

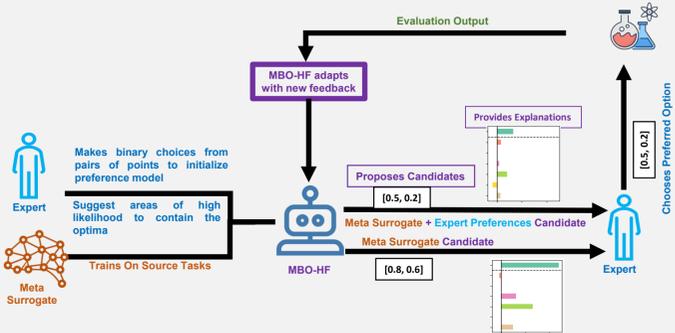
**Abstract**

Due to the high costs and limited experimentation opportunities of some scientific applications, efficient optimization techniques incorporating expert knowledge are crucial.

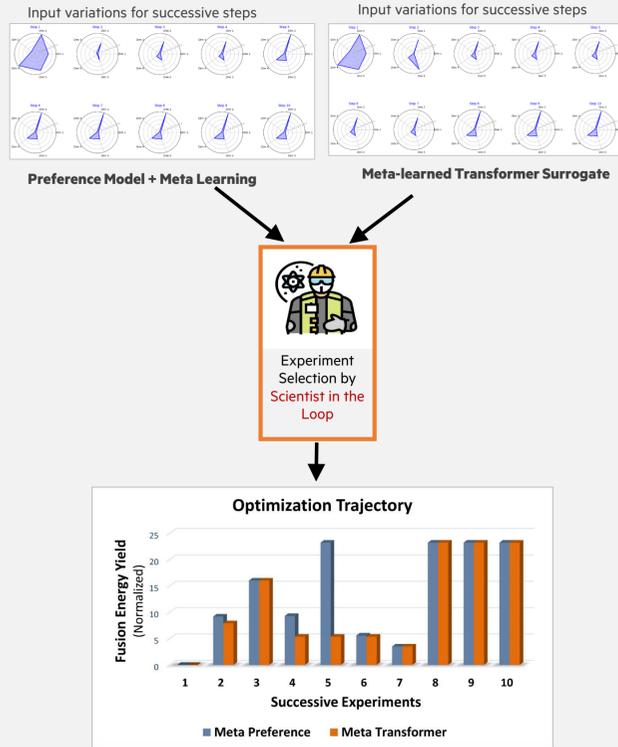
Bayesian Optimization (BO) has shown promise in optimizing complex systems, but trust and sample efficiency remain challenges.

To address the need for trust and sample efficiency, we introduce Meta Bayesian Optimization with Human Feedback (MBO-HF), a framework that:

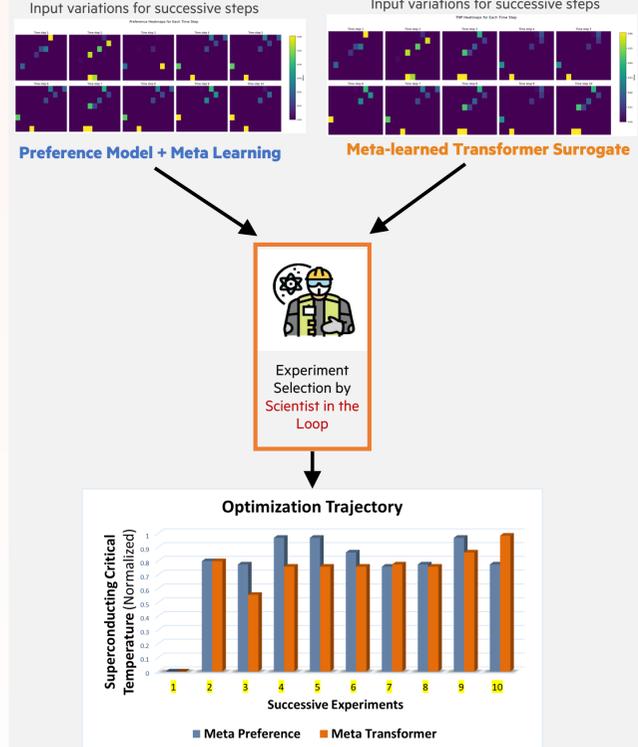
- Integrates expert knowledge
- Provides explainable insights
- Includes the scientist in the decision-making loop.



