RAFE: Reasoning-Augmented Factor Extraction from SEC Filings and Horizon-Targeted Sentiment Alignment via Reinforcement Learning

Anonymous ACL submission

Abstract

SEC filings shape company stock prices, yet standard sentiment tools assign a single polarity score to the entire document — missing the topic-by-topic signals. We present **RAFE**, a pipeline that extracts and scores 14 investment topics from SEC filings, then aligns those sentiment scores with future stock returns.

RAFE first builds a dataset of 23,010 evidence quotes extracted using 68 subfactor questions from 2,230 SEC 10-K/10-Q filings for 62 U.S. technology companies (2015–2025), each labeled by GPT-5-mini and distilled into Qwen3-8B. Two reinforcement-learning stages then align it with subsequent stock returns: DPO uses return-based preference pairs, and GRPO adds a confidence-calibration reward. Together with self-consistency, they reshape the sentiment distribution. Topic pre-selection and aggregation per company further improves portfolio performance.

On held-out 2024–2025 filings, negative sentiment in filings consistently predicts stock underperformance across most learned models. RL alignment further sharpens return-monotone patterns at the factor level. At the filing level, topic pre-selection combined with two-step aggregation yields an average long-short Sharpe ratio of $+0.86$ under daily rebalancing and $+0.62$ under monthly rebalancing across all eight models. Capital intensity is the most robust factor; security/privacy and pricing reverse direction.

1 Introduction

Every quarter, thousands of public companies file annual and quarterly reports with the SEC. Professional investors read these carefully: one section might reveal a looming regulatory threat while another signals a strong acquisition pipeline. Yet existing AI tools collapse all of this into a single positive-or-negative score for the whole document — discarding the per-topic detail that actually drives investment decisions.

Prior NLP sentiment tools score documents as a whole (Malo et al., 2014; Araci, 2019; Wu et al., 2023). Alignment methods such as RLHF (Ouyang et al., 2022), DPO (Rafailov et al., 2023), and GRPO (Shao et al., 2024) use human feedback or rules to shape model behaviour, not financial market outcomes — and stock returns are noisy and delayed by weeks, making them hard to use as a training signal. Our central question is: *Could we train a language model to score each investment topic separately in a company filing and generate signals that might help predict future stock movement?*

We present **RAFE**[1] (Reasoning-Augmented Factor Extraction and Horizon-Targeted Sentiment Alignment via Reinforcement Learning), a pipeline that answers this end-to-end. RAFE first builds a dataset of 23,010 evidence quotes extracted using 68 subfactor questions from 2,230 SEC 10-K/10-Q filings for 62 U.S. technology companies (2015–2025), each labeled very positive, positive, neutral, negative, or very negative with a confidence score and reasoning chain by GPT-5-mini and distilled into Qwen3-8B, which scores new filings without any frontier model at inference time. Two market-alignment stages then push the model beyond GPT-5-mini: DPO corrects labels that conflict with subsequent stock returns via preference pairs, and GRPO further refines predictions using a dual reward — 40% for return alignment and 60% for confidence calibration. Self-consistency at test-time compute further reduces label noise at no training cost. Figure 1 shows the full pipeline. Tested on held-out 2024–2025 filings from large US technology companies, the Loughran–McDonald lexicon shows near-zero predictive power, while our RL-aligned models show that negative sentiment in fil-

[1] Code, 5,000 human-revised samples, preprocessed data, and per-table reproduction scripts are available at `https://anonymous.4open.science/r/RAFE-7884` (anonymous mirror for review; public release upon acceptance).

ings consistently predicts stock underperformance, produce cleaner return-monotone patterns at the factor level, and consistently identify informative factors.

Contributions. We make four contributions:

1. **A factor-level, quote-grounded sentiment dataset for SEC filings.** We apply question-answering to build a dataset of 23,010 evidence quotes from 2,230 10-K and 10-Q filings for 62 S&P 500 IT-sector companies (2015–2025), covering 14 investment factors and 68 subfactors, each grounded in a curated analyst question. Each record pairs one direct quote with a GPT-assigned label, confidence score, and reasoning chain; a 5,000-sample subset is further created to refine factor-specific rationales used in the RL alignment stages.

2. **RL alignment and self-consistency sharpen factor-level signals.** DPO and GRPO help align the model with 21-day excess returns by correcting labels that don't match the market and by rewarding better-calibrated confidence. In addition, test-time self-consistency (TTC)—running the model three times and taking the majority label—helps further reduce noise. Overall, this RL alignment leads to cleaner return-monotone patterns at the factor level: filings with negative labels tend to have lower subsequent returns than neutral, and neutral tend to have lower returns than positive-labelled filings.

3. **IC-based factor pre-selection with two-step aggregation is the key to strong filing-level portfolios.** By limiting each company to its top 3 most return-predictive parent factors (selected using Spearman IC on GPT teacher labels), we observe more than a threefold improvement in portfolio performance: $\mathrm{SFT}_{\mathrm{TTC}}$'s monthly-rebalancing Sharpe rises from $+0.74$ to $+2.37$, the GPT teacher from -0.05 to $+1.31$. Pre-selection paired with two-step aggregation (subfactor $\rightarrow$ parent factor $\rightarrow$ filing label) yields an average long-short Sharpe of $+0.86$ under daily rebalancing and $+0.62$ under monthly rebalancing across all eight models.

4. **Factor-level analysis identifies actionable investment signals.** Capital Intensity appears to be the most robust factor: across all learned models, it shows positive long/short Sharpe ratios and a clear monotonic ordering at both the 1-day and 21-day horizons. Competitive Dynamics and M&A/CorpDev are also consistently profitable across models. In particular, GRPO performs especially well on Regulatory/Legal and Capital Intensity, where RL training seems to identify a high-conviction "Negative" cohort that tends to underperform the market. On the other hand, Security/Privacy and Pricing & Margin do not behave as expected for any of the trained models, and Go-to-Market is the most ambiguous—different training approaches often assign opposite signals to the same filing.

The rest of this paper is organised as follows. Section 2 reviews related work. Section 3 describes the RAFE pipeline. Section 4 presents the dataset, models, and results. Section 5 analyses model behaviour and per-factor findings. Section 6 concludes.

2 Related Work

Financial NLP and sentiment analysis. Early work used word lists or bag-of-words classifiers (Loughran and McDonald, 2011; Malo et al., 2014). FinBERT (Araci, 2019) fine-tunes a pre-trained language model on financial text; BloombergGPT (Wu et al., 2023) pre-trains on a massive financial corpus; FinGPT (Yang et al., 2023) and FinLlama (Konstantinidis et al., 2024) adapt open-weight models via instruction tuning. All assign a single polarity to an entire filing or section. Factor-level sentiment — classifying each investment theme independently within a filing — and its portfolio implications remain unexplored.

RL alignment for language models. RLHF (Ouyang et al., 2022) aligns models to human preference via a reward model trained on pairwise comparisons. DPO (Rafailov et al., 2023) removes the separate reward model by training directly on preferred/rejected pairs. GRPO (Shao et al., 2024) uses group-based advantage normalisation over self-generated completions, enabling RL alignment without a value network. KTO (Ethayarajh et al., 2024) similarly relaxes the pairing requirement. All prior methods optimise for *human* preference. Aligning a language model to *market* preference — realised stock returns that are delayed, noisy ($\sigma \approx 10\%$), and factor-specific — is an open problem not addressed by any of these methods.

Knowledge distillation and factor investing. Knowledge distillation (Hinton et al., 2015) compresses a large teacher into a smaller student; Mag-

ister et al. (2022) show that chain-of-thought rationales transfer effectively. Distilling a frontier teacher into a compact student and then re-aligning the student with market returns — moving beyond teacher imitation toward market-grounded preference — has not been explored. Factor investing decomposes returns into systematic theme exposures (Fama and French, 1993; Harvey et al., 2016); Tetlock (2007) and Loughran and McDonald (2011) show that filing language predicts returns at the document level. Operationalising thematic investment factors at the subfactor level and applying IC selection — restricting each ticker to its top-K most return-predictive themes via Spearman $|\rho|$ — is a standard quantitative practice not previously applied to LLM-derived factor sentiment.

SEC filing analysis. Prior NLP work on SEC filings treats the MD&A section as a whole, predicting return volatility (Kogan et al., 2009) or document-level sentiment (Tetlock, 2007; Loughran and McDonald, 2011). Breaking filing sentiment into separate signals for M&A activity, regulatory exposure, or AI infrastructure — at the language-model alignment level — has not been studied.

Test-time compute and confidence calibration. Majority voting over multiple runs (Wang et al., 2023) improves accuracy on reasoning tasks. Vamvourellis and Mehta (2025) find that extended chain-of-thought reasoning can *hurt* financial sentiment accuracy, with single-pass inference outperforming deliberative models. Isotonic regression (Niculescu-Mizil and Caruana, 2005) is a standard post-hoc calibration tool for correcting miscalibrated confidence scores.

3 The RAFE Pipeline

RAFE converts raw SEC filings into portfolio signals through six steps, shown in Figure 1. Steps 1–2 build the labelled dataset: a small AI model extracts the key quote from each filing per investment topic (§3.1), and a larger GPT model labels each quote with a sentiment score and explanation (§3.2). Steps 3–5 train the model: SFT distills the GPT labels into a compact model (§3.3); DPO and GRPO then use realized return-based signals to align model preferences with subsequent return directions through two rounds of preference learning and reinforcement learning (§3.4–3.5). Step 6 converts the model's outputs into portfolio signals by running the model three times and taking the majority label to form a consensus prediction. Additionally, each company is restricted to topics selected from its in-sample predictive history, implementing the IC-selected aggregation variant evaluated in the backtest. (§3.6). All training uses 2015–2022 data; the 2024–2025 test period is fully held out.

3.1 Investment Factors and Evidence Extraction

We define 14 investment factors and 68 subfactors covering themes commonly used to evaluate technology stocks — M&A activity, tech infrastructure, capital intensity, regulatory risk, security and privacy, and nine others (full list in Appendix E). Each subfactor is paired with a specific question to guide evidence retrieval (e.g., What does the filing say about capital expenditure plans and cyclicality?).

GPT-5-nano reads each SEC filing and extracts the direct quote ranked as most relevant for each subfactor. If no relevant quote exists, no record is created. Each record is one targeted quote tied to one specific factor.

3.2 GPT Teacher Labelling

For each quote, GPT-5-mini reads it together with recent prior filings from the same company and produces three outputs: a sentiment label on a five-point scale (very negative to very positive), a confidence score between 0 and 1, and a short explanation. The full prompt is in Appendix F.

Only this step uses a frontier model; downstream stages reuse the teacher's outputs with human annotation as training targets.

3.3 Supervised Fine-Tuning (SFT)

We distill the GPT teacher into a small, cheap open-weight model (Qwen3-8B; Qwen Team 2025) via SFT. The model first writes out its reasoning, then outputs the sentiment label and confidence score. This stage converts the one-time GPT teacher labels into a Qwen3-8B model that can be run repeatedly over the filing corpus without additional GPT calls. Training details are in Appendix G. The SFT checkpoint is the starting point for all subsequent stages.

3.4 Return-Aligned DPO

DPO introduces a return-based preference signal for cases in which the teacher label differs substantially from the subsequent 21-day excess-return direction. For each training example, we convert the stock's 21-day excess return into a direction label (very negative to very positive) using fixed thresholds.

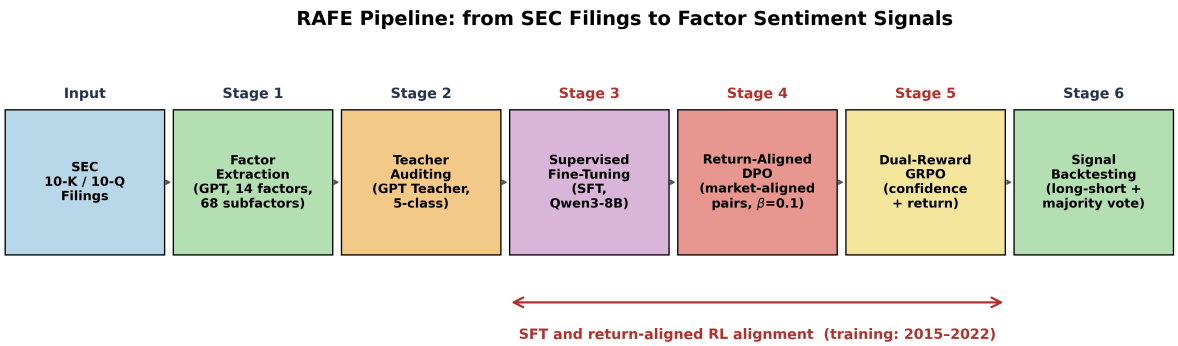

Figure 1: The RAFE pipeline. **Stage 1** (Evidence Extraction): GPT-5-nano extracts the most relevant quote from each SEC filing for each of 14 investment factors. **Stage 2** (GPT Teacher): GPT-5-mini assigns a 5-class sentiment label, confidence score, and reasoning chain to each quote — the only step requiring a frontier model. **Stage 3** (SFT): teacher labels are distilled into a compact Qwen3-8B model via chain-of-thought fine-tuning. **Stage 4** (DPO): return-based preference pairs are used to train the model; ∼5,000 examples were reviewed by human annotators to refine factor-specific rationales. **Stage 5** (GRPO): a dual reward based on return-direction agreement and confidence calibration is applied using the same human-annotated set; the return signal is used only during training. **Stage 6** (Backtesting): factor-level long-short and filing-level majority-vote portfolios are evaluated under daily and monthly rebalancing. The red arrow spans the RL training stages (3–5), all trained on 2015–2022 data.

When the teacher label and the return-direction label agree, the teacher response is used as the preferred response. When they differ by more than one step on the five-point scale (e.g., the teacher label is positive but the return-direction label is very negative), the return-aligned response is used as the preferred response and the teacher response is used as the rejected response. This constructs preference pairs that favor the return-aligned response in cases of large teacher–return disagreement.

Human annotation. Three annotators with equity research backgrounds reviewed a random sample of 5,000 training examples and revised the GPT teacher's outputs — sentiment label, confidence score, and reasoning chain — to reflect factor-specific context, for example distinguishing a company-specific regulatory headwind from a sector-wide one. The revised outputs replace the original GPT outputs in the DPO and GRPO training examples; the SFT stage uses only the original teacher outputs. Each annotator labeled a distinct subset rather than the three jointly labeling the same examples, so we do not report inter-annotator agreement; see Limitations.

Return labels. Preference pairs use the stock's 21-day excess return, mapped to a five-point direction label: returns below -8% are very negative, returns in $[-8\%, -2\%]$ are negative, returns in $[-2\%, 2\%]$ are neutral, returns in $[2\%, 8\%]$ are positive, and returns above 8% are very positive. Because a single filing may contain multiple factor signals, the stock-level return is an imperfect proxy for any individual factor's directional truth. All return labels are drawn from the training period.

3.5 GRPO with Dual Reward

Starting from the DPO checkpoint, we run one additional training stage using reinforcement learning (Shao et al., 2024). The stock return signal is used only during training.

