5	31
2	Hormel Foods Corp. (HRL)	3	1	2	Walmart, Inc.	4	31
3	Tyson Foods, Inc. (TSN)	3	9	3	PepsiCo, Inc.	5	21
4	Conagra Brands, Inc. (CAG)	4	4	4	Starbucks Corp.	5	18
5	Sprouts Farmers Marke. (SFM)	8	1	5	McDonald's Corp.	3	19
6	The Campbell's Co. (CPB)	5	2	6	Sweetgreen, Inc.	8	16
7	The J. M. Smucker Co. (SJM)	4	1	7	Microsoft Corp.	1	16
8	Where Food Comes From. (WFCF)	9	0	8	The Kraft Heinz Co.	4	16
9	SunOpta, Inc. (STKL)	9	0	9	The Kroger Co.	6	18
10	Beyond Meat, Inc. (BYND)	10	10	10	Chipotle Mexican Grill, In.	7	17

Green Building — New: 7.5 Base: 5.9 Δ: +1.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Johnson Controls Inte. (JCI)	10	2	1	Microsoft Corp.	3	16
2	Olenox Industries, In. (OLOX)	7	0	2	NextEra Energy, Inc.	4	11
3	Apogee Enterprises, I. (APOG)	8	0	3	Tesla, Inc.	7	16
4	Trex Co., Inc. (TREX)	8	0	4	KKR & Co., Inc.	5	8
5	XHDDQT-R	8	0	5	Sunrun, Inc.	8	6
6	Beazer Homes USA, Inc. (BZH)	7	0	6	GE Vernova, Inc.	4	6
7	Armstrong World Indus. (AWI)	8	0	7	Berkshire Hathaway, Inc.	6	11
8	Meritage Homes Corp. (MTH)	7	0	8	Enphase Energy, Inc.	9	7
9	SRM57D-R	7	0	9	First Solar, Inc.	4	7
10	Kilroy Realty Corp. (KRC)	5	1	10	SolarEdge Technologies, In.	9	6

Energy Efficiency — New: 8.3 Base: 5.0 Δ: +3.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Constellation Energy . (CEG)	5	3	1	Constellation Energy Corp.	5	94
2	PG&E Corp. (PCG)	7	2	2	Microsoft Corp.	4	105
3	Trane Technologies Plc (TT)	10	3	3	Amazon.com, Inc.	5	90
4	Ameresco, Inc. (AMRC)	10	0	4	Tesla, Inc.	7	97
5	XHDDQT-R	9	0	5	Alphabet, Inc.	6	61
6	Energy Focus, Inc. (EFOI)	9	0	6	GE Vernova, Inc.	8	52
7	Willdan Group, Inc. (WLDN)	9	0	7	Vistra Corp.	4	56
8	Energy Recovery, Inc. (ERII)	10	0	8	NextEra Energy, Inc.	5	49
9	Avista Corp. (AVA)	5	0	9	Oklo, Inc.	4	35
10	Orion Energy Systems,. (OESX)	9	0	10	Meta Platforms, Inc.	2	36

Climate Change Mitigation — New: 6.7 Base: 4.8 Δ: +1.9 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	HA Sustainable Infras. (HASI)	10	1	1	Microsoft Corp.	6	52
2	Chevron Corp. (CVX)	3	5	2	Exxon Mobil Corp.	2	38
3	Occidental Petroleum . (OXY)	2	2	3	Chevron Corp.	3	29
4	Gevo, Inc. (GEVO)	8	0	4	Amazon.com, Inc.	4	25
5	Aemetis, Inc. (AMTX)	8	0	5	Alphabet, Inc.	6	20
6	QWY3DR-R	5	0	6	Occidental Petroleum Corp.	2	28
7	The AES Corp. (AES)	8	0	7	Tesla, Inc.	10	27
8	Hudson Technologies, . (HDSN)	8	0	8	Constellation Energy Corp.	10	15
9	REX American Resource. (REX)	7	0	9	Meta Platforms, Inc.	2	15
10	Tetra Tech, Inc. (TTEK)	8	0	10	Morgan Stanley	3	15

ESG Leaders — New: 8.0 Base: 6.4 Δ: +1.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	MSCI, Inc. (MSCI)	10	8	1	Microsoft Corp.	9	6
2	Morningstar, Inc. (MORN)	10	1	2	Morgan Stanley	5	6
3	Ecolab, Inc. (ECL)	9	1	3	JPMorgan Chase & Co.	6	6
4	HA Sustainable Infras. (HASI)	10	1	4	Exxon Mobil Corp.	1	5
5	Jones Lang LaSalle, I. (JLL)	8	1	5	The Goldman Sachs Group, I.	7	6
6	Affiliated Managers G. (AMG)	5	1	6	State Street Corp.	8	5
7	S&P Global, Inc. (SPGI)	7	1	7	BlackRock, Inc.	8	3
8	First Solar, Inc. (FSLR)	10	2	8	NVIDIA Corp.	4	4
9	CI Health Care Giants. (FHI)	1	0	9	UBS Group AG	6	3
10	NextEra Energy, Inc. (NEE)	10	1	10	Morningstar, Inc.	10	4

Healthcare Innovation — New: 8.1 Base: 9.1 Δ: -1.0 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	UnitedHealth Group, I. (UNH)	7	113	1	Eli Lilly & Co.	9	160
2	iRhythm Holdings, Inc. (IRTC)	9	11	2	Novo Nordisk A/S	9	97
3	DexCom, Inc. (DXCM)	10	16	3	Pfizer Inc.	10	92
4	Intuitive Surgical, I. (ISRG)	10	5	4	Merck & Co., Inc.	10	54
5	Privia Health Group, . (PRVA)	8	3	5	Johnson & Johnson	9	45
6	Oscar Health, Inc. (OSCR)	6	3	6	AstraZeneca PLC	8	44
7	DaVita, Inc. (DVA)	6	6	7	AstraZeneca PLC	8	44
8	Tempus AI, Inc. (TEM)	10	2	8	Moderna, Inc.	10	44
9	Ardent Health, Inc. (ARDT)	6	2	9	Amgen, Inc.	10	31
10	Teladoc Health, Inc. (TDOC)	9	17	10	AbbVie, Inc.	8	31

Biotechnology — New: 8.3 Base: 8.5 Δ: -0.2 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Illumina, Inc. (ILMN)	10	11	1	Pfizer Inc.	9	55
2	Repligen Corp. (RGEN)	8	1	2	Moderna, Inc.	10	35
3	Intellia Therapeutics. (NTLA)	10	19	3	Eli Lilly & Co.	7	23
4	Sangamo Therapeutics,. (SGMO)	8	4	4	BioNTech SE	10	23
5	Cibus, Inc. (CBUS)	5	1	5	Novartis AG	8	21
6	uniQure NV (QURE)	8	6	6	GSK Plc	7	18
7	Editas Medicine, Inc. (EDIT)	9	12	7	CureVac NV	9	14
8	Eli Lilly & Co. (LLY)	7	23	8	Bristol Myers Squibb Co.	9	15
9	BioMarin Pharmaceutic. (BMRN)	9	6	9	Merck & Co., Inc.	8	21
10	Twist Bioscience Corp. (TWST)	9	2	10	Novo Nordisk A/S	8	12

Gene Editing — New: 9.4 Base: 5.5 Δ: +3.9 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	CRISPR Therapeutics AG (CRSP)	10	36	1	Pfizer Inc.	6	9
2	Beam Therapeutics, In. (BEAM)	10	16	2	Novo Nordisk A/S	1	4
3	Editas Medicine, Inc. (EDIT)	10	18	3	CRISPR Therapeutics AG	10	9
4	Intellia Therapeutics. (NTLA)	10	31	4	Intellia Therapeutics, Inc.	10	8
5	Vertex Pharmaceutical. (VRTX)	9	21	5	Eli Lilly & Co.	4	5
6	Caribou Biosciences, . (CRBU)	9	5	6	Vertex Pharmaceuticals, In.	9	8
7	Wave Life Sciences Lt. (WVE)	7	1	7	Novartis AG	5	4
8	Sangamo Therapeutics,. (SGMO)	10	0	8	uniQure NV	4	3
9	Precision BioSciences. (DTIL)	9	0	9	Insmed, Inc.	1	3
10	Prime Medicine, Inc. (PRME)	10	0	10	Sarepta Therapeutics, Inc.	5	7

mRNA Technology — New: 7.0 Base: 6.9 Δ: +0.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Moderna, Inc. (MRNA)	10	159	1	Moderna, Inc.	10	151
2	Arcturus Therapeutics. (ARCT)	10	9	2	Pfizer Inc.	10	114
3	Alnylam Pharmaceutica. (ALNY)	6	4	3	BioNTech SE	10	65
4	Wave Life Sciences Lt. (WVE)	5	1	4	Novavax, Inc.	3	43
5	Sarepta Therapeutics,. (SRPT)	4	1	5	GSK Plc	6	41
6	Novavax, Inc. (NVAX)	3	8	6	CureVac NV	9	15
7	Ernexa Therapeutics, . (ERNA)	8	0	7	Sanofi	7	22
8	M56G62-R	9	0	8	Merck & Co., Inc.	7	23
9	Beam Therapeutics, In. (BEAM)	7	2	9	Johnson & Johnson	3	15
10	Arbutus Biopharma Cor. (ABUS)	8	2	10	AstraZeneca PLC	4	13

