THE WORLD AS A CURATOR: LEARNING FROM MARKET OUTCOMES FOR FINANCIAL REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Supervised fine tuning is widely used to adapt large language models to domain tasks, traditionally relying on high quality examples annotated by experts. Limited coverage of existing datasets and costly expert annotation make expanding high quality domain supervision a persistent challenge. Synthetic instruction and response pairs offer a scalable alternative, but internal judgments cannot independently verify agreement with actual outcomes when generation and selection share model priors without external evidence, while external feedback is often delayed, noisy, and confounded. To this end, we propose **W**orld **I**nformed **S**election with **E**nergy and **R**eliability (WISER), which selects responses for student fine tuning using historical market feedback and its reliability to improve future financial event reasoning. Specifically, we lock candidates before outcome access, train a World Feedback Energy model using nonoverlapping selection and confirmation windows to learn nonadditive interactions among candidate attributes, market compatibility, and feedback reliability, and select its configuration through constrained Bayesian optimization within 64 evaluations. Extensive results on 255 post freeze events across China, the United States, Europe, and Japan show gains of 0.024 in Entity NDCG@5, 0.027 in Direction MCC, and 0.10 in cost adjusted Sharpe over Direct WF Score, demonstrating that selected supervision improves future reasoning and investment performance using event time inputs alone.

1 INTRODUCTION

Supervised fine tuning (SFT) is a central step in adapting pretrained large language models (LLMs) to instruction following and downstream tasks Ouyang et al. (2022). Its effectiveness depends on training examples that connect task instructions with accurate responses and demonstrate the behavior expected when performing these tasks Zhou et al. (2023). In finance, these examples must capture numerical reasoning over financial reports and connect answers to relevant textual and tabular evidence Zhu et al. (2021). FinQA addresses this need through question answering pairs written by financial experts and accompanied by annotated reasoning programs Chen et al. (2021). PIXIU incorporates financial question answering datasets into instruction tuning corpora for adapting general language models to financial tasks Xie et al. (2023). However, the availability of such supervision is constrained by the scarcity of domain experts and the cost of obtaining their annotations Aguda et al. (2024). Limited coverage and diversity complicate dataset expansion, making the construction of high quality domain training data a persistent challenge Sun et al. (2023).

Recent study of synthetic supervision, exemplified by Self-Instruct Wang et al. (2023), shows how generated instruction and response pairs expand training data with less human annotation. Subsequent methods select examples using model scores Chen et al. (2024), self reflective uncertainty Liu et al. (2024a), gradient influence Xia et al. (2024), and target model likelihood Zhang et al. (2026). When generation and selection share model priors without external evidence, internal judgments Zheng et al. (2023); Xu et al. (2024) cannot independently verify agreement with subsequent outcomes, forming the curation loop in Figure 1a. Complementary approaches use answer correctness in STaR Zelikman et al. (2022), environmental feedback in Reflexion Shinn et al. (2023), and estimated effects on downstream metrics in NICE Wang et al. (2025). Financial research connects event information to subsequent returns, trading activity, and responses across firms Li et al. (2024); Koa et al. (2024); Xu et al. (2025). These outcomes arrive with delays and reflect market anticipation, common shocks, liquidity, and concurrent events, making their reliability vary across

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

events and observation windows. They provide relative supervision for comparing responses rather than semantic or causal ground truth, as illustrated in Figure 1b. Figure 1c frames the challenge as selecting fixed candidate responses using historical feedback and its reliability to improve students' reasoning about future events from event time inputs alone.

To this end, we propose **W**orld **I**nformed **S**election with **E**nergy and **R**eliability (WISER), which learns a reliability conditioned energy selector from historical market outcomes to curate supervision for future financial event reasoning. Specifically, we construct and lock K admissible responses from event time evidence using a frozen generator before accessing outcome tables. Early feedback provides market compatibility and observation reliability, while a later, nonoverlapping confirmation window supplies candidate preference targets. The World Feedback Energy model, or WFE, learns nonadditive interactions among candidate attributes, market compatibility, and early reliability from these targets. Training uses exact normalization over each candidate pool, with confirmation reliability weighting each event's supervision. Constrained Bayesian optimization selects capacity, regularization, reliability thresholds, and confirmation temperature within a fixed evaluation budget. Search enforces coverage, effective sample size, evidence completeness, and directional concentration constraints. The frozen selector retains one response per curation event for student SFT using only event time inputs and selected responses. Comparisons share student initialization and effective response token budget. Frozen before future evaluation, the student predicts from event time inputs alone, so historical feedback influences predictions solely through the selected training responses.

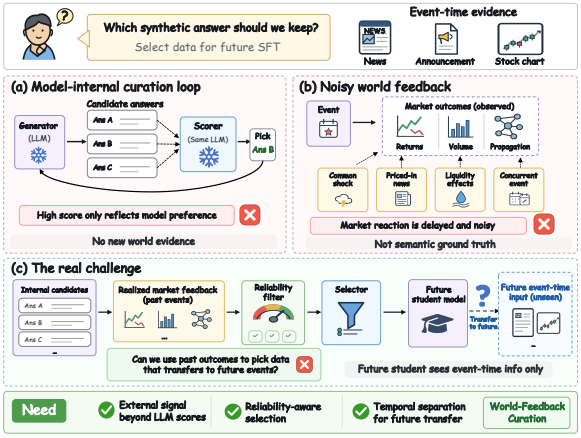


Figure 1. Challenges in world-feedback curation. (a) Model-internal scoring lacks external evidence. (b) Market feedback is delayed and noisy. (c) Reliable selection from historical outcomes must support future inference from event-time inputs alone.

Our main contributions are summarized as follows.

(1) We formulate curation under delayed, noisy world feedback through a protocol that locks candidates before feedback access, uses nonoverlapping selection and confirmation windows, and evaluates students on future events.
(2) We develop reliability-conditioned WFE and constrained BO to learn nonadditive interactions between candidates and outcomes and calibrate selection within 64 queries.
(3) Controlled comparisons, ablations, and transfer analyses support gains in Entity NDCG@5, Direction MCC, and cost-adjusted Sharpe across markets, temporal blocks, and student models without outcome access or selector execution at inference.

Extensive results on 255 post freeze events across China, the United States, Europe, and Japan show gains in Entity NDCG@5, Direction MCC, and cost adjusted Sharpe over matched baselines. Ablation and transfer analyses support reliability conditioned WFE and constrained BO, with gains across temporal blocks and student models that use only event time inputs.

2 Related Work

Synthetic Data Curation. Self-Instruct generates synthetic instructions and responses with rule based filtering Wang et al. (2023), while LIMA demonstrates effective instruction following from small curated datasets Zhou et al. (2023). Selectors use LLM scores Chen et al. (2024), joint quality, complexity, and diversity criteria Liu et al. (2024b), response length Zhao et al. (2024), self reflective uncertainty Liu et al. (2024a), and token information Fu et al. (2026). Others estimate gradient influence Xia et al. (2024), utility for in context learning Agarwal et al. (2025), or target model response likelihood Zhang et al. (2026). Self feedback refines responses Madaan et al. (2023); Yuan

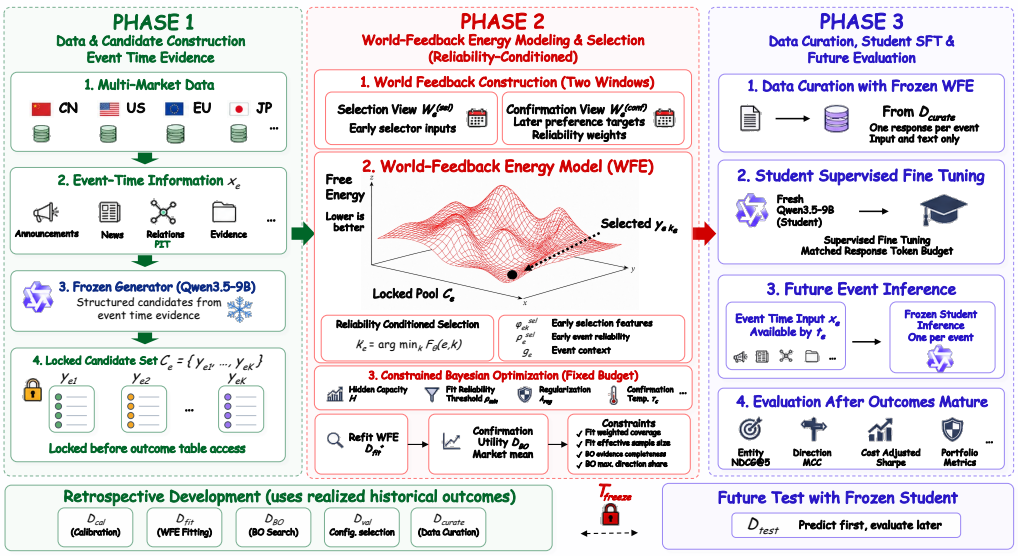


Figure 2: The framework of WISER. **Phase 1** constructs candidates from event time evidence and locks them before accessing outcome tables. **Phase 2** learns WFE from early feedback and later confirmation targets. Constrained Bayesian optimization selects its configuration. **Phase 3** fine tunes a fresh student on selected responses and evaluates its predictions after future outcomes mature.

et al. (2024). WISER isolates the value of historical outcomes under matched candidate pools, student initialization, and effective response token budgets.

Learning from Outcome Feedback. Outcome supervision filters reasoning by answer correctness Zelikman et al. (2022), supports reflection through environmental feedback Shinn et al. (2023), applies rule based corrections Gillman et al. (2024), or estimates example influence on nondifferentiable utility Wang et al. (2025). Data programming estimates errors and dependencies among labeling functions Ratner et al. (2016), while loss correction addresses label noise under explicit assumptions Natarajan et al. (2013). WISER studies downstream utility selection Wang et al. (2025) with delayed, noisy financial market feedback, locking candidates before outcome access, separating selection and confirmation windows, and conditioning comparisons on feedback reliability.

Financial Event Reasoning. Financial benchmarks document challenges in complex reasoning and prediction Xie et al. (2023; 2024). Event research models propagation using global and local news Ang & Lim (2022), company dependencies Luo et al. (2023), and lagged relations Li et al. (2024), while company graphs provide relational resources Cao et al. (2024). Market outcomes guide representation learning Guo & Hauptmann (2024), reflective explanation revision Koa et al. (2024), propagation alignment Xu et al. (2025), agent updates Yu et al. (2024), and research workflows Li et al. (2026). Building on energy models Belanger & McCallum (2016) and constrained Bayesian optimization Snoek et al. (2012); Gardner et al. (2014), WISER learns nonadditive interactions conditioned on reliability under matched training conditions and fixed query budgets. We assess selection gains through students evaluated on future events using event time inputs alone.

3 Methodology

3.1 Task and Overall Framework

World Informed Selection with Energy and Reliability, abbreviated as WISER, learns a response selection rule from historical market outcomes and uses selected responses for student supervised fine tuning, or SFT. For an event e released at t_e, x_e contains announcements, news, company relations, and traceable evidence available by t_e. Curation selects from a locked pool using the event's early market feedback. Confirmation outcomes from earlier development events supervise this rule, while future student evaluation tests the training value of selected responses.

Figure 2 organizes the pipeline into candidate construction, response selection, and student training. The World Feedback Energy model, or WFE, learns later preferences from early feedback; constrained Bayesian optimization, or BO, selects its configuration. The frozen selector then retains one student training response per curation event. Appendix E proves the properties claimed below and pairs each with its diagnostic experiment (Table 12).

The transformation state $\Gamma^\star = \mathrm{Calibrate}(D_{\mathrm{cal}})$ fixes normalization, component orientations, scales, and missing value codes across subsequent partitions. WFE parameters are estimated separately on the fitting partition. Retrospective development uses $D_{\mathrm{dev}}^{(\mathrm{retro})} = D_{\mathrm{cal}} \cup D_{\mathrm{fit}} \cup D_{\mathrm{BO}} \cup D_{\mathrm{val}} \cup D_{\mathrm{curate}}$. Let $T_e^{(\mathrm{used})}$ denote the latest availability time of development feedback actually used, and a_e the prespecified price timestamp starting the first scored test return.

$$D_{\mathrm{cal}} \prec D_{\mathrm{fit}} \prec D_{\mathrm{BO}} \prec D_{\mathrm{val}} \prec D_{\mathrm{curate}}, \qquad \max_{e \in D_{\mathrm{dev}}^{(\mathrm{retro})}} T_e^{(\mathrm{used})} < T_{\mathrm{freeze}} < \min_{e \in D_{\mathrm{test}}} t_e,$$
$$t_e \le T_e^{(\mathrm{output})} < a_e \quad \forall e \in D_{\mathrm{test}}. \tag{1}$$

Partition order requires each development partition's confirmation windows to mature before the next partition begins, and T_{freeze} marks completed development and student training. The fitting, search, and validation partitions provide supervision, configuration scores, and final configuration selection, respectively; the curation partition supplies student records. Test responses are committed after disclosure and before the scoring price timestamp a_e (Assumption E.30). Appendix B.3.1 distinguishes event date indices, feedback availability, and return scoring starts.

3.2 Event Time Candidate Construction

Event snapshots cover China, the United States, Europe, and Japan. A fixed Qwen3.5 9B revision G_ϑ generates $K_{\mathrm{raw}} \ge K$ ordered responses, $\widetilde{y}_{ej} = G_\vartheta(x_e; \zeta_{ej})$, using seeds ζ_{ej}. Each response identifies affected entities, directions, impact horizons, transmission mechanisms, supporting evidence, and confidence values. The deterministic rule A_{adm} checks prespecified schema, entity, and provenance requirements; the first K admissible responses form

$$\mathcal{C}_e = \mathrm{FirstK}([\widetilde{y}_{ej} \mid 1 \le j \le K_{\mathrm{raw}},\ A_{\mathrm{adm}}(x_e, \widetilde{y}_{ej}) = 1]). \tag{2}$$

The operator FirstK retains admissible responses in generation order, indexed as $y_{e1}, \ldots, y_{eK}$. Events with fewer than K admissible responses are excluded before outcome table access. An immutable manifest stores candidate texts and order, seeds, generator and retrieval revisions, evidence cutoffs, and admissibility decisions. Subsequent comparisons share these committed records. Proposition E.29 formalizes the lock: $\mathcal{C}_e$ is a deterministic function of event time inputs, so later selection changes only the retained index.

The same generator supplies internal scores $\boldsymbol{s}_{ek}^{(\mathrm{int})} = S_\vartheta(x_e, y_{ek})$ under fixed scoring prompts and decoding rules. The frozen map $\boldsymbol{b}_{ek} = B_{\Gamma^\star}(x_e, y_{ek}, \boldsymbol{s}_{ek}^{(\mathrm{int})})$ combines internal scores with structural and provenance descriptors, mean confidence, and mean horizon (Appendix B.2). Entity identities and citations are checked against the event time snapshot. The manifest also records internal scores, while market feedback adds observed compatibility to these model derived attributes. Replay controls outcome table access; historical knowledge in pretrained weights remains unverified.

3.3 Two Window World Feedback

World feedback provides candidate compatibility and event observation quality. The historical date index d_e locates the first tradable close after disclosure. The selection view $\mathcal{W}_e^{(\mathrm{sel})}$ covers $[d_e, d_e + 2]$, and the confirmation view $\mathcal{W}_e^{(\mathrm{conf})}$ covers $[d_e + 6, d_e + 20]$, separated by three trading days. Offsets follow each security's market calendar. Early feedback supplies selector inputs, and later feedback supplies development targets for the same candidates, with actual price endpoints recorded separately from date indices.

Expected returns use the preevent estimation interval $[d_e - 120, d_e - 21]$. For each view $v \in \{\mathrm{sel}, \mathrm{conf}\}$, the same structured predictions are compared with standardized abnormal returns over the fixed window. The vector $\boldsymbol{m}_{ek}^{(v)} = [M_{ek}^{(\mathrm{dir},v)}, M_{ek}^{(\mathrm{rank},v)}, M_{ek}^{(\mathrm{cover},v)}]$ measures signed-response

compatibility, cross sectional ranking agreement, and reaction-mass coverage by predicted entities. Evidence descriptors in $\boldsymbol{b}_{ek}$ characterize supporting material. Appendix E.2 specifies these measurements and their validity rules.

Observation reliability is computed over the event's security universe $\mathcal{U}_e$. Normalized components $s_{ie,c}^{(v)} \in [0,1]$ cover estimation completeness, trading status, corporate actions, concurrent events, source consistency, cross sectional resonance, and residual stability. Their geometric mean is $r_{ie}^{(v)} = [\prod_{c\in\mathcal{C}_{\mathrm{rel}}} s_{ie,c}^{(v)}]^{1/|\mathcal{C}_{\mathrm{rel}}|}$, where $\mathcal{C}_{\mathrm{rel}}$ indexes the components. With preevent liquidity weights $\nu_{ei} \geq 0$, event reliability is

$$\rho_e^{(v)} = \mathbb{I}[n_e^{(v)} \geq n_{\min}] \frac{\sum_{i\in\mathcal{U}_e} \nu_{ei} r_{ie}^{(v)}}{\sum_{i\in\mathcal{U}_e} \nu_{ei} + \epsilon}. \tag{3}$$

Here $n_e^{(v)}$ counts securities with valid feedback, $n_{\min}$ sets minimum coverage, and $\epsilon > 0$ stabilizes the denominator. Every candidate for an event shares the resulting reliability score. The selection representation is $\boldsymbol{\phi}_{ek}^{(\mathrm{sel})} = [\boldsymbol{b}_{ek}; \boldsymbol{m}_{ek}^{(\mathrm{sel})}; \rho_e^{(\mathrm{sel})}; \boldsymbol{\phi}_{ek}^{(\mathrm{miss,sel})}]$. Missingness indicators distinguish unavailable observations from neutral market reactions.

Confirmation utility is $c_{ek}^{(\mathrm{conf})} = \alpha_d M_{ek}^{(\mathrm{dir,conf})} + \alpha_r M_{ek}^{(\mathrm{rank,conf})} + \alpha_c M_{ek}^{(\mathrm{cover,conf})}$, with fixed nonnegative weights summing to one, so reliabilities, compatibility components, and $c_{ek}^{(\mathrm{conf})}$ all lie in $[0,1]$ (Proposition E.1). The primary setup uses $\alpha_d = \alpha_r = \alpha_c = 1/3$. Temperature $\tau_c > 0$ converts valid utilities into preferences within each event, with lower temperatures concentrating target mass on higher scoring candidates while preserving their order (Lemma E.9)

$$Q_e^{(\mathrm{conf})}(k) = \frac{\exp(c_{ek}^{(\mathrm{conf})}/\tau_c)}{\sum_{\ell=1}^{K} \exp(c_{e\ell}^{(\mathrm{conf})}/\tau_c)}. \tag{4}$$

Let $V_e = 1$ indicate an available confirmation view, sufficient security coverage, and valid finite confirmation utilities for all K candidates. Define $\Delta_e = \max_k c_{ek}^{(\mathrm{conf})} - \min_k c_{ek}^{(\mathrm{conf})}$ and $D_{\mathrm{fit}}^{+} = \{e \in D_{\mathrm{fit}} \mid V_e = 1, \rho_e^{(\mathrm{conf})} > 0, \Delta_e > 0\}$. Fully tied targets are excluded from preference fitting. This configuration-independent base set has fitting weights $\omega_e^{(\mathrm{conf})} = \rho_e^{(\mathrm{conf})} \mathbb{I}[\rho_e^{(\mathrm{conf})} \geq \rho_{\min}]$. Security reliability enters compatibility; early event reliability conditions WFE, and confirmation reliability weights supervision.

3.4 Reliability Conditioned Energy Selection

Reliability influences candidate ranking through interactions with candidate features. For fixed parameters, an additive term depending only on shared $\rho_e^{(\mathrm{sel})}$ cancels in candidate comparisons (Proposition E.4). WFE models these interactions through an energy selector Belanger & McCallum (2016). Context $\boldsymbol{g}_e$ encodes market, event type, source type, trading session, evidence count, universe size, and preevent liquidity. It determines event dependent biases $\boldsymbol{a}_e = \boldsymbol{a}_0 + A\boldsymbol{g}_e$ and $\boldsymbol{\eta}_e = \boldsymbol{\eta}_0 + B\boldsymbol{g}_e$.

For candidate features $\boldsymbol{\phi}_{ek}^{(\mathrm{sel})} \in \mathbb{R}^d$ and binary hidden variables $\boldsymbol{h} \in \{0,1\}^H$, the joint energy is $E_\theta(e,k,\boldsymbol{h}) = -\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek}^{(\mathrm{sel})} - \boldsymbol{h}^\top(\boldsymbol{\eta}_e + W^\top \boldsymbol{\phi}_{ek}^{(\mathrm{sel})})$. Summing its Boltzmann weights over hidden states and taking the negative logarithm gives the free energy (Proposition E.3)

$$F_\theta(e,k) = -\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek}^{(\mathrm{sel})} - \sum_{j=1}^{H} \mathrm{softplus}\Big(\eta_{e,j} + W_{\cdot j}^\top \boldsymbol{\phi}_{ek}^{(\mathrm{sel})}\Big). \tag{5}$$

The softplus function is $\mathrm{softplus}(u) = \log(1 + \exp u)$, and exact normalization over the committed pool gives $p_\theta(k \mid e) = \mathrm{softmax}_k[-F_\theta(e,\cdot)]$. The selector chooses $\widehat{k}_e = \arg\min_k F_\theta(e,k)$, resolving ties by committed order. Setting $H = 0$ gives a context dependent linear selector; positive capacity allows market compatibility contributions to vary with reliability: Proposition E.5 gives the pairwise reliability response, and Theorem E.6 shows that one hidden unit already reverses a ranking as reliability changes, a reversal that Corollary E.7 places outside the $H = 0$ family. Given features

and context biases, scoring costs $O(Kd(H+1))$. Appendix E.2 gives parameter dimensions, and Appendices E.3–E.4 prove marginalization and the interaction properties.

