
000 PRAGMATIC CURIOSITY: UNIFYING BAYESIAN OPTI-
001 MIZATION AND EXPERIMENTAL DESIGN VIA ACTIVE
002 INFERENCE
003
004
005

006 **Anonymous authors**

007 Paper under double-blind review
008
009

010
011 ABSTRACT

012
013 Bayesian Optimization (BO) and Bayesian Experimental Design (BED) have tradi-
014 tionally offered separate solutions for goal-oriented and information-oriented
015 tasks, respectively, leaving a gap in complex problems where learning and opti-
016 mization are not separate phases but deeply intertwined objectives. In this paper,
017 we provide the first unified framework of BO and BED, which is rooted in the
018 principles of active inference (AIF). We introduce “pragmatic curiosity”, a new
019 paradigm where the classic explore-exploit dilemma is resolved by minimizing a
020 single objective: the Expected Free Energy (EFE), which naturally balances prag-
021 matic (goal-seeking) and epistemic (information-seeking) drives. We demonstrate
022 the power of this approach on a suite of challenging hybrid tasks, including con-
023 strained system identification, targeted active search, and composite optimization
024 with unknown preferences. Empirical results prove the cross-domain adaptability
025 and effectiveness of our proposed framework: our “pragmatic curiosity” paradigm
026 consistently outperforms standard baselines in BO and BED, demonstrating quan-
027 tifiable improvements in key metrics like estimation accuracy, critical region cov-
028 erage, and final solution quality.

029 1 INTRODUCTION
030

031 Engineering and scientific applications often involve expensive optimization tasks aimed at iden-
032 tifying optimal designs or desired system states. Bayesian optimization (BO) seeks to accelerate
033 this search toward a specified *goal* (Shahriari et al., 2016; Frazier, 2018), while Bayesian experi-
034 mental design (BED) aims to maximize the *information* gained about unknown system parameters
035 (Rainforth et al., 2023). Both methodologies leverage probabilistic models and acquisition crite-
036 ria that quantify the utility of evaluating unknown configurations, tailored to either *optimization* or
037 *learning* objectives. **Despite their individual successes and the explosion of research in each field,**
038 **their disconnection creates a vacuum for a broad class of hybrid problems that routinely require**
039 **simultaneously both *seeking knowledge* and *achieving goals*.**

040 For many real-world applications, such as goal-directed planning (Lookman et al., 2019), envi-
041 ronmental monitoring (Konakovic Lukovic et al., 2020), and targeted material design (Matsumoto
042 et al., 2025), *learning* and *optimization* are not separate phases but deeply intertwined objectives.
043 This challenge fundamentally arises across tasks with increasing complexity in terms of both *epis-*
044 *temic consideration* (*i.e.*, from parametric to non-parametric models) and *pragmatic evaluation* (*i.e.*,
045 from known to unknown goals):

046 **(1) Constrained System Identification**, where the *epistemic* desire of precisely learning a system’s
047 parameters is governed by the *pragmatic* need to keep experiments within safe or valid operational
048 bounds (*e.g.*, avoiding sensor saturation, or dangerous chemical reaction). This type of task can
049 be found in numerous applications, including environmental monitoring (Konakovic Lukovic et al.,
050 2020) and catalyst design (Zhong et al., 2020).

051 **(2) Targeted Active Search**, where the *pragmatic* objective to discover regions that meet certain
052 criteria (*e.g.*, system failure modes, or specific performance ranges) requires *epistemic* curiosity
053 to explore the region’s shape, size, and boundaries. Example applications can be found in failure
discovery (Ramanagopal et al., 2018) and medical monitoring (Malkomes et al., 2021).

054 **(3) Composite Bayesian Optimization**, where the *pragmatic* goal is to find an optimal design based
055 on a user’s hidden preferences—a task that is impossible without first being *epistemically* curious
056 about the user’s objectives themselves. Such scenarios commonly arise in simulation-based design
057 (González & Zavala, 2025; Coelho et al., 2025) and A/B testing (Bakshy et al., 2018)

058 Classically, to address these hybrid problems, practitioners have been forced to choose between
059 specialized tools, and accommodate problem-specific adaptation by leveraging information-gain
060 criteria to enhance optimization and vice versa. On the BO side, Russo & Van Roy (2018) pro-
061 posed information-directed sampling (IDS) to online optimization problems. Hvarfner et al. (2023)
062 introduced a statistical distance-based active learning (SAL) criterion into the BO loop to actively
063 learn the Gaussian process hyperparameter even as it searches for the optimum. On the BED (also
064 known as Bayesian active learning, BAL) side, Smith et al. (2023) proposed an expected predictive
065 information gain (EPIG) criterion that focuses on information gain in model predictions, mitigating
066 classical BAL’s tendency to select out-of-distribution or low-relevance queries by accounting for an
067 input data distribution. These efforts begin to show the growing synergy between BO and BED tech-
068 niques, but they remain problem-specific and rarely generalize across categories. More importantly,
069 they lack a principled account of the underlying connections and potential synergy between BO and
070 BED criteria, nor is there a coherent, generalizable framework that unifies them.

071 In this paper, we present a unified perspective that arises naturally from the principles of Active
072 Inference (AIF) (Friston, 2010; Friston et al., 2017). Originally developed in computational neuro-
073 science, AIF has been successfully applied in fields such as robotics (Lanillos et al., 2021), cognitive
074 science (Smith et al., 2022), and ecosystem modeling (Friston et al., 2024). It prescribes behavior
075 that minimizes expected free energy (EFE)—a quantity that naturally balances an *epistemic* drive
076 for information gain with an *pragmatic* drive to achieve preferred outcomes. We prove that many
077 acquisition strategies in BO and BED can all be interpreted as special cases of EFE minimization.

078 Building on this unified framework, we propose a new paradigm, conceptualized as “pragmatic curi-
079 osity”, to offer a principled resolution to the classic explore-exploit dilemma. Under this paradigm,
080 *seeking knowledge* and *achieving goals* are not treated as competing objectives to be balanced, but
081 as two inseparable facets of a single imperative to minimize EFE. As a result, the balance between
082 exploration and exploitation is not manually tuned but emerges from the agent’s beliefs, dynamically
083 adapting based on its current uncertainty and the value of potential rewards.

084 Our “pragmatic curiosity” framework excels at handling the hybrid objectives of *learning* and *op-*
085 *timization* by operating from a single principle. This overcomes a core limitation of pure BO and
086 BED approaches, whose acquisition functions are myopically focused on either exploiting rewards
087 or reducing uncertainty, rendering them ineffective when a task demands a principled combina-
088 tion of both. To validate this, we conduct experiments structured around the three problem classes
089 mentioned above, drawing on applications that include environmental monitoring in plume fields
090 (Konakovic Lukovic et al., 2020), failure detection in autonomous driving scenarios (Ramanagopal
091 et al., 2018), and distributed energy resource allocation in power grids (Kianmehr et al., 2019).

092 The empirical results reveal a consistent pattern of superior performance, demonstrating our frame-
093 work’s advantages in its ability to solve complex, hybrid objectives. In constrained system identi-
094 fication tasks, our algorithm achieved near-perfect estimation accuracy while requiring up to 40%
095 fewer queries than other methods. For targeted active search tasks, it demonstrated a more effective
096 exploration strategy, discovering a crucial 10% more of the critical failure region within the same
097 budget. Most notably, in tasks with unknown user preferences, our approach always successfully
098 learned the underlying objective where other baselines fail to capture. Together, these findings val-
099 idate the power of our unified approach, showing that a principled balance between pragmatic and
100 epistemic drives leads to tangible gains across diverse and challenging problem settings.

101 In summary, our main contributions are the following:

- 102 • A formal unification of BO and BED within the AIF framework, overcoming the myopic exploita-
103 tion of pure BO and the random exploration of pure BED.
- 104 • A “pragmatic curiosity” paradigm, which leads to acquisition functions that dynamically arbitrate
105 between goal-seeking and information-seeking based on the single objective of minimizing EFE.
- 106 • Comprehensive empirical validation across three categories that differ substantially from one an-
107 other in terms of both *epistemic consideration* and *pragmatic evaluation*.

2 PRELIMINARIES

2.1 BAYESIAN OPTIMIZATION

Given an unknown objective function $f : \mathcal{X} \mapsto \mathfrak{R}$, BO seeks to identify the input x^* that maximizes the objective f over an admissible set of queries \mathcal{X} , i.e., $x^* = \arg \max_{x \in \mathcal{X}} f(x)$. To achieve this goal, BO relies on a *surrogate model* that provides a probabilistic representation of the objective f , and uses this information to compute an *acquisition function* to drive the selection of the most promising sample to query.

Surrogate model. We assume the available information regarding the objective function f be stored in the dataset $\mathcal{D}_t := \{(x_1, y_1), \dots, (x_t, y_t)\}$, where $y_t \sim \mathcal{N}(f(x_t), \sigma^2(x_t))$ is the noisy observation of the objective function by assuming the noise follows a zero-mean normal distribution with a standard deviation σ . The surrogate model depicts possible explanations of f as $f(x) \sim p(f(x)|\mathcal{D}_t)$ applying a joint distribution over its behavior at each sample $x \in \mathcal{X}$. In Bayesian inference, the prior distribution of the objective $p(f(x))$ is combined with the likelihood function $p(\mathcal{D}_t|f(x))$ to compute the posterior distribution $p(f(x)|\mathcal{D}_t) \propto p(\mathcal{D}_t|f(x))p(f(x))$, representing the updated beliefs about $f(x)$. Typically, Gaussian processes (GPs) have been widely used as the surrogate model for BO due to their efficient posterior sampling that enables cheap, gradient-based optimization of the acquisition function to propose new query points. GP is specified by a joint normal distribution $p(f(x)|\mathcal{D}_t) = \mathcal{N}(\mu_t(x), \kappa_t(x, x'))$ with mean $\mu_t(x)$ and kernel function $\kappa_t(x, x')$, where $\mu_t(x)$ represents the prediction and $\kappa_t(x, x')$ the associated uncertainty.

Acquisition function. The surrogate model is utilized to decide the next sample $x_{t+1} \in \mathcal{X}$ through the maximization of an acquisition function $\alpha : \mathcal{X} \mapsto \mathfrak{R}$, i.e., $x_{t+1} = \arg \max_{x \in \mathcal{X}} \alpha(x|\mathcal{D}_t)$, where $\alpha(x|\mathcal{D}_t)$ provides a measure of the improvement that the next query is likely to provide with respect to (w.r.t.) the current surrogate model of the objective function. Many acquisition functions have been proposed, including *probability of improvement* (Moćkus, 1975), *expected improvement* (Jones et al., 1998), *upper confidence bound*, and various *entropy search* methods (Hennig & Christian J. Schuler, 2012; Hernández-Lobato et al., 2014; Wang & Jegelka, 2017; Hvarfner et al., 2022a; Neiswanger et al., 2021), as well as practical approaches to optimize them (Wilson et al., 2018).

2.2 BAYESIAN EXPERIMENTAL DESIGN

Rather than optimizing an objective function $f(x)$, the purpose of BED is to sequentially select a set of experimental designs $x \in \mathcal{X}$ and gather outcomes y , to maximize the amount of information obtained about certain *parameters of interest*, denoted as θ . The parameters θ can correspond to some explicit model parameters, or any implicitly defined quantity of interest (e.g., the optimum of a function, the output of an algorithm, or future downstream predictions).

