

Bayesian Optimization Hackathon for Chemistry and Materials

Sterling G. Baird,¹ Mehrad Ansari,¹ Zartashia Afzal,² Qianxiang Ai,³ Alexander Al-Feghali,⁴ Mathieu Alain,⁵ Matias Altamirano,⁶ Thomas Andrews,⁷ Andy Sode Anker,⁸ Rija Ansari,⁹ Samuel Ampofo Appiah,¹⁰ Raul Astudillo,¹¹ Ruhana Azam,¹² Mohammed Azzouzi,¹³ Suneel Kumar BVS,¹⁴ Ben Blaiszik,¹⁵ Anna Borisova,¹⁶ Andres Bran,¹⁶ Pengfei Cai,³ Ting-Yeh Chen,¹⁷ Curtis Chong,¹⁸ Samantha Corapi,¹⁸ Mark Croxall,¹⁸ Gbetondji Dovonon,⁵ Jose Manuel Napoles Duarte,¹⁹ Andrew Falkowski,¹⁹ Giuseppe Fisicaro,²⁰ Martin Fitzner,²¹ Quinn Gallagher,²² Sabah Gaznaghi,²³ Jérôme Genzling,⁴ Christoph Griehl,²⁴ Ryan-Rhys Griffiths,²⁵ Taicheng Guo,²⁶ Kehan Guo,²⁶ Nipun Gupta,¹ Ankur Gupta,²⁷ Mohammad Haddadnia,²⁸ Yuyang Han,²⁹ Joscha Hoche,²¹ Alexander V. Hopp,²¹ Marko Huang,¹⁸ Ayodeji Iijshakin,³⁰ Ramsey Issa,¹⁹ Yeonghun Kang,³¹ Jungtaek Kim,³² Akshay Kudva,¹⁷ Rubén Laplaza,¹⁶ Magdalena Lederbauer,¹⁶ Shi Xuan Leong,¹⁸ Paul W. Leu,³² Viola Muning Li,²¹ Mingxuan Li,³² Tao Liu,⁴ Stanley Lo,¹⁸ Jakub Lála,¹³ Osman Mamun,³³ Owen Melville,¹ Michail Mitsakis,³⁴ Cameron Movassaghi,²⁹ Madhav Reddy Muthyala,¹⁷ Marcel Müller,³⁵ Bozhao Nan,²⁶ Duc Nguyen,¹² Daniele Ongari,³⁶ Anthony Onwuli,¹³ Can Özkan,³⁷ Sergio Pablo-Garcia,¹⁸ Elton Pan,³ Ratish Panda,³⁸ Sean Park,¹⁸ Jaehee Park,¹⁵ Dieter Plessers,³⁹ Tobias Plötz,²¹ Ella M. Rajaonson,¹⁸ Bojana Ranković,¹⁶ Jarett Ren,⁴⁰ Rim Rihana,²¹ Jurgis Ruza,³ Akhil S. Nair,⁴¹ Carter Salbego,¹⁹ Erick Lopez Saldivar,⁴² Arifin San,⁴³ Christina Schenk,⁴⁴ Stefan P. Schmid,⁴⁵ Dylan Schubert,⁴⁶ Philippe Schwaller,¹⁶ Cher-Tian Ser,¹⁸ Maitreyee Sharma Priyadarshini,⁴⁷ Yuxin Shen,⁴⁸ Kevin Shen,⁴⁹ Jiale Shi,³ Farshud Sorourifar,¹⁷ Adrian Šošić,²¹ Taylor Sparks,¹⁹ Jan Christopher Spies,⁵⁰ Felix Strieth-Kalthoff,⁵¹ Suraj Sudhakar,¹⁵ Aditya Sundar,⁵² Alessio Tamburro,⁵³ Clara Tamura,⁵⁴ Yifeng (Oliver) Tang,¹⁵ Dandan Tang,⁵⁵ Nikhil Thota,⁴⁰ Mohammad Erfan Toloue Sadegh Azadi,⁴ Gary Tom,¹⁸ Sang Truong,¹² Ricardo Valencia Alborno,⁴⁸ Luis Walter,⁵⁶ Lawrence Wang,⁵⁷ Fanjin Wang,⁵⁸ Andrew Wang,¹⁸ Yiran Wang,³ Jeffrey Watchorn,¹ Benjamin Weiser,⁴ Geemi Wellawatte,¹⁶ Alexander Wiecezorek,⁵⁹ Tim Würger,⁶⁰ Ilya Yakavets,¹ Jakob Zeitler,⁴⁶ Sylvester Zhang,⁴ Yimu Zhao,¹ Yanqiao Zhu,²⁹ Ruijie (Ray) Zhu,¹⁵ and Yunheng Zou⁶¹

¹Acceleration Consortium, University of Toronto,
80 St George St, Toronto, ON M5S 3H6, Canada.*

²ICodeGuru, Canal Road, Quaid-i-Azam Campus, Lahore, Punjab, Pakistan.