Configuration $\boldsymbol{\xi} = (H, \log \lambda_{\text{reg}}, \rho_{\min}, \log \tau_c)$ determines hidden capacity, regularization, the reliability threshold, and confirmation temperature. WFE minimizes the normalized reliability weighted listwise cross entropy

$$\mathcal{L}_{\text{WFE}}(\theta; \boldsymbol{\xi}) = -\frac{1}{\Omega_{\boldsymbol{\xi}}} \sum_{e \in D_{\text{fit}}^{+}} \omega_e^{(\text{conf})} \sum_{k=1}^{K} Q_e^{(\text{conf})}(k) \log p_\theta(k \mid e) + \lambda_{\text{reg}} \|\theta\|_2^2. \tag{6}$$

Fitting requires positive total supervision weight $\Omega_{\boldsymbol{\xi}} = \sum_{e \in D_{\text{fit}}^{+}} \omega_e^{(\text{conf})} > 0$ (Assumption E.2), and normalization expresses the data term as a reliability-weighted event average. The data term equals a constant plus the reliability weighted KL divergence from the confirmation targets (Proposition E.11). WFE fits later preference distributions from frozen early representations, while configuration search evaluates the selected response's raw confirmation utility. Theorem E.19 connects the two: the confirmation utility regret of the selected candidate is at most $\sqrt{2K\kappa_e(\theta)}/\tanh(1/(2\tau_c))$ for eventwise KL error $\kappa_e(\theta)$, and Theorem E.17 shows that the selected index survives any score perturbation below half the gap between the two lowest free energies. The candidate comparison in Appendix D connects response fields, features, confirmation preferences, and the final choice.

3.5 Constrained Selector Configuration

Each configuration refits WFE on D_{fit}^{+} for 1,000 full-batch Adam updates and records its selections on D_{BO}. Repeated fitting, discrete capacity choices, and top ranked decisions produce a costly, mixed, nonsmooth configuration objective. Constrained BO allocates a fixed evaluation budget Snoek et al. (2012); Gardner et al. (2014). Appendix B.3.1 fixes the partition roles for budgeted evaluation and one time shortlist selection.

Configuration evaluation uses fixed events and fixed weights. For each market m, $\mathcal{V}_{\text{BO}}^{m}$ and $\mathcal{V}_{\text{val}}^{m}$ contain events with valid confirmation targets and positive reliability, including fully tied targets. Every configuration uses $\bar{\omega}_e = \rho_e^{(\text{conf})}$, so identical response choices yield identical evaluation utility. Changing $\rho_{\min}$ affects fitting support while preserving the evaluation denominator. Fully tied events have the uniform target and the same utility under every selector (Proposition E.25), and two configurations differ in utility only through the reliability weighted share of events on which they select different candidates (Proposition E.26).

Fitting support constraints use D_{fit}^{+}, with coverage $R_{\text{cov}} = \sum_e \omega_e^{(\text{conf})} / \sum_e \rho_e^{(\text{conf})}$ and effective sample size $N_{\text{eff}} = (\sum_e \omega_e^{(\text{conf})})^2 / \sum_e (\omega_e^{(\text{conf})})^2$ (Proposition E.24 bounds both and caps each normalized event weight at $1/\sqrt{N_{\text{eff}}}$). On selected D_{BO} records, R_{prov} measures evidence completeness and R_{dom} measures the largest directional share. The market set and thresholds are fixed before opening D_{BO}; market evaluation weights are checked once and remain unchanged. Evaluated configurations are feasible when

$$\Xi_{\text{feas}} = \left\{ \boldsymbol{\xi} \in \Xi_{\text{eval}} \;\middle|\; \begin{array}{l} R_{\text{cov}}(\boldsymbol{\xi}) \geq \kappa_{\text{cov}},\ N_{\text{eff}}(\boldsymbol{\xi}) \geq \kappa_{\text{eff}},\ R_{\text{prov}}(\boldsymbol{\xi}) \geq \kappa_{\text{prov}}, \\ R_{\text{dom}}(\boldsymbol{\xi}) \leq \kappa_{\text{dom}},\ \sum_{e \in \mathcal{V}_{\text{BO}}^{m}} \bar{\omega}_e > 0 \ \ \forall m \in \mathcal{M} \end{array} \right\}. \tag{7}$$

BO computes reliability weighted confirmation utility within each market and then averages markets equally. The fixed reference population makes configuration comparisons reflect differences in response choices

$$J_{\text{BO}}(\boldsymbol{\xi}) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \frac{\sum_{e \in \mathcal{V}_{\text{BO}}^{m}} \bar{\omega}_e \, c_{e, \widehat{k}_e(\boldsymbol{\xi})}^{(\text{conf})}}{\sum_{e \in \mathcal{V}_{\text{BO}}^{m}} \bar{\omega}_e}. \tag{8}$$

The five highest scoring feasible configurations form a shortlist, evaluated once on D_{val} using the same aggregation rule. The selected configuration $\boldsymbol{\xi}^\star$ retains its original fitting parameters $\theta^\star$. Corollary E.27 bounds the gap between J_{BO} and the same population oracle by the BO set KL error, and Proposition E.28 shows that this rule returns a unique $\boldsymbol{\xi}^\star$. Both remain frozen while the selector chooses student training responses from D_{curate}.

3.6 Student Training and Future Evaluation

After each curation event's selection window matures, frozen WFE selects one committed response from its context and early representations. This yields $D_{\mathrm{raw}} = \{(x_e, y_{e\widehat{k}_e}) \mid e \in D_{\mathrm{curate}}\}$. Student records contain event time inputs and selected text, while market measurements and selector states remain in the curation records. Historical outcomes affect student supervision through the selected candidate index. Theorem E.32 makes this path a finite channel: given the manifest and training randomness, $\psi^\star$ is a deterministic function of the selected index vector, and outcome records carry at most $n \log_2 K$ bits into the student for $n = |D_{\mathrm{curate}}|$.

The operator $\mathrm{TokenBudget}_B$ returns the fixed-order sequence D_{SFT} of training occurrences, repeated as needed until B assistant loss evaluations, with the final occurrence masked at the boundary. A fresh student starts from $\psi_0 = \Theta^{(\mathrm{base})}_{\mathrm{Qwen3.5\ 9B}}$ with new training state. For occurrence (x_j, y_j) and unmasked positions $\mathcal{T}_{B,j}$, training minimizes

$$\mathcal{L}_{\mathrm{SFT}}(\psi) = \sum_{j=1}^{|D_{\mathrm{SFT}}|} \sum_{t \in \mathcal{T}_{B,j}} -\log f_\psi(y_{j,t} \mid x_j, y_{j,<t}). \tag{9}$$

Same pool comparisons share candidates and events, student initialization, packing order, optimizer protocol, and effective assistant token budget, so the curation rule enters training only through the selected index at each event (Corollary E.33). Response length determines how many records contribute unmasked positions, so record counts accompany token counts. Optimization produces the final student checkpoint $\psi^\star$.

After the final freeze, the student generates one response $\widehat{y}_e = f_{\psi^\star}(x_e)$ directly from event time inputs for each $e \in D_{\mathrm{test}}$.

Inputs, decoding settings, and responses are timestamped under the commitment boundary in Eq. (1), so each test response is committed before its first scored return (Proposition E.31). After outcomes mature, the frozen evaluator computes Entity NDCG@5, Direction MCC, and prespecified portfolio metrics including cost adjusted Sharpe. The pipeline remains fixed throughout evaluation, and future student predictions test the training value of historically selected responses. Appendix B.3.3 specifies training records; Appendix B.2 gives the budgets and evaluation rules.

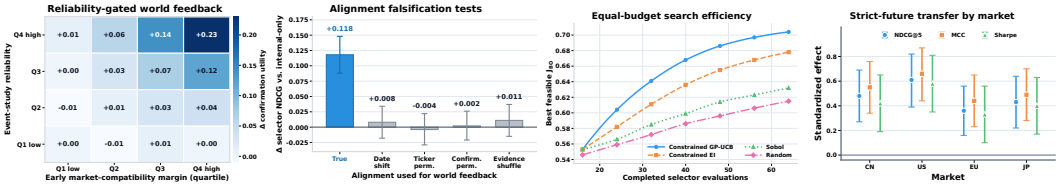
4 Experiments

The benchmark contains 1,690 multilingual financial events across China, the United States, Europe, and Japan, with 1,435 for retrospective development and 255 for testing. Chronological partitions separate calibration, fitting, search, validation, and curation, with prespecified intervals between stages. Panel A selectors retain one response per event from shared pools for 507 curation events. Qwen3.5-9B students share initialization, optimizer settings, effective assistant token budgets, and five paired seeds. Students and evaluation rules are frozen on 15 March 2026 before testing from 16 March to 31 May. Historical market feedback guides selection in WISER, while students use only event time inputs and commit predictions before scored returns begin. Panel B evaluates complete financial systems under the same testing protocol. Primary endpoints are Entity NDCG@5, Direction MCC, and annualized Sharpe after transaction costs of 10 basis points each way. The planned outcome cutoff is 9 July 2026. After outcome maturity, we report event paired 95% bootstrap intervals and Holm adjusted primary comparisons. For details, see Appendix B.2, B.3 .

Performance Analysis. Table 1 shows that WISER achieves the strongest entity ranking, direction prediction, and risk adjusted investment performance across both panels, reaching 0.630 NDCG@5, 0.335 MCC, and 2.88 Sharpe. Direct WF Score ranks second on all three primary endpoints, trailing by 0.024, 0.027, and 0.11, respectively, while Uniform-1 ranks last among the trained curation methods. Reliability conditioned WFE uses later confirmation supervision to learn nonadditive interactions among candidate content, early market compatibility, and feedback reliability, while constrained BO calibrates this selection rule within 64 evaluations. The gains under shared candidate pools and matched student training demonstrate the value of converting delayed and noisy world feedback into useful supervision, allowing students to learn from historical outcomes through selected responses and improve future event reasoning using event time inputs alone.

Method	Post-Freeze Event Reasoning		Cost-Adjusted Investment Utility			
	Entity NDCG@5 ↑	Direction MCC ↑	ARR (%) ↑	AVol (%) ↓	MDD (%) ↓	Sharpe ↑
Panel A: Same-Pool Data Curation						
Qwen3.5-9B Base	0.448 ± 0.016	0.121 ± 0.022	17.8 ± 1.7	15.10	15.8 ± 1.2	0.99 ± 0.10
Uniform-1	0.486 ± 0.014	0.168 ± 0.020	20.6 ± 1.5	11.28	13.7 ± 1.1	1.55 ± 0.09
AlpaGasus Chen et al. (2024)	0.502 ± 0.014	0.181 ± 0.020	21.4 ± 1.5	11.36	13.8 ± 1.1	1.63 ± 0.09
Long Is More Zhao et al. (2024)	0.509 ± 0.014	0.187 ± 0.020	21.8 ± 1.5	11.53	13.2 ± 1.1	1.62 ± 0.09
LESS WF Xia et al. (2024)	0.558 ± 0.013	0.244 ± 0.018	24.6 ± 1.4	11.68	11.3 ± 1.0	1.84 ± 0.08
SelectIT Liu et al. (2024a)	0.535 ± 0.013	0.216 ± 0.019	23.2 ± 1.4	11.61	12.5 ± 1.0	1.76 ± 0.08
DELIFT Agarwal et al. (2025)	0.555 ± 0.013	0.241 ± 0.018	24.4 ± 1.4	11.69	11.5 ± 1.0	1.85 ± 0.08
NICE WF Wang et al. (2025)	0.592 ± 0.012	0.287 ± 0.017	26.5 ± 1.3	11.46	10.9 ± 0.9	2.06 ± 0.07
GRAPE Zhang et al. (2026)	0.573 ± 0.012	0.258 ± 0.018	25.3 ± 1.3	11.61	10.7 ± 0.9	1.93 ± 0.08
T-SHIRT Fu et al. (2026)	0.566 ± 0.012	0.252 ± 0.018	25.0 ± 1.3	11.64	11.0 ± 0.9	1.88 ± 0.08
Direct WF Score	0.606 ± 0.011	0.308 ± 0.016	27.5 ± 1.2	8.78	8.3 ± 0.9	2.77 ± 0.07
WISER	**0.630** ± 0.010	**0.335** ± 0.015	**28.8** ± 1.1	**8.96**	**7.2** ± 0.8	**2.88** ± 0.06
Panel B: Common-Protocol Financial Systems						
FinRipple Xu et al. (2025)	0.584 ± 0.013	0.268 ± 0.019	25.6 ± 1.4	14.17	10.1 ± 0.9	1.62 ± 0.08
CausalStock Li et al. (2024)	0.548 ± 0.013	0.295 ± 0.018	27.2 ± 1.4	12.04	10.9 ± 1.0	2.02 ± 0.08
Newsflow Guo & Hauptmann (2024)	0.521 ± 0.014	0.232 ± 0.020	25.3 ± 1.7	10.62	10.5 ± 1.2	2.11 ± 0.10
SEP Koa et al. (2024)	0.536 ± 0.014	0.276 ± 0.019	25.9 ± 1.4	11.34	8.4 ± 0.9	2.03 ± 0.08
WISER	**0.630** ± 0.010	**0.335** ± 0.015	**28.8** ± 1.1	**8.96**	**7.2** ± 0.8	**2.88** ± 0.06

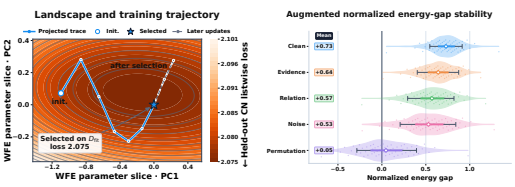
Table 1: Performance on 255 post-freeze events. **Panel A**: shared-pool curation with a base-model reference. **Panel B**: complete financial systems under the same protocol. Metrics: entity ranking (NDCG@5), direction prediction (MCC), annualized return (ARR) and volatility (AVol), maximum drawdown (MDD), and annualized Sharpe.



(a) Reliability conditioning (b) Alignment falsification (c) Fixed budget BO (d) Post freeze transfer

Figure 3: Mechanism and transfer analyses connecting reliable world feedback, WFE selection, constrained calibration, and post freeze student performance.

Figure 3 shows that feedback reliability and correct outcome alignment strengthen candidate selection, constrained GP UCB leads the fixed budget search, and student transfer effects remain positive across four markets. True alignment yields a 0.118 selector NDCG gain over the internal baseline, compared with at most 0.011 under perturbation controls, while standardized transfer effects are largest in the United States and smallest in Europe. WFE uses later confirmation supervision to learn how interactions between candidate content and early market compatibility vary with feedback reliability, while constrained BO calibrates the selector configuration within 64 evaluations. These results support learning response selection rules from historical market outcomes through reliability conditioned WFE and temporally separated confirmation supervision, improving future event reasoning by training students on selected responses while using only event time inputs at inference.



(a) WFE landscape (b) Energy-gap test

Figure 4: WFE optimization and gap correspondence on retrospective development data.

Robustness Analysis. Table 2 shows that WISER outperforms Direct WF Score on NDCG@5, MCC, and Sharpe across four markets, three temporal blocks, and three student models. The United States achieves the highest absolute market scores but the smallest paired gains, while Llama-3.1-8B has the smallest student gains yet retains improvements of 0.018, 0.021, and 0.07 on the three endpoints. WFE conditions candidate selection on event context and feedback reliability, market balanced BO calibrates selection utility with equal market weights,

and the frozen manifest carries the selected supervision into students through retained responses. We further examine candidate-pool size and supervised token-budget sensitivity in Appendix C.3. These gains support reliability aware curation for converting delayed, noisy historical outcomes into supervision that transfers across markets, temporal blocks, and students while preserving future inference from event time inputs alone.

Figure 4a shows that the projected WFE training trajectory reaches a low loss region on held out CN events. The selected point has a listwise loss of 2.075, while subsequent updates move toward higher loss contours. Reliability weighted listwise supervision aligns selection probabilities computed from early candidate representations with later confirmation preferences, reducing the influence of unreliable market observations on fitting. Nonlinear interactions condition candidate rankings on the reliability of feedback. We additionally audit selected reasoning for evidence support and unsupported mechanism claims in Appendix C.2. These results provide optimization evidence for WFE as a learned curation criterion that converts noisy historical feedback into supervision for training response selection through temporally separated confirmation preferences.

Condition	NDCG@5↑	MCC↑	Sharpe↑
Market Robustness			
CN	0.644 [Δ+0.027]	0.351 [Δ+0.031]	2.95 [Δ+0.11]
US	0.652 [Δ+0.019]	0.364 [Δ+0.021]	3.01 [Δ+0.07]
EU	0.621 [Δ+0.026]	0.325 [Δ+0.030]	2.66 [Δ+0.13]
JP	0.604 [Δ+0.024]	0.305 [Δ+0.026]	2.83 [Δ+0.09]
Post Freeze Temporal Blocks			
Early block	0.645 [Δ+0.028]	0.343 [Δ+0.031]	3.05 [Δ+0.12]
Middle block	0.623 [Δ+0.023]	0.338 [Δ+0.027]	2.75 [Δ+0.08]
Late block	0.621 [Δ+0.021]	0.324 [Δ+0.023]	2.86 [Δ+0.10]
Frozen Manifest Student Transfer			
Qwen3.5-9B	**0.630** [Δ+0.024]	**0.335** [Δ+0.027]	**2.88** [Δ+0.10]
Qwen3.5-27B	0.654 [Δ+0.020]	0.358 [Δ+0.023]	3.26 [Δ+0.08]
Llama-3.1-8B	0.608 [Δ+0.018]	0.312 [Δ+0.021]	1.98 [Δ+0.07]

Table 2: Robustness across markets, post-freeze blocks, and frozen-manifest student transfer. Brackets: paired margins over Direct WF Score.

Ablation Study. Table 3 shows that Full WISER achieves the highest mean performance on the retrospective selector diagnostic and all three student endpoints, with every ablation reducing performance. Removing world feedback produces the largest decreases of 0.118 in selector NDCG, 0.059 in student NDCG@5, 0.080 in MCC, and 0.97 in Sharpe, while the variant without BO ranks second on all four metrics. World feedback supplies external outcome supervision, reliability conditioning regulates its influence, WFE learns nonadditive interactions between candidate content and market feedback, and BO calibrates the selector configuration within a fixed budget. We further compare market scoring, selector structure, and configuration search in Appendix C.1. These ablations support combining external supervision, reliability aware selection, and constrained calibration to extract supervision from delayed, noisy historical outcomes through selected responses, improving future event reasoning.

Variant	Retrospective Sel. to Conf. NDCG ↑	Student Entity NDCG@5↑	Direction MCC ↑	Annualized Sharpe ↑
w/o World Feedback	0.624 ± 0.017	0.571 ± 0.012	0.255 ± 0.018	1.91 ± 0.08
w/o Reliability	0.684 ± 0.016	0.599 ± 0.011	0.292 ± 0.017	2.06 ± 0.07
w/o WFE and BO	0.671 ± 0.016	0.594 ± 0.012	0.294 ± 0.018	2.03 ± 0.08
w/o BO	0.718 ± 0.014	0.616 ± 0.011	0.319 ± 0.016	2.43 ± 0.07
w/o WFE	0.702 ± 0.015	0.610 ± 0.011	0.312 ± 0.017	2.30 ± 0.07
Full	**0.742** ± 0.013	**0.630** ± 0.010	0.335 ± 0.015	**2.88** ± 0.06

Table 3: Ablation study of key components with a retrospective selector diagnostic and post freeze student evaluation endpoints.

5 Conclusion

To summarize, this paper presents a novel approach WISER, that uses historical market feedback to select training responses for financial event reasoning. It combines candidate locking before outcome access, nonoverlapping feedback windows, reliability conditioned WFE, and constrained Bayesian optimization. Experiments on 255 post freeze events across four markets show gains in Entity NDCG@5, Direction MCC, and cost adjusted Sharpe. Ablations support the selection components, and transfer analyses show gains across students using only event time inputs.

One potential limitation of WISER is that supervision quality depends on a fixed candidate pool's evidence coverage and diversity and on market feedback potentially confounded by common shocks and concurrent events. Future work will diversify candidates before outcome access, incorporate independent evidence checks, and extend prospective evaluation across markets, event types, and market conditions, assessing whether gains persist and how feedback reliability evolves.

REFERENCES

Ishika Agarwal, Krishnateja Killamsetty, Lucian Popa, and Marina Danilevsky. Delift: Data efficient language model instruction fine-tuning. In *International Conference on Learning Representations*, volume 2025, pp. 100488–100509, 2025.

Toyin D Aguda, Suchetha Siddagangappa, Elena Kochkina, Simerjot Kaur, Dongsheng Wang, and Charese Smiley. Large language models as financial data annotators: A study on effectiveness and efficiency. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 10124–10145, 2024.

Gary Ang and Ee-Peng Lim. Guided attention multimodal multitask financial forecasting with inter-company relationships and global and local news. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6313–6326, 2022.

David Belanger and Andrew McCallum. Structured prediction energy networks. In *International Conference on Machine Learning*, pp. 983–992. PMLR, 2016.

Lele Cao, Vilhelm von Ehrenheim, Mark Granroth-Wilding, Richard Anselmo Stahl, Andrew McCornack, Armin Catovic, and Dhiana Deva Cavalcanti Rocha. Companykg: A large-scale heterogeneous graph for company similarity quantification. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4816–4827, 2024.

Lichang Chen, Shiyang Li, Jun Yan, Hai Wang, Kalpa Gunaratna, Vikas Yadav, Zheng Tang, Vijay Srinivasan, Tianyi Zhou, Heng Huang, et al. Alpagasus: Training a better alpaca with fewer data. In *International Conference on Learning Representations*, volume 2024, pp. 34767–34797, 2024.