Based on the current history of experiments $\mathcal{D}_t := \{(x_1, y_1), \dots, (x_t, y_t)\}$, BED seeks to find the next experimental design x_{t+1} by maximizing the *expected information gain* (EIG) (Chaloner & Verdinelli, 1995) that a potential experimental outcome y_{t+1} can provide about θ , measured in terms of expected entropy reduction of the posterior distribution of θ :

$$\text{EIG}(x|\mathcal{D}_t) = H[p(\theta|\mathcal{D}_t)] - \mathbb{E}_{p(y|x, \mathcal{D}_t)}[H[p(\theta|\mathcal{D}_t \cup (x, y))]] = I(\theta; (x, y)|\mathcal{D}_t),$$

where $H(\cdot)$ and $I(\cdot)$ denote the entropy and mutual information, respectively.

3 A UNIFIED VIEW OF ACQUISITION STRATEGIES

The acquisition strategies in BO typically lead to *goal-directed* behavior, where the (implicit) goal is the optimum of a certain (unknown) objective function. On the contrary, the acquisition strategies in BED encourage *information-seeking* behavior, aiming to gather the maximum amount of information about certain parameters of interest. Although they both can be viewed as realizations of adaptive sampling (Di Fiore et al., 2023), there are no transferable approaches between these two domains due to distinct directives Hvarfner et al. (2025).

In this section, we show that these two seemingly competing imperatives can be naturally balanced through the principles of AIF.

3.1 ACTIVE INFERENCE AS EXPECTED FREE ENERGY MINIMIZATION

We specify a probabilistic surrogate model $q(\cdot)$ to capture the relationship between an outcome y and a decision variable x based on a set of parameters s that are of interest, which factorizes as

$$q(x, y, s) := p(x, y|s)q(s), \quad (1)$$

where $q(x, y, s)$ is a joint probability distribution over (x, y, s) based on a surrogate distribution $q(s)$ over s . We use $q(\cdot)$ to explicitly distinguish this model to be an ‘‘surrogate’’ of the ‘‘true’’ model $p(\cdot)$.

One special aspect of AIF is the way it formalizes goals. Instead of specifying the goal with additional variables related to ‘‘rewards’’ or ‘‘costs’’, AIF directly encodes the preferences over possible outcomes y through a probability distribution $p(y)$. In this distribution, outcomes with higher probabilities are treated as more rewarding. The deviation between observed outcomes and those desired is measured through an information-theoretic quantity known as *self-information* or *surprisal*: $-\log p(y)$. Intuitively, it’s a measure of how unexpected an outcome y is, given a prior preference distribution $p(y)$. Consistent with the intuitive notion of surprise, lower probability events generate higher surprisal values. Then the surprisal incurred from an observed outcome y is quantified by

$$\begin{aligned} -\log p(y) &= -\log \int p(y, s)ds = -\log \int \frac{p(y, s)q(s)}{q(s)} ds \\ &= -\log \mathbb{E}_{q(s)} \left[\frac{p(y, s)}{q(s)} \right] \leq -\mathbb{E}_{q(s)} \left[\log \frac{p(y, s)}{q(s)} \right] = F, \end{aligned} \quad (2)$$

where the last inequality follows *Jensen’s inequality*, which states that the expectation of a logarithm is always less than or equal to the logarithm of an expectation.

The right-hand side of equation 2 is called the *variational free energy* (VFE), whose name arises from the fact that F resembles the *Helmholtz free energy* in physics. We can see that VFE is always greater than or equal to surprisal (*i.e.*, it is an upper bound on surprisal). In machine learning, the sign of VFE is typically reversed, making it an *evidence lower bound* (ELBO). Maximizing the ELBO is a commonly used optimization approach in machine learning (Titsias, 2009).

To formulate a strategy for decision-making, we need to consider the decision variable x and the outcomes that result from the choice of actions. Since the future outcomes have not yet occurred, we resort to examining the expectation of surprisal over predicted outcomes based on a predictive distribution $q(y|x)$:

$$\begin{aligned} -\mathbb{E}_{q(y|x)} \log p(y|x) &\leq -\mathbb{E}_{q(y, s|x)} \left[\log \frac{p(y, s|x)}{q(s|x)} \right] = -\mathbb{E}_{q(y, s|x)} \left[\log \frac{p(s|x, y)p(y|x)}{q(s|x)} \right] \\ &= -\mathbb{E}_{q(y, s|x)} [\log p(s|x, y) - \log q(s|x)] - \mathbb{E}_{q(y, s|x)} \log p(y|x) \\ &= -\mathbb{E}_{q(y, s|x)} [\log p(s|x, y) - \log q(s|x)] - \mathbb{E}_{q(y|x)} \log p(y|x) = G, \end{aligned} \quad (3)$$

where the right-hand side of equation 3 is denoted as the *expected free energy* (EFE).

Theorem 3.1. *When using a surrogate model $q(\cdot) = p(\cdot|\mathcal{D}_t)$ that is considered as the true model constructed from all available data \mathcal{D}_t , G can be simplified as*

$$\begin{aligned} G &= -\mathbb{E}_{q(y|x)} [D_{KL}(q(s|x, y)||q(s))] - \mathbb{E}_{q(y|x)} \log p(y) \\ &= \underbrace{-I_q(s; (x, y))}_{\text{epistemic}} \underbrace{-\mathbb{E}_{q(y|x)} \log p(y)}_{\text{pragmatic}}, \end{aligned} \quad (4)$$

where $I_q(\cdot)$ represents the mutual information given the surrogate model $q(\cdot)$.

Proof. See Appendix B. □

We can see that by construction, EFE strikes a balance between *information-seeking* and *goal-directed* behavior under some prior preferences. It bounds the difference between *epistemic* value (about parameters) and *pragmatic* value (about outcomes), which captures the imperative to maximize the epistemic value (*i.e.*, information gain about latent states), from interactions with the environment, while maximizing the pragmatic value (*i.e.*, expected preference alignment), regarding

Table 1: A unified view of different acquisition strategies in BO and BED, where x^* , y^* represent the true optimal solution and value, respectively, and \hat{y} is the best value observed in \mathcal{D}_t .

Acquisition Strategy	Acquisition Function $\alpha(x)$	Expected Free Energy	
		Parameters s	Preferences $p(y)$
Expected Information Gain (Chaloner & Verdinelli, 1995)	$I(\theta; (x, y) \mathcal{D}_t)$	θ	-
Entropy Search (Hennig & Christian J. Schuler, 2012; Hernández-Lobato et al., 2014)	$I(x^*; (x, y) \mathcal{D}_t)$	x^*	-
Max-value Entropy Search (Wang & Jegelka, 2017)	$I(y^*; (x, y) \mathcal{D}_t)$	y^*	-
Joint Entropy Search (Hvarfner et al., 2022a)	$I((x^*, y^*); (x, y) \mathcal{D}_t)$	(x^*, y^*)	-
Bayesian Algorithm Execution (Neiswanger et al., 2021)	$I(\mathcal{O}_{\mathcal{A}}(f); (x, y) \mathcal{D}_t)$	$\mathcal{O}_{\mathcal{A}}(f)$	-
GP-Upper Confidence Bound	$\mu_t(x) + \beta^{1/2}\sigma_t(x)$	$f_{\mathcal{X}}$	$\exp\{y\}$
Probability of Improvement (Moćkus, 1975)	$p(y \geq \hat{y})$	-	$\exp\{\mathbb{I}(y \geq \hat{y})\}$
Expected Improvement (Jones et al., 1998)	$\mathbb{E}([y - \hat{y}]_+)$	-	$\exp\{[y - \hat{y}]_+\}$

prior preferences. This crucial aspect of AIF effectively addresses the “explore-exploit dilemma” because the imperatives for exploration and exploitation are just two aspects of EFE:

Pragmatic Value (Exploitation): This term encourages goal-directed behavior by favoring actions expected to yield preferred outcomes. Encoded by a prior distribution over desired observations, it functions similarly to a utility or reward function in reinforcement learning (Millidge et al., 2020), driving the agent to exploit its current knowledge to achieve its goals.

Epistemic Value (Exploration): This term promotes information-seeking behavior by favoring actions expected to reduce uncertainty about the underlying system maximally. It quantifies the expected information gain about the model’s parameters, driving the agent to explore the environment to refine its world model.

3.2 REINTERPRETATION OF ACQUISITION STRATEGIES IN BO AND BED

Crucially, minimizing EFE serves as a unifying umbrella principle. Many classic acquisition strategies in BO and BED can be reinterpreted as special cases of minimizing EFE, as shown in Table 1.

The reinterpretations of most of the acquisition strategies in Table 1 are straightforward according to their definitions. However, placing a rather intuitive GP-UCB strategy within this framework seems implicit. To reveal their connection, we rely on the following lemma:

Lemma 3.2. *Let $\mathbf{X} \subseteq \mathcal{X}$ be a subset of inputs, and the corresponding function values evaluated at those inputs be denoted as $f_{\mathbf{X}}$. Given a historical dataset \mathcal{D}_t , and new measurements \mathbf{Y} observed at \mathbf{X} , the mutual information*

$$I(f_{\mathcal{X}}; (\mathbf{X}, \mathbf{Y})|\mathcal{D}_t) = I(f_{\mathbf{X}}; \mathbf{Y}|\mathcal{D}_t),$$

for any (finite or infinite) set \mathcal{X} .

Proof. See Appendix C. □

Then if we assume constant Gaussian noises $\mathcal{N} \sim (0, \sigma^2)$ for the observations, we have

$$I(f_{\mathbf{X}}; y|\mathcal{D}_t) = H(y|\mathcal{D}_t) - H(y|f_{\mathbf{X}}, \mathcal{D}_t) = \frac{1}{2} \log(1 + \sigma^{-2}\sigma_t^2(x)),$$

where $\sigma_t^2(x)$ is the variance evaluated on the GP model $p(f_x|\mathcal{D}_t)$.

When further assuming that the GP kernel $\kappa_t(x, x') \leq 1, \forall x, x' \in \mathcal{X}$, then $0 \leq \sigma_t^2(x) \leq \kappa_t(x, x') \leq 1$, which gives

$$\log(1 + \sigma^{-2}\sigma_t^2(x)) \geq \log(1 + \sigma^{-2})\sigma_t^2(x).$$

If we choose $\beta = \frac{1}{2} \log(1 + \sigma^{-2})$, then the epistemic term in EFE, *i.e.*, $I(f_x; y|\mathcal{D}_t)$, provides an upper bound of the square of the exploration term $\beta^{1/2}\sigma_t(x)$ in GP-UCB.

This reveals the close relationship between GP-UCB and AIF, showing that even a seemingly pure intuitive strategy, like GP-UCB, can have rather rigorous mathematical foundations underlying it.

4 A NEW PARADIGM TO DERIVE ACQUISITION FUNCTIONS

The preference distribution $p(y)$ in Table 1 can be interpreted as a softmax transformation of a time-varying value function over the outcomes. Generalizing from this observation, we introduce “pragmatic curiosity”, a new paradigm for deriving acquisition functions applicable to a wider class of problems than traditional BO and BED.

Grounded on the *Boltzmann distribution* (also called *Gibbs distribution*) in statistical mechanics, we define an Boltzmann operator \mathcal{B}_β that maps an energy function $h(z)$ over domain \mathcal{Z} into a Boltzmann distribution:

$$(\mathcal{B}_\beta[h])(z) := \frac{e^{-h(z)/\beta}}{\int_{\mathcal{Z}} e^{-h(z)/\beta} dz}, \quad (5)$$

where β is called a temperature parameter (in allusion to statistical mechanics). A higher temperature results in a more uniform output distribution (*i.e.* “more random” with higher entropy), while a lower temperature results in a sharper output distribution, with the maximizers of $h(z)$ dominating.

