Information-Theoretic Safe Exploration with Gaussian Processes

Anonymous Author(s)
Affiliation
Address
email

Abstract

We consider a sequential decision making task where we are not allowed to evaluate parameters that violate an *a priori* unknown (safety) constraint. A common approach is to place a Gaussian process prior on the unknown constraint and allow evaluations only in regions that are safe with high probability. Most current methods rely on a discretization of the domain and cannot be directly extended to the continuous case. Moreover, the way in which they exploit regularity assumptions about the constraint introduces an additional critical hyperparameter. In this paper, we propose an information-theoretic safe exploration criterion that directly exploits the GP posterior to identify the most informative safe parameters to evaluate. Our approach is naturally applicable to continuous domains and does not require additional hyperparameters. We theoretically analyze the method and show that we do not violate the safety constraint with high probability and that we explore by learning about the constraint up to arbitrary precision. Empirical evaluations demonstrate improved data-efficiency and scalability.

1 Introduction

In sequential decision making problems, we iteratively select parameters in order to optimize a given performance criterion. However, real-world applications such as robotics (Berkenkamp et al., 2021), mechanical systems (Schillinger et al., 2017) or medicine (Sui et al., 2015) are often subject to additional safety constraints that we are not allowed to violate during the exploration process (Dulac-Arnold et al., 2019). Since it is *a priori* unknown which parameters lead to constraint violations, we need to actively and carefully learn about the constraints without ever violating them. That is, we need to learn about the safety of parameters by only evaluating parameters that are currently known to be safe.

Existing methods by Schreiter et al. (2015); Sui et al. (2015) tackle this problem by placing a Gaussian process (GP) prior over the constraint and only evaluate parameters that do not violate the constraint with high probability. To learn about the safety of parameters, they evaluate the constraint at the parameter with the largest posterior variance. This process is made more efficient by SafeOpt, which restricts exploration to parameters that are close to the boundary of the current set of safe parameters (Sui et al., 2015) at the cost of an additional tuning hyperparameter (Lipschitz constant). However, uncertainty about the constraint is only a proxy objective, which only indirectly learns about the safety of parameters. Consequently, data-efficiency could be improved with an exploration criterion that directly maximizes the information gained about the safety of other parameters.

Our contribution In this paper, we propose Information-Theoretic Safe Exploration (ISE), a safe exploration algorithm that *directly* exploits the information gain about the safety of parameters in order to expand the region of the parameter space that we can classify as safe with high confidence. By directly optimizing for safe information gain, ISE is more data-efficient than existing approaches

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

without manually restricting evaluated parameters to be on the boundary of the safe set. This exploration criterion also means that we do not require additional hyperparameters beyond the GP posterior and that ISE is directly applicable to continuous domains. We theoretically analyze our method and prove that it learns about the safety of reachable parameters to arbitrary precision.

Related work Information-based selection criteria with Gaussian processes models are successfully used in the context of unconstrained Bayesian optimization (BO, Shahriari et al. (2016); Bubeck and Cesa-Bianchi (2012)), where the goal is to find the parameters that maximize an *a priori* unknown function. Hennig and Schuler (2012); Henrández-Lobato et al. (2014); Wang and Jegelka (2017) select parameters that provide the most information about the optimal parameters, while Fröhlich et al. (2020) consider the information under noisy parameters. The success of these information-based approaches also relies in the superior data efficiency that they demonstrated. We draw inspiration from these methods when defining an information-based criterion w.r.t. the safety of parameters to guide safe exploration.

In the presence of constraints that the final solution needs to satisfy, but which we can violate during exploration, Gelbart et al. (2014) propose to combine typical BO acquisition functions with the probability of satisfying the constraint. Instead, Gotovos et al. (2013) propose an uncertainty-based criterion that learns about the feasible region of parameters. When we are not allowed to ever evaluate unsafe parameters, safe exploration is a necessary sub-routine of BO algorithms to learn about the safety of parameters. To safely explore, Schreiter et al. (2015) globally learn about the constraint by evaluating the most uncertain parameters. SAFEOPT by Sui et al. (2015) extends this to joint exploration and optimization and makes it more efficient by explicitly restricting safe exploration to the boundary of the safe set. Sui et al. (2018) proposes STAGEOPT, which additionally separates the exploration and optimization phases. Both of these algorithms assume access to a Lipschitz constant to define parameters close to the boundary of the safe set, which is a difficult tuning parameter in practice. These methods have been extended to multiple constraints by Berkenkamp et al. (2021), while Kirschner et al. (2019) scale them to higher dimensions with LINEBO, which explores in low-dimensional sub-spaces. To improve computational costs, Duivenvoorden et al. (2017) suggest a continuous approximation to SAFEOPT without providing exploration guarantees. All of these methods rely on function uncertainty to drive exploration, while we directly maximize the information gained about the safety of parameters.

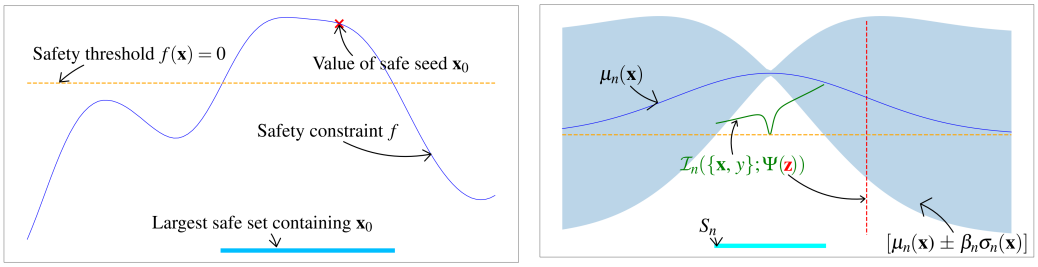
Safe exploration also arises in the context of Markov decision processes (MDP), (Moldovan and Abbeel, 2012; Hans et al., 2008). In particular, Turchetta et al. (2019) traverse the MDP to learn about the safety of parameters selected by some oracle. At its core, their method explores using the same ideas as SAFEOPT and STAGEOPT to select parameters to evaluate. Consequently, our proposed method for safe exploration is also directly applicable to their setting.

