ZEROSHOTOPT: Towards Zero-Shot Pretrained Models for Efficient Black-Box Optimization

Jamison Meindl[1*] **Yunsheng Tian**[1] **Tony Cui**[1] **Veronika Thost**[2] **Zhang-Wei Hong**[2]

Johannes Dürholt[3] **Jie Chen**[2] **Wojciech Matusik**[1]

Mina Konaković Luković[1]

[1]MIT [2]MIT-IBM Watson AI Lab [3]Evonik Operations GmbH

Abstract

Global optimization of expensive, derivative-free black-box functions requires extreme sample efficiency. While Bayesian optimization (BO) is the current state-of-the-art, its performance hinges on surrogate and acquisition function hyper-parameters that are often hand-tuned and fail to generalize across problem landscapes. We present ZEROSHOTOPT, a general-purpose, pretrained model for continuous black-box optimization tasks ranging from 2 D to 20 D. Our approach leverages offline reinforcement learning on large-scale optimization trajectories collected from 12 BO variants. To scale pretraining, we generate millions of synthetic Gaussian process-based functions with diverse landscapes, enabling the model to learn transferable optimization policies. As a result, ZEROSHOTOPT achieves robust zero-shot generalization on a wide array of unseen benchmarks, matching or surpassing the sample efficiency of leading global optimizers, including BO, while also offering a reusable foundation for future extensions and improvements. Our open-source code, dataset, and model are available at https://github.com/jamisonmeindl/zeroshotopt.

1 Introduction

Black-box optimization under tight evaluation budgets is pivotal in many scientific and engineering settings. Since derivatives are often unavailable, classical gradient-based solvers such as Newton's method or conjugate-gradient are not applicable. Practitioners therefore turn to gradient-free heuristics such as genetic algorithms, simulated annealing, and evolutionary strategies, which often require thousands of evaluations to converge [1–3]. When each experiment or simulation is slow or costly, such sample counts become prohibitive. Therefore, we focus on improving continuous black-box optimization under strict evaluation constraints.

Bayesian optimization (BO) [4] reduces this burden by fitting a probabilistic surrogate (typically a Gaussian process) and selecting queries via a heuristic acquisition function that balances exploration and exploitation. BO has driven progress in materials discovery [5], molecular design [6], clinical prognosis [7], and hyper-parameter tuning [8]. However, BO's success hinges on hand-chosen kernels, acquisition functions, and hyper-parameters whose optimal settings vary across landscapes and are difficult to tune without expert insight or extra evaluations.

*Correspondence to: Jamison Meindl <jmeindl@mit.edu>.

ScaleOPT: GPU-Accelerated and Scalable Optimization Workshop at NeurIPS 2025.

Consider a researcher maximizing the efficacy of a drug formulation. Each candidate formulation must pass a clinical assay, capping the budget at *fifty* trials. The formulation parameters span dosage, delivery rate, and co-administered compounds, which are continuous parameters with unknown interactions. Standard heuristics exhaust the budget before converging, while BO would succeed only if its kernel and acquisition hyperparameters were tuned. Therefore, there is a need for a low-budget general-purpose global optimizer that can perform well without tuning.

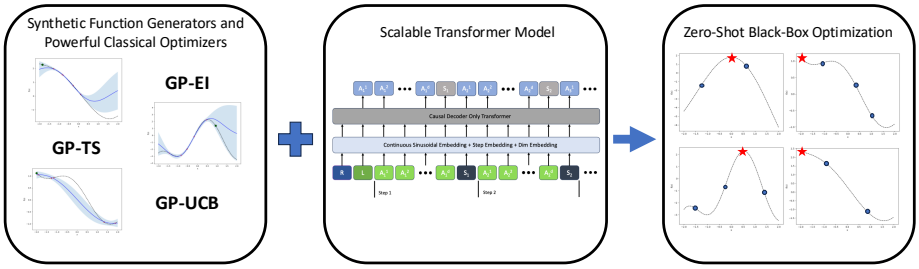


Figure 1: Overall Diagram of ZEROSHOTOPT. We use a combination of diverse synthetic functions and classical optimizers to train a pretrained transformer model for efficient black-box optimization.

The large amounts of data and increasing computing power available today have led to the development of pretrained transformer-based models, used in a variety of data-driven decision making scenarios, such as machine translation [9], regression [10], and robotic control [11]. Recent approaches have begun exploring using pretrained models for optimization, including both transformer architectures and offline pretraining [12–14]. However, the full potential of pretrained models in optimization remains untapped, in part due to the limited availability of massive high-quality optimization datasets [15]. Furthermore, demonstrating robust zero-shot generalization on completely unseen problems within low evaluation budgets remains an open challenge, as most pretrained approaches require training on similar data to which the model is tested. Consequently, traditional optimization, such as Bayesian optimization, continues to be the preferred choice in practical applications.

In this paper, we introduce ZEROSHOTOPT, a *general-purpose pretrained optimizer* for continuous black-box problems up to 20 D. We train a 200 M parameter transformer-based model using offline reinforcement learning on a diverse dataset of trajectories generated using various BO variants on synthetic GP functions. As a result, ZEROSHOTOPT demonstrates a robust understanding of optimization dynamics and provides a valuable framework for further optimization improvements.

In summary, our contributions are as follows.

1. We develop a *synthetic function generator* based on Gaussian processes and collect *large-scale optimization trajectories* from 12 BO variants as pretraining data. We provide this as an open-source dataset, totaling ~ 20 million synthetic trajectories, providing a valuable large dataset of expert optimization trajectories.
2. We design and train a single scalable transformer-based model that operates across dimensions from 2 D to 20 D and up to 50 total evaluations.
3. Our evaluation demonstrates that ZEROSHOTOPT achieves *strong zero-shot generalization* performance across a range of unseen global optimization benchmarks, matching the performance of existing state-of-the-art BO methods, while providing the opportunity for further extensions to our method.

We have made our open-source code, complete training and evaluation dataset, and model available at https://github.com/jamisonmeindl/zeroshotopt.

2 Related work

Bayesian optimization (BO) BO is a powerful global optimization technique due to its ability to efficiently handle expensive black-box functions by balancing exploration and exploitation with carefully designed acquisition functions, such as Expected Improvement (EI), LogEI [16], Upper Confidence Bound (UCB) and Vizier [17, 18]. Designed with different principles of trading-off exploration and exploitation, they suit different types of optimization problems. BO has been

made accessible by a collection of open-source libraries and software including Spearmint [19], BoTorch [20], AutoOED [21], SMAC3 [22], HEBO [23], Openbox [24], etc. However, it remains a challenge for users to determine the most appropriate configuration for their specific problem setting, as the selection is often heuristic and depends on the problem landscape.

Offline reinforcement learning Decision Transformer [11], which models offline RL as sequence learning, is a primary inspiration for our model architecture. We utilize their idea of conditional reward within transformer-based sequence model as the foundation of our model. Follow-up works on Decision Transformer [25, 26] are possible future avenues for improvement. Other insights from offline reinforcement learning [27] could also be used to build upon our methodology.