Immunotherapy — New: 8.4 Base: 8.3 Δ: +0.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	ImmunityBio, Inc. (IBRX)	9	1	1	Pfizer Inc.	5	15
2	Janux Therapeutics, I. (JANX)	9	2	2	Merck & Co., Inc.	10	16
3	Summit Therapeutics, . (SMMT)	8	2	3	Moderna, Inc.	8	14
4	Iovance Biotherapeuti. (IOVA)	10	2	4	Bristol Myers Squibb Co.	10	9
5	Kymera Therapeutics, . (KYMR)	4	1	5	BioNTech SE	10	7
6	Vir Biotechnology, In. (VIR)	6	2	6	CureVac NV	5	5
7	Merck & Co., Inc. (MRK)	10	25	7	Merus NV	9	4
8	Fate Therapeutics, In. (FATE)	9	2	8	AstraZeneca PLC	8	7
9	Bristol Myers Squibb . (BMY)	10	18	9	AstraZeneca PLC	8	7
10	Arcus Biosciences, In. (RCUS)	9	1	10	Genmab A/S	10	3

Neurotechnology — New: 6.9 Base: 3.8 Δ: +3.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	NeuroOne Medical Tech. (NMTC)	8	0	1	NVIDIA Corp.	6	234
2	Neonode, Inc. (NEON)	2	0	2	Meta Platforms, Inc.	8	126
3	HG8197-R	8	0	3	Microsoft Corp.	3	117
4	Firefly Neuroscience,. (AIFF)	9	0	4	Amazon.com, Inc.	2	75
5	NeuroPace, Inc. (NPCE)	10	0	5	Advanced Micro Devices, In.	5	66
6	NVE Corp. (NVEC)	3	0	6	Alphabet, Inc.	5	75
7	Stryker Corp. (SYK)	8	0	7	Apple, Inc.	3	74
8	C3PD4C-R	3	0	8	Broadcom Inc.	2	47
9	D7LJD9-R	9	0	9	Intel Corp.	3	37
10	Solana Co. (HSDT)	9	0	10	Tesla, Inc.	1	31

Aging Population — New: 8.8 Base: 8.3 Δ: +0.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Brookdale Senior Livi. (BKD)	10	1	1	Eli Lilly & Co.	7	45
2	Ventas, Inc. (VTR)	10	3	2	Biogen, Inc.	9	31
3	UnitedHealth Group, I. (UNH)	10	6	3	Novo Nordisk A/S	8	21
4	Genworth Financial, I. (GNW)	9	1	4	Humana, Inc.	10	30
5	The Ensign Group, Inc. (ENSG)	10	1	5	UnitedHealth Group, Inc.	10	35
6	Humana, Inc. (HUM)	10	2	6	Pfizer Inc.	7	11
7	Welltower, Inc. (WELL)	10	3	7	CVS Health Corp.	9	20
8	BDP6Z1-R	1	0	8	Amgen, Inc.	8	7
9	PQLX2F-R	9	0	9	Merck & Co., Inc.	7	6
10	The Pennant Group, In. (PNTG)	9	0	10	AbbVie, Inc.	8	5

Longevity Science — New: 7.0 Base: 8.1 Δ: -1.1 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	D0W54X-R	9	0	1	Eli Lilly & Co.	9	22
2	Longeveron, Inc. (LGVN)	9	0	2	Novo Nordisk A/S	8	14
3	LifeVantage Corp. (LFVN)	5	0	3	EXACT Sciences Corp.	9	3
4	BioAge Labs, Inc. (BIOA)	9	0	4	GRAIL, Inc.	10	3
5	Telomir Pharmaceutica. (TELO)	9	0	5	Guardant Health, Inc.	9	3
6	WHF5LX-R	8	0	6	CRISPR Therapeutics AG	9	7
7	Niagen Bioscience, In. (NAGE)	6	0	7	Vertex Pharmaceuticals, In.	7	7
8	HD2YB8-R	3	0	8	NVIDIA Corp.	6	3
9	FibroBiologics, Inc. (FBLG)	6	0	9	Johnson & Johnson	7	3
10	Geron Corp. (GERN)	6	0	10	Amgen, Inc.	7	3

Digital Health & Telemedicine — New: 8.6 Base: 6.9 Δ: +1.7 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Hims & Hers Health, I. (HIMS)	9	22	1	Eli Lilly & Co.	4	41
2	Teladoc Health, Inc. (TDOC)	10	46	2	UnitedHealth Group, Inc.	7	41
3	iRhythm Holdings, Inc. (IRTC)	9	7	3	CVS Health Corp.	8	38
4	LifeMD, Inc. (LFMD)	10	5	4	Amazon.com, Inc.	8	44
5	DexCom, Inc. (DXCM)	9	12	5	Novo Nordisk A/S	5	28
6	Doximity, Inc. (DOCS)	9	5	6	Microsoft Corp.	7	28
7	UnitedHealth Group, I. (UNH)	7	14	7	Apple, Inc.	8	26
8	Ardent Health, Inc. (ARDT)	4	2	8	Humana, Inc.	7	18
9	ResMed, Inc. (RMD)	9	1	9	Hims & Hers Health, Inc.	9	15
10	Talkspace, Inc. (TALK)	10	2	10	Alphabet, Inc.	6	20

Pharmaceutical R&D — New: 8.5 Base: 9.8 Δ: -1.3 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Viking Therapeutics, . (VKTX)	8	31	1	Pfizer Inc.	10	128
2	Vertex Pharmaceutical. (VRTX)	10	17	2	Eli Lilly & Co.	10	107
3	Insmed, Inc. (INSM)	8	4	3	Novo Nordisk A/S	9	65
4	Altimmune, Inc. (ALT)	8	3	4	Merck & Co., Inc.	10	71
5	Denali Therapeutics, . (DNLI)	9	4	5	AstraZeneca PLC	10	55
6	KalVista Pharmaceutic. (KALV)	7	1	6	AstraZeneca PLC	10	55
7	Agios Pharmaceuticals. (AGIO)	9	2	7	Novartis AG	10	46
8	Madrigal Pharmaceutic. (MDGL)	8	7	8	Johnson & Johnson	9	42
9	Terns Pharmaceuticals. (TERN)	9	2	9	Bristol Myers Squibb Co.	10	40
10	Revolution Medicines,. (RVMD)	9	4	10	Moderna, Inc.	10	40

Mental Health & Wellness — New: 8.1 Base: 7.1 Δ: +1.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Acadia Healthcare Co.. (ACHC)	10	1	1	Johnson & Johnson	7	8
2	Lifestance Health Gro. (LFST)	10	2	2	Hims & Hers Health, Inc.	8	6
3	Teladoc Health, Inc. (TDOC)	9	3	3	Atai Beckley NV	10	4
4	Talkspace, Inc. (TALK)	10	3	4	Novo Nordisk A/S	4	5
5	Neurocrine Bioscience. (NBIX)	9	2	5	Compass Pathways Plc	10	3
6	T4R0BD-R	7	0	6	Eli Lilly & Co.	7	4
7	Universal Health Serv. (UHS)	9	0	7	Microsoft Corp.	4	4
8	UnitedHealth Group, I. (UNH)	8	1	8	Meta Platforms, Inc.	2	5
9	Hims & Hers Health, I. (HIMS)	8	0	9	Talkspace, Inc.	10	3
10	Nakamoto, Inc. (NAKA)	1	0	10	Teladoc Health, Inc.	9	3

Nutrition & Supplements — New: 8.5 Base: 4.9 Δ: +3.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Celsius Holdings, Inc. (CELH)	7	2	1	Novo Nordisk A/S	3	42
2	The Simply Good Foods. (SMPL)	8	2	2	Eli Lilly & Co.	3	38
3	BellRing Brands, Inc. (BRBR)	10	1	3	PepsiCo, Inc.	6	21
4	Hormel Foods Corp. (HRL)	4	1	4	The Kraft Heinz Co.	5	13
5	FitLife Brands, Inc. (FTLF)	9	0	5	Amazon.com, Inc.	8	13
6	LifeVantage Corp. (LFVN)	9	0	6	Conagra Brands, Inc.	5	12
7	Herbalife Ltd. (HLF)	10	0	7	Walmart, Inc.	6	13
8	Natural Alternatives . (NAII)	10	0	8	General Mills, Inc.	5	11
9	Nature's Sunshine Pro. (NATR)	9	0	9	The Hershey Co.	2	11
10	USANA Health Sciences. (USNA)	9	0	10	The Coca-Cola Co.	6	8

US Infrastructure — New: 8.9 Base: 6.4 Δ: +2.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Sterling Infrastructu. (STRL)	9	1	1	NVIDIA Corp.	5	4
2	Granite Construction,. (GVA)	9	1	2	Microsoft Corp.	6	4
3	MasTec, Inc. (MTZ)	9	2	3	Dominion Energy, Inc.	9	4
4	Quanta Services, Inc. (PWR)	10	1	4	Amazon.com, Inc.	6	3
5	MYR Group, Inc. (MYRG)	9	1	5	Algonquin Power & Utilitie.	8	3
6	Shimmick Corp. (SHIM)	8	0	6	Meta Platforms, Inc.	2	2
7	Primoris Services Cor. (PRIM)	9	0	7	KKR & Co., Inc.	5	2
8	Tutor Perini Corp. (TPC)	8	0	8	Blackstone, Inc.	5	2
9	Vulcan Materials Co. (VMC)	10	3	9	Quanta Services, Inc.	10	3
10	Insteel Industries, I. (IIIN)	8	1	10	Consolidated Edison, Inc.	8	3

