Measure Less, Know More: Self-Supervised Test-Time Feature Acquisition

Eeshaan Jain[1] **Linus Bleistein**[1,2] **Bart Deplancke**[2] **Charlotte Bunne**[1,2]

[1]School of Computer and Communication Sciences, EPFL
[2]School of Life Sciences, EPFL
{eeshaan.jain, linus.bleistein, bart.deplancke, charlotte.bunne}@epfl.ch

Abstract

Recent progress in multimodal, high-dimensional learning has enabled foundation models to process heterogeneous, large-scale data. However, at test time, acquiring all features or modalities can be prohibitively costly and often redundant. Sequentially selecting informative modalities is therefore critical, yet challenging when the downstream task or prediction target is unknown. To this end, we introduce ECHO-k, a *task-agnostic* and *self-supervised* learning principle for modality acquisition: we use a deep model's *internal* pretrained representations (e.g., from a foundation model) as proxy targets that summarize cross-modal information. We provide theoretical guarantees in a stylized linear setting that motivate a reinforcement learning (RL) policy for sequential modality selection. Across task-agnostic and label-free acquisition baselines, ECHO-k consistently improves budgeted downstream performance across diverse foundation-model backends. Our method provides a principled route to cost-aware test-time deployment, with implications for any multimodal system where measurements are expensive or time-constrained, and downstream tasks unknown *a priori*. The code is available at github.com/bunnelab/echo-k.

1 Introduction

Multimodal foundation models (FMs) thrive on *more* information: the more modalities we can feed them at test time, the more accurate, robust, and context-aware their representations become (Kaplan et al., 2020; Zhai et al., 2022; Mizrahi et al., 2023; Shukor et al., 2025). In practice, however, not every modality is available by default. Each additional assay, sensor, or data source carries a tangible cost in time, money, throughput, or patient burden. This creates a deployment gap: while models can theoretically produce rich representations by exploiting *multiview* data, acquiring all views for every instance is infeasible.

Modalities are stratified by cost, latency, and availability: in clinical diagnostics, routine measurements are often available immediately, while additional molecular or imaging assays can disambiguate uncertain cases but are constrained by turnaround time, cost, or limited sample material. Similar asymmetries arise in scientific discovery, where inexpensive proxies can be collected at scale, but high-fidelity measurements require specialized instrumentation or scarce experimental time. Measurements are therefore not acquired exhaustively, nor independently, but as structured groups of modalities or features corresponding to coherent sensing processes. We refer to each such group as a *view*. Under a fixed budget, only a small number of views can be queried per instance, and the value of each view depends on which views have already been observed. The acquisition problem is therefore inherently sequential and adaptive: *at test time, given the views observed so far, which view should a foundation model acquire next to most improve its downstream predictions?*

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Existing approaches to sequential acquisition are typically *task*- and *data*-driven: they define utility with respect to a specific downstream objective, such as classification or regression, and optimize expected task loss (Gong et al., 2019; Gadgil et al., 2023; Guney et al., 2025). This paradigm requires labeled data at acquisition time, task-specific reward design, or explicit modeling of task uncertainty. As a result, acquisition policies are tightly coupled to a particular task and dataset, and must be retrained or redesigned when downstream objectives change. This reliance on supervision has left the fully task-agnostic, label-free acquisition setting largely unaddressed (see Related Work in § B). Yet modern pretrained models are increasingly deployed as frozen backbones whose representations support many downstream tasks. This shift suggests a different perspective on acquisition: rather than selecting measurements for a single supervised objective, one can ask which measurements are needed to preserve the representation formed by the pretrained model.

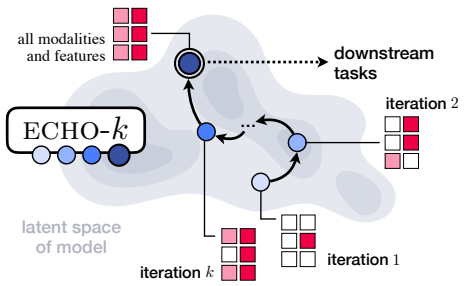


Figure 1: ECHO-k learns an RL policy to iteratively selects views to approximate the full latent representation learned by *any* pretrained model, *without* any knowledge of downstream tasks. Once learned, the policy can be deployed for budgeted feature acquisition.

In this work, we propose to use these latent representations learned by modern foundation models as a task-agnostic signal for view acquisition. When trained on heterogeneous, multiview data, these models learn compressed representations that integrate complementary evidence across views into a compact state (Radford et al., 2021; Siméoni et al., 2025; Wenckstern et al., 2025; Oquab et al., 2023; Lu et al., 2024; Qu et al., 2025; Shmatko et al., 2025). This latent state encodes the information the model deems relevant across many downstream tasks. We use these rich latent states as a learning objective: rather than defining utility with respect to a specific downstream objective, we learn a view selection policy that iteratively selects views to approximate the full latent representation from a minimal number of views.

Learning acquisition policies in representation spaces has two main advantages. On the one hand, it removes dependence on downstream supervision: acquisition decisions can be learned without task labels or task-specific losses. On the other hand, it enables broad transfer: because the representation is shared across tasks, preserving it supports many downstream objectives simultaneously. If the partial-view representation closely matches the full-view representation, then any downstream predictor operating on that representation should incur minimal degradation, independent of the task.

We instantiate this principle with ECHO-k, a reinforcement learning algorithm for long-horizon sequential view acquisition that modularly integrates with any self-supervised (foundation) model (Fig. 1). At each step, the RL agent selects which view to acquire next and the reward quantifies progress toward the corresponding full-view representation. Once trained, the policy is deployed at inference time on unobserved views to allocate acquisition budgets on an instance-by-instance basis.

We support this algorithm with the first theoretical analysis of self-supervised feature acquisition: in a linear model, where we show that representation recovery controls the latent representation regret (Theorem 3.5), and that conditioning on partial observations yields provable gains over fixed orderings. These results in turn justify the RL reward design and motivate the long-horizon acquisition strategy, enabling cost-effective deployment of large multimodal models in settings where full multiview acquisition is impractical.

Contributions. Our contributions are as follows.

1. We introduce a *model-driven* framework for self-supervised feature acquisition, in which acquisition decisions are guided by the internal representations of pretrained models, without access to task-specific supervision.
2. We provide the first theoretical analysis of self-supervised feature acquisition. In a stylized linear setting, we derive performance bounds that relate optimal test-time policies and oracle selection strategies, formally characterizing the gap induced by limited information that motivate an RL-based acquisition strategy.
3. We develop ECHO-k, a cost-efficient reinforcement learning algorithm that learns long-horizon acquisition policies using self-supervised latent representations. The policy is trained to maximize

progress toward the full-view representation, translating the representation-recovery principle into a practical sequential acquisition method.
4. Empirically, we show that ECHO-k learns effective acquisition strategies across diverse data modalities and acquisition horizons. Across image, tabular, single-cell transcriptomics, spatial proteomics, and multimodal geospatial benchmarks, the learned policies rapidly concentrate informative views in early acquisition stages, validating the effectiveness and generality of representation-driven acquisition under realistic budget constraints.

Our contributions bridge theory and practice in unsupervised feature acquisition, enabling foundation models to reason about what to measure next under realistic constraints.

2 Representation-Driven Acquisition as a Markov Decision Process

In the absence of a specified downstream task, acquisition requires a notion of value that is independent of labels. We define this value through the pretrained multiview encoder: acquired views should recover the representation obtained from all views.

Setup. Concretely, we consider n i.i.d. samples, each with K possible views $\boldsymbol{x}^i = (\boldsymbol{m}_1^i, \dots, \boldsymbol{m}_K^i)$ drawn from a joint distribution over the product space of multiple views $\mathcal{M}^1 \times \cdots \times \mathcal{M}^K$. [1] After selection, each view $\boldsymbol{m}_k$ is processed by a view-specific encoder $\Phi_k : \mathcal{M}^k \to \mathbb{R}^{d_k}$ [2] and combined via a fixed aggregation $\mathbf{A}_{\mathbf{W}} : \mathbb{R}^{d_1} \times \dots \mathbb{R}^{d_K} \to \mathbb{R}^p$ parameterized by $\mathbf{W}$, yielding the full-view representation $\boldsymbol{z}^* = \mathbf{A}_{\mathbf{W}}(\Phi_1(\boldsymbol{m}_1), \dots, \Phi_K(\boldsymbol{m}_K))$. This formulation subsumes both multimodal, where each view corresponds to a distinct sensing modality, and unimodal models in which views are partitioned into acquisition units but processed by a shared encoder. For notational simplicity, we absorb encoders into the views and write $\boldsymbol{z}^*$ directly. We let $\boldsymbol{P}_t \subseteq [\![1, K]\!]$ be the acquired set after t steps, with induced partial representation $\boldsymbol{z}(\boldsymbol{P}_t) := \mathbf{A}_{\mathbf{W}}(\{\Phi_k(\boldsymbol{m}_k) : k \in \boldsymbol{P}_t\})$.

Objective. We measure acquisition quality through the representation-recovery loss $\mathcal{L}(\boldsymbol{P}_t) := \|\boldsymbol{z}(\boldsymbol{P}_t) - \boldsymbol{z}^\star\|_2^2$, and seek a sequence $(k_1, \dots, k_T)$ minimizing $\mathcal{L}(\boldsymbol{P}_T)$ under budget $T \ll K$. The ideal one-step oracle greedy rule is

$$k_{t+1} \in \operatorname*{arg\,min}_{k \notin \boldsymbol{P}_t} \|\boldsymbol{z}(\boldsymbol{P}_t \cup \{k\}) - \boldsymbol{z}^*\|_2^2 . \tag{1}$$

From training to test-time deployment. The ideal policy defined above is not deployable at test time. Indeed, it requires both the full-view representation $\boldsymbol{z}^\star$ and forward passes for every candidate view. During *training*, we use unlabeled data to construct a self-supervised representation-recovery signal to train a view-selection policy. In the complete-reference setting, fully observed training samples allow us to compute the full-view representation $\boldsymbol{z}^\star$ and use it as the reference target. More generally, when only a subset of views $R_i \subseteq [\![1, K]\!]$ is available for training instance i, we can use the corresponding reference representation $\boldsymbol{z}(R_i)$ as a partially observed target. In both cases, downstream labels are not used to train the acquisition policy. At *deployment*, the learned policy acts solely from the current acquisition state $s_t = (\boldsymbol{P}_t, \boldsymbol{z}(\boldsymbol{P}_t), b_t)$; $\boldsymbol{z}^\star$ is never accessed.

Downstream tasks. A representation-recovery objective is only meaningful if minimizing it translates to better performance on downstream tasks. If g is an L_g-Lipschitz downstream predictor and the loss ℓ is L_ℓ-Lipschitz in the prediction, then for any acquired set $\boldsymbol{P}$,

$$\ell(g(\boldsymbol{z}(\boldsymbol{P})),\, y) - \ell(g(\boldsymbol{z}^*),\, y) \;\le\; L_\ell L_g \, \|\boldsymbol{z}(\boldsymbol{P}) - \boldsymbol{z}^*\|_2 .$$

This provides a simple sufficient condition under which latent recovery bounds the degradation of a fixed downstream predictor. In our experiments, downstream labels are not used to train the acquisition policy; they are used only post hoc to evaluate whether the recovered representations preserve task-relevant information.

The training-deployment asymmetry shown above raises the central theoretical question: how much is lost when acquisition proceeds from partial observations, relative to an oracle greedy policy that can evaluate the full-information value of each view? A complete MDP formulation is given in § C.

[1] Throughout, we use the term *view* to represent either a modality or a feature group acquired jointly, as our framework applies uniformly to all such settings.

[2] Note, under this framework, multiple views can also share the same joint encoder, or correspond to unimodal encoders.

3 Theoretical Guarantees

To quantify the cost of this partial observability, we compare two idealized strategies in a stylized linear setting: (i) an *oracle policy* that observes the full view vector $\boldsymbol{x}$ and greedily selects the view that most reduces representation error, and (ii) an *optimal test-time policy* that makes the same greedy choice but conditions only on previously acquired views. The gap between such oracle and limited-information strategies has been studied in supervised settings (He et al., 2012; Madasu et al., 2022; Valancius et al., 2023); our contribution is to characterize it in the self-supervised, representation-recovery setting, where no task labels are available.

Let $\boldsymbol{x} \in \mathbb{R}^d$ be a random view vector and $\boldsymbol{z}^* = \boldsymbol{z}(\boldsymbol{x}) := \mathbf{W}\boldsymbol{x} \in \mathbb{R}^m$ a latent representation where $\mathbf{W} = (\mathbf{w}_{ik}) \in \mathbb{R}^{m\times d}$ is a fixed linear aggregation operator. This setting captures the last-layer behavior of linear or locally linearized representation models. Policies iteratively construct a view subset $\boldsymbol{P}_t \subseteq [\![1, d]\!]$; at step t, the quality of the selected set $\boldsymbol{P}_t$ is measured by the conditional expected square reconstruction error

$$\mathcal{L}(\boldsymbol{P}_t \mid \mathcal{I}_t) := \sum_{i=1}^{m} \mathbb{E}\left[\Big(\sum_{k=1}^{d} \mathbf{w}_{ik}x_k - \sum_{k\in\boldsymbol{P}_t} \mathbf{w}_{ik}x_k\Big)^2 \Big|\, \mathcal{I}_t\right] = \sum_{i=1}^{m} \mathbb{E}\left[\Big(\sum_{k\notin\boldsymbol{P}_t} \mathbf{w}_{ik}x_k\Big)^2 \Big|\, \mathcal{I}_t\right],$$

where $\mathcal{I}_t$ denotes the information available to the policy at time t, which is different for each policy.

One-step oracle policy. The *one-step oracle policy* greedily selects views with full access to the realization of $\boldsymbol{x}$. Given a set $\boldsymbol{P}_t^\star$ selected at time t, the oracle chooses

$$k_t^\star \in \operatorname*{arg\,min}_{k\notin\boldsymbol{P}_t^\star} \sum_{i=1}^{m} \left(\sum_{j\notin\boldsymbol{P}_t^\star\cup\{k\}} \mathbf{w}_{ij}x_j\right)^2 = \operatorname*{arg\,min}_{k\notin\boldsymbol{P}_t^\star} \sum_{j\notin\boldsymbol{P}_t^*\cup\{k\}} \underbrace{\|\mathbf{w}_j\|_2^2 x_j^2}_{=:\boldsymbol{E}_j} + \sum_{\substack{j\neq j'\\ j,j'\notin\boldsymbol{P}_t*\cup\{k\}}} \underbrace{\mathbf{w}_j^\top \mathbf{w}_{j'} x_j x_{j'}}_{=:\boldsymbol{C}_{jj'}},$$

where $\mathbf{w}_k$ denotes the k-th column of $\mathbf{W}$. The oracle therefore exploits both individual view contributions $\boldsymbol{E}_k$ (energies) and cross-view interaction terms $C_{kk'}$, made possible by full knowledge of $\boldsymbol{x}$. A test-time policy, by contrast, cannot evaluate the $C_{kk'}$ terms, since they depend on the unobserved views; it must instead rely on conditional expectations given observed views.

Test-time policies. We define the class of *one-step test-time policies* as stochastic mappings that, at each step, select a new view based solely on the history of previously observed views and update their beliefs over the unobserved ones. Formally, such policies are adapted to the filtration $\mathcal{F}_t := \sigma(x_k : k \in \boldsymbol{P}_t)$ and therefore operate under partial observability, in contrast to the oracle. Among this class, the *optimal test-time policy* greedily minimizes the conditional expected representation loss, with perfect knowledge of the conditional distribution of unobserved views given observed ones. Given the current selected set $\widehat{\boldsymbol{P}}_t$, this policy chooses

$$\widehat{k}_t \in \operatorname*{arg\,min}_{k\notin\widehat{\boldsymbol{P}}_t} \left[\sum_{j\notin\widehat{\boldsymbol{P}}_t\cup\{k\}} \mathbb{E}[\boldsymbol{E}_j|\mathcal{F}_t] + \sum_{\substack{j\neq j'\\ j,j'\notin\widehat{\boldsymbol{P}}_t\cup\{k\}}} \mathbb{E}[\boldsymbol{C}_{jj'}|\mathcal{F}_t]\right],$$

where $\boldsymbol{E}_j$ and $\boldsymbol{C}_{jj'}$ are as defined in the oracle decomposition. To quantify performance, we define the *true* loss incurred after selecting $\boldsymbol{P}_t$ as $\mathcal{L}(\boldsymbol{P}_t, \boldsymbol{x}) := \big\|\sum_{k\notin\boldsymbol{P}_t} \mathbf{w}_k x_k\big\|_2^2$, and the latent regret of the test-time policy as $\mathcal{R}_t := \mathbb{E}_{\boldsymbol{x}}\Big[\mathcal{L}(\widehat{\boldsymbol{P}}_t, \boldsymbol{x})\Big] - \mathbb{E}_{\boldsymbol{x}}\Big[\mathcal{L}(\boldsymbol{P}_t^\star, \boldsymbol{x})\Big]$, where $\boldsymbol{P}_t^\star$ denotes the view set selected by the oracle after t steps.

Proxy loss. The true loss includes pairwise interaction terms $\boldsymbol{C}_{jj'}$, so the marginal value of a view depends on the entire set of unobserved views; thus bounding expected regret would require controlling the distribution of orderings in a quadratically coupled system, which is analytically intractable without additional structure. To isolate the effect of partial observability from this combinatorial difficulty, we analyze a proxy objective $\widetilde{\mathcal{L}}(\boldsymbol{P}_t, \boldsymbol{x}) := \sum_{k\notin P_t} \mathbb{E}[\boldsymbol{E}_k \mid \mathcal{I}_t]$ and connect it back to the true regret in Theorem 3.5. Two quantities govern the proxy analysis: the centered variables $\boldsymbol{Y}_{j,s} := \boldsymbol{E}_j - \mathbb{E}[\boldsymbol{E}_j|\mathcal{F}_s]$ measure how far each view's energy deviates from its conditional expectation at step s, and the maximal conditional deviation $\boldsymbol{R}_s := \max_{i,j\notin\widehat{\boldsymbol{P}}_t}(Y_{i,s} - Y_{j,s})$ captures the worst-case spread of these estimation errors among unselected views.

3.1 Pathwise Decomposition of the Proxy Regret

We now study the gap between the one-step oracle policy and the optimal test-time policy under the proxy objective. Lemma 3.1 provides a pathwise bound on their difference.

Lemma 3.1. *For all $t \geq 2$, the proxy loss gap satisfies*

$$\widetilde{\mathcal{L}}(\widehat{\boldsymbol{P}}_t, \boldsymbol{x}) - \widetilde{\mathcal{L}}(\boldsymbol{P}_t^\star, \boldsymbol{x}) \leq \sum_{s=1}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \boldsymbol{R}_s, \tag{2}$$

where B_t denotes the mismatch set between the oracle and test-time selections up to time t.

Lemma 3.1 reduces the proxy regret to a sum of weighted *mistake events*: regret is incurred only when the test-time policy selects a view outside the oracle-selected set, and the magnitude of each mistake is controlled by the conditional deviation $\boldsymbol{R}_s$.

Crucially, extending this bound beyond a pathwise statement requires detailed control over the distribution of the order statistics of $(\boldsymbol{E}_k)_{k=1}^d$. In particular, bounding the expected regret entails enumerating all possible orderings of the view energies and characterizing the frequencies and magnitude of deviation of the two policies. Without structural assumptions on the joint distribution of $(\boldsymbol{E}_k)$, we find such a control to be intractable.We therefore posit a latent-variable model in which the data is driven by a shared discrete factor with Gumbel noise, which induces a tractable Plackett–Luce distribution over view rankings (Yellott Jr, 1977; Ma et al., 2021) and makes the regret analysis possible.

Assumption 3.2 (Gumbel latent correlation model)**.** *The view vector $\boldsymbol{x} \in \mathbb{R}^d$ is strictly positive and satisfies $\log \boldsymbol{x} = \boldsymbol{\eta} + \boldsymbol{\beta} \odot Z + \varepsilon$, where $\boldsymbol{\eta}, \boldsymbol{\beta} \in \mathbb{R}^d$ are deterministic, and the common latent factor Z is a discrete latent variable taking values in $z_1, \ldots, z_M$ with probabilities $p_1, \ldots, p_M$. The noise terms follow $\varepsilon_j \overset{\text{i.i.d.}}{\sim} \mathrm{Gumbel}(0, \sigma)$ with $\sigma \in (0, 1/4)$. We further assume that for all latent states z_m, $\mathbb{P}[Z = z_m \mid \mathcal{F}_s] > 0$.*

Under Assumption 3.2, the conditional ordering of the energies $(\boldsymbol{E}_j)_{j=1}^d$ given $Z = z_m$ follows a Plackett–Luce distribution. Specifically, for any permutation π of $1, \ldots, d$,

$$\mathbb{P}\Big(\boldsymbol{E}_{\pi(1)} \geq \cdots \geq \boldsymbol{E}_{\pi(d)} \mid Z = z_m\Big) = \prod_{s=1}^{d} \frac{e^{h_{\pi(s), z_m}}}{\sum_{r=s}^{d} e^{h_{\pi(r), z_m}}},$$

where $(h_{j,z_m})_{j,m}$ are model-dependent utilities determined by $(\boldsymbol{\eta}, \boldsymbol{\beta}, \sigma)$. Marginally, the induced ranking distribution is therefore a mixture of Plackett–Luce models over the latent states of Z. The role of the Gumbel assumption is not primarily to model the data distribution itself, but to ensure that the ranking of view energies follows a tractable parametric family, which is what makes the regret analysis possible. Fig. 5 provides visualizations of datasets generated under this model.

3.2 Regret Bounds under the Gumbel Model

For each latent state m, define the utilities $u_j^{(m)} := 2\log(\|\mathbf{w}_j\|_2) + 2\eta_j + 2\beta_j z_m$ and let $\gamma := \min_{m \in [M]} \min_{j \neq j'} \left| u_j^{(m)} - u_{j'}^{(m)} \right| > 0$ denote the minimum separation between utilities. A larger γ allows easier identification of view-energy rankings. Also, denote $p_{\min} = \min_{m=1,\ldots,M} p_m$.