³Massachusetts Institute of Technology, Massachusetts Institute of Technology,
77 Massachusetts Ave, Cambridge, MA 02139, United States.

⁴McGill University, 845 Rue Sherbrooke O, Montréal, QC H3A 0G4, Canada.

⁵Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, United Kingdom.

⁶Department of Statistical Science, University College London,
Gower Street, London WC1E 6BT, United Kingdom.

⁷University of Melbourne, CSIRO, Parkville VIC 3052, Australia.

⁸Technical University of Copenhagen and Oxford University,
Anker Engelunds Vej 411, 2800 Kongens Lyngby,
Danmark & Mansfield Rd, Oxford OX1 3TA, United Kingdom.

⁹National Research Council of Canada, 1200 Montreal Road, Ottawa ON, K1A 0R6, Canada.

¹⁰Ghana National Gas Company, Ghana National Gas Company,
Gas House, 7th Avenue, Ridge West, Accra, Ghana.

¹¹California Institute of Technology, 315 Annenberg IST Center,
330 S Chester Ave, Pasadena, CA 91125, United States.

¹²Stanford University, 353 Jane Stanford Way, Stanford, CA, 94305, United States.

¹³Imperial College London, Exhibition Rd, South Kensington, London SW7 2AZ, United Kingdom.

¹⁴Atomica AI solutions private limited, Hyderabad, India.

¹⁵University of Chicago, 5801 S Ellis Ave, Chicago, IL 60637, United States.

¹⁶ETH Zurich, EPFL, Station 06, CH-1015 Lausanne, Switzerland.

¹⁷The Ohio State University, 281 W Lane Ave, Columbus, OH 43210, United States.

¹⁸Department of Chemistry, University of Toronto,
27 King's College Circle, Toronto, ON M5S 1A1, Canada.

¹⁹University of Utah, University of Utah, 201 Presidents Circle, Salt Lake City, UT 84112, United States.

²⁰Institute for Microelectronics and Microsystems (IMM),
National Research Council (CNR), Institute for Microelectronics and Microsystems (IMM),
National Research Council (CNR), Via del Fosso del Cavaliere 100, 00133 Rome, Italy.

²¹Merck KGaA, Darmstadt, Germany, Frankfurter Straße 250, 64293 Darmstadt, Germany.

²²Princeton University, Department of Chemical and Biological Engineering,
Princeton University, Princeton, New Jersey 08540, United States.

²³University of Queensland, Building 75, Cooper Rd, St Lucia QLD 4067, Australia.