Zhiyu Chen, Wenhu Chen, Charese Smiley, Sameena Shah, Iana Borova, Dylan Langdon, Reema Moussa, Matt Beane, Ting-Hao Huang, Bryan R Routledge, et al. Finqa: A dataset of numerical reasoning over financial data. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3697–3711, 2021.

Yanjun Fu, Faisal Hamman, and Sanghamitra Dutta. T-shirt: Token-selective hierarchical data selection for instruction tuning. *Advances in Neural Information Processing Systems*, 38:113932–113958, 2026.

Jacob R Gardner, Matt J Kusner, Zhixiang Eddie Xu, Kilian Q Weinberger, and John P Cunningham. Bayesian optimization with inequality constraints. In *Icml*, volume 2014, pp. 937–945, 2014.

Nate Gillman, Michael Freeman, Daksh Aggarwal, Chia-Hong Hsu, Calvin Luo, Yonglong Tian, and Chen Sun. Self-correcting self-consuming loops for generative model training. *arXiv preprint arXiv:2402.07087*, 2024.

Tian Guo and Emmanuel Hauptmann. Fine-tuning large language models for stock return prediction using newsflow. In *Proceedings of the 2024 conference on empirical methods in natural language processing: industry track*, pp. 1028–1045, 2024.

Kelvin JL Koa, Yunshan Ma, Ritchie Ng, and Tat-Seng Chua. Learning to generate explainable stock predictions using self-reflective large language models. In *Proceedings of the ACM web conference 2024*, pp. 4304–4315, 2024.

Shuqi Li, Yuebo Sun, Yuxin Lin, Xin Gao, Shuo Shang, and Rui Yan. Causalstock: Deep end-to-end causal discovery for news-driven multi-stock movement prediction. *Advances in neural information processing systems*, 37:47432–47454, 2024.

Yuante Li, Xu Yang, Xiao Yang, Xisen Wang, Weiqing Liu, and Jiang Bian. R&d-agent-quant: a multi-agent framework for data-centric factors and model joint optimization. *Advances in Neural Information Processing Systems*, 38, 2026.

Liangxin Liu, Xuebo Liu, Derek F Wong, Dongfang Li, Ziyi Wang, Baotian Hu, and Min Zhang. Selectit: Selective instruction tuning for llms via uncertainty-aware self-reflection. *Advances in Neural Information Processing Systems*, 37:97800–97825, 2024a.

Wei Liu, Weihao Zeng, Keqing He, Yong Jiang, and Junxian He. What makes good data for alignment? a comprehensive study of automatic data selection in instruction tuning. In *International Conference on Learning Representations*, volume 2024, pp. 22353–22373, 2024b.

Di Luo, Weiheng Liao, Shuqi Li, Xin Cheng, and Rui Yan. Causality-guided multi-memory interaction network for multivariate stock price movement prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12164–12176, 2023.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement with self-feedback. *Advances in neural information processing systems*, 36:46534–46594, 2023.

Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. *Advances in neural information processing systems*, 26, 2013.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744, 2022.

Alexander J Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. Data programming: Creating large training sets, quickly. *Advances in neural information processing systems*, 29, 2016.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36:8634–8652, 2023.

Jasper Snoek, Hugo Larochelle, and Ryan Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

Zhiqing Sun, Yikang Shen, Qinhong Zhou, Hongxin Zhang, Zhenfang Chen, David Cox, Yiming Yang, and Chuang Gan. Principle-driven self-alignment of language models from scratch with minimal human supervision. *Advances in Neural Information Processing Systems*, 36:2511–2565, 2023.

Jingtan Wang, Xiaoqiang Lin, Rui Qiao, Pang Wei Koh, Chuan-Sheng Foo, and Bryan Kian Hsiang Low. Nice data selection for instruction tuning in llms with non-differentiable evaluation metric. In *Forty-second International Conference on Machine Learning*, 2025.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)*, pp. 13484–13508, 2023.

Mengzhou Xia, Sadhika Malladi, Suchin Gururangan, Sanjeev Arora, and Danqi Chen. Less: Selecting influential data for targeted instruction tuning. *arXiv preprint arXiv:2402.04333*, 2024.

Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. Pixiu: A comprehensive benchmark, instruction dataset and large language model for finance. *Advances in Neural Information Processing Systems*, 36:33469–33484, 2023.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, et al. Finben: A holistic financial benchmark for large language models. *Advances in neural information processing systems*, 37:95716–95743, 2024.

Wenda Xu, Guanglei Zhu, Xuandong Zhao, Liangming Pan, Lei Li, and William Wang. Pride and prejudice: Llm amplifies self-bias in self-refinement. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15474–15492, 2024.

Yuanjian Xu, Jianing Hao, Kunsheng Tang, Jingnan Chen, Anxian Liu, Peng Liu, and Guang Zhang. Finripple: Aligning large language models with financial market for event ripple effect awareness. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 9377–9398, 2025.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W Suchow, Zhenyu Cui, Rong Liu, et al. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Advances in Neural Information Processing Systems*, 37:137010–137045, 2024.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Xian Li, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. Self-rewarding language models. *arXiv preprint arXiv:2401.10020*, 2024.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488, 2022.

Dylan Zhang, Qirun Dai, and Hao Peng. The best instruction-tuning data are those that fit. *Advances in Neural Information Processing Systems*, 38:141172–141208, 2026.

Hao Zhao, Maksym Andriushchenko, Francesco Croce, and Nicolas Flammarion. Long is more for alignment: A simple but tough-to-beat baseline for instruction fine-tuning. *arXiv preprint arXiv:2402.04833*, 2024.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36:46595–46623, 2023.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, et al. Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36:55006–55021, 2023.

Fengbin Zhu, Wenqiang Lei, Youcheng Huang, Chao Wang, Shuo Zhang, Jiancheng Lv, Fuli Feng, and Tat-Seng Chua. Tat-qa: A question answering benchmark on a hybrid of tabular and textual content in finance. In *Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 1: long papers)*, pp. 3277–3287, 2021.

A Statement on the Use of Large Language Models

We used a large language model (e.g., ChatGPT or Claude) solely to improve the grammar, phrasing, and clarity of the manuscript. The model did not generate technical content, analyses, or results, and all outputs were reviewed and verified by the authors.

B Baseline and Dataset Analysis

B.1 Baseline Adaptations

We adapt published curation mechanisms to our financial dataset as response selectors, retaining one of K locked candidates for each of 507 curation events. Students share initialization, inputs, packing, optimization, and effective assistant token budgets, with evaluation on 255 post freeze events.

Partitions retain calibration, fitting, search, validation, and curation roles, with reference confirmation windows matured before use. Evaluators, prompts, and seeds are frozen before curation. Curation confirmation outcomes remain unavailable for selection. Ties follow locked order, undefined scores rank last, and unscorable pools retain the first admissible response. Students restart from initialization and predict using only event time inputs.

Base uses student initialization. Uniform 1 uses seeded sampling. Direct WF Score maximizes a fixed weighted sum of early compatibility. Its shared measurements retain security-level reliability, while selection weights remain fixed. H0 fits a contextual linear selector with shared features and confirmation supervision, with search disabled by default or restored for 64 BO evaluations.

AlpaGasus adapts quality scoring Chen et al. (2024) with a frozen Qwen3.5 27B instruction evaluator and rubric. Greedy decoding rates each candidate from one to five for task fulfillment, evidence consistency, and unsupported claims without outcomes or other candidate ratings. Each event retains its highest rated response, preserving shared coverage.

Long Is More adapts response length selection Zhao et al. (2024), retaining each event's longest locked response under the fixed student tokenizer. Counts exclude input and prompt control tokens and precede shared packing and global token budget masking.

Both gradient baselines share probe warmup from student initialization for four epochs on outcome independent Uniform 1 responses from D_{fit} using LESS adapters. Epoch checkpoints contribute fixed random 8192 dimensional gradient projections, weighted by mean learning rates. Adam features retain preconditioned loss gradients before the descent minus sign. Confirmation temperature and reliability threshold follow H0 defaults.

LESS WF adapts optimizer aware gradient similarity Xia et al. (2024). Reference loss averages response token negative log likelihood, weights candidates by Q_e^{conf} and fitting events by normalized ω_e^{conf}. Checkpoint aggregated projected cosine similarities between reference gradients and candidate Adam loss gradients determine each event's highest scoring response.

SelectIT adapts multilevel scoring Liu et al. (2024a) using Qwen3.5 9B and 27B and Llama 3.1 8B instruction models. Five equivalent frozen prompts supply rating token probabilities from one to five. Token confidence adjustment, sentence uncertainty coefficient 0.2, and parameter proportional weighting determine each event's winner using only event time inputs and candidate text.

DELIFT adapts facility location Agarwal et al. (2025) using the frozen initial student. Utility is a demonstration's reduction in root mean squared distance between observed response token probabilities and a vector of ones. Negative and within event utilities become zero. Summed maximum utility from selected demonstrations over all curation candidates guides greedy marginal gain selection from uncovered events until each event is represented, continuing through zero gains.

NICE WF adapts gradient influence Wang et al. (2025) using expected confirmation compatibility under a softmax of mean response token log likelihoods, with LESS WF reliability weights across fitting events. Exact differentiation includes finite candidate normalization. Matching negative reward gradients to candidate Adam loss gradients favors reward increasing updates, and each event retains its highest scoring response.

GRAPE adapts the target distribution matching principle of Zhang et al., 2026 to locked candidates. The frozen initial student uses teacher forcing under the same event input to compute mean response token log likelihood, excluding prompt positions, and retains the highest scoring response within each event.

T-SHIRT Fu et al. (2026) uses the frozen student to compute absolute token log probability changes with and without instructions, retaining tokens above the calibration 25th percentile. We apply 30 embedding perturbations at noise scale 5, shortlist up to two candidates with the highest mean selective IFD in $(0, 1)$, and choose minimum neighborhood variance. Common fallbacks apply, and selected texts remain unchanged.

Text selectors use model signals. LESS WF and NICE WF access confirmation outcomes only from fitting events through frozen probe gradients, while Direct WF Score and WISER use early curation feedback. Shared token budgets control student training, with method specific curation costs. Panel B compares financial architectures adapted to our dataset at the system level.

Table 7 compares Uniform-1, Direct WF Score, H0 default, H0 with BO, WFE default, and Full WISER. H0 and WFE compare conditional linear and nonlinear selection under default and searched configurations. Defaults share regularization, reliability thresholds, and target temperature, with fixed positive capacity for WFE default. Learned selectors share pools, early features, confirmation targets, fitting events and budgets, fixed update schedules, optimization protocols, and paired seeds. BO variants also share configuration evaluation budgets.

B.2 Implementation detail

This section specifies the implementation used for the experiments in Section 4. Announcements are collected from CNINFO, SEC EDGAR, JPX TDnet, London Stock Exchange RNS, and Euronext company disclosures; Reuters news archives and LSEG Tick History supply timestamped news and prices. European coverage comprises London, Paris, and Amsterdam. Eligible events concern earnings, guidance, acquisitions, financing, material contracts, or regulatory decisions. Records are ordered by first publication time, grouped by issuer and underlying disclosure, and filtered sequentially for duplicate content, unresolved security identity, unavailable event-time evidence, temporal-boundary violations, and insufficient admissible candidates. Each operation exports its input count, retained count, excluded identifiers, and exclusion reasons. The event universe U_e contains the issuer, explicitly mentioned listed firms, and firms connected through ownership, customer, or supplier relations documented before disclosure; universes exceeding 50 securities retain explicitly mentioned firms first and then rank remaining firms by preevent median dollar turnover. Expected daily returns follow the OLS market model $r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}$, estimated over $[d_e - 120, d_e - 21]$ with at least 80 valid observations, using CSI 300, S&P 500, STOXX Europe 600, and TOPIX returns for the corresponding markets. Returns and benchmarks use the same currency and corporate-action convention. For a window of length L_v, standardized cumulative abnormal returns are $z_{ei}^{(v)} = \mathrm{CAR}_{ei}^{(v)} / \{\widehat{\sigma}_i [L_v + u_v^\top (X_i^\top X_i)^{-1} u_v]^{1/2}\}$, where X_i contains the estimation-period intercept and market returns, u_v sums the corresponding window regressors, and $\widehat{\sigma}_i^2 = \mathrm{RSS}_i/(n_i - 2)$. Selection and confirmation retain the published $[0, 2]$ and $[6, 20]$ offsets. Placebo tail scores compare each absolute standardized reaction with 999 calibration-period security-date observations matched by market, window length, and preevent-volatility quintile, using $(1 + \text{exceedance count})/1000$. Known issuer-event windows are excluded from the reference pool. The seven reliability components are estimation coverage $n_i/100$, tradable-session coverage, verified corporate-action adjustment, $1/(1 + n_{\mathrm{concurrent}})$, cross-source price agreement within one basis point, one minus the fraction of outside-universe securities with $|z| \geq 1.96$, and the ratio of smaller to larger residual variance across the two estimation halves. Missing components receive zero reliability and a missingness flag. Their geometric mean is aggregated using preevent 20-session median dollar turnover. Anchor settings are $n_{\min} = 5$, $\epsilon = 10^{-12}$, $r_{\min} = 0.5$, $q_0 = 0.10$, $s_0 = 1.96$, $s_{\mathrm{scale}} = 2$, $s_{\max} = 5$, and $L = 5$.

The generator, internal scorer, and student initialization use Qwen/Qwen3.5-9B at revision c202236235762e1c871ad0ccb60c8ee5ba337b9a, including its tokenizer and chat template. Generation uses temperature 0.8, top-p 0.95, and at most 1,024 new tokens, retaining the first $K = 8$ admissible responses among at most 32 attempts. The generation instruction is "Using only the supplied event-time evidence, return affected security identifiers, directions, horizons, transmission

mechanisms, evidence identifiers, and confidence values as JSON." The scoring instruction is "Rate this response from 1 to 5 for task fulfillment, consistency with the supplied evidence, and support for its claims; return one integer." Scoring uses greedy decoding and no market outcomes. WFE retains the original softplus free energy. Its ten numerical inputs comprise normalized internal rating, six-field completeness, valid-citation fraction, entity coverage, mean confidence, mean horizon clipped at 20 days, three early compatibility components, and early reliability; ten missingness indicators give $d = 20$. Context encodes market, event type, disclosure session, evidence count, universe size, and preevent liquidity, with transformations fitted on D_{cal}. WFE uses full-batch Adam for 1,000 updates, learning rate 10^{-3}, coefficients $(0.9, 0.999)$, and gradient clipping at 1. Search considers $H \in \{8, 16, 32, 64\}$, $\lambda_{\mathrm{reg}} \in [10^{-6}, 10^{-2}]$, $\rho_{\min} \in [0, 0.6]$, and $\tau_c \in [0.05, 1]$, with logarithmic coordinates for regularization and temperature. Feasibility requires fitting weighted coverage of at least 0.80, fitting ESS of at least 100, selected-response provenance completeness of at least 0.95, and directional concentration no greater than 0.80. Sixteen Sobol evaluations precede 48 constrained GP-UCB queries. The five highest-utility feasible configurations receive one validation evaluation, and the validation winner defines $\xi^\star$ while retaining its fitting checkpoint. Students use rank-16 LoRA with scaling 32 and zero dropout on text-decoder linear layers, with the remaining weights frozen. AdamW uses learning rate 2×10^{-4}, coefficients $(0.9, 0.95)$, weight decay 0.01, five warmup updates, cosine decay, BF16, and gradient clipping at 1. Each of 100 updates accumulates exactly 2,596 assistant loss tokens, giving 259,600 tokens per run; prompt positions are masked. Records follow a fixed event-identifier order, repeated when necessary, with the final loss segment masked at the budget boundary. Seeds are 11, 23, 37, 51, and 71. Each run exports the selected indices, loss-contributing event count, complete and partial response counts, token-length quantiles, market token shares, and cumulative token exposure.

Evaluation uses a fixed event-security population established before scoring. Entity relevance is the confirmation-window reaction mass $g_{ei} = r_{ie}\min(|z_{ei}|, 5)1[i \in A_e]$, and NDCG@5 uses linear gain, logarithmic rank discount, and an ideal ranking over the full evaluable universe. Predictions are ranked by confidence; duplicate entities retain their first occurrence, invalid identifiers receive zero gain, and omitted entities receive zero prediction score. Events with zero ideal gain are excluded from NDCG and counted separately. Direction MCC pools all evaluable event-security pairs, with labels -1, 0, and $+1$ determined by standardized reactions below -1.96, within $[-1.96, 1.96]$, and above 1.96; omitted predictions map to zero, while unavailable outcomes remain outside the fixed evaluable population. Student inference is greedy and uses the same generation instruction and output limit. The reference portfolio is long-only: each event allocates equally across its five highest-confidence positive predictions, with unused allocation held in cash. Positions enter at the first regular-session opening at least five minutes after prediction commitment and expire after 20 local trading sessions. Concurrent events receive equal allocations within each market, duplicate security positions are aggregated, and each market receives one quarter of total capital. Weights are rebalanced at the next local opening, with transaction costs of 10 basis points applied to traded notional on each side. Portfolio accounting uses USD-valued holdings and cash, includes inactive periods, and records execution prices, turnover, overlapping positions, and daily net asset value. With daily net returns r_t, annualization uses $\mathrm{ARR} = 252\bar{r}$, $\mathrm{AVol} = \sqrt{252}\, s(r)$, and $\mathrm{Sharpe} = \sqrt{252}(\bar{r} - 0.03/252)/s(r)$; maximum drawdown is calculated from the cumulative net asset value path. Metrics are calculated separately for each run before aggregation.

B.3 Dataset and Evaluation Protocol

B.3.1 Chronological Partition and Purge Protocol

Table 4 divides 1,690 retained events across four markets into 1,435 development events and 255 test events. Five disjoint development partitions support calibration, WFE fitting, configuration search, validation, and curation. Each of the 507 curation events contributes one selected response for student training. The test partition evaluates student generalization to subsequent events.

Development uses retrospective replay with fixed generator, tokenizer, retrieval, generation code, and evidence filter revisions. Candidate manifests are committed before outcome tables are opened. The pretrained generator's prior knowledge of historical events remains unverified. Following Table 5, the final configuration, WFE parameters, student checkpoint, and evaluation rules are frozen on 15 March 2026. Test events span 16 March to 31 May, with inputs and responses timestamped and committed before scored return windows begin.

Split	Share	Total	CN	US	EU	JP	Role
D_{cal}	10.00%	169	42	41	42	44	Transformation and reliability calibration
D_{fit}	19.94%	337	84	81	84	88	WFE fitting
D_{BO}	14.97%	253	63	61	63	66	Fixed budget configuration evaluation
D_{val}	10.00%	169	42	41	42	44	One time shortlist selection
D_{curate}	30.00%	507	127	122	126	132	Response selection and student training
D_{test}	15.09%	255	64	61	64	66	Evaluation after model freeze
Total	**100.00%**	**1,690**	**422**	**407**	**421**	**440**	Full benchmark

Table 4: Event counts and market composition across development and test partitions. Shares are calculated from the listed total of 1,690 events. Date boundaries are specified in Table 5.

Split	Event date range	Following interval
D_{cal}	2024-05-01 to 2024-06-30	2024-07-01 to 2024-08-11
D_{fit}	2024-08-12 to 2024-11-30	2024-12-01 to 2025-01-12
D_{BO}	2025-01-13 to 2025-03-31	2025-04-01 to 2025-05-06
D_{val}	2025-05-07 to 2025-06-30	2025-07-01 to 2025-08-10
D_{curate}	2025-08-11 to 2026-01-31	2026-02-01 to 2026-03-15
D_{test}	2026-03-16 to 2026-05-31	2026-06-01 to 2026-07-09
Final model freeze		2026-03-15
Planned test outcome cutoff		2026-07-09

Table 5: Event date ranges and subsequent intervals. The first five following intervals are embargoes requiring at least 21 trading days per venue, with confirmation feedback availability checked separately. The final interval is reserved for collecting test outcomes. Student parameters are fixed before testing, and test responses are committed before their scored return windows.

Table 5 defines chronological partitions from 1 May 2024, separated by at least 21 trading days under each trading venue's calendar. Curation ends on 31 January 2026, followed by an embargo from 1 February through 15 March 2026. This embargo contains 24, 29, and 28 trading days under the SSE, NYSE, and JPX calendars, respectively. Confirmation feedback from each development partition must be available before the next partition begins, accounting for data publication delays. Curation applies the frozen selector once selection feedback is available. After training, the student model parameters are fixed on 15 March 2026 and remain unchanged throughout testing.

The purge audit groups records of the same disclosure, issuer action, or source linked event within one partition and excludes violations of prespecified temporal boundaries. The split manifest records event times, information cutoffs, response locks, feedback availability, and exclusion reasons. Test eligibility and evaluation rules are fixed before testing, with test events spanning 16 March through 31 May 2026. Each student response must be committed before its first scored return window, as required by Eq. 1. The following interval from 1 June through 9 July 2026 is reserved for completing test outcome collection. Availability checks cover market specific $d_e + 20$ confirmation endpoints and any longer evaluation windows. Coverage is reported against the fixed test population, and final scores are reported only when the required outcomes are complete.

B.3.2 Event Study Windows

For each event linked security, d_e is the first tradable close after disclosure. Intraday disclosures use the same trading day, while after close, weekend, and market closure disclosures use the next. Offsets follow the security's market calendar.

The date index d_e identifies the first tradable close after disclosure; releases after the close map to the next trading day. Test scoring uses separately recorded price timestamps. For security i, the return is $P_i(b_{ei})/P_i(a_{ei}) - 1$, with $a_e = \min_i a_{ei}$. Responses must precede a_e, and the opening, quoted, or transaction price rule is fixed before testing. Late outputs retain their original deadline.