The model produces two candidate responses per example. Each response is scored on two criteria. The first is directional agreement with the return-derived label (40% weight): responses whose predicted sentiment is closer to the realized stock-return direction receive higher reward, while neutral or opposite-direction responses are penalized. The second is confidence calibration (60% weight): the model is encouraged to report higher confidence when its prediction matches a stronger directional return signal, and lower confidence when the return-based supervision is weaker or more ambiguous. This treats stock returns as a noisy proxy rather than a hard ground-truth label, so the reward favors appropriately calibrated confidence rather than uniformly high confidence. Responses that fail to produce a valid output format are penalised. The higher-scoring response is preferred over the lower-scoring response, without requiring a separate value network.

Neutral Collapse. Because the GRPO reward penalises neutral predictions when the return proxy

moves strongly in one direction, the RL checkpoints produce fewer Neutral labels than the teacher model. We report the resulting label distributions in Table 2 and analyse their implications in §5.2.

3.6 Signal Aggregation and Portfolio Construction

Once the model produces subfactor labels, we convert them into portfolio signals as follows.

Reducing noise with self-consistency at inference time. We run each model three times on the same filing and aggregate the resulting subfactor labels by majority vote. When vote counts are tied, preference is given to the more directional label on the five-point scale. The final confidence score is the mean reported confidence among runs supporting the winning label. This defines the test-time-consensus (TTC) variant used to aggregate the up to 68 subfactor labels into a factor-level or filing-level signal at two levels of granularity.

Factor level (Tables 3–4): for each of the 14 parent factors, we take a majority vote across its subfactor labels to obtain a single factor-level signal per filing.

Filing level (Tables 12–13): we adopt an IC-based hierarchical procedure. For each company, we rank the 14 parent factors by the Spearman correlation between GPT teacher labels and 21-day forward returns on 2015–2023 data, and retain the top three. We then aggregate hierarchically: a majority vote within each retained parent factor produces three parent-level labels, which are combined by a further majority vote to yield the filing-level signal. This IC-selected specification is the primary filing-level variant reported in §5.3; Appendix C reports one-step and no-pre-selection variants for comparison.

4 Experiments

4.1 Setup

Data. We use 49 S&P 500 IT-sector companies, drawing 10-K and 10-Q filings from EDGAR (January 2015 – June 2025); 13 tickers were excluded due to incomplete extractions in test period. The data is split chronologically (Table 1): we train on filings through 2022, tune on 2023 filings, and hold out January 2024 – June 2025 for all portfolio evaluation. During the test period, each company's most recent sentiment label is carried forward until a newer filing arrives, giving 18 non-overlapping monthly hold periods. In Table 1, observations refer to ticker–filing–subfactor records: for a given company filing, each of the 68 subfactor questions can produce one evidence quote, sentiment label, confidence score, and parent-factor assignment used for scoring and portfolio construction.

Table 1: Dataset splits (ticker, subfactor, filing triples). Chronological split: train 2015–2022, val 2023, test 2024–Jun 2025.

Split	Period	**Observations**	Role
Train	2015–2022	8,451	SFT; DPO
Validation	2023	1,034	Tuning
Test	Jan 2024–Jun 2025	1,458	Backtest

Models. We evaluate six learned model variants built on Qwen3-8B: SFT, DPO, GRPO, and test-time compute versions of each (SFT_{TTC}, DPO_{TTC}, $GRPO_{TTC}$), where TTC runs the model three times and takes the majority. TTC is an inference-only self-consistency procedure: the model is sampled three times for the same evidence quote, and the majority label is used as the final prediction. This reduces label noise without changing the trained weights. We also include the GPT teacher (GPT-5-mini, which generates the training labels) as the one baseline reference, and the Loughran–McDonald (LM) word-counting dictionary (Loughran and McDonald, 2011) as a simple baseline scored on the same evidence quotes.

Portfolio construction. We merge very_negative into Neg and very_positive into Pos (extreme labels are $\leq$1.4% of predictions), giving three cohorts: Neg, Neu, Pos. We test three ways of turning these labels into portfolios.

Factor-level long-short (Appendix B): within each of the 14 factors, go long Positive-labelled stocks and short Negative-labelled stocks. This measures how well each individual factor predicts returns.

Filing-level majority vote (Appendix C.1): for each company filing, count Pos and Neg labels across all 14 factors and take the majority direction as the filing label. This gives three portfolio cohorts (Neg, Neu, Pos) at the filing level.

IC-selected cohorts (Appendix C.2): before majority voting, restrict each company to its three most return-predictive parent factors, selected using Spearman IC on the GPT teacher train+validation data (2015–2023). These factors are selected

using Spearman information coefficient (IC) between GPT teacher sentiment scores and 21-day forward excess returns on the train-plus-validation period only (2015–2023), with no use of held-out 2024–2025 returns. This is the strongest construction: focusing on each company's most informative factors substantially improves portfolio quality.

All portfolios enter with a one-day delay ($t+1$ entry) and rebalance either daily or at the start of each month. The benchmark is the equal-weight IT-sector portfolio.

4.2 Results

Figure 2 maps how the 68 subfactor labels flow from sentiment scoring into three portfolio evaluation streams. Appendix A records the full label distribution for all eight models. Appendix B measures factor-level predictability: B.1 reports long-short Sharpe ratios for each of the 14 investment factors under daily and monthly rebalancing; B.2 tests whether models produce a clean Neg < Neu < Pos return ordering within each factor. Appendix C aggregates labels to the filing level and forms cohort portfolios — first without topic filtering (C.1), then with IC-based topic preselection using one-step or two-step aggregation (C.2–C.3). Appendix D checks whether filing-level excess returns survive Fama-French five-factor adjustment. Appendices E–G provide the factor taxonomy, prompt templates, and implementation details.

Together, these experiments answer three questions. *First, do individual investment topics contain return-predictive sentiment? Second, can those topic-level labels be aggregated into filing-level portfolios? Third, do the resulting long-short returns survive standard risk-factor adjustment?* We present the results in that order.

5 Analysis

We report findings from the three evaluation streams: factor-level long-short performance (§5.1), return ordering under RL alignment (§5.2), and filing-level cohort portfolios (§5.3).

5.1 Per-Factor Long-Short Portfolio Summary

Tables 3 and 4 (Appendix B) show how well each model's sentiment ranks stocks within each of the 14 investment factors. For each factor we build a long-short portfolio — long the stocks labelled Positive, short those labelled Negative — and report the Sharpe ratio under daily and monthly rebalancing. Learned models beat the word-counting baseline (LM) on 12 of 14 factors; LM leads only on Security/Privacy and Structural Risks. Because we have only 18 monthly periods, these results should be read as directional evidence, not statistically confirmed alphas. Three patterns stand out.

(1) **Competitive Dynamics and M&A/CorpDev are the most reliable factors.** DPO_{TTC} reaches $+2.47$ monthly on Competitive Dynamics; GPT leads M&A/CorpDev at $+1.99$ monthly. LM is negative on both factors, suggesting that the signal comes from contextual interpretation of filing language rather than simple polarity-word counts. GRPO_{TTC} also performs strongly on Go-to-Market ($+1.91$ monthly) and Management Quality ($+1.86$ monthly), giving the best model-specific results for those two factors.

(2) **Go-to-Market is the most model-dependent factor.** The signal flips sign depending on the model: SFT earns $+1.82$ monthly while GPT loses -3.34 monthly over the same factor and test window. This suggests that Go-to-Market language is especially sensitive to modelling choices: claims about sales execution and channel expansion can plausibly indicate either growth momentum or commercial pressure, so different training objectives extract opposite signals from similar evidence.

(3) **Security/Privacy and Pricing & Margin go in the wrong direction for every learned model.** Every learned model loses money on Security/Privacy (monthly Sharpe -1.05 to -2.41), while LM earns $+1.76$. Pricing & Margin is also negative across all learned models. In this specific sample, they point in the wrong direction for LLM-based extraction. For portfolio construction, they should either be excluded, treated as diagnostic rather than directional, or re-examined with factor-specific prompts.

5.2 Factor Ordering under RL Alignment

RL training produces a visible side effect: it pushes uncertain predictions away from Neutral toward Positive, shrinking the Neutral share from 27% (teacher) to 7–13% across RL models (Table 2). This creates an asymmetry at the filing level. The Neg cohort stays small and high-conviction, while the Pos cohort grows large and noisy. This helps explain why the short side is more reliable than the long side. Across learned models, Negative-labelled filings underperform the benchmark by

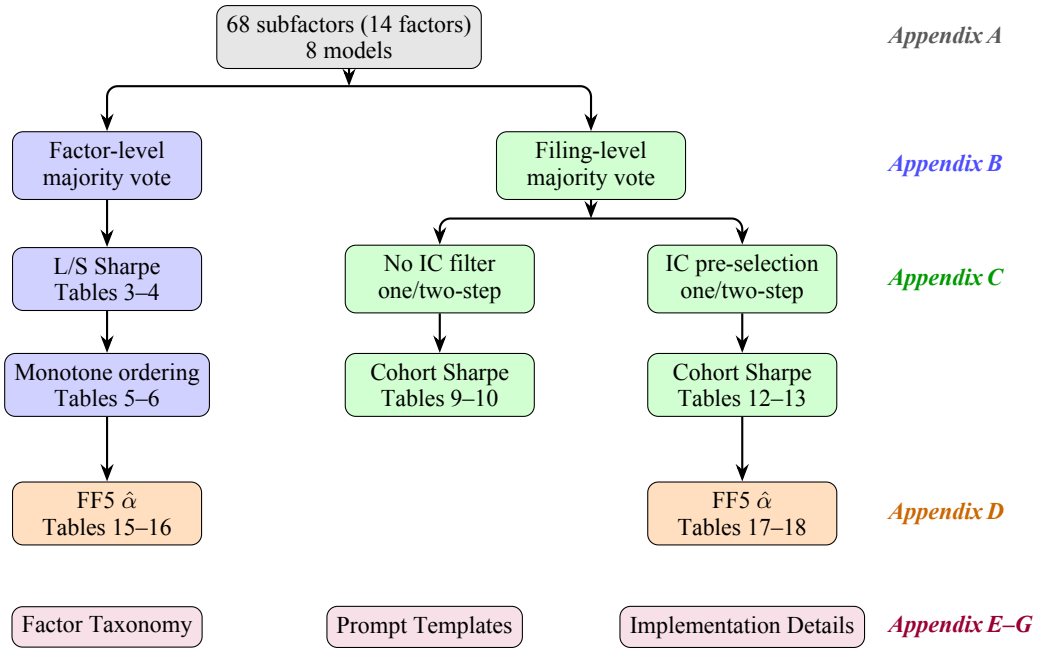

Figure 2: Overview of experiments in Appendices A–G. Blue = Appendix B (B.1: factor-level L/S; B.2: return monotonicity); Green = Appendix C (filing-level cohorts); Orange = Appendix D (FF5 regression); Purple = Apps. E–G (factor taxonomy, prompt templates, implementation details); Gray = input / routing step. IC = Spearman correlation between sentiment score and 21-day forward returns.

7.9% to 19.6% annualised, while the LM baseline points in the opposite direction, producing a −20.7% annualised long-short spread.

The Neg cohort is the stronger signal: GRPO's most confidently negative filings go on to underperform the market by a wide margin, giving GRPO the highest filing-level daily L/S among all learned models (Table 9). GRPO_{TTC} shows the same pattern; the K = 3 majority vote moderates the extremes slightly, yielding a filing-level L/S of +1.39 daily and +0.43 monthly.

We also check whether models rank stocks in the right order within each factor: do negative-labelled filings earn lower returns than neutral ones, which earn lower than positive? Full results are in Appendix B.2. **RL models pass this ordering test more often, especially at the 21-day horizon.** Only 7 of 14 factors have any model passing at 21 days; five of those seven are led by an RL model (DPO or GRPO variant). Capital Intensity is the most consistent factor — three models pass at the 1-day horizon and two (GRPO, GPT) at 21 days. Security/Privacy and Pricing & Margin fail at every horizon for every model, consistent with finding (3) above.

5.3 Filing-Level Majority-Vote Cohorts

For each filing we combine all subfactor labels into one direction by majority vote and form three groups: Neg, Neutral, and Pos. Full results are in Appendix C (Tables 9–13). Four findings stand out.

(1) The Neg cohort is also the stronger signal at filing-level. Across all six learned models, Negative-labelled filings consistently underperform the market (Table 9), and this holds after Fama-French five-factor adjustment (Appendix D). LM shows the opposite — its Neg cohort earns positive excess returns — confirming the signal comes from contextual, topic-grounded reasoning rather than from counting negative finance words.

(2) The Pos cohort is too large to trade on its own. Learned models assign over half of all filings to Positive, making the Pos cohort large but barely different from the benchmark. The actionable edge is on the short side, not the long side.

(3) Factor pre-selection only works when paired with two-step aggregation. Without pre-selection, one-step and two-step aggregation give almost the same result — the order of aggregation barely matters. Once we restrict each ticker to its three most return-predictive parent factors (IC pre-selection), a large gap opens: two-step improves by +0.22 IR daily and +0.61 monthly on aver-

age, while one-step *degrades* by -0.71 IR daily and -0.02 monthly. The reason: one-step votes across only 3 subfactors per filing, while two-step retains coverage by voting within each selected parent factor before aggregating across the three parent labels. The practical rule: *always pair IC pre-selection with two-step aggregation*. Using pre-selection with one-step is worse than using no selection at all.

(4) GRPO_{TTC} is the exception. For every other model, two-step beats one-step after IC pre-selection. GRPO_{TTC} reverses this pattern: one-step gives $+1.02$ daily and $+1.65$ monthly versus two-step's $+0.80$ and $+0.63$ — the strongest one-step outcome across all models. With 1,373 predictions, GRPO_{TTC} retains enough post-filter coverage for the strict one-step filter to form a high-purity signal that two-step would dilute with weaker subfactor evidence.

6 Conclusion

We presented RAFE, a pipeline that reads SEC filings one investment topic at a time, trains Qwen3-8B to align those labels with future stock returns, and turns the resulting signals into portfolios. Three findings stand out.

Negative-labelled filings show directional underperformance in our test sample. Across most learned models, Negative-labelled filings underperform the benchmark by 7.9% to 19.6% annualised, while the word-counting baseline (LM) points in the opposite direction (-20.7% annualised L/S). This pattern is consistent with contextual reasoning, rather than keyword counting, driving the signal.

Topic pre-selection appears to matter more than model choice. Without topic filtering, the Positive cohort is too large to trade cleanly — RL training pushes Neutral predictions down from 27% to 7–13%, inflating the Positive group with low-conviction labels. Restricting each company to its three strongest topics and using two-step aggregation improves Sharpe ratios in our sample: SFT_{TTC} monthly long-short Sharpe rises from $+0.74$ to $+2.37$, and the GPT teacher from -0.05 to $+1.31$. The practical rule we observe: IC pre-selection should be paired with two-step aggregation; combining it with one-step aggregation made results worse than no filtering at all in our experiments.

Some topics carry stronger signals than others. Capital Intensity appears the most robust factor in our test sample — GRPO earns $+2.34$ monthly with consistent return predictability at both 1-day and 21-day horizons. Competitive Dynamics shows the next strongest signal ($+2.47$ monthly, DPO_{TTC}). Security/Privacy and Pricing & Margin point in the wrong direction for every learned model and warrant exclusion or investigation before deployment.