Then, for any pre-defined energy function $h(y|\mathcal{D}_t)$, we can transfer it into a Boltzmann distribution and use it as the preference distribution $p(y)$ in equation 4:

$$\begin{aligned} G &= -I(s; (x, y)|\mathcal{D}_t) - \mathbb{E}_{p(y|x, \mathcal{D}_t)}[\log e^{-h(y|\mathcal{D}_t)/\beta}] + Z \\ &= -I(s; (x, y)|\mathcal{D}_t) + \frac{1}{\beta} \mathbb{E}_{p(y|x, \mathcal{D}_t)}[h(y|\mathcal{D}_t)] + Z, \end{aligned} \quad (6)$$

where $p(y|x, \mathcal{D}_t)$ is the predictive distribution of a surrogate model constructed from the historical data \mathcal{D}_t , and $Z = \log \int_{\mathcal{Y}} e^{-h(y|\mathcal{D}_t)/\beta} dy$ is a normalization constant independent of y .

Therefore, we propose a new acquisition paradigm by minimizing the EFE, *i.e.*,

$$x_{t+1} = \arg \max_{x \in \mathcal{X}} \{\beta_t I(s; (x, y)|\mathcal{D}_t) - \mathbb{E}_{p(y|x, \mathcal{D}_t)}[h(y|\mathcal{D}_t)]\}, \quad (7)$$

where $\beta_t \geq 0$ is a parameter that modulates the degree of “curiosity”, and $h(y|\mathcal{D}_t)$ is a problem-dependent energy function that captures the goal (in terms of regret w.r.t. a certain outcome y).

A key innovation of our framework is the problem-dependent energy function, $h(y|\mathcal{D}_t)$, which provides the flexibility to explicitly model uncertainty about the goals themselves. This allows us to derive acquisition functions for a class of complex problems often ignored by standard methods, including tasks with evolving goals (where conditions change over time) or implicit goals (which are not defined a priori). The advantages of this approach on such tasks are demonstrated in the following section.

5 EXPERIMENTS

In this section, we exemplify the benefits and variability of the proposed acquisition strategy by performing experiments across three categories that differ substantially from one another. [These problems originate from distinct literature within BO and BED, and none of them are canonical BO or BED tasks.](#) Thus each of them is evaluated against a different set of BO-type (optimization-focused) and BED-type (learning-focused) baselines appropriate for that specific task, ensuring a fair and rigorous comparison.

All acquisitions are evaluated via Monte Carlo estimation using the built-in samplers in BoTorch.

5.1 PARAMETRIC MODELS WITH KNOWN INVARIANT GOALS

We first consider a simplest task where the goals are known and invariant, and we have a decent understanding of the model, where only certain finite parameters are unknown.

One typical problem structure is the *constrained system identification* whose objective is to precisely learn unknown system parameters, θ , when valid observations can only be gathered under specific operational constraints, quantified by $C(y) \leq 0$. For instance, when monitoring a chemical plume, the goal is to identify its source (θ) while ensuring that sensor measurements (y) do not exceed a saturation threshold ($C(y)$).

In such cases, it is straightforward to choose the interested state as $s = \theta$, and the energy function as $h(y|\mathcal{D}_t) = \mathbb{I}(C(y) > 0)$. This leads to an acquisition function:

$$\alpha(x|\mathcal{D}_t) = \beta I(\theta; (x, y)|\mathcal{D}_t) - \mathbb{E}_{p(y|x, \mathcal{D}_t)} [\mathbb{I}(C(y) > 0)]. \quad (8)$$

Tasks. We perform experiments on a real-world environmental monitoring problem in 2d plume fields where the sensors have a saturation threshold y_{max} (i.e., $C(y) = y - y_{max}$) (Detailed settings and hyper-parameter choices in Appendix D.2). We consider three types of monitoring tasks: (a) locating the unknown source location; (b) estimating unknown wind direction and strength; and (c) identifying the active sources in fields with multiple sources.

Baselines. We compare our proposed acquisition strategy (AIF) with BO-type and BED-type baselines tailored this task: (a) Random; (b) Greedy by choosing the point that leads to the least probability of violating the constraint (BO-type); and (c) EIG about the unknown parameter (BED-type).

Evaluations. We evaluate the performance from both epistemic and pragmatic perspectives: (a) estimation accuracy; and (b) constraint violations.

Results. Fig. 1 shows that our approach achieves consistently stronger query efficiency over baselines while respecting all operational constraints, with cumulative constraint violation always being zero. This advantage is especially evident in the source localization task, where the drive for information and the need to satisfy constraints create opposing pressures. Resolving this conflict, our approach reaches near-perfect estimation using up to 40% fewer queries than competing methods.

Choices of β . To study the effect of β , we perform an ablation over different β values and report the estimation error in Fig. 2 (cumulative constraint violations are zero in all cases). The results show that the optimal β is task-dependent. From task (a) to (c), the correlation between sensor measurements and latent parameters weakens, so the mutual information term in equation 8 shrinks. In these less informative regimes, a larger β is needed to rescale the information-gain term. However, if β is too large, the acquisition becomes overly exploratory, which in turn degrades estimation performance. This yields a simple guideline: start with a moderate β , increase it only when the mutual information term in equation 8 is consistently small (e.g., due to weak sensor-parameter coupling), and stop or decrease β once further increases no longer reduce, or start to increase, the estimation error. In practice, β should be just large enough for the curiosity term to matter, but not so large that it induces unnecessary exploration.

5.2 NON-PARAMETRIC MODELS WITH KNOWN EVOLVING GOALS

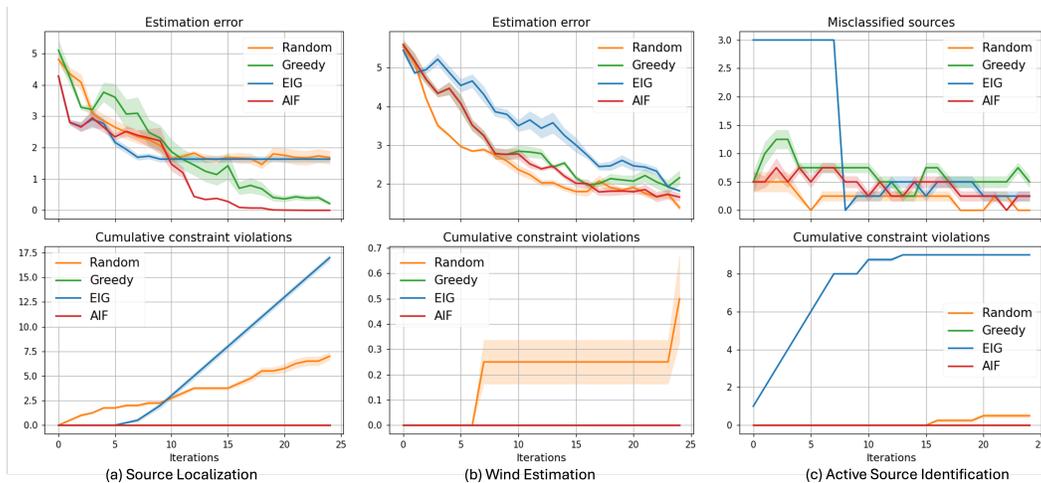
Next, we consider a more challenging setting where the task condition evolves, and the model is fully black-box such that we need to resort to a non-parametric model (e.g., GPs).

One such example is the *targeted active search* in multi-objective design problems. The objectives are viewed as metrics where specific ranges carry particular significance, and the goal is to design experiments that maximize coverage of these important regions \mathcal{S} .

Adopted from (Malkomes et al., 2021), we assume that the experimental design problems possess a sense of known resolution, such as simulation accuracy or manufacturing precision/tolerance, and any outcome within distance δ of another does not convey extra information about \mathcal{S} . Similar to (Malkomes et al., 2021), we define the coverage neighborhood of any y as $\mathbb{C}_\delta(y) := \{y' : d(y, y') < \delta\}$. And the coverage neighborhood of a set of points Y is defined as $\mathbb{C}_\delta(Y) := \bigcup_{y \in Y} \mathbb{C}_\delta(y)$.

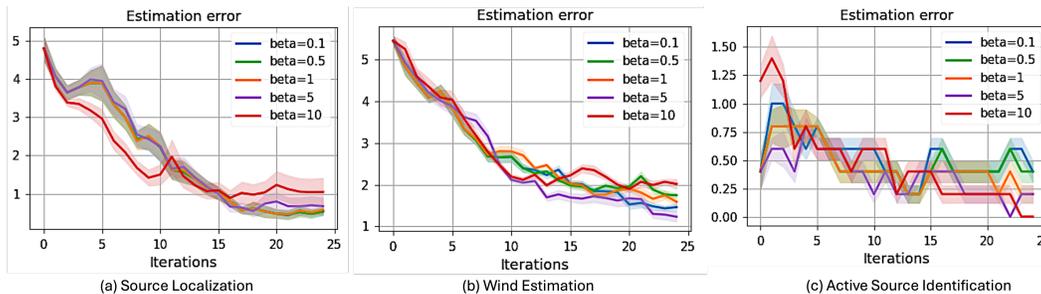
Since the goal is to cover as much volume of the interested region \mathcal{S} as possible, $h(y|\mathcal{D}_t)$ can be chosen as $h(y|\mathcal{D}_t) = \text{Vol}(\mathcal{S}) - \text{Vol}(\mathbb{C}_\delta(Y \cup y) \cap \mathcal{S})$, where Y contains all the history of outcomes.

378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393



394 Figure 1: Performance evaluation for constrained system identification on environmental monitoring
395 in 2d plume fields. Error bars represent ± 1 std over 5 seeds.
396

397
398
399
400
401
402
403
404
405
406
407



408 Figure 2: Ablation studies of β for constrained system identification on environmental monitoring
409 in 2d plume fields. Error bars represent ± 1 std over 5 seeds.
410

411
412
413
414

Similarly, coverage over the design configurations can be flexibly incorporated by expanding the interested state into the whole input space, *i.e.*, $s = f_{\mathcal{X}}$. With non-parametric models like GPs, we can leverage Lemma 3.2 and derive an acquisition function:

415
416

$$\alpha(x|\mathcal{D}_t) = \beta I(f_x; y|\mathcal{D}_t) + \mathbb{E}_{p(y|x, \mathcal{D}_t)}[\text{Vol}(\mathcal{C}_\delta(Y \cup y) \cap \mathcal{S})]. \quad (9)$$

417
418
419
420
421
422

Tasks. We perform experiments on a real-world failure discovery problem in autonomous driving scenarios where the perception module, a YOLO detector, may fail due to multiple causes, which could potentially lead to collisions (3d input-2d output). We consider three target sets with decreasing volume, *i.e.*, Target Set 1 \supset Target Set 2 \supset Target Set 3 (Detailed settings and hyper-parameter choices in Appendix D.3).

423
424
425

Baselines. We again compare our proposed acquisition strategy (AIF) with BO-type and BED-type baselines tailored this task: (a) Random; (b) Greedy by maximizing coverage volume in metrics space (BO-type); and (c) EIG by maximizing coverage volume in parameter space (BED-type).

426
427
428

Evaluations. We evaluate the performance by inspecting the coverage volume over both spaces: (a) metrics space ($\mathcal{C}_\delta(Y)$); and (b) parameter space ($\mathcal{C}_\delta(X)$).