Transportation & Logistics Tech — New: 8.6 Base: 8.3 Δ: +0.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	C.H. Robinson Worldwi. (CHRW)	9	8	1	Amazon.com, Inc.	9	74
2	J.B. Hunt Transport S. (JBHT)	9	17	2	FedEx Corp.	10	68
3	XPO, Inc. (XPO)	9	24	3	United Parcel Service, Inc.	10	70
4	Norfolk Southern Corp. (NSC)	7	15	4	Uber Technologies, Inc.	9	46
5	Old Dominion Freight . (ODFL)	9	5	5	Tesla, Inc.	9	39
6	Union Pacific Corp. (UNP)	8	18	6	Walmart, Inc.	5	44
7	Ryder System, Inc. (R)	8	7	7	Alphabet, Inc.	8	32
8	Knight-Swift Transpor. (KNX)	8	5	8	Union Pacific Corp.	8	31
9	FedEx Corp. (FDX)	10	30	9	Norfolk Southern Corp.	7	34
10	Landstar System, Inc. (LSTR)	9	1	10	Lyft, Inc.	8	19

Construction & Housing — New: 9.4 Base: 8.8 Δ: +0.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Lennar Corp. (LEN)	10	24	1	Lennar Corp.	10	32
2	D.R. Horton, Inc. (DHI)	10	27	2	D.R. Horton, Inc.	10	34
3	PulteGroup, Inc. (PHM)	10	17	3	PulteGroup, Inc.	10	22
4	Toll Brothers, Inc. (TOL)	9	18	4	Toll Brothers, Inc.	9	18
5	Builders FirstSource,. (BLDR)	10	5	5	CRH Plc	10	7
6	KB Home (KBH)	9	8	6	CoStar Group, Inc.	6	10
7	Beazer Homes USA, Inc. (BZH)	9	1	7	Berkshire Hathaway, Inc.	8	5
8	Boise Cascade Co. (BCC)	9	4	8	Equity Residential	9	8
9	Weyerhaeuser Co. (WY)	9	5	9	CBRE Group, Inc.	7	9
10	Meritage Homes Corp. (MTH)	9	8	10	Invitation Homes, Inc.	9	8

Smart Cities — New: 8.1 Base: 7.0 Δ: +1.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Itron, Inc. (ITRI)	10	0	1	NVIDIA Corp.	9	70
2	Iveda Solutions, Inc. (IVDA)	9	0	2	Alphabet, Inc.	8	73
3	Q2VHYL-R	9	0	3	Tesla, Inc.	7	64
4	Beam Global (BEEM)	9	0	4	Amazon.com, Inc.	7	56
5	Datasea, Inc. (DTSS)	3	0	5	Microsoft Corp.	8	59
6	Rekor Systems, Inc. (REKR)	9	0	6	Meta Platforms, Inc.	3	37
7	SmartRent, Inc. (SMRT)	7	0	7	Uber Technologies, Inc.	7	39
8	Willdan Group, Inc. (WLDN)	8	0	8	Constellation Energy Corp.	7	23
9	Acuity, Inc. (AYI)	8	0	9	General Motors Co.	8	27
10	Tyler Technologies, I. (TYL)	9	0	10	Oracle Corp.	6	17

Utilities — New: 9.4 Base: 4.9 Δ: +4.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Northwest Natural Hol. (NWN)	9	0	1	Amazon.com, Inc.	1	14
2	QMX32N-R	9	0	2	Constellation Energy Corp.	10	13
3	National Fuel Gas Co. (NFG)	8	0	3	GE Vernova, Inc.	6	11
4	DTLWYY-R	10	0	4	Alphabet, Inc.	1	11
5	The AES Corp. (AES)	9	0	5	Meta Platforms, Inc.	1	9
6	Dominion Energy, Inc. (D)	10	0	6	Vistra Corp.	9	10
7	Avista Corp. (AVA)	9	0	7	Talen Energy Corp.	8	10
8	Entergy Corp. (ETR)	10	0	8	Microsoft Corp.	1	9
9	WEC Energy Group, Inc. (WEC)	10	0	9	NuScale Power Corp.	6	9
10	American Electric Pow. (AEP)	10	0	10	Centrus Energy Corp.	6	8

Real Estate — New: 8.4 Base: 8.3 Δ: +0.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Douglas Elliman, Inc. (DOUG)	8	5	1	News Corp.	6	3
2	Lennar Corp. (LEN)	9	5	2	Public Storage	10	2
3	American Homes 4 Rent (AMH)	10	9	3	Equinix, Inc.	9	2
4	CoStar Group, Inc. (CSGP)	9	2	4	NNN REIT, Inc.	9	1
5	Meritage Homes Corp. (MTH)	8	2	5	Realty Income Corp.	10	1
6	Zillow Group, Inc. (Z)	8	10	6	Sabra Health Care REIT, In.	9	1
7	D.R. Horton, Inc. (DHI)	9	5	7	AT&T, Inc.	2	1
8	eXp World Holdings, I. (EXPI)	7	1	8	American Tower Corp.	9	1
9	RE/MAX Holdings, Inc. (RMAX)	6	2	9	Americold Realty Trust, In.	10	1
10	Sun Communities, Inc. (SUI)	10	1	10	Crown Castle, Inc.	9	1

Data Centers — New: 8.1 Base: 9.0 Δ: -0.9 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Digital Realty Trust,. (DLR)	10	10	1	NVIDIA Corp.	10	202
2	Equinix, Inc. (EQIX)	10	24	2	Microsoft Corp.	10	185
3	Vertiv Holdings Co. (VRT)	10	5	3	Amazon.com, Inc.	10	132
4	Applied Digital Corp. (APLD)	9	2	4	Meta Platforms, Inc.	8	100
5	Core Scientific, Inc. (CORZ)	7	1	5	Oracle Corp.	8	80
6	Cogent Communications. (CCOI)	7	2	6	Alphabet, Inc.	9	83
7	DigitalBridge Group, . (DBRG)	9	3	7	CoreWeave, Inc.	10	33
8	American Tower Corp. (AMT)	3	2	8	Constellation Energy Corp.	6	43
9	Entergy Corp. (ETR)	6	2	9	Advanced Micro Devices, In.	9	28
10	NVIDIA Corp. (NVDA)	10	11	10	Vertiv Holdings Co.	10	23

Global Commodities — New: 9.3 Base: 8.3 Δ: +1.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Archer-Daniels-Midlan. (ADM)	10	1	1	Exxon Mobil Corp.	10	54
2	Bunge Global SA (BG)	10	1	2	Chevron Corp.	10	48
3	CME Group, Inc. (CME)	8	1	3	BP Plc	10	24
4	iShares S&P GSCI Comm. (GSG)	10	1	4	Shell Plc	10	19
5	Invesco DB Commodity . (DBC)	10	1	5	Newmont Corp.	10	15
6	abrdn Physical Pallad. (PALL)	10	1	6	The Goldman Sachs Group, I.	5	21
7	Invesco DB Agricultur. (DBA)	9	0	7	JPMorgan Chase & Co.	5	17
8	Newmont Corp. (NEM)	10	1	8	TotalEnergies SE	10	13
9	Intercontinental Exch. (ICE)	8	0	9	Freeport-McMoRan, Inc.	10	12
10	Canlan Ice Sports Cor. (ICE)	8	0	10	Costco Wholesale Corp.	3	8

Gold & Precious Metals — New: 7.5 Base: 7.4 Δ: +0.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Newmont Corp. (NEM)	10	128	1	Newmont Corp.	10	101
2	SPDR Gold Shares (GLD)	10	52	2	SPDR Gold Shares	10	43
3	Royal Gold, Inc. (RGLD)	10	7	3	Agnico Eagle Mines Ltd.	10	32
4	Hecla Mining Co. (HL)	8	1	4	JPMorgan Chase & Co.	3	29
5	abrdn Physical Pallad. (PALL)	9	1	5	The Goldman Sachs Group, I.	5	22
6	iShares Silver Trust (SLV)	10	3	6	Barrick Mining Corp.	10	17
7	NovaGold Resources, I. (NG)	8	3	7	Costco Wholesale Corp.	2	21
8	Freeport-McMoRan, Inc. (FCX)	6	2	8	BlackRock, Inc.	5	13
9	Southern Copper Corp. (SCCO)	3	1	9	Franco-Nevada Corp.	9	12
10	Invesco DB Agricultur. (DBA)	1	0	10	Wheaton Precious Metals Co.	10	10

Oil & Gas — New: 8.4 Base: 9.1 Δ: -0.7 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Occidental Petroleum . (OXY)	10	83	1	Exxon Mobil Corp.	10	412
2	ConocoPhillips (COP)	10	59	2	Chevron Corp.	10	342
3	Devon Energy Corp. (DVN)	10	25	3	Shell Plc	10	159
4	Expand Energy Corp. (EXE)	1	17	4	BP Plc	10	153
5	Crescent Energy Co. (CRGY)	9	2	5	Occidental Petroleum Corp.	10	141
6	Matador Resources Co. (MTDR)	8	10	6	ConocoPhillips	10	116
7	Comstock Resources, I. (CRK)	9	10	7	Diamondback Energy, Inc.	10	82
8	EQT Corp. (EQT)	10	17	8	TotalEnergies SE	10	68
9	Range Resources Corp. (RRC)	8	12	9	Expand Energy Corp.	1	57
10	APA Corp. (APA)	9	8	10	EQT Corp.	10	51