Overall, these results provide directional evidence that factor-level sentiment signals can be extracted from SEC filings and combined into return-predictive portfolios on our held-out 2024–2025 sample. They do not establish robust alpha: the test window is short (18 monthly periods) and limited to the IT sector; confirmation requires longer horizons and out-of-sample replication across sectors.

Limitations

Scope and statistical power. The evaluation covers 49 S&P 500 technology companies over an 18-month test window (2024–2025), with fewer than 50 events per topic per model. No individual per-topic result survives multiple-testing correction, so findings should be read as directional evidence rather than statistically confirmed alphas. Fama-French 5-factor regressions nonetheless show that Competitive Dynamics and M&A activity alpha survive risk adjustment. Extending RAFE to a broader sector universe and longer backtest is the natural next step.

Label noise from stock returns. Using a single 21-day stock return to label an individual topic is a simplification: if a filing is positive on acquisitions but negative on regulation, the return conflates both. Remedies include topic-level contrastive rewards or recalibrating confidence after majority voting. Scaling to the full S&P 500 and longer horizons is the natural next step, though costly in GPT extraction calls.

AI model training cutoffs. GPT-5-nano and GPT-5-mini knowledge cutoffs may overlap the Jan 2024–Jun 2025 test window, so these models may have prior knowledge of test-period events — a source of potential look-ahead bias that is difficult to measure or control.

Annotation design. The 5,000 human-revised training examples were partitioned across three annotators rather than jointly labeled, prioritising coverage over inter-annotator agreement; we therefore cannot report Cohen's κ or Krippendorff's α. Future iterations should reserve a small overlap (5–10%) for IAA estimation.

Ethics Statement

RAFE processes publicly available SEC filings and publicly traded equity prices. No private or personally identifiable information is used. The portfolio backtest is conducted on historical data for research purposes; it does not constitute investment advice. Automated extraction relies on GPT APIs; outputs should be reviewed by domain experts before operational deployment.

Reproducibility

The complete RAFE artefact — full training and inference codebase, 5,000 human-revised training samples, preprocessed EDGAR feature panels, model checkpoints, and the backtesting harness — is released at `https://anonymous.4open.science/r/RAFE-7884` (anonymous mirror for double-blind review; public GitHub release upon acceptance). A dedicated `paper_reproduction/` subfolder provides one script per numerical table (Tables 1–19), so reviewers can independently reproduce any reported number.

References

Dogu Araci. 2019. FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. 2024. KTO: Model alignment as prospect theoretic optimization. In *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*.

Eugene F. Fama and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Campbell R. Harvey, Yan Liu, and Heqing Zhu. 2016. …and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.

Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531.

Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. 2009. Predicting risk from financial reports with regression. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL (NAACL 2009)*, pages 272–280.

Thanos Konstantinidis, Giorgos Iacovides, Mingxue Xu, Tony G. Constantinides, and Danilo Mandic. 2024. FinLlama: Financial sentiment classification for algorithmic trading applications.

Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. 2022. Teaching small language models to reason. *arXiv preprint*.

Pekka Malo, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala. 2014. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, 65(4):782–796.

Alexandru Niculescu-Mizil and Rich Caruana. 2005. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML 2005)*, pages 625–632. ACM.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelley, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*.

Qwen Team. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Paul C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Dimitris Vamvourellis and Dhagash Mehta. 2025. Reasoning or overthinking: Evaluating large language models on financial sentiment analysis. *arXiv preprint arXiv:2506.04574*.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models. In *Proceedings of the 11th International Conference on Learning Representations (ICLR 2023)*.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023.

BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023. FinGPT: Open-source financial large language models. *arXiv preprint arXiv:2306.06031*.

A Label Statistics

Table 2 shows how sentiment labels are distributed across models and data splits (train 2015–2022, val 2023, test Jan 2024–Jun 2025) at the subfactor level. Four patterns stand out. First, extreme labels (VN and VP) are $\leq 1.4\%$ everywhere, so collapsing to three classes (Neg, Neu, Pos) loses almost nothing. Second, the Neg share is stable at 32–37% across all models, confirming that RL training does not distort the negative signal. Third, RL training dramatically shrinks the Neutral class — from 27% (GPT teacher test) to 7–13% across learned models — shifting probability into Pos (37% $\rightarrow$ 53–58%); GRPO shows the strongest collapse (7.7% Neu). This Neutral collapse explains why over half of filings land in the Pos cohort in subsequent portfolio tables. Fourth, GRPO_{TTC} has the highest test-period coverage (1,373 predictions); single-run SFT and DPO have the lowest (1,037–1,041), reflecting occasional output parsing failures. Figure 2 maps how these subfactor labels flow through all experiments in Appendices B–D.

Table 2: Label distribution across splits (subfactor-event level). Train and Val show GPT teacher labels; Test rows show each model's valid predictions. LM = Loughran–McDonald dictionary scored at the factor-audit level (one label per factor–filing pair; VN and VP are absent by construction). VN = very_negative; VP = very_positive.

Split / Model	Total	VN (%)	Neg (%)	Neu (%)	Pos (%)	VP (%)	VN+VP (%)
Train (GPT teacher)	8,451	24 (0.3)	2,895 (34.3)	2,143 (25.4)	3,334 (39.5)	55 (0.7)	79 (0.9)
Val (GPT teacher)	1,034	8 (0.8)	370 (35.8)	282 (27.3)	373 (36.1)	1 (0.1)	9 (0.9)
Test – LM	1,458	0 (0.0)	529 (36.3)	596 (40.9)	333 (22.8)	0 (0.0)	0 (0.0)
Test – GPT teacher	1,458	3 (0.2)	499 (34.2)	397 (27.2)	544 (37.3)	15 (1.0)	18 (1.2)
Test – SFT	1,037	4 (0.4)	341 (32.9)	136 (13.1)	545 (52.6)	11 (1.1)	15 (1.4)
Test – DPO	1,041	3 (0.3)	331 (31.8)	125 (12.0)	571 (54.9)	11 (1.1)	14 (1.3)
Test – GRPO	1,169	1 (0.1)	388 (33.2)	90 (7.7)	677 (57.9)	13 (1.1)	14 (1.2)
Test – SFT_{TTC}	1,179	0 (0.0)	418 (35.5)	81 (6.9)	679 (57.6)	1 (0.1)	1 (0.1)
Test – DPO_{TTC}	1,189	0 (0.0)	438 (36.8)	90 (7.6)	660 (55.5)	1 (0.1)	1 (0.1)
Test – GRPO_{TTC}	1,373	0 (0.0)	477 (34.7)	123 (9.0)	772 (56.2)	1 (0.1)	1 (0.1)

B Per-Factor Long-Short Results

This appendix evaluates factor-level sentiment signals in two complementary ways. Appendix B.1 forms a long-short portfolio within each of the 14 parent investment factors, going long stocks labelled Positive and short stocks labelled Negative, to measure whether factor-specific sentiment ranks firms by subsequent return. Appendix B.2 tests a stricter condition: whether mean excess returns follow the expected ordering Neg $<$ Neu $<$ Pos at the individual filing-factor level. Together, these tests separate factor-level tradability from basic directional rank ordering.

B.1 Long-Short Portfolio Performance by Factor

For each factor, we build a daily long-short portfolio: go long stocks labelled Positive and short stocks labelled Negative on that factor, rebalanced daily with a one-day entry delay. To get one factor label per (ticker, filing), we collapse the model's five-class output to three directions (Neg, Neu, Pos) and take a majority vote across all subfactors within the parent factor; ties go to Neu. Tables 3 and 4 report the Sharpe ratio of this Pos$-$Neg spread under daily and monthly rebalancing.

Competitive Dynamics and M&A/CorpDev are the most reliable factors. The strongest learned-model results are concentrated in these themes; DPO_{TTC} reaches $+2.47$ monthly Sharpe on Competitive Dynamics, GPT reaches $+1.99$ on M&A/CorpDev, and GRPO reaches $+2.34$ on Capital Intensity. The LM baseline is weak or negative on these factors, suggesting that the edge comes from contextual interpretation of filing language rather than simple polarity-word counts. GRPO_{TTC} also delivers the strongest model-specific results for Go-to-Market ($+1.91$ monthly) and Management Quality ($+1.86$ monthly).

Go-to-Market is the most model-dependent factor.

The sign of the signal changes sharply across models. SFT earns $+1.82$ monthly Sharpe, while GPT loses -3.34 over the same factor and test window. This suggests the Go-to-Market language is

especially ambiguous. Discussion of sales execution, channel expansion, demand generation, and customer acquisition can indicate either genuine growth momentum or defensive commercial pressure, so different training objectives extract opposite investment signals from similar evidence.

Security/Privacy and Pricing & Margin go in the wrong direction for every learned model. Every learned model loses money on Security/Privacy, with monthly Sharpe ratios from -1.05 to -2.41, while LM earns $+1.76$. Pricing & Margin is also negative across all learned models; LM is not positive there, but it is the least negative model at -0.24 monthly. These two factors are therefore not merely weak in this sample: for LLM-extracted sentiment, they are directionally unreliable and should either be excluded, treated as diagnostic rather than tradable, or revisited with factor-specific prompts.

Table 3: Per-Factor Long-Short (Pos−Neg) Performance: Daily Rebalancing (Jan 2024–Jun 2025). *Shr* = annualised Sharpe of Pos−Neg zero-cost spread (RF = 0; standard convention for self-financing portfolios); *MDD* = max drawdown of Pos cohort (%). Subfactor labels are aggregated to parent-factor level by majority vote: for each (ticker, filing, factor), each subfactor label is mapped to coarse Neg/Neu/Pos and the direction with more votes wins (ties → Neu). LM = Loughran–McDonald dictionary applied to factor-specific evidence quotes (Stage 1 output). Bold = best Sharpe per factor row. "—" = cohort absent ($<$ 3 periods). Models: LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

	LM		GPT		SFT		DPO		GRPO		SFT_{TTC}		DPO_{TTC}		GRPO_{TTC}	
Factor	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*
AI Infra Cycle	+1.08	−17.3	**+1.40**	−20.4	+1.00	−26.2	+0.71	−19.9	−0.75	−25.6	−0.49	−26.9	−0.45	−26.9	−0.77	−32.7
Capital Intensity	−0.84	−35.9	+0.47	−13.0	+0.83	−8.8	+0.57	−9.6	**+1.24**	−15.0	+0.67	−23.6	+0.07	−14.3	+0.24	−15.8
Competitive Dynamics	−0.10	−29.4	+0.42	−13.8	**+1.06**	−13.8	+0.14	−14.8	+0.48	−17.1	+0.12	−30.1	+1.01	−13.2	+0.44	−31.7
Data Moat†	—	—	**+0.41**	−36.6	—	—	—	—	+0.37	−36.8	—	—	—	—	−1.87	−36.5
Financial Quality	−1.26	−28.8	+0.79	−13.7	+0.15	−18.2	−0.69	−15.7	+0.29	−12.3	−0.11	−25.0	**+1.02**	−8.2	+0.74	−17.3
Go-to-Market	−1.66	−53.1	−2.29	−45.2	**+1.62**	−22.4	+0.18	−15.9	−0.15	−29.4	−0.34	−28.6	−1.44	−44.4	+1.28	−29.7
M&A / CorpDev	+0.09	−10.2	**+1.34**	−13.2	+1.20	−11.1	+0.55	−8.2	+0.36	−12.4	+0.57	−9.4	+0.34	−11.4	+0.15	−11.5
Mgmt. Quality	+0.20	−18.5	−0.89	−33.5	**+0.88**	−13.4	−0.04	−19.1	−0.47	−21.1	−0.19	−20.9	+0.23	−14.1	+0.79	−11.4
Pricing & Margin	−0.52	−19.3	−0.67	−22.0	−0.56	−26.6	−1.11	−28.7	−0.86	−26.1	**−0.17**	−14.0	−0.99	−24.8	−0.75	−24.2
Product Platform‡	−1.36	−42.6	−0.61	−23.6	**+0.85**	−27.8	—	—	−0.21	−3.2	—	—	—	—	+0.42	−19.4
Regulatory / Legal	−0.31	−40.1	−0.69	−31.5	−1.00	−28.8	−0.63	−27.4	**+0.37**	−18.5	−0.06	−23.9	+0.14	−23.9	−0.72	−24.6
Security / Privacy	**+0.63**	−19.1	−0.81	−31.3	−1.95	−64.5	−1.43	−40.0	−1.37	−56.4	−1.60	−43.9	−1.16	−49.8	−1.30	−58.0
Structural Risks	**+0.74**	−9.1	+0.08	−26.8	+0.33	−33.1	+0.23	−20.1	+0.12	−31.6	+0.42	−17.4	−0.08	−27.1	+0.17	−23.0
Tech Infra	−0.79	−30.1	−0.66	−34.1	+0.05	−23.8	**+0.56**	−19.1	+0.36	−21.9	+0.21	−26.7	−0.38	−32.4	+0.40	−24.4

† Data Moat: only GPT, GRPO, and GRPO_{TTC} produce a non-empty Neg cohort; LM and others shown as "—". ‡ Product Platform: DPO, SFT_{TTC}, DPO_{TTC} have zero Neg-cohort periods; LM, GPT, SFT, GRPO, GRPO_{TTC} only.

Table 4: Per-Factor Long-Short (Pos−Neg) Performance: Monthly Rebalancing (18 periods, Jan 2024–Jun 2025). *Shr* = annualised Sharpe of Pos−Neg zero-cost spread (RF = 0; standard convention for self-financing portfolios); *MDD* = max drawdown of Pos cohort (%). Subfactor labels aggregated to parent-factor level by majority vote (same as Table 3). LM = Loughran–McDonald dictionary applied to factor-specific evidence quotes (Stage 1 output). Bold = best Sharpe per factor row. "—" = cohort absent (< 3 periods). Models: LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, $GRPO_{TTC}$.