Table 6 separates estimation, immediate reaction, and later market response. Selection spans $[d_e, d_e + 2]$, followed by a three trading day gap and 15 trading day confirmation at $[d_e + 6, d_e + 20]$.

Window	Trading day interval	Purpose
Normal return estimation	$[d_e - 120, d_e - 21]$	Estimate expected returns from about 100 observations
Selection	$[d_e, d_e + 2]$	Measure the immediate market reaction
Event gap	$[d_e + 3, d_e + 5]$	Reduce mechanical dependence between views
Confirmation	$[d_e + 6, d_e + 20]$	Measure later abnormal returns over 15 trading days
Split embargo	21 trading days	Required minimum; verify against each market calendar

Table 6: Prespecified market windows for the primary analysis.

Pre-event estimation supplies expected returns, and the gap reduces dependence between views. Early compatibility thus predicts later candidate ordering through distinct measurements.

B.3.3 Student Transfer and Information Audit

After curation selection windows mature, frozen WFE uses $\phi_{ek}^{(\mathrm{sel})}$ and event context to select $(x_e, y_{e\widehat{k}_e})$, with event identifiers stored separately for audit. Prices, abnormal returns, reliability, energies, selector probabilities, confirmation outcomes, BO observations, and configuration states are removed before training. World feedback shapes supervision through response selection.

Every method initializes a fresh student from the same frozen Qwen3.5-9B base revision, sharing the optimizer protocol, event coverage target, update count, and effective assistant token budget. Fixed order packing masks assistant loss positions beyond budget B. Same pool comparisons share candidates and event identities, isolating response selection under common initialization and supervision budgets.

Student comparisons share candidate pools, base revision, optimizer protocol, and effective assistant token budget. Fixed-order training counts each assistant loss evaluation toward B, including repeated records; the final occurrence is masked at the boundary. Selected records retain event IDs, original text, and hashes, alongside counts of events and tokens contributing supervision. The freeze binds the student checkpoint, decoding settings, evaluation prices, and cost rules. Future evaluation uses returns beginning after response commitment and waits for the corresponding outcome windows to mature.

Test inference uses only event inputs and the frozen student. Input snapshots, decoding settings, and outputs are hashed and timestamped within the required boundary. After maturity, the frozen evaluator computes prespecified entity, direction, portfolio metrics, and paired statistics. Training, selection, thresholds, windows, and portfolio rules remain frozen throughout testing.

C Supplementary Experiments

The supplementary experiments evaluate how historical market feedback improves response selection, reasoning quality, market transfer, and training efficiency. Section B.3 specifies the data partitions, event windows, and student training protocol.

C.1 Market Scoring, Nonlinear Selection, and Configuration Search

Direct WF Score applies a fixed market compatibility rule. H0 default is the joint ablation without WFE and BO, with a learned conditional linear selector and zero hidden capacity. H0 with BO retains configuration search for that linear selector, while WFE default retains nonlinear selection with a default configuration. Full WISER combines nonlinear selection and configuration search.

All methods share candidate pools, event identities, selected response counts, the Qwen3.5-9B initialization, student optimizer, update count, effective assistant token budget, and five paired seeds. H0 default and WFE default share preset regularization, reliability thresholds, and target temperature, with a fixed positive hidden capacity for WFE. Each BO run uses 16 Sobol initial evalua-

tions and 48 constrained GP UCB queries, with zero hidden capacity for H0 and searched capacity for Full WISER.

Method	Selector	Config.	Retrospective Sel. NDCG ↑	Student Entity NDCG@5 ↑	Direction MCC ↑	Sharpe ↑
Panel A Reference selectors						
Uniform-1	Random	Fixed	0.512	0.486	0.168	1.56
Direct WF Score	Fixed score	Fixed	0.665	0.606	0.308	2.78
Panel B Selector structure and configuration search						
H0 default	Linear	Default	0.671	0.594	0.294	2.03
H0 with BO	Linear	BO 64	0.702	0.610	0.312	2.30
WFE default	Nonlinear	Default	0.718	0.616	0.319	2.43
WISER	**Nonlinear**	**BO 64**	**0.742**	**0.630**	**0.335**	**2.88**

Table 7: Selector and student results under the main evaluation protocol. Panel A compares reference selectors, and Panel B separates nonlinear selection from configuration search. Selector NDCG uses retrospective data, while student metrics use the 255 test events.

Table 7 shows that WISER achieves the strongest retrospective selector diagnostic and student performance, reaching 0.742 selector NDCG, 0.630 student NDCG@5, 0.335 MCC, and 2.88 Sharpe. WFE default ranks second on selector NDCG and both student reasoning metrics, while Direct WF Score achieves the second highest Sharpe at 2.78. The gains from replacing H0 with WFE and subsequently adding BO support complementary roles for nonlinear modeling of candidate and feedback interactions and configuration calibration within the fixed 64 evaluation budget. This comparison demonstrates the value of learning and calibrating outcome guided selection under a shared student training protocol, turning delayed and noisy historical feedback into supervision that improves future event reasoning.

Comparison	Δ NDCG@5 ↑	Δ MCC ↑	Δ Sharpe ↑
Fixed market scoring			
Direct WF Score over **Uniform-1**	+0.120	+0.140	+1.22
Nonlinear selection			
WFE default over **H0 default**	+0.022	+0.025	+0.40
Full WISER over **H0 with BO**	+0.020	+0.023	+0.58
Configuration search			
H0 with BO over **H0 default**	+0.016	+0.018	+0.27
WISER over **WFE default**	+0.014	+0.016	+0.45
Overall improvement			
WISER over **H0 default**	+0.036	+0.041	+0.85
WISER over Direct WF Score	**+0.024**	**+0.027**	**+0.10**

Table 8: Absolute differences between the means in Table 7. Each row subtracts the second method from the first, separating fixed scoring, nonlinear selection, and configuration search under shared student training.

Table 8 shows positive mean gains in student NDCG@5, MCC, and Sharpe from fixed market scoring, nonlinear selection, and configuration search under shared student training. Direct WF Score over Uniform-1 produces the largest gains of 0.120, 0.140, and 1.22 on the three endpoints, while BO improves Sharpe by 0.45 for WFE compared with 0.27 for H0. Fixed scoring supplies outcome based candidate preferences, WFE learns reliability conditioned interactions, and constrained BO calibrates selector capacity and supervision settings within 64 evaluations. The additional gains over Direct WF Score demonstrate the value of learned selection and constrained calibration for turning noisy historical outcomes into useful supervision through selected training responses, improving future event reasoning.

C.2 Semantic Audit of Selected Reasoning

Response source	Evidence quality: Entailed evidence ↑	Evidence quality: Unsupported mechanism ↓	Agreement: Alpha	Reasoning audit: Correct and unsupported ↓	Reasoning audit: WISER preference ↑
Internal	68.2 59.3 to 76.4	37.6 31.1 to 43.9	0.731	23.3 14.2 to 33.3	71.2 64.0 to 77.6
Direct WF	74.4 65.6 to 81.4	31.8 25.4 to 38.3	0.702	17.5 12.5 to 22.5	72.2 65.2 to 78.2
WISER with replaced rationale	41.6 34.0 to 49.9	74.5 69.1 to 79.7	0.768	58.5 51.3 to 65.5	87.8 82.6 to 93.3
WISER	**81.6** **75.4 to 87.6**	**22.0** **17.8 to 26.8**	**0.762**	**15.1** **11.0 to 19.2**	**Reference**

Table 9: Semantic audit on 120 events with 95% intervals. Rates are percentages, and Alpha measures annotator agreement. WISER preference compares the original WISER response with each alternative.

Table 9 shows that WISER achieves the highest evidence entailment at 81.6%, the lowest unsupported mechanism rate at 22.0%, and the lowest rate of correct yet unsupported responses at 15.1%. Replacing the WISER rationale reduces entailment to 41.6% and raises unsupported mechanisms to 74.5%, with the original response preferred in 87.8% of comparisons. WFE evaluates outcome compatibility jointly with candidate content, evidence provenance, and feedback reliability, incorporating the evidential context of market agreement into selection. These results connect world feedback curation to better supported explanations, demonstrating the value of historical outcomes for selecting grounded training responses and addressing the challenge of extracting reliable reasoning supervision from noisy market observations.

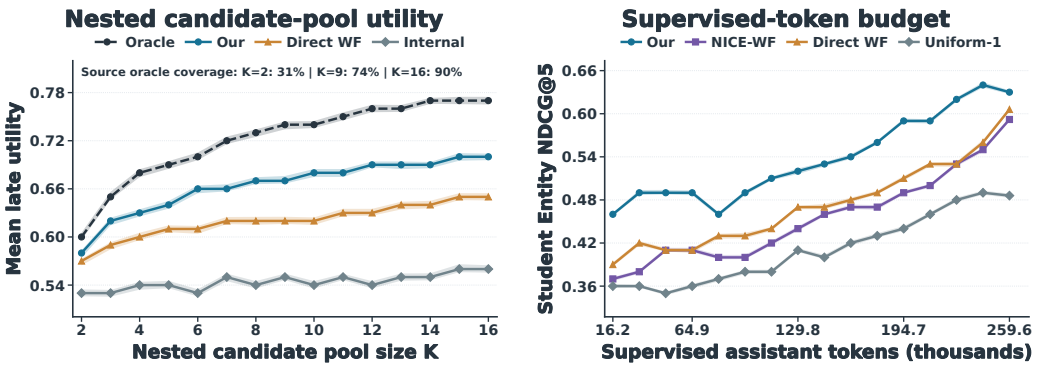
C.3 Candidate Availability and Token Budgets

Figure 5 shows that WISER leads the nonoracle selectors in mean late utility as candidate pools expand and achieves the highest student Entity NDCG@5 across the evaluated token budgets. Direct WF ranks second among nonoracle selectors, Internal and Uniform-1 rank last in their respective panels, and WISER reaches 0.630 NDCG@5 at the main budget of 259.6 thousand tokens. Larger pools offer more response alternatives, while reliability conditioned WFE learns which combinations of candidate content and market feedback predict later compatibility, selecting responses before the shared token budget is applied. These results support world feedback as a curation signal that improves both response selection and student data efficiency, addressing the challenge of extracting useful supervision from synthetic candidates under limited training budgets.

C.4 Interpretation of Selection and Student Transfer

The common candidate pools distinguish generation coverage from selection quality. Every curator can choose only among the same admissible responses, so differences arise from the supervision retained and its subsequent training effect. The selector cannot insert an entity or explanation absent from the locked responses, but it can favor responses that the generator can express without consistently preferring them. Locking candidates before inspecting their corresponding market outcomes restricts retrospective editing within the study workflow, while the post freeze test additionally commits student responses before the market reaction. The retrospective commitment governs the use of inputs and outcome tables within this workflow without assuming that generator or evaluator pretraining excluded historical events.

GRAPE and the internal score have different operational definitions. GRAPE measures mean response token log likelihood under the initial student. We define the internal score as the fixed generator rating under the accuracy rubric used by the AlpaGasus adaptation, with the generator checkpoint as evaluator and the prompt and decoding rule fixed before candidate locking. The internal representation additionally includes structural completeness and evidence provenance descriptors. The two scores therefore measure distribution matching and model quality judgment, respectively,



(a) Candidate pool

(b) Token budget

Figure 5: Sensitivity to candidate availability and supervised token budgets. Panel a reports mean late utility; panel b reports student Entity NDCG@5.

and need not induce identical rankings even with identical model weights. WFE combines these internal descriptors with early market compatibility, observation reliability, and event context in its energy function.

The common token budget counts assistant loss evaluations, including repeated occurrences, while raw event coverage counts distinct records. Fixed-order training repeats records as needed and masks the final occurrence at the budget boundary, so response length determines the number of occurrences and distinct events contributing loss. The Long Is More comparison consequently measures length selection under the common budget operator. Retaining one record per event does not imply identical numbers of complete responses, effective training events, or market token compositions. This interpretation preserves the meaning of budget matching without equating it with identical supervision content.

When candidate pools, student initialization, and training procedures are fixed, and market feedback affects the student only through the selected candidate indices, each event conveys at most $\log_2 K$ bits of additional market information. The corresponding selection channel for 507 curation events has a bound of $507 \log_2 K$ bits. This bound concerns transmitted information and does not itself bound predictive improvement, because the model and candidate texts already contain substantial prior structure. The 507 records constitute the raw student training set, while other historical partitions support calibration, fitting, search, and validation of the selector. These stages remain subject to the same selection channel bound when their influence on the student is confined to the retained responses.

The change from the base reference to Uniform 1 measures the combined effect of supervised fine tuning on admitted candidates. It can include task adaptation, response format learning, and the influence of candidate admissibility rules. Generation from the same model family does not ensure that the supervision distribution matches the initial student output distribution at evaluation, since prompts, sampling, and admissibility can change the retained responses. A zero shot generator reference equals the initial student reference only when weights, input information, prompts, and decoding settings coincide. The present comparison consequently supports analysis of performance before and after supervised adaptation without separately identifying the contributions of formatting compliance and semantic prediction.

Different references support different levels of interpretation. The complete method has reported mean gains over Uniform 1 of 0.144 in NDCG@5 and 0.167 in MCC, measuring the overall value of selected supervision over random supervision. Its corresponding mean gains over Direct WF Score are 0.024 and 0.027, with a Sharpe gain of 0.10, measuring additional value over fixed market scoring. Direct WF Score and H0 share access to early curation feedback with WISER and provide the relevant references for separating fixed scoring, learned interactions, and search. The structural comparison further distinguishes linear and nonlinear selection under default and searched configurations with the common student protocol. Information access groups identify supervision sources, while component contributions follow from matched comparisons rather than the overall ranking of different methods.

Effect sizes are interpreted in relation to their metric scales and reference methods. Relative percentages depend on the baseline level and are not directly comparable to accuracy point improvements on other benchmarks. Event reasoning and investment metrics provide different perspectives, while investment metrics derived from the same return sequence are statistically dependent. When AVol is reconstructed from ARR and Sharpe, it is a derived quantity whose propagated error differs from an empirical standard deviation across runs. The results support performance comparisons under the stated protocol, while improved market compatibility does not automatically establish the semantic or causal correctness of generated explanations.

D Candidate Comparison for a Financial Event

NVIDIA's earnings release on 19 November 2025 reported revenue of 57.0 billion dollars, including 51.2 billion dollars from Data Center. The date falls within the curation interval. Table 10 contrasts broad chip growth spillovers, gains among production and systems partners, and negative valuation adjustment over 20 trading days.

Table 10: Candidate comparison for the public earnings event.

Field	**Candidate 1**	**Selected candidate 2**	**Candidate 3**
Entities and directions	NVDA, AMD, AVGO positive	**NVDA, TSM, DELL positive**	AMD, DELL negative
Mechanism	Broad demand spillovers	**Production and systems links**	Valuation adjustment
Response quality b_k	0.450	**0.900**	0.700
Early compatibility	$(0.900, 0.820, 0.800)$	$\mathbf{(0.780, 0.760, 0.710)}$	$(0.430, 0.380, 0.390)$
Later compatibility	$(0.620, 0.510, 0.610)$	$\mathbf{(0.830, 0.790, 0.780)}$	$(0.480, 0.430, 0.500)$
Confirmation utility c_k	0.580	**0.800**	0.470
Target preference Q_k	0.218	**0.656**	0.126
WFE free energy F_k	-1.407	$\mathbf{-2.144}$	-1.254
Selection probability p_k	0.253	**0.529**	0.217

Response entities, directions, and confidence become q_{eki}, while provenance, structure, and internal scores form $\boldsymbol{b}_{ek}$. Candidate 2 assigns positive confidence values of 0.85, 0.75, and 0.65 to NVDA, TSM, and DELL. The calculation uses the tabulated quality coordinate b_k, shared early reliability $\rho = 0.65$, and zero missingness indicators. Equal weighting of later components gives c_k, and temperature $\tau_c = 0.20$ gives Q_k, connecting structured responses to later preferences.

Energy calculation uses $H = 1$ and visible weight 0.4 on b. Hidden weights on b, the three early compatibility coordinates, and reliability are $2, 2/3, 2/3, 2/3, 2$, with bias -3. All remaining weights are zero, reducing the full energy expression to

$$F_k = -0.4b_k - \log\left[1 + \exp\left(-3 + 2b_k + \frac{2}{3}\sum_{u\in\{\mathrm{dir,rank,cover}\}} M_{ek}^{(u,\mathrm{sel})} + 2\rho\right)\right]. \tag{10}$$

Candidate 2 has the lowest energy, so the selected index is $\widehat{k}_e = 2$. Candidate 1 has higher mean early compatibility, 0.840 versus 0.750, but the joint score retains candidate 2. Later preferences assess this choice retrospectively, and the curation rule passes (x_e, y_{e2}) with the original response text to the student. Matched evaluation on future events measures the resulting student training value.

E Mathematical Foundations of WISER

This appendix develops the mathematical properties of the finite-candidate selector used by WISER. The argument follows the same order as the method itself. We first establish that the world-feedback quantities and candidate preferences are well defined. We then derive the exact WFE free energy,

characterize the additional ranking behavior introduced by reliability-conditioned hidden interactions, and identify the statistical target of the listwise objective. These results lead to two decision guarantees: a score-space bound that controls confirmation-utility loss under model and input error, and a loss-space bound that converts listwise KL error directly into candidate-selection regret. The final subsections connect these guarantees to constrained configuration, temporal commitment, and the selection-only path through which historical outcomes reach the student.

The theoretical chain is

$$\begin{aligned}&\text{bounded world feedback} \Longrightarrow \text{valid event preferences} \Longrightarrow \text{reliability-conditioned WFE}\\ &\Longrightarrow \text{preference fitting} \Longrightarrow \text{selection guarantees} \Longrightarrow \text{fixed-population configuration}\\ &\Longrightarrow \text{frozen student transfer.}\end{aligned} \tag{11}$$

All statements are pointwise or finite-sample unless a probability distribution is introduced explicitly. This matches the experimental protocol in Methodology, which fits WFE on fixed historical partitions, configures the selector on fixed development events, and evaluates the final student on post-freeze events. Thresholds, windows, weights, and the candidate manifest are supplied by the same registered configuration used in evaluation.

E.1 Algorithmic Overview

Algorithm 1 summarizes WISER using the registered protocol $\mathcal{P}$, which fixes the chronological partitions, the two feedback windows, the calibration rules, the configuration space Ξ with its evaluation budget N_{BO} and feasibility thresholds, the student training specification, and the freeze time T_{freeze}. Let $D_{\mathrm{dev}}^{(\mathrm{retro})} = D_{\mathrm{cal}} \cup D_{\mathrm{fit}} \cup D_{\mathrm{BO}} \cup D_{\mathrm{val}} \cup D_{\mathrm{curate}}$ denote the development events. Phase 1 constructs and locks candidate pools before any outcome table is opened. Phase 2 measures world feedback, fits WFE for every evaluated configuration, and freezes the configuration chosen by constrained search. Phase 3 curates the student records, trains the student, and commits post-freeze test responses.

Commit writes the signed manifest: candidate texts and order, seeds, generator and retrieval revisions, evidence cutoff, admissibility decisions, and internal scores. WorldFeedback fits expected returns on $[d_e - 120, d_e - 21]$, computes standardized abnormal returns over the view's window, security and event reliability as in Equation (18), the calibrated anchor set and reaction mass, and the three compatibility components of Equations (21)–(23); it marks ranking and coverage as missing when the valid reaction mass is zero and returns the validity flag V_e and the spread Δ_e for the confirmation view. Propose applies the registered constrained acquisition rule, GP-UCB in the primary setting, to the completed evaluations. Evaluate computes R_{cov} and N_{eff} on D_{fit}^{+}, R_{prov} and R_{dom} on the selected records, and the fixed-population utility of Equation (8) with weights $\bar{\omega}_e = \rho_e^{(\mathrm{conf})}$; on D_{val} it returns J_{val} by the same aggregation. Shortlist keeps the highest scoring feasible configurations under the prespecified shortlist rule. FitWFE optimizes Equation (6) under the common fitting protocol and returns its checkpoint. Train optimizes Equation (9) from ψ_0 using the fixed optimizer settings and randomness R_{train}.

Every $\min\{k : \cdots\}$ implements the committed-order tie rule. A configuration whose supervision weight $\Omega_{\boldsymbol{\xi}}$ is zero is never fitted and therefore never enters Ξ_{eval}, which contains completed evaluations only. Market evaluation denominators are checked once and then held fixed, and $\mathcal{V}_{\mathrm{BO}}^{m}$ and $\mathcal{V}_{\mathrm{val}}^{m}$ retain fully tied events while D_{fit}^{+} excludes them. Curation on Line 28 reads only the selection view and the event context; market measurements, energies, selector probabilities, and search traces stay outside D_{SFT}. Window offsets follow each security's market calendar, and the confirmation view of a partition matures before the next partition opens.

Line 5 is the commitment formalized by Proposition E.29. Lines 10 and 13 produce the bounded quantities of Proposition E.1 and the ordered targets of Lemma E.9. Line 19 fits the KL-projection objective of Proposition E.11, whose mathematical minimum exists by Proposition E.13, and Line 20 is the selection whose regret and stability are controlled by Theorems E.19 and E.17. Lines 21–25 use the statistics of Proposition E.24, the utility of Proposition E.26 and Corollary E.27, and the uniqueness of Proposition E.28. Lines 28–31 form the channel of Theorem E.32 and Corollary E.33, and Line 33 satisfies Assumption E.30 and Proposition E.31.