Natural Gas — New: 9.0 Base: 8.0 Δ: +1.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Expand Energy Corp. (EXE)	3	57	1	Exxon Mobil Corp.	8	63
2	EQT Corp. (EQT)	10	86	2	Chevron Corp.	8	58
3	The Williams Cos., In. (WMB)	10	19	3	EQT Corp.	10	35
4	Cheniere Energy, Inc. (LNG)	10	38	4	Shell Plc	10	38
5	Comstock Resources, I. (CRK)	10	17	5	Expand Energy Corp.	3	31
6	Antero Resources Corp. (AR)	10	22	6	BP Plc	9	27
7	Range Resources Corp. (RRC)	10	24	7	TotalEnergies SE	9	22
8	Devon Energy Corp. (DVN)	8	5	8	Cheniere Energy, Inc.	10	15
9	Kinder Morgan, Inc. (KMI)	10	14	9	Occidental Petroleum Corp.	7	19
10	National Fuel Gas Co. (NFG)	9	1	10	Constellation Energy Corp.	6	13

USD Strength / FX Hedge — New: 8.9 Base: 5.5 Δ: +3.4 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Invesco DB US Dollar . (UUP)	10	0	1	Bank of America Corp.	6	7
2	Invesco CurrencyShare. (FXF)	10	0	2	Newmont Corp.	3	8
3	Invesco CurrencyShare. (FXE)	10	0	3	The Goldman Sachs Group, I.	8	6
4	Invesco Currencyshare. (FXY)	8	0	4	Agnico Eagle Mines Ltd.	6	5
5	WisdomTree, Inc. (WT)	9	0	5	CME Group, Inc.	8	3
6	VSJ28G-R	10	0	6	JPMorgan Chase & Co.	8	4
7	Philip Morris Interna. (PM)	9	4	7	Deutsche Bank AG	4	4
8	The Western Union Co. (WU)	8	0	8	Tesla, Inc.	3	3
9	iShares Silver Trust (SLV)	7	1	9	Costco Wholesale Corp.	5	3
10	CME Group, Inc. (CME)	8	2	10	BlackRock, Inc.	4	2

Inflation Protection — New: 7.7 Base: 6.3 Δ: +1.4 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	SPDR Gold Shares (GLD)	10	4	1	Bank of America Corp.	6	16
2	Invesco DB Agricultur. (DBA)	8	0	2	Costco Wholesale Corp.	7	16
3	abrdn Physical Platin. (PPLT)	8	0	3	The Goldman Sachs Group, I.	4	13
4	iShares S&P GSCI Comm. (GSG)	9	0	4	UBS Group AG	4	11
5	Invesco DB US Dollar . (UUP)	2	0	5	Newmont Corp.	9	13
6	Invesco DB Precious M. (DBP)	9	0	6	General Mills, Inc.	6	14
7	Invesco CurrencyShare. (FXE)	4	0	7	The Kraft Heinz Co.	6	13
8	iShares Silver Trust (SLV)	8	0	8	The Coca-Cola Co.	7	11
9	Invesco DB Commodity . (DBC)	10	0	9	Morgan Stanley	4	7
10	abrdn Precious Metals. (GLTR)	9	0	10	SPDR Gold Shares	10	11

Impact Investing — New: 6.8 Base: 4.5 Δ: +2.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	KKR & Co., Inc. (KKR)	7	3	1	The Carlyle Group Inc.	4	11
2	GCM Grosvenor, Inc. (GCMG)	4	3	2	KKR & Co., Inc.	7	11
3	TPG, Inc. (TPG)	8	5	3	BlackRock, Inc.	8	7
4	WisdomTree, Inc. (WT)	3	2	4	Microsoft Corp.	4	7
5	Affiliated Managers G. (AMG)	5	2	5	Exxon Mobil Corp.	2	9
6	BlackRock, Inc. (BLK)	8	2	6	Brookfield Asset Managemen.	8	9
7	Ridgepost Capital, In. (RPC)	8	0	7	The Goldman Sachs Group, I.	4	11
8	Devvstream Corp. (DEVS)	9	0	8	Chevron Corp.	1	7
9	HA Sustainable Infras. (HASI)	10	0	9	Blackstone, Inc.	1	9
10	Invesco Ltd. (IVZ)	6	1	10	Bank of America Corp.	6	8

ESports & Gaming — New: 7.9 Base: 7.8 Δ: +0.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Electronic Arts, Inc. (EA)	10	170	1	Electronic Arts, Inc.	10	89
2	Take-Two Interactive . (TTWO)	10	70	2	Microsoft Corp.	10	78
3	DraftKings, Inc. (DKNG)	8	13	3	Take-Two Interactive Softw.	10	53
4	Microsoft Corp. (MSFT)	10	109	4	DraftKings, Inc.	8	61
5	GameStop Corp. (GME)	7	10	5	Sony Group Corp.	9	54
6	PENN Entertainment, I. (PENN)	3	16	6	The Walt Disney Co.	4	43
7	Flutter Entertainment. (FLUT)	6	11	7	Flutter Entertainment Plc	6	37
8	Madison Square Garden. (MSGS)	5	1	8	Netflix, Inc.	3	34
9	Roblox Corp. (RBLX)	10	10	9	Roblox Corp.	10	23
10	NVIDIA Corp. (NVDA)	10	11	10	Amazon.com, Inc.	8	28

Sports Betting & Gambling — New: 8.9 Base: 6.1 Δ: +2.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	DraftKings, Inc. (DKNG)	10	97	1	DraftKings, Inc.	10	133
2	Flutter Entertainment. (FLUT)	10	117	2	Flutter Entertainment Plc	10	77
3	PENN Entertainment, I. (PENN)	9	58	3	The Walt Disney Co.	5	82
4	Churchill Downs, Inc. (CHDN)	9	5	4	Netflix, Inc.	1	66
5	MGM Resorts Internati. (MGM)	10	46	5	Amazon.com, Inc.	2	47
6	Boyd Gaming Corp. (BYD)	9	5	6	Fox Corp.	8	38
7	Caesars Entertainment. (CZR)	9	38	7	MGM Resorts International	10	37
8	Wynn Resorts Ltd. (WYNN)	9	5	8	Comcast Corp.	4	42
9	Rush Street Interacti. (RSI)	9	2	9	Robinhood Markets, Inc.	2	26
10	FuboTV, Inc. (FUBO)	5	4	10	PENN Entertainment, Inc.	9	38

Media & Streaming — New: 9.0 Base: 8.8 Δ: +0.2 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Netflix, Inc. (NFLX)	10	53	1	Netflix, Inc.	10	359
2	Warner Bros. Discover. (WBD)	9	23	2	The Walt Disney Co.	10	348
3	Roku, Inc. (ROKU)	10	12	3	Warner Bros. Discovery, In.	9	282
4	Amazon.com, Inc. (AMZN)	9	26	4	Comcast Corp.	9	221
5	FuboTV, Inc. (FUBO)	9	1	5	Amazon.com, Inc.	9	235
6	Paramount Skydance Co. (PSKY)	9	2	6	Fox Corp.	9	114
7	Charter Communication. (CHTR)	7	1	7	Apple, Inc.	8	108
8	Comcast Corp. (CMCSA)	9	21	8	Alphabet, Inc.	10	61
9	Apple, Inc. (AAPL)	8	17	9	Charter Communications, In.	7	52
10	The Walt Disney Co. (DIS)	10	40	10	News Corp.	7	55

Social Media — New: 7.9 Base: 7.4 Δ: +0.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Meta Platforms, Inc. (META)	10	109	1	Meta Platforms, Inc.	10	59
2	Snap, Inc. (SNAP)	10	44	2	Pinterest, Inc.	10	14
3	Trump Media & Technol. (DJT)	9	3	3	Alphabet, Inc.	10	14
4	Pinterest, Inc. (PINS)	10	19	4	Reddit, Inc.	10	11
5	Reddit, Inc. (RDDT)	10	4	5	Snap, Inc.	10	19
6	Alphabet, Inc. (GOOGL)	10	25	6	Apple, Inc.	4	12
7	Sprout Social, Inc. (SPT)	10	0	7	Amazon.com, Inc.	6	8
8	IZEA Worldwide, Inc. (IZEA)	8	0	8	Applovin Corp.	4	3
9	Datacentrex, Inc. (DTCX)	1	0	9	Trump Media & Technology G.	9	4
10	Strive, Inc. (United . (ASST)	1	0	10	Mizuho Financial Group, In.	1	2

Digital Advertising — New: 8.6 Base: 7.0 Δ: +1.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	The Trade Desk, Inc. (TTD)	10	14	1	Meta Platforms, Inc.	10	237
2	Nexstar Media Group, . (NXST)	5	2	2	Amazon.com, Inc.	9	148
3	Teads Holding Co. (TEAD)	9	1	3	Alphabet, Inc.	10	135
4	PubMatic, Inc. (PUBM)	9	5	4	Apple, Inc.	5	110
5	Pinterest, Inc. (PINS)	9	13	5	Omnicom Group, Inc.	9	53
6	Omnicom Group, Inc. (OMC)	9	42	6	WPP Plc	1	57
7	USA TODAY Co., Inc. (TDAY)	8	6	7	Netflix, Inc.	2	65
8	The Arena Group Holdi. (AREN)	7	2	8	Microsoft Corp.	6	59
9	Alphabet, Inc. (GOOG)	10	268	9	Walmart, Inc.	9	48
10	Alphabet, Inc. (GOOGL)	10	268	10	Reddit, Inc.	9	30