Pretrained optimization methods Causal transformers have been used for learning to optimize, including BONET [13], OptFormer [12], and RIBBO [28]. These methods train transformer-based models on a variety of synthetic and real-world data and show strong test performance on their test suites. However, these methods primarily rely on specific training sets with similar distributions to their test sets. Further comparisons to ZEROSHOTOPT are available in Appendix B. The goal of ZEROSHOTOPT is to provide a novel approach to the challenge of current black-box optimization methods, which suffer from requiring hyperparameter tuning or domain specific adaptation.

Therefore, we propose ZEROSHOTOPT, a general-purpose, zero-shot, transformer-based optimizer verified on continuous black-box optimization problems ranging from 2 D–20 D. Unlike BONET, OptFormer and RIBBO, which each excel only when the test distribution resembles the data they were trained on, ZEROSHOTOPT is pretrained once on ~ 20 million synthetic trajectories and then can be deployed unchanged on out-of-distribution tests. It retains competitive performance against strong BO baselines on *unseen* synthetic and real-world test suites, while also demonstrating the ability to be fine-tuned to specific domains. Therefore, ZEROSHOTOPT makes progress towards closing the outstanding gap left by prior transformer-based optimizers by providing a robust out-of-distribution optimization method that provides a base for further improvement.

3 Approach: ZEROSHOTOPT

Problem Statement: For generic black-box global optimization problems, the goal is to solve $\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^d$ is a vector of decision variables, $\mathcal{X} \subseteq \mathbb{R}^d$ represents the feasible search space, $f : \mathcal{X} \to \mathbb{R}$ is the black-box objective function, which is typically expensive to evaluate and lacks gradient information, and $\mathbf{x}^*$ is the global minimum of $f(\mathbf{x})$.

Goal: We aim to develop a pretrained model that serves as a "plug-and-play" optimizer, capable of outperforming traditional black-box optimizers without hyperparameter tuning.

Challenge: Typically, training a model of this nature requires a large dataset of examples as most approaches are based on supervised learning [29–31]. Therefore, we need expert demonstrations illustrating how to choose evaluation points that minimize functions in low-budget scenarios. Unfortunately, no publicly available dataset provides such demonstrations.

3.1 Learning to optimize via offline RL

Optimization as sequential decision-making: We frame black-box optimization as a sequential decision-making problem, where the optimizer acts as an agent interacting with the black-box function f. The optimizer observes a set of initial samples: $\{(\mathbf{x}_1, f(\mathbf{x}_1)), \ldots, (\mathbf{x}_m, f(\mathbf{x}_m))\}$, where each $\mathbf{x}_i \in \mathcal{X}$ is randomly sampled from the search space. At each step t, the optimizer observes all previously evaluated points and their function values, $\{\mathbf{x}_1, f(\mathbf{x}_1), \ldots, \mathbf{x}_{m+t}, f(\mathbf{x}_{m+t})\}$, selects the next evaluation point $\mathbf{x}_{m+t+1}$ based on its policy π, receives the function value $f(\mathbf{x}_{m+t+1})$, and repeats this process until the number of steps t hits the user-defined limit. The agent's objective is to find a point $\mathbf{x}$ that minimizes the objective value $f(\mathbf{x})$ over its interactions with the environment.

Implementation: We adapt Decision Transformer (DT) [11], a scalable, transformer-based, offline RL algorithm, to learn an optimizer policy from static datasets. DT is a transformer-based model that takes the history of states, actions, and trajectory quality as input and predicts actions to achieve trajectories of the specified quality. We define trajectory quality based on normalized regret R relative to the set of optimization methods run on the same function.

Our adapted DT takes as input past query points $\mathbf{x}_i$ and their function values $f(\mathbf{x}_i)$, along with trajectory regret and length, and predicts the next query point required to generate a trajectory with the given regret and length. Like supervised learning, DT learns from demonstrations in static datasets but differs by conditioning on trajectory regret and length. This is crucial for allowing the model to distinguish between effective and ineffective optimization trajectories. At inference time, we specify zero as the desired regret R to initiate generation, encouraging the model to act as the best methods do. We also specify parameter L as the evaluation budget for the method.

3.2 Synthetic data generation

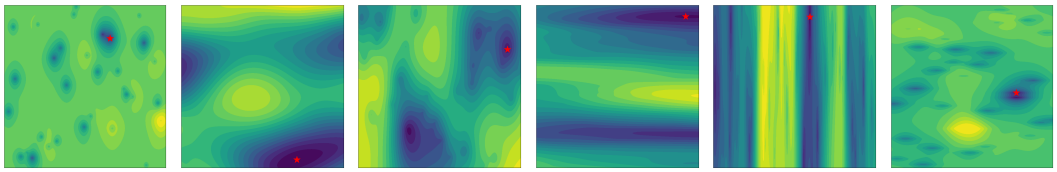


Figure 2: 2D synthetic functions generated by our GP-based function generator with various kernels and parameters. The red star is the global minimum and darker color signifies lower function value.

Gaussian process as function generators: Inspired by Chen et al. [32], we use Gaussian Processes (GPs) as flexible function generators. The intuition here is that GP-based BO methods perform well over a wide variety of function spaces. By basing our training data on GP functions as well, we hope to learn a policy that also adapts well to other function spaces. While GPs may not perfectly represent real-world function spaces, our empirical results show they provide a good basis for learning a policy that works on a wide range of function spaces. To introduce variability, we randomize the kernel type, length scale, and other initialization parameters. Further work in generating synthetic functions could improve the function diversity and therefore results. Figure 2 illustrates example generated functions.

Trajectory generation: For each of the 1,600,000 synthetic functions generated ranging from 2 D to 20 D, we run BO with 12 kernel–acquisition variants. These include the Expected Improvement (EI), LogEI [16], Upper Confidence Bound (UCB), Joint Entropy Search (JES), Max-value Entropy Search (MES), and Thompson Sampling (TS) acquisition functions with RBF and Matern kernels. Each model is initialized with 10 random points before iteratively fitting the GP and selecting new points using the acquisition function. We generate 12 trajectories per environment, each consisting of 10 initial samples followed by 40 optimization steps, for a total of 50 evaluations.

3.3 Model architecture

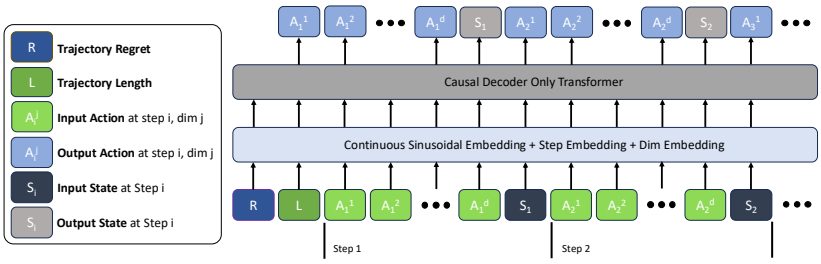


Figure 3: Illustration of the ZEROSHOTOPT architecture. The model embeds continuous inputs using a fixed sinusoidal embedding and 2 learned positional embeddings. The main model is a causal decoder only transformer, trained on loss on the binned action and state space.