Algorithm 1 WISER: Candidate Locking, Reliability-Conditioned Selection, and Student Transfer

Require: Generator G_ϑ with admissibility rule A_{adm} and pool size K; protocol $\mathcal{P}$; partitions $D_{\mathrm{cal}} \prec D_{\mathrm{fit}} \prec D_{\mathrm{BO}} \prec D_{\mathrm{val}} \prec D_{\mathrm{curate}}$ and test events D_{test}; student base ψ_0 with token budget B_{tok}

Ensure: Frozen selector $(\boldsymbol{\xi}^\star, \theta^\star)$, curated records D_{SFT}, student $\psi^\star$, and committed test responses $\{\widehat{y}_e\}_{e \in D_{\mathrm{test}}}$

/* Phase 1: event-time candidate construction and commitment */
for $e \in D_{\mathrm{dev}}^{(\mathrm{retro})}$ **do**
  $\widetilde{y}_{ej} \leftarrow G_\vartheta(x_e; \zeta_{ej}),\ j = 1, \dots, K_{\mathrm{raw}}$
  $\mathcal{C}_e \leftarrow \mathrm{FirstK}\big([\widetilde{y}_{ej} \mid A_{\mathrm{adm}}(x_e, \widetilde{y}_{ej}) = 1]\big);\ \boldsymbol{s}_{ek}^{(\mathrm{int})} \leftarrow S_\vartheta(x_e, y_{ek}),\ k = 1, \dots, K$
  COMMIT$(\mathcal{C}_e, \zeta_{e\cdot}, \boldsymbol{s}_{e\cdot}^{(\mathrm{int})}; \mathcal{P})$ ▷ Signed manifest before outcome access
end for
/* Phase 2: world feedback, preference fitting, and constrained configuration */
$\Gamma^\star \leftarrow \mathrm{Calibrate}(D_{\mathrm{cal}})$
for $e \in D_{\mathrm{fit}} \cup D_{\mathrm{BO}} \cup D_{\mathrm{val}} \cup D_{\mathrm{curate}}$, once $\mathcal{W}_e^{(\mathrm{sel})}$ matures **do**
  $(\rho_e^{(\mathrm{sel})}, \boldsymbol{m}_{e\cdot}^{(\mathrm{sel})}) \leftarrow$ WORLDFEEDBACK$(e, \mathrm{sel}; \Gamma^\star);\ \boldsymbol{\phi}_{ek} \leftarrow [\boldsymbol{b}_{ek}; \boldsymbol{m}_{ek}^{(\mathrm{sel})}; \rho_e^{(\mathrm{sel})}; \boldsymbol{\phi}_{ek}^{(\mathrm{miss,sel})}]$
end for
for $e \in D_{\mathrm{fit}} \cup D_{\mathrm{BO}} \cup D_{\mathrm{val}}$, once $\mathcal{W}_e^{(\mathrm{conf})}$ matures **do**
  $(\rho_e^{(\mathrm{conf})}, \boldsymbol{m}_{e\cdot}^{(\mathrm{conf})}) \leftarrow$ WORLDFEEDBACK$(e, \mathrm{conf}; \Gamma^\star);\ c_{ek} \leftarrow$ confirmation utility of Equation (27)
end for
$D_{\mathrm{fit}}^+ \leftarrow \{e \in D_{\mathrm{fit}} : V_e = 1,\ \rho_e^{(\mathrm{conf})} > 0,\ \Delta_e > 0\}$ ▷ Fully tied targets leave fitting only
for $t = 1, \dots, N_{\mathrm{BO}}$ **do**
  $\boldsymbol{\xi}_t \leftarrow$ PROPOSE$(\{(\boldsymbol{\xi}_s, J_{\mathrm{BO}}(\boldsymbol{\xi}_s))\}_{s<t}; \mathcal{P})$ ▷ $\boldsymbol{\xi} = (H, \log \lambda_{\mathrm{reg}}, \rho_{\min}, \log \tau_c)$
  $Q_e(k) \leftarrow \mathrm{softmax}_k(c_{e\cdot}/\tau_c);\ \omega_e \leftarrow \rho_e^{(\mathrm{conf})}\, \mathbb{I}[\rho_e^{(\mathrm{conf})} \geq \rho_{\min}],\ e \in D_{\mathrm{fit}}^+$
  $\widehat{\theta}_{\boldsymbol{\xi}_t} \leftarrow$ FITWFE$(D_{\mathrm{fit}}^+; \boldsymbol{\xi}_t, \mathcal{P})$ ▷ Common fitting protocol
  $\widehat{k}_e(\boldsymbol{\xi}_t) \leftarrow \min\{k : F_{\widehat{\theta}_{\boldsymbol{\xi}_t}}(e, k) = \min_\ell F_{\widehat{\theta}_{\boldsymbol{\xi}_t}}(e, \ell)\},\ e \in D_{\mathrm{BO}}$
  $(R_{\mathrm{cov}}, N_{\mathrm{eff}}, R_{\mathrm{prov}}, R_{\mathrm{dom}}, J_{\mathrm{BO}})(\boldsymbol{\xi}_t) \leftarrow$ EVALUATE$(\widehat{k}_\cdot(\boldsymbol{\xi}_t), D_{\mathrm{fit}}^+, D_{\mathrm{BO}})$
end for
$\Xi_{\mathrm{feas}} \leftarrow \{\boldsymbol{\xi} \in \Xi_{\mathrm{eval}} : \text{Equation (7) holds}\}$
$\Xi_{\mathrm{short}} \leftarrow$ SHORTLIST$(\Xi_{\mathrm{feas}}, J_{\mathrm{BO}});\ J_{\mathrm{val}}(\boldsymbol{\xi}) \leftarrow$ EVALUATE$(\widehat{k}_\cdot(\boldsymbol{\xi}), D_{\mathrm{val}}),\ \boldsymbol{\xi} \in \Xi_{\mathrm{short}}$ ▷ Once
$\boldsymbol{\xi}^\star \leftarrow \arg\max_{\boldsymbol{\xi} \in \Xi_{\mathrm{short}}} J_{\mathrm{val}}(\boldsymbol{\xi})$ under the deterministic tie rule; $\theta^\star \leftarrow \widehat{\theta}_{\boldsymbol{\xi}^\star}$
/* Phase 3: curation, student training, and post-freeze commitment */
for $e \in D_{\mathrm{curate}}$ **do**
  $I_e \leftarrow \min\{k : F_{\theta^\star}(e, k) = \min_\ell F_{\theta^\star}(e, \ell)\}$
end for
$D_{\mathrm{raw}} \leftarrow \{(x_e, y_{eI_e}) : e \in D_{\mathrm{curate}}\};\ D_{\mathrm{SFT}} \leftarrow \mathrm{TokenBudget}_{B_{\mathrm{tok}}}(D_{\mathrm{raw}})$
$\psi^\star \leftarrow$ TRAIN$(\psi_0, D_{\mathrm{SFT}}; R_{\mathrm{train}});$ freeze $(\boldsymbol{\xi}^\star, \theta^\star, \psi^\star)$ at T_{freeze}
for $e \in D_{\mathrm{test}}$, at $t_e > T_{\mathrm{freeze}}$ **do**
  $\widehat{y}_e \leftarrow \mathrm{Decode}(\psi^\star, x_e)$, committed at $T_e^{(\mathrm{output})} < a_e$
end for
return $(\boldsymbol{\xi}^\star, \theta^\star)$, D_{SFT}, $\psi^\star$, $\{\widehat{y}_e\}_{e \in D_{\mathrm{test}}}$, and the manifest, feedback, search, and evaluation records

E.2 Setup and Bounded World Feedback

This subsection fixes the notation used throughout the appendix and proves that every quantity entering the WFE objective lies in a controlled domain. The boundedness result is the starting point for the preference and regret analysis in later subsections.

Notation. For an event e, let t_e denote its release time and let x_e contain the event-time information available by t_e. A fixed generator G_ϑ produces ordered raw responses. After the deterministic admissibility rule A_{adm} in Equation (2), the first K admissible responses form the committed candidate pool

$$\mathcal{C}_e = \{y_{e1}, \ldots, y_{eK}\}, \qquad K \geq 2. \tag{12}$$

The candidate pool and its order are fixed before the corresponding outcome table is opened. Table 11 lists the symbols used in the theoretical analysis.

To keep the two uses of the letter B in Methodology distinct, this appendix writes the context-to-hidden matrix as B_g and the student token budget as B_{tok}. The WFE parameters are therefore

$$\theta = (\boldsymbol{a}_0, A, \boldsymbol{\eta}_0, B_g, W), \tag{13}$$

with dimensions

$$\boldsymbol{a}_0 \in \mathbb{R}^d, \quad A \in \mathbb{R}^{d\times q}, \quad \boldsymbol{\eta}_0 \in \mathbb{R}^H, \quad B_g \in \mathbb{R}^{H\times q}, \quad W \in \mathbb{R}^{d\times H}. \tag{14}$$

The context-dependent biases are

$$\boldsymbol{a}_e = \boldsymbol{a}_0 + A\boldsymbol{g}_e, \qquad \boldsymbol{\eta}_e = \boldsymbol{\eta}_0 + B_g\boldsymbol{g}_e. \tag{15}$$

We use the parameter norm

$$\|\theta\|_2^2 = \|\boldsymbol{a}_0\|_2^2 + \|A\|_F^2 + \|\boldsymbol{\eta}_0\|_2^2 + \|B_g\|_F^2 + \|W\|_F^2. \tag{16}$$

Reliability and compatibility measurements. For a valid feedback view, security-level reliability is the geometric mean

$$r_{ie}^{(v)} = \left(\prod_{c\in\mathcal{C}_{\mathrm{rel}}} s_{ie,c}^{(v)}\right)^{1/|\mathcal{C}_{\mathrm{rel}}|}, \tag{17}$$

and event reliability is

$$\rho_e^{(v)} = \mathbb{I}[n_e^{(v)} \geq n_{\min}] \frac{\sum_{i\in\mathcal{U}_e} \nu_{ei} r_{ie}^{(v)}}{\sum_{i\in\mathcal{U}_e} \nu_{ei} + \epsilon}, \tag{18}$$

which restates Equation (3). Let $z_{ei}^{(v)}$ denote the standardized abnormal return used in view v. Candidate k assigns confidence $\gamma_{eki} \in [0,1]$ and direction $d_{eki} \in \{-1,+1\}$ to each security in its structured affected-entity field. Its prediction coordinate is

$$q_{eki} = \gamma_{eki} d_{eki} \mathbb{I}[i \in S_{ek}], \tag{19}$$

where S_{ek} is the predicted entity set. Both views evaluate the same q_{eki} over their respective market windows. The transformed market response is

$$v_{ei}^{(v)} = \tanh\left(\frac{z_{ei}^{(v)}}{s_{\mathrm{scale}}}\right), \qquad s_{\mathrm{scale}} > 0. \tag{20}$$

The three compatibility components are

$$M_{ek}^{(\mathrm{dir},v)} = 1 - \frac{\sum_i r_{ie}^{(v)} (q_{eki} - v_{ei}^{(v)})^2}{4\sum_i r_{ie}^{(v)} + \epsilon}, \tag{21}$$

$$M_{ek}^{(\mathrm{rank},v)} = \mathrm{NDCG@}\,L\left(|q_{ek\cdot}|; g_{e\cdot}^{(v)}\right), \tag{22}$$

and

$$M_{ek}^{(\mathrm{cover},v)} = \frac{\sum_i g_{ei}^{(v)} \mathbb{I}[i \in S_{ek}]}{\sum_i g_{ei}^{(v)} + \epsilon}. \tag{23}$$

Symbol	Definition
e	event index
$k \in \{1, \dots, K\}$	candidate index within the committed pool
$v \in \{\mathrm{sel}, \mathrm{conf}\}$	selection or confirmation feedback view
$\mathcal{U}_e$	finite security universe associated with event e
$\mathcal{C}_{\mathrm{rel}}$	index set of security-level reliability components
$s_{ie,c}^{(v)} \in [0,1]$	normalized reliability component c for security i
$r_{ie}^{(v)}$	security-level observation reliability
$\nu_{ei} \geq 0$	pre-event liquidity weight
$n_e^{(v)}$	number of securities with valid feedback
$n_{\min}$	minimum valid-security coverage
$\epsilon > 0$	denominator stabilizer in the reliability and compatibility formulas
$\rho_e^{(v)}$	event-level observation reliability
$\Gamma^\star$	frozen transformation state $\mathrm{Calibrate}(D_{\mathrm{cal}})$
$\boldsymbol{b}_{ek}$	outcome-blind candidate descriptors after frozen transformation $\Gamma^\star$
$\boldsymbol{m}_{ek}^{(v)}$	world-feedback compatibility vector in view v
S_{ek}	securities in candidate k's structured affected-entity field
γ_{eki}	candidate confidence for security i
$d_{eki} \in \{-1, +1\}$	encoded candidate direction for a mentioned security i
$z_{ei}^{(v)}$	standardized abnormal return
$s_{\mathrm{scale}} > 0$	frozen scale used to transform standardized abnormal returns
$s_{\max} > 0$	cap on the magnitude entering reaction mass
$\mathcal{A}_e^{(v)}$	calibrated set of valid reaction anchors
$r_{\min}, q_0, s_0$	calibrated anchor thresholds for reliability, placebo probability, and reaction magnitude
$q_{ei}^{(\mathrm{pl},v)}$	calibrated placebo probability for security i
L	NDCG ranking cutoff
$g_{ei}^{(v)}$	nonnegative reaction mass used by ranking and coverage
$\boldsymbol{\phi}_{ek}$	WFE selection representation $\boldsymbol{\phi}_{ek}^{(\mathrm{sel})}$
$\boldsymbol{\phi}_{ek}^{(\mathrm{miss,sel})}$	missingness indicators for unavailable early-feedback measurements
$\boldsymbol{g}_e \in \mathbb{R}^q$	event context vector
c_{ek}	confirmation utility $c_{ek}^{(\mathrm{conf})}$
$\alpha_d, \alpha_r, \alpha_c$	nonnegative confirmation-utility weights summing to one
$\tau_c > 0$	confirmation-target temperature
$Q_e(k)$	confirmation preference probability $Q_e^{(\mathrm{conf})}(k)$
$H \in \mathbb{N}_0$	number of binary WFE hidden units
$\boldsymbol{h} \in \{0,1\}^H$	hidden state
θ	WFE parameter vector
$F_\theta(e,k)$	candidate free energy
$S_\theta(e,k) = -F_\theta(e,k)$	candidate score
$p_\theta(k \mid e)$	WFE candidate probability
$\widehat{k}_e$	selected candidate index
$\lambda_{\mathrm{reg}} > 0$	WFE ℓ_2 regularization coefficient
$\rho_{\min}$	minimum confirmation reliability used for fitting support
ω_e	confirmation supervision weight $\omega_e^{(\mathrm{conf})}$
$\bar{\omega}_e$	fixed confirmation weight used in BO and validation evaluation
$\boldsymbol{\xi}$	selector configuration $(H, \log \lambda_{\mathrm{reg}}, \rho_{\min}, \log \tau_c)$

Table 11: Notation used in the theoretical analysis. Symbols follow Methodology; c_{ek}, $Q_e(k)$, ω_e, and $\boldsymbol{\phi}_{ek}$ abbreviate their confirmation-view and selection-view counterparts in the main text.

Here

$$g_{ei}^{(v)} = r_{ie}^{(v)} \min\{|z_{ei}^{(v)}|, s_{\max}\} \mathbb{I}[i \in \mathcal{A}_e^{(v)}], \qquad s_{\max} > 0, \tag{24}$$

and the calibrated anchor set is

$$\mathcal{A}_e^{(v)} = \left\{ i \in \mathcal{U}_e : r_{ie}^{(v)} \geq r_{\min},\ q_{ei}^{(\mathrm{pl},v)} \leq q_0,\ |z_{ei}^{(v)}| \geq s_0 \right\}. \tag{25}$$

Here $r_{\min}$ is the minimum security-level reliability, $q_{ei}^{(\mathrm{pl},v)}$ is the calibrated placebo probability, q_0 is its maximum accepted value, s_0 is the minimum standardized reaction magnitude, and L is the ranking cutoff in NDCG@L. The validity rule treats ranking and coverage as missing when the valid reaction mass is zero, so Equation (22) is used only when its ideal discounted gain is positive.

The selection representation is

$$\boldsymbol{\phi}_{ek} = [\boldsymbol{b}_{ek}; \boldsymbol{m}_{ek}^{(\mathrm{sel})}; \rho_e^{(\mathrm{sel})}; \boldsymbol{\phi}_{ek}^{(\mathrm{miss,sel})}] \in \mathbb{R}^d. \tag{26}$$

Confirmation utility is

$$c_{ek} = \alpha_d M_{ek}^{(\mathrm{dir,conf})} + \alpha_r M_{ek}^{(\mathrm{rank,conf})} + \alpha_c M_{ek}^{(\mathrm{cover,conf})}, \tag{27}$$

where $\alpha_d, \alpha_r, \alpha_c \geq 0$ and $\alpha_d + \alpha_r + \alpha_c = 1$. The primary setting uses $\alpha_d = \alpha_r = \alpha_c = 1/3$.

Proposition E.1 (Bounded feedback and confirmation utility)**.** *For every valid event-view pair,*

$$0 \leq r_{ie}^{(v)} \leq 1, \qquad 0 \leq \rho_e^{(v)} < 1. \tag{28}$$

For every valid candidate measurement,

$$0 \leq M_{ek}^{(\mathrm{dir},v)},\ M_{ek}^{(\mathrm{rank},v)},\ M_{ek}^{(\mathrm{cover},v)} \leq 1. \tag{29}$$

Consequently,

$$0 \leq c_{ek} \leq 1. \tag{30}$$

Proof. Each reliability component lies in $[0, 1]$, so their geometric mean also lies in $[0, 1]$. Because $\nu_{ei} \geq 0$,

$$0 \leq \sum_i \nu_{ei} r_{ie}^{(v)} \leq \sum_i \nu_{ei}. \tag{31}$$

The denominator in Equation (18) is strictly larger than $\sum_i \nu_{ei}$ because $\epsilon > 0$. Multiplication by the coverage indicator preserves the lower bound and gives $0 \leq \rho_e^{(v)} < 1$.

Equation (19) gives $q_{eki} \in [-1, 1]$, while Equation (20) gives $v_{ei}^{(v)} \in (-1, 1)$. Hence

$$(q_{eki} - v_{ei}^{(v)})^2 \leq 4. \tag{32}$$

The numerator subtracted in Equation (21) is therefore at most $4\sum_i r_{ie}^{(v)}$, which gives $0 \leq M_{ek}^{(\mathrm{dir},v)} \leq 1$. Valid NDCG with nonnegative gains is normalized by its ideal ranking and lies in $[0, 1]$. The reaction mass in Equation (24) is nonnegative, and the numerator of Equation (23) is at most its denominator before the positive stabilizer is added, so $M_{ek}^{(\mathrm{cover},v)} \in [0, 1]$. Equation (27) is a convex combination of these three quantities, which proves Equation (30). □

The bounded interval in Equation (30) is used directly by the selection-regret theorems in Section E.6. It also implies that every market-level BO utility in Section E.7 is bounded.

Confirmation preference distribution. For $\tau_c > 0$, WISER converts valid confirmation utilities into the event-level target of Equation (4),

$$Q_e(k) = \frac{\exp(c_{ek}/\tau_c)}{\sum_{\ell=1}^{K} \exp(c_{e\ell}/\tau_c)}. \tag{33}$$

Every coordinate is strictly positive and the coordinates sum to one. Define the confirmation spread

$$\Delta_e = \max_k c_{ek} - \min_k c_{ek} \in [0, 1]. \tag{34}$$

Let $V_e = 1$ denote an available confirmation view with sufficient security coverage and finite confirmation utilities for all K candidates. The configuration-independent fitting base set is

$$D_{\mathrm{fit}}^{+} = \left\{ e \in D_{\mathrm{fit}} : V_e = 1,\ \rho_e^{(\mathrm{conf})} > 0,\ \Delta_e > 0 \right\}. \tag{35}$$

The configuration-dependent fitting weight is

$$\omega_e = \rho_e^{(\mathrm{conf})} \mathbb{I}[\rho_e^{(\mathrm{conf})} \geq \rho_{\min}]. \tag{36}$$

Assumption E.2 (Positive supervision)**.** For every fitted configuration $\boldsymbol{\xi}$,

$$\Omega_{\boldsymbol{\xi}} = \sum_{e \in D^{+}_{\mathrm{fit}}} \omega_e > 0. \tag{37}$$

This is the positive-supervision condition already enforced by the fitting protocol in Methodology. Configurations that violate it are never fitted, so every result below that conditions on a fitted configuration inherits Equation (37).

E.3 Exact Finite-Candidate WFE

This subsection derives the free energy used by the selector from its joint energy. The finite candidate pool permits exact hidden-state marginalization and exact normalization over candidates, so the WFE likelihood is evaluated directly.

For candidate k and hidden state $\boldsymbol{h} \in \{0,1\}^H$, the joint energy is

$$E_\theta(e, k, \boldsymbol{h}) = -\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek} - \boldsymbol{h}^\top (\boldsymbol{\eta}_e + W^\top \boldsymbol{\phi}_{ek}). \tag{38}$$

Write

$$u_{ekj} = \eta_{e,j} + W_{\cdot j}^\top \boldsymbol{\phi}_{ek}. \tag{39}$$

Proposition E.3 (Exact marginalization and free energy)**.** *For every event e and candidate k,*

$$\sum_{\boldsymbol{h} \in \{0,1\}^H} \exp[-E_\theta(e, k, \boldsymbol{h})] = \exp(\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek}) \prod_{j=1}^{H} (1 + \exp u_{ekj}). \tag{40}$$

The corresponding free energy is

$$F_\theta(e, k) = -\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek} - \sum_{j=1}^{H} \mathrm{softplus}(u_{ekj}), \tag{41}$$

where $\mathrm{softplus}(u) = \log(1 + e^u)$*, in agreement with Equation* (5)*. Exact normalization over the committed pool gives*

$$p_\theta(k \mid e) = \frac{\exp[-F_\theta(e, k)]}{\sum_{\ell=1}^{K} \exp[-F_\theta(e, \ell)]}. \tag{42}$$

Thus $p_\theta(\cdot \mid e)$ is a strictly positive probability distribution on the K committed candidates.