A key step is to bound the number of disagreements between the oracle and optimal test-time policies. Let G_t (resp. B_t) denote the set of views selected only by the oracle (resp. optimal test-time) policy at time t, and define

$$R_{\max} := \max_{j \in [d]} \frac{\max_m e^{2\beta_j z_m}}{\min_m e^{2\beta_j z_m}} < \infty, \text{ and } \varepsilon_\gamma := \frac{e^\gamma - 1}{R_{\max} + e^\gamma - 1} \in (0, 1).$$

Lemma 3.3. *The expected number of oracle-test-time mismatches satisfies, for any $\lambda > 0$,*

$$\mathbb{E}|B_t| = \mathbb{E}|G_t| \leq D_t(\lambda) := \min\left\{ t, d - t, \frac{C_\lambda}{1 - \rho_\lambda} + \frac{1}{e^{\rho_*} - 1} \right\} \tag{3}$$

where $C_\lambda := (M-1)\left(\frac{1-\varepsilon_\gamma}{\varepsilon_\gamma p_{min}}\right)^\lambda$, $\rho_\star := 1 - e^{-\gamma/(2\sigma)}$, and ρ_λ is a contraction constant (see § F.4).

Proof outline. The analysis proceeds in three steps: (i) we bound the number of oracle–test-time disagreements (ii) this yields a proxy regret bound via the pathwise decomposition; (iii) we relate proxy and true losses through a structural assumption on the embedding weights. For proofs, see § E.

The regret on the proxy controls errors due to partial observability, but says nothing about the gap between the proxy and true losses. Bridging this gap requires a condition on the geometry of the embedding weights: if the normalized columns of $\mathbf{W}$ are approximately orthogonal, the cross-view interaction terms $\boldsymbol{C}_{jj'}$, that the proxy ignores, remain small.

Assumption 3.4 (Near Isometry)**.** *There exists constants C, $p_{\text{eff}} > 0$ such that for any subset $J \subseteq \{1, \dots, d\}$ of size m, the matrix $\boldsymbol{B}_J = \left(\mathbf{w}_u / \|\mathbf{w}_u\|\right)_{u \in J}$ satisfies*

$$\left\|\boldsymbol{B}_J^\top \boldsymbol{B}_J - \boldsymbol{I}_m\right\|_{\text{op}} \leq C\left(\sqrt{\frac{m}{p_{\text{eff}}}} + \frac{m}{p_{\text{eff}}}\right).$$

This condition enforces approximate orthogonality of embedding directions and is empirically satisfied by large-scale vision transformers such as the DINOv2 family (Oquab et al., 2023) and by trained autoencoders across a range of dimensions (§ H.9.2). It ensures that the cross-view interaction terms ignored by the proxy objective remain controlled, so that the proxy loss remains aligned with the true latent reconstruction loss. We can now state the main result.

Theorem 3.5. *Take $p \in \left(1, \frac{1}{2\sigma}\right)$ and let $q = \frac{p}{p-1}$. Under Assumptions 3.2 and 3.4, for any $\lambda > 0$, the latent regret satisfies*

$$\mathcal{R}_t \leq \left(\sum_{i=1}^{d} M_{(i)} \min(t, d-i+1)\right)^{1/p} D_t^{1/q} + C'\left(\sqrt{\frac{N_t}{p_{\text{eff}}}} + \frac{N_t}{p_{\text{eff}}}\right)\sqrt{D_t N_t}, \tag{4}$$

where $M_j := \mathbb{E}[\boldsymbol{E}_j^p]$ and $M_{(1)} \geq \cdots \geq M_{(d)}$ be the sorted p-th moments, and $N_t = d - t + D_t$.

The bound separates two sources of error. The first term is an *oracle-mismatch term*: it captures the cost of partial observability through the disagreement envelope D_t. This term decreases as the utility gap γ increases, reflecting that well-separated energy rankings are easier to infer from partial observations. The second term is a *spectral misalignment term*: it controls the error introduced by replacing the true latent loss with the marginal-energy proxy. Consequently, the bound is boundary-correct: since $D_t = 0$ at $t = 0$ and $t = d$, both terms vanish at the endpoints, matching the reverse-U regret profile observed empirically in Fig. 2.

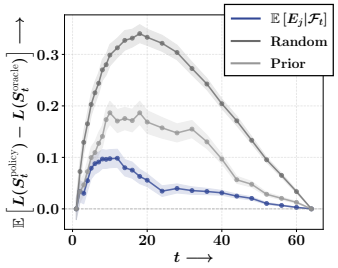


Figure 2: Expected true loss gap curve between the oracle and the test-time optimal, random, and prior-based ($\mathbb{E}[\boldsymbol{E}_j]$) policies.

4 Methodology

4.1 From Theoretical Test-Time Policies to Learnable Policies

The theoretical analysis in § 3 yields three insights that directly inform algorithm design. First, Theorem 3.5 shows that representation-recovery errors control latent regret up to an oracle-mismatch term and a spectral misalignment term, which becomes small when embedding directions are approximately orthogonal. This motivates using progress toward the full-view representation as a self-supervised acquisition signal. Second, the posterior concentration result (Lemma F.4) shows that conditioning on observed views can reduce the gap to a full-information oracle, motivating adaptive, instance-specific policies that update after each acquisition. Third, the analysis is restricted to *myopic* policies: the oracle itself is one-step greedy, and the regret bound characterizes the gap between two greedy strategies. In practice, the conditional law of unobserved views given $\mathcal{F}_t$ is unknown, and acquisition is a finite-horizon decision problem in which early choices affect the value of later measurements, especially when views are correlated (Shim et al., 2018; Zannone et al., 2019; Kachuee et al., 2019; Schütz et al., 2025). This motivates moving beyond the greedy regime to long-horizon planning.

Together, these insights culminate in ECHO-k, a reinforcement-learning algorithm for self-supervised, long-horizon view acquisition. We use the theory to define a reward, and learn a parameterized policy

that optimizes cumulative improvement in representation quality over an entire acquisition trajectory, rather than approximating the one-step oracle. Formally, we aim to learn a parameterized policy $\widehat{\pi}_{\boldsymbol{\theta}}$ that maximizes cumulative discounted improvement starting from any step s

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{t=s}^{d} \alpha^{t-s}\mathbb{E}\big[R_t\Big(\pi_{\boldsymbol{\theta}}(\boldsymbol{P}_{t-1})\Big)\big],$$

where R_t denotes the incremental improvement in latent representation quality at step t, $\alpha \in (0, 1]$ is a discount factor that allows the policy to trade off immediate gains against future benefits. The self-supervised reward requires a reference representation against which partial representations are compared. We consider two training regimes. In the *complete-reference* setting, each training instance is fully observed, so the reference is the full-view representation $\boldsymbol{z}_i^\star = \boldsymbol{z}_i(\llbracket 1, K\rrbracket)$. In the *partially observed-reference* setting, training instance i contains only a subset of available views $R_i \subseteq \llbracket 1, K\rrbracket$; in this case, we use $\boldsymbol{z}_i^{\mathrm{ref}} = \boldsymbol{z}_i(R_i)$ as a noisy but label-free reference target.

4.2 Learning Representation-Driven Policies

Reward. Given the reference representation $\boldsymbol{z}_i^{\mathrm{ref}}$, we define the representation error for instance i as $\ell_i(S) := \|\boldsymbol{z}_i^{\mathrm{ref}} - \boldsymbol{z}_i(S)\|_2$. The step-wise reward is the normalized improvement in the error,

$$R_t(\widehat{\pi}_{\boldsymbol{\theta}}) := \omega_t \frac{1}{N}\sum_{i=1}^{N} \frac{\ell_i(\boldsymbol{P}_{t-1}^i) - \ell_i(\boldsymbol{P}_{t-1}^i \cup \{k_t^i\})}{\ell_i(\varnothing) + \epsilon}, \tag{5}$$

where N is the minibatch size, and $k_t^i \sim \widehat{\pi}_{\boldsymbol{\theta}}(\cdot \mid s_{t-1}^i)$ is the view selected for instance i at step t. We use the unsquared norm rather than the squared loss analyzed in § 3, as this provides more stable reward scaling in practice. The reward 5 retains the same structure as the representation-recovery objective studied theoretically: it measures marginal progress toward a reference latent representation without using downstream labels. The factor ω_t optionally upweights early acquisition steps, where marginal gains are typically largest, and normalization prevents high-error instances from dominating the learning signal. Moreover, when views have nonuniform acquisition costs c_k, we use a cost-aware reward of the form $r_t^{\mathrm{cost}} = r_t - \lambda_c \frac{1}{N}\sum_{i=1}^{N} c_{k_t^i}$ with costs normalized across views.

Policy input. At each step, the policy receives the current partial representation $\boldsymbol{z}_i(\boldsymbol{P}_t^i)$, the observed view values with unobserved entries masked, the acquisition mask, and the remaining budget. In heterogeneous-cost experiments, the state is additionally augmented with the view-cost vector and remaining cost budget; the full state specification is given in § G.5.

Optimization. We optimize $\widehat{\pi}_{\boldsymbol{\theta}}$ using Proximal Policy Optimization (PPO) (Schulman et al., 2017), with invalid action masking (Huang & Ontañón, 2020) to prevent re-selection of already acquired views. During both training and deployment, the policy has access only to the frozen representation model and partial observations induced by the current acquisition set.

5 Experimental Results

We evaluate ECHO-k across controlled synthetic experiments, vision, and tabular encoders, and three frozen foundation-model settings: single-cell transcriptomics with TABULA (Ding et al., 2025), spatial proteomics with VirTues (Wenckstern et al., 2025), and multimodal geospatial sensing with a self-supervised multimodal encoder (Gordon et al., 2026). The experiments are designed to answer four questions: (i) whether representation-driven acquisition improves downstream performance under limited budgets; (ii) whether the same label-free acquisition principle composes with heterogeneous pretrained backbones; (iii) whether learned acquisition policies transfer to downstream tasks unseen during policy learning; and (iv) whether the method remains effective under realistic deployment constraints, including partially observed reference data and heterogeneous acquisition costs. We discuss the results for foundation models, grouped by application domain, here. Additional controlled benchmarks on synthetic data, and vision datasets such as MNIST, Fashion-MNIST, and MiniBooNE, a physics benchmark, along with incorporating costs and adaptive stopping, are shown in § H.

Evaluation. Across all experiments, the representation model is frozen and receives partially observed inputs. ECHO-k learns an acquisition policy using only self-supervised rewards derived

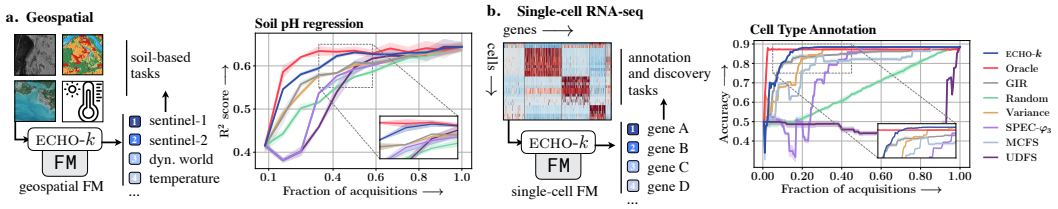


Figure 3: **a.** Results of ECHO-k on acquiring geospatial modalities to predict soil pH regression on MMEarth-Bench ($d = 12$) **b.** ECHO-k can operate on single-cell RNA-seq foundation models to sequentially acquire informative genes that support broad set of downstream tasks. We further present comparison of ECHO-k to baselines, evaluated on a cell type annotation task on hPancreas (with $d = 200$).

from progress toward a reference latent representation. In the complete-reference setting, this reference is the full-view embedding; in the partially observed-reference setting, it is computed from the subset of views available for each training instance. Downstream labels are never used for policy learning and are reserved exclusively for post-hoc evaluation. At deployment, the policy observes only the values of the views acquired so far and the acquisition mask, and never accesses the full-view representation or downstream labels. To measure downstream performance, we train a single linear probe per downstream task on partial embeddings using random masking augmentation, and keep it fixed across acquisition budgets. The same probe is applied to the partial embeddings induced by every acquisition method, so that differences in performance reflect only which views were acquired. Moreover, a single acquisition policy is learned per backbone and reused across all budgets and downstream tasks, including tasks that were not specified during policy learning. All results are averaged over three random seeds (§ G.3).

Baselines and metrics. We compare against 12 baselines spanning oracle/random references, model-driven heuristics (`GIR`, ESM2-based orderings), and unsupervised feature selection methods (`Laplacian Score`, SPEC-φ_3, `MCFS`, `UDFS`, `SVD-Ent`, `Variance`) adapted as acquisition orders (see § G.4 for details). The `Oracle` greedily selects the view that maximally reduces the distance to the full-view embedding using unobserved values; it is not deployable at test time and serves as an upper bound on greedy acquisition. Domain-specific baselines are evaluated only where they are defined, e.g., ESM2-based orderings require protein-sequence embeddings and are therefore restricted to protein markers. Since most of these baselines induce a fixed, instance-independent acquisition order, we additionally compare against label-free instance-adaptive policies (§ 5.3) and supervised acquisition methods that use downstream labels (§ H.4). For regression, we report R^2 or MSE, and for other real-world tasks, classification accuracy as a function of the acquisition budget $t \in [\![1, d]\!]$, reported as a fraction of the total number of views. To directly assess the training objective, we further report the normalized representation recovery (§ 5.4). For full evaluation details, refer to § G.

5.1 Evaluation on Foundation Model Backends

Single-cell foundation models. In single-cell RNA-seq, selecting the *right* genes is a critical bottleneck, and the same panel must often support multiple analyses such as cell type annotation, trajectory inference, or gene program discovery that may not be specified at design time (Fig. 3a). We pair ECHO-k with TABULA (Ding et al., 2025), a FM for single-cell transcriptomics (§ G.2).

Results. We evaluate on hPancreas, a curated single-cell transcriptomics (RNA-seq) collection from human pancreatic tissue (Ding et al., 2025), retaining the top 200 highly variable genes across 14 cell types. We use cell type annotation as a representative downstream task; it is the standard evaluation objective in single-cell analysis and requires the gene subset to preserve fine-grained phenotypic distinctions (§ G.1). ECHO-k underperforms slightly during the first $\sim$2–3% of acquisitions but rapidly recovers and consistently outperforms `Variance`, `GIR`, and `Random` thereafter (Fig. 3b). `Random` performs poorly throughout, reflecting the sparsity of informative genes in high-dimensional expression data.

Spatial proteomics foundation models. In spatial proteomics, multiplexed protein imaging characterizes tissue through successive rounds of antibody staining, each revealing a new marker and adding to assay cost, imaging time, and tissue degradation risk. Panel design must therefore

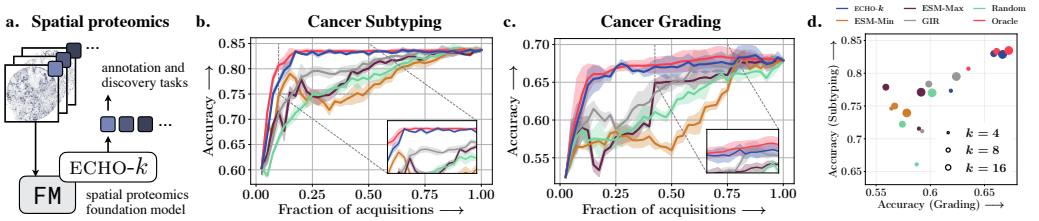


Figure 4: **Results of ECHO-k on the NSCLC dataset. a.** Overview of the experimental setting: given high-dimensional multiplex imaging data, we study how foundation models trained on such data can be leveraged to sequentially acquire informative markers for downstream annotation and discovery tasks. **b.** Comparison of ECHO-k with baseline methods on the cancer subtyping task. **c.** Comparison of ECHO-k with baseline methods on the cancer grading task. **d.** Scatter plot illustrating the accuracy trade-off between cancer subtyping and grading when using the same set of views.

support multiple diagnostic objectives simultaneously (e.g., cancer subtyping *and* grading), with different patients potentially benefiting from different markers (Fig. 4a). We compose ECHO-k with the VirTues foundation model (Wenckstern et al., 2025), which embeds multiplex imaging data into tissue-level representations. At deployment, the policy mirrors the iterative staining workflow: selecting which marker to acquire next, updating its partial representation, and using the updated embedding to inform the next selection (§ G.5).

Results. We evaluate on a non-small cell lung cancer (NSCLC) patient cohort (Cords et al., 2024) profiled by imaging mass cytometry with 40 protein markers spanning tumor, immune, stromal, and vascular compartments. We consider two clinically relevant downstream tasks: cancer subtyping (2 classes) and cancer grading (3 classes), the latter being substantially more challenging (§ G.1). ECHO-k consistently outperforms baselines on both tasks (Figs. 4b,c). The scatter plot in Fig. 4d shows that ECHO-k achieves favorable accuracy on *both* tasks under identical acquisition orders, demonstrating task-agnostic panel design. A qualitative analysis (§ H.10) reveals that ECHO-k prioritizes clinically established markers such as `panCK`, `Ki-67`, and `CD4` early in the process, consistent with domain knowledge (Martin et al., 2004; Su et al., 2006), despite receiving no supervision. Acquisition orders differ substantially across samples, with mean pairwise Kendall's $\tau = 0.06$. Across training seeds, orders for the same sample show greater agreement: mean $\tau = 0.512$ and 61.6% top-five overlap, compared with 12.8% under random ordering.

Multimodal geospatial foundation models. Earth observation integrates heterogeneous sensing products, from optical and radar imagery to terrain, land-cover, canopy-height, meteorological, and geolocation-derived variables. We evaluate ECHO-k on MMEarth-Bench (Gordon et al., 2026), a globally distributed benchmark of 12 aligned geospatial modalities, treating each modality as a separately acquirable view.

Results. We evaluate on soil pH regression as an unseen downstream task, reporting R^2 as a function of the number of modalities. ECHO-k outperforms non-oracle baselines consistently under limited modality budgets, reaching strong predictive performance after acquiring only 20-30% of the views (Fig. 3b). We also notice that `GIR` and Variance-based baselines are the second-best performers.

5.2 Robustness to Partially Observed Reference Data

In Table 1, we evaluate whether ECHO-k remains effective when trained from incomplete views by randomly masking a fraction m of views before computing the reference embedding, and train the policy against this partially observed target. For single-cell transcriptomics FMs, performance is nearly unchanged for m =10%, remains strong for m =20%, and still recovers high accuracy under heavier missingness once moderate acquisition budgets are allowed. For proteomics FMs, performance degrades more gradually with m, but partially observed-reference policies remain competitive with strong baselines across budgets. We also notice that for high budgets, `GIR` is competitive with ECHO-k. Overall, ECHO-k degrades smoothly rather than collapsing, suggesting that useful acquisition policies can be learned from incomplete reference representations.

Table 1: **Results of robustness to partially observed reference data.** Performance of ECHO-k when an m-fraction of views is randomly masked while computing the training reference representation. Accuracy is reported at different acquisition budgets on hPancreas and NSCLC grading. Lower m corresponds to more complete reference data. denotes best and second best performance, apart from the Oracle.

Method	hPancreas				Method	NSCLC (Grading)			
	20%	40%	60%	80%		20%	40%	60%	80%
Oracle	0.872±0.002	0.872±0.001	0.872±0.001	0.872±0.001	Oracle	0.665±0.006	0.672±0.010	0.680±0.012	0.681±0.010
Var	0.679±0.004	0.823±0.004	0.873±0.001	0.874±0.001	ESM-Max	0.564±0.005	0.649±0.004	0.655±0.007	0.676±0.008
SPEC-φ_3	0.476±0.009	0.421±0.012	0.696±0.001	0.819±0.004	ESM-Min	0.582±0.005	0.569±0.008	0.597±0.015	0.684±0.009
MCFS	0.694±0.000	0.762±0.004	0.840±0.006	0.813±0.009	GIR	0.607±0.004	0.636±0.019	0.678±0.013	0.683±0.009
Random	0.500±0.007	0.516±0.010	0.566±0.005	0.608±0.013	Random	0.573±0.012	0.611±0.008	0.640±0.011	0.661±0.005
ECHO-k	0.808±0.003	0.880±0.001	0.884±0.002	0.884±0.002	ECHO-k	0.658±0.006	0.671±0.009	0.674±0.009	0.683±0.016
ECHO-k (m =10%)	0.812±0.003	0.880±0.002	0.884±0.003	0.884±0.002	ECHO-k (m =10%)	0.655±0.015	0.671±0.014	0.669±0.013	0.678±0.012
ECHO-k (m =20%)	0.808±0.003	0.871±0.002	0.883±0.002	0.883±0.002	ECHO-k (m =20%)	0.641±0.009	0.658±0.009	0.660±0.010	0.672±0.010
ECHO-k (m =30%)	0.783±0.004	0.865±0.004	0.882±0.002	0.882±0.002	ECHO-k (m =30%)	0.615±0.007	0.653±0.011	0.656±0.008	0.671±0.018
ECHO-k (m =40%)	0.751±0.005	0.851±0.006	0.864±0.004	0.871±0.003	ECHO-k (m =40%)	0.609±0.010	0.636±0.011	0.640±0.007	0.664±0.009
ECHO-k (m =50%)	0.733±0.007	0.832±0.005	0.861±0.004	0.872±0.002	ECHO-k (m =50%)	0.595±0.012	0.622±0.008	0.636±0.007	0.669±0.015

5.3 Comparison with Adaptive Acquisition

Beyond fixed acquisition orderings, we compare ECHO-k with label-free, instance-adaptive policies under the same information setting: GSMRL-AIR (Li & Oliva, 2021), based on conditional generative imputations; Adaptive Latent Gain (ALG), which predicts one-step representation gain; and Cluster Information Gain (CIG), which estimates view-wise information gain under a Gaussian mixture model. ECHO-k outperforms all three across every budget on MiniBooNE (Table 2) and hPancreas (§H.3). Notably, ALG shares ECHO-k 's state and reward but optimizes only one-step gain, highlighting the benefit of long-horizon acquisition. Finally, ECHO-k remains competitive with supervised acquisition policies that have access to downstream labels (§ H.4), quantifying the cost of task agnosticism.