429
430
431

Results. As shown in Fig. 3, our AIF algorithm effectively balances coverage of both the parameter and metrics spaces. This capability is especially impactful for smaller target sets where the search is more difficult. In the most challenging case (Target Set 3), our method identifies nearly 10% more of the critical failure region compared to the leading baseline.

432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485

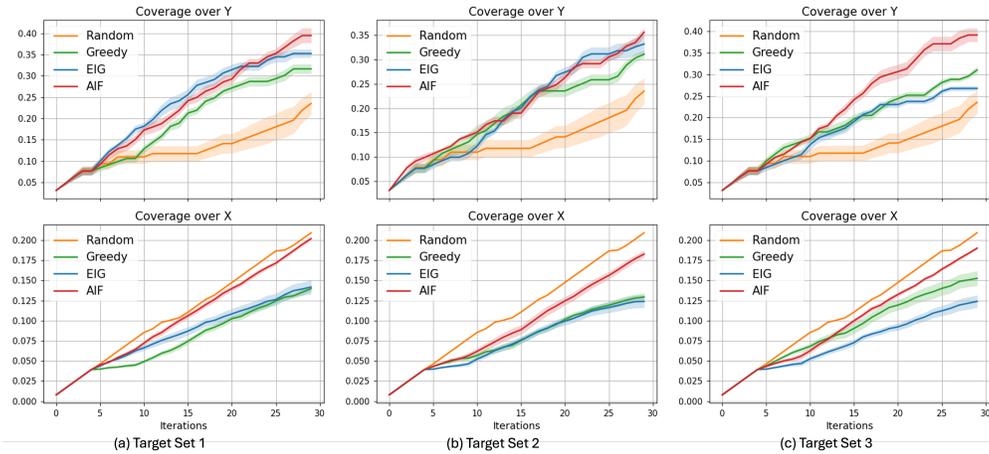


Figure 3: Performance evaluation for targeted active search on failure discovery in autonomous driving scenarios. Error bars represent ± 1 std over 4 seeds.

5.3 NON-PARAMETRIC MODELS WITH UNKNOWN GOALS

Finally, we look into the most difficult setting where both the models and goals are black-box.

One practical scenario arises from the *composite BO* in multi-objective optimization problems. The objectives are weighted by a preference function $g(y)$ that is unknown *a priori* and must be simultaneously learned during the optimization process.

As assumed in (Lin et al., 2022), we can estimate the unknown preference function $g(y)$ by asking a decision-maker (DM) to express preferences over pairs of outcomes (y_1, y_2) . Let $z(y_1, y_2) \in \{1, 2\}$ indicate whether the DM preferred the first or second outcome offered. Following (Chu & Ghahramani, 2005), we assume that the DM’s responses are distributed according to a *probit* likelihood: $L(z(y_1, y_2) = 1 | g(y_1), g(y_2)) = \Phi(\frac{g(y_1) - g(y_2)}{\sqrt{2\lambda}})$, where λ is a hyperparameter, and Φ is the standard normal CDF.

Taking the joint exploration and exploitation of the preference function $g(y)$ into consideration, we can choose $h(y | \mathcal{D}_t)$ as the EFE of the preference model, *i.e.*, $h(y | \mathcal{D}_t) = \text{EFE}(g(y))$, which naturally extends the acquisition function into a nested structure:

$$\alpha(x | \mathcal{D}_t) = \gamma I(f_x; y | \mathcal{D}_t) + \mathbb{E}_{p(y|x, \mathcal{D}_t)}[\beta I(g_y; z | \mathcal{D}_t) + \mathbb{E}_{p(g_y|y, \mathcal{D}_t)} g_y], \quad (10)$$

where $x = [x_1, x_2]$ are two jointly evaluated candidates, and $\gamma, \beta \geq 0$.

Tasks. We perform experiments on three real-world problems, including vehicle safety (5d-3d), penicillin production simulator (7d-3d), and distributed energy resource allocation in power grids (40d-4d) (Detailed settings and hyper-parameter choices in Appendix D.4).

Baselines. Different from (Lin et al., 2022), which separates and iterates the *preference exploration* and *experimentation* stages, our approach unifies the exploration and exploitation of both outcome and preference models simultaneously. In its best setting (*i.e.*, switch stages and update models at each step), their acquisition strategy reduces to a special case of our proposed equation 10 with $\beta = \gamma = 0$. Thus, we design a set of baselines to investigate the effect that each component in equation 10 plays for the joint exploration and exploitation of hierarchical models. We ablate the full AIF strategy (g + g IG + f IG) by (a) removing the last term (EIG); (b) removing the first term (g + g IG); and (c) removing the first two terms (g only).

Evaluations. We evaluate their performance by inspecting the best preference of all collected outcomes using the true preference function $g(y)$.

Results. Fig. 4 illustrates our AIF approach’s outstanding ability to learn unknown preference functions, a key advantage over baselines that often fail due to ill-directed queries. As the tasks become more complex and noisy (from (a) to (c)), our results demonstrate that every component of our ac-

486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539

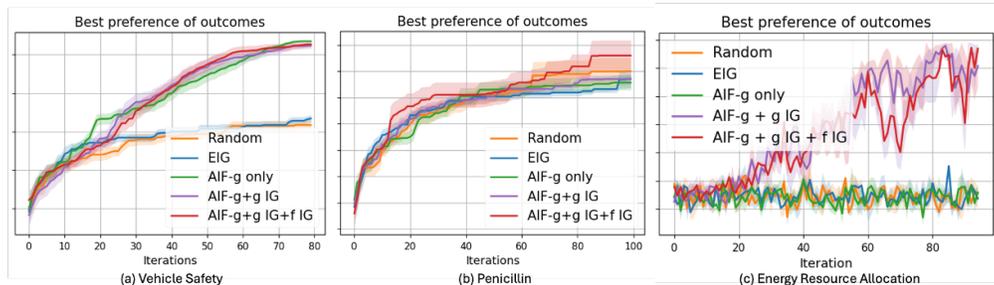


Figure 4: Performance evaluation for composite BO with unknown preferences. Error bars represent ± 1 std over 20 seeds for vehicle safety, penicillin, and 5 for energy resource allocation.

quisition function (10) plays an irreplaceable role in achieving optimal performance. This benefit is most stark in the energy resource allocation task, where competing methods failed to capture any meaningful preference model while our approach consistently succeeded.

Choices of β and γ . We also conducted an ablation study to investigate how the curiosity weights β and γ affect the learning dynamics. Detailed plots and analysis are included in Appendix D.4.3. Based on those results, we suggested the following design guideline. When the preference model $g(y)$ is complex or weakly informed by the initial data, β should be set relatively large so that the algorithm explicitly allocates queries to reduce uncertainty in g . At the same time, γ should be kept in a moderate-to-large range to ensure sufficient exploration of the outcome model $f(x)$. In simpler or more informative tasks, smaller values of β and γ already suffice, and the method is relatively insensitive to their precise tuning.

Benefits of joint learning and optimization. To highlight the benefits of jointly learning and optimizing, rather than separating these into stages, we compare our method against several BOPE variants from Lin et al. (2022) that use different stage-wise design choices. Detailed experimental setups, design choices, plots, and analyses are provided in Appendix D.4.4. The results show that our method naturally balances exploration and exploitation at each step and consistently discovers higher-preference regions, whereas the BOPE variants are highly sensitive to how the stages are configured. Consequently, stage-wise approaches like BOPE require careful manual tuning of these choices, while our unified formulation automates this trade-off and is therefore more amenable to higher-order hierarchical models.

6 CONCLUSION AND LIMITATION

In this work, we unified BO and BED by introducing “pragmatic curiosity”, a paradigm grounded in AIF. Our framework resolves the classic explore-exploit challenge by treating goal-seeking and information-seeking as two facets of a single, principled objective: minimizing EFE. We validated this approach across a suite of challenging hybrid tasks, demonstrating that our unified approach consistently outperforms standard baselines.

The potential impact of this work is to promote a conceptual shift from handling the explore-exploit dilemma with ad hoc heuristics to treating it as a unified inference problem. Our contribution is a concrete, principled step in this direction: we introduce an AIF-based acquisition that jointly accounts for learning and optimization, and demonstrate its benefits on tasks where these two objectives are tightly intertwined. We view this as a foundational rather than final step toward unifying *goal-directed* and *information-seeking* objectives, and hope it will invite more attention and follow-up work on this perspective.

The main limitation of our framework is the computational cost of estimating the EFE, which can be prohibitive in high-dimensional or time-critical applications. This points to clear avenues for future research, in particular developing more scalable and efficient approximations of the EFE so that the proposed framework can be applied to larger and more complex real-world problems.

540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593

ETHICS STATEMENT

This research is purely computational and does not involve human subjects, animal testing, or personally identifiable data. The primary ethical considerations relate to the potential downstream applications of our proposed “pragmatic curiosity” framework. We have focused on applications with clear societal benefits: enhancing environmental monitoring, improving safety in autonomous systems, and optimizing energy grid stability.

REPRODUCIBILITY STATEMENT

We are committed to the full reproducibility of our research.

Code: The complete source code for our framework, all baseline algorithms, and the experimental testbeds will be made publicly available in a GitHub repository upon publication of this manuscript. The code is implemented in Python and primarily relies on standard libraries such as PyTorch, GPyTorch, and BoTorch.

Experiments: All experimental results reported in this paper are averaged over multiple random seeds to ensure statistical robustness. The exact experiment environments, configurations, and hyper-parameters used to generate every figure and table are provided in Appendix D, allowing for the direct replication of our findings.

REFERENCES

- Eytan Bakshy, Lili Dworkin, Brian Karrer, Konstantin Kashin, Benjamin Letham, Ashwin Murthy, and Shaun Singh. AE: A domain-agnostic platform for adaptive experimentation. In *Conference on neural information processing systems*, pp. 1–8, 2018.
- Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. *Advances in neural information processing systems*, 33:21524–21538, 10 2020.
- Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical science*, pp. 273–304, 1995.
- Wei Chu and Zoubin Ghahramani. Preference learning with Gaussian processes. In *Proceedings of the 22nd international conference on Machine learning*, pp. 137–144, 2005.
- RP Cardoso Coelho, A Francisca Carvalho Alves, TM Nogueira Pires, and FM Andrade Pires. A composite bayesian optimisation framework for material and structural design. *Computer Methods in Applied Mechanics and Engineering*, 434:117516, 2025.
- Francesco Di Fiore, Michela Nardelli, and Laura Mainini. Active learning and bayesian optimization: A unified perspective to learn with a goal. *arXiv preprint arXiv:2303.01560*, 2023.
- Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Conference on robot learning*, pp. 1–16, 2017.
- Tommaso Dreossi, Daniel J Fremont, Shromona Ghosh, Edward Kim, Hadi Ravanbakhsh, Marcell Vazquez-Chanlatte, and Sanjit A Seshia. Verifai: A toolkit for the formal design and analysis of artificial intelligence-based systems. In *International Conference on Computer Aided Verification*, pp. 432–442, 2019.
- Peter I. Frazier. A Tutorial on Bayesian Optimization. (Section 5):1–22, 2018. URL <http://arxiv.org/abs/1807.02811>.
- Daniel J Fremont, Tommaso Dreossi, Shromona Ghosh, Xiangyu Yue, Alberto L Sangiovanni-Vincentelli, and Sanjit A Seshia. Scenic: a language for scenario specification and scene generation. In *Proceedings of the 40th ACM SIGPLAN conference on programming language design and implementation*, pp. 63–78, 2019.