2 Problem Statement

In this section, we introduce the problem and notation that we use throughout the paper. We are given an unknown and expensive to evaluate safety constraint $f : \mathcal{X} \to \mathbb{R}$ s.t. parameters that satisfy $f(\boldsymbol{x}) \geq 0$ are classified as safe, while others are unsafe. To start exploring safely, we also have access to an initial safe parameter $\boldsymbol{x}_0$ that satisfies the safety constraint, $f(\boldsymbol{x}_0) \geq 0$. We sequentially select safe parameters $\boldsymbol{x}_n \in \mathcal{X}$ where to evaluate f in order to learn about the safety of parameters beyond $\boldsymbol{x}_0$. At each iteration n, we obtain a noisy observation of $y_n \coloneqq f(\boldsymbol{x}_n) + \nu_n$ that is corrupted by additive homoscedastic Gaussian noise $\nu_n \sim \mathcal{N}\big(0, \sigma_\nu^2\big)$. We illustrate the task in Figure 1a, where starting from $\boldsymbol{x}_0$ we aim to safely explore the domain so that we ultimately classify as safe all the safe parameters that are reachable from $\boldsymbol{x}_0$.

As f is unknown and the evaluations y_n are noisy, it is not feasible to select parameters that are safe with certainty and we provide high-probability safety guarantees instead. To this end, we assume that the safety constraint f has bounded norm in the Reproducing Kernel Hilbert Space (RKHS) (Schölkopf and Smola, 2002) $\mathcal{H}_k$ associated to some kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with $k(\boldsymbol{x}, \boldsymbol{x}') \leq 1$. This assumption allows us to to model f as a Gaussian process (GP) (Srinivas et al., 2010).

Gaussian Processes A GP is a stochastic process specified by a mean function $\mu : \mathcal{X} \to \mathbb{R}$ and a kernel k (Rasmussen and Williams, 2006). It defines a probability distribution over real-valued functions on $\mathcal{X}$, such that any finite collection of function values at parameters $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ is distributed as a multivariate normal distribution. The GP prior can then be conditioned on (noisy)



(a) Problem components.

(b) ISE mutual information.

Figure 1: In (a) we illustrate the safe exploration task. Based on the unknown safety constraint f, we are only allowed to evaluate safe parameters $\boldsymbol{x}$ with values $f(\boldsymbol{x}) \geq 0$ above the safety threshold (dashed line). Starting from a safe seed $\boldsymbol{x}_0$ a safe exploration strategy needs to discover the largest reachable safe region of the parameter space containing $\boldsymbol{x}_0$. In (b) we show the mutual information $I_n(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z}))$ in green for different $\boldsymbol{x}$ inside the safe set and for a fixed $\boldsymbol{z}$ outside (red dashed line). ISE maximizes this information jointly over $\boldsymbol{x}$ and $\boldsymbol{z}$.

function evaluations $\mathcal{D}_n = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$. If the noise is Gaussian, then the resulting posterior is also a GP and with posterior mean and variance

$$\begin{aligned} \mu_n(\boldsymbol{x}) &= \mu(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})^\top (\boldsymbol{K} + \boldsymbol{I}\sigma_\nu^2)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \\ \sigma_n^2(\boldsymbol{x}) &= k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x})^\top (\boldsymbol{K} + \boldsymbol{I}\sigma_\nu^2)^{-1}\boldsymbol{k}(\boldsymbol{x}), \end{aligned} \tag{1}$$

where $\boldsymbol{\mu} := [\mu(\boldsymbol{x}_1), \dots \mu(\boldsymbol{x}_n)]$ is the mean vector at parameters $\boldsymbol{x}_i \in \mathcal{D}_n$ and $[\boldsymbol{y}]_i := y(\boldsymbol{x}_i)$ the corresponding vector of observations. We have $\left[\boldsymbol{k}(\boldsymbol{x})\right]_i := k(\boldsymbol{x}, \boldsymbol{x}_i)$, the kernel matrix has entries $[\boldsymbol{K}]_{ij} := k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $\boldsymbol{I}$ is the identity matrix.

Safe set Using the previous assumptions, we can construct high-probability confidence intervals on the function values $f(\boldsymbol{x})$. Concretely, for any $\delta > 0$ it is possible to find a sequence of positive numbers $\{\beta_n\}$ such that $f(\boldsymbol{x}) \in \left[\mu_n(\boldsymbol{x}) \pm \beta_n \sigma_n(\boldsymbol{x})\right]$ with probability at least $1 - \delta$, jointly for all $\boldsymbol{x} \in \mathcal{X}$ and $n \geq 1$. For a proof and more details see (Chowdhury and Gopalan, 2017). We use these confidence intervals to define a *safe set*

$$S_n := \{\boldsymbol{x} \in \mathcal{X} : \mu_n(\boldsymbol{x}) - \beta_n \sigma_n(\boldsymbol{x}) \geq 0\} \cup \{\boldsymbol{x}_0\}, \tag{2}$$

which contains all parameters whose β_n-lower confidence bound is above the safety threshold and the initial safe parameter $\boldsymbol{x}_0$. Consequently, we know that all parameters in S_n are safe, $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in S_n$, with probability at least $1 - \delta$ jointly over all iterations n.