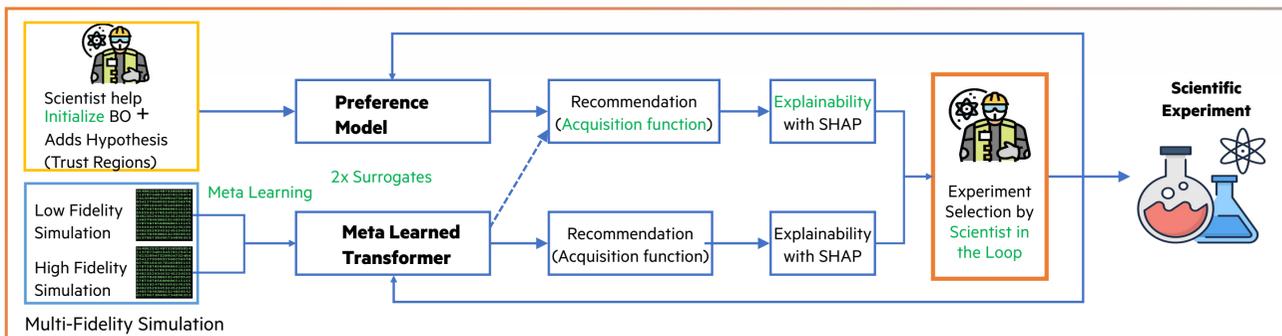
**MBO-HF in Laser Pulse Shaping for Inertial Confinement Fusion**



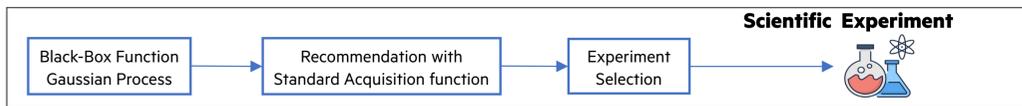
**MBO-HF in Superconducting Temperature Maximization**



**Proposed Method**



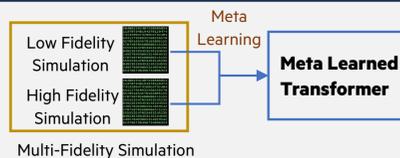
**Explainable Scientist in the Loop Optimization with Meta Learning from Multi-Fidelity Simulation**



Standard Bayesian Optimization – how it has been done

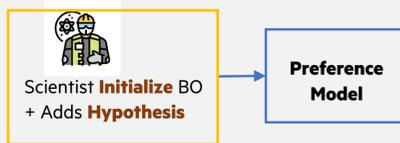
**Meta Learning:**

MBO-HF introduces meta-learning-based optimization leveraging Transformer Neural Processes (TNPs). By training meta-surrogates on source tasks, MBO-HF incorporates prior knowledge into the optimization of target functions.



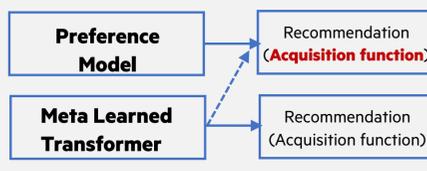
**Preference Learning:**

MBO-HF makes use of a binary preference model which is trained using expert judgments on pairs of candidate points. Moreover, Hypothesis-driven sampling is employed to focus candidate selection within promising subregions.



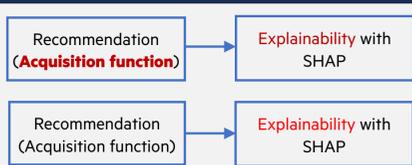
**Acquisition Function:**

MBO-HF uses a novel Expected Improvement (EI)-based acquisition function that combines the meta-surrogate and preference model, balancing exploration and exploitation. This formulation adjusts the influence of expert preferences dynamically, ensuring adaptability over time.