Luxury Goods — New: 8.2 Base: 6.9 Δ: +1.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Capri Holdings Ltd. (CPRI)	9	37	1	Tapestry, Inc.	9	38
2	Tapestry, Inc. (TPR)	9	35	2	Capri Holdings Ltd.	9	36
3	Ralph Lauren Corp. (RL)	9	15	3	Ralph Lauren Corp.	9	25
4	The Estée Lauder Comp. (EL)	9	5	4	Ferrari NV	10	25
5	Macy's, Inc. (M)	6	5	5	The RealReal, Inc.	8	24
6	RH (RH)	9	1	6	Tesla, Inc.	8	26
7	Signet Jewelers Ltd. (SIG)	7	2	7	NIKE, Inc.	4	21
8	Coty, Inc. (COTY)	8	1	8	The Estée Lauder Companies.	9	20
9	Vince Holding Corp. (VNCE)	7	0	9	Walmart, Inc.	2	12
10	Movado Group, Inc. (MOV)	9	0	10	Bank of America Corp.	1	12

Pets Economy — New: 9.4 Base: 4.6 Δ: +4.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Petco Health & Wellne. (WOOF)	10	12	1	Chewy, Inc.	10	20
2	Chewy, Inc. (CHWY)	10	11	2	Costco Wholesale Corp.	5	6
3	Freshpet, Inc. (FRPT)	9	1	3	Petco Health & Wellness Co.	10	7
4	Trupanion, Inc. (TRUP)	10	2	4	Freshpet, Inc.	9	3
5	Zoetis, Inc. (ZTS)	10	4	5	Bank of America Corp.	1	3
6	Tractor Supply Co. (TSCO)	8	1	6	Walmart, Inc.	6	2
7	N36LSM-R	10	0	7	UBS Group AG	1	2
8	PetMed Express, Inc. (PETS)	9	0	8	Mastercard, Inc.	2	2
9	Central Garden & Pet . (CENT)	9	0	9	Morgan Stanley	1	2
10	Colgate-Palmolive Co. (CL)	9	1	10	GameStop Corp.	1	2

Food Technology — New: 7.0 Base: 5.5 Δ: +1.5 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Sensient Technologies. (SXT)	9	1	1	PepsiCo, Inc.	5	17
2	Hormel Foods Corp. (HRL)	5	1	2	Starbucks Corp.	8	10
3	Tyson Foods, Inc. (TSN)	6	6	3	Amazon.com, Inc.	7	10
4	Conagra Brands, Inc. (CAG)	5	7	4	Mondelez International, In.	5	9
5	The Campbell's Co. (CPB)	4	5	5	The Kraft Heinz Co.	5	8
6	Neogen Corp. (NEOG)	9	1	6	McDonald's Corp.	8	9
7	McCormick & Co., Inc. (MKC)	8	3	7	General Mills, Inc.	5	8
8	JBT Marel Corp. (JBTM)	10	0	8	Walmart, Inc.	5	8
9	General Mills, Inc. (GIS)	5	10	9	Target Corp.	3	6
10	The Middleby Corp. (MIDD)	9	0	10	The J. M. Smucker Co.	4	6

Online Education — New: 8.5 Base: 6.4 Δ: +2.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Stride, Inc. (LRN)	10	4	1	Coursera, Inc.	10	3
2	Strategic Education, . (STRA)	8	1	2	Microsoft Corp.	8	3
3	Grand Canyon Educatio. (LOPE)	9	2	3	Chegg, Inc.	9	2
4	American Public Educa. (APEI)	9	1	4	Alphabet, Inc.	8	2
5	Chegg, Inc. (CHGG)	9	7	5	News Corp.	3	1
6	Coursera, Inc. (COUR)	10	4	6	NVIDIA Corp.	5	2
7	HHK128-R	10	0	7	Salesforce, Inc.	4	1
8	M2QYMM-R	1	0	8	Amazon.com, Inc.	4	2
9	Perdoceo Education Co. (PRDO)	9	0	9	Duolingo, Inc.	10	1
10	Duolingo, Inc. (DUOL)	10	1	10	Meta Platforms, Inc.	3	1

Remote Work Tech — New: 7.1 Base: 5.7 Δ: +1.4 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Microsoft Corp. (MSFT)	10	24	1	Microsoft Corp.	10	49
2	Zoom Communications, . (ZM)	10	15	2	Amazon.com, Inc.	8	47
3	Logitech Internationa. (LOGI)	9	2	3	Alphabet, Inc.	9	20
4	8x8, Inc. (EGHT)	8	1	4	Meta Platforms, Inc.	4	19
5	RingCentral, Inc. (RNG)	9	2	5	International Business Mac.	6	12
6	Cisco Systems, Inc. (CSCO)	8	2	6	NVIDIA Corp.	5	11
7	Lantronix, Inc. (LTRX)	5	0	7	Gartner, Inc.	4	13
8	VirnetX Holding Corp. (VHC)	3	0	8	Salesforce, Inc.	7	12
9	TaoWeave, Inc. (TWAV)	1	0	9	Tesla, Inc.	1	10
10	eXp World Holdings, I. (EXPI)	8	0	10	Palantir Technologies, Inc.	3	7

Women in Leadership — New: 6.6 Base: 5.3 Δ: +1.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	The Estée Lauder Comp. (EL)	6	4	1	The Goldman Sachs Group, I.	4	13
2	Victoria's Secret & C. (VSCO)	8	1	2	Morgan Stanley	5	5
3	Bumble, Inc. (BMBL)	8	3	3	Citigroup, Inc.	10	4
4	Siebert Financial Cor. (SIEB)	7	0	4	JPMorgan Chase & Co.	7	4
5	HBS5L0-R	6	0	5	Bank of America Corp.	4	3
6	Merck & Co., Inc. (MRK)	4	3	6	UBS Group AG	3	2
7	Procter & Gamble Co. (PG)	3	6	7	Alphabet, Inc.	6	3
8	Korn Ferry (KFY)	8	0	8	Exxon Mobil Corp.	2	2
9	Professional Diversit. (IPDN)	10	0	9	NVIDIA Corp.	6	2
10	TherapeuticsMD, Inc. (TXMD)	6	0	10	The PNC Financial Services.	6	2

Social Inclusion & Justice — New: 7.4 Base: 4.4 Δ: +3.0 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	KWD40N-R	9	0	1	KKR & Co., Inc.	4	1
2	Geodrill Ltd. (GEO)	8	0	2	Blackstone, Inc.	5	2
3	CoreCivic, Inc. (CXW)	1	0	3	The Goldman Sachs Group, I.	4	2
4	AudioEye, Inc. (AEYE)	10	0	4	JPMorgan Chase & Co.	5	2
5	MAXIMUS, Inc. (MMS)	9	0	5	Netflix, Inc.	5	2
6	The GEO Group, Inc. (GEO)	1	0	6	Harley-Davidson, Inc.	1	1
7	Professional Diversit. (IPDN)	10	0	7	Molson Coors Beverage Co.	1	1
8	Molina Healthcare, In. (MOH)	9	0	8	Morningstar, Inc.	7	1
9	NMG4ST-R	10	0	9	CF Industries Holdings, In.	4	1
10	Tyler Technologies, I. (TYL)	7	0	10	Hologic, Inc.	8	1

Entrepreneurial Capital — New: 8.7 Base: 6.5 Δ: +2.2 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Capital Southwest Corp (CSWC)	9	0	1	Microsoft Corp.	8	17
2	Greenpro Capital Corp. (GRNQ)	8	0	2	Meta Platforms, Inc.	8	11
3	Netcapital, Inc. (NCPL)	10	0	3	The Goldman Sachs Group, I.	9	8
4	Gladstone Investment . (GAIN)	8	0	4	Strategy, Inc.	2	7
5	Trinity Capital, Inc. (TRIN)	10	0	5	NVIDIA Corp.	6	9
6	Main Street Capital C. (MAIN)	9	0	6	JPMorgan Chase & Co.	6	8
7	PhenixFIN Corp. (PFX)	7	0	7	Apollo Global Management, .	9	6
8	Great Elm Capital Cor. (GECC)	7	0	8	PayPal Holdings, Inc.	8	6
9	MDB Capital Holdings . (MDBH)	9	0	9	BlackRock, Inc.	5	5
10	Ares Capital Corp. (ARCC)	10	0	10	Palantir Technologies, Inc.	4	5

Urbanization & Housing — New: 9.7 Base: 8.8 Δ: +0.9 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Equity Residential (EQR)	10	5	1	Lennar Corp.	10	15
2	AvalonBay Communities. (AVB)	10	4	2	News Corp.	7	10
3	Toll Brothers, Inc. (TOL)	9	4	3	Equity Residential	10	7
4	Essex Property Trust,. (ESS)	10	3	4	CoStar Group, Inc.	7	7
5	Invitation Homes, Inc. (INVH)	10	7	5	D.R. Horton, Inc.	10	12
6	UDR, Inc. (UDR)	9	1	6	Invitation Homes, Inc.	10	8
7	D.R. Horton, Inc. (DHI)	10	10	7	CBRE Group, Inc.	8	8
8	LGI Homes, Inc. (LGIH)	9	3	8	AvalonBay Communities, Inc.	10	5
9	PulteGroup, Inc. (PHM)	10	9	9	Camden Property Trust	10	4
10	Mid-America Apartment. (MAA)	10	1	10	Ventas, Inc.	6	5