Safe exploration Given the safe set S_n, the next question is which parameters in S_n to evaluate in order to efficiently expand it. Most existing safe exploration methods rely on uncertainty sampling over subsets of S_n. SAFEOPT-like approaches, for example, use the Lipschitz assumption on f to identify parameters in S_n that could expand the safe set, and then select the parameter that has the biggest uncertainty among those. In the next sections, we present and analyze our safe exploration strategy, ISE, that instead uses an information gain measure to identify the parameters that allow us to efficiently learn about the safety of parameters outside of S_n.

3 Information-Theoretic Safe Exploration

We present Information-Theoretic Safe Exploration (ISE), which guides the safe exploration by using an information-theoretic criterion. Our goal is to design an exploration strategy that directly exploits the properties of GPs to learn about the safety of parameters outside of S_n. We draw inspiration from Hennig and Schuler (2012); Wang and Jegelka (2017), who exploit information-theoretic insights to design data-efficient BO acquisition functions for their respective optimization objectives.

Information gain measure In our case, we want to evaluate f at safe parameters that are maximally informative about the safety of other parameters, in particular of those where we are uncertain about whether they are safe or not. To this end, we need a corresponding measure of information

Algorithm 1 Information-Theoretic Safe Exploration

Input: GP prior (μ_0, k, σ_ν), Safe seed $\boldsymbol{x}_0$
for $n = 0, \dots$ **do**
  $x_{n+1} \leftarrow \arg\max_{\boldsymbol{x} \in S_n} \max_{\boldsymbol{z} \in \mathcal{X}} I_n\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$
  $y_{n+1} \leftarrow f(\boldsymbol{x}_{n+1}) + \nu$
  Update GP posterior with $(\boldsymbol{x}_{n+1}, y_{n+1})$

gain. We define such a measure using the binary variable $\Psi(\boldsymbol{x}) = \mathbb{I}_{\{f(\boldsymbol{x}) \geq 0\}}$, which is equal to one iff $f(\boldsymbol{x}) \geq 0$. Its entropy is given by

$$H_n\big[\Psi(\boldsymbol{z})\big] = -p_n^-(\boldsymbol{z}) \ln\big(p_n^-(\boldsymbol{z})\big) - \big(1 - p_n^-(\boldsymbol{z})\big) \ln\big(1 - p_n^-(\boldsymbol{z})\big) \tag{3}$$

where $p_n^-(\boldsymbol{z})$ is the probability of $\boldsymbol{z}$ being unsafe: $p_n^-(\boldsymbol{z}) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\Big(-\frac{1}{\sqrt{2}} \frac{\mu_n(\boldsymbol{z})}{\sigma_n(\boldsymbol{z})}\Big)$. The random variable $\Psi(\boldsymbol{z})$ has high-entropy when we are uncertain whether a parameter is safe or not; that is, its entropy decreases monotonically as $|\mu_n(\boldsymbol{z})|$ increases and the GP posterior moves away from the safety threshold, and also decreases monotonically as $\sigma_n(\boldsymbol{z})$ decreases and we become more certain about the constraint. This behaviour also implies that the entropy goes to zero as the confidence about the safety of $\boldsymbol{z}$ increases, as desired.

Given our definition of Ψ, we consider the mutual information $I\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$ between an observation y at a parameter $\boldsymbol{x}$ and the value of Ψ at another parameter $\boldsymbol{z}$. Since Ψ is the indicator function of the safe regions of the parameter space, the quantity $I\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$ measures how much information about the safety of $\boldsymbol{z}$ we gain by evaluating the safety constraint f at $\boldsymbol{x}$, averaged over all possible observed values y. This interpretation follows directly from the definition of mutual information: $I\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big) = H_n\big[\Psi(\boldsymbol{z})\big] - \mathbb{E}_y\Big[H_{n+1}\big[\Psi(\boldsymbol{z}) \big| \{\boldsymbol{x}, y\}\big]\Big]$, where $H_n[\Psi(\boldsymbol{z})]$ is the entropy of $\Psi(\boldsymbol{z})$ according to the GP posterior at iteration n, while $H_{n+1}\big[\Psi(\boldsymbol{z}) \big| \{\boldsymbol{x}, y\}\big]$ is its entropy at iteration $n+1$, conditioned on a measurement y at $\boldsymbol{x}$ at iteration n. Intuitively, $I\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$ is negligible whenever the confidence about the safety of $\boldsymbol{z}$ is high or, more generally, whenever an evaluation at $\boldsymbol{x}$ does not have the potential to substantially change our belief about the safety of $\boldsymbol{z}$. The mutual information is large whenever an evaluation at $\boldsymbol{x}$ on average causes the confidence about the safety of $\boldsymbol{z}$ to increase significantly. As an example, in Figure 1 we plot $I\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$ as a function of $\boldsymbol{x} \in S_n$ for a specific choice of $\boldsymbol{z}$ and for an RBF kernel. As one would expect, we see that the closer it gets to $\boldsymbol{z}$, the bigger the mutual information becomes, and that it vanishes in the neighborhood of previously evaluated parameters, where the posterior variance is negligible.

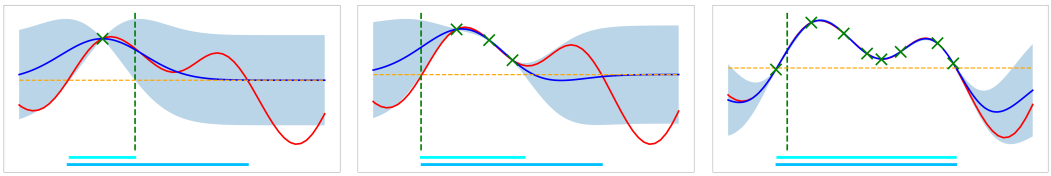
To compute $I\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$, we need to average (3) conditioned on an evaluation y over all possible values of y. However, the resulting integral is intractable given the expression of $H_n[\Psi(\boldsymbol{z})]$ in (3). In order to get a tractable result, we derive a close approximation of (3), which recovers almost exactly its true behavior (see Appendix B for details),

$$H_n\big[\Psi(\boldsymbol{z})\big] \approx \ln(2) \exp\left\{-\frac{1}{\pi \ln(2)} \left(\frac{\mu_n(\boldsymbol{z})}{\sigma_n(\boldsymbol{z})}\right)^2\right\}. \tag{4}$$