Proof. Expanding the negative energy gives

$$\exp[-E_\theta(e, k, \boldsymbol{h})] = \exp(\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek}) \exp\left(\sum_{j=1}^{H} h_j u_{ekj} \right). \tag{43}$$

The hidden coordinates are binary, so the finite sum factorizes:

$$\begin{aligned} \sum_{\boldsymbol{h} \in \{0,1\}^H} \exp\left(\sum_{j=1}^{H} h_j u_{ekj} \right) &= \prod_{j=1}^{H} \sum_{h_j \in \{0,1\}} e^{h_j u_{ekj}} \\ &= \prod_{j=1}^{H} (1 + e^{u_{ekj}}). \end{aligned} \tag{44}$$

Multiplying by the visible factor proves Equation (40). Taking the negative logarithm gives Equation (41). Every free energy is finite for finite inputs and parameters, so every exponential in Equation (42) is positive and the finite denominator is positive. Summing Equation (42) over k gives one. □

The selector uses

$$\widehat{k}_e = \min\left\{k : F_\theta(e,k) = \min_{1\le \ell \le K} F_\theta(e,\ell)\right\}, \tag{45}$$

where the outer minimum implements the committed-order tie rule. A minimum exists because the candidate set is finite, and the tie rule makes the returned index unique.

For $H = 0$, Equation (41) reduces to

$$F_\theta(e,k) = -\boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek}, \tag{46}$$

which is the context-dependent linear selector used as the $H = 0$ structural reference. For $H > 0$, each hidden unit contributes a softplus of an affine function of the candidate representation.

Given $\boldsymbol{a}_e$ and $\boldsymbol{\eta}_e$, one candidate requires $O(dH + d)$ operations, and all K candidates require

$$O(Kd(H+1)). \tag{47}$$

Computing the context-dependent biases once per event adds $O((d+H)q)$ operations.

E.4 Reliability-Conditioned Expressive Capacity

This subsection isolates the mathematical role of reliability conditioning. The shared reliability scalar enters every candidate representation for the same event. A purely additive use of that scalar cancels from pairwise comparisons. The hidden WFE terms create candidate-dependent reliability responses, which allow the ranking itself to change with observation quality.

Write the selection representation as

$$\boldsymbol{\phi}_{ek} = (\boldsymbol{\chi}_{ek}, \rho_e), \qquad \rho_e = \rho_e^{(\mathrm{sel})}, \tag{48}$$

where $\boldsymbol{\chi}_{ek}$ collects the remaining candidate-specific coordinates. In this subsection, $\boldsymbol{g}_e$, $\boldsymbol{\chi}_{ek}$, and θ are fixed when a partial derivative with respect to ρ_e is taken.

Proposition E.4 (Additive event terms cancel from candidate comparisons)**.** *Suppose a candidate score has the form*

$$S(e,k) = T(e,k) + b(e), \tag{49}$$

where $b(e)$ is shared by all candidates of event e. Then S and T induce the same pairwise score differences, the same softmax probabilities, and the same selected index under the committed-order tie rule. For the $H = 0$ WFE,

$$S_\theta(e,k) = \boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek} = \boldsymbol{a}_{e,\chi}^\top \boldsymbol{\chi}_{ek} + a_{e,\rho}\rho_e. \tag{50}$$

Therefore, with the main-text representation fixed, the explicit shared reliability coordinate has no effect on within-event ranking when $H = 0$.

Proof. For any k and ℓ,

$$S(e,k) - S(e,\ell) = T(e,k) - T(e,\ell). \tag{51}$$

The softmax probabilities satisfy

$$\frac{e^{T(e,k)+b(e)}}{\sum_\ell e^{T(e,\ell)+b(e)}} = \frac{e^{T(e,k)}}{\sum_\ell e^{T(e,\ell)}}. \tag{52}$$

Hence pairwise order, normalized probabilities, and the deterministic tie rule are unchanged. Equation (50) is a direct application with $b(e) = a_{e,\rho}\rho_e$. □

This result gives the linear reference a precise interpretation: reliability becomes ranking-relevant only through interactions with candidate-varying coordinates.

Proposition E.5 (Pairwise reliability response of WFE)**.** *Let $w_{\rho j}$ denote the reliability coordinate of hidden weight $W_{\cdot j}$. For two candidates k and ℓ,*

$$\frac{\partial}{\partial \rho_e}\left[S_\theta(e,k) - S_\theta(e,\ell)\right] = \sum_{j=1}^{H} w_{\rho j}\left[\sigma(u_{ekj}) - \sigma(u_{e\ell j})\right], \tag{53}$$

where $\sigma(u) = (1+e^{-u})^{-1}$. If χ_r is a candidate-specific coordinate with hidden weight w_{rj}, then

$$\frac{\partial^2 S_\theta}{\partial \rho\, \partial \chi_r} = \sum_{j=1}^{H} w_{\rho j} w_{rj} \sigma(u_j)[1-\sigma(u_j)]. \tag{54}$$

Proof. The score is

$$S_\theta(e,k) = \boldsymbol{a}_e^\top \boldsymbol{\phi}_{ek} + \sum_{j=1}^{H} \mathrm{softplus}(u_{ekj}). \tag{55}$$

Because $\mathrm{softplus}'(u) = \sigma(u)$,

$$\frac{\partial S_\theta(e,k)}{\partial \rho_e} = a_{e,\rho} + \sum_{j=1}^{H} w_{\rho j}\sigma(u_{ekj}). \tag{56}$$

Subtracting the corresponding expression for candidate ℓ removes the shared visible term and gives Equation (53). Differentiating once more with respect to χ_r and using $\sigma'(u) = \sigma(u)[1-\sigma(u)]$ gives Equation (54). □

Equation (53) states the mechanism in pairwise form. Candidate-dependent hidden activations make the effect of reliability depend on the candidate representation. Equation (54) gives the local interaction strength between reliability and any candidate feature.

Theorem E.6 (One hidden unit can reverse a candidate ranking as reliability changes)**.** *Assume the representation contains at least one candidate-varying coordinate. Consider two candidates whose value on one such scalar coordinate is $z \in \{0,1\}$ and a shared reliability coordinate $\rho \in [0,1]$. For any $\beta > 0$, the one-hidden-unit score*

$$S_\beta(z,\rho) = -\frac{\beta}{2}z + \mathrm{softplus}\left(\beta z + 2\beta\rho - \frac{3\beta}{2}\right) \tag{57}$$

is realizable by an $H = 1$ WFE. Define the score gap between the $z = 1$ and $z = 0$ candidates by

$$G_\beta(\rho) = S_\beta(1,\rho) - S_\beta(0,\rho). \tag{58}$$

Then

$$G_\beta(1/4) = -m_\beta < 0, \qquad G_\beta(3/4) = m_\beta > 0, \tag{59}$$

where

$$m_\beta = \log\cosh(\beta/2) > 0. \tag{60}$$

Thus the preferred candidate changes as reliability moves from $1/4$ to $3/4$. Every $H = 0$ WFE keeps this pairwise ranking fixed when the remaining inputs are fixed.

Proof. Equation (57) is obtained by taking visible weight $-\beta/2$ on z, hidden weight β on z, hidden weight 2β on ρ, and hidden bias $-3\beta/2$. All parameters are finite.

At $\rho = 1/4$,

$$G_\beta(1/4) = -\frac{\beta}{2} + \mathrm{softplus}(0) - \mathrm{softplus}(-\beta). \tag{61}$$

At $\rho = 3/4$,

$$G_\beta(3/4) = -\frac{\beta}{2} + \mathrm{softplus}(\beta) - \mathrm{softplus}(0). \tag{62}$$

Using

$$\mathrm{softplus}(-\beta) = \mathrm{softplus}(\beta) - \beta \tag{63}$$

and

$$\mathrm{softplus}(\beta) = \frac{\beta}{2} + \log 2 + \log\cosh(\beta/2), \tag{64}$$

we obtain Equation (59). Since $\beta > 0$, $\cosh(\beta/2) > 1$, so $m_\beta > 0$. Proposition E.4 gives the $H = 0$ statement. □

Corollary E.7 (The nonlinear WFE strictly contains the conditional linear selector)**.** *Under the representation condition of Theorem E.6, let $\mathcal{P}_H$ be the family of within-event candidate probability maps induced by WFE with hidden capacity H. Then*

$$\mathcal{P}_0 \subsetneq \mathcal{P}_H \qquad \textit{for every } H \geq 1. \tag{65}$$

Proof. Any $H = 0$ model can be embedded into an $H \geq 1$ model by setting the added hidden weights and biases to zero. Each added unit then contributes the same constant $\log 2$ to every candidate score, which cancels under softmax. Hence $\mathcal{P}_0 \subseteq \mathcal{P}_H$.

The construction in Theorem E.6 produces a probability ratio

$$\log \frac{p_\theta(z = 1 \mid \rho)}{p_\theta(z = 0 \mid \rho)} = G_\beta(\rho) \tag{66}$$

whose sign changes with reliability. Proposition E.4 shows that no $H = 0$ model with the same representation can realize this map. The inclusion is therefore strict. □

Corollary E.8 (Reliability-gated targets can yield a strict fit advantage)**.** *Fix $\tau_c > 0$ and choose $\beta > 0$ such that $\tau_c m_\beta \leq 1$, where m_β is defined in Equation* (60)*. Such a choice always exists because $m_\beta \to 0$ as $\beta \to 0$. Consider two equally weighted events with the same context vector, the same candidate-specific coordinates $z = 0, 1$, and reliability values $1/4$ and $3/4$. Assign bounded confirmation utilities*

$$\begin{aligned} (c_0^-, c_1^-) &= (\tau_c m_\beta, 0), \quad Q^- = \mathrm{softmax}(0, -m_\beta), \\ (c_0^+, c_1^+) &= (0, \tau_c m_\beta), \quad Q^+ = \mathrm{softmax}(0, m_\beta), \end{aligned} \tag{67}$$

where the first coordinate corresponds to $z = 0$. Consider the listwise data term of Equation (6) *with the regularization term removed. The $H = 1$ construction in Theorem E.6 matches both targets exactly. Every $H = 0$ model predicts the same candidate probability ratio for the two events, and its minimum average cross-entropy is $\log 2$. The $H = 1$ construction attains*

$$h(\sigma(m_\beta)) < \log 2, \tag{68}$$

where $h(p) = -p \log p - (1 - p) \log(1 - p)$ is binary entropy.

Proof. Theorem E.6 gives score gaps $-m_\beta$ and m_β, so the target probabilities in Equation (67) are matched exactly. An $H = 0$ model produces a common prediction $(r, 1 - r)$ at both reliability values. Because the average of Q^- and Q^+ is $(1/2, 1/2)$, the average cross-entropy is

$$-\frac{1}{2} \log r - \frac{1}{2} \log(1 - r) \geq \log 2, \tag{69}$$

with equality at $r = 1/2$. Exact matching with $H = 1$ gives the target entropy $h(\sigma(m_\beta))$, which is strictly below $\log 2$ because $m_\beta > 0$ and therefore $\sigma(m_\beta) \neq 1/2$. □

Theorem E.6 and its corollaries provide the theoretical counterpart to two empirical comparisons in the paper. Figure 3a tests whether early compatibility becomes more informative as event reliability increases, while Table 7 compares the $H = 0$ selector with positive-capacity WFE under matched candidate pools and training protocols. With regularization present, the fitted checkpoint trades the fit advantage against the parameter norm, which Proposition E.13 below bounds explicitly.

E.5 Confirmation Preferences and Preference Fitting

This subsection identifies the object learned by WFE. The confirmation temperature preserves candidate order while controlling target sharpness. The listwise cross-entropy is exactly a reliability-weighted KL projection onto these event-level target distributions. This gives a direct bridge from the confirmation utilities to the fitted WFE scores.

Lemma E.9 (Preference geometry of the confirmation softmax)**.** *For every valid event and every candidate pair k, ℓ,*

$$\log \frac{Q_e(k)}{Q_e(\ell)} = \frac{c_{ek} - c_{e\ell}}{\tau_c}. \tag{70}$$

Hence

$$c_{ek} > c_{e\ell} \quad \Longleftrightarrow \quad Q_e(k) > Q_e(\ell), \tag{71}$$

and equality is preserved as well. A fully tied utility vector gives $Q_e(k) = 1/K$ for every k. If the utility maximum is unique at $k_e^\star$, then

$$\lim_{\tau_c \to 0^+} Q_e(k_e^\star) = 1, \qquad \lim_{\tau_c \to \infty} Q_e(k) = \frac{1}{K}. \tag{72}$$

Proof. Taking the ratio of two coordinates in Equation (33) cancels the shared normalization and gives Equation (70). Since $\tau_c > 0$, the ratio is greater than, equal to, or less than one exactly when the corresponding utility difference has the same sign. The tied case follows immediately.

For a unique maximum, divide numerator and denominator by $\exp(c_{e,k_e^\star}/\tau_c)$. Every competing term becomes

$$\exp\left(-\frac{c_{e,k_e^\star} - c_{e\ell}}{\tau_c}\right) \tag{73}$$

and tends to zero as $\tau_c \to 0^+$. For $\tau_c \to \infty$, every exponent c_{ek}/τ_c tends to zero, so every unnormalized weight tends to one. □

The temperature therefore controls how strongly later confirmation differences are expressed in the listwise target while preserving their ranking.

Theorem E.10 (Exact matching is equivalent to matching confirmation-utility differences)**.** *Fix a valid event and finite WFE scores. The following statements are equivalent:*

(1) $p_\theta(k \mid e) = Q_e(k)$ *for every* k.

(2) For every k, ℓ,

$$S_\theta(e,k) - S_\theta(e,\ell) = \frac{c_{ek} - c_{e\ell}}{\tau_c}. \tag{74}$$

(3) There exists a scalar $b_e \in \mathbb{R}$ *such that*

$$S_\theta(e,k) = \frac{c_{ek}}{\tau_c} + b_e \qquad \text{for every } k. \tag{75}$$

(4) For every k, ℓ,

$$F_\theta(e,k) - F_\theta(e,\ell) = -\frac{c_{ek} - c_{e\ell}}{\tau_c}. \tag{76}$$

Proof. Assume statement (1). Softmax probabilities are strictly positive, so

$$\log \frac{p_\theta(k \mid e)}{p_\theta(\ell \mid e)} = S_\theta(e,k) - S_\theta(e,\ell). \tag{77}$$

Applying Lemma E.9 gives statement (2).

Assume statement (2) and fix a reference candidate k_0. Define

$$b_e = S_\theta(e,k_0) - \frac{c_{ek_0}}{\tau_c}. \tag{78}$$

Taking $\ell = k_0$ in Equation (74) and rearranging gives statement (3).

Assume statement (3). Substituting Equation (75) into the candidate softmax yields

$$p_\theta(k \mid e) = \frac{e^{c_{ek}/\tau_c + b_e}}{\sum_\ell e^{c_{e\ell}/\tau_c + b_e}} = Q_e(k), \tag{79}$$

which recovers statement (1). Finally, $F_\theta = -S_\theta$, so statements (2) and (4) are equivalent. □

The shared offset b_e is the expected softmax invariance: event-level score differences are identifiable, while adding the same constant to every candidate leaves the distribution unchanged.

Proposition E.11 (Reliability-weighted cross-entropy is a KL projection)**.** *Define the per-event cross-entropy*

$$\ell_e(\theta) = -\sum_{k=1}^{K} Q_e(k) \log p_\theta(k \mid e), \tag{80}$$

the target entropy

$$\mathsf{H}(Q_e) = -\sum_{k=1}^{K} Q_e(k) \log Q_e(k), \tag{81}$$

and the eventwise preference-fit error

$$\kappa_e(\theta) = \mathrm{KL}\left(Q_e \| p_\theta(\cdot \mid e)\right). \tag{82}$$

Let

$$\pi_e = \frac{\omega_e}{\Omega_{\boldsymbol{\xi}}}, \qquad \sum_{e \in D_{\mathrm{fit}}^+} \pi_e = 1. \tag{83}$$

Then the WFE objective of Equation (6) *can be written exactly as*

$$\mathcal{L}_{\mathrm{WFE}}(\theta; \boldsymbol{\xi}) = C_{\boldsymbol{\xi}} + \widehat{\mathcal{K}}_{\boldsymbol{\xi}}(\theta) + \lambda_{\mathrm{reg}} \|\theta\|_2^2, \tag{84}$$

where

$$C_{\boldsymbol{\xi}} = \sum_e \pi_e \mathsf{H}(Q_e), \qquad \widehat{\mathcal{K}}_{\boldsymbol{\xi}}(\theta) = \sum_e \pi_e \kappa_e(\theta) \geq 0. \tag{85}$$

Moreover, $\widehat{\mathcal{K}}_{\boldsymbol{\xi}}(\theta) = 0$ *exactly when* $p_\theta(\cdot \mid e) = Q_e$ *for every fitting event with positive weight.*

Proof. For each event,

$$\begin{aligned} \ell_e(\theta) &= -\sum_k Q_e(k) \log Q_e(k) + \sum_k Q_e(k) \log \frac{Q_e(k)}{p_\theta(k \mid e)} \\ &= \mathsf{H}(Q_e) + \kappa_e(\theta). \end{aligned} \tag{86}$$

KL divergence is nonnegative. One direct proof uses $\log x \leq x - 1$:

$$\mathrm{KL}(Q \| p) = -\sum_k Q_k \log \frac{p_k}{Q_k} \geq \sum_k Q_k \left(1 - \frac{p_k}{Q_k}\right) = 0. \tag{87}$$

Equality holds only when $p_k / Q_k = 1$ for every k. Multiplying the per-event identity by π_e and summing gives Equations (84) and (85). □

The data term therefore fits later confirmation preferences in a proper listwise sense, with confirmation reliability controlling each event's contribution. We write $\widehat{\theta}_{\boldsymbol{\xi}}$ for the checkpoint returned by the common fitting protocol under configuration $\boldsymbol{\xi}$, and $\theta^\dagger_{\boldsymbol{\xi}}$ for any global minimizer of the mathematical objective.

Proposition E.12 (Exact gradient of the listwise data term)**.** *Let* $s_{ek} = S_\theta(e, k)$. *Then*

$$\frac{\partial \ell_e}{\partial s_{ek}} = p_\theta(k \mid e) - Q_e(k). \tag{88}$$

Consequently,

$$\nabla_\theta \mathcal{L}_{\mathrm{WFE}}(\theta; \boldsymbol{\xi}) = \sum_e \pi_e \sum_k \left[p_\theta(k \mid e) - Q_e(k)\right] \nabla_\theta S_\theta(e, k) + 2\lambda_{\mathrm{reg}} \theta. \tag{89}$$

Proof. The cross-entropy can be written as

$$\ell_e = \log \sum_j e^{s_{ej}} - \sum_j Q_e(j) s_{ej}, \tag{90}$$

because $\sum_j Q_e(j) = 1$. Differentiation with respect to s_{ek} gives Equation (88). Applying the chain rule and the normalized event weights gives Equation (89). □

Equation (89) shows how later confirmation preferences train the early-feedback score function: the update is driven by the difference between the current candidate probability and the confirmation target.

Proposition E.13 (The regularized WFE objective attains a global minimum)**.** *Fix a configuration with finite H, finite fitting data, Assumption E.2, and $\lambda_{\mathrm{reg}} > 0$. Then*

$$\arg\min_{\theta} \mathcal{L}_{\mathrm{WFE}}(\theta; \boldsymbol{\xi}) \neq \varnothing. \tag{91}$$

Every global minimizer $\theta^{\dagger}_{\boldsymbol{\xi}}$ satisfies

$$\|\theta^{\dagger}_{\boldsymbol{\xi}}\|_2 \leq \sqrt{\frac{\log K}{\lambda_{\mathrm{reg}}}}. \tag{92}$$

More generally, any fitted checkpoint $\widehat{\theta}_{\boldsymbol{\xi}}$ satisfying

$$\mathcal{L}_{\mathrm{WFE}}(\widehat{\theta}_{\boldsymbol{\xi}}; \boldsymbol{\xi}) \leq \log K + \varepsilon_{\mathrm{chk}} \tag{93}$$

obeys

$$\|\widehat{\theta}_{\boldsymbol{\xi}}\|_2 \leq \sqrt{\frac{\log K + \varepsilon_{\mathrm{chk}}}{\lambda_{\mathrm{reg}}}}. \tag{94}$$

Proof. The objective is continuous in the finite-dimensional parameter vector. Since cross-entropy is nonnegative,

$$\mathcal{L}_{\mathrm{WFE}}(\theta; \boldsymbol{\xi}) \geq \lambda_{\mathrm{reg}} \|\theta\|_2^2. \tag{95}$$

At $\theta = 0$, every candidate receives the same score. Hence $p_0(k \mid e) = 1/K$ and

$$\mathcal{L}_{\mathrm{WFE}}(0; \boldsymbol{\xi}) = \log K. \tag{96}$$

The sublevel set

$$\{\theta : \mathcal{L}_{\mathrm{WFE}}(\theta; \boldsymbol{\xi}) \leq \log K\} \tag{97}$$

is nonempty, closed, and bounded by Equation (95). It is compact in finite dimension. A continuous function attains its minimum on this set, and every point outside the set has objective value greater than $\log K$. Thus a global minimizer exists. For any minimizer,

$$\lambda_{\mathrm{reg}} \|\theta^{\dagger}_{\boldsymbol{\xi}}\|_2^2 \leq \mathcal{L}_{\mathrm{WFE}}(\theta^{\dagger}_{\boldsymbol{\xi}}; \boldsymbol{\xi}) \leq \log K, \tag{98}$$

which proves Equation (92). Equation (94) follows from the same lower bound applied to Equation (93). □

The existence result characterizes the mathematical objective. The experimental optimizer produces a concrete checkpoint $\widehat{\theta}_{\boldsymbol{\xi}}$, and Equation (94) turns its recorded objective value into a parameter-norm certificate when needed for the stability analysis below.