Our model takes 4 different types of token inputs: regret, length, action (proposed points), and state (function values). The actions are split into individual tokens for each dimension. These are provided as normalized continuous values to a fixed sinusoidal embedding. We add 2 learned positional embeddings, an embedding of the dimension and an embedding of the step. These embeddings are passed on to a causal decoder-only transformer model. This is trained using cross-entropy loss on the binned action and state values, where the normalized inputs are split into 2,000 linearly spaced bins. The main architecture of ZEROSHOTOPT is illustrated in Figure 3.

3.4 Training and inference

We train a single model with 200 million parameters on synthetic GP data ranging from $2\,\mathrm{D}$ to $20\,\mathrm{D}$, with step counts of 10, 20, 30, or 40 model steps. We train for a total of 250,000 steps with a batch size of 1,024 trajectories. The model is trained with cross-entropy loss on the binned action and state values. We also test the ability to fine-tune our base model to a new domain, specifically the Hyperparameter Optimization Benchmark (HPO-B) dataset [33]. We split the HPO-B dataset into the specified train and test datasets and generate trajectories on 200 environments from each of the 13 HPO-B classes. We use this small dataset to fine-tune the model to the new domain.

At inference time, we begin by sampling 10 random steps and evaluating these points, as done during dataset generation. We feed this data into the model and iteratively generate 4 proposed actions with the model, along with a predicted state distribution for each action. Similar to OptFormer [12], we use an expected improvement acquisition function to select an action for that iteration. We then evaluate the action, and feed the new information into the model until the step count is exhausted. We normalize the actions via their allowed range and the state via a probable scaled range.

4 Experiments

4.1 Benchmarks and experimental setup

We utilize our GP function generator to create a test set of diverse GP functions to test our models in-distribution performance. To test out-of-distribution performance, we use the Virtual Library of Simulated Experiments (VLSE) [34] and Black-Box Optimization Benchmark (BBOB) [35]. These both contain synthetic functions that are traditionally used as standard benchmarks for global black-box optimization. We test our method on $2\,\mathrm{D}$, $5\,\mathrm{D}$, $10\,\mathrm{D}$ and $20\,\mathrm{D}$ function spaces. To evaluate our model on real-world scenarios, we use the HPO-B dataset [33]. HPO-B contains optimization benchmarks built on real-world experiments from OpenML [36, 37] that have been cleaned up for benchmarking. We show results over the 13 search spaces ranging from $2\,\mathrm{D}$ to $10\,\mathrm{D}$.

We utilize the same global optimizers used to generate training data as our baselines. This includes the following acquisition functions: EI, LogEI, UCB, JES, MES, and TS, all with both RBF and Matern kernels. These baselines were implemented in BoTorch [20]. We also compare to other gradient-free optimizers, including Covariance matrix adaptation evolution strategy (CMA-ES), particle swarm optimization (PSO), and differential evolution (DE). We show the performance of each method on these benchmarks using normalized performance. We measure this with $P = (f^* - \min(\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_L)\}))/(f_m - f^*)$, where f^* is the global minimum and f_m is the median of the initial randomly sampled points. We report the mean performance over tested functions.

4.2 Model performance

We show the mean normalized performance after 50 total evaluations on three datasets in Table 1. ZEROSHOTOPT consistently achieves top-tier performance across the GP test suite and out-of-distribution test suites (BBOB and VLSE). This is important because ZEROSHOTOPT was trained on the GP data, so the BBOB and VLSE tests are completely zero-shot. We see that ZEROSHOTOPT is the top performing method for both synthetic test suites. Additionally, we see that the fine-tuned ZEROSHOTOPT model performs second best on the HPO-B dataset. We also see the challenge of selecting a BO method here, as the top performing methods are different than on the synthetic datasets. While the original ZEROSHOTOPT remains competitive with the top BO methods, fine-tuning further turns ZEROSHOTOPT into a top-performing method.

4.3 Runtime results

We compare the runtime of the best BO method on the synthetic data, GP-LogEI (Matern), to the runtime of ZEROSHOTOPT across various dimensions in Table 2. We report the average runtime per evaluation, measured on an NVIDIA H100 GPU using our GP benchmark suite. The runtime of both ZEROSHOTOPT and GP-LogEI (Matern) increases with higher dimensions, with the exception of GP-LogEI (Matern) at 20D (due to inherent variability in GP fitting and acquisition optimization, smoother acquisition landscapes, or reduced convergence that occurs in higher dimensions). However, ZEROSHOTOPT remains substantially faster than GP-LogEI (Matern) across all dimensions.

Table 1: Mean normalized performance after 50 total evaluations (10 initial + 40 model steps), averaged over 2 D, 5 D, 10 D and 20 D synthetic functions (GP, BBOB & VLSE) and HPO-B functions ranging from 2D to 10D. We evaluate mean and standard deviation across 5 seed splits.

	GP		**BBOB & VLSE**		**HPO-B**	
Method	Mean ± SD	Rank	Mean ± SD	Rank	Mean ± SD	Rank
ZEROSHOTOPT	**0.647 ± 0.011**	**1**	**0.881 ± 0.003**	**1**	0.885 ± 0.009	8
GP-LogEI (Matern)	0.629 ± 0.008	2	0.878 ± 0.006	2	0.900 ± 0.007	5
GP-MES (Matern)	0.622 ± 0.012	3	0.868 ± 0.006	6	0.887 ± 0.010	7
GP-UCB (Matern)	0.610 ± 0.004	4	0.866 ± 0.004	7	0.829 ± 0.013	16
GP-EI (Matern)	0.604 ± 0.004	5	0.878 ± 0.004	2	0.866 ± 0.007	11
GP-UCB (RBF)	0.603 ± 0.011	6	0.865 ± 0.006	8	0.845 ± 0.011	14
GP-LogEI (RBF)	0.598 ± 0.016	7	0.870 ± 0.004	4	0.903 ± 0.003	4
GP-MES (RBF)	0.589 ± 0.007	8	0.861 ± 0.005	9	0.890 ± 0.008	6
GP-EI (RBF)	0.580 ± 0.008	9	0.869 ± 0.002	5	0.868 ± 0.004	10
GP-JES (Matern)	0.518 ± 0.010	10	0.804 ± 0.005	11	**0.915 ± 0.003**	**1**
GP-JES (RBF)	0.512 ± 0.016	11	0.798 ± 0.005	12	0.913 ± 0.004	3
GP-TS (Matern)	0.508 ± 0.013	12	0.795 ± 0.004	14	0.844 ± 0.004	15
GP-TS (RBF)	0.506 ± 0.016	13	0.797 ± 0.007	13	0.848 ± 0.007	13
PSO	0.468 ± 0.009	14	0.790 ± 0.006	15	0.856 ± 0.006	12
CMA-ES	0.442 ± 0.007	15	0.827 ± 0.007	10	0.740 ± 0.014	17
DE	0.420 ± 0.008	16	0.718 ± 0.005	16	0.877 ± 0.005	9
ZEROSHOTOPT Fine-Tune	–	–	–	–	0.913 ± 0.002	2

Table 2: Average runtime (seconds) per evaluation over different dimensions.