Since the posterior mean at $\boldsymbol{z}$ after an evaluation at $\boldsymbol{x}$ depends linearly on $\mu_n(\boldsymbol{x})$, and since the probability density of y depends exponentially on $-\mu_n^2(\boldsymbol{x})$, using (4) reduces the conditional entropy $\mathbb{E}_y\Big[H_{n+1}\big[\Psi(\boldsymbol{z}) \big| \{\boldsymbol{x}, y\}\big]\Big]$ to a Gaussian integral with the exact solution

$$\begin{aligned} &\mathbb{E}_y\Big[H_{n+1}\big[\Psi(\boldsymbol{z}) \big| \{\boldsymbol{x}, y\}\big]\Big] = \\ &\quad \ln(2) \sqrt{\frac{\sigma_\nu^2 + \sigma_n^2(\boldsymbol{x})(1 - \rho_n^2(\boldsymbol{x}, \boldsymbol{z}))}{\sigma_\nu^2 + \sigma_n^2(\boldsymbol{x})(1 + c_2 \rho_n^2(\boldsymbol{x}, \boldsymbol{z}))}} \exp\left\{-c_1 \frac{\mu_n^2(\boldsymbol{z})}{\sigma_n^2(\boldsymbol{z})} \frac{\sigma_\nu^2 + \sigma_n^2(\boldsymbol{x})}{\sigma_\nu^2 + \sigma_n^2(\boldsymbol{x})(1 + c_2 \rho_n^2(\boldsymbol{x}, \boldsymbol{z}))}\right\}, \end{aligned} \tag{5}$$

where $\rho_n(\boldsymbol{x}, \boldsymbol{z})$ is the linear correlation coefficient between $f(\boldsymbol{x})$ and $f(\boldsymbol{z})$, and with c_1 and c_2 given by $c_1 \coloneqq 1/\ln(2)\pi$ and $c_2 \coloneqq 2c_1 - 1$. In the following, $I_n\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$ refers to the mutual information based on (4) and (5) and all our theoretical results in Section 4 are proven for these expressions.



(a) Iteration 1 (b) Iteration 3 (c) Iteration 8

Figure 2: Example run of ISE at different iterations. The GP's posterior mean (blue line) and confidence interval $\mu_n \pm \beta_n \sigma_n$ (blue shaded) approximate the true safety constraint f (red line) based on selected data points (green crosses) and, together with the safety threshold at zero (orange line), identify the current safe set S_n (light blue bar, bottom). The vertical green line indicates the location of the next parameter $\boldsymbol{x}_{n+1}$ selected by (6). Initially, ISE evaluates parameters on the boundary of the safe set, but eventually also evaluates inside the safe set if that provides the most information. ISE quickly discover the largest reachable safe set (dark blue bar, bottom).

ISE acquisition function Now that we have defined a way to measure and compute the information gain about the safety of parameters, we can use it to design an exploration strategy that selects the next parameters to evaluate. The natural choice for such selection criterion is to select the parameter that maximizes the information gain; that is, we select $\boldsymbol{x}_{n+1}$ according to

$$\boldsymbol{x}_{n+1} \in \operatorname*{arg\,max}_{\boldsymbol{x} \in S_n} \max_{\boldsymbol{z} \in \mathcal{X}} I_n\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big), \tag{6}$$

where we jointly optimize over $\boldsymbol{x}$ in the safe set S_n and an unconstrained second parameter $\boldsymbol{z}$. Evaluating f at $\boldsymbol{x}_{n+1}$ according to (6) maximizes the information gained about the safety of some parameter $\boldsymbol{z} \in \mathcal{X}$, so that it allows us to efficiently learn about parameters that are not yet known to be safe. While $\boldsymbol{z}$ can lie in the whole domain, the parameters where we are the most uncertain about the safety constraint lie outside the safe set. By leaving $\boldsymbol{z}$ unconstrained, we show in our theoretical analysis in Section 4 that once we have learned about the safety of parameters outside the safe set, (6) resorts to learning about the constraint function also inside S_n. An overview of ISE can be found in Algorithm 1 and we show an example run of a one-dimensional illustration of the algorithm in Figure 2.

4 Theoretical Results

In this section, we study the expression for $I_n\big(\{\boldsymbol{x}, y\}; \Psi(\boldsymbol{z})\big)$ obtained using (4) and (5) and analyze the properties of the ISE exploration criterion (6). By construction of S_n in (2) and the assumptions on f in Section 2, we know that any parameter selected according to (6) is safe with high probability, see Appendix A for details. To show that we also learn about the safe set, we first need to define what it means to successfully explore starting from $\boldsymbol{x}_0$. The main challenge is that it is difficult to analyze how a GP generalizes based on noisy observations, so that it is difficult to define a notion of convergence that is not dependent on the specific run. In SafeOpt, they avoid this issue by expanding the safe set not based on the GP, but only using the Lipschitz constant L. Contrary to their approach, we depend on the GP to generalize from the safe set. In this case, the equivalent natural notion of convergence is provided by the the posterior variance. In particular, we say that at iteration n we have explored the safe set up to ε-accuracy if $\sigma_n^2(\boldsymbol{x}) \leq \varepsilon$ for all parameters $\boldsymbol{x}$ in S_n. In the following, we show that ISE asymptotically leads either to convergence or to indefinite expansion of the safe set. The proofs can be found in Appendix A.