Table 2: **Comparison with unsupervised adaptive acquisition on MiniBooNE.** Accuracy (%) at different acquisition budgets. denotes best and second best performance.

Method	20%	40%	60%
CIG	79.80	84.87	86.17
ALG	77.50	83.60	86.12
GSMRL-AIR	79.18	84.78	86.40
ECHO-k	81.70	86.87	87.60

5.4 Comparison of representation-recovery

While our main results focus on downstream performance, we also directly evaluate the representation-recovery objective optimized by ECHO-k (Eq. 5). For each instance i, we define normalized recovery after t acquisitions as

$$\rho_{i,t} := 1 - \frac{\|\boldsymbol{z}_i(\boldsymbol{P}_t^i) - \boldsymbol{z}_i^\star\|_2}{\|\boldsymbol{z}_i(\varnothing) - \boldsymbol{z}_i^\star\|_2},$$

We report ρ_t averaged over test instances. Table 3 shows that ECHO-k achieves higher recovery than `Variance`, the strongest competing baseline on both hPancreas and MiniBooNE, across all budgets. Moreover, recovery and downstream accuracy exhibit the same ordering (Fig. 3b; Fig. 7): better preservation of the full-view representation corresponds to better downstream performance. This provides empirical support for the connection established in Theorem 3.5 between representation recovery and downstream error.

Table 3: **Results of representation recovery.** Comparison of ρ_t of ECHO-k and `Variance` at different budgets. Higher ρ_t denotes better recovery. denotes best performance.

Dataset	Method	20%	40%	60%	80%
hPancreas	Var	0.38	0.53	0.65	0.84
	ECHO-k	0.41	0.58	0.70	0.89
MiniBooNE	Var	0.35	0.54	0.60	0.73
	ECHO-k	0.41	0.60	0.71	0.80

6 Conclusion

We presented ECHO-k, a task-agnostic and label-free framework for sequential test-time view acquisition that uses a pretrained model's latent representation as a self-supervised proxy: under a cost budget, acquire views so that the induced partial representation rapidly approaches the full-view representation. We support this formulation with a theoretical analysis in a linear setting that characterizes the gap between oracle and feasible test-time policies, and we introduce a practical cost-aware RL algorithm that learns long-horizon acquisition strategies. Experiments with controlled synthetic data and multimodal foundation models across single-cell transcriptomics, spatial proteomics, and geospatial sensing show improved budgeted acquisition, provide evidence for transfer to downstream tasks unseen during policy learning, robustness to partially observed reference data, and extensibility to heterogeneous acquisition costs. The broader implication is that foundation models need not remain passive predictors over pre-collected inputs; they can become resource-aware systems that actively shape the evidence available for inference.

Acknowledgments

This work was supported by the Swiss AI Initiative through a Swiss National Supercomputing Centre (CSCS) allocation under the Swiss AI Large Grants `a146` and `ab007` ("Virtual Patient Platform"), an EPFL AI Center Postdoctoral Fellowship, an AI2050 Early Career Fellowship, and by the ISREC Foundation.

References

Ash, J. T., Zhang, C., Krishnamurthy, A., Langford, J., and Agarwal, A. Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds. *arXiv preprint arXiv:1906.03671*, 2019.

Blau, T., Bonilla, E. V., Chades, I., and Dezfouli, A. Optimizing Sequential Experimental Design with Deep Reinforcement Learning. In *International Conference on Machine Learning*, pp. 2107–2128. PMLR, 2022.

Blau, T., Chades, I., Dezfouli, A., Steinberg, D. M., and Bonilla, E. V. Cross-entropy estimators for sequential experiment design with reinforcement learning. In *NeurIPS 2023 Workshop on Adaptive Experimental Design and Active Learning in the Real World*, 2023.

Blum, A. L. and Langley, P. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2):245–271, 1997.

Breiman, L. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

Cai, D., Zhang, C., and He, X. Unsupervised feature selection for multi-cluster data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 333–342, 2010.

Carrara, N., Leurent, E., Laroche, R., Urvoy, T., Maillard, O.-A., and Pietquin, O. Budgeted Reinforcement Learning in Continuous State Space. *Advances in Neural Information Processing Systems*, 32, 2019.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., and Kujala, J. V. Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22 (4):887–905, 2010.

Chakrabarti, D., Kumar, R., Radlinski, F., and Upfal, E. Mortal Multi-Armed Bandits. *Advances in Neural Information Processing Systems*, 21, 2008.

Chaloner, K. and Verdinelli, I. Bayesian Experimental Design: A Review. *Statistical Science*, pp. 273–304, 1995.

Chen, Y., Hassani, S. H., Karbasi, A., and Krause, A. Sequential Information Maximization: When is Greedy Near-optimal? In *Conference on Learning Theory*, pp. 338–363. PMLR, 2015.

Claici, S., Genevay, A., and Solomon, J. Wasserstein Measure Coresets. *arXiv preprint arXiv:1805.07412*, 2018.

Coleman, C., Yeh, C., Mussmann, S., Mirzasoleiman, B., Bailis, P., Liang, P., Leskovec, J., and Zaharia, M. Selection via Proxy: Efficient Data Selection for Deep Learning. *arXiv preprint arXiv:1906.11829*, 2019.

Cords, L., Engler, S., Haberecker, M., Rüschoff, J. H., Moch, H., de Souza, N., and Bodenmiller, B. Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer. *Cancer Cell*, 42(3):396–412, 2024.

Covert, I. C., Qiu, W., Lu, M., Kim, N. Y., White, N. J., and Lee, S.-I. Learning to Maximize Mutual Information for Dynamic Feature Selection. In *International Conference on Machine Learning*, pp. 6424–6447. PMLR, 2023.

Ding, J., Lin, J., Jiang, S., Wang, Y., Miao, Z., Fang, Z., Tang, J., Li, M., and Qiu, X. Tabula: A Tabular Self-Supervised Foundation Model for Single-Cell Transcriptomics. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Ducoffe, M. and Precioso, F. Adversarial Active Learning for Deep Networks: a Margin Based Approach. *arXiv preprint arXiv:1802.09841*, 2018.

Dulac-Arnold, G., Denoyer, L., Preux, P., and Gallinari, P. Datum-Wise Classification: A Sequential Approach to Sparsity. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 375–390. Springer, 2011.

Feldman, D. Core-sets: Updated survey. In *Sampling Techniques for Supervised or Unsupervised Tasks*, pp. 23–44. Springer, 2019.

Fisher, A., Rudin, C., and Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20(177):1–81, 2019.

Foster, A., Ivanova, D. R., Malik, I., and Rainforth, T. Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design. In *International Conference on Machine Learning*, pp. 3384–3395. PMLR, 2021.

Gadgil, S., Covert, I., and Lee, S.-I. Estimating Conditional Mutual Information for Dynamic Feature Selection. *arXiv preprint arXiv:2306.03301*, 2023.

Golovin, D. and Krause, A. Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Gong, W., Tschiatschek, S., Nowozin, S., Turner, R. E., Hernández-Lobato, J. M., and Zhang, C. Icebreaker: Element-wise Efficient Information Acquisition with a Bayesian Deep Latent Gaussian Model. *Advances in Neural Information Processing Systems*, 32, 2019.

Gordon, L., Belongie, S., Igel, C., and Lang, N. MMEarth-Bench: Global Model Adaptation via Multimodal Test-Time Training. *arXiv preprint arXiv:2602.06285*, 2026.

Guha, S. and Munagala, K. Multi-armed bandits with limited exploration. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2007.

Guney, O. B., Saichandran, K. S., Elzokm, K., Zhang, Z., and Kolachalama, V. B. Active feature acquisition via explainability-driven ranking. In *International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pp. 20748–20765. PMLR, 2025.

Gupta, S., Chaudhari, S., Joshi, G., and Yağan, O. Multi-armed bandits with correlated arms. *IEEE Transactions on Information Theory*, 67(10):6711–6732, 2021.

Gupte, S. R., Aklilu, J., Nirschl, J. J., and Yeung-Levy, S. Revisiting Active Learning in the Era of Vision Foundation Models. *arXiv preprint arXiv:2401.14555*, 2024.

Guyon, I. and Elisseeff, A. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(Mar):1157–1182, 2003.

He, H., Eisner, J., and Daume, H. Imitation Learning by Coaching. *Advances in Neural Information Processing Systems*, 25, 2012.

He, X., Cai, D., and Niyogi, P. Laplacian Score for Feature Selection. *Advances in Neural Information Processing Systems*, 18, 2005.

Holzmüller, D., Zaverkin, V., Kästner, J., and Steinwart, I. A Framework and Benchmark for Deep Batch Active Learning for Regression. *Journal of Machine Learning Research*, 24(164):1–81, 2023.

Huang, D., Guo, Y., Acerbi, L., and Kaski, S. Amortized Bayesian Experimental Design for Decision-Making. *Advances in Neural Information Processing Systems*, 37:109460–109486, 2024.

Huang, H.-T., Dinh, D., and Oliva, J. B. Information Templates: A New Paradigm for Intelligent Active Feature Acquisition. *arXiv preprint arXiv:2508.18380*, 2025.

Huang, S. and Ontañón, S. A Closer Look at Invalid Action Masking in Policy Gradient Algorithms. *arXiv preprint arXiv:2006.14171*, 2020.

Huang, S.-J., Xu, M., Xie, M.-K., Sugiyama, M., Niu, G., and Chen, S. Active feature acquisition with supervised matrix completion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1571–1579, 2018.

Jain, E., Nandy, T., Aggarwal, G., Tendulkar, A., Iyer, R., and De, A. Efficient Data Subset Selection to Generalize Training Across Models: Transductive and Inductive Networks. *Advances in Neural Information Processing Systems*, 36:4716–4740, 2023.

Jain, E., Wenckstern, J., von Querfurth, B., and Bunne, C. Test-Time View Selection for Multi-Modal Decision Making. In *ICLR 2025 Workshop on Machine Learning for Genomics Explorations*, 2025.

Kachuee, M., Goldstein, O., Karkkainen, K., Darabi, S., and Sarrafzadeh, M. Opportunistic Learning: Budgeted Cost-Sensitive Learning from Data Streams. *arXiv preprint arXiv:1901.00243*, 2019.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling Laws for Neural Language Models. *arXiv preprint arXiv:2001.08361*, 2020.

Killamsetty, K., Durga, S., Ramakrishnan, G., De, A., and Iyer, R. GRAD-MATCH: Gradient Matching based Data Subset Selection for Efficient Deep Model Training. In *International Conference on Machine Learning*, pp. 5464–5474. PMLR, 2021.

Kim, Y. and Shin, B. In defense of core-set: A density-aware core-set selection for active learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 804–812, 2022.

Kirsch, A., Van Amersfoort, J., and Gal, Y. BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Kobayashi, Y., Jing, Z., Yao, J., Namkoong, H., and Joshi, S. Learning-To-Measure: In-context Active Feature Acquisition. *arXiv preprint arXiv:2510.12624*, 2025.

Kohavi, R. and John, G. H. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2): 273–324, 1997.

Kolossov, G., Montanari, A., and Tandon, P. Towards a statistical theory of data selection under weak supervision. *arXiv preprint arXiv:2309.14563*, 2023.

LeCun, Y., Cortes, C., and Burges, C. J. C. The MNIST database of handwritten digits. https://yann.lecun.org/exdb/mnist/index.html, 1998. URL https://yann.lecun.org/exdb/mnist/index.html.

Li, D., Wang, Z., Chen, Y., Jiang, R., Ding, W., and Okumura, M. A Survey on Deep Active Learning: Recent Advances and New Frontiers. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):5879–5899, 2025.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., and Liu, H. Feature Selection: A Data Perspective. *ACM Computing Surveys (CSUR)*, 50(6):1–45, 2017.

Li, J., Wu, L., Dani, H., and Liu, H. Unsupervised Personalized Feature Selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Li, Y. and Oliva, J. Active Feature Acquisition with Generative Surrogate Models. In *International Conference on Machine Learning*, pp. 6450–6459. PMLR, 2021.

Lifshitz, S., Lindenbaum, O., Mishne, G., Meir, R., and Benisty, H. Unsupervised Feature Selection Through Group Discovery. *arXiv preprint arXiv:2511.09166*, 2025.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Lu, M. Y., Chen, B., Williamson, D. F., Chen, R. J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L. P., Gerber, G., et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024.

Lundberg, S. M. and Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 2017.

Ma, C., Tschiatschek, S., Palla, K., Hernández-Lobato, J. M., Nowozin, S., and Zhang, C. EDDI: Efficient Dynamic Discovery of High-Value Information with Partial VAE. In *International Conference on Machine Learning (ICML)*, 2019.

Ma, J., Yi, X., Tang, W., Zhao, Z., Hong, L., Chi, E., and Mei, Q. Learning-to-Rank with Partitioned Preference: Fast Estimation for the Plackett-Luce Model. In *International Conference on Artificial Intelligence and Statistics*, pp. 928–936. PMLR, 2021.

Madasu, A., Oliva, J., and Bertasius, G. Learning to Retrieve Videos by Asking Questions. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 356–365, 2022.

Martin, B., Paesmans, M., Mascaux, C., Berghmans, T., Lothaire, P., Meert, A.-P., Lafitte, J.-J., and Sculier, J.-P. Ki-67 expression and patients survival in lung cancer: systematic review of the literature with meta-analysis. *British Journal of Cancer*, 91(12):2018–2025, 2004.

Melville, P., Saar-Tsechansky, M., Provost, F., and Mooney, R. Active feature-value acquisition for classifier induction. In *Fourth IEEE International Conference on Data Mining (ICDM'04)*, pp. 483–486. IEEE, 2004.

Mizrahi, D., Bachmann, R., Kar, O., Yeo, T., Gao, M., Dehghan, A., and Zamir, A. 4M: Massively Multimodal Masked Modeling. *Advances in Neural Information Processing Systems*, 36:58363–58408, 2023.

Nielsen, F. Chernoff information of exponential families. *arXiv preprint arXiv:1102.2684*, 2011.

Norcliffe, A., Lee, C., Imrie, F., Van Der Schaar, M., and Liò, P. Stochastic Encodings for Active Feature Acquisition. *arXiv preprint arXiv:2508.01957*, 2025.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. DINOv2: Learning Robust Visual Features without Supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Piskorz, J., Astorga, N., Berrevoets, J., and van der Schaar, M. Active Feature Acquisition for Personalised Treatment Assignment. In *International Conference on Artificial Intelligence and Statistics*, pp. 4330–4338. PMLR, 2025.

Qu, J., Holzmüller, D., Varoquaux, G., and Le Morvan, M. TabICL: A Tabular Foundation Model for In-Context Learning on Large Data. *arXiv preprint arXiv:2502.05564*, 2025.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning*, pp. 8748–8763. PmLR, 2021.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., and Dormann, N. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 22 (268):1–8, 2021.

Rahbar, A., Aronsson, L., and Chehreghani, M. H. A Survey on Active Feature Acquisition Strategies. *arXiv preprint arXiv:2502.11067*, 2025.

Rainforth, T., Foster, A., Ivanova, D. R., and Bickford Smith, F. Modern Bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024.

Roe, B. MiniBooNE particle identification. UCI Machine Learning Repository, 2005. DOI: https://doi.org/10.24432/C5QC87.

Roy, O. and Vetterli, M. The effective rank: A measure of effective dimensionality. In *2007 15th European Signal Processing Conference*, pp. 606–610. IEEE, 2007.

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Schütz, V., Wu, H., Rezvan, R., Aronsson, L., and Chehreghani, M. H. AFABench: A Generic Framework for Benchmarking Active Feature Acquisition. *arXiv preprint arXiv:2508.14734*, 2025.

Sebastiani, P. and Wynn, H. P. Maximum entropy sampling and optimal Bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):145–157, 2000.

Sener, O. and Savarese, S. Active Learning for Convolutional Neural Networks: A Core-Set Approach. *arXiv preprint arXiv:1708.00489*, 2017.

Settles, B. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3): 379–423, 1948.

Shim, H., Hwang, S. J., and Yang, E. Joint Active Feature Acquisition and Classification with Variable-Size Set Encoding. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

Shmatko, A., Jung, A. W., Gaurav, K., Brunak, S., Mortensen, L. H., Birney, E., Fitzgerald, T., and Gerstung, M. Learning the natural history of human disease with generative transformers. *Nature*, 647(8088):248–256, 2025.

Shui, C., Zhou, F., Gagné, C., and Wang, B. Deep Active Learning: Unified and Principled Method for Query and Training. In *International Conference on Artificial Intelligence and Statistics*, pp. 1308–1318. PMLR, 2020.

Shukor, M., Fini, E., da Costa, V. G. T., Cord, M., Susskind, J., and El-Nouby, A. Scaling Laws for Native Multimodal Models. In *International Conference on Computer Vision (ICCV)*, pp. 12–23, 2025.

Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al. DINOv3. *arXiv preprint arXiv:2508.10104*, 2025.

Solorio-Fernández, S., Carrasco-Ochoa, J. A., and Martínez-Trinidad, J. F. A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2):907–948, 2020.

Sristi, R. D., Lindenbaum, O., Lifshitz, S., Lavzin, M., Schiller, J., Mishne, G., and Benisty, H. Contextual Feature Selection with Conditional Stochastic Gates. *arXiv preprint arXiv:2312.14254*, 2023.

Su, Y.-C., Hsu, Y.-C., and Chai, C.-Y. Role of TTF-1, CK20, and CK7 immunohistochemistry for diagnosis of primary and secondary lung adenocarcinoma. *The Kaohsiung Journal of Medical Sciences*, 22(1):14–19, 2006.

Tharwat, A. and Schenck, W. A survey on active learning: State-of-the-art, practical challenges and research directions. *Mathematics*, 11(4):820, 2023.

Valancius, M., Lennon, M., and Oliva, J. Acquisition Conditioned Oracle for Nongreedy Active Feature Acquisition. *arXiv preprint arXiv:2302.13960*, 2023.

Varshavsky, R., Gottlieb, A., Linial, M., and Horn, D. Novel unsupervised feature filtering of biological data. *Bioinformatics*, 22(14):e507–e513, 2006.

Wang, D. and Shang, Y. A new active labeling method for deep learning. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 112–119. IEEE, 2014.

Wei, K., Liu, Y., Kirchhoff, K., Bartels, C., and Bilmes, J. Submodular subset selection for large-scale speech training data. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3311–3315. IEEE, 2014.

Wenckstern, J., Jain, E., Vasilev, K., Pariset, M., Wicki, A., Gut, G., and Bunne, C. AI-powered virtual tissues from spatial proteomics for clinical diagnostics and biomedical discovery. *arXiv preprint arXiv:2501.06039*, 2025.

Wolf, F. A., Angerer, P., and Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, 2018.

Xiao, H., Rasul, K., and Vollgraf, R. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Yang, Y., Shen, H. T., Ma, Z., Huang, Z., and Zhou, X. $\ell_{2,1}$-norm regularized discriminative feature selection for unsupervised learning. In *IJCAI International Joint Conference on Artificial Intelligence*, 2011.

Yehuda, O., Dekel, A., Hacohen, G., and Weinshall, D. Active Learning Through a Covering Lens. *Advances in Neural Information Processing Systems*, 35:22354–22367, 2022.

Yellott Jr, J. I. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2): 109–144, 1977.

Yin, C., Qian, B., Cao, S., Li, X., Wei, J., Zheng, Q., and Davidson, I. Deep similarity-based batch mode active learning with exploration-exploitation. In *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 575–584. IEEE, 2017.

Zannone, S., Hernández-Lobato, J. M., Zhang, C., and Palla, K. Odin: Optimal discovery of high-value information using model-based deep reinforcement learning. In *ICML Real-world Sequential Decision Making Workshop*, 2019.

Zhai, X., Kolesnikov, A., Houlsby, N., and Beyer, L. Scaling Vision Transformers. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12104–12113, 2022.

Zhan, X., Wang, Q., Huang, K.-h., Xiong, H., Dou, D., and Chan, A. B. A Comparative Survey of Deep Active Learning. *arXiv preprint arXiv:2203.13450*, 2022.

Zhao, Z. and Liu, H. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th International Conference on Machine Learning*, pp. 1151–1157, 2007.

Zhdanov, F. Diverse mini-batch Active Learning. *arXiv preprint arXiv:1901.05954*, 2019.

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, 2017.

Zubek, V. B. and Dietterich, T. G. Pruning Improves Heuristic Search for Cost-Sensitive Learning. In *International Conference on Machine Learning (ICML)*, 2002.