	LM		GPT		SFT		DPO		GRPO		SFT_{TTC}		DPO_{TTC}		$GRPO_{TTC}$	
Factor	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*	*Shr*	*MDD*
AI Infra Cycle	+1.05	−12.7	**+1.30**	−10.6	+0.18	−22.5	+0.18	−15.5	−0.56	−18.6	−0.34	−22.4	−0.48	−17.5	−0.73	−27.4
Capital Intensity	−1.19	−36.2	+0.63	−7.6	+0.53	−9.8	+1.04	−10.4	**+2.34**	−8.1	+0.35	−21.8	−0.44	−19.6	−0.02	−12.6
Competitive Dynamics	−1.16	−30.8	+0.84	−10.0	+2.37	−4.6	−1.56	−14.6	+0.69	−18.5	+0.27	−22.3	**+2.47**	−4.5	+0.65	−20.2
Data Moat[†]	—	—	+1.06	−16.8	—	—	—	—	**+1.15**	−14.4	—	—	—	—	−2.05	−29.2
Financial Quality	−1.39	−26.8	**+0.85**	−4.9	+0.39	−9.3	−0.56	−7.8	−0.11	−6.8	−0.27	−10.6	+0.42	−14.0	−0.04	−12.3
Go-to-Market	−2.77	−46.9	−3.34	−44.9	+1.82	−6.1	−0.05	−7.8	+0.04	−28.3	−0.43	−11.4	−1.25	−36.1	**+1.91**	−10.4
M&A / CorpDev	−0.16	−8.9	**+1.99**	−3.3	+1.77	−3.0	−0.16	−3.2	−0.19	−6.3	+0.69	−0.2	+0.02	−8.0	−0.32	−7.9
Mgmt. Quality	−0.19	−14.8	−0.82	−22.4	+1.38	−7.1	+0.56	−10.0	+0.50	−6.3	−0.06	−15.2	+1.15	−5.7	**+1.86**	−4.7
Pricing & Margin	**−0.24**	−20.0	−0.78	−24.8	−0.74	−24.3	−1.13	−29.1	−0.95	−24.8	−0.42	−22.3	−1.15	−25.3	−0.84	−24.8
Product Platform[‡]	−2.27	−31.6	−0.41	−21.1	**+1.27**	−19.4	—	—	—	—	—	—	—	—	+0.42	−17.7
Regulatory / Legal	−0.01	−27.0	−1.26	−24.2	−0.95	−32.5	−0.47	−26.8	**+0.72**	−13.9	−0.14	−16.4	+0.17	−19.0	−0.91	−24.4
Security / Privacy	**+1.76**	−9.4	−1.05	−28.9	−2.41	−57.5	−1.85	−45.0	−1.43	−39.6	−1.78	−42.0	−1.81	−48.4	−1.18	−46.0
Structural Risks	+0.74	−8.5	+0.54	−22.2	+0.38	−28.1	+0.19	−20.6	+0.24	−14.1	**+1.16**	−17.7	+0.02	−22.5	+0.54	−18.0
Tech Infra	−1.23	−30.6	−0.91	−34.0	−0.33	−19.1	+0.10	−11.8	−0.20	−19.3	**+0.27**	−15.5	−0.50	−31.5	+0.10	−15.3

[†] Data Moat: empty Neg cohort for LM/SFT/DPO/SFT_{TTC}/DPO_{TTC}; GPT, GRPO, and $GRPO_{TTC}$ produce results. [‡] Product Platform: empty Neg for DPO/GRPO/SFT_{TTC}/DPO_{TTC}; LM, GPT, SFT, $GRPO_{TTC}$ only.

B.2 Parent-Factor Return Monotonicity

This section checks whether each model assigns sentiment labels that line up with actual stock returns at the parent-factor level. For each (company, filing date, parent factor) we aggregate subfactor labels via majority vote to get one label — Neg, Neu, or Pos — then record the stock's excess return over the IT sector benchmark starting from that filing date. Each data point is one filing by one company; companies with multiple filings contribute multiple observations. The excess return is *not* annualised and is *not* directly comparable to the IR/Sharpe in Tables 9–13, which use a full daily-rebalancing portfolio with a one-day entry delay.

A model passes the monotonicity test for a factor if its mean excess returns satisfy Neg < Neu < Pos (negative-labelled filings earn the lowest returns, positive the highest) with at least three filings per cohort. For each factor we show the passing model with the largest Pos−Neg spread.

Table 5 uses a 1-day excess return with a 1-day entry delay (t+1 to t+2), consistent with how the portfolio tables enter trades. Table 6 uses a 21-day excess return from the filing date (t to t+21), a longer window that captures more of the information in the filing.

Table 5: Best monotone model per factor — **1-day excess return, 1-day entry delay** (t+1 to t+2). Test period Jan 2024–Jun 2025. Each row is one (filing, parent factor) pair; excess return (relative to equal-weight IT sector) measured over one trading day starting the day after the filing. A model passes if mean returns satisfy Neg < Neu < Pos with ≥3 filings per cohort. "—" = no model passes for that factor.

Factor	Best Model	Neg (%)	Neu (%)	Pos (%)	Pos−Neg (%)	Also monotone
AI Infra Cycle	DPO_{TTC}	+0.49	+0.62	+0.83	+0.34	—
Capital Intensity	DPO	−0.68	+0.17	+0.75	+1.42	SFT, GRPO
Competitive Dynamics	—	—	—	—	—	—
Data Moat	—	—	—	—	—	—
Financial Quality	DPO_{TTC}	−0.19	−0.10	+0.33	+0.52	DPO, SFT_{TTC}, GPT
Go-to-Market	SFT	+0.03	+0.70	+0.73	+0.70	—
M&A / CorpDev	—	—	—	—	—	—
Mgmt. Quality	—	—	—	—	—	—
Pricing & Margin	GPT	−0.20	−0.10	+0.63	+0.83	DPO
Product Platform	GPT	−0.42	+0.19	+0.55	+0.97	—
Regulatory / Legal	SFT	−0.21	−0.14	−0.02	+0.19	—
Security / Privacy	GRPO	−1.02	−0.97	−0.14	+0.88	—
Structural Risks	—	—	—	—	—	—
Tech Infra	GRPO	−0.23	−0.05	+0.19	+0.42	—

Table 6: Best monotone model per factor — **21-day excess return from filing date** (t to t+21). Test period Jan 2024–Jun 2025. Each row is one (filing, parent factor) pair; excess return (relative to equal-weight IT sector) measured over 21 trading days from the filing date. A model passes if mean returns satisfy Neg < Neu < Pos with ≥3 filings per cohort. "—" = no model passes for that factor.

Factor	Best Model	Neg (%)	Neu (%)	Pos (%)	Pos−Neg (%)	Also monotone
AI Infra Cycle	—	—	—	—	—	—
Capital Intensity	GRPO	−1.21	+0.33	+0.79	+2.00	GPT
Competitive Dynamics	—	—	—	—	—	—
Data Moat	—	—	—	—	—	—
Financial Quality	DPO_{TTC}	−1.29	−1.10	+1.98	+3.27	GRPO, GPT, GRPO_{TTC}
Go-to-Market	DPO	−2.66	−2.03	+0.14	+2.80	—
M&A / CorpDev	GPT	−1.70	+0.68	+1.21	+2.91	GRPO
Mgmt. Quality	GRPO_{TTC}	−2.40	−2.18	−1.12	+1.28	—
Pricing & Margin	—	—	—	—	—	—
Product Platform	—	—	—	—	—	—
Regulatory / Legal	DPO_{TTC}	−2.23	−1.52	+0.88	+3.11	—
Security / Privacy	—	—	—	—	—	—
Structural Risks	—	—	—	—	—	—
Tech Infra	GPT	−3.06	−1.17	+1.12	+4.18	SFT, GRPO

Key findings. RL models rank stocks in the right order more often, especially over longer horizons. At the 21-day horizon (Table 6), only half of the 14 factors have any model passing the monotonicity test, and the majority of those passing factors are led by RL-trained models (DPO or GRPO variants). The sentiment signal takes time to be reflected in prices — RL alignment helps models extract the directional content that matters over weeks, not just overnight.

At the 1-day horizon, more factors pass but the signal is weaker. Nine factors have at least one passing model over the 1-day delayed window, but the Pos−Neg spreads are smaller and several passes come from non-RL models. This is consistent with short-term filing reactions capturing obvious information, while the 21-day test better reflects whether sentiment labels rank firms by sustained post-filing performance.

It is the only factor with multiple models passing the monotonicity test at both horizons: DPO, SFT, and GRPO at 1 day, and GRPO and GPT at 21 days. This consistency matches the portfolio results, where Capital Intensity is one of the most robust long-short factors.

Security/Privacy and Pricing & Margin have no sustained signal. Both fail the ordering test at the 21-day horizon for every model. Any apparent 1-day reaction in Security/Privacy appears to mean-revert quickly. These two factors are not suitable for sentiment-based trading regardless of the model used.

C Filing-Level Majority-Vote Cohort Analysis

This appendix evaluates filing-level portfolio signals built by collapsing all subfactor sentiment labels into a single filing-level direction via majority vote. **C.1** presents results using all subfactor labels with no pre-filtering, testing one-step and two-step aggregation side by side. **C.2** applies IC-based factor selection — restricting each ticker to its most return-predictive subfactors or parent themes before aggregating. **C.3** compares the two approaches directly and summarises the key findings. Table 7 shows the filing-level label distribution shared across all three subsections.

Table 7: Filing-level label counts after majority-vote aggregation across factors (test period Jan 2024–Jun 2025). VN/VP are merged into Neg/Pos respectively ($<$ 1% of all labels). LM = Loughran–McDonald dictionary; one filing label from majority vote across factor-audit LM scores. Percentages in parentheses are relative to total filings per model.

Model	Total	Neg (%)	Neu (%)	Pos (%)
LM	255	83 (32.5)	137 (53.7)	35 (13.7)
GPT	254	93 (36.6)	56 (22.0)	105 (41.3)
SFT	246	64 (26.0)	45 (18.3)	137 (55.7)
DPO	242	57 (23.6)	47 (19.4)	138 (57.0)
GRPO	244	52 (21.3)	41 (16.8)	151 (61.9)
SFT_{TTC}	245	60 (24.5)	43 (17.6)	142 (58.0)
DPO_{TTC}	251	69 (27.5)	52 (20.7)	130 (51.8)
$GRPO_{TTC}$	253	61 (24.1)	37 (14.6)	155 (61.3)

Most very_negative and very_positive labels are rare ($<$3 filings per model), so they are folded into Neg and Pos respectively, leaving three cohorts. The LM dictionary is inverted: it labels 54% of filings Neutral and only 14% Positive, while every learned model assigns 52–62% Positive. As a result, LM's only useful signal is the Neg cohort, and its L/S spread is negative — the opposite of every learned model.

C.1 No IC-Based Subfactor and Parent Selection

Both aggregation methods in this section use all subfactor labels without any pre-filtering. **One-step** counts a direct majority vote across every subfactor label for a given filing and assigns the winning direction as the filing signal. **Two-step** first votes within each of the 14 parent investment themes to produce one label per theme, then takes a majority vote across those 14 theme labels. Both methods see identical data, so portfolio sizes and performance are nearly the same (see design in Table 8).

Table 8: Aggregation design for Tables 9–10: all subfactor labels used, no IC pre-selection.

	One-step	**Two-step**
Input	All subfactor labels per (ticker, filing)	All subfactor labels per (ticker, filing)
Aggregation	Direct majority vote across all subfactors $\rightarrow$ filing label	**Step 1**: majority vote within each parent factor $\rightarrow$ 1 label per parent. **Step 2**: majority vote across up to 14 parent labels $\rightarrow$ filing label
Avg N (daily)	$\approx$8–29 tickers	$\approx$8–29 tickers

Key findings. Tables 9–10 report filing-level cohort performance without any IC-based pre-selection. **Aggregation order matters little when all subfactors are retained.** One-step and two-step aggregation produce similar long-short performance because both methods see essentially the same underlying information.

Among learned models, the strongest long-short results are driven primarily by the short side: filings labelled Negative consistently underperform the equal-weight IT benchmark. The Positive cohort is much larger and usually closer to the benchmark, especially after RL alignment shifts probability mass away from Neutral and toward Positive. DPO_{TTC} Neutral is a notable exception: split-vote filings under this model earn unusually strong returns, suggesting that ambiguous filings may still contain useful information, though this result should be interpreted cautiously given the smaller cohort size.

Table 9: 3-cohort performance with *daily* rebalancing and *filing-level majority-vote* sentiment (Pos = positive ∪ very_positive; Jan 2024–Jun 2025). Two aggregation methods shown side by side: **One-step** = majority vote across all subfactor labels directly; **Two-step** = majority vote subfactors→parent factor first, then across 14 parent factors (consistent with Tables 3–4). IR = annualised Information Ratio vs. equal-weight IT benchmark; Shr = annualised Sharpe (cohort: minus RF; L/S: RF = 0); Exc = annualised excess return (%); Vol = annualised volatility (%); MDD = max drawdown (%); IC = Spearman(S, next-period return); Avg N = average tickers per period. Δ Shr = Two-step minus One-step Sharpe. **Bold** = best IR per model per method. Models include LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

		One-step							Two-step							
Model	Cohort	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	Δ Shr
LM	Neg	+**0.57**	+1.00	+3.7	6.5	−28.4	−0.01	23.9	+**0.52**	+1.02	+3.5	6.7	−28.2	−0.00	23.0	+0.02
	Neu	+0.36	+1.02	+3.1	8.8	−26.2	+0.01	9.7	+0.32	+1.00	+2.8	8.8	−26.3	+0.01	10.4	−0.02
	Pos	−1.90	+0.44	−16.8	8.8	−29.6	−0.02	12.1	−2.12	+0.39	−18.2	8.6	−29.4	−0.01	12.2	−0.05
	L/S	−1.73	−1.73	−19.5	11.3	−24.5	−0.02	35.6	−1.87	−1.87	−20.7	11.1	−27.6	−0.02	34.8	−0.14
GPT	Neg	−0.86	+0.68	−6.9	8.0	−27.3	+0.02	16.7	−1.01	+0.64	−7.9	7.8	−27.1	+0.02	16.3	−0.04
	Neu	+0.28	+0.98	+2.7	9.3	−27.7	+0.00	10.1	+0.28	+0.96	+2.8	9.8	−27.5	+0.00	10.3	−0.02
	Pos	−0.43	+0.93	−3.1	7.1	−29.4	+0.01	19.1	−0.33	+0.96	−2.3	7.1	−29.1	+0.00	18.9	+0.03
	L/S	+**0.42**	+0.42	+4.5	10.7	−13.6	+0.01	35.5	+**0.60**	+0.60	+6.3	10.5	−13.1	+0.00	34.9	+0.18
SFT	Neg	−1.06	+0.43	−11.9	11.2	−37.0	−0.01	11.0	−0.97	+0.45	−11.1	11.5	−37.7	−0.01	10.8	+0.02
	Neu	−0.59	+0.69	−6.1	10.2	−28.5	+0.01	7.1	−0.53	+0.76	−5.3	10.0	−28.4	+0.00	8.3	+0.07
	Pos	+0.17	+1.12	+1.1	6.4	−26.7	+0.00	26.8	+0.09	+1.08	+0.6	6.5	−26.6	+0.00	25.8	−0.03
	L/S	+**1.02**	+1.02	+13.9	13.6	−17.3	+0.01	37.6	+**0.91**	+0.91	+12.5	13.8	−19.2	+0.00	36.4	−0.11
DPO	Neg	−0.92	+0.48	−10.9	11.8	−34.6	+0.00	10.2	−0.89	+0.48	−10.6	11.8	−37.2	+0.00	9.8	−0.00
	Neu	−0.24	+0.76	−3.2	13.3	−24.9	+0.01	7.7	−0.28	+0.77	−3.6	12.6	−22.6	+0.02	8.4	+0.01
	Pos	−0.64	+0.93	−4.0	6.2	−27.9	+0.01	26.9	−0.61	+0.94	−3.8	6.2	−27.9	+0.00	26.6	+0.01
	L/S	+**0.56**	+0.56	+7.7	13.7	−13.3	+0.01	36.9	+**0.56**	+0.56	+7.5	13.5	−16.0	+0.01	36.3	+0.00
GRPO	Neg	−1.26	+0.26	−17.2	13.6	−43.5	+0.01	8.4	−1.52	+0.19	−19.6	12.9	−44.6	+0.02	8.9	−0.07
	Neu	−0.06	+0.77	−0.7	12.5	−33.5	−0.01	7.9	+0.08	+0.88	+0.8	10.2	−30.7	+0.01	8.8	+0.11
	Pos	−0.27	+1.04	−1.6	6.0	−26.6	+0.01	28.7	−0.11	+1.06	−0.7	6.1	−26.3	+0.01	27.4	+0.03
	L/S	+**1.06**	+1.06	+16.4	15.5	−17.6	+0.01	37.0	+**1.32**	+1.32	+19.7	14.9	−17.5	+0.01	36.1	+**0.26**
SFT_{TTC}	Neg	−1.66	+0.22	−18.9	11.4	−40.2	−0.00	10.7	−1.72	+0.22	−19.0	11.1	−39.4	+0.00	10.5	+0.00
	Neu	+0.29	+0.94	+3.4	11.6	−29.1	−0.01	6.9	+1.02	+1.17	+12.8	12.6	−28.6	+0.02	7.6	+0.23
	Pos	+0.06	+1.10	+0.3	6.3	−26.8	+0.00	26.8	−0.35	+1.01	−2.2	6.3	−26.9	+0.01	26.3	−0.09
	L/S	+**1.46**	+1.46	+20.1	13.7	−15.5	+0.01	37.3	+**1.31**	+1.31	+17.6	13.4	−17.4	+0.01	36.6	−0.15
DPO_{TTC}	Neg	−1.47	+0.25	−17.6	12.0	−41.3	+0.00	12.2	−1.57	+0.23	−18.3	11.6	−41.1	+0.01	12.1	−0.02
	Neu	+**2.16**	+1.59	+22.5	10.4	−27.6	+0.01	7.8	+**2.28**	+1.62	+22.2	9.8	−27.1	+0.00	9.6	+0.03
	Pos	−0.45	+0.99	−3.1	6.8	−25.3	−0.01	25.2	−0.68	+0.91	−4.7	6.9	−25.6	+0.00	23.6	−0.08
	L/S	+1.02	+1.02	+15.3	15.0	−17.3	+0.00	37.2	+1.00	+1.00	+14.2	14.3	−17.2	+0.01	35.5	−0.02
GRPO_{TTC}	Neg	−1.61	+0.21	−19.2	11.9	−40.6	−0.02	10.1	−1.55	+0.27	−17.3	11.2	−40.2	−0.00	10.1	+0.06
	Neu	−0.12	+0.70	−1.5	12.8	−31.9	−0.01	7.2	+0.04	+0.77	+0.5	11.3	−29.7	−0.00	8.8	+0.07
	Pos	+0.06	+1.13	+0.4	6.4	−25.4	−0.00	28.1	+0.04	+1.13	+0.3	6.6	−25.0	−0.01	26.6	+0.00
	L/S	+**1.39**	+1.39	+20.4	14.6	−18.6	+0.00	38.0	+**1.29**	+1.29	+18.4	14.2	−18.5	−0.00	36.5	−0.10