-
- 594 Karl Friston. The free-energy principle: A unified brain theory?, 2 2010. ISSN 1471003X.
595
- 596 Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo.
597 Active Inference: A Process Theory. *Neural Computation*, 29(1):1–49, 1 2017. ISSN 0899-7667.
598 doi: 10.1162/NECO{_}_a{_}_00912.
- 599 Karl J Friston, Maxwell JD Ramstead, Alex B Kiefer, Alexander Tschantz, Christopher L Buckley,
600 Mahault Albarracín, Riddhi J Pitliya, Conor Heins, Brennan Klein, Beren Millidge, Dalton AR
601 Sakthivadivel, Toby St Clere Smithe, Magnus Koudahl, Safae Essafi Tremblay, Capm Petersen,
602 Kaiser Fung, Jason G Fox, Steven Swanson, Dan Mapes, and Gabriel René. Designing ecosystems
603 of intelligence from first principles. *Collective Intelligence*, 3(1), 1 2024. ISSN 2633-9137. doi:
604 10.1177/26339137231222481.
- 605 Jacob R. Gardner, Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wil-
606 son. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration.
607 *Advances in neural information processing systems*, 31, 9 2018.
- 608
- 609 Leonardo D González and Victor M Zavala. Implementation of a bayesian optimization framework
610 for interconnected systems. *Industrial & Engineering Chemistry Research*, 64(4):2168–2182,
611 2025.
- 612 Philipp Hennig and Christian J. Schuler. Entropy search for information-efficient global optimiza-
613 tion. *Journal of Machine Learning Research*, 13(6), 2012.
- 614
- 615 José Miguel Hernández-Lobato, W. Matthew Hoffman, and Zoubin Ghahramani. Predictive entropy
616 search for efficient global optimization of black-box functions. In *Advances in neural information
617 processing systems*, 2014.
- 618 Carl Hvarfner, Frank Hutter, and Luigi Nardi. Joint Entropy Search for Maximally-Informed
619 Bayesian Optimization. 6 2022a. URL <http://arxiv.org/abs/2206.04771>.
- 620
- 621 Carl Hvarfner, Frank Hutter, and Luigi Nardi. Joint entropy search for maximally-informed bayesian
622 optimization. *Advances in Neural Information Processing Systems*, 35:11494–11506, 2022b.
- 623
- 624 Carl Hvarfner, Erik Hellsten, Frank Hutter, and Luigi Nardi. Self-Correcting Bayesian Optimization
625 through Bayesian Active Learning. In *Conference on Neural Information Processing Systems
(NeurIPS)*, 2 2023. URL <http://arxiv.org/abs/2304.11005>.
- 626
- 627 Carl Hvarfner, David Eriksson, Eytan Bakshy, and Max Balandat. Informed initialization for
628 bayesian optimization and active learning. *arXiv preprint arXiv:2510.23681*, 2025.
- 629
- 630 Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. A Review of Yolo algorithm devel-
631 opments. *Procedia computer science*, 199:1066–1073, 2022.
- 632
- 633 Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient Global Optimization of Ex-
634 pensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, 1998. ISSN
09255001. doi: 10.1023/A:1008306431147.
- 635
- 636 Ehsan Kianmehr, Saman Nikkhah, Vahid Vahidinasab, Damian Giaouris, and Philip C. Taylor. A
637 Resilience-Based Architecture for Joint Distributed Energy Resources Allocation and Hourly Net-
638 work Reconfiguration. *IEEE Transactions on Industrial Informatics*, 15(10):5444–5455, 10 2019.
ISSN 1551-3203. doi: 10.1109/TII.2019.2901538.
- 639
- 640 Mina Konakovic Lukovic, Yunsheng Tian, and Wojciech Matusik. Diversity-guided multi-objective
641 bayesian optimization with batch evaluations. In *Advances in Neural Information Processing
642 Systems*, pp. 17708–17720, 2020.
- 643
- 644 Pablo Lanillos, Cristian Meo, Corrado Pezzato, Ajith Anil Meera, Mohamed Baioumy, Wataru
645 Ohata, Alexander Tschantz, Beren Millidge, Martijn Wisse, Christopher L. Buckley, and Jun
646 Tani. Active Inference in Robotics and Artificial Agents: Survey and Challenges. 12 2021. URL
647 <http://arxiv.org/abs/2112.01871>.
- Qiaohao Liang and Lipeng Lai. Scalable bayesian optimization accelerates process optimization of
penicillin production. In *NeurIPS 2021 AI for Science Workshop*, 2021.

-
- 648 Zhiyuan Jerry Lin, Raul Astudillo, Peter I. Frazier, and Eytan Bakshy. Preference Exploration for
649 Efficient Bayesian Optimization with Multiple Outcomes. In *International Conference on Arti-*
650 *ficial Intelligence and Statistics*, pp. 4235–4258, 3 2022. URL [http://arxiv.org/abs/](http://arxiv.org/abs/2203.11382)
651 [2203.11382](http://arxiv.org/abs/2203.11382).
- 652 Turab Lookman, Prasanna V Balachandran, Dezhen Xue, and Ruihao Yuan. Active learning in
653 materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj*
654 *Computational Materials*, 5(1):21, 2019.
- 656 Gustavo Malkomes, Bolong Cheng, Eric Hans Lee, and Michael Mccourt. Beyond the Pareto Ef-
657 ficient Frontier: Constraint Active Search for Multiobjective Experimental Design. Technical
658 report, 2021.
- 659 Takazumi Matsumoto, Kentaro Fujii, Shingo Murata, and Jun Tani. Active inference with dynamic
660 planning and information gain in continuous space by inferring low-dimensional latent states.
661 *Entropy*, 27(8):846, 2025.
- 663 Beren Millidge, Alexander Tschantz, Anil K. Seth, and Christopher L. Buckley. On the relationship
664 between active inference and control as inference. In *Communications in Computer and Informa-*
665 *tion Science*, volume 1326, pp. 3–11. Springer Science and Business Media Deutschland GmbH,
666 2020. ISBN 9783030649180. doi: 10.1007/978-3-030-64919-7{_}1.
- 667 J. Močkus. On Bayesian Methods for Seeking the Extremum. In *Optimization Techniques IFIP*
668 *Technical Conference*, pp. 400–404. Springer Berlin Heidelberg, Berlin, Heidelberg, 1975. doi:
669 10.1007/978-3-662-38527-2{_}55.
- 671 Willie Neiswanger, Ke Alexander Wang, and Stefano Ermon. Bayesian algorithm execution: Esti-
672 mating computable properties of black-box functions using mutual information. In *International*
673 *Conference on Machine Learning*, pp. 8005–8015, 2021.
- 674 Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern Bayesian
675 Experimental Design. 2 2023. URL <http://arxiv.org/abs/2302.14545>.
- 677 Manikandasriram Srinivasan Ramanagopal, Cyrus Anderson, Ram Vasudevan, and Matthew
678 Johnson-Roberson. Failing to Learn: Autonomously Identifying Perception Failures for Self-
679 Driving Cars. *IEEE Robotics and Automation Letters*, 3(4):3860–3867, 10 2018. ISSN 2377-
680 3766. doi: 10.1109/LRA.2018.2857402.
- 681 Joseph Redmon and Ali Farhadi. YOLOv3: An Incremental Improvement. *arXiv preprint*
682 *arXiv:1804.02767*, 4 2018.
- 684 Daniel Russo and Benjamin Van Roy. Learning to Optimize via Information-Directed Sampling.
685 *Operations Research*, 66(1):230–252, 2 2018. ISSN 0030-364X. doi: 10.1287/opre.2017.1663.
686 URL <https://pubsonline.informs.org/doi/10.1287/opre.2017.1663>.
- 687 Mark J. Schervish. *Theory of Statistics*. Springer New York, New York, NY, 1995. ISBN 978-1-
688 4612-8708-7. doi: 10.1007/978-1-4612-4250-5.
- 690 Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. Taking the
691 human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):
692 148–175, 2016. ISSN 15582256. doi: 10.1109/JPROC.2015.2494218.
- 693 Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom
694 Rainforth. Prediction-Oriented Bayesian Active Learning. In *International Conference on Artifi-*
695 *cial Intelligence and Statistics (AISTATS)*, 2023.
- 697 Ryan Smith, Karl J. Friston, and Christopher J. Whyte. A step-by-step tutorial on active inference
698 and its application to empirical data. *Journal of Mathematical Psychology*, 107, 4 2022. ISSN
699 10960880. doi: 10.1016/j.jmp.2021.102632.
- 700 Ryoji Tanabe and Hisao Ishibuchi. An Easy-to-use Real-world Multi-objective Optimization Prob-
701 lem Suite. 9 2020. doi: 10.1016/j.asoc.2020.106078.

702 Michalis Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *Artificial*
703 *intelligence and statistics*, pp. 567–574, 2009.
704

705 Ben Tu, Axel Gandy, Nikolas Kantas, and Behrang Shafei. Joint entropy search for multi-objective
706 bayesian optimization. *Advances in Neural Information Processing Systems*, 35:9922–9938,
707 2022.

708 Jianfeng Wang, Zhengyuan Yang, Xiaowei Hu, Linjie Li, Kevin Lin, Zhe Gan, Zicheng Liu, Ce Liu,
709 and Lijuan Wang. GIT: A Generative Image-to-text Transformer for Vision and Language. 5
710 2022.
711

712 Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient Bayesian optimization. In
713 *International Conference on Machine Learning*, pp. 3627–3635, 2017.

714 James T. Wilson, Frank Hutter, and Marc Peter Deisenroth. Maximizing acquisition functions for
715 Bayesian optimization. In *Conference on Neural Information Processing Systems (NeurIPS)*, 12
716 2018. URL <http://arxiv.org/abs/1805.10196>.
717

718 Miao Zhong, Kevin Tran, Yimeng Min, Chuanhao Wang, Ziyun Wang, Cao-Thang Dinh, Phil
719 De Luna, Zongqian Yu, Armin Sedighian Rasouli, Peter Brodersen, Song Sun, Oleksandr Voznyy,
720 Chih-Shan Tan, Mikhail Askerka, Fanglin Che, Min Liu, Ali Seifitokaldani, Yuanjie Pang, Shen-
721 Chuan Lo, Alexander Ip, Zachary Ulissi, and Edward H. Sargent. Accelerated discovery of CO2
722 electrocatalysts using active machine learning. *Nature*, 581(7807):178–183, 5 2020. ISSN 0028-
723 0836. doi: 10.1038/s41586-020-2242-8.
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

A RELATED WORK

Several independent attempts have been made to handle the hybrid objectives of goal seeking and learning. Below, we discuss some key prior works, originating from three viewpoints: (1) BO based settings, (2) BED (also known as Bayesian active learning, BAL) based works, and (3) some works that have acknowledged the synergy and attempted to present a unifying viewpoint, using adaptive sampling and AIF based approaches.

On the BO side, Russo & Van Roy (2018) proposed information-directed sampling (IDS) to online optimization problems by minimizing the ratio between squared expected single-period regret and a measure of information gain. Hvarfner et al. (2023) introduced a statistical distance-based active learning (SAL) criterion into the BO loop to actively learn the Gaussian process hyperparameter even as it searches for the optimum. Tu et al. (2022); Hvarfner et al. (2022b) propose information-theoretic acquisition function for multi-objective BO, which considers the joint information gain for the optimal set of inputs and outputs. Overall, these works acknowledge and demonstrate the value of information gain in enhancing the capabilities of goal-seeking strategies, with an ultimate focus on optimization. However, they do not explicitly comment on the synergy of goal seeking and learning, or formalize the acquisition strategies as a way to explicitly balance both objectives.