Appendix

A Impact and Limitations **18**

B Related Work **18**

C Extended Mathematical Framework **19**

D Notation **20**

E Proofs **23**

E.1 Proof of Lemma 3.1 23

E.2 Proof of Proposition C.1 23

E.3 Proof of Theorem 3.5 24

F Technical Lemmas **25**

F.1 Technical Lemma 1: Inclusion of Mismatch Events 25

F.2 Technical Lemma 2: Moment Bound on Conditional Deviations 25

F.3 Technical Lemma 3: Bounding Disagreement Probability 26

F.4 Technical Lemma 4: Posterior Concentration 26

F.5 Technical Lemma 5: Spectral Misalignment Bound 27

G Experimental Details **28**

G.1 Datasets 28

G.2 Architectures and Backbone Details 30

G.3 Evaluation Protocol 31

G.4 Baselines 31

G.5 Implementation Details of ECHO-k 33

H Additional Experiments **36**

H.1 Controlled Synthetic Experiment 36

H.2 Extended Baseline Comparisons 36

H.3 Comparison with Adaptive Acquisition Policies 36

H.4 Comparison with Supervised Acquisition 37

H.5 Cost of Task Agnosticism 37

H.6 Robustness to Partially Observed Data 38

H.7 Heterogeneous Acquisition Costs 38

H.8 Adaptive Stopping 39

H.9 Empirical Validation of the Near Isometry Assumption 40

H.10 Analysis of Markers Acquired by ECHO-k on Spatial Proteomics 42

A Impact and Limitations

ECHO-k enables multimodal representation models to operate under test-time measurement budgets by learning task-agnostic, label-free acquisition policies. Instead of requiring downstream labels to decide which view to acquire next, the method uses the model's own full-view representation as a self-supervised reference. This addresses a practical deployment bottleneck: multimodal and scientific foundation models often benefit from additional inputs, but acquiring all modalities can be infeasible due to cost, latency, limited tissue or sample material, scarce instrumentation time, or experimental constraints. In scientific settings, ECHO-k could help prioritize which measurements, assays, markers, sensors, or experiments to run next under limited resources. In clinical settings, it could support budget-aware acquisition of follow-up tests or multimodal patient measurements, while keeping downstream decision-making within validated and expert-supervised workflows. The main limitations concern deployment specificity rather than the general formulation. Acquisition costs and constraints may vary across domains, so practical use requires defining the relevant budget, availability, and cost model for the target setting. The method also benefits from reference representations that are informative enough to guide policy learning; our robustness experiments suggest that these references need not always be complete, but the appropriate level of missingness may depend on the application and backbone. Finally, as with any method intended for scientific or clinical use, policies should be evaluated on the intended endpoint and deployment distribution before adoption. Moreover, ECHO-k preserves whatever information the pretrained backbone encodes, and does not assert that the full-view representation $\boldsymbol{z}^\star$ is optimal for every downstream task; any task-irrelevant information in $\boldsymbol{z}^\star$ is therefore inherited from the backbone. Our experiments evaluate transfer across downstream objectives, but do not test explicit distribution shift between the reference-construction data and the deployment domain. Similarly, the full-view representation is not always the best training target: masked references outperform it on MNIST-14 at 20% and 40% budgets and on hPancreas at 20% (§ H), suggesting that masking can act as a mild regularizer during training. Our robustness sweep considers randomly masked reference views and does not identify where partial references fail, since even at 50% masking ECHO-k remains competitive with strong baselines. A natural extension is structured incompleteness, where an entire assay or modality is missing for a subset of samples, and understanding how the reference representation interacts with that structure is an interesting direction for future work. These considerations leave substantial room for future work on richer cost models, uncertainty-aware acquisition, and prospective validation in real-world acquisition workflows.

B Related Work

Classical feature selection methods aim to find a subset of relevant features that maximizes model performance while excluding irrelevancies (Kohavi & John, 1997; Guyon & Elisseeff, 2003). Early surveys (Blum & Langley, 1997) reviewed both empirical advances and theoretical definitions of feature relevance, formalizing the notions of strong and weak relevance. Several works cast the problem as submodular optimization (Killamsetty et al., 2021; Jain et al., 2023; Wei et al., 2014; Coleman et al., 2019), though their end goal is to improve performance under limited budget during training rather than at test time. Previous work has also centered around greedy heuristics with theoretical guarantees for sequential search (Golovin & Krause, 2011). Another line of work focuses on instance-specific or personalized feature selection (Melville et al., 2004; Li et al., 2018). Beyond selecting a subset of features, a related field is **feature importance quantification**: given a fixed dataset and (most commonly) a target, the goal is to isolate features that contribute significantly to predictions (Guyon & Elisseeff, 2003; Settles, 2009; Li et al., 2017; Lundberg & Lee, 2017; Sristi et al., 2023; Jain et al., 2023; Lifshitz et al., 2025). Model-specific methods such as (Breiman, 2001) leverage the structure of the model to assign importance ranks, while model-agnostic methods (Fisher et al., 2019) rely on importance metrics applicable to any model.

Active learning (AL) incorporates a sample selection policy in the training loop in an online fashion: at each time t, a new sample is queried to augment the existing dataset (Settles, 2009; Zhan et al., 2022; Tharwat & Schenck, 2023; Li et al., 2025). AL methods are particularly well suited for settings where acquiring new samples is expensive. They can be broadly split into three types of strategies (Zhan et al., 2022). *Uncertainty-based* querying strategies select points based on the estimated information they provide to the learning algorithm, including maximum entropy selection (Shannon, 1948), confidence-based selection (Wang & Shang, 2014), and adversarial methods (Ducoffe & Precioso, 2018). *Diversity-based strategies* aim to increase the coverage of samples in the dataset,

for instance through core-set approaches that build subsets approximating the full dataset (Sener & Savarese, 2017; Claici et al., 2018; Feldman, 2019; Kim & Shin, 2022). Lastly, *combined* strategies mix both approaches (Yin et al., 2017; Zhdanov, 2019; Shui et al., 2020).

Feature acquisition (FA). Our method focuses on designing instance-specific *feature* acquisition strategies rather than *sample* acquisition policies (Settles, 2009; Kirsch et al., 2019; Holzmüller et al., 2023; Kolossov et al., 2023; Gupte et al., 2024). FA strategies can be broadly split (Rahbar et al., 2025) into *greedy* strategies that sequentially acquire information by maximizing expected utility, and *non-greedy* methods that plan acquisition over multiple time steps (Zubek & Dietterich, 2002; Dulac-Arnold et al., 2011; Shim et al., 2018; Huang et al., 2018, 2025; Norcliffe et al., 2025; Piskorz et al., 2025), often using reinforcement learning (Shim et al., 2018; Zannone et al., 2019; Kachuee et al., 2019). This problem has also been studied under the name of *Bayesian experiment design* (BED), which aims to optimize the distribution of ξ, the random variable controlling the experiment, in order to maximize the information gain obtained by running the experiment parametrized by ξ (Chaloner & Verdinelli, 1995; Sebastiani & Wynn, 2000; Cavagnaro et al., 2010; Ryan et al., 2016; Rainforth et al., 2024). Classical BED methods sequentially update a model from a stream of incoming experiments; recently, the focus has shifted towards policy learning (Foster et al., 2021; Blau et al., 2022, 2023) and decision-aware experiment design (Huang et al., 2024).

Task-agnostic acquisition. Contrarily to previous work, our proposed method is fundamentally task-agnostic. While some active learning and feature selection strategies have considered weakly or unlabeled settings (Ash et al., 2019; Yehuda et al., 2022; Lifshitz et al., 2025; Kobayashi et al., 2025), task-agnostic feature acquisition has received little attention. Closest to our setup are Li & Oliva (2021), who consider a partially supervised setup and tackle this problem by training a surrogate reward model, and introduce an active instance recognition task in which the goal is to select features to reduce uncertainty.

Theoretical aspects of FA have received little attention, both in the supervised and unsupervised setting. Chen et al. (2015) study a probabilistic model inspired by Bayesian experiment design, in which the goal is to learn a policy that sequentially acquires features to infer the value of a target. Kobayashi et al. (2025) and Covert et al. (2023) provide theoretical insights in the supervised case. Our work is the first to consider the unsupervised case and to provide a bound on the gap between the greedy oracle and the optimal test-time policy.

C Extended Mathematical Framework

Data generation. We consider n i.i.d. instances. For each $i \in [\![1, n]\!]$, the data $\boldsymbol{x}^i = (\boldsymbol{m}_1^i, \dots, \boldsymbol{m}_K^i)$ is drawn from a joint distribution over $\mathcal{M}^1 \times \cdots \times \mathcal{M}^K$. We use *view*, *modality*, and *feature group* interchangeably, as our framework applies uniformly to all settings.

Representation model. Each view $\boldsymbol{m}_k$ is processed by a view-specific encoder $\Phi_k : \mathcal{M}^k \to \mathbb{R}^{d_k}$. The encoded views are combined through a fixed aggregation function $\mathbf{A}_\mathbf{W} : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_K} \to \mathbb{R}^p$, yielding the full-view representation $\boldsymbol{z}^* = \mathbf{A}_\mathbf{W}(\Phi_1(\boldsymbol{m}_1), \dots, \Phi_K(\boldsymbol{m}_K))$. This subsumes both multimodal models (each $\boldsymbol{m}_k$ is a distinct modality: image, audio, protein marker, ...) and unimodal models (features partitioned into jointly acquired groups processed by a shared encoder). For notational simplicity, we drop the encoders in subsequent expressions and treat views as directly mapped into the latent space.

Sequential acquisition. Let $\boldsymbol{P}_t \subseteq [\![1, K]\!]$ denote the acquired set after t steps and $\boldsymbol{z}(\boldsymbol{P}_t) := \mathbf{A}_\mathbf{W}(\{\Phi_k(\boldsymbol{m}_k) : k \in \boldsymbol{P}_t\})$ the induced partial representation, with unobserved views treated as missing. The finite-horizon acquisition problem is to select a sequence $(k_1, \dots, k_T)$, with $k_{t+1} \notin \boldsymbol{P}_t$, minimizing the representation-recovery loss

$$\mathcal{L}(\boldsymbol{P}_t) := \|\boldsymbol{z}(\boldsymbol{P}_t) - \boldsymbol{z}^\star\|_2^2.$$

Equivalently, the policy maximizes cumulative progress toward $\boldsymbol{z}^\star$ along the acquisition trajectory. The ideal one-step greedy rule is

$$k_{t+1} \in \arg\min_{k \notin \boldsymbol{P}_t} \|\boldsymbol{z}(\boldsymbol{P}_t \cup \{k\}) - \boldsymbol{z}^*\|_2^2, \tag{6}$$

which is not deployable at test time, since it requires access to both $\boldsymbol{z}^\star$ and the representation after acquiring each candidate view.

Markov decision process formulation. We frame sequential acquisition as an instance-specific MDP with a continuous state space and evolving action set. Dropping the instance superscript for clarity, the components are:

- **State.** At step t, the state is $s_t = (M_t, \boldsymbol{x}_{\boldsymbol{P}_t}, \boldsymbol{z}(\boldsymbol{P}_t), b_t)$, where $M_t \in \{0,1\}^K$ is the acquisition mask, $\boldsymbol{x}_{\boldsymbol{P}_t}$ are the observed view values, $\boldsymbol{z}(\boldsymbol{P}_t)$ is the current partial representation, and b_t is the remaining budget. To maintain a fixed state space across steps, unobserved views are set to `NA`, so the augmented space is $\big(\mathcal{M}^1 \cup \{\texttt{NA}\}\big) \times \cdots \times \big(\mathcal{M}^K \cup \{\texttt{NA}\}\big)$ (Norcliffe et al., 2025).
- **Actions.** The available actions at step t are the unacquired views: $\mathcal{A}(s_t) = \{k \in [\![1, K]\!] : k \notin \boldsymbol{P}_t\}$.
- **Transitions.** Upon selecting k_{t+1}, its value is drawn from $\boldsymbol{m}_{k_{t+1}} \sim p_{\texttt{data}}(\cdot \mid \boldsymbol{m}_k, k \in \boldsymbol{P}_t)$, the state updates to $\boldsymbol{P}_{t+1} = \boldsymbol{P}_t \cup \{k_{t+1}\}$, and $\boldsymbol{z}(\boldsymbol{P}_{t+1})$ is recomputed. The transition probability between states differing only in coordinate k_t is
$$\mathbb{P}[\boldsymbol{P}_t \mid \boldsymbol{P}_{t-1}] = \mathbb{P}[a_t = k_t \mid \boldsymbol{P}_{t-1}] \cdot p_{\texttt{data}}(\boldsymbol{m}_{k_t} \mid \boldsymbol{m}_k,\, k \in \boldsymbol{P}_{t-1})\,.$$
- **Reward.** The reward at step $t+1$ quantifies progress toward the reference-view representation:
$$r_{t+1}(\boldsymbol{P}_{t+1}) := \|\boldsymbol{z}(\boldsymbol{P}_t) - \boldsymbol{z}^{\text{ref}}\|_2^2 - \|\boldsymbol{z}(\boldsymbol{P}_{t+1}) - \boldsymbol{z}^{\text{ref}}\|_2^2.$$
where $\boldsymbol{z}^{\text{ref}}$ can be the full view embedding $\boldsymbol{z}^*$ or the embedding from all the views (can be a subset of full views) measured for the underlying data point.

Selection policies. A stochastic selection policy is a map $\pi_\theta(\cdot \mid s_t) \in \Delta(\mathcal{A}(s_t))$, where $\Delta(\mathcal{A}(s_t))$ is the probability simplex over available actions. During training, fully observed samples allow $\boldsymbol{z}^\star$ to be computed and used as a self-supervised target — no downstream labels are required. When only a reference subset R_i is observed for instance i, $\boldsymbol{z}(R_i)$ serves as a partial target. At deployment, the learned policy is fixed and acts solely from the current state s_t.

Connection to sequential decision making. This MDP is an instance of a K-armed bandit with correlated arms and a single-pull constraint per arm (Gupta et al., 2021); see also Guha & Munagala (2007); Chakrabarti et al. (2008); Carrara et al. (2019) for related setups. The sequential decision problem is Markovian with respect to $(\boldsymbol{P}_t, \boldsymbol{z}(\boldsymbol{P}_t))$: given the current partial representation and acquired set, future acquisition values are independent of the full history of selections.

D Notation

Notation	Meaning
Representation-acquisition setup	
n	Number of i.i.d. samples.
K	Number of possible views in the general acquisition problem.
$\boldsymbol{m}_k^i$	The k-th view of sample i. A view may be a modality or a jointly acquired feature group.
$\mathcal{M}^k$	Space of possible values for view k.
$\Phi_k : \mathcal{M}^k \to \mathbb{R}^{d_k}$	View-specific encoder for view k.
$\mathbf{A}_\mathbf{W}$	Fixed aggregation map combining encoded views into a representation.
$\boldsymbol{z}^\star$	Full-view representation, obtained from all available views.
$\boldsymbol{P}_t \subseteq \{1, \ldots, K\}$	Set of acquired views after t acquisition steps in the general MDP formulation.
$\boldsymbol{z}(\boldsymbol{P}_t)$	Partial representation induced by the acquired set $\boldsymbol{P}_t$.
T	Acquisition budget.
$s_t = (\boldsymbol{P}_t, \boldsymbol{z}(\boldsymbol{P}_t), b_t)$	Deployment state of the acquisition policy.
b_t	Remaining or current budget component in the MDP state.

Notation	Meaning
$\mathcal{L}(\boldsymbol{P}_t)$	Representation-recovery loss, $\mathcal{L}(\boldsymbol{P}_t) = \|\boldsymbol{z}(\boldsymbol{P}_t) - \boldsymbol{z}^\star\|_2^2$.
Stylized linear model	
d	Number of views/features in the stylized linear theory.
m	Dimension of the latent linear representation.
$\boldsymbol{x} \in \mathbb{R}^d$	Random view vector in the linear theory.
x_j	The j-th component of the view vector $\boldsymbol{x}$.
$\mathbf{W} \in \mathbb{R}^{m \times d}$	Fixed linear aggregation matrix.
$\mathbf{w}_j$	The j-th column of $\mathbf{W}$.
$\boldsymbol{z}^* = \mathbf{W}\boldsymbol{x}$	Latent representation in the linear theory.
$\mathcal{I}_t$	Information available to a policy at time t. This differs between oracle and test-time policies.
$\mathcal{F}_t = \sigma(x_k : k \in \boldsymbol{P}_t)$	Filtration generated by the views observed up to time t.
$\mathcal{L}(\boldsymbol{P}_t \mid \mathcal{I}_t)$	Conditional expected squared reconstruction error of selected set $\boldsymbol{P}_t$ given information $\mathcal{I}_t$.
Policies and selected sets	
$\boldsymbol{P}_t^\star$	Set selected by the oracle policy after t steps.
$\widehat{\boldsymbol{P}}_t$	Set selected by the optimal test-time policy after t steps.
$k_t^\star$	View selected by the oracle policy at step t.
$\widehat{k}_t$	View selected by the optimal test-time policy at step t.
$\widehat{\boldsymbol{P}}_{-s}$	Set of views still available to the test-time policy at step s.
Losses, energies, and regrets	
$\boldsymbol{E}_j$	Energy of view j, defined by $\boldsymbol{E}_j := \|\mathbf{w}_j\|_2^2 x_j^2$.
$\boldsymbol{C}_{jj'}$	Cross-view interaction term, defined by $\boldsymbol{C}_{jj'} := \mathbf{w}_j^\top \mathbf{w}_{j'} x_j x_{j'}$.
$\mathcal{L}(\boldsymbol{P}_t, \boldsymbol{x})$	True latent reconstruction loss of selected set $\boldsymbol{P}_t$ on realization $\boldsymbol{x}$.
$\widetilde{\mathcal{L}}(\boldsymbol{P}_t, \boldsymbol{x})$	Proxy loss based only on marginal energy terms.
$\mathcal{R}_t$	True latent regret of the optimal test-time policy relative to the oracle policy.
$\widetilde{\mathcal{R}}_t$	Proxy regret of the optimal test-time policy relative to the oracle policy.
$\boldsymbol{Y}_{j,s}$	Centered energy error at step s, $\boldsymbol{Y}_{j,s} := \boldsymbol{E}_j - \mathbb{E}[\boldsymbol{E}_j \mid \mathcal{F}_s]$.
$\boldsymbol{R}_s$	Maximal conditional deviation spread among available views at step s.
$M_j^{(p)}$	p-th moment of the energy of view j, $M_j^{(p)} := \mathbb{E}[\boldsymbol{E}_j^p]$.
$M_{(1)}^{(p)} \geq \cdots \geq M_{(d)}^{(p)}$	Sorted p-th energy moments.
A_t	Moment envelope, $A_t := \sum_{i=1}^d M_{(i)}^{(p)} \min(t, d-i+1)$.
p, q	Hölder conjugate exponents, with $p \in (1, 1/(2\sigma))$ and $q = p/(p-1)$.
Mismatch and ranking events	
G_t	Views selected only by the oracle policy at time t.
B_t	Views selected only by the optimal test-time policy at time t.
$\mathrm{RK}_A(k)$	Rank of $\boldsymbol{E}_k$ among the energy values $(\boldsymbol{E}_j)_{j \in A}$.
$\mathrm{BAD}(s,t)$	Bad ranking event, $\mathrm{BAD}(s,t) := \{\mathrm{RK}_{\widehat{\boldsymbol{P}}_{-s}}(\widehat{k}_{s+1}) > t - s\}$.
$D_t(\lambda)$	Disagreement envelope controlling the expected number of oracle–test-time mismatches.
N_t	Effective size term in the true regret bound, $N_t := d - t + D_t$.
Gumbel latent correlation model	

Notation	Meaning
Z	Discrete latent factor governing correlations between views.
$z_1, \dots, z_M$	Possible latent states of Z.
p_m	Prior probability of latent state z_m.
$p_{\min}$	Minimum latent-state probability, $p_{\min} := \min_{m\in[M]} p_m$.
$\boldsymbol{\eta}, \boldsymbol{\beta} \in \mathbb{R}^d$	Deterministic location and latent-loading vectors in the Gumbel model.
ε_j	Independent Gumbel noise term for view j.
σ	Gumbel scale parameter.
a_j	Squared column norm, $a_j := \|\mathbf{w}_j\|_2^2$, when writing utilities in the form $\log a_j + 2\eta_j + 2\beta_j z_m$.
$u_j^{(m)}$	Utility of view j under latent state z_m, $u_j^{(m)} := \log a_j + 2\eta_j + 2\beta_j z_m$.
$\theta_j^{(m)}$	Plackett–Luce weight, $\theta_j^{(m)} := \exp(u_j^{(m)}/(2\sigma))$.
γ	Minimum utility separation, $\gamma := \min_{m\in[M]} \min_{j\neq j'} \lvert u_j^{(m)} - u_{j'}^{(m)}\rvert$.
$R_{\max}$	Maximum latent-state variation in view weights, $R_{\max} := \max_{j\in[d]} \frac{\max_m e^{2\beta_j z_m}}{\min_m e^{2\beta_j z_m}}$.
ε_γ	Posterior accuracy threshold, $\varepsilon_\gamma := \frac{e^\gamma - 1}{R_{\max} + e^\gamma - 1}$.
$\rho_{\mathrm{PL}}(k)$	Minimum Plackett–Luce selection probability over subsets of size k.
$\rho_\star$	Ranking-concentration parameter, $\rho_\star := 1 - e^{-\gamma/(2\sigma)}$.
$\alpha_m(\mathcal{F}_s)$	Posterior probability of latent state z_m after observing $\mathcal{F}_s$, $\alpha_m(\mathcal{F}_s) := \mathbb{P}[Z = z_m \mid \mathcal{F}_s]$.
$E_{\mathrm{post}}(s, z_m)$	Posterior failure event under true state z_m.
$c_j(m, m'; \lambda)$	Chernoff coefficient between the conditional distributions of view j under states m and m'.
ρ_λ	Worst-case Chernoff contraction coefficient, $\rho_\lambda := \max_{j\in[d]} \max_{m\neq m'} c_j(m, m'; \lambda)$.
C_λ	Posterior concentration constant, $C_\lambda := (M-1)\left(\frac{1-\varepsilon_\gamma}{\varepsilon_\gamma p_{\min}}\right)^\lambda$.
Spectral misalignment quantities	
$\mathcal{M}(\boldsymbol{P})$	Difference between true and proxy losses for selected set $\boldsymbol{P}$, $\mathcal{M}(\boldsymbol{P}) := \mathcal{L}(\boldsymbol{P}) - \widetilde{\mathcal{L}}(\boldsymbol{P})$.
$u(\boldsymbol{P})$	Vector of unobserved weighted view magnitudes, $u(\boldsymbol{P}) := [\mathbf{1}_{\{j\notin \boldsymbol{P}\}} \|\mathbf{w}_j\| x_j]_{j=1}^d$.
$\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')$	Set of indices not jointly selected by $\boldsymbol{P}$ and $\boldsymbol{P}'$, $\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}') := (\boldsymbol{P} \cap \boldsymbol{P}')^c$.
$\Gamma(\boldsymbol{P}, \boldsymbol{P}')$	Symmetric difference between two selected sets, $\Gamma(\boldsymbol{P}, \boldsymbol{P}') := \boldsymbol{P} \Delta \boldsymbol{P}'$.
$\boldsymbol{B}_J$	Matrix of normalized embedding columns indexed by J, $\boldsymbol{B}_J := (\mathbf{w}_u/\|\mathbf{w}_u\|)_{u\in J}$.
p_{eff}	Effective dimension parameter in the near-isometry assumption.
C, C'	Constants appearing in the near-isometry and spectral misalignment bounds.
B	Matrix of all normalized embedding columns, $B := [\mathbf{w}_j/\|\mathbf{w}_j\|]_{j=1}^d$.
H	Gram deviation matrix, $H := B^\top B - I$.
H_U	Restriction of the Gram deviation matrix to the index set $\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')$.