E.6 Selection Guarantees and Perturbation Stability

This subsection converts preference fitting into guarantees on the selected candidate. Two complementary certificates are useful. The first works in score space and makes the role of score alignment and perturbation explicit. The second starts from the listwise KL error already present in the WFE objective and bounds the selected candidate's confirmation-utility regret directly.

For event e, define the confirmation-optimal candidate set

$$\mathcal{K}^{\star}_e = \arg\max_{1 \leq k \leq K} c_{ek}, \tag{99}$$

and the confirmation-utility regret of candidate j by

$$\mathcal{R}_e(j) = \max_k c_{ek} - c_{ej}. \tag{100}$$

By Proposition E.1,

$$0 \leq \mathcal{R}_e(j) \leq \Delta_e \leq 1. \tag{101}$$

Score alignment after removing the softmax offset. Theorem E.10 shows that exact preference matching is equivalent to matching c_{ek}/τ_c up to an event-specific constant. This suggests the following alignment error:

$$\varepsilon_e^{(\mathrm{align})} = \inf_{b\in\mathbb{R}} \max_k \left| S_\theta(e,k) - \frac{c_{ek}}{\tau_c} - b \right|. \tag{102}$$

Lemma E.14 (Closed form of the alignment error)**.** *Let*

$$\varsigma_{ek} = S_\theta(e,k) - \frac{c_{ek}}{\tau_c}, \qquad \varsigma_e^{\max} = \max_k \varsigma_{ek}, \qquad \varsigma_e^{\min} = \min_k \varsigma_{ek}. \tag{103}$$

Then

$$b_e^\star = \frac{\varsigma_e^{\max} + \varsigma_e^{\min}}{2} \tag{104}$$

attains the infimum in Equation (102)*, and*

$$\varepsilon_e^{(\mathrm{align})} = \frac{\varsigma_e^{\max} - \varsigma_e^{\min}}{2}. \tag{105}$$

In particular,

$$\varepsilon_e^{(\mathrm{align})} = 0 \quad \Longleftrightarrow \quad p_\theta(\cdot \mid e) = Q_e. \tag{106}$$

Proof. For any scalar b,

$$\varsigma_e^{\max} - \varsigma_e^{\min} \le |\varsigma_e^{\max} - b| + |\varsigma_e^{\min} - b| \le 2 \max_k |\varsigma_{ek} - b|. \tag{107}$$

Hence the objective in Equation (102) is at least $(\varsigma_e^{\max} - \varsigma_e^{\min})/2$. Choosing the midpoint in Equation (104) makes every ς_{ek} lie within this radius, so the lower bound is attained. The error is zero exactly when every ς_{ek} equals the same constant, which is statement (3) of Theorem E.10. □

Theorem E.15 (Score alignment controls confirmation-utility regret)**.** *Let $\widetilde{S}_{ek}$ be the scores used for selection and suppose*

$$\max_k |\widetilde{S}_{ek} - S_\theta(e,k)| \le \eta_e, \qquad \eta_e \ge 0. \tag{108}$$

Let

$$\widetilde{k}_e \in \arg\max_k \widetilde{S}_{ek}. \tag{109}$$

Then

$$\mathcal{R}_e(\widetilde{k}_e) \le \min\left\{ \Delta_e, 2\tau_c \left(\varepsilon_e^{(\mathrm{align})} + \eta_e \right) \right\}. \tag{110}$$

Proof. Choose $k_e^\star \in \mathcal{K}_e^\star$. By Lemma E.14, for every candidate k,

$$S_\theta(e,k) = \frac{c_{ek}}{\tau_c} + b_e^\star + \delta_{ek}, \qquad |\delta_{ek}| \le \varepsilon_e^{(\mathrm{align})}. \tag{111}$$

Therefore

$$\frac{c_{ek_e^\star} - c_{e\widetilde{k}_e}}{\tau_c} \le S_\theta(e,k_e^\star) - S_\theta(e,\widetilde{k}_e) + 2\varepsilon_e^{(\mathrm{align})}. \tag{112}$$

Because $\widetilde{k}_e$ maximizes the perturbed score,

$$\widetilde{S}_{ek_e^\star} - \widetilde{S}_{e\widetilde{k}_e} \le 0. \tag{113}$$

Using Equation (108) at the two compared candidates gives

$$S_\theta(e,k_e^\star) - S_\theta(e,\widetilde{k}_e) \le 2\eta_e. \tag{114}$$

Combining Equations (112) and (114) and multiplying by τ_c proves the second bound in Equation (110). The first bound follows from Equation (101). □

Theorem E.15 separates two sources of decision error. The alignment term measures how closely the fitted score differences reproduce later confirmation-utility differences. The perturbation term measures how much the score changes after the representation or numerical computation changes.

Lipschitz control of WFE scores. The perturbation term in Theorem E.15 can be bounded from the WFE parameters.

Proposition E.16 (Candidate-representation Lipschitz bound)**.** *Fix θ and event context $\boldsymbol{g}_e$. Define*

$$\Lambda_e^{(\phi)} = \|\boldsymbol{a}_e\|_2 + \sum_{j=1}^{H} \|W_{\cdot j}\|_2. \tag{115}$$

For any two candidate representations $\boldsymbol{\phi}$ and $\widetilde{\boldsymbol{\phi}}$,

$$\left| S_\theta(\widetilde{\boldsymbol{\phi}}, \boldsymbol{g}_e) - S_\theta(\boldsymbol{\phi}, \boldsymbol{g}_e) \right| \le \Lambda_e^{(\phi)} \|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_2. \tag{116}$$

Moreover,

$$\Lambda_e^{(\phi)} \le \sqrt{1 + \|\boldsymbol{g}_e\|_2^2 + H}\, \|\theta\|_2. \tag{117}$$

Proof. The derivative of softplus is the logistic function and lies in $(0, 1)$. The mean-value theorem therefore gives

$$|\operatorname{softplus}(u) - \operatorname{softplus}(v)| \le |u - v|. \tag{118}$$

Subtract the two WFE scores and apply Cauchy–Schwarz to the visible term and each hidden term:

$$\begin{aligned} |S_\theta(\widetilde{\boldsymbol{\phi}}, \boldsymbol{g}_e) - S_\theta(\boldsymbol{\phi}, \boldsymbol{g}_e)| &\le |\boldsymbol{a}_e^\top (\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi})| + \sum_j |W_{\cdot j}^\top (\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi})| \\ &\le \Lambda_e^{(\phi)} \|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_2. \end{aligned} \tag{119}$$

For Equation (117),

$$\|\boldsymbol{a}_e\|_2 = \|\boldsymbol{a}_0 + A\boldsymbol{g}_e\|_2 \le \sqrt{\|\boldsymbol{a}_0\|_2^2 + \|A\|_F^2} \sqrt{1 + \|\boldsymbol{g}_e\|_2^2}, \tag{120}$$

and

$$\sum_j \|W_{\cdot j}\|_2 \le \sqrt{H}\, \|W\|_F. \tag{121}$$

Applying Cauchy–Schwarz once more to these two terms yields Equation (117). □

If

$$\max_k \|\widetilde{\boldsymbol{\phi}}_{ek} - \boldsymbol{\phi}_{ek}\|_2 \le \delta_e^{(\phi)}, \tag{122}$$

then Equation (108) holds with

$$\eta_e = \Lambda_e^{(\phi)} \delta_e^{(\phi)}. \tag{123}$$

Combining Proposition E.16 with the checkpoint norm bound in Equation (94) gives the explicit certificate

$$\Lambda_e^{(\phi)} \le \sqrt{\frac{(1 + \|\boldsymbol{g}_e\|_2^2 + H)(\log K + \varepsilon_{\mathrm{chk}})}{\lambda_{\mathrm{reg}}}}. \tag{124}$$

The same argument can include context perturbations. For two input pairs $(\boldsymbol{\phi}, \boldsymbol{g})$ and $(\widetilde{\boldsymbol{\phi}}, \widetilde{\boldsymbol{g}})$,

$$\begin{aligned} |S_\theta(\widetilde{\boldsymbol{\phi}}, \widetilde{\boldsymbol{g}}) - S_\theta(\boldsymbol{\phi}, \boldsymbol{g})| \le &\Lambda^{(\phi)}(\boldsymbol{g}) \|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_2 \\ &+ \Lambda^{(g)}(\widetilde{\boldsymbol{\phi}}) \|\widetilde{\boldsymbol{g}} - \boldsymbol{g}\|_2, \end{aligned} \tag{125}$$

with

$$\Lambda^{(g)}(\widetilde{\boldsymbol{\phi}}) = \|A^\top \widetilde{\boldsymbol{\phi}}\|_2 + \sum_{j=1}^{H} \|(B_g)_{j\cdot}\|_2. \tag{126}$$

Equation (125) follows by inserting the intermediate score $S_\theta(\widetilde{\boldsymbol{\phi}}, \boldsymbol{g})$ and applying the same one-Lipschitz argument to the context-dependent affine terms.

Theorem E.17 (Positive energy gap certifies selection stability)**.** *Let k_e^S be the unique maximum-score candidate and define the score margin*

$$\mathsf{Gap}_e = \min_{\ell \neq k_e^S} \left[S_\theta(e, k_e^S) - S_\theta(e, \ell) \right] > 0. \tag{127}$$

Equivalently, Gap_e is the gap between the lowest and second-lowest free energies. If the perturbed scores satisfy Equation (108) *and*

$$2\eta_e < \mathsf{Gap}_e, \tag{128}$$

then k_e^S remains the unique selected candidate after perturbation.

Proof. For any competitor $\ell \neq k_e^S$,

$$\begin{aligned} \widetilde{S}_{ek_e^S} - \widetilde{S}_{e\ell} &\geq S_\theta(e, k_e^S) - S_\theta(e, \ell) - 2\eta_e \\ &\geq \mathsf{Gap}_e - 2\eta_e > 0. \end{aligned} \tag{129}$$

The inequality holds for every competitor, so k_e^S remains the unique maximizer. □

The margin condition gives a direct interpretation to the energy-gap perturbation analysis in Figure 4b: the measured gap is the empirical quantity that determines how much score perturbation the selected index can absorb before it changes.

From listwise KL error to confirmation-utility regret. The alignment certificate is useful when score differences are inspected directly. The next theorem connects the WFE training loss to candidate utility through the eventwise KL error $\kappa_e(\theta)$.

Lemma E.18 (KL cost of reversing a target pair)**.** *Let Q and p be strictly positive probability vectors. Suppose $Q_i > Q_j$ while $p_j \geq p_i$. Define $s = Q_i + Q_j$. Then*

$$\mathrm{KL}(Q\|p) \geq \mathcal{B}(Q_i, Q_j), \tag{130}$$

where

$$\mathcal{B}(a, b) = a \log \frac{2a}{a+b} + b \log \frac{2b}{a+b}. \tag{131}$$

Moreover,

$$\mathcal{B}(a, b) \geq \frac{(a-b)^2}{2(a+b)}. \tag{132}$$

Proof. Define a probability vector p° by

$$p_i^\circ = p_j^\circ = s/2, \qquad p_k^\circ = Q_k \quad (k \notin \{i, j\}). \tag{133}$$

Direct expansion gives

$$\mathrm{KL}(Q\|p) = \mathcal{B}(Q_i, Q_j) + \mathrm{KL}(p^\circ\|p) + \frac{Q_i - Q_j}{2} \log \frac{p_j}{p_i}. \tag{134}$$

The final two terms are nonnegative because $p_j \geq p_i$, which proves Equation (130).

For Equation (132), write $r = a/(a+b) \in (1/2, 1)$ and

$$f(r) = r \log(2r) + (1-r) \log(2(1-r)). \tag{135}$$

Then $f(1/2) = f'(1/2) = 0$ and

$$f''(r) = \frac{1}{r(1-r)} \geq 4. \tag{136}$$

Integrating this lower bound twice from $1/2$ to r yields

$$f(r) \geq 2(r - 1/2)^2. \tag{137}$$

Multiplying by $a+b$ and substituting $r - 1/2 = (a-b)/[2(a+b)]$ gives Equation (132). □

Theorem E.19 (Listwise KL error controls confirmation-utility regret)**.** *Let $\widehat{k}_e$ be the WFE-selected candidate. Then*

$$\mathcal{R}_e(\widehat{k}_e) \le \min\left\{\Delta_e, \frac{\sqrt{2K}}{\tanh(1/(2\tau_c))}\sqrt{\kappa_e(\theta)}\right\}. \tag{138}$$

Equivalently, any positive regret $\mathcal{R}_e(\widehat{k}_e)$ requires

$$\kappa_e(\theta) \ge \frac{\tanh^2(1/(2\tau_c))}{2K}\,\mathcal{R}_e(\widehat{k}_e)^2. \tag{139}$$

Proof. If $\mathcal{R}_e(\widehat{k}_e) = 0$, the statement follows. Suppose $\mathcal{R}_e(\widehat{k}_e) > 0$. Choose $i \in \mathcal{K}_e^\star$, let $j = \widehat{k}_e$, and write

$$\Delta_e(i,j) = c_{ei} - c_{ej} = \mathcal{R}_e(\widehat{k}_e) > 0. \tag{140}$$

Lemma E.9 gives $Q_e(i) > Q_e(j)$. Since j maximizes the WFE probability, $p_\theta(j \mid e) \ge p_\theta(i \mid e)$. Lemma E.18 therefore applies with $s = Q_e(i) + Q_e(j)$, and Equation (132) gives

$$\kappa_e(\theta) \ge \frac{s}{2}\left(\frac{Q_e(i) - Q_e(j)}{Q_e(i) + Q_e(j)}\right)^2. \tag{141}$$

The target odds from Lemma E.9 give

$$\frac{Q_e(i) - Q_e(j)}{Q_e(i) + Q_e(j)} = \tanh\left(\frac{\Delta_e(i,j)}{2\tau_c}\right). \tag{142}$$

Candidate i has maximum target probability, so $Q_e(i) \ge 1/K$ and therefore $s \ge 1/K$. Combining Equations (141) and (142),

$$\kappa_e(\theta) \ge \frac{1}{2K}\tanh^2\left(\frac{\Delta_e(i,j)}{2\tau_c}\right). \tag{143}$$

Proposition E.1 gives $0 \le \Delta_e(i,j) \le 1$. The function $\tanh x$ is concave and increasing on $[0,\infty)$ with $\tanh 0 = 0$. Hence

$$\tanh\left(\frac{\Delta_e(i,j)}{2\tau_c}\right) \ge \Delta_e(i,j)\tanh\left(\frac{1}{2\tau_c}\right). \tag{144}$$

Substitution into Equation (143) gives Equation (139), and rearranging gives Equation (138). □

Theorem E.19 makes the listwise objective decision-relevant: small KL error forces the selected candidate to have confirmation utility close to the best candidate available in the committed pool.

Corollary E.20 (Eventwise certificate for exact confirmation-optimal selection)**.** *Assume the confirmation-optimal candidate $k_e^\star$ is unique and define*

$$b_e^{(\mathrm{ord})} = \min_{j \ne k_e^\star} \mathcal{B}(Q_e(k_e^\star), Q_e(j)). \tag{145}$$

Then $b_e^{(\mathrm{ord})} > 0$, and

$$\kappa_e(\theta) < b_e^{(\mathrm{ord})} \quad \Longrightarrow \quad \widehat{k}_e = k_e^\star. \tag{146}$$

Proof. If WFE selected some $j \ne k_e^\star$, then $p_\theta(j \mid e) \ge p_\theta(k_e^\star \mid e)$. Lemma E.18 would imply

$$\kappa_e(\theta) \ge \mathcal{B}(Q_e(k_e^\star), Q_e(j)) \ge b_e^{(\mathrm{ord})}, \tag{147}$$

contradicting the assumed strict inequality. □

Corollary E.21 (Aggregate utility bound under fixed event weights)**.** *Let $\mathcal{E}$ be a finite set of valid events with fixed weights $\mu_e \ge 0$ and $\sum_{e\in\mathcal{E}} \mu_e = 1$. Define*

$$J_\mu(\theta) = \sum_{e\in\mathcal{E}} \mu_e c_{e,\widehat{k}_e}, \tag{148}$$

$$J_\mu^{\mathrm{oracle}} = \sum_{e\in\mathcal{E}} \mu_e \max_k c_{ek}, \tag{149}$$

and

$$\mathcal{K}_\mu(\theta) = \sum_{e\in\mathcal{E}} \mu_e \kappa_e(\theta). \tag{150}$$

Then

$$0 \le J_\mu^{\mathrm{oracle}} - J_\mu(\theta) \le \min\left\{ \sum_e \mu_e \Delta_e, C_{\tau_c,K}\sqrt{\mathcal{K}_\mu(\theta)} \right\}, \tag{151}$$

where

$$C_{\tau_c,K} = \frac{\sqrt{2K}}{\tanh(1/(2\tau_c))}. \tag{152}$$

Proof. Apply Theorem E.19 eventwise, multiply by μ_e, and sum. The KL term satisfies

$$\sum_e \mu_e \sqrt{\kappa_e} \le \sqrt{\sum_e \mu_e \kappa_e} \tag{153}$$

by Cauchy–Schwarz and $\sum_e \mu_e = 1$. The spread term follows from $\mathcal{R}_e \le \Delta_e$. □

For the WFE fitting set, choose $\mu_e = \pi_e$. Proposition E.11 then gives

$$\mathcal{K}_\pi(\theta) = \mathcal{L}_{\mathrm{WFE}}(\theta;\boldsymbol{\xi}) - C_{\boldsymbol{\xi}} - \lambda_{\mathrm{reg}}\|\theta\|_2^2. \tag{154}$$

Equations (151) and (154) provide the direct finite-sample link from the fitted WFE objective to the reliability-weighted confirmation utility of its selected candidates on the same event set.

Unified certificate under score perturbation. The KL and margin views can be combined. Let $\widetilde{p}_e = \mathrm{softmax}(\widetilde{S}_e)$ for scores satisfying Equation (108).

Lemma E.22 (Score perturbation changes eventwise KL by at most $2\eta_e$)**.** *For fixed target Q_e,*

$$|\mathrm{KL}(Q_e\|\widetilde{p}_e) - \mathrm{KL}(Q_e\|p_\theta(\cdot \mid e))| \le 2\eta_e. \tag{155}$$

Proof. The uniform score bound gives

$$\left| \log \sum_k e^{\widetilde{S}_{ek}} - \log \sum_k e^{S_{ek}} \right| \le \eta_e. \tag{156}$$

It also gives

$$\left| \sum_k Q_e(k)(\widetilde{S}_{ek} - S_{ek}) \right| \le \eta_e. \tag{157}$$

Cross-entropy equals log-sum-exp minus the target-weighted score. Its change is therefore at most $2\eta_e$. The target entropy is fixed, so KL changes by the same amount. □

Theorem E.23 (Unified WFE selection certificate)**.** *Under Equation* (108)*, every maximum perturbed-score candidate $\widetilde{k}_e$ satisfies*

$$\mathcal{R}_e(\widetilde{k}_e) \le \min\left\{ \Delta_e,\ 2\tau_c(\varepsilon_e^{(\mathrm{align})} + \eta_e),\ C_{\tau_c,K}\sqrt{\kappa_e(\theta) + 2\eta_e} \right\}. \tag{158}$$

If the original score margin is positive and $2\eta_e < \mathsf{Gap}_e$, then the selected index is unchanged.

Proof. The alignment term is Theorem E.15. Apply Theorem E.19 to the perturbed distribution $\widetilde{p}_e$, whose maximizer is $\widetilde{k}_e$, then use Lemma E.22 to bound its KL by $\kappa_e(\theta) + 2\eta_e$. The final statement is Theorem E.17. □

Theorem E.23 is the central finite-candidate guarantee. It connects the design quantities that are directly observable in WISER: confirmation spread, listwise fit, energy gap, and representation sensitivity. Figure 4 measures the energy-gap side of this certificate under evidence, relation, and noise perturbations. The alignment falsification analysis in Figure 3b complements the theorem by testing whether the learned score structure disappears when candidate-outcome correspondence is broken.

E.7 Constrained Configuration on Fixed Reference Populations

This subsection formalizes the quantities used by constrained BO. The fitting constraints regulate how much reliable supervision supports a configuration, while BO and validation compare configurations on fixed event sets with fixed confirmation weights. This separation makes configuration scores comparable even when $\rho_{\min}$ changes the fitting support.