Travel & Leisure — New: 9.0 Base: 9.9 Δ: -0.9 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Expedia Group, Inc. (EXPE)	10	26	1	Royal Caribbean Group	10	91
2	Carnival Corp. (CCL)	10	18	2	Delta Air Lines, Inc.	10	89
3	Carnival Plc (CUK)	10	18	3	United Airlines Holdings, .	9	63
4	Hyatt Hotels Corp. (H)	1	13	4	Carnival Corp.	10	54
5	TripAdvisor, Inc. (TRIP)	9	10	5	Carnival Plc	10	53
6	Royal Caribbean Group (RCL)	10	21	6	Norwegian Cruise Line Hold.	10	58
7	Norwegian Cruise Line. (NCLH)	10	11	7	American Airlines Group, I.	10	46
8	Travel + Leisure Co. (TNL)	10	3	8	Marriott International, In.	10	51
9	Marriott Internationa. (MAR)	10	24	9	Booking Holdings, Inc.	10	33
10	Booking Holdings, Inc. (BKNG)	10	21	10	Airbnb, Inc.	10	43

Defense Technology — New: 9.1 Base: 8.5 Δ: +0.6 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	AeroVironment, Inc. (AVAV)	9	9	1	Lockheed Martin Corp.	10	348
2	Northrop Grumman Corp. (NOC)	10	23	2	Northrop Grumman Corp.	10	184
3	General Dynamics Corp. (GD)	10	14	3	RTX Corp.	8	149
4	Kratos Defense & Secu. (KTOS)	9	6	4	General Dynamics Corp.	10	131
5	Lockheed Martin Corp. (LMT)	10	25	5	L3Harris Technologies, Inc.	10	115
6	L3Harris Technologies. (LHX)	10	13	6	The Boeing Co.	9	133
7	Leonardo DRS, Inc. (DRS)	9	2	7	Palantir Technologies, Inc.	9	49
8	RTX Corp. (RTX)	8	12	8	AeroVironment, Inc.	9	45
9	Mercury Systems, Inc. (MRCY)	9	1	9	Tesla, Inc.	1	33
10	Skywater Technology, . (SKYT)	7	1	10	Kratos Defense & Security .	9	20

Retail & E-commerce — New: 9.3 Base: 9.4 Δ: -0.1 **Base wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	The TJX Cos., Inc. (TJX)	10	38	1	Walmart, Inc.	10	288
2	Walmart, Inc. (WMT)	10	280	2	Amazon.com, Inc.	10	252
3	Target Corp. (TGT)	9	120	3	Costco Wholesale Corp.	9	89
4	Amazon.com, Inc. (AMZN)	10	317	4	Target Corp.	9	82
5	Macy's, Inc. (M)	10	79	5	Shopify, Inc.	10	54
6	Gap, Inc. (GAP)	9	23	6	Etsy, Inc.	10	41
7	Kohl's Corp. (KSS)	9	40	7	Macy's, Inc.	10	63
8	Abercrombie & Fitch C. (ANF)	8	22	8	The TJX Cos., Inc.	10	42
9	American Eagle Outfit. (AEO)	8	14	9	The Home Depot, Inc.	10	36
10	Best Buy Co., Inc. (BBY)	10	54	10	Meta Platforms, Inc.	6	42

Consumer Staples — New: 9.6 Base: 7.3 Δ: +2.3 **New wins**

	New Scoring				Baseline (theme_v2)		
#	Company	Rel	Art	#	Company	Rel	Art
1	Conagra Brands, Inc. (CAG)	9	66	1	Walmart, Inc.	10	153
2	The J. M. Smucker Co. (SJM)	9	46	2	Amazon.com, Inc.	5	76
3	General Mills, Inc. (GIS)	10	73	3	Costco Wholesale Corp.	9	71
4	The Campbell's Co. (CPB)	9	61	4	Target Corp.	7	52
5	McCormick & Co., Inc. (MKC)	10	10	5	PepsiCo, Inc.	10	46
6	The Clorox Co. (CLX)	10	27	6	The Home Depot, Inc.	1	40
7	The Kroger Co. (KR)	10	24	7	Procter & Gamble Co.	10	45
8	The Kraft Heinz Co. (KHC)	10	60	8	The Kraft Heinz Co.	10	32
9	Hormel Foods Corp. (HRL)	9	12	9	General Mills, Inc.	10	32
10	Kimberly-Clark Corp. (KMB)	10	26	10	The TJX Cos., Inc.	1	27

Consumer Discretionary — New: 8.1 Base: 7.0 Δ: +1.1 **New wins**

	New Scoring				Baseline (theme_v2)		
#	**Company**	**Rel**	**Art**	**#**	**Company**	**Rel**	**Art**
1	Urban Outfitters, Inc. (URBN)	8	1	1	Walmart, Inc.	4	74
2	Dick's Sporting Goods. (DKS)	8	5	2	Amazon.com, Inc.	10	50
3	Abercrombie & Fitch C. (ANF)	9	3	3	Costco Wholesale Corp.	6	45
4	The TJX Cos., Inc. (TJX)	9	4	4	PepsiCo, Inc.	4	38
5	Ralph Lauren Corp. (RL)	9	4	5	NIKE, Inc.	10	24
6	Macy's, Inc. (M)	9	8	6	McDonald's Corp.	10	26
7	Ross Stores, Inc. (ROST)	9	2	7	Mondelez International, In.	5	28
8	Vera Bradley, Inc. (VRA)	7	1	8	The TJX Cos., Inc.	9	18
9	Coty, Inc. (COTY)	5	2	9	Starbucks Corp.	9	22
10	Steven Madden Ltd. (SHOO)	8	1	10	Conagra Brands, Inc.	3	24

Utilities & Renewables — New: 8.2 Base: 6.4 Δ: +1.8 **New wins**

	New Scoring				Baseline (theme_v2)		
#	**Company**	**Rel**	**Art**	**#**	**Company**	**Rel**	**Art**
1	NextEra Energy, Inc. (NEE)	10	65	1	Constellation Energy Corp.	9	140
2	Consolidated Edison, . (ED)	9	20	2	Microsoft Corp.	3	106
3	Dominion Energy, Inc. (D)	8	30	3	Vistra Corp.	8	81
4	Constellation Energy . (CEG)	9	28	4	Amazon.com, Inc.	4	92
5	Duke Energy Corp. (DUK)	9	37	5	GE Vernova, Inc.	9	55
6	Entergy Corp. (ETR)	9	13	6	Alphabet, Inc.	4	57
7	TXNM Energy, Inc. (TXNM)	2	3	7	Oklo, Inc.	9	45
8	PG&E Corp. (PCG)	9	34	8	NextEra Energy, Inc.	10	61
9	The AES Corp. (AES)	9	6	9	Meta Platforms, Inc.	1	44
10	Vistra Corp. (VST)	8	25	10	Tesla, Inc.	7	63

C Cross-Model Evaluation: Full Per-Theme Results

Table 22: Cross-model score agreement: Claude Sonnet 4.6 vs. GPT-5, all 81 themes. ρ: Spearman rank correlation. MAD: mean absolute difference. $\bar{s}$: mean relevance score across all candidates for that theme.