Theorem 1. *Assume that $\boldsymbol{x}_{n+1}$ is chosen according to* (6)*, and that there exists $\hat{n}$ such that for all $n \geq \hat{n}$ $S_{n+1} \subseteq S_n$. Moreover, assume that for all $n \geq \hat{n}$, $|\mu_n(\boldsymbol{x})| \leq M$ for some $M > 0$ for all $\boldsymbol{x} \in S_n$. Then, for all $\varepsilon > 0$ there exists N_ε such that $\sigma_n^2(\boldsymbol{x}) \leq \varepsilon$ for every $\boldsymbol{x} \in S_n$ if $n \geq \hat{n} + N_\varepsilon$. The smallest of such N_ε is given by*

$$N_\varepsilon = \min\left\{ N \in \mathbb{N} : F^{-1}\left(\frac{C\gamma_N}{N}\right) \leq \varepsilon \right\}, \tag{7}$$

where $F(\eta) \coloneqq \ln(2)\exp\left\{-c_1 \frac{M^2}{\eta}\right\}\left[1-\sqrt{\frac{\sigma_\nu^2}{2c_1\eta+\sigma_\nu^2}}\right]$, $\gamma_N = \max_{D\subset\mathcal{X};|D|=N} I\big(\boldsymbol{f}(D);\boldsymbol{y}(D)\big)$ *is the maximum information capacity of the chosen kernel (Srinivas et al., 2010; Contal et al., 2014), and* $C = \ln(2)/\sigma_\nu^2 \ln\big(1+\sigma_\nu^{-2}\big)$.

Theorem 1 tells us that if at some point the set of safe parameters S_n stops expanding, then the posterior variance over the safe set vanishes eventually. The intuition behind Theorem 1 is that if there were a parameter $\boldsymbol{x}$ in the safe set whose posterior mean remained finite and whose posterior variance remained bounded from below, then an evaluation of f at such $\boldsymbol{x}$ would yield a non negligible average information gain about the safety of $\boldsymbol{x}$, so that, since $\boldsymbol{x}$ is in the safe set, at some point ISE will be forced to choose to evaluate $\boldsymbol{x}$, reducing its posterior variance. This intuition also offers an insight about how to interpret the result: although Theorem 1 guarantees that ISE asymptotically explores the safe set up to an arbitrary ε-accuracy, in practice it will first focus on reducing the uncertainty in areas of the safe set that are most informative about parameters whose classification is still uncertain (e.g. areas close to the boundary of the safe set), and only eventually turns to learning about the inside of the safe set. This behaviour is what ultimately leads to the posterior variance to decay over the whole S_n. We can therefore read Theorem 1 as an exploration guarantee for ISE, as it cannot happen that it gets stuck in some subset of S_n if that causes σ_n^2 to stay finite over the safe set.

Theorem 1 requires a bound on the GP posterior mean function, which is always satisfied with high probability based on our assumptions about f. Specifically, we have that $|\mu_n(\boldsymbol{x})| \leq 2\beta_n$ with probability of at least $1-\delta$, for all n (see Appendix A for details). Therefore, it does not represent an additional restrictive assumption for f. Finally, we also note that the the constant N_ε defined by (7) always exists since F is monotonically increasing, as long as γ_N grows sublinearly in N. Srinivas et al. (2010) prove that this is the case for commonly-used kernel and, more generally, it is a prerequisite for data-efficient learning with GP models.

5 Discussion and Limitations

ISE drives exploration of the parameter space by selecting the parameters to evaluate according to (6). An alternative but conceptually similar approach to this criterion would be to consider the parameter that yields the biggest information gain *on average* over the domain, i.e., substituting the inner $\max$ in (6) with an average over $\mathcal{X}$. The resulting integral, however, is intractable and would require further approximations. Moreover, the parameter found by solving (6) will also yield a high average information gain over the domain, due to the regularity of all involved objects.

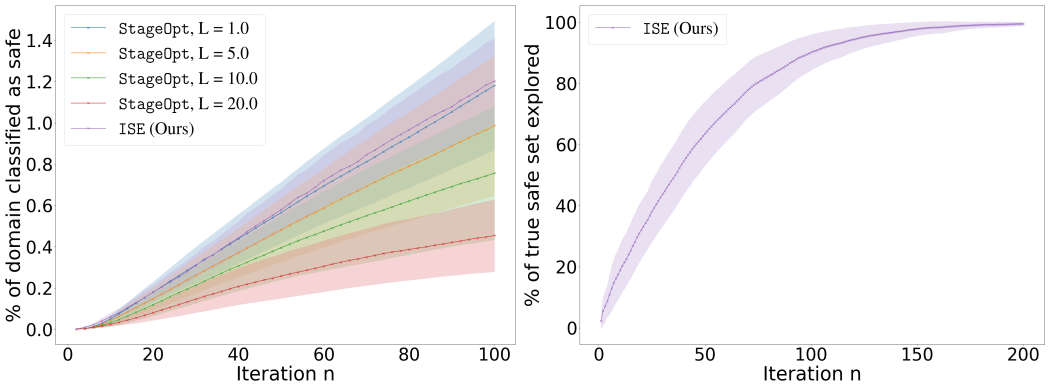
Being able to work in a continuous domain, ISE scales better than algorithms requiring a discrete parameter space. However, as noted in Section 4, finding $\boldsymbol{x}_{n+1}$ as in (6) means to solve a non convex optimization problem with twice the dimension of the the parameter space, which can also become a computationally challenging problem as the dimension grows. In a high-dimensional setting, we follow LINEBO by Kirschner et al. (2019), which at each iteration selects a random one-dimensional subspace to which it restricts the optimization of the acquisition function.