E Proofs

E.1 Proof of Lemma 3.1

Proof. Let G_t be the set of features selected only by the oracle policy at time t, and B_t the set of features selected only by the optimal test-time policy. First, observe that we have the trivial bound

$$|B_t| = |G_t| \leq \min(d-t, t)$$

We have

$$\widetilde{\mathcal{L}}(\widehat{\boldsymbol{P}}_t, \boldsymbol{x}) - \widetilde{\mathcal{L}}(\boldsymbol{P}_t^\star, \boldsymbol{x}) = \sum_{g \in G_t} \boldsymbol{E}_g - \sum_{b \in B_t} \boldsymbol{E}_b.$$

Write $\widehat{k}_2, \dots, \widehat{k}_t$ for the indices of the features selected by the optimal test-time policy. Since G_t and B_t are of equal size, there exists a permutation σ between B_t and G_t. Also, by definition of G_t, we know that $G_t \cap B_t = \{\varnothing\}$. By iterating over time steps, we have

$$\sum_{g \in G_t} \boldsymbol{E}_g - \sum_{b \in B_t} \boldsymbol{E}_b \leq \sum_{s=2}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \Big(\boldsymbol{E}_{\sigma(\widehat{k}_s)} - \boldsymbol{E}_{\widehat{k}_s} \Big).$$

Now, we also have

$$\begin{aligned}
&\sum_{s=2}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \Big(\boldsymbol{E}_{\sigma(\widehat{k}_s)} - \boldsymbol{E}_{\widehat{k}_s} \Big) \\
=&\sum_{s=2}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \Big(\boldsymbol{E}_{\sigma(\widehat{k}_s)} - \mathbb{E}\big[\boldsymbol{E}_{\sigma(\widehat{k}_s)} \mid \mathcal{F}_s\big] + \mathbb{E}\big[\boldsymbol{E}_{\sigma(\widehat{k}_s)} \mid \mathcal{F}_s\big] - \boldsymbol{E}_{\widehat{k}_s} - \mathbb{E}\big[\boldsymbol{E}_{\widehat{k}_s} \mid \mathcal{F}_s\big] + \mathbb{E}\big[\boldsymbol{E}_{\widehat{k}_s} \mid \mathcal{F}_s\big] \Big) \\
=&\sum_{s=2}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \Big(Y_{\sigma(\widehat{k}_s),s} - Y_{\widehat{k}_s,s} + \underbrace{\mathbb{E}\big[\boldsymbol{E}_{\sigma(\widehat{k}_s)} \mid \mathcal{F}_s\big] - \mathbb{E}\big[\boldsymbol{E}_{\widehat{k}_s} \mid \mathcal{F}_s\big]}_{\leq 0} \Big) \\
\leq&\sum_{s=2}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \Big(Y_{\sigma(\widehat{k}_s),s} - Y_{\widehat{k}_s,s} \Big).
\end{aligned}$$

where the last equality comes from optimality of the test time policy and the fact that $\sigma(\widehat{k}_s)$ is available, i.e. $\sigma(\widehat{k}_s) \in \widehat{\boldsymbol{P}}_{-s}$. By definition of $\boldsymbol{R}_s$, we have

$$\widetilde{\mathcal{L}}(\widehat{\boldsymbol{P}}_t, \boldsymbol{x}) - \widetilde{\mathcal{L}}(\boldsymbol{P}_t^\star, \boldsymbol{x}) \leq \sum_{s=2}^{t} \mathbf{1}_{\{\widehat{k}_s \in B_t\}} \boldsymbol{R}_s.$$

E.2 Proof of Proposition C.1

We prove the following Proposition.

Proposition E.1. *Take $p \in \big(1, \frac{1}{2\sigma}\big)$ and let $q = \frac{p}{p-1}$. Under Assumptions, we have*

$$\widetilde{\mathcal{R}}_t \leq A_t^{1/p} \left(\min \left\{ t, d-t, \frac{C_\lambda}{1-\rho_\lambda} + \frac{1}{e^{\rho_0} - 1} \right\} \right)^{1/q}$$

where $A_t = \sum_{i=1}^{d} M_{(i)} \min(t, d-i+1)$.

Proof. By Lemma 3.1, we have

$$\widetilde{\mathcal{R}}_t \leq \sum_{s=2}^{t} \mathbb{E}_{\boldsymbol{x}}\big[\mathbf{1}_{\{\widehat{k}_s \in B_t\}} \boldsymbol{R}_s\big].$$

By Hölder's inequality, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}}\big[\mathbf{1}_{\{\widehat{k}_s \in B_t\}} \boldsymbol{R}_s\big] \leq& \mathbb{E}_{\boldsymbol{x}}\big[\boldsymbol{R}_s^p\big]^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\big[\mathbf{1}_{\widehat{k}_s \in B_t}^q\big]^{\frac{1}{q}} \\
=& \mathbb{E}_{\boldsymbol{x}}\big[\boldsymbol{R}_s^p\big]^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\big[\mathbf{1}_{\widehat{k}_s \in B_t}\big]^{\frac{1}{q}}
\end{aligned}$$

for any $p \in \left(1, \frac{1}{2\sigma}\right)$ and $q = \frac{p}{p-1}$. Using Lemma F.1, we obtain

$$\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{R}_s^p\right]^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_{\widehat{k}_s \in B_t}\right]^{\frac{1}{q}} \leq \mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{R}_s^p\right]^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_{\mathrm{BAD}(s,t)}\right]^{\frac{1}{q}}$$

Now, using Lemma F.2, we have

$$\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{R}_s^p\right]^{\frac{1}{p}} \leq 4 \left(\sum_{i=1}^{d-s} M_{(i)}^{(p)}\right)^{\frac{1}{p}}$$

where $M_{(j)}^{(p)} := \mathbb{E}[\boldsymbol{E}_j^p]$, and $M_{(1)}^{(p)} \geq ... \geq M_{(d)}^{(p)}$ are the sorted p-moments. Hence, we have

$$\widetilde{\mathcal{R}}_t \leq \sum_{s=2}^{t} \mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{R}_s^p\right]^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_{\mathrm{BAD}(s,t)}\right]^{\frac{1}{q}} \leq \sum_{s=2}^{t} 4 \left(\sum_{i=1}^{d-s} M_{(i)}^{(p)}\right)^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_{\mathrm{BAD}(s,t)}\right]^{\frac{1}{q}}.$$

For any $s \leq d$, let $S_{d-s} := \sum_{i=1}^{d-s} M_{(i)}^{(p)}$ and $A_t = \sum_{s=0}^{t-1} S_{d-s} = \sum_{s=0}^{t-1} \sum_{i=1}^{d-s} M_{(i)}^{(p)} = \sum_{i=1}^{d} M_{(i)}^{(p)} \min(t, d-i+1)$. Then,

$$\sum_{s=2}^{t} 4 \left(\sum_{i=1}^{d-s} M_{(i)}^{(p)}\right)^{\frac{1}{p}} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_{\mathrm{BAD}(s,t)}\right]^{\frac{1}{q}} \leq A_t^{1/p} \left(\sum_{s=0}^{t-1} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_{\mathrm{BAD}(s,t)}\right]\right)^{1/q} = A_t^{1/p} \left(\sum_{s=0}^{t-1} \mathbb{P}\big(\mathrm{BAD}(s,t)\big)\right)^{1/q}$$

by Hölder's inequality. Now, combining this with Lemma F.3, we have the upper bound

$$A_t^{1/p} \left(\sum_{s=0}^{t-1} \mathbb{P}\big(\mathrm{BAD}(s,t)\big)\right)^{1/q} \leq A_t^{1/p} \left(\min\left\{t, d-t, \sum_{s=0}^{t-1} \mathbb{E}_Z\Big[\mathbb{P}\big(E_{\mathrm{post}}(s, z_{m^\star})\big)\Big] + e^{-(t-s)\rho_{\mathrm{PL}}(d-t)}\right\}\right)^{1/q}$$

which can in turn be bounded from above by

$$A_t^{1/p} \left(\sum_{s=0}^{t-1} \mathbb{E}_Z\Big[\mathbb{P}\big(E_{\mathrm{post}}(s, z_{m^\star})\big)\Big] + e^{-k\rho_{\mathrm{PL}}(d-t)}\right)^{1/q} \leq A_t^{1/p} \left(\min\left\{t, d-t, \sum_{s=1}^{t} C_\lambda \rho_\lambda^s + e^{-(t-s)\rho_{\mathrm{PL}}(d-t)}\right\}\right)^{1/q}$$

where $C_\lambda := (M-1)\left(\frac{1-\varepsilon_\gamma}{\varepsilon_\gamma p_{\min}}\right)^\lambda$ using Lemma F.4. The sum in the right-hand side term of this bound can be further upper bounded by

$$\frac{C_\lambda}{1-\rho_\lambda} + \frac{1}{e^{\rho_0} - 1}.$$

This yields the final bound

$$\widetilde{\mathcal{R}}_t \leq A_t^{1/p} \left(\min\left\{t, d-t, \frac{C_\lambda}{1-\rho_\lambda} + \frac{1}{e^{\rho_0}-1}\right\}\right)^{1/q}.$$

E.3 Proof of Theorem 3.5

Proof. From Lemma F.5, we have

$$\mathcal{L}(\widehat{\boldsymbol{P}}_t, \boldsymbol{x}) - \mathcal{L}(\boldsymbol{P}_t^\star, \boldsymbol{x}) \leq \widetilde{\mathcal{L}}(\widehat{\boldsymbol{P}}_t, \boldsymbol{x}) - \widetilde{\mathcal{L}}(\boldsymbol{P}_t^\star, \boldsymbol{x}) + C' \left(\sqrt{\frac{|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|}{p_{\mathrm{eff}}}} + \frac{|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|}{p_{\mathrm{eff}}}\right) \sqrt{\sum_{j \in \boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')} \boldsymbol{E}_j} \sqrt{\sum_{j \in \Gamma(\boldsymbol{P}, \boldsymbol{P}')} \boldsymbol{E}_j}.$$

Taking expectations over $\boldsymbol{x}$ and using the Cauchy-Schwarz inequality, we obtain

$$\mathcal{R}_t \leq \widetilde{\mathcal{R}}_t + C' \left(\frac{\mathbb{E}_{\boldsymbol{x}}\left(\sqrt{|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|}\right)}{p_{\mathrm{eff}}} + \frac{\mathbb{E}_{\boldsymbol{x}}\left(|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|\right)}{p_{\mathrm{eff}}}\right) \mathbb{E}_{\boldsymbol{x}} \sqrt{\sum_{j \in \boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')} \boldsymbol{E}_j} \mathbb{E}_{\boldsymbol{x}} \sqrt{\sum_{j \in \Gamma(\boldsymbol{P}, \boldsymbol{P}')} \boldsymbol{E}_j}$$

and using Jensen's inequality, we obtain

$$\mathcal{R}_t \leq \widetilde{\mathcal{R}}_t + C' \left(\frac{\sqrt{\mathbb{E}_{\boldsymbol{x}}|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|}}{p_{\mathrm{eff}}} + \frac{\mathbb{E}_{\boldsymbol{x}}\left(|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|\right)}{p_{\mathrm{eff}}}\right) \sqrt{\mathbb{E}_{\boldsymbol{x}} \sum_{j \in \boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')} \boldsymbol{E}_j} \cdot \sqrt{\mathbb{E}_{\boldsymbol{x}} \sum_{j \in \Gamma(\boldsymbol{P}, \boldsymbol{P}')} \boldsymbol{E}_j}.$$

We can in turn bound this using F.3 by

$$\widetilde{\mathcal{R}}_t + C'\left(\frac{\sqrt{\mathbb{E}_{\boldsymbol{x}}|\boldsymbol{U}(\boldsymbol{P},\boldsymbol{P}')|}}{p_{\text{eff}}} + \frac{\mathbb{E}_{\boldsymbol{x}}\Big(|\boldsymbol{U}(\boldsymbol{P},\boldsymbol{P}')|\Big)}{p_{\text{eff}}}\right)\sqrt{\mathbb{E}_{\boldsymbol{x}}\sum_{j\in\boldsymbol{U}(\boldsymbol{P},\boldsymbol{P}')}\boldsymbol{E}_j}$$
$$\leq \widetilde{\mathcal{R}}_t + C'\left(\frac{\sqrt{d-t+D_t}}{p_{\text{eff}}} + \frac{d-t+D_t}{p_{\text{eff}}}\right)\sqrt{D_t(d-t+D_t)}.$$

F Technical Lemmas

F.1 Technical Lemma 1: Inclusion of Mismatch Events

Recall that for any subset $A \subseteq \{1,\dots,d\}$, $\text{RK}_A(k)$ is the rank of $\boldsymbol{E}_k$ among all $(\boldsymbol{E}_j)_{j\in A}$. Consider the event

$$\text{BAD}(s,t) := \Big\{\text{RK}_{\widehat{\boldsymbol{P}}_{-s}}(\widehat{k}_{s+1}) > t-s\Big\} = \Big\{\boldsymbol{E}_{\widehat{k}_{s+1}} < \boldsymbol{E}_{(t-s)}\Big\}.$$

Lemma F.1. *We have*

$$\sum_{s=2}^{t}\mathbf{1}_{\{\widehat{k}_s\in B_t\}}\boldsymbol{R}_s \leq \sum_{s=2}^{t}\mathbf{1}_{\text{BAD}(s,t)}\boldsymbol{R}_s.$$

Proof. At time $s \leq t$, the test-time policy has selected at most $s-1$ items from the top $t-1$ values of $(\boldsymbol{E}_k)_{k=1}^d$, the most favorable case being when it selects the same items as the oracle policy. Hence, there are at least $t-s-1$ items from the top $t-1$ values of $(\boldsymbol{E}_k)_{k=1}^d$ in the set $\widehat{\boldsymbol{P}}_{-s}$. Therefore, if $\widehat{k}_s \in B_t$, then by definition it is not in the top $t-s-1$ items, because it would otherwise also belong to $G_t = \boldsymbol{P}_t^\star \setminus B_t$. Hence,

$$\boldsymbol{E}_{\widehat{k}_s} \leq \boldsymbol{E}_{(t-s-1)}.$$

We have hence proved that $\{\widehat{k}_s \in B_t\} \subseteq \text{BAD}(s,t)$, which proves our claim.

F.2 Technical Lemma 2: Moment Bound on Conditional Deviations

Lemma F.2. *Let $M_j^{(p)} := \mathbb{E}[\boldsymbol{E}_j^p]$, and $M_{(1)}^{(p)} \geq \dots \geq M_{(d)}^{(p)}$ be the sorted (p)-moment order. Then, for every s and for any $p \in \big(1, \frac{1}{2\sigma}\big)$*

$$\mathbb{E}[\boldsymbol{R}_s^p]^{1/p} \leq 4\left(\sum_{i=1}^{d-s} M_{(i)}\right)^{\frac{1}{p}}$$

Proof. Recall that $\boldsymbol{R}_s := \max_{i,j\in\widehat{\boldsymbol{P}}_{-t}}(\boldsymbol{Y}_{i,s} - \boldsymbol{Y}_{j,s})$. We have

$$\boldsymbol{R}_s \leq \max_{i,j\in\widehat{\boldsymbol{P}}_{-t}}\big(|\boldsymbol{Y}_{i,s}| + |\boldsymbol{Y}_{j,s}|\big) \leq 2\max_{i,j\in\widehat{\boldsymbol{P}}_{-t}}|\boldsymbol{Y}_{j,s}|.$$

Take $p \in \big(1, \frac{1}{2\sigma}\big)$. Since $\max_{i,j\in\widehat{\boldsymbol{P}}_{-t}}|\boldsymbol{Y}_{j,s}| \leq \Big(\sum_{j\in\widehat{\boldsymbol{P}}_{-t}}|\boldsymbol{Y}_{j,s}|^p\Big)^{1/p}$, we have $\boldsymbol{R}_s^p \leq 2^p\sum_{j\in\widehat{\boldsymbol{P}}_{-t}}|\boldsymbol{Y}_{j,s}|^p$. Using the inequality for any $k\in\{1,\dots,d\}$

$$|\boldsymbol{E}_k - \mathbb{E}[\boldsymbol{E}_k \mid \mathcal{F}_s]|^p \leq 2^{p-1}\big(|\boldsymbol{E}_k|^p + |\mathbb{E}[\boldsymbol{E}_k \mid \mathcal{F}_s]|^p\big),$$

together with Jensen's inequality, we obtain $\mathbb{E}|\boldsymbol{Y}_{j,s}|^p \leq 2^p\mathbb{E}[|\boldsymbol{E}_j|^p] = 2^p M_j^{(p)}$. Therefore,

$$\mathbb{E}[\boldsymbol{R}_s^p] \leq 2^p\sum_{j\in\widehat{\boldsymbol{P}}_{-t}}\mathbb{E}|\boldsymbol{Y}_{j,s}|^p \leq 4^p\sum_{j\in\widehat{\boldsymbol{P}}_{-t}}M_j^{(p)}.$$

Since $|\widehat{\boldsymbol{P}}_{-t}| = d-s$ deterministically and $\sum_{j\in\widehat{\boldsymbol{P}}_{-t}}M_j^{(p)} \leq \sum_{i=1}^{d-s}M_{(i)}^{(p)}$, the claimed bound follows.

F.3 Technical Lemma 3: Bounding Disagreement Probability

For each state $m \in \{1, \dots, M\}$, define

$$u_j^{(m)} := \log a_j + 2\eta_j + 2\beta_j z_m, \qquad \theta_j^{(m)} := \exp\left(\frac{u_j^{(m)}}{2\sigma}\right).$$

Also let $\gamma := \min_{m\in[M]} \min_{j\neq j'} \left|u_j^{(m)} - u_{j'}^{(m)}\right| > 0$ and $R_{\max} := \max_{j\in[d]} \frac{\max_m e^{2\beta_j z_m}}{\min_m e^{2\beta_j z_m}} < \infty$. Finally, define $\varepsilon_\gamma := \frac{e^\gamma - 1}{R_{\max} + e^\gamma - 1} \in (0,1)$ and $\rho_{\mathrm{PL}}(k) := \min_{m\in[M]} \min_{\substack{J\subset[d]\\|J|=k}} \frac{\max_{j\in J} \theta_j^{(m)}}{\sum_{\ell\in J} \theta_\ell^{(m)}} \geq \frac{1}{k}$.

Lastly, define $\alpha_m(\mathcal{F}_s) := \mathbb{P}\Big[Z = z_m \mid \mathcal{F}_s\Big]$ and

$$E_{\mathrm{post}}(s, z_m) := \left\{\mathbb{P}\Big[Z = z_m \mid \mathcal{F}_s\Big] < 1 - \varepsilon_\gamma \mid Z = z_m\right\},$$
$$E_{\mathrm{post}}(s, Z) := \left\{\mathbb{P}\Big[Z = z_m \mid \mathcal{F}_s\Big] < 1 - \varepsilon_\gamma \mid Z\right\}.$$

Lemma F.3. *Assume that $\mathbb{P}\big[Z = z_m \mid \mathcal{F}_s\big] > 0$ for all z_m and $s \geq 1$. We have*

$$\mathbb{P}(\textsc{Bad}(s,t)) \leq \mathbb{E}_Z\Big[\mathbb{P}\big(E_{\mathrm{post}}(s, Z)\big)\Big] + e^{-k\rho_{\mathrm{PL}}(d-t)}.$$

Proof. Let us first work conditionally on $Z = z_{m^\star}$. Define $j^\star(z_{m^\star}) := \arg\max_{j\in\widehat{\boldsymbol{P}}_{-t}} u_j^{(m^\star)}$. Since $\gamma > 0$, this index is uniquely defined. We claim that on $E_{\mathrm{post}}(s, z_{m^\star})^c$, the test-time optimal policy selects $\widehat{k}_{s+1} = j^\star(z_{m^\star})$. To see this, note that $\mathbb{E}[\boldsymbol{E}_j \mid \mathcal{F}_s] \propto e^{\log a_j + 2\eta_j} \sum_{m=1}^M \alpha_m(\mathcal{F}_s) e^{2\beta_j z_m}$. Define $T_j := \sum_{m=1}^M \alpha_m(\mathcal{F}_s) e^{2\beta_j z_m}$.

On $E_{\mathrm{post}}(s, z_{m^\star})^c$, we have $\alpha_{m^\star}(\mathcal{F}_s) \geq 1 - \varepsilon_\gamma$. Hence:

$$T_j = \alpha_{m^\star}(\mathcal{F}_s) e^{2\beta_j z_{m^\star}} + \underbrace{\sum_{m\neq m^\star} \alpha_m(\mathcal{F}_s) e^{2\beta_j z_m}}_{>0} > (1-\varepsilon_\gamma) e^{2\beta_j z_{m^\star}}$$

where the strict inequality uses our assumption $\mathbb{P}\big[Z = z_m \mid \mathcal{F}_s\big] > 0$. On the other hand,

$$T_j \leq (1-\varepsilon_\gamma) e^{2\beta_j z_{m^\star}} + \varepsilon_\gamma R_{\max} e^{2\beta_j z_{m^\star}} = e^{2\beta_j z_{m^\star}} \left(1 - \varepsilon_\gamma + \varepsilon_\gamma R_{\max}\right).$$

Thus, up to a positive multiplicative constant: $(1-\varepsilon_\gamma) e^{u_j^{(m^\star)}} \leq \mathbb{E}[\boldsymbol{E}_j \mid \mathcal{F}_s] \leq (1 - \varepsilon_\gamma + \varepsilon_\gamma R_{\max})\, e^{u_j^{(m^\star)}}$.