Theme	n	ρ	MAD	GPT	Cl.	**Theme**	n	ρ	MAD	GPT	Cl.
Semiconductors	436	.932	0.55	4.3	4.2	Longevity Science	581	.789	1.47	4.6	3.6
Aerospace & Defense	437	.926	0.69	4.5	4.2	Digital Payments	447	.784	1.12	4.7	4.2
Consumer Staples	482	.915	0.76	4.2	4.2	Neurotechnology	685	.784	0.77	2.7	2.4
Global Commodities	638	.914	0.85	5.1	4.7	3D Printing	427	.770	0.66	2.5	2.4
Gold & Prec. Metals	310	.909	0.54	4.3	4.3	Wind Energy	537	.770	0.85	3.0	2.7
Consumer Discr.	810	.909	0.75	5.2	5.3	Water Sustainability	339	.773	0.90	3.9	3.5
Immunotherapy	566	.897	0.84	4.4	4.3	Green Building	436	.768	1.13	4.1	3.6
AI Chips & Hardware	690	.897	0.62	3.0	2.7	Mental Health	341	.768	1.15	4.1	3.7
Cloud Computing	610	.896	0.79	4.8	4.8	Precision Agri.	408	.762	0.84	2.9	2.8
Constr. & Housing	427	.896	0.92	5.7	5.3	Sports Betting	332	.753	0.73	2.8	2.7
5G / Next-Gen	502	.891	0.83	4.6	4.4	Sustainable Food	321	.745	1.09	3.9	3.9
Biotechnology	788	.891	0.64	5.8	6.1	Social Media	490	.746	0.81	2.9	2.6
Travel & Leisure	279	.890	0.80	5.4	5.3	Pets Economy	341	.746	0.63	2.6	2.4
Autonomous Vehicles	440	.880	0.73	3.6	3.5	mRNA Technology	486	.746	0.60	2.4	2.1
Cybersecurity	378	.874	0.80	4.1	4.1	Food Technology	319	.743	1.19	4.5	4.1
Util. & Renewables	482	.873	0.95	5.2	4.8	Carbon Capture	518	.742	0.94	3.2	2.7
Natural Gas	399	.872	0.86	5.3	5.4	InsurTech	518	.735	1.14	3.4	2.7
Energy Infra.	568	.871	0.99	5.7	5.3	Impact Investing	787	.722	1.36	4.4	3.4
Gene Editing	479	.870	0.68	2.8	2.9	Social Incl. & Justice	550	.705	1.54	4.7	3.6
Digital Advertising	376	.868	0.86	4.6	4.4	Entrepreneurial Cap.	609	.702	1.53	4.6	4.0
Media & Streaming	343	.868	0.91	4.4	4.2	Blockchain Technology	584	.689	0.97	3.2	2.9
Digital Health	522	.866	0.82	3.9	3.9	Hydrogen Economy	570	.655	1.09	3.2	2.6
Def. Technology	346	.863	0.86	5.1	5.0	Women in Ldshp.	483	.496	1.82	4.2	2.8
Battery Technology	459	.863	0.86	4.0	3.6	USD Str. / FX Hedge	630	.471	1.56	3.9	3.2
Utilities	728	.861	0.49	2.9	2.8						
Oil & Gas	564	.855	0.89	5.5	5.6						
Data Centers	568	.849	0.93	4.0	3.7						
Real Estate	505	.851	0.99	5.6	5.7						
Retail & E-commerce	548	.849	0.96	5.9	5.6						
Urbanization	538	.846	1.08	5.2	4.8						
US Infrastructure	598	.846	1.30	5.8	4.9						
Space Exploration	375	.846	0.80	3.1	3.0						
Luxury Goods	336	.845	0.82	3.6	3.5						
Clean Energy	509	.842	0.95	5.1	4.8						
Aging Population	573	.838	1.03	5.3	4.8						
Cryptocurrency	409	.839	0.67	3.2	3.0						
Fintech	580	.835	0.95	4.9	4.6						
IoT	638	.835	0.97	4.8	4.4						
Water Infra.	391	.834	0.80	3.5	3.4						
Robotics & Auto.	729	.831	0.97	4.0	3.7						
ESports & Gaming	316	.828	0.91	3.5	3.4						
Adv. Manufacturing	501	.826	0.96	5.5	5.2						
Circular Economy	578	.827	0.97	4.2	3.9						
AI	940	.825	1.04	4.8	4.2						
Transp. & Logistics	491	.824	1.15	4.7	4.2						
Inflation Prot.	493	.820	1.17	5.3	4.3						
Pharma. R&D	601	.818	0.77	6.5	6.8						
Healthcare Innov.	672	.875	0.76	5.7	5.8						
Online Education	260	.811	0.87	3.5	3.4						
ML Applications	1318	.805	1.11	4.7	4.2						
Nutrition & Suppl.	346	.797	1.00	3.5	3.1						
Climate Mitigation	497	.796	1.07	4.8	4.5						
Smart Cities	541	.791	1.54	5.2	4.0						
Remote Work	495	.829	0.88	3.7	3.7						
Energy Efficiency	586	.815	0.92	4.4	4.6						
ESG Leaders	966	.790	1.08	5.0	4.7						

Overall: $\rho = 0.853$, $r = 0.859$, **MAD = 0.96** **GPT-5** $\bar{s} = 4.35$, **Claude** $\bar{s} = 4.04$

Table 23: Per-theme basket quality under GPT-5 vs. Claude Sonnet 4.6 (final-retrain, 10 companies/theme). Δ = Claude − GPT-5.

Theme	GPT	Cl.	Δ	Theme	GPT	Cl.	Δ
3D Printing	7.7	7.9	+0.2	Longevity Science	7.0	7.6	+0.6
5G / Next-Gen	8.7	7.4	−1.3	Luxury Goods	8.2	8.1	−0.1
AI Chips & Hardware	8.8	8.9	+0.1	ML Applications	8.4	8.1	−0.3
Adv. Manufacturing	8.1	7.7	−0.4	Media & Streaming	9.0	8.6	−0.4
Aerospace & Defense	9.6	9.6	0.0	Mental Health	8.1	7.6	−0.5
Aging Population	8.8	8.7	−0.1	Natural Gas	9.0	9.5	+0.5
AI	9.1	9.3	+0.2	Neurotechnology	6.9	7.5	+0.6
Autonomous Vehicles	8.0	8.1	+0.1	Nutrition & Suppl.	8.5	8.6	+0.1
Battery Technology	9.3	9.0	−0.3	Oil & Gas	8.4	9.6	+1.2
Biotechnology	8.3	9.0	+0.7	Online Education	8.5	9.2	+0.7
Blockchain	7.5	7.7	+0.2	Pets Economy	9.4	9.2	−0.2
Carbon Capture	8.6	7.1	−1.5	Pharma. R&D	8.5	8.8	+0.3
Circular Economy	8.8	8.5	−0.3	Precision Agri.	7.4	7.3	−0.1
Clean Energy	9.4	9.2	−0.2	Real Estate	8.4	8.5	+0.1
Climate Mitigation	6.7	7.0	+0.3	Remote Work	7.1	8.1	+1.0
Cloud Computing	9.0	9.0	0.0	Retail & E-comm.	9.3	9.3	0.0
Constr. & Housing	9.4	9.8	+0.4	Robotics & Auto.	9.4	8.6	−0.8
Consumer Discr.	8.1	9.0	+0.9	Semiconductors	9.3	9.3	0.0
Consumer Staples	9.6	9.6	0.0	Smart Cities	8.1	7.8	−0.3
Cryptocurrency	8.5	9.3	+0.8	Social Incl.	7.4	5.9	−1.5
Cybersecurity	8.7	9.3	+0.6	Social Media	7.9	9.1	+1.2
Data Centers	8.1	8.3	+0.2	Solar Energy	9.6	9.3	−0.3
Defense Technology	9.1	9.5	+0.4	Space Exploration	9.1	8.7	−0.4
Digital Advertising	8.6	8.9	+0.3	Sports Betting	8.9	8.4	−0.5
Digital Health	8.6	8.5	−0.1	Sustainable Food	6.3	6.7	+0.4
Digital Payments	9.3	9.3	0.0	Transp. & Logistics	8.6	8.8	+0.2
ESG Leaders	8.0	8.1	+0.1	Travel & Leisure	9.0	9.0	0.0
ESports & Gaming	7.9	7.6	−0.3	US Infrastructure	8.9	9.3	+0.4
Energy Efficiency	8.3	7.9	−0.4	USD Str. / FX Hedge	8.9	7.9	−1.0
Energy Infrastructure	9.4	8.9	−0.5	Urbanization	9.7	9.6	−0.1
Entrepreneurial Cap.	8.7	7.7	−1.0	Utilities	9.4	9.7	+0.3
Fintech	9.0	9.0	0.0	Util. & Renewables	8.2	8.9	+0.7
Food Technology	7.0	7.0	0.0	Water Infrastructure	9.1	9.1	0.0
Gene Editing	9.4	9.3	−0.1	Water Sustainability	9.2	9.1	−0.1
Global Commodities	9.3	8.8	−0.5	Wind Energy	7.7	7.1	−0.6
Gold & Prec. Metals	7.5	7.6	+0.1	Women in Ldshp.	6.6	5.5	−1.1
Green Building	7.5	7.5	0.0	mRNA Technology	7.0	6.2	−0.8
Healthcare Innov.	8.1	7.6	−0.5				
Hydrogen Economy	8.1	7.0	−1.1				
Immunotherapy	8.4	8.5	+0.1				
Impact Investing	6.8	7.4	+0.6				
Inflation Protection	7.7	7.8	+0.1				
InsurTech	7.5	6.0	−1.5				
IoT	9.4	8.8	−0.6				
All 81 themes: GPT-5 = 8.43				**Claude = 8.37** (Δ = −0.06)			

D Complete Scoring Algorithm

Algorithm 2 Complete Basket Generation (Weighted Mode)

Require: Theme query q, parameter vector $\boldsymbol{\theta} = (\tau_s, p, \beta_b, \tau_c, w_n, w_d, w_a, w_t, w_e, w_{\text{sa}}, \gamma)$
Require: Cached retrieval results: $\mathcal{A}$ (news), $\mathcal{D}$ (descriptions), $\mathcal{B}$ (areas), $\mathcal{H}$ (themes), $\mathcal{E}$ (earnings), $\mathcal{S}$ (SA)
Ensure: Ranked basket of top-K companies

// Phase 1: News signal per company
for all article $(s_i, C_i, t_i) \in \mathcal{A}$ **do**
  if $s_i < \tau_s$ **then continue**
  end if
  $\text{age}_i \leftarrow t_{\text{query}} - t_i$ (in days)
  if $\text{age}_i < 0$ or $\text{age}_i > 2922$ **then continue**
  end if
  $\delta_i \leftarrow \exp(-\ln 2 \cdot \text{age}_i / 1461)$
  $\tilde{s}_i \leftarrow s_i \cdot \delta_i$
  for all company $c \in C_i$ with `about=true` and valid ISIN **do**
    Append $\tilde{s}_i$ to scores$[c]$; increment count$[c]$; B_c += δ_i
  end for
end for
for all company c with $n_c = |\text{scores}[c]| > 0$ **do**
  $I_c \leftarrow (\text{mean}(\tilde{s}^{\,p}))^{1/p}$ ▷ power mean
  $N_c \leftarrow I_c \cdot (1 + \beta_b \cdot \ln(1 + B_c))$ ▷ breadth amplification
end for