LM goes in the wrong direction. Its long-short spread is negative because its polarity labels are inverted relative to subsequent stock outcomes. This contrast reinforces the main result: useful filing-level sentiment comes from contextual reasoning, not from counting negative finance words.

C.2 IC-Based Subfactor and Parent Selection

This section adds a pre-filtering step: before aggregating, each ticker is restricted to its most return-predictive factors. Selection is based on the Information Coefficient (IC) — the Spearman correlation between a factor's sentiment score and the next 21-day excess return — estimated once on the GPT teacher's train and validation data (2015–2023) and applied uniformly to all models. Table 11 shows the two designs.

One-step. For each ticker, select the 3 subfactors with the highest absolute IC across all 68 subfactors. At test time, only those 3 subfactors' labels vote, and the majority direction becomes the filing signal. This is a strict filter that produces small portfolios (4–25 tickers daily).

Two-step. For each ticker, select the 3 parent investment themes whose subfactors show the highest absolute IC. At test time, keep all subfactors within those 3 themes. Aggregate in two stages: majority within each theme first, then majority across the 3 theme labels. Keeping all subfactors within a theme

Table 10: 3-cohort performance with *monthly* rebalancing and *filing-level majority-vote* sentiment (Pos = positive ∪ very_positive; 18 periods, Jan 2024–Jun 2025). Two aggregation methods shown side by side: **One-step** = majority vote across all subfactor labels directly; **Two-step** = majority vote subfactors→parent factor first, then across 14 parent factors (consistent with Tables 3–4). IR = annualised Information Ratio vs. equal-weight IT benchmark; Shr = annualised Sharpe (cohort: minus RF; L/S: RF = 0); Exc = annualised excess return (%); Vol = annualised volatility (%); MDD = max drawdown (%); IC = Spearman(S, next-period return); − − − = not computed for L/S; Avg N = average tickers per period. Δ Shr = Two-step minus One-step Sharpe. **Bold** = best IR per model per method. Models include LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

		One-step							Two-step							
Model	Cohort	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	Δ Shr
LM	Neg	+**1.20**	+1.38	+6.3	5.8	−11.5	−0.04	23.7	+**1.20**	+1.41	+8.8	7.4	−10.9	−0.01	22.7	+0.03
	Neu	+0.38	+1.34	+3.8	9.5	−8.4	+0.07	9.5	+0.16	+1.29	+1.5	9.1	−10.9	−0.00	10.4	−0.05
	Pos	−2.26	+0.27	−18.2	8.4	−18.7	−0.09	11.9	−2.26	+0.25	−18.5	8.2	−18.3	−0.04	12.0	−0.02
	L/S	−2.44	−2.44	−24.5	10.1	−22.5	− − −	35.6	−2.27	−2.27	−27.4	12.1	−24.9	− − −	34.7	+0.17
GPT	Neg	−0.00	+0.79	−0.1	11.0	−13.8	+0.01	15.8	−0.00	+1.04	−0.0	10.0	−14.2	−0.01	15.8	+0.25
	Neu	+**0.66**	+1.15	+6.9	8.1	−16.1	−0.00	9.6	+**0.66**	+1.24	+4.8	7.3	−15.3	+0.15	9.9	+0.09
	Pos	−0.09	+0.81	−2.0	9.0	−11.0	+0.02	19.5	−0.09	+1.11	−0.8	8.7	−11.1	+0.02	19.4	+0.30
	L/S	−0.01	−0.01	−0.2	14.6	−2.0	− − −	35.6	−0.05	−0.05	−0.8	14.6	−2.2	− − −	35.2	−0.04
SFT	Neg	−0.11	+0.84	−1.6	13.7	−21.9	−0.10	10.9	−0.15	+0.81	−2.0	13.6	−22.7	−0.06	10.8	−0.03
	Neu	−0.54	+0.83	−4.9	9.0	−15.5	+0.03	6.8	−0.85	+0.67	−8.6	10.1	−17.1	−0.00	7.8	−0.16
	Pos	+0.05	+1.28	+0.4	7.3	−9.9	+0.05	26.9	+**0.25**	+1.37	+1.9	7.7	−9.4	+0.05	26.0	+0.09
	L/S	+**0.10**	+0.10	+2.0	19.0	−10.3	− − −	37.8	+0.21	+0.21	+4.0	18.9	−9.4	− − −	36.8	+0.11
DPO	Neg	−0.26	+0.77	−3.5	13.3	−18.9	−0.07	9.9	−0.41	+0.68	−5.3	12.8	−22.2	+0.01	9.5	−0.09
	Neu	+**0.13**	+0.79	+2.0	15.0	−16.4	−0.04	7.7	+**0.19**	+1.00	+2.3	12.0	−10.6	−0.02	8.4	+0.21
	Pos	−0.51	+1.08	−3.8	7.3	−13.1	+0.06	27.2	−0.50	+1.08	−3.7	7.3	−13.1	+0.05	26.9	+0.00
	L/S	−0.01	−0.01	−0.3	18.3	−4.1	− − −	37.1	+0.09	+0.09	+1.6	18.1	−4.1	− − −	36.5	+0.10
GRPO	Neg	−0.38	+0.61	−5.8	15.2	−24.0	+0.07	8.1	−0.89	+0.37	−12.9	14.5	−29.0	+0.13	8.5	−0.24
	Neu	+**0.40**	+1.19	+4.5	11.2	−14.8	+0.02	7.8	+**0.75**	+1.38	+7.5	10.0	−13.3	+0.11	8.6	+0.19
	Pos	−0.22	+1.19	−1.8	8.3	−10.2	+0.08	28.8	−0.12	+1.21	−1.0	8.4	−9.9	+0.09	27.6	+0.02
	L/S	+0.20	+0.20	+4.0	20.5	−8.3	− − −	36.9	+0.60	+0.60	+11.9	20.0	−4.8	− − −	36.1	+**0.40**
SFT_{TTC}	Neg	−0.93	+0.39	−12.8	13.8	−28.1	−0.04	10.5	−0.82	+0.48	−10.6	12.9	−26.3	−0.04	10.4	+0.09
	Neu	+0.33	+1.19	+3.5	10.5	−14.9	−0.14	6.6	+0.40	+1.25	+4.6	11.5	−13.8	+0.11	7.2	+0.05
	Pos	+0.54	+1.44	+3.5	6.6	−9.9	+0.06	26.9	+0.31	+1.36	+2.0	6.6	−10.2	+0.06	26.5	−0.08
	L/S	+**0.91**	+0.91	+16.3	17.8	−5.7	− − −	37.5	+**0.74**	+0.74	+12.6	17.1	−6.8	− − −	36.9	−0.17
DPO_{TTC}	Neg	−1.04	+0.48	−11.7	11.2	−25.4	+0.02	12.1	−1.04	+0.48	−11.6	11.1	−25.1	−0.01	11.9	+0.00
	Neu	+**2.20**	+1.84	+23.4	10.7	−11.0	+0.02	7.5	+**1.72**	+1.72	+19.5	11.4	−11.9	+0.11	9.2	−0.12
	Pos	−0.69	+1.02	−5.4	7.8	−10.4	−0.02	25.3	−0.77	+0.94	−6.3	8.2	−10.5	−0.02	23.8	−0.08
	L/S	+0.41	+0.41	+6.3	15.4	−6.8	− − −	37.4	+0.36	+0.36	+5.4	14.8	−7.3	− − −	35.7	−0.05
GRPO_{TTC}	Neg	−0.48	+0.62	−6.3	13.1	−25.0	−0.11	9.8	−0.44	+0.63	−5.8	13.3	−25.4	−0.05	9.8	+0.01
	Neu	−0.05	+0.92	−0.7	12.6	−17.0	+0.16	7.1	+0.31	+1.11	+3.6	11.4	−12.0	+0.19	8.6	+0.19
	Pos	+0.21	+1.41	+1.3	6.4	−9.3	+0.01	28.2	+0.18	+1.40	+1.2	6.6	−9.7	−0.02	26.7	−0.01
	L/S	+**0.43**	+0.43	+7.7	17.9	−8.2	+0.07	38.1	+**0.38**	+0.38	+7.0	18.5	−9.0	+0.05	36.5	−0.05

gives larger portfolios (7–34 tickers daily) while still excluding the weakest themes.

Key findings. Tables 12–13 show results after applying IC-based factor selection. **Choosing which factors to include now matters a great deal.** When only 3 subfactors per ticker vote (one-step), the portfolios are too thin and performance degrades. When the 3 strongest parent themes are selected instead and all their subfactors are kept (two-step), coverage stays healthy and performance improves substantially — a gap that simply does not exist in C.1.

Among learned models, **SFT_{TTC} two-step is the strongest result overall**, and GRPO and GPT Teacher also benefit meaningfully from two-step selection. **The word-count baseline (LM) stays negative regardless of which factors are selected**, confirming its labels are systematically opposite to market outcomes, not salvageable by better filtering.

GRPO_{TTC} is the only exception — for this model, one-step actually beats two-step. The reason is size: GRPO_{TTC} has far more predictions than any other model, so even after the strict 3-subfactor filter enough labels survive to form a reliable signal. For every other model, two-step is the right aggregation strategy once IC pre-selection is applied.

C.3 Comparison: Effect of IC-Based Factor Selection

Table 14 summarises the impact of IC-based factor selection by averaging L/S IR across all eight models.

Table 11: Aggregation design for Tables 12–13: IC selection on GPT teacher train+val 2015–2023, applied uniformly to all models.

	One-step	**Two-step**
IC selection	Top-3 subfactors per ticker by highest absolute Spearman correlation (globally across all subfactors)	Top-3 parent factors per ticker by highest absolute Spearman correlation observed among their subfactors
Test-time input	3 selected subfactor labels	All subfactors within the 3 selected parents
Aggregation	Direct majority vote → filing label	**Step 1**: majority vote within each selected parent → 1 parent label. **Step 2**: majority vote across 3 parent labels → filing label
Avg N (daily)	≈4–25 tickers	≈7–34 tickers

Key insights. **Without factor selection, aggregation order barely matters.** When all subfactors are used, one-step and two-step deliver nearly identical long-short performance. Every theme gets an equal vote, so the order in which votes are counted does not change the outcome much.

Factor selection changes everything — but only if paired with two-step aggregation. Selecting the strongest factors hurts one-step and helps two-step, and the gap opens up at both daily and monthly horizons. One-step suffers because restricting to three subfactors per ticker leaves too few votes per filing, making portfolios thin and unreliable. Two-step benefits because it selects three parent *themes* rather than individual subfactors, keeping all their subfactors in play; coverage stays healthy while the weakest themes (such as Security/Privacy and Pricing & Margin) are filtered out.

The practical rule: pair IC selection with two-step aggregation. Two-step with IC selection is the best-performing combination on average across all models and both rebalancing frequencies. Applying IC selection to one-step is counterproductive — it produces worse results than using no selection at all.