Method	**2D**	**5D**	**10D**	**20D**
ZEROSHOTOPT	0.368	0.778	1.889	5.556
GP-LogEI (Matern)	30.597	50.030	92.865	87.752

5 Conclusion, limitations, and future work

In this work, we proposed ZEROSHOTOPT, a novel approach for black-box global optimization using a pretrained transformer model trained with offline RL. We formulated the optimization task as a sequential decision-making problem and utilized GPs to create a wide variety of synthetic functions for training. Learning from high-quality trajectories from multiple expert optimizers, our model shows strong performance on unseen optimization problems without the need for task-specific tuning. This work highlights the potential of data-driven pretrained models with GPU acceleration for advancing global optimization and naturally paves the way for promising future extensions.

Our approach is currently limited to continuous, single-objective optimization under 20D, which restricts its applicability to problems involving combinatorial or mixed-integer decision spaces, as well as multi-objective scenarios. Extending the model to tackle these broader categories could significantly enhance its versatility. Additionally, reducing model size through parameter-efficient techniques or pruning could improve deployment efficiency. Lastly, while ZEROSHOTOPT is competitive with *all* baseline BO variants and matches or beats the best BO variants, it is occasionally outperformed by the best variants for individual evaluation distributions. Further improvement beyond the performance of the top BO methods would increase the usefulness of the model.

However, there are many advantages of our transformer-based approach in comparison to BO and opportunities for future work. First, if training data is available, we show that ZEROSHOTOPT can be fine-tuned to specific domains. Additionally, ZEROSHOTOPT could be augmented to include semantic information that is difficult to incorporate with BO-based methods. Including information such as historical data from previous tests or parameter definitions could provide valuable information that further improves models based on ZEROSHOTOPT. Lastly, scaling up with additional training data from a broader range of functions, as well as utilizing larger networks, is expected to improve performance. Therefore, ZEROSHOTOPT provides a base that can be utilized to extend to different contexts or quantities of information in a way that BO cannot.

Acknowledgments

We would like to thank Ashim Sitoula for early experiments on RL and environment generation. This work is supported by Jane Street and the MIT-IBM Watson AI Lab.

References

[1] John H Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[2] Scott Kirkpatrick, C Daniel Gelatt Jr, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[3] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

[4] 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*, 104(1):148–175, 2015.

[5] Timothy Erps, Michael Foshey, Mina Konaković Luković, Wan Shou, Hanns Hagen Goetzke, Herve Dietsch, Klaus Stoll, Bernhard von Vacano, and Wojciech Matusik. Accelerated discovery of 3d printing materials using data-driven multiobjective optimization. *Science advances*, 7(42):eabf7435, 2021.

[6] Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained bayesian optimization for automatic chemical design using variational autoencoders. *Chemical science*, 11(2):577–586, 2020.

[7] Ahmed Alaa and Mihaela Schaar. Autoprognosis: Automated clinical prognostic modeling via bayesian optimization with structured kernel learning. In *International conference on machine learning*, pages 139–148. PMLR, 2018.

[8] Ryan Turner, David Eriksson, Michael McCourt, Juha Kiili, Eero Laaksonen, Zhen Xu, and Isabelle Guyon. Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020. In *NeurIPS 2020 Competition and Demonstration Track*, pages 3–26. PMLR, 2021.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.

[10] Xingyou Song, Oscar Li, Chansoo Lee, Daiyi Peng, Sagi Perel, Yutian Chen, et al. Omnipred: Language models as universal regressors. *arXiv preprint arXiv:2402.14547*, 2024.

[11] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. *Advances in neural information processing systems*, 34:15084–15097, 2021.

[12] Yutian Chen, Xingyou Song, Chansoo Lee, Zi Wang, Richard Zhang, David Dohan, Kazuya Kawakami, Greg Kochanski, Arnaud Doucet, Marc'aurelio Ranzato, et al. Towards learning universal hyperparameter optimizers with transformers. *Advances in Neural Information Processing Systems*, 35:32053–32068, 2022.

[13] Siddarth Krishnamoorthy, Satvik Mehul Mashkaria, and Aditya Grover. Generative pretraining for black-box optimization. *arXiv preprint arXiv:2206.10786*, 2022.

[14] Alexandre Maraval, Matthieu Zimmer, Antoine Grosnit, and Haitham Bou Ammar. End-to-end meta-bayesian optimisation with transformer neural processes. *Advances in Neural Information Processing Systems*, 36, 2024.

[15] Xingyou Song, Yingtao Tian, Robert Tjarko Lange, Chansoo Lee, Yujin Tang, and Yutian Chen. Position paper: Leveraging foundational models for black-box optimization: Benefits, challenges, and future directions. *arXiv preprint arXiv:2405.03547*, 2024.

[16] Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for bayesian optimization. *Advances in Neural Information Processing Systems*, 36:20577–20612, 2023.

[17] Daniel Golovin, Benjamin Solnik, Subhodeep Moitra, Greg Kochanski, John Karro, and David Sculley. Google vizier: A service for black-box optimization. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1487–1495, 2017.

[18] Xingyou Song, Qiuyi Zhang, Chansoo Lee, Emily Fertig, Tzu-Kuo Huang, Lior Belenki, Greg Kochanski, Setareh Ariafar, Srinivas Vasudevan, Sagi Perel, et al. The vizier gaussian process bandit algorithm. *arXiv preprint arXiv:2408.11527*, 2024.

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

[20] 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. In *Advances in Neural Information Processing Systems 33*, 2020. URL `https://proceedings.neurips.cc/paper/2020/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html`.

[21] Yunsheng Tian, Mina Konaković Luković, Timothy Erps, Michael Foshey, and Wojciech Matusik. Autooed: Automated optimal experiment design platform. *arXiv preprint arXiv:2104.05959*, 2021.

[22] Marius Lindauer, Katharina Eggensperger, Matthias Feurer, André Biedenkapp, Difan Deng, Carolin Benjamins, Tim Ruhkopf, René Sass, and Frank Hutter. Smac3: A versatile bayesian optimization package for hyperparameter optimization. *Journal of Machine Learning Research*, 23(54):1–9, 2022. URL `http://jmlr.org/papers/v23/21-0888.html`.

[23] Alexander Cowen-Rivers, Wenlong Lyu, Rasul Tutunov, Zhi Wang, Antoine Grosnit, Ryan-Rhys Griffiths, Alexandre Maravel, Jianye Hao, Jun Wang, Jan Peters, and Haitham Bou Ammar. Hebo: Pushing the limits of sample-efficient hyperparameter optimisation. *Journal of Artificial Intelligence Research*, 74, 07 2022.

[24] Huaijun Jiang, Yu Shen, Yang Li, Beicheng Xu, Sixian Du, Wentao Zhang, Ce Zhang, and Bin Cui. Openbox: A python toolkit for generalized black-box optimization. *Journal of Machine Learning Research*, 25(120):1–11, 2024.

[25] Taku Yamagata, Ahmed Khalil, and Raul Santos-Rodriguez. Q-learning decision transformer: Leveraging dynamic programming for conditional sequence modelling in offline rl. In *International Conference on Machine Learning*, pages 38989–39007. PMLR, 2023.

[26] Yueh-Hua Wu, Xiaolong Wang, and Masashi Hamaya. Elastic decision transformer. *Advances in Neural Information Processing Systems*, 36, 2024.

[27] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems, 2020. URL `https://arxiv.org/abs/2005.01643`.

[28] Lei Song, Chenxiao Gao, Ke Xue, Chenyang Wu, Dong Li, Jianye Hao, Zongzhang Zhang, and Chao Qian. Reinforced in-context black-box optimization. *arXiv preprint arXiv:2402.17423*, 2024.