Lastly, we reiterate that the theoretical safety guarantees offered by ISE are derived under the assumption that f is a bounded norm element of the RKHS space associated with the GP's kernel. In applications, therefore, the choice of the kernel function becomes even more crucial when safety is an issue. For details on how to construct and choose kernels see (Garnett, 2022).

6 Experiments

In this section we empirically evaluate ISE. Additional details about the experiments and setup can be found in Appendix C.

GP samples For the first part of the experiments, we evaluate ISE on samples from a GP. This allows us to test ISE under the assumptions of the theory and we compare its performance to that of the exploration part of STAGEOPT (Sui et al., 2018). STAGEOPT is a modified version of SAFEOPT, in which the exploration and optimization parts are performed separately: first the SAFEOPT exploration exploration strategy is used to expand the safe set as much as possible, then the objective function is optimized within the discovered safe set. We select 100 samples from a two-dimensional GP with RBF kernel, defined in $[-2.5, 2.5] \times [-2.5, 2.5]$ and run ISE and STAGEOPT for 100 iterations for



(a) Comparison with multiple STAGEOPT instances. (b) Convergence to maximally reachable safe set.

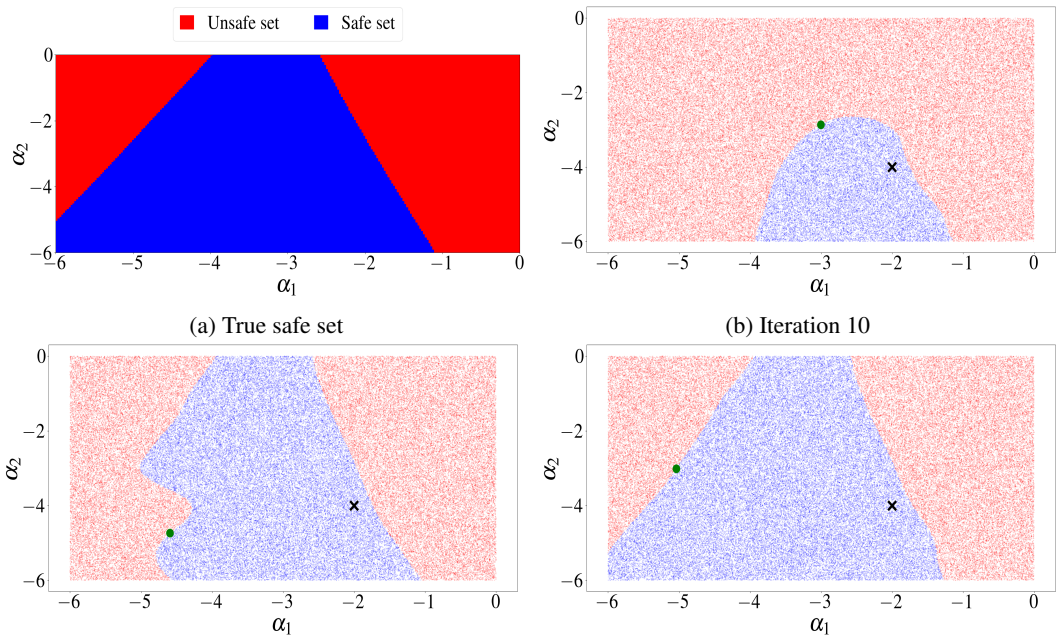
Figure 3: Average expansion of safe set over 100 two-dimensional GP samples. (a) The percentage of the domain classified as safe is plotted as a function of n. The expansion promoted by ISE is compared with the one promoted by STAGEOPT for multiple values of the Lipshcitz constant L. We can see that ISE obtains an higher sample efficiency than the best instance of STAGEOPT, whose performance, on the other hand, is heavily affected by the choice of L. (b) The percentage of the maximally reachable safe set classified as safe is plotted as function of n: in agreement with the theory, we can see that ISE eventually leads to discover the full reachable safe set.

each sample. As STAGEOPT requires a discretization of the domain, we use this discretization to compare the sample efficiency of the two methods, by computing, at each iteration, what percentage of the discretized domain is classified as safe. The results of this comparison can be seen in Figure 3a. For the STAGEOPT run, the kernel metric has been used to compute the set of potential expanders, for different values of the Lipschitz constant L. From Figure 3a we see not only that ISE performs as well or better than all tested instances of STAGEOPT, but also how the choice of L affects the performance of the latter. This plot makes also evident how crucial the choice of the Lipschitz constant is for STAGEOPT and SAFEOPT-like algorithms in general. Moreover, we use GP samples also to test our theoretical results, according to which ISE eventually leads to the discovery of the maximally reachable safe set. This can be seen in Figure 3b, where we plot the average percentage of the maximally reachable safe set classified as safe by ISE: we can see that eventually the whole set is explored and classified as safe.

OpenAI Gym Control After investigating the performance of ISE under the hypothesis of the theory, we apply it to two popular classic control tasks from the OpenAI Gym framework (Brockman et al., 2016), with the goal of finding the set of parameters of a controller that satisfy some safety constraint. In particular we considered linear controllers for the inverted pendulum and cart pole tasks.

For the inverted pendulum task, the linear controller is given by $u_t = \alpha_1 \theta_t + \alpha_2 \dot{\theta}_t$, where u_t is the control signal at time t, while θ_t and $\dot{\theta}_t$ are, respectively, the angular position and the angular velocity of the pendulum. Starting from a position close to the upright equilibrium, the considered controller's task is the stabilization of the pendulum, subject to a safety constraint on the maximum velocity in one episode. For some given initial controller configuration $\boldsymbol{\alpha}_0 := (\alpha_1^0, \alpha_2^0)$, we want to explore the controller's parameter space avoiding configurations that lead the pendulum to swing with a too high velocity. We apply ISE to explore the $\boldsymbol{\alpha}$-space with $\boldsymbol{x}_0 = \boldsymbol{\alpha}_0$ and the safety constraint being the maximum angular velocity reached by the pendulum in an episode of fixed length. In this case the safety threshold is not at zero, but some finite value $\dot{\theta}_M$, and the safe parameters are those for which the maximum velocity is below $\dot{\theta}_M$. The formalism developed in the previous sections can be easily applied to this scenario if we consider $f(\boldsymbol{\alpha}) = -(\max_t \dot{\theta}_t(\boldsymbol{\alpha}) - \dot{\theta}_M)$. in Figure 4a we show the true safe for this problem, while in Figure 4b to Figure 4d one can see ISE in action and the safe set it discovers.