Table 12: 3-cohort performance with *daily* rebalancing and IC-selected subfactors (Pos = positive ∪ very_positive; Jan 2024–Jun 2025). GPT teacher train+val (2015–2023) Spearman IC used to select subfactors; applied uniformly to all models. Two aggregation methods shown side by side: **One-step** = IC selects top-3 subfactors per ticker, direct majority vote → filing label; **Two-step** = IC selects top-3 parent factors per ticker; all subfactors within those parents are used: majority vote subfactors→parent factor, then majority vote across parent factors → filing label. IR = annualised Information Ratio vs. benchmark; Shr = annualised Sharpe (cohort: minus RF; L/S: RF = 0); Exc = annualised excess return (%); Vol = annualised volatility (%); MDD = max drawdown (%); IC = Spearman(S, next-period return); "− − −" = IC undefined; Avg N = average tickers per period. Δ Shr = Two-step minus One-step Sharpe. **Bold** = best L/S Sharpe per model. Models include LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

		One-step (IC top-3 subfactors per ticker, direct MV)							**Two-step** (IC top-3 parents per ticker, all-subfactor 2-step MV)							
Model	Cohort	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	Δ Shr
LM	Neg	+0.97	+1.22	+11.5	11.8	−28.7	−0.047	8.3	+0.24	+0.96	+1.6	6.9	−27.3	−0.030	15.8	−0.26
	Neu	−1.25	+0.69	−10.8	8.7	−27.2	+0.007	14.1	−0.59	+0.92	−4.6	7.8	−27.3	−0.000	16.0	+0.23
	Pos	−0.96	+0.54	−9.1	9.5	−34.9	−0.000	7.9	−0.04	+0.77	−0.5	12.0	−32.8	−0.000	10.0	+0.23
	L/S	−1.32	−1.32	−21.0	15.9	−39.2	−0.030	15.9	−0.25	−0.25	−3.6	14.4	−26.2	−0.020	25.4	+1.07
GPT	Neg	+0.60	+1.09	+6.4	10.6	−28.7	+0.003	8.2	−1.69	+0.31	−16.2	9.6	−31.7	−0.000	12.8	−0.78
	Neu	−1.38	+0.56	−14.1	10.2	−27.8	+0.002	10.5	+0.12	+1.18	+1.1	9.5	−24.3	+0.010	11.1	+0.61
	Pos	+0.13	+0.85	+1.6	12.3	−28.1	+0.008	11.8	+0.62	+1.08	+4.1	6.6	−28.8	−0.000	17.7	+0.23
	L/S	−0.50	−0.50	−8.8	17.5	−25.1	+0.000	19.6	**+1.50**	+1.50	+19.3	12.9	−8.8	+0.010	30.4	+2.00
SFT	Neg	−0.82	+0.37	−12.4	15.2	−37.7	+0.007	6.1	−1.47	+0.36	−15.0	10.2	−34.2	+0.000	9.3	−0.01
	Neu	−0.57	+0.73	−9.2	16.3	−24.0	−0.020	4.0	−0.17	+0.83	−2.7	15.6	−31.8	−0.009	7.3	+0.10
	Pos	−0.34	+0.87	−2.6	7.6	−26.2	+0.008	15.5	+0.42	+1.06	+2.4	5.7	−25.0	+0.010	24.2	+0.19
	L/S	+0.77	+0.77	+14.5	18.8	−16.5	+0.010	21.4	**+1.24**	+1.24	+16.4	13.2	−10.6	+0.014	33.5	+0.47
DPO	Neg	−0.63	+0.48	−9.0	14.3	−35.3	+0.025	5.8	−0.86	+0.53	−9.7	11.3	−29.5	−0.001	7.5	+0.05
	Neu	+0.57	+1.27	+8.5	14.9	−27.2	−0.014	4.7	−0.86	+0.59	−10.3	12.0	−34.1	+0.003	7.6	−0.68
	Pos	−0.53	+0.74	−4.8	9.1	−25.8	−0.017	14.3	−0.15	+0.97	−0.8	5.8	−24.4	+0.012	24.4	+0.23
	L/S	+0.17	+0.17	+3.2	18.7	−13.8	−0.008	20.0	**+0.65**	+0.65	+9.0	13.9	−11.6	+0.008	31.7	+0.48
GRPO	Neg	−0.73	+0.66	−7.3	9.9	−30.7	+0.007	7.6	−0.90	+0.55	−9.6	10.6	−29.1	+0.025	7.1	−0.10
	Neu	−0.49	+0.74	−8.4	16.9	−26.2	−0.022	4.1	−1.12	+0.60	−12.3	10.9	−29.5	+0.022	7.1	−0.14
	Pos	−1.04	+0.68	−7.5	7.2	−27.6	+0.001	16.6	+0.83	+1.09	+4.2	5.1	−26.0	−0.003	26.1	+0.41
	L/S	+0.06	+0.06	+0.7	12.8	−9.2	−0.001	24.1	**+1.12**	+1.12	+14.0	12.4	−7.5	+0.002	33.1	+1.06
SFT_{TTC}	Neg	−0.69	+0.59	−7.9	11.4	−33.3	+0.003	7.2	−1.40	+0.33	−16.1	11.5	−30.0	+0.003	7.4	−0.26
	Neu	−1.25	−0.27	−39.2	31.3	−49.6	− − −	2.0	−1.29	+0.43	−15.8	12.2	−36.8	+0.009	7.0	+0.69
	Pos	−0.98	+0.70	−6.4	6.5	−26.2	+0.010	17.6	+0.81	+1.13	+4.2	5.1	−24.6	−0.004	26.0	+0.43
	L/S	+0.06	+0.06	+0.8	14.4	−15.3	+0.008	24.7	**+1.63**	+1.63	+21.1	13.0	−7.1	+0.003	33.3	+1.57
DPO_{TTC}	Neg	+0.62	+1.22	+9.1	14.6	−29.9	−0.021	5.3	−0.17	+0.84	−1.6	9.8	−26.9	−0.007	8.1	−0.37
	Neu	−1.24	+0.29	−19.3	15.6	−36.0	+0.012	4.1	−0.45	+0.78	−5.0	11.1	−33.9	+0.010	8.9	+0.49
	Pos	−0.42	+0.82	−2.8	6.8	−27.6	+0.020	18.9	+0.24	+1.02	+1.4	5.6	−23.8	+0.005	23.3	+0.19
	L/S	−0.77	−0.77	−12.9	16.9	−34.4	+0.003	24.2	**+0.16**	+0.16	+2.0	12.0	−12.2	+0.001	31.4	+0.93
GRPO_{TTC}	Neg	−1.23	+0.39	−14.1	11.5	−34.2	−0.036	6.6	−0.74	+0.59	−7.6	10.3	−31.2	−0.007	9.3	+0.20
	Neu	−0.41	+0.68	−6.3	15.2	−29.7	+0.017	4.5	−0.95	+0.54	−12.0	12.6	−32.9	+0.023	7.6	−0.14
	Pos	−0.39	+0.90	−2.6	6.5	−25.3	−0.009	18.3	+0.50	+1.10	+2.7	5.4	−24.3	+0.000	24.7	+0.19
	L/S	**+1.02**	+1.02	+14.0	13.7	−7.4	+0.000	24.7	+0.80	+0.80	+10.4	13.1	−6.8	+0.005	33.9	−0.22

Table 13: 3-cohort performance with *monthly* rebalancing and IC-selected subfactors (Pos = positive ∪ very_positive; 18 periods, Jan 2024–Jun 2025). GPT teacher train+val (2015–2023) Spearman IC used to select subfactors; applied uniformly to all models. Two aggregation methods shown side by side: **One-step** = IC selects top-3 subfactors per ticker, direct majority vote → filing label; **Two-step** = IC selects top-3 parent factors per ticker; all subfactors within those parents are used: majority vote subfactors→parent factor, then majority vote across parent factors → filing label. IR = annualised Information Ratio vs. benchmark; Shr = annualised Sharpe (cohort: minus RF; L/S: RF = 0); Exc = annualised excess return (%); Vol = annualised volatility (%); MDD = max drawdown (%); IC = Spearman(S, next-period return); "− − −" = IC undefined; Avg N = average tickers per period. Δ Shr = Two-step minus One-step Sharpe. **Bold** = best L/S Sharpe per model. Models include LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

		One-step (IC top-3 subfactors per ticker, direct MV)							**Two-step** (IC top-3 parents per ticker, all-subfactor 2-step MV)							
Model	Cohort	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	IR	Shr	Exc (%)	Vol (%)	MDD (%)	IC	Avg N	Δ Shr
LM	Neg	+0.48	+1.28	+7.8	16.2	−9.9	−0.230	7.9	+0.10	+1.23	+0.7	7.5	−12.2	−0.080	15.8	−0.05
	Neu	−1.00	+0.37	−16.7	16.7	−11.5	−0.020	14.2	−0.02	+1.39	−0.1	8.5	−10.2	+0.000	15.8	+1.02
	Pos	−0.96	+0.34	−10.0	10.5	−21.2	+0.060	7.9	−0.08	+0.57	−1.2	16.1	−21.3	−0.050	10.1	+0.23
	L/S	−1.41	−1.41	−24.5	17.4	−39.8	−0.110	16.2	−0.22	−0.22	−4.4	19.8	−27.1	−0.060	26.7	+1.19
GPT	Neg	+0.26	+1.29	+3.4	13.1	−13.5	−0.050	8.1	−1.28	+0.62	−9.6	7.5	−19.0	−0.040	12.5	−0.67
	Neu	−1.25	+0.31	−17.2	13.7	−15.1	+0.060	10.5	+0.77	+1.66	+6.8	8.8	−9.3	+0.010	11.2	+1.34
	Pos	−0.04	+0.79	−0.4	10.5	−13.5	−0.020	11.7	+0.34	+1.27	+2.1	6.1	−11.2	−0.040	17.4	+0.48
	L/S	−0.65	−0.65	−9.8	15.1	−22.6	−0.070	20.2	**+1.31**	+1.31	+11.7	8.9	−3.0	+0.030	29.9	+1.96
SFT	Neg	−0.67	+0.37	−8.9	13.3	−20.0	+0.080	6.1	−0.40	+0.77	−4.5	11.2	−19.8	−0.000	9.1	+0.40
	Neu	−0.59	+0.38	−13.3	22.4	−13.2	−0.200	3.9	−0.03	+0.84	−0.6	18.2	−16.2	+0.010	7.2	+0.47
	Pos	−0.36	+1.14	−3.5	9.9	−12.0	−0.010	15.2	−0.27	+1.30	−2.3	8.5	−9.5	−0.040	23.8	+0.16
	L/S	**+0.65**	+0.65	+10.9	16.9	−14.1	+0.020	22.1	+0.14	+0.14	+2.2	16.3	−9.0	+0.010	32.9	−0.51
DPO	Neg	−0.18	+0.65	−2.1	12.0	−17.4	+0.120	5.7	−0.62	+0.55	−7.0	11.2	−14.3	+0.100	7.6	−0.10
	Neu	−0.20	+1.08	−3.0	15.6	−9.3	−0.010	4.6	−0.33	+0.76	−4.0	12.1	−19.5	+0.010	7.5	−0.32
	Pos	−0.28	+0.96	−2.3	8.4	−13.0	−0.090	14.2	−0.05	+1.40	−0.3	6.1	−9.9	+0.010	24.1	+0.43
	L/S	−0.02	−0.02	−0.2	12.5	−8.8	−0.070	19.9	**+0.91**	+0.91	+10.4	11.3	−6.5	+0.010	33.0	+0.93
GRPO	Neg	−0.73	+0.60	−7.1	9.6	−17.9	+0.080	7.8	−0.59	+0.65	−5.0	8.4	−17.1	+0.120	7.2	+0.05
	Neu	−0.36	+0.64	−7.9	21.9	−13.1	−0.140	4.0	−1.04	+0.66	−10.7	10.3	−17.6	+0.100	7.1	+0.02
	Pos	−1.01	+0.76	−8.5	8.4	−15.0	−0.020	16.1	+0.40	+1.29	+2.2	5.6	−10.6	+0.030	25.7	+0.53
	L/S	+0.41	+0.41	+3.9	9.5	−5.6	−0.010	24.9	**+0.87**	+0.87	+9.0	10.4	−6.7	+0.010	34.2	+0.46
SFT_{TTC}	Neg	−0.67	+0.53	−7.5	11.1	−21.4	+0.000	7.3	−1.31	+0.17	−14.8	11.3	−18.7	−0.020	7.4	−0.36
	Neu	−1.36	−0.57	−43.2	31.8	−36.0	− − −	1.9	−1.17	+0.07	−21.4	18.3	−27.4	+0.050	7.1	+0.64
	Pos	−1.03	+0.82	−8.9	8.6	−12.3	+0.000	17.3	+1.29	+1.57	+6.2	4.8	−8.2	−0.030	25.6	+0.76
	L/S	+0.36	+0.36	+4.2	11.9	−7.9	+0.040	25.6	**+2.37**	+2.37	+21.9	9.2	−2.0	+0.050	34.3	+2.01
DPO_{TTC}	Neg	+0.68	+1.65	+10.3	15.3	−10.7	−0.060	5.2	+1.04	+1.54	+6.4	6.2	−10.2	+0.010	8.0	−0.11
	Neu	−1.23	−0.04	−23.9	19.5	−17.1	+0.050	4.1	+0.33	+1.13	+3.8	11.5	−17.8	+0.040	9.0	+1.16
	Pos	−1.05	+0.71	−10.4	9.9	−15.5	+0.100	18.4	−0.46	+1.14	−2.7	5.9	−11.5	+0.040	22.8	+0.43
	L/S	−1.52	−1.52	−20.8	13.7	−31.2	−0.020	23.5	−1.03	−1.03	−9.1	8.9	−12.1	+0.010	30.8	+0.49
GRPO_{TTC}	Neg	−1.31	+0.25	−12.4	9.4	−20.7	−0.163	6.6	−0.39	+0.64	−3.6	9.2	−17.5	+0.062	9.5	+0.39
	Neu	−0.80	+0.45	−13.2	16.5	−13.0	+0.135	4.6	−0.76	+0.56	−9.3	12.2	−19.7	+0.065	7.7	+0.11
	Pos	−0.28	+1.24	−2.5	8.6	−12.0	−0.088	17.8	+0.27	+1.43	+1.2	4.6	−8.3	+0.044	24.2	+0.20
	L/S	**+1.65**	+1.65	+15.5	9.4	−5.9	+0.026	25.4	+0.63	+0.63	+6.5	10.4	−5.0	+0.047	34.9	−1.02

Table 14: Average L/S IR across all eight models under each aggregation method and rebalancing frequency. *No IC* = all subfactors used (C.1, Tables 9 and 10). *IC* = top-3 subfactor / parent selection (C.2, Tables 12 and 13). IC Effect = IC − No IC.

	No IC (C.1)		**IC (C.2)**		**IC Effect**	
	Daily	Monthly	Daily	Monthly	Daily	Monthly
One-step	+0.65	−0.05	−0.06	−0.07	−0.71	−0.02
Two-step	+0.64	+0.01	**+0.86**	**+0.62**	+0.22	+0.61
Δ (Two−One)	−0.01	+0.06	+0.92	+0.69		

D Fama-French Five-Factor Regression

Tables 15 and 16 run Fama–French five-factor regressions on the Pos−Neg long-short spread for each of the 14 parent factors under daily and monthly rebalancing. The regression is

$$r_t = \alpha + \beta_{\text{MKT}}\,\text{MKT-RF}_t + \beta_{\text{SMB}}\,\text{SMB}_t + \beta_{\text{HML}}\,\text{HML}_t + \beta_{\text{RMW}}\,\text{RMW}_t + \beta_{\text{CMA}}\,\text{CMA}_t + \varepsilon_t,$$

with HAC standard errors using Newey–West lags of 21 for daily returns and 1 for monthly returns. Each cell reports annualised $\hat{\alpha}$ in percent, with the HAC t-statistic in brackets and $\hat{\beta}_{\text{MKT}}$ in italics.

Four findings stand out.

(1) The long-short spreads are approximately market-neutral.

This is expected for within-sector Pos−Neg portfolios. Some monthly regressions show larger market loadings, but these should be interpreted cautiously because the monthly test contains only 18 periods and some cohorts are thin.

(2) M&A/CorpDev alpha is real for LLMs but absent for the word-count baseline.

GPT and SFT_{TTC} earn positive monthly alpha on M&A/CorpDev after controlling for market, size, value, profitability, and investment factors. The LM baseline does not show a comparable alpha on the same factor, supporting the interpretation that the signal comes from contextual filing analysis rather than keyword polarity.

(3) Go-to-Market is the most model-dependent factor.

Some models produce positive adjusted alpha on Go-to-Market, while others produce large negative alpha. This mirrors the raw Sharpe results and confirms that Go-to-Market language is not a stable directional signal unless the model extracts the correct context.