On the BED (also known as Bayesian active learning, BAL) side, Smith et al. (2023) proposed an expected predictive information gain (EPIG) criterion that focuses on information gain in model predictions, mitigating classical BAL’s tendency to select out-of-distribution or low-relevance queries by accounting for an input data distribution. Lookman et al. (2019) discusses the performance of various BAL and BO strategies for material design application, noting their connection with adaptive sampling, and simultaneous exploration and exploitation. This shows that realistic application of BO/BED based approaches often necessitates a formalized viewpoint unifying each distinct objectives.

These efforts begin to show the growing synergy between BO and BED techniques, where information-gain criteria are leveraged to enhance optimization and vice versa. In search of a unifying perspective for these techniques, Di Fiore et al. (2023) views BO and BAL as symbiotic adaptive sampling techniques driven by common principles of ‘goal-driven learning’. However, their viewpoint is limited to BO and BED as separate realizations of adaptive sampling, and they do not discuss a way to combine both objectives in a single acquisition strategy.

In this paper, we present a unified perspective that arises naturally from the principles of Active Inference (AIF) (Friston, 2010; Friston et al., 2017). Originally developed in computational neuroscience, AIF has been successfully applied in fields such as robotics (Lanillos et al., 2021), cognitive science (Smith et al., 2022), ecosystem modeling (Friston et al., 2024), and goal-directed planning in partially observable environments Matsumoto et al. (2025). These works leverage the ability of AIF to accommodate goal-directed and epistemic considerations, which we utilize as the backbone for our unifying mechanism discussed in this paper.

B PROOF OF THEOREM 3.1

Proof. Recall that

$$G = \underbrace{-\mathbb{E}_{q(y,s|x)} [\log p(s|x, y) - \log q(s|x)]}_{\text{Term 1}} \underbrace{-\mathbb{E}_{q(y|x)} \log p(y|x)}_{\text{Term 2}}$$

Since our purpose is to align the true outcome distribution given the decision x (i.e., $p(y|x)$) with the prior preference distribution $p(y)$, we replace the second term of G by $-\mathbb{E}_{q(y|x)} \log p(y)$. It is commonly referred to as *pragmatic* value, which scores the anticipated surprisal of the future outcomes given a predictive distribution $q(y|x)$.

For the first term of G , add and subtract $\log q(s|x, y)$ inside the expectation, we get

$$\begin{aligned}
\text{Term 1} &= -\mathbb{E}_{q(y, s|x)} [\log p(s|x, y) - \log q(s|x, y) + \log q(s|x, y) - \log q(s|x)] \\
&= -\mathbb{E}_{q(y|x)} \mathbb{E}_{q(s|x, y)} [\log p(s|x, y) - \log q(s|x, y) + \log q(s|x, y) - \log q(s|x)] \\
&= -\mathbb{E}_{q(y|x)} [\mathbb{E}_{q(s|x, y)} [\log p(s|x, y) - \log q(s|x, y)] + \mathbb{E}_{q(s|x, y)} [\log q(s|x, y) - \log q(s|x)]] \\
&= -\mathbb{E}_{q(y|x)} \left[\underbrace{-D_{\text{KL}}(q(s|x, y)||p(s|x, y))}_{\text{Term 3}} + \underbrace{D_{\text{KL}}(q(s|x, y)||q(s|x))}_{\text{Term 4}} \right].
\end{aligned} \tag{11}$$

Since the surrogate model $q(s)$ usually will not be updated until a new outcome y is observed, it doesn't depend on the decision variable x solely, i.e., $q(s|x) = q(s)$. Thus Term 4 becomes $D_{\text{KL}}(q(s|x, y)||q(s))$, which is commonly referred to as the *epistemic* value, or the expected information gain of a state when it is conditioned on expected observations.

Term 3 is, in general, intractable since it depends on the true posterior distribution $p(s|x, y)$ that is often unknown. The best surrogate model we can have is $q(\cdot) = p(\cdot|\mathcal{D}_t)$ that is constructed from all available data \mathcal{D}_t . Thus if we treat this surrogate model $q(s)$ as the best approximation of the true model $p(s)$, i.e., $q(s) \approx p(s)$, and the surrogate and true models both follow the Bayes' rule, i.e., $q(s|x, y) = \frac{p(x, y|s)q(s)}{\int p(x, y|s)q(s)ds}$, $p(s|x, y) = \frac{p(x, y|s)p(s)}{\int p(x, y|s)p(s)ds}$, then Term 3 vanishes. \square

C PROOF OF LEMMA 3.2

To prove Lemma 3.2, we first introduce a lemma:

Lemma C.1. For any (finite or infinite) set \mathcal{X} , after a few new measurements (\mathbf{X}, \mathbf{Y}) , $\mathbf{X} \subseteq \mathcal{X}$ the KL divergence between $p(f_{\mathcal{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))$ and $p(f_{\mathcal{X}}|\mathcal{D})$ is

$$D_{\text{KL}}[p(f_{\mathcal{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))||p(f_{\mathcal{X}}|\mathcal{D})] = \mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})} \left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \right],$$

where $L(\mathbf{Y}|f_{\mathbf{X}})$ is the likelihood of observations, $L(\mathbf{Y}) = \int L(\mathbf{Y}|f_{\mathbf{X}})p(f_{\mathbf{X}}|\mathcal{D})df_{\mathbf{X}}$ the marginal likelihood.

Proof. We first consider the case where \mathcal{X} is finite to provide an intuitive insight. The KL divergence between $p(f_{\mathcal{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))$ and $p(f_{\mathcal{X}}|\mathcal{D})$ can be given as

$$\begin{aligned}
&D_{\text{KL}}[p(f_{\mathcal{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))||p(f_{\mathcal{X}}|\mathcal{D})] \\
&= D_{\text{KL}}[p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))||p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D})] \\
&= \int p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y})) \log \frac{p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))}{p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D})} df_{\mathcal{X} \setminus \mathbf{X}} df_{\mathbf{X}} \\
&= \int p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y})) \log \frac{p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D})L(\mathbf{Y}|f_{\mathbf{X}})}{p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D})L(\mathbf{Y})} df_{\mathcal{X} \setminus \mathbf{X}} df_{\mathbf{X}} \\
&= \int p(f_{\mathcal{X} \setminus \mathbf{X}}, f_{\mathbf{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y})) \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} df_{\mathcal{X} \setminus \mathbf{X}} df_{\mathbf{X}} \\
&= \int p(f_{\mathbf{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y})) \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} df_{\mathbf{X}} \\
&= \int \frac{p(f_{\mathbf{X}}|\mathcal{D})L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} df_{\mathbf{X}} \\
&= \mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})} \left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \right],
\end{aligned} \tag{12}$$

where $L(\mathbf{Y}|f_{\mathbf{X}})$ is the likelihood of observations, $L(\mathbf{Y}) = \int L(\mathbf{Y}|f_{\mathbf{X}})p(f_{\mathbf{X}}|\mathcal{D})df_{\mathbf{X}}$ the marginal likelihood.

864 Now we move to the more general cases where \mathcal{X} is *infinite*. In such cases, there is no useful infinite-
865 dimensional Lebesgue measure with respect to an “infinite-dimensional vector” $f_{\mathcal{X}}$. Thus, we need
866 to resort to a more general definition for KL divergence based on the Radon-Nikodym derivative:
867

868
869 **Definition C.2.** If P and Q are probability measures over a set \mathcal{X} , and P is absolutely continuous
870 with respect to Q , then the KL divergence from P to Q is defined as
871

$$872 \quad D_{\text{KL}}[P||Q] = \int_{\mathcal{X}} \log\left(\frac{dP}{dQ}\right)dP,$$

873 where $\frac{dP}{dQ}$ is the Radon–Nikodym derivative of P with respect to Q , and provided the expression on
874 the right-hand side exists.
875

876
877
878
879
880 According to the measure-theoretic definition of Bayes’ theorem for a dominated model (Schervish,
881 1995), the Radon-Nikodym derivative of the posterior $P(\cdot) := p(\cdot|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))$ with respect to the
882 prior $\hat{P}(\cdot) := p(\cdot|\mathcal{D})$ is given as
883

$$884 \quad \frac{dP}{d\hat{P}}(f_{\mathcal{X}}) = \frac{L(\mathbf{Y}|f_{\mathcal{X}})}{L(\mathbf{Y})}.$$

885
886
887 Since the dataset \mathbf{Y} is finite, so similar to previous, we restrict the likelihood to only depend on the
888 finite dataset:
889

$$890 \quad \frac{dP}{d\hat{P}}(f_{\mathcal{X}}) = \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}.$$

891
892
893 Now the KL divergence between \hat{P} and P is quantified as
894

$$\begin{aligned} 895 & D_{\text{KL}}[P(f_{\mathcal{X}})||\hat{P}(f_{\mathcal{X}})] \\ 896 &= \int_{f_{\mathcal{X}}} \log\left(\frac{dP}{d\hat{P}}(f_{\mathcal{X}})\right)dP(f_{\mathcal{X}}) \\ 897 &= \int_{f_{\mathcal{X}}} \log\left(\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right)dP(f_{\mathcal{X}}) \\ 898 &= \int_{f_{\mathbf{X}}} \log\left(\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right) \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} d\hat{P}(f_{\mathbf{X}}) \quad (13) \\ 899 &= \int_{f_{\mathbf{X}}} \log\left(\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right) \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} d\hat{P}(f_{\mathbf{X}}) \\ 900 &= \mathbb{E}_{\hat{P}(f_{\mathbf{X}})} \left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \right], \end{aligned}$$

901
902
903
904
905
906
907
908 which has the exact same form as equation 12.

909
910
911 Therefore, we can conclude that regardless of the set \mathcal{X} being finite or infinite, the KL divergence
912 between the prior and posterior only depends on the evaluations of the observed data. That is to say,
913 whilst we are in fact quantifying the KL divergence between the full distributions, we only need to
914 keep track of the distribution over finite function values $f_{\mathbf{X}}$. \square
915
916
917

Now we are ready to prove Lemma 3.2.

918 *Proof.* According to Lemma C.1,
 919

$$\begin{aligned}
 920 \quad I(f_{\mathcal{X}}; (\mathbf{X}, \mathbf{Y})|\mathcal{D}) &= \mathbb{E}_{p(\mathbf{Y}|\mathbf{X}, \mathcal{D})}[D_{\text{KL}}[p(f_{\mathcal{X}}|\mathcal{D} \cup (\mathbf{X}, \mathbf{Y}))||p(f_{\mathcal{X}}|\mathcal{D})]] \\
 921 &= \mathbb{E}_{p(\mathbf{Y}|\mathbf{X}, \mathcal{D})}\mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})}\left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right] \\
 922 &= \mathbb{E}_{L(\mathbf{Y})}\mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})}\left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right] \\
 923 &= \mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})}\mathbb{E}_{L(\mathbf{Y})}\left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right] \\
 924 &= \mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})} \int \left[\frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})}\right] L(\mathbf{Y}) d\mathbf{Y} \\
 925 &= \mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})} \int L(\mathbf{Y}|f_{\mathbf{X}}) \log \frac{L(\mathbf{Y}|f_{\mathbf{X}})}{L(\mathbf{Y})} d\mathbf{Y} \\
 926 &= \mathbb{E}_{p(f_{\mathbf{X}}|\mathcal{D})}[D_{\text{KL}}[L(\mathbf{Y}|f_{\mathbf{X}})||L(\mathbf{Y})]] \\
 927 &= I(f_{\mathbf{X}}; \mathbf{Y}|\mathcal{D}).
 \end{aligned} \tag{14}$$

935 □

936 D EXPERIMENTAL DETAILS

937 This appendix provides a comprehensive overview of the simulation environment, model parameters, and hyper-parameter choices used to generate the results in this paper. The experiments were designed to be reproducible given the configurations outlined below.