The setting for the cart pole task is similar, but the parameter space in this case has three dimensions. The controller we consider is given by $u_t = \alpha_1 \theta_t + \alpha_2 \dot{\theta}_t + \alpha_3 \dot{s}_t$, where θ_t and $\dot{\theta}_t$ are, respectively,



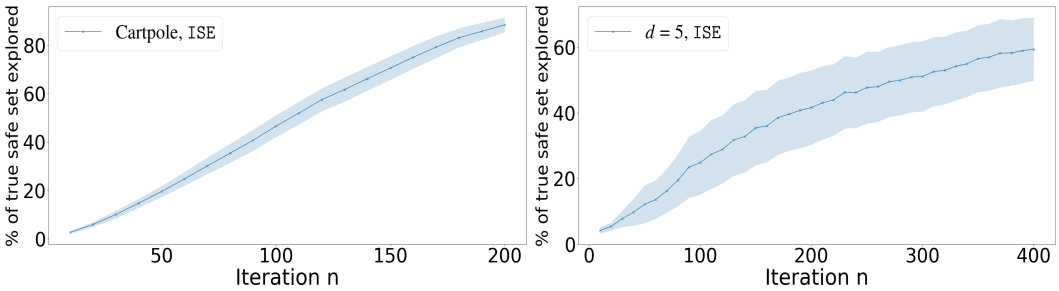
(a) True safe set

(b) Iteration 10

(c) Iteration 30

(d) Iteration 50

Figure 4: Safe exploration of the linear controller's parameter space in the inverted pendulum experiment. In (a) we see the true safe set, while in (b-d) we see the safe set as identified by ISE after 10, 30 and 50 iterations. The blue region is the safe set, while the red one is the unsafe region. The point marked by the green dot is the one the algorithm is going to evaluate next according to (6), while the black cross is the initial safe seed $\boldsymbol{\alpha}_0$.

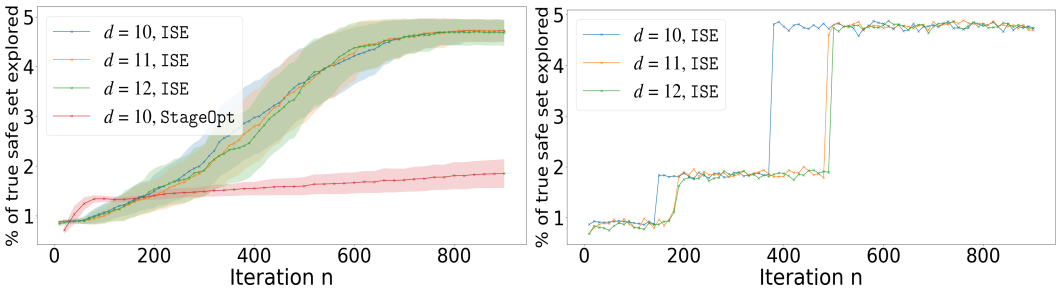


(a) Exploration of the safe set for the cart pole task.

(b) Expansion of the safe set in dimension five.

Figure 5: In (a) we plot the percentage of the true safe set of the cart pole task classified as safe by ISE for the first 200 iterations, while in (b) we see the expansion of the five-dimensional safe set promoted by ISE, for the safety constraint f used in the high dimensional experiments. In both cases, we can observe how ISE drives the exploration, leading to an always greater fraction of parameters to be classified as safe.

the angular position and angular velocity of the pole at time t, while $\dot{s}_t$ is the cart's velocity. In this case, we set the initial state as one having zero angular and linear velocity and with the pole close to the vertical position, with the controller's goal being of keeping the pole stable in the upright position. A combination of the three parameters α_1, α_2 and α_3 is considered safe if the angle of the pole does not exceed a given threshold. Again, we can easily cast this safety constraint in terms of the formalism developed in the paper: $f(\boldsymbol{\alpha}) = -(\max_t \theta_t(\boldsymbol{\alpha}) - \theta_M)$, where θ_M is the maximum allowed angle. Figure 5a shows the expansion of the cart pole $\boldsymbol{\alpha}$ space promoted by ISE.



(a) Average safe optimum over multiple runs.

(b) Safe optimum for a single run.

Figure 6: Example of high dimensional exploration, for d =10, 11, 12. To be able to estimate how much of the domain is classified as safe, after each ten iterations, we perform safe UCB within the current safe set and plot the found optimum against the number of iterations. In (a) we plot the average over 100 runs, while in (b) a single example run is shown. The safe seed $\boldsymbol{x}_0$ is chosen close to the border of the real safe set, in the vicinity of the lowest optimum. We can see that this adapted version of ISE promotes expansion of the safe set, until also the region including the location of the highest optimum - the furthest from x_0 - is classified as safe. In (a) we compare the results with those obtained with STAGEOPT (in conjunction with LINEBO) and observe that in such high dimensional setting ISE achieves a better sample efficiency.