[29] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[30] Florian Bordes, Richard Yuanzhe Pang, Anurag Ajay, Alexander C Li, Adrien Bardes, Suzanne Petryk, Oscar Mañas, Zhiqiu Lin, Anas Mahmoud, Bargav Jayaraman, et al. An introduction to vision-language modeling. *arXiv preprint arXiv:2405.17247*, 2024.

[31] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

[32] Yutian Chen, Matthew W Hoffman, Sergio Gómez Colmenarejo, Misha Denil, Timothy P Lillicrap, Matt Botvinick, and Nando Freitas. Learning to learn without gradient descent by gradient descent. In *International Conference on Machine Learning*, pages 748–756. PMLR, 2017.

[33] Sebastian Pineda Arango, Hadi S Jomaa, Martin Wistuba, and Josif Grabocka. Hpo-b: A large-scale reproducible benchmark for black-box hpo based on openml. *arXiv preprint arXiv:2106.06257*, 2021.

[34] S. Surjanovic and D. Bingham. Virtual library of simulation experiments: Test functions and datasets. Retrieved September 20, 2024, from `http://www.sfu.ca/~ssurjano`, 2013.

[35] Ouassim Elhara, Konstantinos Varelas, Duc Nguyen, Tea Tusar, Dimo Brockhoff, Nikolaus Hansen, and Anne Auger. Coco: the large scale black-box optimization benchmarking (bbob-largescale) test suite. *arXiv preprint arXiv:1903.06396*, 2019.

[36] Joaquin Vanschoren, Jan N Van Rijn, Bernd Bischl, and Luis Torgo. Openml: networked science in machine learning. *ACM SIGKDD Explorations Newsletter*, 15(2):49–60, 2014.

[37] Bernd Bischl, Giuseppe Casalicchio, Taniya Das, Matthias Feurer, Sebastian Fischer, Pieter Gijsbers, Subhaditya Mukherjee, Andreas C Müller, László Németh, Luis Oala, et al. Openml: Insights from 10 years and more than a thousand papers. *Patterns*, 2025.

[38] GPy. GPy: A gaussian process framework in python. `http://github.com/SheffieldML/GPy`, since 2012.

[39] Andrej Karpathy. NanoGPT. `https://github.com/karpathy/nanoGPT`, 2022.

[40] Ke Li and Jitendra Malik. Learning to optimize. *arXiv preprint arXiv:1606.01885*, 2016.

[41] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems*, 29, 2016.

[42] Tianlong Chen, Xiaohan Chen, Wuyang Chen, Howard Heaton, Jialin Liu, Zhangyang Wang, and Wotao Yin. Learning to optimize: A primer and a benchmark. *Journal of Machine Learning Research*, 23(189): 1–59, 2022.

[43] Sofian Chaybouti, Ludovic Dos Santos, Cedric Malherbe, and Aladin Virmaux. Meta-learning of black-box solvers using deep reinforcement learning. In *NeurIPS 2022, MetaLearn Workshop*, 2022.

[44] Xiaobin Li, Kai Wu, Yujian Betterest Li, Xiaoyu Zhang, Handing Wang, and Jing Liu. Pretrained optimization model for zero-shot black box optimization. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[45] Samuel Müller, Matthias Feurer, Noah Hollmann, and Frank Hutter. Pfns4bo: In-context learning for bayesian optimization. In *International Conference on Machine Learning*, pages 25444–25470. PMLR, 2023.

[46] Zi Wang, George E Dahl, Kevin Swersky, Chansoo Lee, Zachary Nado, Justin Gilmer, Jasper Snoek, and Zoubin Ghahramani. Pre-trained gaussian processes for bayesian optimization. *Journal of Machine Learning Research*, 25(212):1–83, 2024.

[47] Michael Volpp, Lukas P Fröhlich, Kirsten Fischer, Andreas Doerr, Stefan Falkner, Frank Hutter, and Christian Daniel. Meta-learning acquisition functions for transfer learning in bayesian optimization. *arXiv preprint arXiv:1904.02642*, 2019.

[48] Bing-Jing Hsieh, Ping-Chun Hsieh, and Xi Liu. Reinforced few-shot acquisition function learning for bayesian optimization. *Advances in Neural Information Processing Systems*, 34:7718–7731, 2021.

[49] Robert Lange, Yingtao Tian, and Yujin Tang. Large language models as evolution strategies. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 579–582, 2024.

[50] Tennison Liu, Nicolás Astorga, Nabeel Seedat, and Mihaela van der Schaar. Large language models to enhance bayesian optimization. *arXiv preprint arXiv:2402.03921*, 2024.

A Implementation details

A.1 Data generation

The ability to generate diverse functions, while scaling up in terms of complexity and dimensionality, is crucial to our model's success. Inspired by Chen et al. [32], we use the expressive power of Gaussian Processes (GPs) as a general function generator. Specifically, we randomize the kernel type ($k(\mathbf{x}, \mathbf{x}')$) and its length scale to create variability. We randomly sample n input points $\{\mathbf{x}_i\}_{i=1}^n \subset \mathcal{X}$ within the search space and assign random objective values $\{y_i\}_{i=1}^n$. Using these samples, we fit a GP to obtain the posterior mean $\mu(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$, resulting in a posterior function that serves as the training function.

To evaluate the function at a specific point $\mathbf{x}$, we perform inference on the GP posterior, yielding a sample from the predictive distribution: $f(\mathbf{x}) \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$, where $\mu(\mathbf{x})$ is the mean of the posterior and $\sigma^2(\mathbf{x})$ is the variance derived from the kernel. We sample a single variance input used over the whole environment to generate a continuous function. This approach allows us to efficiently generate training functions with varied and complex behavior.

In our implementation based on GPy [38], we randomly select the kernel type from the set of RBF, Matern 3/2, Matern 5/2, Exponential, Cosine and Quadratic kernels. We also further randomize the kernel by combining up to two kernels together, by adding or multiplying randomly selected kernels. In total, we include 78 different kernel types by combining kernels. We also randomize the length scale between 0.1 and 10, and the number of input points for fitting the GP between $10 * d$ and $30 * d$, where d is the dimension of the action space. This diverse setup ensures variability and richness in the generated functions. A few examples of a generated functions are shown in Figure 2.

We use a variety of global optimization algorithms in each environment to generate expert trajectories. We vary the acquisition function and the kernel used to fit each model. We utilize both RBF and matern kernels for each process, with Expected Improvement (EI), Log Expected Improvement (LogEI), Upper Confidence Bound (UCB), Joint Entropy Search (JES), Max-value Entropy Search (MES), and Thompson Sampling (TS) as acquisition functions. We begin by sampling 10 initial points randomly to initialize each model. We then fit the GP model to the set of points and use the specified acquisition function to sample a prospective point. After evaluating that point, we refit the model and continue iteratively. In total, we generate 12 trajectories for each environment. These trajectories contain the 10 initial samples and then a series of points generated by each method. We generate trajectories of length 40 steps, not including the initial samples, for a total of 50 evaluations. We use these trajectories as a baseline for our input regret value and include all as training data for our model.