For any $j \neq j^\star(z_{m^\star})$ in $\widehat{\boldsymbol{P}}_{-t}$:

$$\frac{\mathbb{E}[\boldsymbol{E}_{j^\star} \mid \mathcal{F}_s]}{\mathbb{E}[\boldsymbol{E}_j \mid \mathcal{F}_s]} > e^{u_{j^\star}^{(m^\star)} - u_j^{(m^\star)}} \frac{1-\varepsilon_\gamma}{1 - \varepsilon_\gamma + \varepsilon_\gamma R_{\max}} \geq e^\gamma \frac{1-\varepsilon_\gamma}{1 - \varepsilon_\gamma + \varepsilon_\gamma R_{\max}},$$

and this ratio exceeds 1 by definition of ε_γ, proving our claim.

From Assumption 3.2, conditional on $Z = z_{m^\star}$, we have $\log \boldsymbol{E}_j = u_j^{(m^\star)} + \xi_j$ where $\xi_j \sim \mathrm{Gumbel}(0, 2\sigma)$. The ranking follows a Plackett–Luce distribution with weights $\{\theta_j^{(m^\star)}\}$ (Yellott Jr, 1977; Ma et al., 2021). Under the PL sequential sampling view, the probability that $j^\star(z_{m^\star})$ is not selected in the first s draws is bounded as $\mathbb{P}\left(\mathrm{RK}_{\widehat{\boldsymbol{P}}_{-s}}(j^\star) > t - s \mid Z = z_{m^\star}\right) \leq e^{-sp^\star(z_{m^\star})} \leq e^{-s\rho_{\mathrm{PL}}(d-s)}$, where $p^\star(z_{m^\star}) := \theta_{j^\star}^{(m^\star)} / \sum_{\ell\in J_s} \theta_\ell^{(m^\star)} \geq \rho_{\mathrm{PL}}(d-s)$. This bound does not depend on $z_{m^\star}$, so we may average over Z.

Finally, decomposing $\textsc{Bad}(s,t)$ by latent states and conditioning on $E_{\mathrm{post}}(s, z_{m^\star})$ vs. its complement:

$$\mathbb{P}(\textsc{Bad}(s,t)) \leq \mathbb{P}(E_{\mathrm{post}}(s)) + e^{-k\rho_{\mathrm{PL}}(d-t)}.$$

F.4 Technical Lemma 4: Posterior Concentration

Recall that given $Z = z_m$, we have $X_j \sim \mathrm{Gumbel}(\eta_j + \beta_j z_m, \sigma)$ with density $f_{j,m}$. Fix $\lambda \in (0,1)$ and define the Chernoff coefficient (Nielsen, 2011) between states m and m' for feature j as

$c_j(m, m'; \lambda) := \int f_{j,m}(x)^{1-\lambda} f_{j,m'}(x)^{\lambda}\, dx$. For two Gumbel distributions with locations μ, μ' and common scale σ, this has the closed form $c(\mu, \mu'; \lambda) = \frac{e^{-\lambda\Delta}}{(1-\lambda)+\lambda e^{-\Delta}}$, where $\Delta := \frac{\mu-\mu'}{\sigma}$.

Let $\rho_\lambda := \max_{j\in[d]} \max_{m\neq m'} c_j(m, m'; \lambda) < 1$ and $p_{\min} := \min_{m\in[M]} p_m > 0$.

Lemma F.4. *For any value of $Z = z_{m^\star}$, any $\varepsilon \in (0, 1)$, any $s \geq 1$, and $\lambda \geq 1$,*

$$\mathbb{P}\left(\alpha_{m^\star}(\mathcal{F}_s) \leq 1 - \varepsilon \mid Z = z_{m^\star}\right) \leq (M-1)\left(\frac{1-\varepsilon}{\varepsilon p_{\min}}\right)^{\lambda} \rho_\lambda^s.$$

In particular, setting $\varepsilon = \varepsilon_\gamma$, we have $\mathbb{P}\big(E_{\text{post}}(s, z_{m^\star})\big) \leq C_\lambda \rho_\lambda^s$, where $C_\lambda := (M-1)\left(\frac{1-\varepsilon_\gamma}{\varepsilon_\gamma p_{\min}}\right)^{\lambda}$.

Proof. For any state $m \neq m^\star$, define the likelihood ratio $\kappa_s(m) := \frac{p_m}{p_{m^\star}} \prod_{r=1}^{s} \frac{f_{i_r,m}(X_{i_r})}{f_{i_r,m^\star}(X_{i_r})}$, where $X_{i_r} = \log x_{i_r}$. By Bayes' rule, $\alpha_{m^\star}(\mathcal{F}_s) = \frac{1}{1+\sum_{m\neq m^\star} \kappa_s(m)}$. If $\alpha_{m^\star}(\mathcal{F}_s) \leq 1 - \varepsilon$, then $\sum_{m\neq m^\star} \kappa_s(m) \geq \frac{\varepsilon}{1-\varepsilon}$. By the pigeonhole principle, there exists some $m \neq m^\star$ such that $\kappa_s(m) \geq \frac{\varepsilon}{(1-\varepsilon)(M-1)}$.

By a union bound and Markov's inequality with exponent $\lambda \geq 1$:

$$\mathbb{P}\left(\alpha_{m^\star}(\mathcal{F}_s) \leq 1 - \varepsilon \mid Z = z_{m^\star}\right) \leq \sum_{m\neq m^\star} \frac{\mathbb{E}[\kappa_s(m)^\lambda \mid Z = z_{m^\star}]}{\left(\frac{\varepsilon}{(1-\varepsilon)(M-1)}\right)^{\lambda}}.$$

Since $X_{i_1}, \ldots, X_{i_s}$ are independent draws from $f_{i_1,m^\star}, \ldots, f_{i_s,m^\star}$ given $Z = z_{m^\star}$, each factor satisfies $\mathbb{E}\left[\left(\frac{f_{i_r,m}(X_{i_r})}{f_{i_r,m^\star}(X_{i_r})}\right)^{\lambda} \,\middle|\, Z = z_{m^\star}\right] = c_{i_r}(m^\star, m; \lambda) \leq \rho_\lambda$. Therefore, $\mathbb{E}\left[\kappa_s(m)^\lambda \mid Z = z_{m^\star}\right] \leq p_{\min}^{-\lambda} \rho_\lambda^s$. Summing over all $M-1$ wrong states yields the claim. Since the bound does not depend on $z_{m^\star}$, it holds for any latent state.

F.5 Technical Lemma 5: Spectral Misalignment Bound

Define for any subset $\boldsymbol{P} \subset \{1, \ldots, d\}$: $\mathcal{M}(\boldsymbol{P}) := \mathcal{L}(\boldsymbol{P}) - \widetilde{\mathcal{L}}(\boldsymbol{P})$ and $u(\boldsymbol{P}) := \left[\mathbf{1}_{\{j\notin\boldsymbol{P}\}} \|\mathbf{w}_j\| \boldsymbol{x}_j\right]_{j=1}^{d}$.

Lemma F.5. *For any two sets $\boldsymbol{P}, \boldsymbol{P}'$, let $\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}') := \big(\boldsymbol{P} \cap \boldsymbol{P}'\big)^c$ and $\Gamma(\boldsymbol{P}, \boldsymbol{P}') = \boldsymbol{P} \Delta \boldsymbol{P}'$. We have*

$$|\mathcal{M}(\boldsymbol{P}) - \mathcal{M}(\boldsymbol{P}')| \leq C' \left(\sqrt{\frac{|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|}{p_{\text{eff}}}} + \frac{|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')|}{p_{\text{eff}}}\right) \sqrt{\sum_{j\in\boldsymbol{U}(\boldsymbol{P},\boldsymbol{P}')} \boldsymbol{E}_j} \sqrt{\sum_{j\in\Gamma(\boldsymbol{P},\boldsymbol{P}')} \boldsymbol{E}_j}$$

Proof. Let $B := \left[\frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}\right]_{j=1}^{d}$ and define $H := B^T B - I$. We have $\mathcal{M}(\boldsymbol{P}) - \mathcal{M}(\boldsymbol{P}') = (u(\boldsymbol{P}) - u(\boldsymbol{P}'))^\top H_U (u(\boldsymbol{P}) + u(\boldsymbol{P}'))$ where $H_U := B_U^\top B_U - I$ is the restriction of H to $\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')$. Therefore:

$$\begin{aligned} |\mathcal{M}(\boldsymbol{P}) - \mathcal{M}(\boldsymbol{P}')| &\leq \|H_U\|_{\text{op}} \|u(\boldsymbol{P}) - u(\boldsymbol{P}')\|_2 \|u(\boldsymbol{P}) + u(\boldsymbol{P}')\|_2 \\ &\leq \|H_U\|_{\text{op}} \|u(\boldsymbol{P}) - u(\boldsymbol{P}')\|_2 (\|u(\boldsymbol{P})\|_2 + \|u(\boldsymbol{P}')\|_2). \end{aligned}$$

Note that $|\boldsymbol{U}(\boldsymbol{P}, \boldsymbol{P}')| = d - t + k_t$, and $\|u(\boldsymbol{P}) - u(\boldsymbol{P}')\|_2^2 = \sum_{j\in\Gamma(\boldsymbol{P},\boldsymbol{P}')} \|w_j\|_2^2 x_j^2 = \sum_{j\in\boldsymbol{U}(\boldsymbol{P},\boldsymbol{P}')} V_j$. Using $\|u(\boldsymbol{P}) + u(\boldsymbol{P}')\|_2^2 \leq 4 \sum_{j\in U(\boldsymbol{P},\boldsymbol{P}')} \boldsymbol{E}_j$ and Assumption 3.4 concludes the proof.

Table 5: Overview of datasets used for experiments.

Dataset	Task	Train/Val/Test	# classes	# views	Source
Synthetic	Random isotropic regression	200/-/200	——	64	——
NSCLC patient cohort	Cancer grading	1.6k/400/400	3	40	Cords et al. (2024)
	Cancer subtyping		2	40	Cords et al. (2024)
hPancreas	Cell type annotation	9.5k/1k/4.2k	14	200	Ding et al. (2025)
MiniBooNE	Particle identification	65k/5k/6k	2	33	Roe (2005)
MNIST	Image classification	50k/10k/10k	10	196	LeCun et al. (1998)
Fashion-MNIST	Image classification	50k/10k/10k	10	196	Xiao et al. (2017)
MMEarth-Bench	Soil pH regression	5.7k/1.2k/350	——	12	Gordon et al. (2026)

G Experimental Details

G.1 Datasets

We evaluate ECHO-k and all baselines on six datasets spanning seven downstream tasks. Table 5 provides a summary. Below we describe each dataset in detail, including the downstream task, preprocessing choices, and the specific aspects that make it relevant for evaluating budgeted feature acquisition.

G.1.1 Vision Datasets: MNIST and Fashion-MNIST

MNIST (LeCun et al., 1998) and Fashion-MNIST (Xiao et al., 2017) are standard computer vision benchmarks consisting of images of handwritten digits and Zalando article images, respectively. Both have 10-class classification as the downstream task. Following Li & Oliva (2021), we downsample images from 28×28 to 14×14 resolution, treating each pixel as an independently acquirable feature and yielding $d = 196$ acquisition units. This downsampling increases the information density per pixel and makes the acquisition problem more challenging, since fewer pixels suffice to reach a given accuracy. We use a 70:15:15 split (50k/10k/10k). To ensure a controlled comparison, all acquisition methods are initialized from the same top-corner pixel (■), so differences in accuracy reflect only the quality of the subsequent acquisition strategy.

G.1.2 Physics: MiniBooNE

The MiniBooNE dataset (Roe, 2005) originates from the Fermilab MiniBooNE neutrino experiment. The goal is binary classification: distinguishing electron-neutrino signal events from muon-neutrino background events. The dataset contains $d = 33$ reconstructed physics variables derived from detector signals, including Cherenkov ring characteristics, event energy, and track reconstruction outputs. In practice, computing these reconstructed variables from raw detector data is expensive and the variables exhibit substantial mutual correlations, making it undesirable to rely on the full feature set at test time. We adopt the preprocessing protocol of Norcliffe et al. (2025), which standardizes features and removes duplicate events, and use a 70:15:15 split (65k/5k/6k). This dataset is particularly relevant for evaluating ECHO-k because the strong inter-feature correlations create a setting where greedy, myopic acquisition can be suboptimal: selecting the individually most informative variable first may be redundant with the second-most informative one, motivating the long-horizon planning that ECHO-k provides.

G.1.3 Single-Cell Transcriptomics: hPancreas

The hPancreas dataset (Ding et al., 2025) is a curated single-cell RNA-sequencing collection from human pancreatic tissue with 14 cell type labels spanning endocrine, exocrine, and immune populations. This dataset is widely used as a benchmark in the single-cell foundation model literature. From

the 1,200 genes available, we retain the top 200 most variable genes using Cell Ranger (Zheng et al., 2017) and ScanPy (Wolf et al., 2018), following standard practice for reducing noise from uninformative transcripts while preserving biologically meaningful variation. We use the preprocessing, gene filtering, and data splits provided by (Ding et al., 2025) (9.5k/1k/4.2k). Cell-type annotation serves as a representative downstream task: it is the most common evaluation objective in single-cell analysis and requires the acquired gene subset to preserve fine-grained phenotypic distinctions between closely related cell populations. Crucially, the ECHO-k policy is trained without any access to cell-type labels; the annotation task is used purely for post-hoc evaluation of the acquired gene panels.

G.1.4 Spatial Proteomics: NSCLC Patient Cohort

The NSCLC clinical cohort (Cords et al., 2024) consists of spatial proteomics data acquired via imaging mass cytometry (IMC) at 1 μm/px resolution with $d = 40$ semantically meaningful protein markers. The marker panel includes tumor markers (e.g., `panCK`), proliferation markers (e.g., `Ki-67`), immune cell markers (e.g., `CD4`, `CD8`, `CD68`), stromal markers, and vascular markers, providing a comprehensive view of the tumor microenvironment. Clinical annotations include cancer subtyping (2 classes: adenocarcinoma vs. squamous cell carcinoma) and cancer grading (3 classes), the latter being substantially more challenging due to subtler morphological and molecular differences between grades. We adopt the preprocessing pipeline described in (Wenckstern et al., 2025), which includes cell segmentation, marker normalization, and quality filtering. The experimental setup is inherently sequential: ECHO-k iteratively selects one protein marker at a time, the VirTues backbone updates its tissue embedding given the newly observed marker, and the updated embedding informs the next acquisition decision. This mirrors the real-world workflow in multiplexed imaging, where tissue sections are characterized through successive rounds of antibody staining and each additional round carries a cost. The acquisition budget is set to $T = 10$ markers (25% of the full panel), reflecting realistic constraints on antibody panel size in clinical multiplex imaging.

G.1.5 Geospatial Earth Observations: MMEarth-Bench

The MMEarth-Bench dataset (Gordon et al., 2026) is a global multimodal Earth-observation benchmark to evaluate models under heterogeneous sensing conditions. Each location is represented by 12 aligned geospatial views, including optical satellite imagery, radar measurements, terrain and elevation products, land-cover maps, canopy-height information, meteorological variables, and geolocation metadata. It provides a multimodal acquisition setting: each view corresponds to a distinct sensing product or environmental data source that may differ in availability, cost, spatial resolution, and information content. We treat each modality group as one acquisition unit and pair ECHO-k with a frozen multimodal geospatial representation model trained on the full set of available views. We evaluate downstream transfer on soil pH regression, reporting R^2 as a function of the acquired modality budget.

G.1.6 Synthetic Dataset

We generate synthetic data under the data generating process (DGP) of Assumption 3.2. We consider a gradient correlation family shared by a common latent state Z, where $Z \sim \text{Discrete}\left(\{z_m\}_{m=1}^M, \{p_m\}_{m=1}^M\right)$, $\varepsilon_j \sim \text{Gumbel}(0, \sigma)$ being independent noise. Each feature is generated as $X_j \sim \eta_j + \beta_j Z + \varepsilon_j$, where η_j and β_j are deterministic coefficients, and we develop a gradient-like correlation structure by letting

$$\beta_j := \beta_{\max}\left(1 - \alpha\frac{j-1}{d-1}\right),$$

and set $x_j = e^{X_j}$. In Fig. 5a, we show the data distribution when the support of Z has cardinality 3, and $z_m \in \{-1, 0, 1\}$. Here $\beta_{\max}$ is the maximum strength and $\alpha \in [0, 1]$ controls the decay rate. Under this construction, early features exhibit strong mutual correlations through Z, while

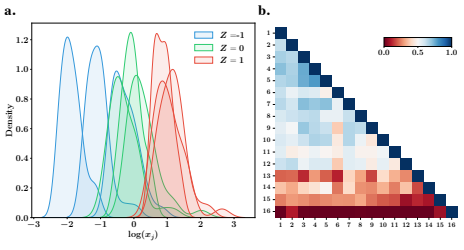


Figure 5: **Overview of the Latent Gumbel Correlation Model. a.** Distribution of $\log(\boldsymbol{x}_j)$ for 3 equiprobable latent states Z with support $\{-1, 0, 1\}$. **b.** Correlation of the features under the gradient correlation family described in § G.1.6.

later features are increasingly dominated by independent noise. The induced correlation structure is:

$$\mathrm{Corr}(X_i, X_j) = \frac{\beta_i \beta_j \mathrm{Var}(Z)}{\sqrt{\left(\beta_j^2 \mathrm{Var}(Z) + \frac{\pi^2 \sigma^2}{6}\right)\left(\beta_k^2 \mathrm{Var}(Z) + \frac{\pi^2 \sigma^2}{6}\right)}}.$$

Regret bound experiment (Fig. 2 We set $p = 25, d = 32$ (i.e., $\boldsymbol{W} \in \mathbb{R}^{p \times d}$), $\sigma = 0.2$, $M = 5$ with $z_m \in \{-1, -1/2, 0, 1/2, 1\}$ and $p_m = 0.2\ \forall m$. We consider $\eta_j = \frac{j-1}{d-1} - \frac{1}{2}$ for $j \in 1, \ldots, d$. For the gradient correlation family, we set $\beta_{\max} = 1.5$ and $\alpha = 0.9$.

Random isotropic regression experiment (Fig. 6). We increase the acquisition horizon by setting $d = 64$ and $p = 50$, keeping all other hyperparameters fixed. We sample the ground-truth regression weights as $\boldsymbol{W}^* \sim \mathcal{N}(0, \frac{1}{d} I)$ and generate labels as $\boldsymbol{y} = \boldsymbol{W}^* \boldsymbol{x} + \boldsymbol{\varepsilon'}$ where $\boldsymbol{\varepsilon'} \in \mathbb{R}^p$ has i.i.d. entries sampled from $\mathrm{Gumbel}(0, 1)$.

G.2 Architectures and Backbone Details

G.2.1 Masked Autoencoder (MNIST, Fashion-MNIST, MiniBooNE)

For image and tabular data, we train a masked autoencoder via self-supervised learning on the full dataset. Given an input x, a random subset of features is revealed via a binary mask $m \in \{0, 1\}^d$, and the model learns an encoder $z = f_\theta(x \odot m, m)$ and decoder $\hat{x} = g_\theta(z)$ to reconstruct the full input using an MSE objective. The encoder receives both the masked input and the mask itself, allowing it to distinguish between zero-valued features and unobserved ones. The masking ratio is sampled uniformly during training to expose the model to a wide range of partial observations. After pretraining, the encoder is frozen and used as the representation backbone for ECHO-k. For downstream evaluation, a linear probe is trained on z for classification under random masking to simulate partial observations; this ensures that the probe is calibrated to the type of partial inputs encountered during sequential acquisition. The same frozen encoder is shared across all baselines and ECHO-k.

G.2.2 TABULA Foundation Model

For single-cell RNA-seq, we use the pretrained TABULA foundation model (Ding et al., 2025), a self-supervised transformer designed for single-cell transcriptomic data. TABULA is trained on large-scale single-cell atlases and can embed partially observed gene expression profiles into cell-level representations. When a subset of genes is unobserved, TABULA treats these entries as missing and produces an embedding conditioned only on the available genes. This native support for partial observations makes it well-suited as a backbone for ECHO-k. The downstream task of cell-type annotation is evaluated via a linear probe trained on top of the cell embeddings. We use TABULA without any fine-tuning; the model weights are frozen throughout both policy learning and evaluation.

G.2.3 VirTues Foundation Model

For the NSCLC patient cohort, we deploy ECHO-k with the VirTues spatial proteomics foundation model (Wenckstern et al., 2025). VirTues is a graph-based transformer that operates on cell-level protein expression profiles within tissue sections, producing both cell-level and tissue-level representations. Tissue-level representations are obtained by averaging the cell-level summary tokens across all cells in the tissue sample. When a protein marker is unobserved, its expression value is treated as missing in the cell-level input, and VirTues produces an embedding conditioned on the available markers. Linear probes are trained for cancer subtyping and grading using the tissue-level embeddings. Importantly, the same ECHO-k policy is used for both downstream tasks: the acquisition order is determined once by the policy and then evaluated on both tasks, demonstrating task-agnostic panel design.

G.2.4 GeoSpatial Encoder

For MMEarth-Bench, a masked encoder is used. Concretely, each of the 12 Earth-observation modalities is first converted to a modality-level vector by global average pooling for spatial inputs and identity for tile-level inputs, then projected to a shared dimension and multiplied by a binary

modality mask indicating which modalities are observed. The masked per-modality projections are concatenated together with the mask itself and passed through an MLP encoder to produce a latent representation z. This encoder is trained under random modality masking, so it learns to make reconstructions from arbitrary partial modality subsets while explicitly distinguishing unobserved modalities from zero-valued features.

Backbone-agnosticism. Across all settings, both ECHO-k and the baselines have access only to the pretrained models during the feature selection phase. Linear probes are used exclusively for downstream evaluation after feature acquisition. ECHO-k interfaces with each backbone only through the partial embedding $\boldsymbol{z}(S_t)$ and the full-information target $\boldsymbol{z}^\star$; no architecture-specific modifications are needed. This design means that ECHO-k can be deployed with any new backbone simply by providing the embedding function, without rewriting any acquisition logic.

G.3 Evaluation Protocol

Each dataset is partitioned into training, validation, and test splits. For image and tabular tasks, we use a 70:15:15 split. For cell-type annotation, we adopt the predefined splits released with TABULA (Ding et al., 2025), and for cancer subtyping and grading, we follow the test splits defined in (Wenckstern et al., 2025).