// Phase 2: Card signals per company
for all collection $k \in \{\text{desc}, \text{area}, \text{theme}, \text{earn}, \text{sa}\}$ **do**
  for all $(s_j, \texttt{tilt\_id}_j) \in \mathcal{D}_k$ **do**
    if $s_j \geq \tau_c$ **then**
      $V^{(k)}_{\texttt{tilt\_id}_j} \leftarrow \max(V^{(k)}_{\texttt{tilt\_id}_j},\ s_j - \tau_c)$
    end if
  end for
end for

// Phase 3: Aggregation
for all candidate company c **do**
  $W_c \leftarrow w_d \cdot V_c^{(\text{desc})} + w_a \cdot V_c^{(\text{area})} + w_t \cdot V_c^{(\text{theme})} + w_e \cdot V_c^{(\text{earn})} + w_{\text{sa}} \cdot V_c^{(\text{sa})}$
  $\mathcal{V}_c \leftarrow$ sorted desc. positive values of $\{V_c^{(\text{desc})}, V_c^{(\text{area})}, V_c^{(\text{theme})}, V_c^{(\text{earn})}, V_c^{(\text{sa})}\}$
  if $|\mathcal{V}_c| \geq 2$ **then**
    W_c += $\mathcal{V}_c[0] \cdot \mathcal{V}_c[1] \cdot \gamma$ ▷ card synergy
  end if
  $S_c \leftarrow w_n \cdot N_c + W_c + \gamma \cdot N_c \cdot W_c$ ▷ cross-source interaction
end for
return Top-K companies by S_c

E Mathematical Notation Summary

Table 24: Notation used throughout this document.

Symbol	Meaning
q	User-supplied theme query string
$\mathbf{e}(x)$	Voyage 4 Large embedding of text x
s_i	Cosine similarity of article/card i to the query
δ_i	Exponential time-decay factor for article i
$\tilde{s}_i$	Effective score: $s_i \cdot \delta_i$
τ_s	Minimum cosine similarity threshold (news)
τ_c	Minimum cosine similarity threshold (cards)
p	Power mean exponent
β_b	Breadth amplification coefficient
$T_{1/2}$	Half-life in days (1,461 $\approx$ 4 years)
$T_{\max}$	Maximum lookback in days (2,922 $\approx$ 8 years)
n_c	Number of filtered articles mentioning company c
B_c	Decay-weighted article count for company c
I_c	Power mean intensity for company c
N_c	Final news score for company c
$V_c^{(k)}$	Card signal for company c from collection k
W_c	Weighted card sum (with synergy) for company c
S_c	Final composite score for company c
$w_n, w_d, w_a, w_t, w_e, w_{\text{sa}}$	Source weights (news, desc., area, theme, earnings, SA)
γ	Cross-source interaction strength
$R(c, \tau)$	LLM relevance score for company c on theme τ
$f(\boldsymbol{\theta}; \mathcal{T})$	Objective function (average basket relevance)
K	Basket size (default 10)
k	Number of CV folds (default 5)
B	CMA-ES optimization budget (rounds per fold)

F LLM Prompts

This appendix reproduces, verbatim, the LLM prompts used in company-card extraction (Section 2.1) and the evaluation oracle (Section 5.1).

F.1 Company Card Extraction (System Prompt)

Listing 1: System prompt for SEC 10-K company card extraction.

```
You are a financial analyst. Given sections from a company's SEC 10-K
filing, produce a structured company profile.

Your output must be ONLY valid JSON (no markdown, no code fences) with
these fields:
{
  "description": "<A neutral 2-4 sentence description of what the company
   does, its scale, and market position. No promotional language.>",
```

```
  "business_areas": ["<area1>", "<area2>", ...],
  "themes": ["<theme1>", "<theme2>", ...]
}

Rules:
- description: Be factual and concise. Mention the industry, key
  products/services, geographic reach, and rough scale if evident.
- business_areas: List the company's main lines of business or revenue
  segments (e.g. "Cloud Computing", "Semiconductor Manufacturing",
  "Retail Pharmacy"). Typically 2-8 items.
- themes: List investment themes the company has POSITIVE exposure to
  -- meaning the company BENEFITS from or is a TAILWIND beneficiary of
  that theme.

  IMPORTANT -- themes must represent POSITIVE exposure only:
  - GOOD theme: "artificial intelligence" (company builds AI products)
  - GOOD theme: "electric vehicles" (company makes EV batteries)
  - GOOD theme: "aging population" (company sells senior care)
  - BAD theme: "regulatory risk" -- this is a RISK, not a positive
    exposure. Do NOT include it.
  - BAD theme: "cybersecurity risk" -- same. Only include
    "cybersecurity" if the company SELLS cybersecurity products.

  Include macro themes (e.g. "US manufacturing reshoring"), technology
  themes (e.g. "quantum computing", "CRISPR"), and sector themes
  (e.g. "aging population", "renewable energy"). Aim for 5-30 themes
  depending on the company's breadth. Use lowercase.
```

F.2 Company Card Extraction (User Prompt Template)

Listing 2: User prompt template for card extraction.

```
Company: {company_name} ({ticker})
Filing Date: {filing_date}

=== Business Overview (Section 1) ===
{business_section[:6000]}

=== Management Discussion & Analysis (Section 7) ===
{mda_section[:6000]}

Analyze the above SEC 10-K sections and produce the company profile JSON.
```

F.3 Evaluation Oracle (System Prompt)

Listing 3: System prompt for per-company thematic relevance scoring.

```
You are an expert portfolio manager evaluating individual securities for
thematic relevance. Given a theme and a list of securities with their
SEC 10-K filing context, rate EACH security's relevance.

Return JSON:
{
  "evaluations": [
    {"ticker": "MSFT", "relevance_score": 9, "reasoning": "..."}
```

```
  ]
}

Scoring guidelines:
- 9-10: Core/canonical pick. The company is a leader or pure-play
        in this theme.
- 7-8:  Clearly relevant. Significant exposure to the theme.
- 5-6:  Tangentially related. Some connection but not a primary play.
- 3-4:  Weak connection. Stretch to include in this theme.
- 1-2:  Irrelevant. No meaningful connection to the theme.

Use the SEC filing context (description, business areas, themes) to
make informed judgments. Provide a concise 1-sentence reasoning.
```

G API Reference

The scoring system is exposed via a REST API. A single-theme query takes the form:

Listing 4: Single-theme API call.

```
curl -s -X POST "https://sig-api.tilt.io/api/theme-new" \
    -H "Content-Type: application/json" \
    -H "X-API-Key: <API_KEY>" \
    -d '{
      "themes": [
        {"query": "artificial intelligence", "weight": 1.0}
      ],
      "as_of_date": "2026-03-18",
      "limit": 20,
      "normalization": "sum_to_one"
    }'
```

Multi-theme blending (Section 4.6) is specified by passing multiple `(query, weight)` entries. The response includes per-company scores, normalized weights, and a `theme_scores` breakdown showing which theme drives each company's inclusion.

Disclaimer. This document describes a quantitative methodology for constructing thematic investment baskets. It does not constitute investment advice, a solicitation, or a recommendation to buy or sell any securities. Past performance and backtested results do not guarantee future performance. The use of large language models introduces inherent uncertainty and potential biases in both data extraction and evaluation. All parameter values and performance figures are as of the date indicated and may change with updated data or model versions.

References

[1] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[2] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed. Cambridge University Press, 1952.

[3] P. S. Bullen, *Handbook of Means and Their Inequalities*. Kluwer Academic Publishers, 2003.

[4] N. Hansen and A. Ostermeier, “Completely derandomized self-adaptation in evolution strategies,” *Evolutionary Computation*, vol. 9, no. 2, pp. 159–195, 2001.

[5] N. Hansen, “The CMA evolution strategy: A tutorial,” arXiv preprint arXiv:1604.00772, 2016 (originally circulated 2006).

[6] J. Rapin and O. Teytaud, “Nevergrad – A gradient-free optimization platform,” `https://github.com/facebookresearch/nevergrad`, 2018.

[7] G. V. Cormack, C. L. A. Clarke, and S. Büttcher, “Reciprocal rank fusion outperforms Condorcet and individual rank learning methods,” in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 758–759, 2009.

[8] Y. A. Malkov and D. A. Yashunin, “Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 824–836, 2020.

[9] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, “Dense passage retrieval for open-domain question answering,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781, 2020.

[10] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3982–3992, 2019.

[11] R. Nogueira and K. Cho, “Passage re-ranking with BERT,” arXiv preprint arXiv:1901.04085, 2019.

[12] J. Lin, R. Nogueira, and A. Yates, *Pretrained Transformers for Text Ranking: BERT and Beyond*. Morgan & Claypool Publishers, 2021.

[13] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, and I. Stoica, “Judging LLM-as-a-judge with MT-Bench and Chatbot Arena,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.

[14] S. Rendle, “Factorization machines,” in *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM)*, pp. 995–1000, 2010.

[15] M. Stone, “Cross-validatory choice and assessment of statistical predictions,” *Journal of the Royal Statistical Society: Series B*, vol. 36, no. 2, pp. 111–133, 1974.

[16] S. Geisser, “The predictive sample reuse method with applications,” *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 320–328, 1975.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.

[18] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.