We generate trajectories using CPU machines, primarily on an Intel Xeon Platinum 8260 system. We find that, although BoTorch supports GPU acceleration, the most cost-efficient manner to generate trajectories is to use parallel CPU processes. However, even with optimizations, running our baseline global optimizers is slow, especially for higher dimensions. Therefore, we total $\sim 150,000$ vCPU hours for our data generation. Additionally, we were limited by the amount of trajectories we could generate and more data may help improve the model.

A.2 Model architecture and training

Trajectory quality: Regret is defined as the trajectory quality in our paper. We process the regret by the following transformation for m initial samples and L overall steps:

$$R(\mathbf{x}_1, f(\mathbf{x}_1), \ldots, \mathbf{x}_L, f(\mathbf{x}_L)) = \sqrt{\frac{\min(\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_L)\}) - f_L^*}{\text{Median}(\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_m)\}) - f_L^*}} \tag{1}$$

Note that this trajectory quality definition differs from that used in Decision Transformer [11], which employs a return-to-go $R_t = \sum_{t'=t}^{T} r_{t'}$ at each step of the trajectory where r_t denotes the reward at times step t. In the context of our sequence, unlike in robotic control or Atari games, the state itself reflects the function value, providing direct information about the return. Thus, there is no need for an additional reward signal that differs from the function value. In other words, the state sufficiently conveys the outcomes of prior actions, giving the agent enough context to effectively propose new actions.

Training: Our full training dataset contains $\sim 20,000,000$ total trajectories. We use data augmentation to expand this and enable further generalization. We augment by swapping axes and flipping the action space to provide additional training data. This augmentation greatly expands the trajectory space, particularly for higher dimensions. We also include shortened trajectories (i.e., the first 20 steps of a 40 step trajectory) to expand the dataset. Overall, Table 3 contains the number of environments and therefore trajectories of each type before augmentation. We input trajectories of length 10, 20, 30, and 40 to the model, where length does not include the 10 random initial steps.

Table 3: Number of functions and trajectories for each dimension.

Dimension	Functions	Trajectories
2D	500,000	6,000,000
3D	200,000	2,400,000
4D	100,000	1,200,000
5D–20D	50,000 each	600,000 each
Total	**1,600,000**	**19,200,000**

Our overall training framework is based on the NanoGPT repository [39], which provides a simple and fast GPT implementation, though additional engineering optimizations could further improve efficiency. We adapted this model with our embedding strategy and data pipeline. We train a model with 16 layers, 16 heads, and an embedding dimension of 1024, totaling ~ 200 million parameters. The action and state space is split into 2,000 evenly spaced bins for our loss calculation. We use trajectories ranging from 2D to 20D and from 10 steps to 40 steps, not including the 10 initial random steps, to train this model. Due to computational constraints and slow generation of higher dimensional data, this dataset contains more data from lower dimensions and fewer steps.

We use the hyperparameters shown in Table 4 to train the model. The base model is trained using the AdamW optimizer using 4 Nvidia H100 GPUs. In total, training takes ~ 3 days on this system.

Table 4: Hyperparameters for model training. The training implementation is largely based on NanoGPT [39], with adaptations for our specific architecture and data input.

Hyperparameter	Value
Total Parameters	200 million
Number of Transformer Heads	16
Transformer Embedding Dimension	1024
Transformer Layers	16
Learning Rate	6×10^{-4} with cosine scaling
Weight Decay	1×10^{-1}
Batch Size	1024 trajectories
Total Iterations	250,000
Precision	BF16

A.3 Inference

A.3.1 Overall Methodology

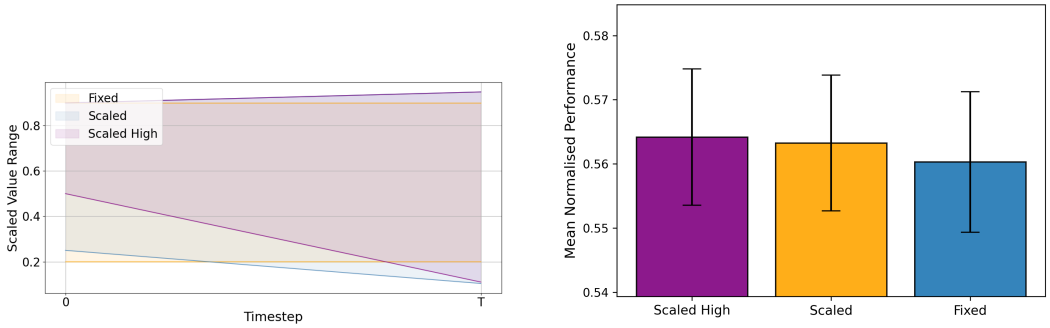
At inference time, we specify zero as the desired regret R to initiate generation, encouraging the agent to act as the best baseline methods do. We then provide the 10 random initial samples, as provided to the global optimization methods to initialize the model. To perform a step, we perform 4 parallel passes of the model at the same time. For each pass, the model generates a distribution across bins for the first dimension. We sample from the distribution using top-p sampling with $p = 0.9$ and convert the bin to a continuous action using the center of the range of the selected bin. We iteratively select actions for each dimension. Once we have sampled the full action dimension for each pass, we use the model to predict a state distribution. Based on these state distributions, we use an expected improvement acquisition function to select the best of the 4 possible actions. We then evaluate the selected action and proceed to the next step. We continue to sample and select actions for the specified number of steps.

A.3.2 Inference Scaling

We show the performance of different inference strategies using the same model in Figure 4. We scale our states between 0 and 1 during training using the highest and lowest values achieved by any method on the function space, but we do not know these values at inference time. Therefore, we need to develop a scaling strategy for inference. We do this by changing the scaling of the input states at step $t+1$, using parameters C_u and C_l in the following equation:

$$S_i' = (S_i - min(S_0, ..., S_t))/(max(S_0, ..., S_t) - min(S_0, ..., S_t)) * (1 - C_u - C_l) + C_l.$$

We test various methodologies, from fixed values to scaled values that decrease as the trajectory progresses. We show the ranges of scaled values inputed over T timesteps with 3 methodologies in Figure 4a and the results of these 3 methodologies in Figure 4b, which shows a high C_l that decreases with progression performs the best. However, the model is fairly robust to different initializations.



(a) Scaled range of input over steps for each method.

(b) Results on 2D, 5D, 10D, and 20D GP functions.

Figure 4: Ablation on scaling strategies during inference.

We define C_u and C_l differently for each method, where t is the current step and L is the number of overall steps. We trial multiple different methodologies for scaling states during inference. These methods vary the way we normalize the state values, given we have less information about the sequence at inference time.

Fixed: $C_u : 0.1, C_l : 0.2$

Scaled: $C_u : 0.05 + 0.05 * \frac{L-t}{L}, C_l : 0.1 + 0.15 * \frac{L-t}{L}$

Scaled High: $C_u : 0.05 + 0.05 * \frac{L-t}{L}, C_l : 0.1 + 0.4 * \frac{L-t}{L}$

As shown in Figure 4, we find that **Scaled High** methodology produces the best results. Therefore, we use this scaling for all presented results. Keeping our scaling configuration consistent across tests allows for flexible model usage and improved utility of our model.