Coverage and effective sample size. On D_{fit}^{+}, define

$$R_{\text{cov}}(\boldsymbol{\xi}) = \frac{\sum_e \omega_e}{\sum_e \rho_e^{(\text{conf})}}, \tag{159}$$

and

$$N_{\text{eff}}(\boldsymbol{\xi}) = \frac{\left(\sum_e \omega_e\right)^2}{\sum_e \omega_e^2}. \tag{160}$$

Let

$$\mathcal{I}_+(\boldsymbol{\xi}) = \{e : \omega_e > 0\}, \qquad n_+(\boldsymbol{\xi}) = |\mathcal{I}_+(\boldsymbol{\xi})|. \tag{161}$$

Proposition E.24 (Coverage and ESS bounds)**.** *For every fitted configuration,*

$$0 < R_{\text{cov}}(\boldsymbol{\xi}) \le 1, \tag{162}$$

and

$$1 \le N_{\text{eff}}(\boldsymbol{\xi}) \le n_+(\boldsymbol{\xi}). \tag{163}$$

If the positive weights are normalized as

$$\pi_e = \frac{\omega_e}{\sum_j \omega_j}, \tag{164}$$

then

$$N_{\text{eff}}(\boldsymbol{\xi}) = \frac{1}{\sum_e \pi_e^2}, \tag{165}$$

and the largest normalized event weight satisfies

$$\max_e \pi_e \le \frac{1}{\sqrt{N_{\text{eff}}(\boldsymbol{\xi})}}. \tag{166}$$

Proof. Because $\omega_e = \rho_e^{(\text{conf})} \mathbb{I}[\rho_e^{(\text{conf})} \ge \rho_{\min}]$,

$$0 \le \omega_e \le \rho_e^{(\text{conf})}. \tag{167}$$

The denominator of Equation (159) is positive on D_{fit}^{+}, and Assumption E.2 gives a positive numerator. This gives Equation (162).

For ESS, the lower bound follows from

$$\left(\sum_e \omega_e\right)^2 = \sum_e \omega_e^2 + 2\sum_{i<j} \omega_i \omega_j \ge \sum_e \omega_e^2. \tag{168}$$

For the upper bound, Cauchy–Schwarz on the n_+ positive weights gives

$$\left(\sum_e \omega_e\right)^2 \le n_+ \sum_e \omega_e^2. \tag{169}$$

Equation (165) follows by normalization. Finally,

$$\max_e \pi_e^2 \le \sum_e \pi_e^2 = \frac{1}{N_{\text{eff}}}, \tag{170}$$

which proves Equation (166). □

The ESS constraint therefore limits event-weight concentration. Stronger filtering reduces retained support and lowers ESS, which is exactly the quantity controlled by Equation (163).

Fixed-population BO utility. The remaining feasibility statistics are also fixed by the main protocol. R_{prov} is the evidence-completeness rate among selected BO records, R_{dom} is the largest directional share, and $\kappa_{\mathrm{cov}}, \kappa_{\mathrm{eff}}, \kappa_{\mathrm{prov}}, \kappa_{\mathrm{dom}}$ are the prespecified feasibility thresholds.

Proposition E.25 (Fully tied events are selector-neutral under raw-utility evaluation)**.** *Suppose an evaluation event satisfies $\Delta_e = 0$. Then $c_{ek} = c_e$ for every k, so its weighted contribution to BO or validation utility is*

$$\bar{\omega}_e c_e, \tag{171}$$

independent of the selected candidate. The same event has the uniform confirmation target $Q_e(k) = 1/K$.

Proof. The condition $\Delta_e = 0$ means that the maximum and minimum confirmation utilities coincide. Every candidate therefore has the same utility c_e. Selecting any index k yields the same term $\bar{\omega}_e c_e$ in the fixed-population utility. Lemma E.9 gives $Q_e(k) = 1/K$ because all confirmation utilities are equal. □

Preference fitting omits these events because their target carries no within-event ordering signal. BO and validation retain them because they preserve the fixed reference population while contributing the same raw utility under every selector. This separates the role of a preference target from the role of a fixed evaluation record.

For market m, let $\mathcal{V}_{\mathrm{BO}}^m$ be the fixed BO evaluation set and let

$$\bar{\omega}_e = \rho_e^{(\mathrm{conf})} > 0 \tag{172}$$

be the fixed evaluation weight. The BO utility of Equation (8) is

$$J_{\mathrm{BO}}(\boldsymbol{\xi}) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \frac{\sum_{e \in \mathcal{V}_{\mathrm{BO}}^m} \bar{\omega}_e c_{e,\widehat{k}_e(\boldsymbol{\xi})}}{\sum_{e \in \mathcal{V}_{\mathrm{BO}}^m} \bar{\omega}_e}. \tag{173}$$

Every denominator is checked once and kept fixed across configurations.

Proposition E.26 (Boundedness and selector comparability of J_{BO})**.** *For every evaluated configuration with valid market denominators,*

$$0 \le J_{\mathrm{BO}}(\boldsymbol{\xi}) \le 1. \tag{174}$$

If two configurations select the same candidate for every BO event, then they have identical BO utility. More generally, define

$$d_e(\boldsymbol{\xi}, \boldsymbol{\xi}') = \mathbb{I}[\widehat{k}_e(\boldsymbol{\xi}) \neq \widehat{k}_e(\boldsymbol{\xi}')]. \tag{175}$$

Then

$$|J_{\mathrm{BO}}(\boldsymbol{\xi}) - J_{\mathrm{BO}}(\boldsymbol{\xi}')| \le \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \frac{\sum_{e \in \mathcal{V}_{\mathrm{BO}}^m} \bar{\omega}_e d_e(\boldsymbol{\xi}, \boldsymbol{\xi}')}{\sum_{e \in \mathcal{V}_{\mathrm{BO}}^m} \bar{\omega}_e}. \tag{176}$$

Proof. Each market term in Equation (173) is a positive weighted average of confirmation utilities in $[0, 1]$, so it lies in $[0, 1]$. Averaging across markets proves Equation (174).

For the second result, the contribution of event e is identical when the two configurations select the same candidate. When their selections differ, Proposition E.1 gives

$$|c_{e,\widehat{k}_e(\boldsymbol{\xi})} - c_{e,\widehat{k}_e(\boldsymbol{\xi}')}| \le 1. \tag{177}$$

Taking the weighted difference market by market gives Equation (176). □

Equation (176) makes the fixed-reference design explicit: configuration-score differences arise through different candidate choices on the same weighted events.

Corollary E.27 (BO utility regret is controlled by BO-set KL)**.** *Define normalized BO weights*

$$\mu_e^{\mathrm{BO}} = \frac{1}{|\mathcal{M}|} \frac{\bar{\omega}_e}{\sum_{e' \in \mathcal{V}_{\mathrm{BO}}^m} \bar{\omega}_{e'}} \quad \textit{for } e \in \mathcal{V}_{\mathrm{BO}}^m. \tag{178}$$

These weights sum to one. Define

$$\mathcal{K}_{\mathrm{BO}}(\boldsymbol{\xi}) = \sum_e \mu_e^{\mathrm{BO}} \, \mathrm{KL}\left(Q_e^{(\tau_c(\boldsymbol{\xi}))} \| p_{\widehat{\theta}_{\boldsymbol{\xi}}}(\cdot \mid e) \right), \tag{179}$$

and the same-population oracle utility

$$J_{\mathrm{BO}}^{\mathrm{oracle}} = \sum_e \mu_e^{\mathrm{BO}} \max_k c_{ek}. \tag{180}$$

Then

$$0 \le J_{\mathrm{BO}}^{\mathrm{oracle}} - J_{\mathrm{BO}}(\boldsymbol{\xi}) \le C_{\tau_c(\boldsymbol{\xi}),K} \sqrt{\mathcal{K}_{\mathrm{BO}}(\boldsymbol{\xi})}. \tag{181}$$

Proof. Equation (173) is exactly J_μ from Corollary E.21 under the weights in Equation (178). Apply Equation (151). □

The same construction applies to the fixed validation set. The BO and validation partitions therefore evaluate the decision quality of a fitted selector on event populations that remain fixed as the fitting threshold changes.

Proposition E.28 (Final configuration is well defined on the evaluated set)**.** *Let Ξ_{eval} be the finite set of completed configuration evaluations. Define the feasible subset by the main-text constraints of Equation* (7)*,*

$$\Xi_{\mathrm{feas}} = \left\{ \boldsymbol{\xi} \in \Xi_{\mathrm{eval}} : \begin{array}{l} R_{\mathrm{cov}}(\boldsymbol{\xi}) \ge \kappa_{\mathrm{cov}}, \\ N_{\mathrm{eff}}(\boldsymbol{\xi}) \ge \kappa_{\mathrm{eff}}, \\ R_{\mathrm{prov}}(\boldsymbol{\xi}) \ge \kappa_{\mathrm{prov}}, \\ R_{\mathrm{dom}}(\boldsymbol{\xi}) \le \kappa_{\mathrm{dom}}, \\ \sum_{e \in \mathcal{V}_{\mathrm{BO}}^m} \bar{\omega}_e > 0 \quad \forall m \in \mathcal{M} \end{array} \right\}. \tag{182}$$

Assume Ξ_{feas} is nonempty and every shortlisted configuration has a finite validation utility. Then the prespecified BO ranking, finite shortlist, one-time validation score, and deterministic tie rule return a unique final configuration $\boldsymbol{\xi}^\star$. The retained WFE parameters are the original fitting parameters $\theta^\star = \widehat{\theta}_{\boldsymbol{\xi}^\star}$.

Proof. A nonempty finite set has at least one maximizer of the finite BO utility. The shortlist rule retains finitely many feasible configurations. Each shortlisted configuration receives one finite validation value. The deterministic ranking and tie rule therefore select one index uniquely. Parameter retention is part of the selection rule, so $\theta^\star$ is uniquely determined by the chosen configuration and its stored fitting checkpoint. □

The constrained search in WISER is therefore a finite-budget allocation strategy over Ξ_{eval}, followed by a deterministic constrained choice from completed evaluations. Figure 3c compares search policies under the same completed-evaluation budget, while Equation (182) defines the common admissible region used by those comparisons.

E.8 Temporal Commitment and the Selection-Only Student Channel

This subsection formalizes the information path of WISER. Historical outcomes enter the final student through the candidate indices chosen by the frozen selector. Test responses are generated from a student frozen before the test events and are committed before the first scored return timestamp.

Candidate commitment. For a development event, the committed pool can be written as the deterministic mapping

$$\mathcal{C}_e = \mathcal{G}_{\vartheta, A_{\mathrm{adm}}}(x_e, \zeta_{e1:K_{\mathrm{raw}}}), \tag{183}$$

where the mapping includes the fixed generator revision, decoding rule, admissibility rule, and the FirstK operator of Equation (2). The signed manifest stores the resulting text, order, seeds, model revision, retrieval revision, and evidence cutoff before outcome-table access.

Proposition E.29 (Candidate texts are fixed before event outcomes enter selection)**.** *Conditioned on the event-time snapshot, frozen generator state, generation seeds, and admissibility rule, $\mathcal{C}_e$ is fixed by Equation* (183)*. Later selection and confirmation measurements can change the selected index, while the candidate texts and their committed order remain the manifest values.*

Proof. Every argument of $\mathcal{G}_{\vartheta, A_{\mathrm{adm}}}$ in Equation (183) is available at candidate-construction time. The FirstK rule is deterministic once the ordered raw responses and admissibility decisions are fixed. The manifest commits the output before the workflow reads the corresponding outcome table. Subsequent modules receive $\mathcal{C}_e$ as an input and choose among its elements. □

This proposition is a statement about the computational dependency graph enforced by retrospective replay. It is the formal basis for same-pool comparisons throughout the experiments.

Post-freeze non-anticipation. Let $\mathcal{F}_t^{\mathrm{pipe}}$ denote the information available to the WISER evaluation pipeline by time t, including frozen model states and timestamped event inputs. Write

$$D_{\mathrm{dev}}^{(\mathrm{retro})} = D_{\mathrm{cal}} \cup D_{\mathrm{fit}} \cup D_{\mathrm{BO}} \cup D_{\mathrm{val}} \cup D_{\mathrm{curate}}. \tag{184}$$

For a development event, $T_e^{(\mathrm{used})}$ is the latest availability time of feedback actually used by the pipeline. For a test event, $T_e^{(\mathrm{output})}$ is the response commitment time and a_e is the prespecified price timestamp that starts the first scored return interval, as in Methodology.

Assumption E.30 (Temporal protocol)**.** The protocol satisfies

$$\max_{e \in D_{\mathrm{dev}}^{(\mathrm{retro})}} T_e^{(\mathrm{used})} < T_{\mathrm{freeze}} < \min_{e \in D_{\mathrm{test}}} t_e, \tag{185}$$

and, for each test event,

$$t_e \le T_e^{(\mathrm{output})} < a_e. \tag{186}$$

Equations (185) and (186) are the last two lines of Equation (1).

Proposition E.31 (Test outputs are non-anticipating with respect to scored returns)**.** *Assume the final student $\psi^\star$ and test decoding rule are frozen at T_{freeze}, with any stochastic decoding seed fixed or recorded by the output time. Under Assumption E.30, for each $e \in D_{\mathrm{test}}$, the response*

$$\widehat{y}_e = \mathrm{Decode}(\psi^\star, x_e) \tag{187}$$

is measurable with respect to $\mathcal{F}^{\mathrm{pipe}}_{T_e^{(\mathrm{output})}}$. Since $T_e^{(\mathrm{output})} < a_e$, the scored return interval begins after the response is committed.

Proof. The frozen checkpoint is available before every test event by Equation (185). The event-time input becomes available at t_e, and the response is a function of the frozen checkpoint, the event-time input, and the fixed decoding rule. Equation (186) places this computation before the first scored price timestamp. Hence every variable used to generate the response belongs to the pipeline information set at the commitment time, which precedes the scored return interval. □

The proposition formalizes the timestamp guarantee used by the confirmatory test. Pretrained weights enter this statement as part of the frozen model state, exactly as they do in the replay protocol of Appendix B.3.1.

Selection-only path to the student. Let $n = |D_{\text{curate}}|$ and define the selected-index vector

$$\boldsymbol{I} = (I_1, \ldots, I_n) \in \{1, \ldots, K\}^n, \qquad I_e = \widehat{k}_e. \tag{188}$$

Let $\mathfrak{M}$ collect the outcome-blind curation manifest and the fixed student-training specification: event-time inputs, committed candidate texts and order, packing order, student initialization, optimizer settings, and token budget B_{tok}. Let O_{dev} denote the historical outcome-side records used by calibration, fitting, configuration, and curation selection. Let R_{train} collect the fixed or seeded randomness of student training.

The raw student records are

$$D_{\text{raw}} = \{(x_e, y_{eI_e}) : e \in D_{\text{curate}}\}. \tag{189}$$

The fixed packing operator produces

$$D_{\text{SFT}} = \text{TokenBudget}_{B_{\text{tok}}}(D_{\text{raw}}), \tag{190}$$

and the student objective of Equation (9) is

$$\mathcal{L}_{\text{SFT}}(\psi) = \sum_{j=1}^{|D_{\text{SFT}}|} \sum_{t \in \mathcal{T}_{B_{\text{tok}},j}} -\log f_\psi(y_{j,t} \mid x_j, y_{j,<t}), \tag{191}$$

where j indexes training occurrences and $\mathcal{T}_{B_{\text{tok}},j}$ contains their unmasked assistant positions; repetitions count separately toward B_{tok}.

The WISER pipeline factors as

$$O_{\text{dev}} \xrightarrow{\mathcal{S}} \boldsymbol{I} \xrightarrow{\mathcal{T}_{\mathfrak{M}}} D_{\text{SFT}} \xrightarrow{\text{Train}(\cdot; R_{\text{train}})} \psi^\star. \tag{192}$$

The middle map selects one committed response per event and applies the fixed token-budget operator. Market measurements, energies, selector probabilities, and configuration traces remain outside the student records.

Theorem E.32 (The final curation decisions form a finite outcome-to-student channel)**.** *Condition on $\mathfrak{M}$ and R_{train}. Then the final student checkpoint is a deterministic function of the selected-index vector:*

$$\psi^\star = \mathcal{T}(\boldsymbol{I}; \mathfrak{M}, R_{\text{train}}). \tag{193}$$

The number of student checkpoints reachable by varying the curation decisions is at most

$$K^n. \tag{194}$$

If $\mathsf{H}_2(\cdot)$ denotes Shannon entropy in bits, then

$$\mathsf{H}_2(\boldsymbol{I} \mid \mathfrak{M}, R_{\text{train}}) \le n \log_2 K. \tag{195}$$

Furthermore, the conditional mutual information between the historical outcome-side records and the final student satisfies

$$\mathsf{I}_2(O_{\text{dev}}; \psi^\star \mid \mathfrak{M}, R_{\text{train}}) \le n \log_2 K. \tag{196}$$

Proof. For fixed $\mathfrak{M}$, the selected indices determine the retained candidate texts. The fixed packing rule and B_{tok} determine D_{SFT}. Conditioning on R_{train} makes the optimization trajectory and final checkpoint deterministic, which proves Equation (193).

There are K^n possible index vectors, and a deterministic map cannot have more outputs than inputs. This proves Equation (194). Entropy is maximized by the uniform distribution on a finite support, so

$$\mathsf{H}_2(\boldsymbol{I} \mid \mathfrak{M}, R_{\text{train}}) \le \log_2 K^n = n \log_2 K. \tag{197}$$

For the information bound, Equation (193) gives

$$\mathsf{H}_2(\psi^\star \mid \boldsymbol{I}, \mathfrak{M}, R_{\text{train}}) = 0. \tag{198}$$

Using the chain rule for conditional mutual information,

$$\begin{aligned} &\mathsf{I}_2(O_{\text{dev}}; \boldsymbol{I}, \psi^\star \mid \mathfrak{M}, R_{\text{train}}) = \mathsf{I}_2(O_{\text{dev}}; \boldsymbol{I} \mid \mathfrak{M}, R_{\text{train}}), \\ &\mathsf{I}_2(O_{\text{dev}}; \boldsymbol{I}, \psi^\star \mid \mathfrak{M}, R_{\text{train}}) = \mathsf{I}_2(O_{\text{dev}}; \psi^\star \mid \mathfrak{M}, R_{\text{train}}) + \mathsf{I}_2(O_{\text{dev}}; \boldsymbol{I} \mid \psi^\star, \mathfrak{M}, R_{\text{train}}). \end{aligned} \tag{199}$$

The second term in the final line is nonnegative. Therefore

$$\mathsf{I}_2(O_{\text{dev}}; \psi^\star \mid \mathfrak{M}, R_{\text{train}}) \le \mathsf{I}_2(O_{\text{dev}}; \boldsymbol{I} \mid \mathfrak{M}, R_{\text{train}}) \le \mathsf{H}_2(\boldsymbol{I} \mid \mathfrak{M}, R_{\text{train}}), \tag{200}$$

and Equation (195) completes the proof. □

Theoretical property	Derived quantity	Empirical diagnostic
Reliability changes candidate ranking through hidden interactions	pairwise derivative in Equation (53); ranking reversal in Theorem E.6	Figure 3a reliability-conditioning analysis; Table 3 reliability ablation; Table 7 $H = 0$ versus WFE
Candidate-outcome correspondence drives preference learning	Q_e, κ_e, and score-utility matching in Theorem E.10	Figure 3b alignment falsification tests
Listwise fit controls candidate decision quality	eventwise regret bound in Equation (138); aggregate bound in Equation (151)	retrospective selection diagnostics and raw confirmation-utility evaluation in Appendix C.1
Positive energy margin stabilizes the selected index	Gap_e and perturbation radius in Theorem E.17	Figure 4b normalized energy-gap perturbation test
Regularization bounds parameter scale	Equations (92), (94), and (124)	recorded WFE objectives and optimization trajectory in Figure 4a
Reliability threshold trades supervision quality against support	R_{cov} and N_{eff} in Proposition E.24	Fitting-support constraints in Section 3.5
Configuration comparison uses a common reference population	Equation (173) and disagreement bound Equation (176)	Figure 3c equal-budget BO comparison
Historical outcomes reach the student through selected responses	channel factorization Equation (192) and Corollary E.33	information audit in Appendix B.3.3; Table 1 Panel A and Table 7 matched same-pool comparisons
Final predictions precede scored test returns	Assumption E.30 and Proposition E.31	post-freeze protocol in Appendix B.3.1; Figure 3d; Tables 1 and 2

Table 12: Theoretical properties, their derived quantities, and the empirical diagnostics that measure them. Each row pairs one result of this appendix with one protocol element or matched experiment in the paper.

The channel result formalizes the central design choice of WISER: outcome-side information reaches the student through which committed responses are retained. The semantic content already present in the candidate texts belongs to the locked manifest, while the additional outcome-dependent signal is carried by the selected indices.

Corollary E.33 (Matched same-pool student comparisons isolate the selection vector)**.** *Consider two curation methods A and B evaluated under the same manifest $\mathfrak{M}$ and the same paired student-training randomness R_{train}. Let their selected-index vectors be $\boldsymbol{I}^A$ and $\boldsymbol{I}^B$. Then*

$$\psi_A^\star = \mathcal{T}(\boldsymbol{I}^A; \mathfrak{M}, R_{\mathrm{train}}), \qquad \psi_B^\star = \mathcal{T}(\boldsymbol{I}^B; \mathfrak{M}, R_{\mathrm{train}}). \tag{201}$$

Consequently, under the matched protocol, the method-specific input to student training is the selected candidate index at each event. If $\boldsymbol{I}^A = \boldsymbol{I}^B$, the two methods produce the same packed SFT data and the same final checkpoint for the paired run.

Proof. Theorem E.32 expresses the final checkpoint as a deterministic function of the selection vector once $\mathfrak{M}$ and R_{train} are fixed. Applying Equation (193) to the two methods gives Equation (201). Equal selection vectors retain the same response text for every event; the common packing rule then produces identical D_{SFT}, and the paired training trajectory is identical. □

This corollary is the protocol-level basis for the same-pool comparisons in Table 1 Panel A and the selector-structure comparisons in Table 7: the generator outputs, event identities, student initialization, packing rule, optimizer protocol, and effective assistant-token budget are shared, while the curation rule determines which committed responses enter supervision.

E.9 Theoretical Predictions and Empirical Diagnostics

The preceding results produce observable quantities that are already measured by the experimental design. Table 12 records the connection between each theoretical property and its empirical diagnostic.

These links give the appendix a single role in the paper: each theorem explains a property that is either enforced by the protocol or measured by a matched experiment. The empirical sections then evaluate how strongly those properties are realized on the financial-event data.