(4) Two-step aggregation produces more reliable alpha after IC pre-selection. Tables 17 and 18 apply the same FF5 regression to IC-pre-selected cohort portfolios. Under two-step aggregation, several models deliver positive filing-level L/S alpha that survives factor adjustment, all with near-zero market beta. Under one-step aggregation, some models show large negative L/S alpha, particularly at monthly frequency, because the strict subfactor filter leaves too few filings in the Neg cohort to form a stable signal. The exception is GRPO_{TTC}, whose large prediction set means even the strict one-step filter retains enough data to produce significant positive L/S alpha — consistent with the Sharpe results in Tables 12–13.

Filing-Level Majority-Vote Cohort FF5 Regressions

Tables 17 and 18 run the same FF5 regression on the IC-pre-selected filing-level cohort portfolios from Tables 12–13, separately for one-step (Panel A) and two-step (Panel B) aggregation. Three observations stand out. First, the Neg, Neu, and Pos cohorts are long-only, so each carries substantial market exposure ($\hat{\beta}_{\text{MKT}} \approx 1.2$–$1.5$); the L/S spread cancels most of this out and is nearly market-neutral ($|\hat{\beta}_{\text{MKT}}| \leq 0.32$ across all models and panels). Second, under two-step daily rebalancing (Panel B), the Neg cohort earns negative alpha for most learned models (-6% to -16%), meaning negative-sentiment stocks underperform even after controlling for market, size, value, profitability, and investment factors. LM shows the opposite ($+1\%$), consistent with its tendency to mislabel sentiment. Third, GRPO_{TTC}'s one-step L/S alpha ($+16\%^{*}$ daily, $+20\%^{**}$ monthly) is the highest one-step result across all models and holds up after FF5 adjustment, ruling out common factor exposures as the source.

Table 15: **Factor-level.** FF5-adjusted annualised $\hat{\alpha}$ (%) for the per-factor Pos−Neg long-short spread, *daily* rebalancing (Jan 2024–Jun 2025). Rows = 14 investment factors. For each parent factor, subfactor labels are aggregated via majority vote to one factor-level label per (ticker, filing date); the resulting Pos−Neg equal-weight long-short spread is rebalanced daily (delay = 1 trading day) and regressed on the Fama–French five factors with HAC standard errors (lag = 21). RF = 0 (self-financing spread). $\hat{\beta}_{\text{MKT}}$ reported in italics. $^{\sim}p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. "—" = <30 overlapping observations. Models: LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

Factor	LM	GPT	SFT	DPO	GRPO	SFT_{TTC}	DPO_{TTC}	GRPO_{TTC}
	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]
AI Infra Cycle	+34	+46	+16	+8	−22	−16	−9	−21
	[+1.32] *−0.15*	[+1.48] *−0.09*	[+0.39] *−0.63*	[+0.29] *−0.42*	[−0.87] *−0.13*	[−0.56] *−0.12*	[−0.54] *−0.19*	[−0.66] *−0.48*
Capital Intensity	−3	−2	+18	+2	+25	+3	−2	−5
	[−0.12] *−0.25*	[−0.19] *−0.45*	[+1.50] *−0.34*	[+0.19] *−0.21*	[+1.62] *−0.45*	[+0.23] *−0.26*	[−0.21] *−0.26*	[−0.47] *−0.31*
Competitive Dynamics	−16	+0	+22	+18	+5	+9	+18	+15
	[−0.58] *+0.56*	[+0.01] *−0.00*	[+1.20] *−0.16*	[+0.48] *−0.61*	[+0.23] *−0.29*	[+0.43] *−0.48*	[+0.94] *−0.05*	[+0.72] *−0.44*
Data Moat	—	+17	—	—	+12	—	—	−51
		[+0.35] *+0.47*			[+0.26] *+0.48*			[−1.15] *−0.49*
Financial Quality	−2	$+12^{\sim}$	+0	+0	+10	−1	+1	+1
	[−0.19] *−0.01*	[+1.71] *+0.21*	[+0.01] *−0.04*	[−0.03] *+0.02*	[+1.30] *−0.11*	[−0.09] *−0.06*	[+0.09] *+0.18*	[+0.13] *+0.18*
Go-to-Market	-48^{**}	-51^{**}	$+43^{*}$	−10	−8	−9	$-40^{\sim}$	+14
	[−3.08] *+0.17*	[−2.69] *−0.24*	[+2.16] *+0.14*	[−0.59] *−0.04*	[−0.55] *+0.25*	[−0.58] *+0.09*	[−1.82] *−0.08*	[+1.03] *+0.24*
M&A / CorpDev	+8	$+17^{\sim}$	+9	+5	+5	+5	+3	+5
	[+0.77] *−0.05*	[+1.67] *+0.13*	[+1.14] *+0.01*	[+0.69] *+0.04*	[+0.68] *+0.16*	[+0.73] *+0.04*	[+0.43] *+0.16*	[+0.52] *+0.08*
Mgmt. Quality	+2	−20	+12	−2	−15	−11	−2	+12
	[+0.17] *+0.02*	[−0.98] *−0.09*	[+0.55] *−0.19*	[−0.09] *−0.18*	[−1.02] *−0.20*	[−0.61] *−0.10*	[−0.11] *−0.07*	[+0.79] *−0.28*
Pricing & Margin	−12	−16	−9	−16	−16	+1	−16	−13
	[−0.94] *+0.00*	[−0.96] *−0.44*	[−0.51] *−0.71*	[−1.07] *−0.50*	[−1.08] *−0.42*	[+0.04] *−0.30*	[−1.36] *−0.32*	[−0.83] *−0.52*
Product Platform	−21	−15	+37	—	−13	—	—	+13
	[−1.10] *+0.05*	[−0.99] *−0.26*	[+1.43] *−0.71*		[−0.66] *+0.12*			[+0.65] *−0.43*
Regulatory / Legal	−15	−32	−31	−16	+2	−7	−2	−16
	[−0.72] *+0.49*	[−1.06] *−0.76*	[−1.56] *+0.05*	[−1.04] *−0.02*	[+0.11] *+0.30*	[−0.46] *+0.21*	[−0.10] *+0.25*	[−0.81] *−0.15*
Security / Privacy	+16	−17	-59^{**}	-26^{*}	−36	-31^{*}	$-35^{\sim}$	−36
	[+0.75] *−0.08*	[−0.72] *−0.26*	[−2.64] *+0.38*	[−2.23] *+0.02*	[−1.45] *+0.04*	[−2.26] *+0.30*	[−1.92] *+0.52*	[−1.51] *+0.20*
Structural Risks	+17	+8	+12	+5	+6	+11	+2	+8
	[+1.21] *+0.20*	[+0.38] *−0.37*	[+0.34] *−0.31*	[+0.30] *−0.16*	[+0.31] *−0.24*	[+0.46] *−0.08*	[+0.11] *−0.46*	[+0.40] *−0.36*
Tech Infra	−13	−14	+4	+7	+10	+4	−4	+9
	[−1.37] *−0.12*	[−0.75] *−0.30*	[+0.20] *−0.29*	[+0.42] *−0.04*	[+0.60] *−0.32*	[+0.15] *−0.07*	[−0.22] *−0.26*	[+0.49] *−0.26*

Table 16: **Factor-level.** FF5-adjusted annualised $\hat{\alpha}$ (%) for the per-factor Pos−Neg long-short spread, *monthly* rebalancing (18 periods, Jan 2024–Jun 2025). Rows = 14 investment factors. For each parent factor, subfactor labels are aggregated via majority vote to one factor-level label per (ticker, filing date); the resulting Pos−Neg equal-weight long-short spread is rebalanced monthly (delay = 1 trading day) and regressed on the Fama–French five factors with HAC standard errors (lag = 1). RF = 0 (self-financing spread). $\hat{\beta}_{\text{MKT}}$ reported in italics. $^{\sim}p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. "—" = <8 overlapping observations. Models: LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

Factor	LM	GPT	SFT	DPO	GRPO	SFT_{TTC}	DPO_{TTC}	GRPO_{TTC}
	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]
AI Infra Cycle	+19	+26	—	—	−22	−12	−8	−22
	[+1.15] *+0.04*	[+1.39] *−0.27*			[−1.27] *−0.51*	[−0.82] *−1.17*	[−0.47] *−0.55*	[−1.45] *−1.51*
Capital Intensity	$+29^{*}$	−1	$+34^{**}$	+14	$+43^{**}$	+17	+6	−2
	[+2.09] *−1.31*	[−0.40] *−0.32*	[+2.60] *−1.14*	[+1.42] *−0.59*	[+2.72] *−1.07*	[+1.23] *−0.87*	[+0.77] *−0.72*	[−0.31] *−0.55*
Competitive Dynamics	-40^{***}	+14	+26	—	+10	+22	+10	$+36^{\sim}$
	[−5.13] *+0.30*	[+0.69] *−1.07*	[+1.39] *−0.34*		[+0.43] *−0.99*	[+1.14] *−0.92*	[+0.95] *+0.94*	[+1.67] *−0.63*
Data Moat	—	−40	—	—	−37	—	—	—
		[−1.01] *+1.20*			[−1.16] *+1.29*			
Financial Quality	+13	+12	+8	+10	+4	−9	−2	+9
	[+1.43] *−0.15*	[+1.52] *+0.10*	[+0.59] *−0.25*	[+0.86] *−0.49*	[+0.42] *−0.06*	[−0.80] *−0.27*	[−0.24] *−0.05*	[+0.94] *−0.24*
Go-to-Market	-60^{***}	-70^{***}	$+60^{***}$	−6	−7	−23	-53^{***}	$+26^{*}$
	[−6.48] *+0.16*	[−3.73] *+0.69*	[+3.67] *−0.59*	[−0.56] *−0.25*	[−0.38] *+0.45*	[−1.41] *+0.34*	[−4.32] *+0.74*	[+2.27] *−0.14*
M&A / CorpDev	+6	$+24^{*}$	+4	−1	−4	$+17^{\sim}$	+4	+2
	[+0.63] *−0.01*	[+2.17] *−0.07*	[+0.56] *+0.04*	[−0.10] *−0.22*	[−0.32] *−0.03*	[+1.87] *−0.36*	[+0.30] *−0.15*	[+0.19] *−0.32*
Mgmt. Quality	−2	−21	+16	+4	−12	+10	+3	$+26^{*}$
	[−0.16] *+0.03*	[−1.43] *+0.20*	[+0.79] *+0.75*	[+0.22] *+0.56*	[−0.77] *+0.98*	[+0.54] *−0.41*	[+0.27] *+0.21*	[+2.23] *−0.03*
Pricing & Margin	-24^{**}	−32	−27	−28	−33	−26	−30	−30
	[−2.62] *+0.17*	[−1.36] *+0.45*	[−1.17] *+0.27*	[−1.21] *+0.25*	[−1.55] *+0.63*	[−1.09] *+0.85*	[−1.62] *+0.39*	[−1.34] *+0.58*
Product Platform	-47^{**}	−29	+11	—	—	—	—	+16
	[−3.18] *+0.91*	[−1.47] *+0.51*	[+0.75] *+0.14*					[+1.17] *−0.25*
Regulatory / Legal	$+23^{\sim}$	—	−26	−10	+1	−7	−12	-29^{*}
	[+1.73] *−0.21*		[−0.70] *−0.55*	[−0.60] *+0.05*	[+0.07] *+0.57*	[−0.61] *+0.30*	[−0.66] *+0.71*	[−2.15] *+0.33*
Security / Privacy	$+43^{***}$	−25	-41^{***}	-34^{***}	−19	-29^{**}	-36^{*}	−15
	[+3.31] *+0.01*	[−1.52] *−0.52*	[−3.53] *−0.15*	[−4.61] *+0.24*	[−1.07] *−1.04*	[−3.07] *+0.03*	[−2.23] *+0.19*	[−0.95] *−0.27*
Structural Risks	$+18^{**}$	+22	+0	−2	+23	+41	+8	+27
	[+2.74] *−0.18*	[+0.83] *−0.21*	[−0.00] *+0.15*	[−0.06] *−0.11*	[+1.06] *−0.78*	[+1.52] *+0.23*	[+0.33] *−0.61*	[+1.06] *−0.73*
Tech Infra	$-20^{\sim}$	−18	+22	$+34^{*}$	+15	$+47^{**}$	+10	$+28^{\sim}$
	[−1.95] *+0.05*	[−1.07] *−0.59*	[+1.36] *−1.17*	[+2.18] *−1.19*	[+1.08] *−0.88*	[+2.70] *−1.06*	[+0.41] *−0.31*	[+1.91] *−1.17*

Table 17: **Company filing-level.** FF5-adjusted annualised $\hat{\alpha}$ (%) for IC-selected cohort portfolios (Neg / Neu / Pos / L/S), *daily* rebalancing (Jan 2024–Jun 2025). IC selection uses GPT teacher train+val (2015–2023) Spearman correlations applied uniformly to all models. **Panel A (One-step)**: top-3 subfactors per ticker selected by IC, direct majority vote → filing label. **Panel B (Two-step)**: top-3 parent factors per ticker selected by IC; all subfactors within those parents used; majority vote within each parent, then across parents → filing label. Dependent variable: cohort return minus RF (L/S: raw spread, RF = 0). HAC t-statistic in brackets (lag = 21); $\hat{\beta}_{\text{MKT}}$ in italics. $^{\sim}p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Models: LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

Cohort	LM	GPT	SFT	DPO	GRPO	SFT_{TTC}	DPO_{TTC}	GRPO_{TTC}
	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]
Panel A: One-step — IC top-3 subfactors per ticker, direct MV → filing label								
Neg	+8	+7	−10	−9	−3	−5	+13	−13
	[+0.84] *+1.33*	[+0.72] *+1.33*	[−0.90] *+1.48*	[−0.87] *+1.49*	[−0.51] *+1.34*	[−0.68] *+1.42*	[+1.07] *+1.27*	[−1.51] *+1.37*
Neu	−8	−11	−7	+11	−8	$-38^{\sim}$	−21	−9
	[−1.16] *+1.26*	[−1.26] *+1.25*	[−0.52] *+1.08*	[+0.96] *+1.24*	[−0.65] *+1.08*	[−1.65] *+1.07*	[−1.56] *+1.48*	[−0.77] *+1.38*
Pos	−7	+0	−0	−1	−5	−6	−3	+0
	[−0.65] *+1.34*	[+0.05] *+1.40*	[−0.07] *+1.25*	[−0.10] *+1.19*	[−0.68] *+1.29*	[−0.95] *+1.28*	[−0.46] *+1.27*	[+0.08] *+1.25*
L/S	−18	−10	+15	+8	−0	+0	−16	$+16^{*}$
	[−1.45] *+0.02*	[−0.98] *+0.07*	[+1.17] *−0.23*	[+0.87] *−0.32*	[−0.03] *−0.05*	[+0.00] *−0.13*	[−1.22] *+0.00*	[+2.02] *−0.12*
Panel B: Two-step — IC top-3 parent factors per ticker, all-subfactor 2-step MV → filing label								
Neg	+1	$-16^{\sim}$	−14	−7	−6	−15	+1	−8
	[+0.20] *+1.30*	[−1.78] *+1.42*	[−1.39] *+1.38*	[−0.66] *+1.38*	[−0.85] *+1.37*	[−1.48] *+1.34*	[+0.06] *+1.30*	[−0.82] *+1.43*
Neu	−2	+4	−3	−11	−10	−15	−7	−13
	[−0.27] *+1.24*	[+0.52] *+1.21*	[−0.26] *+1.31*	[−1.05] *+1.44*	[−1.35] *+1.25*	[−1.53] *+1.41*	[−0.77] *+1.42*	[−1.31] *+1.39*
Pos	−2	+5	+4	+1	+4	+5	+3	+4
	[−0.15] *+1.45*	[+0.69] *+1.27*	[+0.66] *+1.23*	[+0.12] *+1.20*	[+0.63] *+1.28*	[+0.88] *+1.23*	[+0.48] *+1.22*	[+0.85] *+1.22*
L/S	−5	$+21^{*}$	$+18^{\sim}$	+9	$+12^{\sim}$	$+23^{**}$	+2	$+13^{\sim}$
	[−0.47] *+0.14*	[+2.34] *−0.15*	[+1.82] *−0.16*	[+1.02] *−0.18*	[+1.71] *−0.10*	[+2.64] *−0.10*	[+0.22] *−0.07*	[+1.93] *−0.21*

Table 18: **Company filing-level.** FF5-adjusted annualised $\hat{\alpha}$ (%) for IC-selected cohort portfolios (Neg / Neu / Pos / L/S), *monthly* rebalancing (18 periods, Jan 2024–Jun 2025). IC selection uses GPT teacher train+val (2015–2023) Spearman correlations applied uniformly to all models. **Panel A (One-step)**: top-3 subfactors per ticker selected by IC, direct majority vote → filing label. **Panel B (Two-step)**: top-3 parent factors per ticker selected by IC; all subfactors within those parents used; majority vote within each parent, then across parents → filing label. Dependent variable: cohort return minus RF (L/S: raw spread, RF = 0). HAC t-statistic in brackets (lag = 1); $\hat{\beta}_{\text{MKT}}$ in italics. $^{\sim}p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Models: LM, GPT, SFT, DPO, GRPO, SFT_{TTC}, DPO_{TTC}, GRPO_{TTC}.