A.3.3 Inference Runtime

We show complete results comparing the runtime of ZeroShotOpt to our baseline methods in Table 5. We report the average runtime per evaluation, measured on an NVIDIA H100 GPU using our GP benchmark suite. We see that there is some variance across BO methods due to the nature of different acquisition functions, but ZeroShotOpt is generally faster than most BO methods and can be made more efficient by improved techniques in transformer GPU acceleration. CMA-ES, DE, and PSO are all much faster due to their evolutionary nature, but this comes with much worse performance, which is not preferred for optimizations with low evaluation budgets.

A.4 Model Scaling

We complete an ablation on model size, suggesting that further increasing the size of ZeroShotOpt could further improve results. We compare the performance of ZeroShotOpt with different model sizes in Figure 5. In particular, we compare our full-size ZeroShotOpt model with 200M parameters with a smaller model with 90M parameters. This comparison is done without state

Table 5: Average runtime (seconds) per evaluation over different dimensions, sorted by 20D runtime.

Method	**2D**	**5D**	**10D**	**20D**
CMA-ES	0.017	0.022	0.040	0.083
PSO	0.032	0.043	0.090	0.208
DE	0.033	0.044	0.091	0.209
GP-JES (RBF)	2.012	1.972	1.888	1.905
GP-JES (Matern)	2.018	2.067	2.061	2.116
ZEROSHOTOPT	0.368	0.778	1.889	5.556
GP-TS (Matern)	13.041	20.094	24.262	24.326
GP-TS (RBF)	12.460	18.486	26.779	24.905
GP-UCB (Matern)	20.088	31.973	43.356	36.811
GP-UCB (RBF)	19.016	32.738	44.413	37.279
GP-EI (RBF)	15.804	37.344	57.042	46.994
GP-EI (Matern)	17.093	37.621	67.082	54.672
GP-LogEI (Matern)	30.597	50.030	92.865	87.752
GP-MES (RBF)	21.625	54.192	85.005	89.352
GP-LogEI (RBF)	29.871	54.489	85.576	94.265
GP-MES (Matern)	23.166	57.048	86.132	95.136

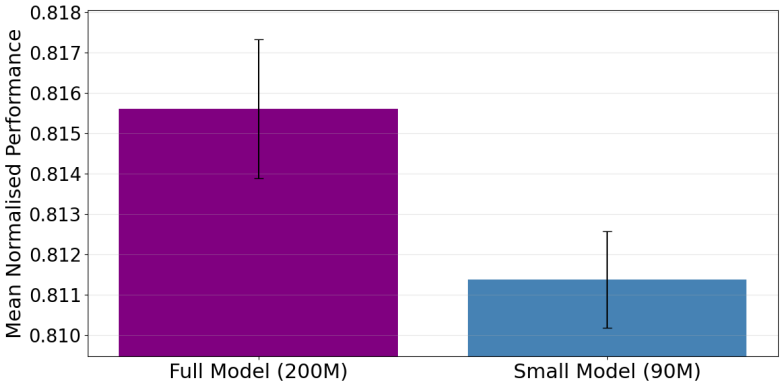


Figure 5: Ablation on Model Size. Mean normalized performance over steps 10, 20, 30 and 40 on 2 D, 5 D, 10 D and 20 D BBOB and VLSE functions. We test over 500 functions from each dataset and evaluate standard deviation across 5 independent seed splits.

prediction and the expected improvement acquisition function due to compute requirements. All training is completed with the same number of iterations and training data. We find that the larger model performs better due to improved capacities, suggesting that further scaling may continue to improve results.

B Model Comparison

We aim provide further comparison to other proposed learning to optimize methods below. The idea of learning to optimize has been studied relatively early in continuous gradient-based optimization [40–42]. Many early works reformulate optimization as a sequence prediction problem, training recurrent neural network (RNN) to predict the next point to evaluate. Yet, gradients need to be provided as additional information and they mostly assume in-domain settings [42]. One notable exception are Chen et al. [32], who show that trained RNNs are able to generalize from simple objective functions to a variety of other unseen test functions without gradient information.

Chaybouti et al. [43] train a transformer optimizer using online meta RL, but still limited to learning task-specific solvers without generalization to a wide variety of unseen tasks. Similarly, neural acquisition process (NAP) [14] also considers a similar online RL-based formulation of the problem. They propose an additional training objective to explicitly predicting acquisition function values to

guide RL exploration more effectively. They also focus on evaluating transfer learning but, different from us, still train individually for different tasks on selected, task-related data.

Pretrained Optimization Model (POM) [44] introduces a population-based model for zero-shot optimization by formulating optimization as an evolutionary algorithm and using meta-learning for training the model. However, they focus on population-based, high-dimensional problems (>100 dimensions), which are beyond the scope we are considering. Given the close connection with reinforcement learning, there are numerous other works that separately study the problem of learning the surrogate [45, 46] or the acquisition function [47, 48]. Large language models have also been applied out of the box for global optimization; for instance, by using in-context learning [49, 50].

We also include more detailed comparison between ZEROSHOTOPT and other pretrained optimizers (BONET, RIBBO, OptFormer, and PFNs4BO) below to further emphasize the differences and impact of ZEROSHOTOPT.

BONET [13]: BONET applies a causal transformer and RL-inspired pretraining to specific optimization tasks. They introduce a novel approach by synthesizing optimization trajectories by reordering samples in a given dataset, which creates high quality synthetic trajectories. They show promising results on domain-specific extrapolation, outperforming classical methods in their test domain. In comparison to ZEROSHOTOPT, they focus on domains with higher budgets (in the hundreds) and show results on a small set of high-dimensional problems. Additionally, the study only evaluates domain-specific extrapolation based on a relatively large-scale dataset, instead of tackling unseen optimization problems with just a few dozens of samples. This is an important distinction, as BONET uses more initial points to initialize the model and trains and tests on similar distributions. Therefore, the general problem being solved by BONET has different limitations. Overall, they provide a methodology that advances ideas in transformer-based optimization, but work towards different problems than ones ZEROSHOTOPT is designed for.

OptFormer [12]: OptFormer is a text-based transformer framework designed for hyperparameter optimization for general problem space beyond continuous. The model is trained on a proprietary hyperparameter optimization database and exploits the textual nature by leveraging hyperparameter names and descriptions. The method shows promise for transformer-based models trained on large datasets by outperforming solutions such as BO on the given test cases, which are drawn from similar distributions as training data. They train a model on a proprietary dataset, BBOB, and HPO-B data. This shows strong results on the test sets of each respective dataset. As their dataset contains a diverse set of functions and their model shows the ability to learn a variety of different algorithms, their contributions show promise towards the idea of a general-purpose transformer-based optimizer. However, the model shows significant performance degradation when tested on distributions different from its training data. Overall, its strong performance on a large dataset marks an important step towards general-purpose transformer-based optimization, but does not truly work as a general purpose optimizer.