938 D.1 SIMULATION ENVIRONMENT

939 All simulations for the perception failure evaluation in CARLA (Section D.3) were conducted on a Linux workstation with Ubuntu 22.04 LTS equipped with an Intel 13th Gen Core i7-13700KF CPU (16 cores, 24 threads, up to 5.4 GHz) and an NVIDIA GeForce RTX 4090 GPU (24 GB VRAM). The system ran CARLA simulations using CUDA 12.2 and NVIDIA driver version 535.230.02.

940 All other experiments were run on a MacBook Pro equipped with an Apple M2 Pro processor (10-core CPU, 16-core GPU) and a 3024 × 1964 Retina display. The GPU supports Metal 3, and the system was used as-is without external accelerators.

941 All experiments were conducted using Python 3.9. The core scientific computing libraries utilized were:

- 942 • BoTorch (Balandat et al., 2020)
- 943 • GPyTorch (Gardner et al., 2018)

944 In all experiments, we utilized the built-in Monte Carlo sampler in Botorch for the optimization of acquisition functions. The Monte Carlo samples are drawn from the posteriors for each model to approximate the expectations of acquisition functions.

945 D.2 CONSTRAINED SYSTEM IDENTIFICATION

946 D.2.1 PLUME FIELD MODEL AND PARAMETERS

947 We consider the monitoring of a chemical plume field, where multiple plume sources generate plume particles that can be measured by sensors. The field function is represented by the rate of hits, defined as the average number of particles per unit time measured by the sensor at a certain location.

948 The rate of hits for a chemical plume source is given as:

$$949 \quad R_{\theta}(x) = \frac{R_s}{\log \frac{\gamma}{a}} \exp\left(-\frac{\langle \theta - x, V \rangle}{2D}\right) K_0\left(\frac{\|\theta - x\|_2}{\gamma}\right),$$

972 where θ is the location of the plume source, R_s is the rate at which the plume source releases the
 973 plume particles in the environment, $\gamma = \sqrt{D\tau/(1 + \frac{\|V\|^2\tau}{4D})}$ is the average distance traveled by a
 974 plume particle in its lifetime, a is the size of the sensor detecting plume particles, V is the average
 975 wind velocity, D is the diffusivity of the plume particles, and K_0 is the Bessel function of zeroth
 976 order.
 977

978 The measurement y , *i.e.*, the number of particles measured, is modeled as a Poisson random variable
 979 with $R_\theta(x)\Delta t$ as the rate parameter, which leads to a likelihood model as
 980

$$981 L_\theta(y|x) = \frac{\exp(-R_\theta(x)\Delta t)(R_\theta(x)\Delta t)^y}{y!},$$

982 where Δt is the time taken to obtain a measurement.
 983

984 **General Field and Sensor Parameters:**

- 985 • Grid Size: 100 × 100 units.
- 986 • Sensor Size (a): 1.0 units.
- 987 • Measurement Time (Δt): 1.0 seconds.

988 **Source Parameters:** Eight potential plume sources were defined for the experiments. Their specific
 989 physical parameters are detailed as:
 990

SOURCE	θ	R_s	γ	V	D
1	[20.0, 20.0]	100.0	50.0	[0.5, 0.5]	10
2	[30.0, 80.0]	100.0	60.0	[-0.3, 0.2]	15
3	[45.0, 55.0]	15.0	50.0	[0.5, 0.5]	10
4	[50.0, 50.0]	18.0	30.0	[-0.3, 0.2]	15
5	[55.0, 45.0]	16.0	40.0	[0.2, -0.4]	12
6	[48.0, 52.0]	17.0	35.0	[0.1, 0.1]	11
7	[52.0, 55.0]	14.0	45.0	[-0.1, -0.1]	13
8	[52.0, 52.0]	18.0	40.0	[0.1, -0.1]	13

1002
 1003 **D.2.2 TASK-SPECIFIC CONFIGURATIONS**

1004 **Source Localization.**

- 1005 • Goal: Estimate the unknown location $\theta = [x, y]$ of a single active source.
- 1006 • Ground Truth: Source 1 was configured as the single active source.
- 1007 • Maximum Sensor Threshold (y_{max}): 60.0 hits/second.
- 1008 • Hypothesis Space: A discrete grid of potential locations spanning [0,100]×[0,100] with a
 1009 resolution of 5 units, resulting in a 20 × 20 grid of 400 hypotheses.
- 1010 • Degree of curiosity: $\beta = 0.5$.

1011 **Wind Estimation.**

- 1012 • Goal: Estimate the unknown wind vector $V = [v_x, v_y]$ for a single source with a known
 1013 location.
- 1014 • Ground Truth: Source 2 was used, with its true wind vector set to [-0.3, 0.2].
- 1015 • Maximum Sensor Threshold (y_{max}): 60.0 hits/second.
- 1016 • Hypothesis Space: A discrete grid of potential wind vectors spanning [-1,1]×[-1,1] with a
 1017 resolution of 0.1 units, resulting in a 20 × 20 grid of 400 hypotheses.
- 1018 • Degree of curiosity: $\beta = 1.0$.

1019 **Active Source Identification.**

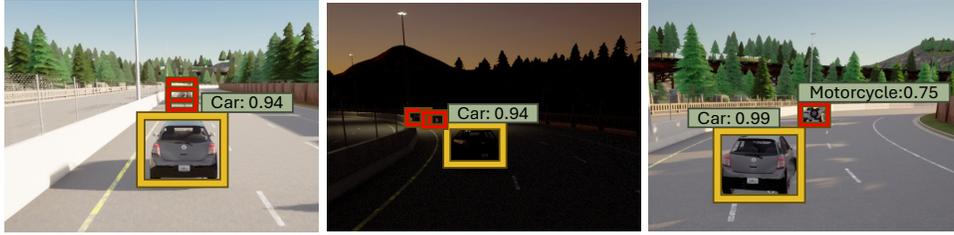


Figure 5: Examples of missed object detection by YOLO due to two reasons considered in perception failure case study in Section D.3. Fig (left to right): example of Failure-1 (distance), Failure-2 (poor light) and Failure-1 and Failure-2 both in one scene (distance and poor light), respectively. Bounding boxes for detected objects (mis-detections) shown in yellow (red) with detection confidence numbers. Each scene has two cars and a pedestrian.

- Goal: Identify the subset of active sources from a set of six potential sources with known locations.
- Ground Truth: Sources 3-8 were used, with Sources 3, 5, 6, and 8 were set as active.
- Maximum Sensor Threshold (y_{max}): 30.0 hits/second.
- Hypothesis Space: The set of all possible on/off combinations for the six sources. This is a discrete space with $2^6 = 64$ unique hypotheses.
- Degree of curiosity: $\beta = 5.0$.

D.3 CONSTRAINED ACTIVE SEARCH

D.3.1 PERCEPTION IN SELF DRIVING SIMULATION CARLA

We consider the failure discovery for YOLO-based object detection (Jiang et al., 2022; Redmon & Farhadi, 2018) in the CARLA simulator (Dosovitskiy et al., 2017).

This requires the generation of various scenarios in the environment using CARLA simulator. The environment is a composition of a static context and scenario ϕ . We use the probabilistic programming framework Scenic (Fremont et al., 2019) for sampling scenarios with varying contextual information for a fixed scenario variable ϕ . For the simulations presented in this paper, we used a publically available, pre-existing environment (Dreossi et al., 2019), which consists of the ego vehicle maneuvering on the road with two non-ego agents— two non-ego cars and a pedestrian crossing the road. We use YOLO object detection model to detect all non-ego agents in a scene. The generated scenario is seeded for reproducibility, so that for a given scenario ϕ , the environment can be treated as a deterministic quantity. Each scenario is defined using $\phi = [b_e, b_l, s]$, where $b_e, b_l \in [5, 15]$ represent the braking threshold of the ego car and lead car (non-ego car in-front of ego car) (m) and $s \in [0, \pi/2]$ denotes the sun altitude angle (rad). Each of these quantities is normalized to be within $[0, 1]$, and the normalized scenario is chosen as the decision variable x for active inference.

We are interested in environment variables (scenarios) that lead to two specific types of failures— failure to detect non-ego agents due to large distance from ego vehicle (Failure 1), and failure to detect non-ego agents due to poor scene lighting (Failure 2). Fig. 5 shows examples of the discussed object detection failures we aim to discover.

D.3.2 LLM-BASED EVALUATION FOR CARLA

We use LLM to perform evaluations for whether a generated scene corresponds to a failure due to specific type.

Each evaluation for a specific scenario corresponds to $T = 60$ steps of simulations. Images recorded from the camera view are used for object detection and classification at every 10 steps using YOLO-v3 (Redmon & Farhadi, 2018), and the classified image are used as inputs to GiT (Wang et al., 2022) to predict a likely failure type based on fine-tuning data.

1080 Results obtained from YOLO along-with the reports from GiT are used as inputs to GPT3 model
1081 for failure evaluation, which is queried 6 times per evaluation, and assigns binary scores pertaining
1082 to each failure mode for each scene (camera image). The average value reported across 6 scenes is
1083 used to construct $c_i : \mathcal{Z} \rightarrow \mathbb{R}$ for $i = 1, 2$ as: $c_i(z) = \frac{1}{T} \sum_{t=1}^T b_t^i$. Here $b_t^i \in \{0, 1\}$ is a scene-
1084 specific binary evaluation provided by the LLM based on report generated by GiT to assess if an
1085 object detection failure is observed in a given scene and corresponds to Failure- i .

1086 We use GPT-3.5 Turbo model for LLM-based binary evaluations, with each evaluation we query the
1087 LLM 6 times and combine the binary evaluations for all 6 runs. Note that usage of LLM is not a
1088 core part of our methodology and is only used to as a subjective evaluator.

1089 We show the prompt used for failure evaluation of each scene using the LLM in the box below. In-
1090 formation shown in red and blue is obtained from GiT captioning system, and CARLA respectively.
1091 The output of the LLM is used to obtain a binary number for each scene which is composed to give
1092 a scenario specific cost function c_1, c_2 .

Prompt Used for CARLA evaluation

1096 You will be provides the analysis of YOLO Object detection on an image that was taken from
1097 the camera feed of CARLA simulator. The simulator is simulating a pedestrian crossing the
1098 road before a car in front of the ego car. There are two cars and one pedestrian in each image.
1099 The information provided:

- 1100 1. *Objects detected*: List of objects detected by YOLO in the image. This list should
1101 have atleast one object from the *Objects to detect* list
- 1102 2. *Objects to detect* list:
 - 1103 • One object with one of the following labels: ‘car’, ‘truck’,
 - 1104 • One object with one of the following labels: ‘car’,
 - 1105 ‘truck’, ‘bus’, ‘motorcycle’, ‘bicycle’
 - 1106 • One object with one of the following labels: ‘person’
- 1107 3. Reason: The reason is a brief explanation of the failure to detect all objects, if that
1108 happens, and is generated by a pre-trained GiT model in the form of captions for
1109 the image.