Acquisition curves. The primary evaluation metric is the acquisition curve: for each method, we plot downstream performance (MSE or classification accuracy) as a function of the number of features acquired $t \in [\![1, d]\!]$. This allows us to compare methods across the full range of acquisition budgets and identify regimes where differences are most pronounced. In practice, the early regime (5–25% of features) is of greatest interest, since this is where cost savings are largest.

Linear probes. For all real-world datasets, downstream performance is evaluated by training a linear probe on the partial embeddings $\boldsymbol{z}(S_t)$ produced by the frozen backbone. A single linear probe is trained per downstream task using random masking augmentation and held fixed across acquisition budgets. This ensures that the evaluation is task-specific (it depends on the downstream labels) while the acquisition policy remains task-agnostic.

Reproducibility. All experiments are run with three random seeds for both the ECHO-k policy and the downstream evaluation. The `Random` baseline is additionally averaged over five independent runs per seed to reduce variance. Error bands in all figures correspond to standard error across seeds.

G.4 Baselines

We compare ECHO-k against a total of 15 baselines spanning synthetic (S), image (I), tabular (T), single-cell RNA-seq (C), and multiplex imaging (M) datasets. Below, we describe each baseline and indicate the settings in which it is evaluated.

G.4.1 Oracle and Random Baselines

`Oracle` **(S, I, T).** A one-step myopic oracle that greedily selects the next feature maximizing the immediate reduction in distance between the current embedding and the full (gold) embedding. This policy is anti-causal: it has access to unobserved features and corresponds to the oracle analyzed in our theoretical results (§3). It serves as an upper bound on greedy acquisition performance but is not deployable at test time.

`Random` **(S, I, T, C, M).** Selects the next view uniformly at random from the set of unobserved features. Provides a minimal reference point for all experiments.

G.4.2 Model-Driven Heuristic Baselines

`Prior` **(S).** Specific to the synthetic setting. Ranks features according to their unconditional expected contribution $\mathbb{E}[E_j]$, which under Assumption 3.2 admits the closed form:

$$\mathbb{E}[\boldsymbol{E}_j] = \|\mathbf{w}_j\|_2^2 \Gamma(1-2\sigma) e^{2\eta_j} \sum_{m=1}^{M} p_m e^{2\beta_j z_m} \propto \|\mathbf{w}_j\|_2^2 e^{2\eta_j} \sum_{m=1}^{M} p_m e^{2\beta_j z_m}.$$

Unlike ECHO-k, this baseline does not condition on observed features and therefore captures only population-level feature importance.

`GIR` **(I, T, S, M).** Global Impact Ranking is a practical extension of `Prior` that does not rely on a known generative model. For each feature i, a validation-set score $s_i := \|\boldsymbol{z}_{\text{full}} - \boldsymbol{z}_{-i}\|_2$ is computed, where $\boldsymbol{z}_{\text{full}}$ denotes the self-supervised embedding obtained using all features and $\boldsymbol{z}_{-i}$ denotes the embedding obtained when feature i is withheld. This induces a fixed, instance-independent acquisition order that captures population-level importance but not individual-level adaptivity.

`ESM-Max`, `ESM-Min` **(M).** For multiplex imaging only. Protein markers are embedded using ESM2 (Lin et al., 2023). For each marker, pairwise Euclidean distances are computed in the ESM2 embedding space. Markers are ranked by either minimum (`ESM-Min`) or maximum (`ESM-Max`) distance, yielding fixed acquisition orders based on semantic similarity or dissimilarity, following (Jain et al., 2025).

G.4.3 Unsupervised Feature Selection Baselines

The following methods were originally designed to identify informative feature *subsets* given full feature access (Solorio-Fernández et al., 2020). We adapt them to the sequential acquisition setting by using their induced feature rankings as acquisition orders. While not designed for instance-adaptive or sequential settings, they provide strong unsupervised baselines for ranking-based selection. All scores are computed on the validation set and induce a fixed acquisition order applied at test time.

`Variance` **(I, T, S).** Ranks features by their marginal variance over the validation set, selecting in decreasing order of variance.

SPEC-φ_3 (I, T). Ranks features based on alignment with the leading $k-1$ eigenvectors of the graph Laplacian, where k is the number of clusters (Zhao & Liu, 2007). Features with stronger alignment to these optimal soft cluster indicators receive higher scores.

`Laplacian Score` **(I, T).** The Laplacian Score (He et al., 2005) evaluates features by how well they preserve local geometric structure encoded by a nearest-neighbor graph. For each feature r, it computes $L_r = \frac{\widetilde{f}_r^\top L \widetilde{f}_r}{\widetilde{f}_r^\top D \widetilde{f}_r}$, where $L = D - S$ is the graph Laplacian, S is the similarity matrix, and $\widetilde{f}_r$ is the mean-centered weighted feature vector. Features with smaller scores are ranked higher.

`MCFS` **(I, T).** Multi-Cluster Feature Selection (Cai et al., 2010) selects features via spectral analysis of the graph Laplacian. It first solves the generalized eigenvalue problem $Ly = \lambda Dy$ to obtain K eigenvectors $Y = [y_1, \ldots, y_K]$. For each y_k, it solves an ℓ_1-regularized regression $\min_{a_k} \|y_k - X^T a_k\|^2 + \beta|a_k|$ to obtain sparse coefficients. Features are ranked by $\texttt{MCFS}(j) = \max_k |a_{k,j}|$ and selected in decreasing order.

`UDFS` **(I, T).** Unsupervised Discriminative Feature Selection (Yang et al., 2011) learns a sparse projection matrix $W \in \mathbb{R}^{d \times c}$ by minimizing local discriminative scores with $\ell_{2,1}$-norm regularization: $\min_{W^T W = I} \text{Tr}(W^T M W) + \gamma \|W\|_{2,1}$, where M encodes local neighborhood structure via k-nearest neighbors. Features are ranked by the row norms $\|w_i\|_2$ of W in descending order.

`SVD-Ent` **(I, T).** SVD-Entropy (Varshavsky et al., 2006) ranks features by their contribution to SVD-entropy. For each feature i, its contribution is $\text{CE}_i = E(A) - E(A_{-i})$, where $E = -\frac{1}{\log N} \sum_{j=1}^{N} V_j \log V_j$ and $V_j = s_j^2 / \sum_k s_k^2$ are normalized squared singular values. Features with larger CE_i are ranked higher.

G.4.4 Adaptive Label-Free Baselines

The baselines above induce fixed, instance-independent acquisition orders. To compare against policies that condition acquisition on the current instance, we additionally evaluate the following label-free adaptive baselines on MiniBooNE and hPancreas (§ H.3). As for ECHO-k, these methods observe only the currently acquired values and the acquisition mask at deployment, and never access the full-view embedding or downstream labels.

`GSMRL-AIR` **(T, C).** An instance-adaptive RL method based on acquisition with generative surrogate models (Li & Oliva, 2021), adapted to the label-free setting. The policy conditions its action selection on the per-sample conditional imputations of unobserved features produced by a generative surrogate, so that acquisition decisions depend on the values observed for each instance.

`ALG` **(T, C).** Adaptive Latent Gain is a learned regressor that predicts, for each unobserved view, the conditional one-step improvement in representation recovery directly from the current partial observation, and greedily acquires the view with the largest predicted gain. `ALG` shares the reward and state representation of ECHO-k, but optimizes only the immediate gain rather than a long-horizon objective, and therefore isolates the contribution of sequential planning.

`CIG` **(T).** Cluster Information Gain fits a Gaussian mixture model to the training data and scores each unobserved feature by the mutual information between the feature and the mixture component under the posterior given the currently observed features. The feature with the largest information gain is acquired next. Evaluated on MiniBooNE only.

G.4.5 Summary Table

Table 6: Summary description of implemented baselines. Methods marked with (†) rely on unsupervised feature selection and have oracle access to all features at ranking time. Methods marked with a ($\star$) are only deployed on the cancer grading and typing datasets, since they rely on a protein sequence foundation model (Lin et al., 2023). Methods marked with (‡) are label-free adaptive baselines that select instance-specific feature sets through continuous updates of their selection policies.

Name	Model Driven?	Selection Criterion	Adaptive?	Source
`Oracle`	✓	1-step Oracle knowledge	✓	—
`GIR`	✓	Average contribution to latent state	✗	—
`Random`	✗	Uniform distribution over unseen features	✗	—
`Laplacian Score` †	✗	Locality preservation	✗	He et al. (2005)
SPEC-φ_3 †	✗	Graph-based ranking	✗	Zhao & Liu (2007)
`UDFS` †	✗	Local discriminative structure	✗	Yang et al. (2011)
`MCFS` †	✗	Multi-cluster structure	✗	Cai et al. (2010)
`SVD-Ent` †	✗	SVD-entropy contribution	✗	Varshavsky et al. (2006)
`Variance` †	✗	Maximum variance	✗	—
`ESM-Max` ($\star$)	✓	ESM2-based	✗	Lin et al. (2023)
`ESM-Min` ($\star$)	✓	ESM2-based	✗	Lin et al. (2023)
`GSMRL-AIR` (‡)	✗	Generative surrogate imputations	✓	Li & Oliva (2021)
`ALG` (‡)	✓	Predicted one-step latent improvement	✓	—
`CIG` (‡)	✗	GMM mutual information	✓	—
ECHO-k	✓	Learned policy via expected latent improvement	✓	This paper

G.5 Implementation Details of ECHO-k

We cast sequential feature acquisition in ECHO-k as a finite-horizon decision problem over partial observations, closely aligned with the conditional surrogate objective analyzed in § 3. The goal of the agent is to sequentially select features so as to minimize the discrepancy between the latent representation induced by the acquired subset and the full-information latent representation of a frozen self-supervised model.

Table 7: Hyperparameters used for ECHO-k across datasets. The remaining hyperparameters follow the defaults in `Stable-Baselines3`.

Parameter	MNIST/Fashion-MNIST	MiniBooNE	hPancreas	NSCLC	MMEarth-Bench	Synthetic
Acquisition budget T	10	7	25	10	5	varies
Step threshold t_0	$T/2$	$T/2$	$T/2$	$T/2$	$T/2$	$T/2$
Gain factor λ	1.5	1.5	1.5	1.5	1.5	1.5
Discount factor γ	0.98	0.98	0.99	0.98	0.98	0.98
Training steps	300k	300k	500k	300k	200k	300k

G.5.1 MDP Formulation

Let d denote the total number of features and T the acquisition budget. An episode corresponds to sequentially selecting T features without replacement. At step t, the agent has observed a subset $S_t \subseteq \{1, \dots, d\}$. We define a finite-horizon Markov decision process $(\mathcal{S}, \mathcal{A}, P, r)$ as follows.

State. The observable state summarizes the acquisition history and the induced latent representation:

$$s_t := (S_t, \boldsymbol{z}(S_t), t/T, \|\boldsymbol{z}(S_t)\|_2, \boldsymbol{x} \odot \boldsymbol{m}_t)$$

where $\boldsymbol{m}_t \in \{0,1\}^d$ is the binary acquisition mask corresponding to S_t, and $\boldsymbol{z}(S_t)$ denotes the latent embedding produced by the frozen self-supervised encoder given the partially observed input. This state representation renders the process Markov with respect to the surrogate objective.

Action. At each step, the agent selects a feature index $a_t \in \mathcal{A}_t := \{1, \dots, d\} \setminus S_t$, corresponding to acquiring a previously unobserved feature. We apply invalid action masking (Huang & Ontañón, 2020), assigning zero probability to actions in S_t during policy evaluation and optimization.

Transition. Transitions are deterministic: selecting $a_t = j$ yields $S_{t+1} = S_t \cup \{j\}$ and updates the latent embedding to $\boldsymbol{z}(S_{t+1})$.

G.5.2 Reward Design

Let $\boldsymbol{z}^*$ denote the latent embedding obtained from the full feature set. Consistent with the surrogate loss analyzed in § 3, we define the instantaneous reward as the normalized reduction in representation error:

$$r_t = (1 + \mathbb{1}\{t < t_0\}\lambda) \left(\frac{\|\boldsymbol{z}^* - \boldsymbol{z}(S)\|_2 - \|\boldsymbol{z}^* - \boldsymbol{z}(S \cup \{j\})\|_2}{\|\boldsymbol{z}^* - \boldsymbol{z}(S_0)\|_2} \right),$$

where $\lambda > 0$ is the gain factor and S_0 denotes the singleton initial feature set. This reward is a shaped version of the marginal gain $\Delta(a_t \mid S_t) = \|\boldsymbol{z}^* - \boldsymbol{z}(S_t)\|_2 - \|\boldsymbol{z}^* - \boldsymbol{z}(S_t \cup \{a_t\})\|_2$, which serves as a proxy for the conditional expected contribution $\mathbb{E}[E_{a_t} \mid \mathcal{F}_t]$ appearing in our theoretical analysis. The step-dependent amplification factor $(1 + \lambda \mathbb{1}\{t < t_0\})$ encourages refinement in early acquisition stages, where marginal reductions are typically largest.

G.5.3 Policy Learning

We learn a stochastic policy $\pi_\theta(a_t \mid s_t)$ using PPO (Schulman et al., 2017) with invalid action masking (Huang & Ontañón, 2020). During both training and inference, the policy has access only to the frozen self-supervised encoder and the induced latent representations. Downstream task labels and linear probes are used exclusively for post-hoc evaluation and do not influence the policy or reward.

G.5.4 Hyperparameters

We implement ECHO-k using the `Stable-Baselines3 (SB3)` (Raffin et al., 2021) framework with `sb3-contrib` for support of masked action spaces. Unless mentioned, all PPO parameters are default of SB3. All experiments were conducted on NVIDIA A100 GPUs (80GB) with 128 AMD EPYC 7543 32-core processors. Table 7 summarizes the key hyperparameters across datasets.

The step threshold t_0 is uniformly set to half of the episode length T. Across experiments, the acquisition budget T is chosen to lie between 5% and 25% of the total number of available features,

depending on the dataset. We set $\gamma = 0.98$ for small-horizon acquisition and $\gamma = 0.99$ for long-horizon (hPancreas). For datasets with fewer features, we train for 300,000 steps; for larger feature spaces, we train for 500,000 steps (except for MMEarth-Bench, where we train for 200,000 steps).

H Additional Experiments

H.1 Controlled Synthetic Experiment

We first isolate the core question in a controlled setting with known data-generating process: *how much does conditioning on partial observations help?* For the synthetic data, we consider a randomized isotropic linear regression task with $\boldsymbol{W} \sim \mathcal{N}(0, \frac{1}{d}\boldsymbol{I}_d)$ and $\varepsilon' \sim \text{Gumbel}(\mu', \sigma')$ yielding labels $y = \boldsymbol{W}^\top \boldsymbol{x} + \varepsilon'$. We fit $\widehat{\boldsymbol{W}}$ via ridge regression on the full feature set (see Assumption 3.2), with labels from isotropic linear regression (details in § G.1.6). Fig. 6 confirms that the optimal test-time policy consistently outperforms both non-oracle baselines and closely tracks the oracle in early stages, validating the theoretical separation of §3 and motivating the long-horizon RL strategy of ECHO-k.

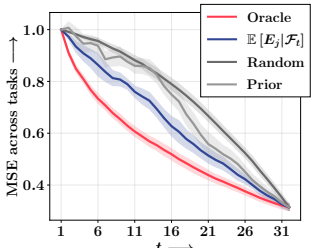


Figure 6: Downstream performance of random, prior-based ($\mathbb{E}[\boldsymbol{E}_j]$), optimal-test-time, and oracle policies on a random isotropic regression task

H.2 Extended Baseline Comparisons

In vision and tabular settings, features are individually cheap but collectively redundant; a policy that identifies the most informative subset early reduces measurement overhead at test time. For MNIST, Fashion-MNIST, and MiniBooNE, we train a masked autoencoder and freeze the encoder as ECHO-k's backbone. For hPancreas, we use TABULA (Ding et al., 2025) scRNA foundation model. ECHO-k interfaces with it only through $\boldsymbol{z}(\cdot)$ and $\boldsymbol{z}^\star$, requiring no architecture-specific changes (see § G.2 for backbone details).

Results on MNIST and Fashion-MNIST. Both datasets are downsampled to 14×14 following Li & Oliva (2021), yielding $d = 196$ pixel-features with 10-class classification as the downstream task. The goal is to classify each image (handwritten digit or clothing item) by sequentially revealing as few pixels as possible. All methods start from the same top-corner pixel (Fig. 8). ECHO-k rapidly concentrates its acquisitions on discriminative image regions, achieving gains of up to 12% over all non-oracle baselines in the early regime (10–20% of pixels). No single baseline is consistently second-best: `Variance` leads on MNIST while `Random` is surprisingly competitive on Fashion-MNIST.

Results on MiniBooNE. MiniBooNE contains $d = 33$ reconstructed physics variables from a neutrino experiment, with binary signal-vs-background classification as the downstream task; these variables are costly to compute and highly correlated (see § G.1). ECHO-k outperforms all heuristic baselines and the one-step oracle (Fig. 7). We attribute this to the oracle's myopic strategy, which over-selects correlated variables early and suffers diminishing returns, whereas ECHO-k optimizes over a long horizon and learns to exploit complementary information.

H.3 Comparison with Adaptive Acquisition Policies

Most baselines in the main text are fixed acquisition orderings. To verify that the gains of ECHO-k are not solely due to comparison against fixed rankings, we compare against label-free controls that condition acquisition on the current instance and match the information setting of ECHO-k:

- **GSMRL-AIR**, an instance-adaptive RL method that conditions action selection on the per-sample conditional imputations of a generative surrogate (Li & Oliva, 2021), adapted to the label-free setting.
- **Adaptive Latent Gain (ALG)**, a learned regressor that predicts the conditional one-step representation improvement directly from the current partial observation. ALG shares the reward and state representation of ECHO-k, but not its long-horizon objective.
- **Cluster Information Gain (CIG)**, which scores candidate views by mutual information under a fitted Gaussian mixture model posterior (MiniBooNE only).

All methods observe only the currently acquired values and the acquisition mask, and never access the full-view embedding or downstream labels at deployment.

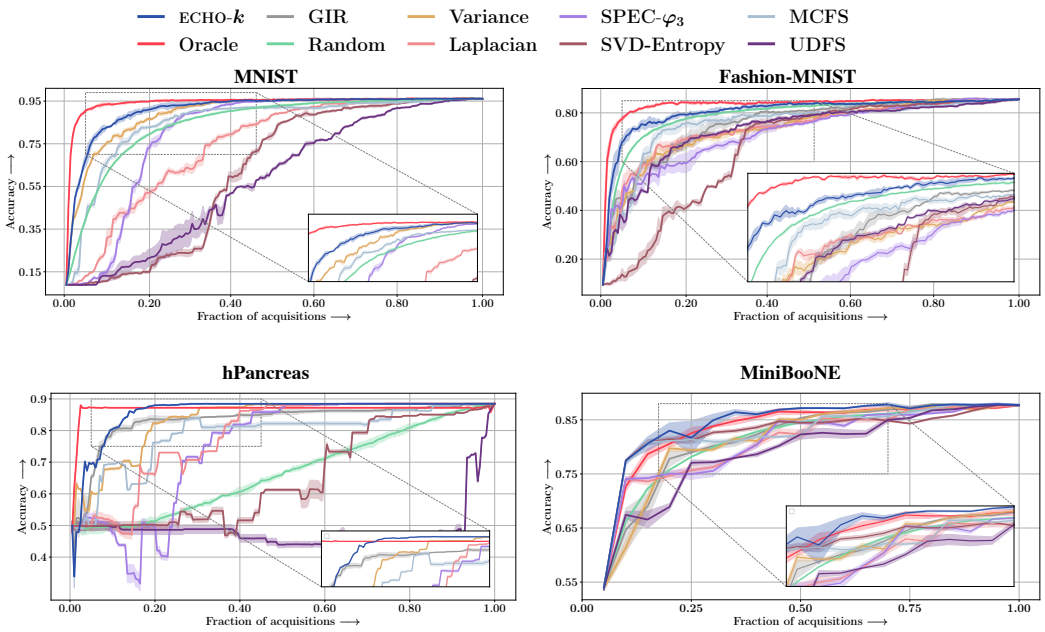


Figure 7: Results of ECHO-k across all baselines for the MNIST (top-left), Fashion-MNIST (top-right), hPancreas (bottom-left), and MiniBooNE (bottom-right) datasets.

Table 8: **Comparison with adaptive acquisition policies.** Accuracy (%) at different acquisition budgets on MiniBooNE and hPancreas. All policies operate without downstream labels. denotes best and second best performance.

Method	MiniBooNE				Method	hPancreas			
	20%	**40%**	**60%**	**80%**		**20%**	**40%**	**60%**	**80%**
CIG	79.80	84.87	86.17	86.97					
ALG	77.50	83.60	86.12	87.26	ALG	73.38	80.16	83.12	86.33
GSMRL-AIR	79.18	84.78	86.40	86.26	GSMRL-AIR	77.23	81.87	84.41	86.40
ECHO-k	81.70	86.87	87.60	87.83	ECHO-k	80.81	88.04	88.43	88.45

Table 8 shows that ECHO-k outperforms all adaptive controls at every budget on both datasets, including GSMRL-AIR, a fully adaptive RL baseline operating under the same label-free constraints. The comparison with ALG is particularly informative: both methods use representation improvement to guide acquisition, but ECHO-k optimizes gains over a longer horizon. This indicates that the advantage of ECHO-k comes from long-horizon planning rather than from the recovery signal alone.

H.4 Comparison with Supervised Acquisition

Table 9 reports the full comparison of ECHO-k with five supervised acquisition methods on MiniBooNE: DIME (Gadgil et al., 2023), GDFS (Covert et al., 2023), SEFA (Norcliffe et al., 2025), GSMRL (Li & Oliva, 2021), and EDDI (Ma et al., 2019). Unlike ECHO-k, these methods use downstream labels during policy learning.