We utilize the pretrained model and code provided by OptFormer to compare to our results on BBOB data. Although we used the default parameters and setup provided in the original documentation, we find the OptFormer model very slow on our hardware, which limited the scope of our experiments. Therefore, we test one environment for each of the 24 BBOB categories for each dimension in $2\,\mathrm{D}$, $5\,\mathrm{D}$, and $10\,\mathrm{D}$. Access to TPU machines may negate these issues. We choose the BBOB tests because although they are contained within the OptFormer training data, we can create general synthetic tests. Therefore, we can limit the impact of including exactly the same data in our test set as OptFormer was trained on. We show the results of our experiments in Figure 6. We find that ZEROSHOTOPT outperforms OptFormer, although there is a high standard deviation due to the limited number of tests. Additionally, it is important to note that OptFormer includes BBOB tests within its training data, whereas ZEROSHOTOPT does not. OptFormer trains on 16 out of the 24 BBOB classes, while specifying 5 examples for out-of-distribution tests. Therefore, this shows the effectiveness of ZEROSHOTOPT as a general-purpose optimizer that can surpass previous learned optimizers, even when they are trained on task-specific data.

RIBBO [28]: RIBBO is a transformer-based framework trained using offline reinforcement learning. While similarly inspired by decision transformer, RIBBO is trained on the benchmark datasets taken from existing optimization tasks. RIBBO shows strong performance and is able to outperform the algorithms it was trained on on many of its test datasets, which is an important contribution. This shows the promising aspect of conditioned transformed-based models. However, RIBBO

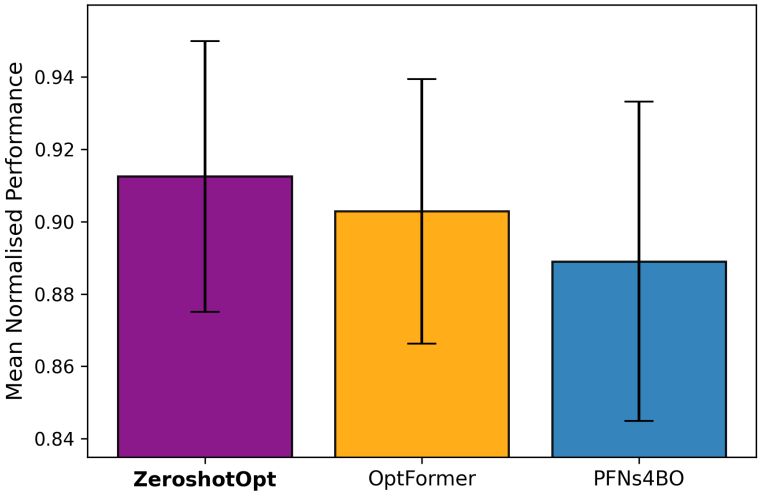


Figure 6: Evaluation Comparison with OptFormer and PFNs4BO. Mean normalized performance over steps 10, 20, 30 and 40 on 2 D, 5 D, and 10 D BBOB functions. We test over 24 functions from each dataset and evaluate standard deviation across 4 independent seed splits.

uses fixed dimensionality, an inflexible tokenization scheme, and a relatively small scale training scheme. Therefore, RIBBO can only evaluate results for small training and evaluation datasets from similar distributions. This differs from ZEROSHOTOPT, which is designed to be scalable and adapt to different dimension counts and function environments. This scalable nature and use of competitive data generation methods on synthetic functions allows ZEROSHOTOPT to generalize to different distributions, a feature which is minimal within RIBBO. However, the ideas of conditioned transformer-based models and the improved performance RIBBO shows over baselines on individual datasets provides a step towards outperforming BO in transformer-based optimization.

PFNs4BO [45]: PFNs4BO use prior-data fitted networks (PFNs) as a surrogate model for Bayesian Optimization. They show the ability to fix to flexible priors, including Gaussian Processes and Bayesian Neural Networks. Similar to us, they train on synthetic datasets with the goal of creating a general-purpose optimizer. They show promising results for their method across a range of test suites, while also allowing the user to add a prior to further improve the model. We show comparison results to ZEROSHOTOPT and OptFormer in Figure 6. We see the that transformer-based models outperform PFNs4BO on the BBOB set. However, this method is another promising approach in the learning to optimize space.

C Evaluations

C.1 Benchmarks

GP: We use the function generator used for generating training data as the initial baseline for our method. This contains diverse GP functions over a flexible dimension.

BBOB [35]: The Black-Box Optimization Benchmarking suite was developed as part of the COCO (Comparing Continuous Optimizers) platform to provide a rigorous and standardized environment for evaluating continuous, unconstrained optimization algorithms. BBOB includes 24 benchmark functions that represent a wide range of challenges encountered in real-world optimization, such as separability, multimodality, ill-conditioning, and non-convexity. Each function is parameterized with randomized shifts, scalings, and rotations to prevent algorithms from overfitting to specific patterns. The suite was carefully designed through mathematical constructions and transformations of base functions to create controlled yet diverse test cases. It supports varying dimensions and is widely used in the black-box optimization community.

VLSE [34]: The Virtual Library of Simulation Experiments is a benchmark suite aimed at simulating real-world optimization problems where the objective function is defined by computational simulations rather than closed-form expressions. We use the optimization test problems from this set.

HPO-B [33]: The Hyperparameter Optimization Benchmark was created to support fast and reproducible evaluation of hyperparameter optimization methods by providing surrogate models that approximate the behavior of real machine learning training processes. Built from large-scale logging of real hyperparameter tuning runs on algorithms like XGBoost, SVMs, feedforward neural networks, and others across multiple datasets, HPO-B allows researchers to benchmark optimization strategies without incurring the computational cost of retraining models for each evaluation. It enables fair comparisons across different optimizers under controlled experimental conditions. There are a few different evaluation sets provided by HPO-B. We utilize the provided train and test sets for our evaluations.

C.2 Baselines

We utilize the same global optimizers used to generate training data as our baselines. This includes the following acquisition functions for BO:

- **Expected Improvement (EI)**: Selects point that is expected to improve upon the current best observation, balancing exploration and exploitation.
- **Log Expected Improvement (LogEI)**: A variant of EI that operates in log space, making it more suitable for objectives with large dynamic ranges or multiplicative noise.
- **Upper Confidence Bound (UCB)**: Prioritizes points with high predicted mean and uncertainty, controlled by a trade-off parameter.
- **Joint Entropy Search (JES)**: Reduces uncertainty about both the location and value of the global minimum by selecting the point expected to maximally reduce the joint entropy of the posterior over the minimum.
- **Max-value Entropy Search (MES)**: Reduces uncertainty about the value of the global minimum by selecting query points that are expected to most reduce the entropy of its posterior distribution.
- **Thompson Sampling (TS)**: Samples functions from the posterior and optimizes them directly, encouraging diverse sampling over time.

Each acquisition function is used with both RBF and Matern kernels, providing a good baseline across different smoothness assumptions. We use the default parameterizations of these methods from BoTorch [20], covering a broad range of BO variants.

We also compare to other gradient-free global optimizers:

- **Covariance Matrix Adaptation Evolution Strategy (CMA-ES)**: An evolutionary algorithm that adapts the sampling distribution using covariance information for efficient search.
- **Particle Swarm Optimization (PSO)**: A population-based stochastic optimizer inspired by the social behavior of birds and fish, adjusting candidate solutions based on personal and global bests.
- **Differential Evolution (DE)**: A simple yet powerful method that perturbs candidate solutions using scaled differences between population members.

These classical methods typically require thousands of iterations to converge but provide a strong point of comparison to highlight the performance of our model.