High dimensional domains Many interesting applications have a high dimensional parameter space. While SAFEOPT-like methods are difficult to apply already with dimension > 3 due to the discretization of the domain, ISE can perform well also in four or five dimensions. To see this, we applied ISE to the constraint function $f(\boldsymbol{x}) = e^{-\boldsymbol{x}^2} + 2e^{-(\boldsymbol{x}-\boldsymbol{x}_1)^2} + 5e^{-(\boldsymbol{x}-\boldsymbol{x}_2)^2} - 0.2$, which has three local optima of different heights at 0, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, with $\|\boldsymbol{x}_1\| < \|\boldsymbol{x}_2\|$. Starting from a safe seed close to the border of the safe set in the vicinity of the origin, ISE was able to explore the safe set until also $\boldsymbol{x}_2$ was classified as safe. Figure 6b shows the percentage of the true safe set discovered by ISE as function of the iteration. As discussed in Section 5, for even higher dimensions, we can follow a similar approach to LINEBO, limiting the optimization to a randomly selected one-dimensional subspace. We applied this method to the same constraint function used for the five dimensional case and found that indeed we were able to explore the domain until the discovery of the location of the global optimum. The corresponding results can be seen in Figure 6.

7 Conclusion and Societal Impact

We have introduced Information-Theoretic Safe Exploration (ISE), a novel approach to safely explore a space in a sequential decision task where the safety constraint is *a priori* unknown. ISE efficiently and safely explores by evaluating only parameters that are safe with high probability and by choosing those parameters that yield the greatest information gain about the safety of other parameters. We theoretically analyzed ISE and showed that it leads to arbitrary reduction of the uncertainty in the largest reachable safe set containing the starting parameter. Our experiments support these theoretical results and demonstrate an increased sample efficiency and scalability of ISE compared to SAFEOPT-based approaches.

In many safety sensitive applications the shape of the safety constraints is unknown, so that an important prerequisite for any kind of process is to identify what parameters are safe to evaluate. By providing a principled way to do this, the contributions of this paper allow to deal with safety in a broad range of applications, which can favor the usage of ML approaches also in safety sensitive settings. On the other hand, misuse of the proposed method cannot be prevented in general.

References

Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, 2021.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 2012.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning (ICML)*, 2017.

Emile Contal, Vianney Perchet, and Nicolas Vayatis. Gaussian process optimization with mutual information. In *International Conference on International Conference on Machine Learning (ICML)*, 2014.

Rikky R.P.R. Duivenvoorden, Felix Berkenkamp, Nicolas Carion, Andreas Krause, and Angela P. Schoellig. Constrained Bayesian Optimization with Particle Swarms for Safe Adaptive Controller Tuning. *IFAC-PapersOnLine*, 2017.

Gabriel Dulac-Arnold, Daniel J. Mankowitz, and Todd Hester. Challenges of real-world reinforcement learning. In *International Conference on International Conference on Machine Learning (ICML)*, 2019.

Lukas Fröhlich, Edgar Klenske, Julia Vinogradska, Christian Daniel, and Melanie Zeilinger. Noisy-input entropy search for efficient robust bayesian optimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2022. in preparation.

Michael A. Gelbart, Jasper Snoek, and Ryan P. Adams. Bayesian optimization with unknown constraints. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.

Alkis Gotovos, Nathalie Casati, Gregory Hitz, and Andreas Krause. Active learning for level set estimation. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.

Alexander Hans, Daniel Schneegass, Anton Schäfer, and Steffen Udluft. Safe exploration for reinforcement learning. In *European Symposium on Artificial Neural Networks (ESANN)*, 2008.

Philipp Hennig and Christian J. Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, 2012.

José Miguel Henrández-Lobato, Matthew W. Hoffman, and Zoubin Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *International Conference on Neural Information Processing Systems (NIPS)*, 2014.

Johannes Kirschner, Mojmir Mutný, Nicole Hiller, Rasmus Ischebeck, and Andreas Krause. Adaptive and safe bayesian optimization in high dimensions via one-dimensional subspaces. In *International Conference for Machine Learning (ICML)*, 2019.

Teodor Mihai Moldovan and Pieter Abbeel. Safe exploration in markov decision processes. In *International Coference on International Conference on Machine Learning (ICML)*, 2012.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, Cambridge, MA, USA, 2006.

Mark Schillinger, Benjamin Hartmann, Patric Skalecki, Mona Meister, Duy Nguyen-Tuong, and Oliver Nelles. Safe active learning and safe bayesian optimization for tuning a pi-controller. *IFAC-PapersOnLine*, 2017.

Bernhard Schölkopf and Alexander J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge, MA, USA, 2002.

Jens Schreiter, Duy Nguyen-Tuong, Mona Eberts, Bastian Bischoff, Heiner Markert, and Marc Toussaint. Safe exploration for active learning with gaussian processes. In *European Conference on Machine Learning and Knowledge Discovery in Databases (ECMLPKDD)*, 2015.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 2016.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *International Conference on Machine Learning (ICML)*, 2010.

Yanan Sui, Alkis Gotovos, Joel W. Burdick, and Andreas Krause. Safe exploration for optimization with Gaussian processes. In *International Conference on Machine Learning (ICML)*, 2015.

Yanan Sui, Vincent Zhuang, Joel Burdick, and Yisong Yue. Stagewise safe Bayesian optimization with Gaussian processes. In *International Conference on Machine Learning (ICML)*, 2018.

Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. Safe exploration for interactive machine learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient bayesian optimization. In *International Conference on Machine Learning (ICML)*, 2017.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Limitations are discussed in Section 5
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section 7
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 4.
   (b) Did you include complete proofs of all theoretical results? [Yes] See Appendix A.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] We will include the url to the code with the camera-ready version of the paper. Experiment details to reproduce our results can be found in Appendix C.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Appendix C.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] They are in the respective plots.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Appendix C.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 6 and Appendix C
   (b) Did you mention the license of the assets? [Yes] See Appendix C
   (c) Did you include any new assets either in the supplemental material or as a URL? [No] No new assets have been used nor introduced.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]