At 10% acquisition, ECHO-k is within 0.2 points of the best supervised method and exceeds four of the five, despite never observing downstream labels. The gap is largest at 20%, where GDFS leads by 3.0 points, and narrows to 0.9 and 0.5 points at 40% and 60%, respectively. Given that ECHO-k never observes downstream labels during training, we view this as the expected cost of task agnosticism.

H.5 Cost of Task Agnosticism

When a single downstream task is fixed and sufficient labels are available, a task-specific policy may be preferable. To quantify this trade-off, we train supervised acquisition policies for NSCLC

Table 9: **Comparison with supervised acquisition on MiniBooNE.** Accuracy (%) at different acquisition budgets. All methods except ECHO-k use downstream labels. denotes best and second best performance.

Method	10%	20%	40%	60%
DIME	79.97	84.23	86.87	87.07
GDFS	80.23	84.73	87.50	87.80
SEFA	78.67	84.70	87.73	88.13
GSMRL	75.73	83.27	87.37	87.77
EDDI	72.30	76.70	83.50	85.07
ECHO-k	80.03	81.70	86.87	87.60

subtyping (S) and grading (G) using the same architecture and training budget as ECHO-k, replacing the latent reward with the downstream-loss improvement

$$r_t^{\text{sup}} = \ell\big(g(\boldsymbol{z}(\boldsymbol{P}_{t-1})), y\big) - \ell\big(g(\boldsymbol{z}(\boldsymbol{P}_t)), y\big).$$

PPO-Subtype and PPO-Grade optimize their respective endpoints, while PPO-Multitask averages the two task-specific losses at each step. We evaluate each policy on both tasks.

Table 10: **Task-specific and task-agnostic acquisition on NSCLC.** Accuracy (%) for subtyping (S), grading (G), and their mean at different acquisition budgets. Supervised PPO policies use task labels during policy learning, while ECHO-k uses only the representation-recovery reward.

Policy	20%			40%			60%		
	S	G	Mean	S	G	Mean	S	G	Mean
PPO-Subtype	83.2	62.8	73.0	85.4	64.2	74.8	86.3	65.0	75.7
PPO-Grade	78.5	67.8	73.1	82.2	68.2	75.2	83.1	68.0	75.5
PPO-Multitask	82.9	66.5	74.7	84.1	67.2	75.6	84.8	67.9	76.3
ECHO-k	82.8	65.8	74.3	83.8	67.1	75.5	85.2	67.4	76.3

At 20% acquisition, ECHO-k trails the corresponding task-specific policies by 0.4 points on subtyping and 2.0 points on grading (Table 10). The specialized policies also transfer poorly across endpoints, losing approximately 3–5 points when evaluated on the other task. More importantly, ECHO-k is within 0.4, 0.1, and 0.0 points of PPO-Multitask in mean accuracy at 20%, 40%, and 60%, respectively, despite using neither task label. This quantifies the expected cost of task agnosticism, while showing that preserving the pretrained representation nearly matches direct multitask supervision.

H.6 Robustness to Partially Observed Data

Table 11 provides additional robustness results under partially observed reference data on MNIST-14 and FMNIST-14. We train ECHO-k after randomly masking an m-fraction of views when computing the reference representation used for the self-supervised reward. Across both datasets, performance degrades smoothly as m increases, rather than collapsing. On MNIST-14, partially observed-reference policies remain close to or even slightly above the complete-reference policy across several budgets, indicating that the acquisition policy can tolerate substantial noise in the reference target. On FMNIST-14, the degradation is more visible but remains gradual: even with 50% missing reference views, ECHO-k stays competitive with strong baselines and preserves most of its early-budget advantage. These results complement the biological experiments in Table 1, showing that the method's robustness to incomplete reference data is not specific to a single modality or dataset.

H.7 Heterogeneous Acquisition Costs

To test whether ECHO-k can adapt when views have nonuniform acquisition costs, we run a controlled cost-aware experiment on MNIST-14. We assign each pixel a sample-dependent acquisition cost based on its intensity, rescaled to the range $[0.25, 2.0]$, so that the policy must trade off representation recovery against cumulative cost. We compare a cost-aware variant of ECHO-k, trained with a cost-penalized reward and cost information in the state, against the standard cost-unaware policy.

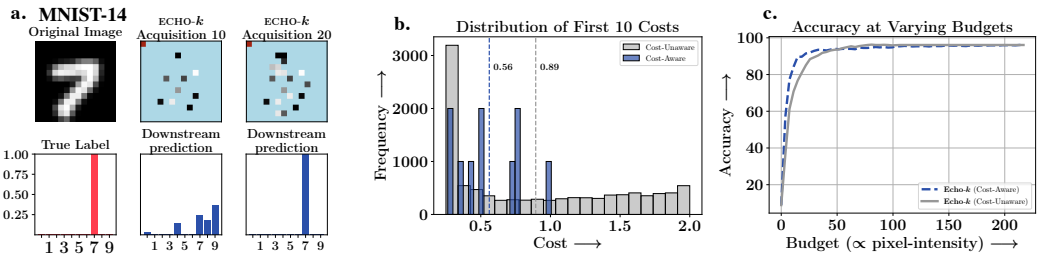


Figure 8: **a.** Overview of the task and acquisition behavior of ECHO-k: we illustrate how ECHO-k sequentially acquires pixels in an unsupervised manner, progressively concentrating the (unobserved) downstream class probability mass. **b.** Distribution of acquisition costs among the first ten selected pixels; the cost-aware policy selects lower-cost pixels early. **c**. Accuracy as a function of cumulative acquisition cost; cost-aware training improves performance in the low-budget regime, while both policies converge at larger budgets.

Table 11: We train ECHO-k after randomly masking an m-fraction of views when computing the reference representation. Accuracy is reported at different acquisition budgets on MNIST-14 and FMNIST-14. denotes best and second best performance, apart from the Oracle.

Method	MNIST-14				FMNIST-14			
	20%	40%	60%	80%	20%	40%	60%	80%
Oracle	0.937±0.005	0.951±0.002	0.953±0.002	0.956±0.002	0.824±0.003	0.843±0.006	0.840±0.003	0.842±0.004
Variance	0.759±0.007	0.865±0.005	0.928±0.005	0.951±0.006	0.565±0.022	0.670±0.008	0.725±0.016	0.741±0.005
SPEC-φ_3	0.168±0.014	0.754±0.018	0.894±0.003	0.942±0.002	0.543±0.036	0.600±0.039	0.681±0.012	0.738±0.013
MCFS	0.678±0.010	0.829±0.005	0.904±0.004	0.918±0.002	0.631±0.029	0.760±0.016	0.767±0.013	0.793±0.010
Random	0.627±0.009	0.800±0.005	0.871±0.003	0.907±0.003	0.708±0.004	0.779±0.004	0.805±0.002	0.821±0.005
ECHO-k	0.830±0.004	0.913±0.012	0.935±0.009	0.951±0.003	0.749±0.019	0.793±0.007	0.818±0.007	0.833±0.009
ECHO-k (m =10%)	0.845±0.015	0.921±0.002	0.941±0.002	0.949±0.003	0.742±0.024	0.780±0.008	0.814±0.008	0.821±0.011
ECHO-k (m =20%)	0.862±0.002	0.913±0.006	0.939±0.007	0.948±0.003	0.735±0.020	0.767±0.013	0.803±0.006	0.813±0.008
ECHO-k (m =30%)	0.836±0.008	0.912±0.003	0.940±0.008	0.950±0.005	0.721±0.018	0.769±0.014	0.798±0.009	0.814±0.008
ECHO-k (m =40%)	0.830±0.017	0.903±0.014	0.934±0.001	0.950±0.011	0.716±0.024	0.776±0.003	0.787±0.003	0.801±0.009
ECHO-k (m =50%)	0.823±0.019	0.900±0.011	0.938±0.007	0.949±0.002	0.711±0.013	0.765±0.018	0.792±0.006	0.815±0.002

Fig. 8 shows that cost-aware training changes the acquisition behavior. Over the first ten acquisitions, the cost-aware policy selects substantially cheaper pixels on average than the cost-unaware policy, reducing the mean acquisition cost from 0.89 to 0.56. This shift also improves performance in the low-budget regime: when accuracy is plotted against cumulative cost rather than number of pixels, the cost-aware policy reaches high accuracy with fewer cost units. As the budget grows, both policies converge, indicating that cost-awareness is most useful when acquisition resources are limited. This controlled experiment confirms that ECHO-k can incorporate heterogeneous costs by changing the reward and state, without modifying the representation-recovery objective.

H.8 Adaptive Stopping

A fixed acquisition horizon may be restrictive in deployment, where the policy should also decide whether the currently acquired views are already sufficient. We extend ECHO-k with a STOP action, converting the fixed-horizon MDP into an optimal-stopping problem. For a measurement action a, the reward is

$$r_t = \frac{e(\boldsymbol{P}_{t-1}) - e(\boldsymbol{P}_t)}{e(\boldsymbol{P}_0) + \epsilon} - \lambda c_a, \qquad e(\boldsymbol{P}) = \|\boldsymbol{z}(\boldsymbol{P}) - \boldsymbol{z}^\star\|_2,$$

while STOP terminates the episode with zero additional reward. Since STOP yields zero future reward, the policy continues only when the expected cumulative representation gain outweighs the acquisition cost, and with $\gamma = 1$, the learned stopping time τ directly trades recovery against total acquisition cost. As in ECHO-k, $\boldsymbol{z}^\star$ is used only to construct the label-free training reward; at deployment, both acquisition and stopping depend only on the current partial state.

On MiniBooNE, ECHO-k with stopping reaches 86.06% accuracy at a mean of 9.86 acquired features, showing instance-dependent stopping (Table 12).

Table 12: **Adaptive stopping on MiniBooNE.** Accuracy (%) and number of acquired features.

Policy	# Features	Accuracy
ECHO-k (fixed horizon)	10	86.11
ECHO-k (STOP)	9.86 (mean)	86.06

This matches the fixed-horizon policy at a comparable budget (86.11% at $t = 10$), while requiring no budget to be specified in advance. Together with the heterogeneous-cost experiment above, this shows that the representation-recovery objective of ECHO-k can be combined with cost-aware rewards to decide both which views to acquire and when to stop.

H.9 Empirical Validation of the Near Isometry Assumption

A key ingredient in our theoretical analysis is Assumption 3.4 (Near Isometry), which requires that the normalized columns of the embedding weight matrix are approximately orthogonal. We validate this assumption empirically on two classes of models.

H.9.1 DinoV2 Foundation Models

We examine the DINOv2 family of vision transformers (Oquab et al., 2023), including ViT-S/14, ViT-B/14, ViT-L/14, and ViT-G/14 (with register tokens). For each model, we: (1) extract the final MLP projection weights $\boldsymbol{W} \in \mathbb{R}^{p \times d}$ from the last transformer block's feed-forward network (fc2 layer); (2) normalize columns to get $\boldsymbol{B}$ and compute the Gram matrix $\boldsymbol{G} = \boldsymbol{B}^\top \boldsymbol{B}$; (3) analyze the distribution of the off-diagonal elements of $\boldsymbol{G} - \boldsymbol{I}_d$. As shown in Fig. 9, these values concentrate tightly near zero across all model scales, confirming approximate orthogonality. Larger models exhibit tighter concentration, consistent with theoretical expectations for high-dimensional random-like weight matrices.

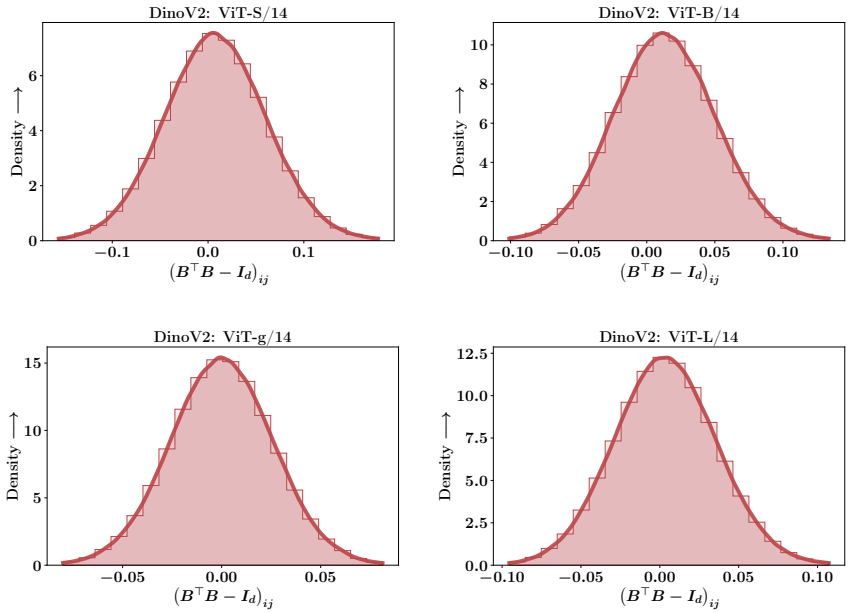


Figure 9: Distribution of $\boldsymbol{B}^\top \boldsymbol{B} - \boldsymbol{I}_d$ for 4 different DinoV2 models of increasing capacity.

H.9.2 Trained Autoencoders

We further validate the assumption on trained autoencoders across varying dimensions d from 32 to 4096. For each dimension, we generate low-rank data $\boldsymbol{X}^\top = \boldsymbol{U}\boldsymbol{V}^\top + \boldsymbol{\epsilon}$ where $\boldsymbol{U} \in \mathbb{R}^{n \times r}$, $\boldsymbol{V} \in \mathbb{R}^{d \times r}$. We extract the last layer weight matrix $\boldsymbol{W}$ and compute the effective rank p_{eff} using the entropy of normalized singular values (Roy & Vetterli, 2007):

$$p_{\text{eff}} = \exp\left(-\sum_i \frac{s_i}{\sum_k s_k} \log \frac{s_i}{\sum_k s_k}\right).$$

We sample random subsets $J \subseteq \{1, \ldots, d\}$ of varying sizes m, normalize the corresponding weight vectors to form $\boldsymbol{B}_J$, and compute $\|\boldsymbol{B}_J^\top \boldsymbol{B}_J - \boldsymbol{I}_m\|_{\text{op}}$. We compare against the theoreti-

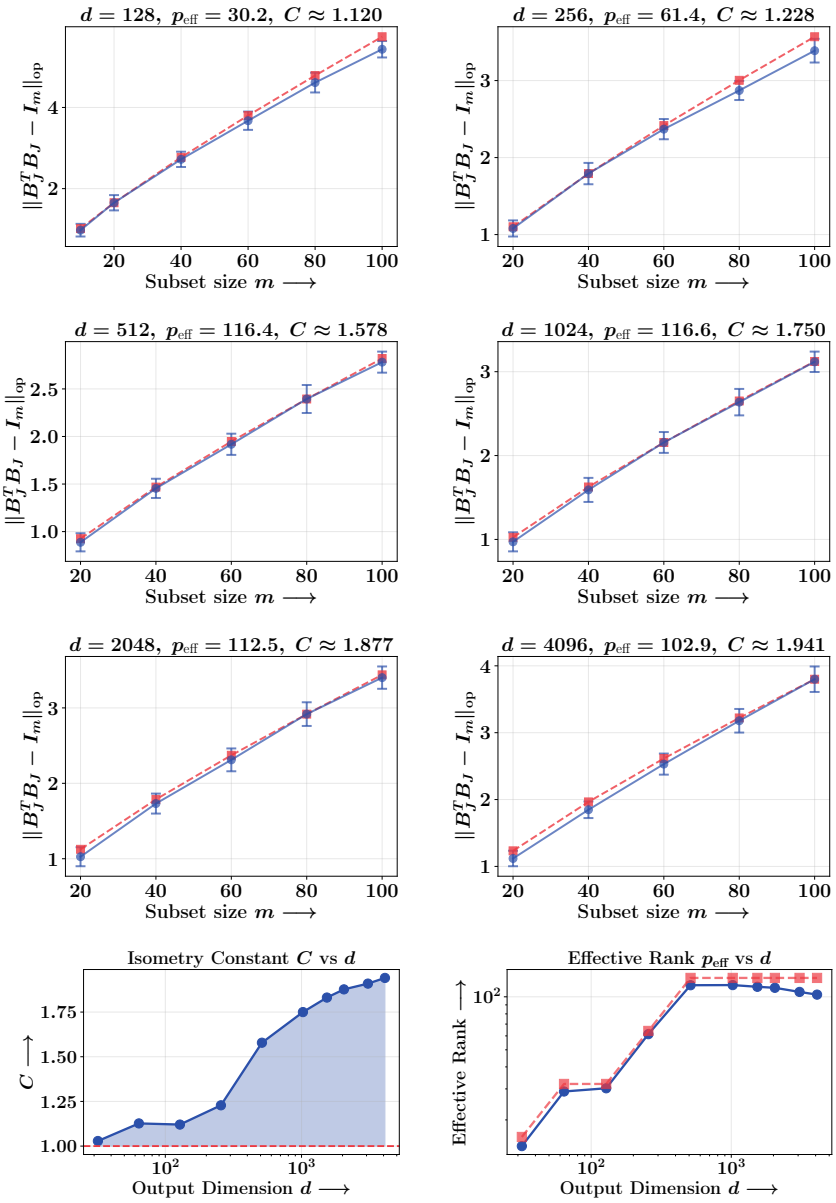


Figure 10: Variation of $\|\boldsymbol{B}^\top \boldsymbol{B} - \boldsymbol{I}_d\|_{\text{op}}$ for varying values of d on trained autoencoder models and variation of the empirical estimation of C with d.

cal bound $C\left(\sqrt{m/p_{\text{eff}}} + m/p_{\text{eff}}\right)$. Across 50 repetitions, the near isometry property holds with $C \in [0.93, 1.73]$ across all tested dimensions. As d increases, the standard deviation of off-diagonal elements decreases, validating that trained neural network weights naturally form near-orthonormal systems (Fig. 10).

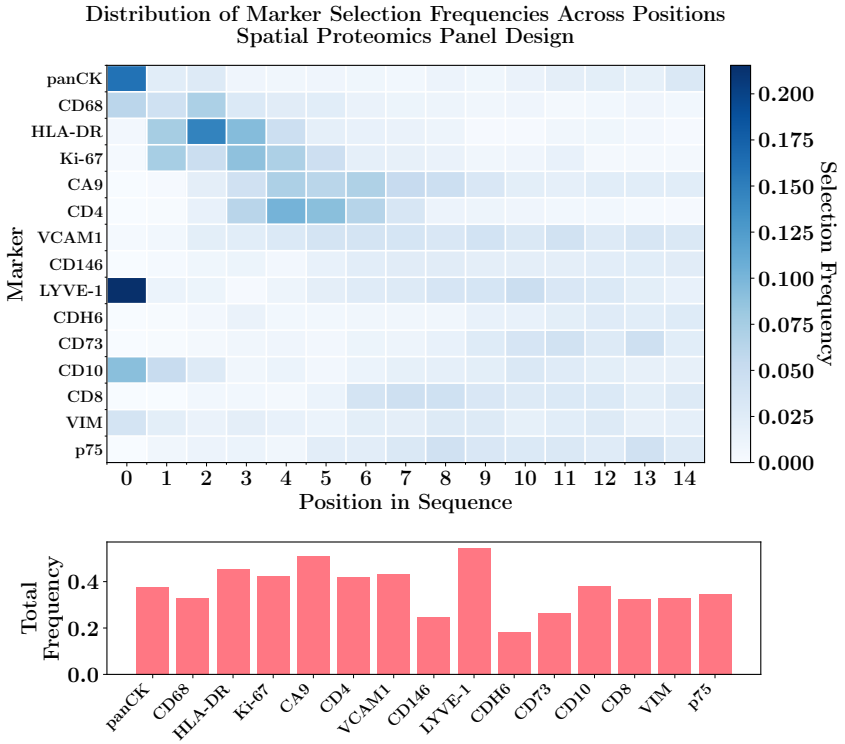


Figure 11: Analysis of the top-15 markers and their frequency, as chosen by ECHO-k jointly for the cancer subtyping and grading task on the NSCLC spatial proteomics dataset.

H.10 Analysis of Markers Acquired by ECHO-k on Spatial Proteomics

In Fig. 11, we analyze the distribution of markers chosen by ECHO-k on the NSCLC dataset tasks: cancer subtyping and grading from the marker panel provided by (Cords et al., 2024). In an unsupervised fashion, ECHO-k prioritizes markers that are well-established in the clinical literature for their roles in tumor characterization and microenvironmental context:

- `panCK`: a pan-cytokeratin marker for epithelial (tumor) cell identity, consistently selected first.
- `Ki-67`: a proliferation marker widely used in clinical grading of lung cancers (Martin et al., 2004).
- `CD4`: a marker of helper T-cell infiltration, reflecting the adaptive immune response in the tumor microenvironment.
- `LYVE-1`: a marker of lymphatic endothelial cells, relevant to lymphatic involvement and metastatic potential.

This early selection pattern suggests that ECHO-k effectively identifies markers with strong relevance to both tumor biology and immune context, thereby supporting both subtyping and grading tasks without any supervision. The alignment with clinical domain knowledge provides additional validation that the self-supervised acquisition objective captures diagnostically meaningful information.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The paper introduces ECHO-k, provides theoretical justifications, and experiments, as reflected in the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: The paper contains an impact and limitations statement in the appendix which discusses the limitations of the work.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper provides a central theorem bridging the oracle-test-time-policy gap and assumptions in the main, along with detailed proof in the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides all experimental details, evaluation protocols, hyperparameters, hardware architecture, and Python libraries.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The data /models are publicly available, and we attach the code to train and evaluate as a zip file in the supplement.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.

- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All experimental details are provided in the main and appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All experiments have been run multiple times for statistical significance, and the error bars have been provided for all experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All hardware details for the experiments have been provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Yes, the research conducted conforms with the Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper contains an impact and limitations statement which discusses the impact of the work.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The work uses standard datasets published in other works, and does not use generative models.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All original creators of code, data, and models are cited thoroughly in the paper.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: No new assets are introduced in the paper.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not contain crowdsourcing experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not contain research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were only used for editing of grammar in the paper.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.