1110 We are looking to discover images where YOLO fails to detect an object due to **bad light**
1111 and/or **large distance**. If the list of *objects detected* has an object missing from the *objects*
1112 *to detect* list, look at the reason. The reason can have other components as well, but it can
1113 ‘only’ be considered as **bad light** if at least one of the objects was failed to detect strictly
1114 due to **bad light**. Similarly, the reason can have other components as well, but it can ‘only’
1115 be considered as **large distance** if the reason contains the phrase ‘far away’. Follow the
1116 response instructions while responding.

1117 Response Instructions: Respond should be an integer 0, 1, 2, 3 or 4:

- 1118 • 0 indicating that at least one object was missing from the ‘objects to detect’ list, but
1119 the reason provided does not correspond to bad light or large distance.
- 1120 • 1 indicating that an object was not detected and the reason provided corresponds to
1121 bad light only.
- 1122 • 2 indicating that an object was not detected, and the reason corresponds to large
1123 distance only.
- 1124 • 3 indicating that an object was not detected, and the reason corresponds to both
1125 large distance and bad light.
- 1126 • 4 indicating all objects are detected. Do not provide explanation.

1127 Response format: Response: [integer], where integer = 0,1,2,3,4.

1128 The list of objects detected and reason for incomplete detection for the image are as follows:

- 1129 • Objects detected: {objects}
- 1130 • Reason: {reason}

1134 D.3.3 TASK-SPECIFIC CONFIGURATIONS

- 1135 • Goal: We consider two cost functions c_1, c_2 associated with each type of failure. The goal
1136 is to sample from the set $\Omega = \{z | c_1(z) \geq C_1, c_2(z) \geq C_2\}$, and we consider three target
1137 sets defined by $C_1 = C_2 = 0.1$, $C_1 = C_2 = 0.5$, and $C_1 = C_2 = 0.8$.
1138
- 1139 • Degree of curiosity: $\beta = 20.0$.
1140

1141 D.4 COMPOSITE BAYESIAN OPTIMIZATION

1142 D.4.1 PREFERENCE EVALUATION

1144 We simulate human-in-the-loop or policy-driven decision-making via pairwise preference queries.
1145 That is, for selected pairs of outcomes (y_1, y_2) , a preference function indicates which design is
1146 preferred. These preferences are generated based on a latent utility function, not revealed to the
1147 optimizer.

1148 An initial set of 1 pairwise preferences is randomly sampled to initialize the model. Each step of the
1149 optimization selects new pairs to query, guided by the used acquisition strategy.
1150

1151 D.4.2 TASK-SPECIFIC CONFIGURATIONS

1152 **Vehicle Safety.**

- 1153 • Goal: Optimize vehicle crash-worthiness.
- 1154 • Testbed: See (Tanabe & Ishibuchi, 2020) for details.
- 1155 • Ground Truth Preference: $g(y) = -(y - y^*)^2$, where $y^* = [1864.7202, 11.8199, 0.2904]$.
- 1156 • Degree of curiosity: $\beta = \gamma = 1.0$.
1157
1158
1159

1160 **Penicillin.**

- 1161 • Goal: Maximize the penicillin yield while minimizing time to ferment and the CO2 byprod-
1162 uct.
- 1163 • Testbed: See (Liang & Lai, 2021) for details.
- 1164 • Ground Truth Preference: $g(y) = -(y - y^*)^2$, where $y^* = [25.935, 57.612, 935.5]$.
- 1165 • Degree of curiosity: $\beta = \gamma = 1.0$.
1166
1167
1168

1169 Performance Metrics	Definition
1170 Voltage Fairness	Measures the variance in bus voltages across the network; lower variance im- 1171 plies more equitable voltage delivery.
1172 Total Cost	Combines capital expenditures for DER installation and operational costs re- 1173 lated to reactive power support.
1174 Priority Area Coverage	Quantifies the share of power delivered to high-priority buses, such as rural or 1175 underserved regions.
1176 Resilience	Assesses the percentage of time that all bus voltages remain within safe operat- 1177 ing limits under perturbations (e.g., load uncertainty or line outages).

1178 **Energy Resource Allocation.**

- 1179 • Goal: Identify deployment strategies for Distributed Energy Resources (DERs) in Optimal
1180 Power Flow (OPF) that align with implicit ethical preferences across multiple performance
1181 dimensions detailed in the following table.
- 1182 • Testbed: IEEE 30-bus network in pandapower library.
- 1183 • Ground Truth Preference: $g(y) = a^\top y$, where $a = [1, -1, 2, 1]$
- 1184 • Degree of curiosity: $\beta = \gamma = 1.0$.
1185
1186
1187

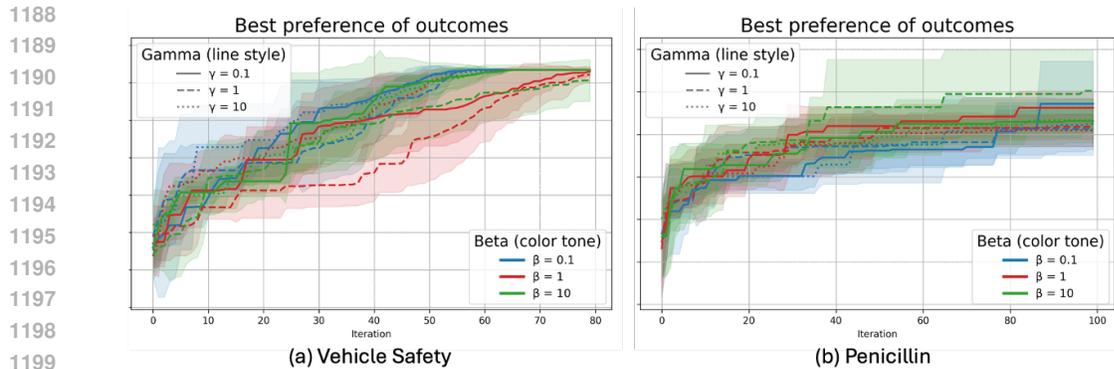


Figure 6: Ablation studies on β and γ for vehicle safety and penicillin. Error bars represent ± 1 std over 5 seeds.

D.4.3 ABLATION STUDIES ON β AND γ

We further conduct an ablation study to investigate how the curiosity weights β and γ affect the learning dynamics. Recall that β controls the mutual information term for the preference model $g(y)$, while γ controls the mutual information term for the outcome model $f(x)$. Figure 6 reports the evolution of the best observed preference value $g(y)$ over iterations, with line style encoding γ and color encoding β .

For the penicillin task (right), performance is clearly sensitive to both β and γ . Runs with a small preference-curiosity weight ($\beta = 0.1$, blue) consistently underperform those with larger β , indicating that actively reducing uncertainty in the preference model is crucial. Increasing β to 10 (green) yields the highest final preference values, especially when combined with moderate or large outcome curiosity ($\gamma = 1$ or $\gamma = 10$); these settings converge faster and reach higher plateaus than configurations with $\gamma = 0.1$. This suggests that, in problems where the mapping from outcomes to preferences is non-trivial, it is beneficial to invest in both exploring the outcome space and actively refining the preference model.

For the vehicle safety task (left), the effect of β and γ is more modest. Most configurations eventually reach similar best preference values, and the differences appear mainly in the early iterations: larger γ (dotted lines) tends to accelerate initial improvement, whereas small β (blue) already suffices to achieve desired exploration. Overall, this task is more robust to the precise choice of β and γ , consistent with an easier or more informative outcome–preference structure.

D.4.4 EXTENDED BASELINE COMPARISON WITH BOPE

To highlight the benefits of jointly learning and optimizing, rather than separating these into stages, we extend the baseline comparison with **BOPE** from Lin et al. (2022).

The original BOPE framework is intentionally flexible and leaves many problem-specific design choices open, especially regarding how and when to switch between preference exploration and experimentation. In our comparison, we consider four representative stage-wise variants:

- **BOPE-I:** A two-phase strategy that starts with qEUBO (preference-focused exploration) and switches to qNEI in the second half (objective-driven refinement), illustrating the effect of premature exploitation.
- **BOPE-II:** A two-phase strategy that starts with qNEI (exploring the objective space) and switches to qEUBO in the second half (exploiting the learned preference model), using a frozen outcome snapshot for qEUBO.
- **BOPE-III:** A qEUBO-only variant where experiments are selected by qEUBO with a newly sampled objective realization at each iteration, encouraging stronger exploration through objective variation.

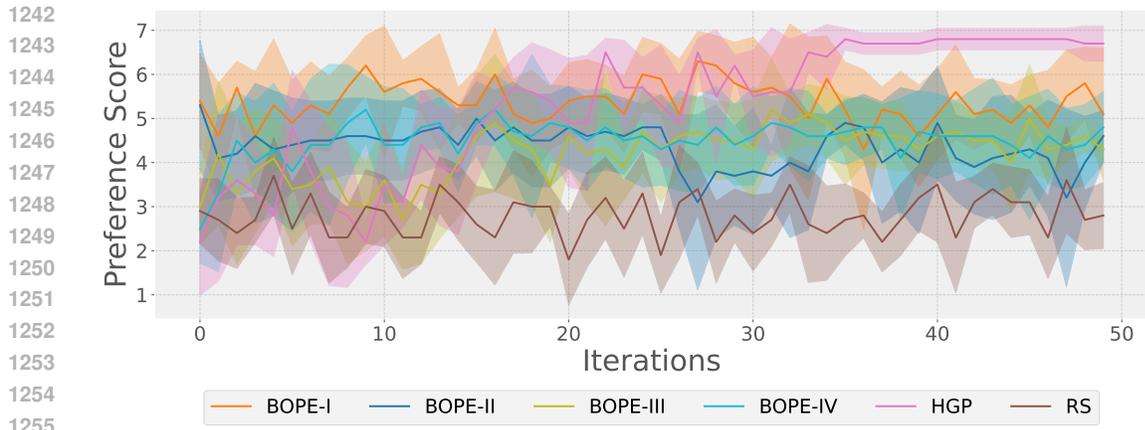


Figure 7: Extended baseline comparison with BOPE for energy resource allocation. Error bars represent ± 1 standard deviation over 5 seeds.

- **BOPE-IV**: A BOPE variant that uses standard preference exploration (qEUBO) and selects experiments exclusively with qNEI, fully refitting both outcome and preference GPs after each update.

Figure 7 reports their preference scores over random sampling (**RS**). It is evident that our active-inference method for hierarchical GPs (**HGP**) consistently discovers higher-preference regions after a brief initial exploration phase, while BOPE variants are highly sensitive to their stage-wise design choices. **BOPE-I**, being preference-driven in the first phase, initially attains higher preference values but fails to balance exploration and exploitation, leading to poor convergence in the second half; its performance is also sensitive to the precise switching point. **BOPE-II** explores first and then optimizes, achieving better final performance than BOPE-I, but the strict separation between exploration and exploitation still yields suboptimal outcomes. **BOPE-III** mixes both aspects but remains more exploitation-centric, performing better than BOPE-I/II yet still below HGP. **BOPE-IV** and HGP share the idea of refitting both models at each step, but BOPE-IV converges to a lower-preference solution. In contrast, our acquisition strategy jointly leverages information from both the outcome and preference models at every iteration, leading to higher sample efficiency and more reliable discovery of high-preference regions.

E USE OF LARGE LANGUAGE MODELS

We use LLM to revise and polish drafts of the introduction, abstract, conclusion, and limitations sections to enhance clarity, flow, and accessibility for a broader audience.

We also use LLM as an evaluator for the experiments of failure discovery in CARLA, which is detailed in Appendix D.3.