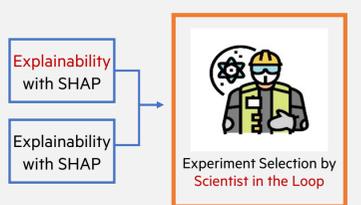
**Explainability:**

MBO-HF employs Shapley values to provide insights into the acquisition function's decisions, offering a breakdown of feature contributions to the optimization.



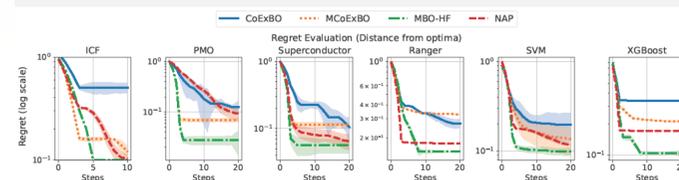
**Optimization Process:**

MBO-HF presents candidates to the expert, and the expert selects the candidate they prefer to evaluate. The surrogates are updated with this new evaluation. This process is repeated until we run out of experimental budget, or the optima has been reached.



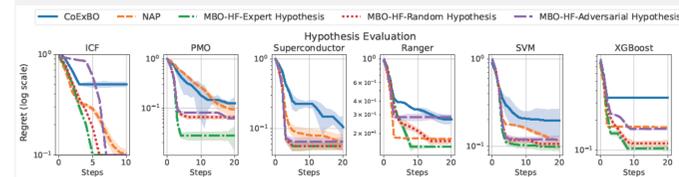
**Results**

**MBO-HF Regret Minimization**



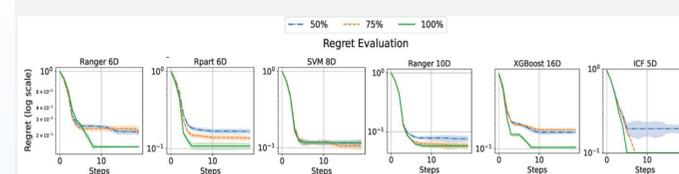
MBO-HF outperforms baselines on all 6 benchmarks, lower the better. Shaded areas: +/- 1SD. Regrets are normalized.

**Hypothesis Evaluation**



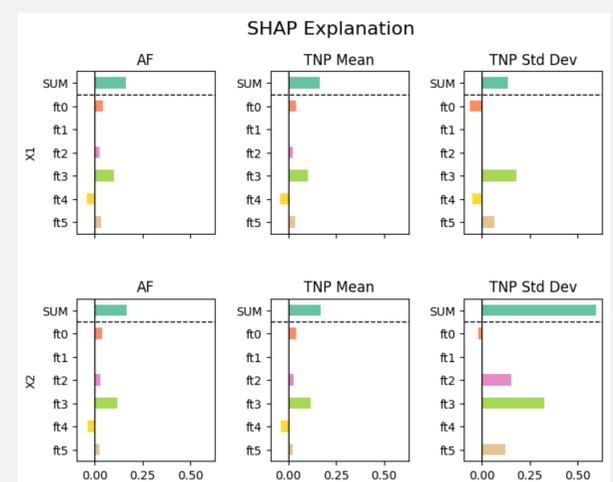
Comparison of different hypothesis: "Good" best-case, "Bad" worst-case, and Random sampling in preference learning initialization.

**Accuracy Evaluation**



Results of the comparison between different expert's accuracy percentages.

**SHAP Explanations**



Feature attribution using for the two candidate points proposed by MBO-HF. Bars represents Shapley value of the feature w.r.t. the BO metrics (AF, Mean, SD).

This material is based upon work supported by the Department of Energy National under Award No. DE-SC0024408, Hewlett Packard Enterprise and the University of Rochester.