Cohort	LM	GPT	SFT	DPO	GRPO	SFT_{TTC}	DPO_{TTC}	GRPO_{TTC}
	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]	$\hat{\alpha}$% [t]
Panel A: One-step — IC top-3 subfactors per ticker, direct MV → filing label								
Neg	$+8$	$+13$	$-9^{\sim}$	-11	$+4$	$+7^{\sim}$	$+23^{*}$	-19^{*}
	[$+0.61$] *+1.06*	[$+1.46$] *+0.84*	[-1.84] *+1.46*	[-1.27] *+1.74*	[$+0.95$] *+0.88*	[$+1.77$] *+0.88*	[$+2.09$] *+0.84*	[-2.05] *+1.50*
Neu	-10	-8	-17	$+14$	-10	-14	-15	-2
	[-0.65] *+0.52*	[-0.78] *+0.70*	[-0.73] *+0.73*	[$+1.03$] *+0.40*	[-0.53] *+0.78*	[-1.01] *−0.48*	[-1.06] *+0.34*	[-0.13] *+0.59*
Pos	-13^{*}	-20^{*}	-5	-1	-15	-15	-10	-1
	[-2.37] *+1.40*	[-2.43] *+1.76*	[-0.76] *+0.99*	[-0.25] *+1.28*	[-1.63] *+1.30*	[-1.46] *+1.15*	[-1.13] *+0.99*	[-0.11] *+0.91*
L/S	$-24^{\sim}$	-36^{**}	$+6$	$+11$	-18^{***}	-20^{***}	-33^{***}	$+20^{**}$
	[-1.69] *−0.14*	[-3.14] *+0.52*	[$+1.14$] *−0.16*	[$+1.33$] *−0.46*	[-7.94] *+0.71*	[-7.37] *+0.57*	[-4.56] *+0.16*	[$+2.59$] *−0.30*
Panel B: Two-step — IC top-3 parent factors per ticker, all-subfactor 2-step MV → filing label								
Neg	-6	$-11^{\sim}$	-0	-2	-2	$-10^{\sim}$	$+6$	-4
	[-0.64] *+1.18*	[-1.86] *+1.18*	[-0.18] *+1.26*	[-0.18] *+0.70*	[-0.33] *+0.88*	[-1.86] *+0.85*	[$+0.76$] *+1.10*	[-0.45] *+1.22*
Neu	$+8$	$+0$	-7	-11	-5	$-18^{\sim}$	-2	-20
	[$+1.05$] *+0.68*	[$+0.01$] *+1.09*	[-0.35] *+0.72*	[-1.09] *+1.38*	[-0.40] *+0.95*	[-1.66] *+0.62*	[-0.14] *+1.25*	[-1.24] *+1.13*
Pos	-22^{*}	-2	-4	-3	-5	$+2$	-7	-4
	[-2.31] *+2.14*	[-0.31] *+1.22*	[-0.52] *+0.96*	[-0.41] *+1.05*	[-0.76] *+1.23*	[$+0.36$] *+1.24*	[-1.36] *+1.14*	[-1.07] *+1.13*
L/S	$-17^{\sim}$	$+9$	-4	$+0$	-2	$+13^{***}$	-12	$+0$
	[-1.85] *+0.82*	[$+0.99$] *+0.04*	[-0.39] *−0.30*	[$+0.04$] *+0.56*	[-0.37] *+0.50*	[$+3.61$] *+0.49*	[-1.55] *+0.05*	[$+0.06$] *+0.01*

E Factor Taxonomy

Table 19: Full factor taxonomy: 14 parent factors and 68 subfactors with extraction focus and record count across 2,230 10-K/10-Q filings for 62 S&P 500 IT-sector tickers (January 2015 – June 2025; train 2015–2022, val 2023, test Jan 2024–Jun 2025). Only records with a non-empty extracted quote are counted; grand total = 23,010.

Parent Factor	Subfactor	What It Extracts	N
tech_infra_platforms			**1,958**
	ecosystem_control	Control over silicon/infra/software/distribution layers	132
	compute_intensity	AI/datacenter demand driving workload growth	101
	platform_dependency	Dependency on hyperscalers, OEMs, or platform gatekeepers	115
	customer_concentration	Revenue concentration among large customers	400
	capital_allocation_signal	M&A vs. buyback vs. R&D capital deployment rationale	1210
ai_infrastructure_cycle			**217**
	ai_datacenter_buildout	Capex cycle exposure from AI datacenter build-outs	159
	inference_vs_training_mix	Shift between AI training vs. inference workloads	2
	supply_constraints_ai	HBM/packaging/substrate/foundry bottlenecks	21
	hyperscaler_concentration	Revenue exposure to large AI infrastructure buyers	25
	ai_monetization_realism	Real vs. pilot/narrative AI revenue	10
product_platform_evolution			**877**
	roadmap_conviction	Next-gen product cycle visibility and revenue contribution	43
	attach_rate_expansion	Cross-sell/suite expansion within installed base	30
	ecosystem_lockin	Switching costs, interoperability, and stickiness	74
	platform_unification	Integrated platform vs. point-solution strategy	35
	r_and_d_intensity	R&D spend intensity and innovation prioritization	695
pricing_and_margin_power			**684**
	infrastructure_pricing	Pricing power in mission-critical silicon/enterprise software	88
	silicon_asp_mix	ASP expansion from mix shift to premium silicon/tiers	10
	cost_curve_advantage	Process-node leadership, scale economics, software leverage	23
	gross_margin_mix	Margin expansion/compression from product mix or AI workloads	563
capital_intensity_and_cycle			**1,514**
	capex_signal	Capex trends and strategic purpose	606
	cyclicality_exposure	Semiconductor/enterprise IT/hardware spending cycle exposure	101
	inventory_normalization	Inventory levels, channel digestion, normalization	303
	working_capital_discipline	Working capital management and cash conversion	504
m_and_a_and_corpdev			**6,085**
	strategic_stack_expansion	Vertical/horizontal acquisitions within the tech stack	822
	ai_capability_acquisition	AI capabilities, proprietary data, or model development acquired	46
	integration_execution_risk	Integration risk, restructuring costs, operational friction	2761
	inorganic_growth_mix	Inorganic vs. organic growth proportion	746
	divestiture_focus	Portfolio pruning and strategic divestitures	1282
	synergy_realization	Expected or realized revenue/cost synergies	428
structural_risks_tech			**788**
	export_controls_silicon	Export controls and geopolitical restrictions on advanced chips	380
	open_source_disruption	Open-source competition threatening pricing or monetization	27
	platform_policy_risk	App store/cloud marketplace/ecosystem governance changes	1
	supply_chain_concentration	Geographic supplier concentration and mitigation	380
go_to_market_and_distribution			**1,004**
	sales_cycle_duration	Sales cycle length, deal timing, and procurement friction	334
	partner_channel_leverage	Channel/SI/MSP reliance, margin, and customer control	504
	marketplace_take_rates	Cloud marketplace contribution, take rates, and constraints	1
	public_sector_procurement	Government exposure, procurement cycles, compliance requirements	139

Table 19 (continued)

Parent Factor	Subfactor	What It Extracts	N
	sales_productivity_unit_econ	Rep ramp, quota attainment, CAC payback trends	26
competitive_dynamics_and_substitution			**582**
	competitive_displacement	Competitive wins/losses and displacement evidence	64
	bundle_vs_best_of_breed	Customer shift toward bundled suites vs. point tools	17
	hyperscaler_native_substitution	Competition from hyperscaler-native products	4
	price_competition_discounting	Competitive pricing pressure, discounting, deal term changes	488
	adjacent_entrants_new_modes	New entrants or architectural shifts threatening relevance	9
security_privacy_and_resilience			**226**
	breach_incident_history	Security incidents, remediation, and customer/financial impact	111
	uptime_outages_sla_risk	Outages, SLA performance, and reliability investments	19
	data_residency_and_privacy	Data residency, privacy compliance, and go-to-market constraints	55
	certifications_and_audits	SOC2/ISO/FedRAMP/PCI status, gaps, and costs	33
	supply_chain_security	SBOM, third-party vendor security, and mitigation strategies	8
regulatory_legal_and_antitrust			**955**
	antitrust_platform_scrutiny	Antitrust/platform-regulation exposure and business model risk	74
	ai_regulation_compliance	AI-specific regulation (safety, transparency) and compliance cost	3
	privacy_enforcement_risk	GDPR/CCPA enforcement exposure and risk management	131
	ip_litigation_patents	Patent posture, IP litigation risk, and royalty/licensing deps	687
	cross_border_data_trade	Cross-border data transfer rules and product/regional impact	60
data_moat_and_ecosystem_gravity			**175**
	data_rights_access	Proprietary data access, customer data rights, training constraints	13
	developer_adoption_metrics	API usage, SDK adoption, and developer community growth	47
	integration_density	Integration breadth/depth and its role in retention/expansion	75
	ecosystem_revenue_shares	Rev-share, fees, partner incentives, and tradeoffs	17
	data_feedback_loops	Telemetry/usage loops that improve model/product over time	23
financial_quality_and_balance_sheet			**6,799**
	liquidity_and_runway	Cash, revolver, and downside runway under stress	2711
	debt_maturity_interest_sensitivity	Debt maturities, refinancing risk, and rate sensitivity	1426
	stock_based_comp_dilution	SBC trends, dilution, and buyback offset	516
	cash_conversion_quality	CFO vs. NI, collections quality, and payment term changes	349
	contingent_liabilities	Legal, tax, warranty, and restructuring contingencies	1797
management_quality_and_incentives			**1,146**
	kpi_selection_and_changes	ARR/NRR/usage/backlog definitions and metric definition changes	323
	guidance_philosophy	Management guidance conservatism vs. optimism language	233
	incentive_alignment	Executive comp structure (growth vs. margin vs. FCF targets)	19
	governance_risk_controls	Internal controls, material weaknesses, and remediation	237
	execution_track_record	Delivery vs. prior commitments on products, integrations, costs	334

F Prompt Templates

Representative prompts for each pipeline stage are shown below. Placeholders in `{braces}` are filled at runtime.

Stage 1 — Evidence Extraction (GPT-5-nano).

```
System: You are a Financial Data Engineer. Extract original quotes from the provided
↪  source text that answer the specific questions in the target factor. For each
↪  answer you find, return the original_quote and a relevance_confidence from 0 to
↪  1. If the target factor questions are not answered by the filing text, return an
↪  empty extractions list. Output only valid JSON.
User: Ticker: {ticker} | Filing: {form} | Date: {report_date}
Text: {filing_markdown}
Target Factor Block: Factor: {factor_name}
Candidate Questions For This Factor:
- {question_key_1}
  {subfactor_question_1}
- ...
Output: {"metadata": {...}, "extractions": [{"factor": "...", "question_key": "...",
↪  "original_quote": "...", "relevance_confidence": 0.0}]}
```

Stage 2 — Sentiment Audit (GPT-5-mini teacher).

```
System: You are a Senior Equity Strategist. Audit the current evidence by comparing
↪  it against the historical context. Analyze the deviation from the baseline and
↪  sector context. Score the sentiment as one of: very_negative, negative, neutral,
↪  positive, very_positive. Provide a reasoning_chain, sentiment_label, and
↪  confidence_score. Output only valid JSON.
User: Factor: {factor_name}
Current evidence: {current_evidence}
Historical context: {historical_context}
Output: {"reasoning_chain": "...", "sentiment_label": "...", "confidence_score":
↪  0.0-1.0}
```

Stages 3–5 — SFT / DPO / GRPO Inference (Qwen3-8B). All three learned models share the same inference prompt; they differ only in their trained weights.

```
System: You are a Senior Equity Strategist specializing in factor-level sentiment
↪  analysis of SEC filings.
User: Ticker: {ticker} | Filing: {form} | Date: {report_date}
Factor: {factor_name}
Current evidence: {extracted_quotes}
Historical context: {historical_context}
Classify sentiment. Think step by step, then output JSON.
Output: <think>step-by-step reasoning...</think>{"sentiment_label": "...",
↪  "confidence_score": 0.0-1.0}
```

Stages 3–5 (TTC variant) — Majority-Vote Sampling. SFT_{TTC}, DPO_{TTC}, and GRPO_{TTC} use the same prompt as above with temperature $T{=}0.7$ and sample $K{=}3$ completions. The majority `sentiment_label` across completions is taken as the final prediction; ties are broken by highest mean `confidence_score`.

G Implementation Details

Table 20: Hyperparameters for SFT, DPO, and GRPO training.

Hyperparameter	SFT	DPO	GRPO
Base model	Qwen3-8B	Qwen3-8B	Qwen3-8B
LoRA rank / α	16 / 32	16 / 16	16 / 16
Max seq length	4096	4096	5200
Learning rate	2×10^{-4}	1.3×10^{-5}	5×10^{-6}
Epochs	3	1	1
Batch size (device)	12	4	8
Grad accum steps	6	6	6
β	–	0.05	0.02
Num generations	–	–	2
Precision	bf16	bf16	fp16
Optimizer	AdamW-8bit	AdamW-8bit	AdamW-8bit
Random seed	42	42	42

All three training stages (SFT, DPO, GRPO) use random seed 42 for weight initialisation and data shuffling, ensuring reproducibility. DPO is initialised from the SFT checkpoint; GRPO is initialised from the DPO checkpoint. The complete training, inference, and evaluation pipelines — together with all data preprocessing scripts — are available at the anonymous mirror linked in the Reproducibility statement.