- ²⁴Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39112 Magdeburg, Germany.
- ²⁵FutureHouse, 1405 Minnesota St, San Francisco, CA 94107, United States.
- ²⁶University of Notre Dame, Holy Cross Dr, Notre Dame, IN 46556, United States.
- ²⁷Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720.
- ²⁸University Of Harvard, Harvard University, Massachusetts Hall, Cambridge, MA 02138, United States.
- ²⁹University of California, Los Angeles, 607 Charles E Young Dr E, Los Angeles, CA 90093, United States.
- ³⁰Department of Medical Physics & Biomedical Engineering,
University College London, Gower Street, London WC1E 6BT, United Kingdom.
- ³¹KAIST, KAIST (Korea Advanced Institute of Science and Technology),
291 Daehak-ro, Yuseong-gu, Daejeon 34141, South Korea.
- ³²University of Pittsburgh, 3700 O'Hara St, Pittsburgh, PA 15261, United States.
- ³³Cornell University, Ithaca, NY 14850, United States
- ³⁴Technical University of Denmark (DTU), Artus 64 & Potamou 77, Kifissia, Athens, Greece.
- ³⁵Merck KGaA, Darmstadt, Germany / University of Bonn,
Frankfurter Straße 250, 64293 Darmstadt, Germany.
- ³⁶Syngco, Viale Lombardia 20, Bollate, Italy.
- ³⁷TU Delft, Mekelweg 5, 2628CD Delft, the Netherlands.
- ³⁸Birla Institute of Technology, Mesra, Jharkhand 835215, India.
- ³⁹KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium.
- ⁴⁰Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, United States.
- ⁴¹Fritz-Haber Institute of the Max-Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany.
- ⁴²Autonomous University of Chihuahua, Autonomous University of Chihuahua,
Av. Escorza 900, Colonia Centro, 31000 Chihuahua, Chih., Mexico.
- ⁴³JSR Corporation, JSR Bioscience and informatics R&D center (JSR BiRD),
3-103-9 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-0821, Japan.
- ⁴⁴IMDEA Materials Institute, Calle Eric Kandel 2, Tecogetafe, 28906 Getafe (Madrid), Spain.
- ⁴⁵ETH Zurich, Vladimir-Prelog-Weg 1, 8093 Zurich, Switzerland.
- ⁴⁶Matterhorn Studio, Walton Street, Oxford, United Kingdom.
- ⁴⁷Virginia Tech, Virginia Tech, 925 Prices Fork Road, Blacksburg, VA 24061, United States.
- ⁴⁸University of Edinburgh, CH Waddington Building, King's Buildings,
Max Born Crescent, Edinburgh, EH9 3BF, United Kingdom.
- ⁴⁹NobleAI, 345 California St Suite 600, San Francisco, CA 94104, United States.
- ⁵⁰University of Muenster, Corrensstraße 40, 48149 Münster, Germany.
- ⁵¹University of Wuppertal, Germany, Gaußstraße 20, 42119 Wuppertal, Germany.
- ⁵²Argonne National Laboratory, Lemont, IL 60439, United States
- ⁵³Syngco, 350 George Patterson blvd, Bristol, PA 19007, United States.
- ⁵⁴University of Washington, 3900 E Stevens Way NE, Seattle, WA 98195, United States.
- ⁵⁵University of Virginia, University of Virginia, 1827 University Ave, Charlottesville, VA 22903, United States.
- ⁵⁶Heidelberg University, Philosophenweg 19, 69120 Heidelberg.
- ⁵⁷Erthos, 2735 Matheson Blvd E 5, Mississauga, ON L4W 4M8, Canada.
- ⁵⁸Department of Mechanical Engineering, University College London, London, WC1E 7JE, United Kingdom.
- ⁵⁹Empa – Swiss Federal Laboratories for Materials Science and Technology,
Ueberlandstrasse 129, Dübendorf, CH-8600 Switzerland.
- ⁶⁰Helmholtz-Zentrum Hereon, Max-Planck-Str. 1, 21502 Geesthacht, Germany.
- ⁶¹University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1., Canada.

The Acceleration Consortium and Merck KGaA hosted a 2-day virtual hackathon on March 27-28, 2024, bringing together scientists to explore, collaborate, and innovate in the field of Bayesian optimization for the physical sciences. Participants were encouraged to select or develop Bayesian optimization algorithms, apply them to benchmarking tasks, design new benchmarks, create instructional tutorials, and describe real-world applications. With over 100 participants across 69 academic, industry, and government organizations located in 59 cities, 19 countries, and 4 continents, this was a global event. The outputs from this event, including developed algorithms, benchmarks, and tutorials, will serve as valuable resources for the research community, in addition to the new skills learned and connections formed. Released projects and general information are available at <https://ac-bo-hackathon.github.io/> and other locations linked from individual project pages. This event demonstrates the potential of community-driven research efforts to accelerate advances in Bayesian optimization in chemistry and materials science.

* Corresponding authors: {sterling.baird, mehrad.ansari}@utoronto.ca

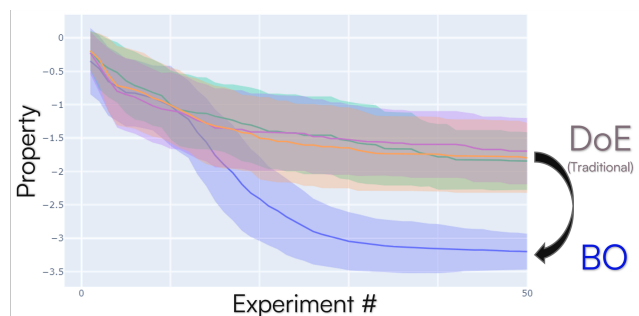


Figure 1. Optimization traces for traditional design of experiment (DoE) methods compared with BO, typically outperforms. BO uses a smart model to predict where to look next in an experiment to find the best results with few experiments [1].

I. Introduction

Bayesian optimization (BO) has emerged as a powerful tool in optimizing complex and expensive-to-evaluate functions, often outperforming traditional search methods in a variety of scientific domains such as optimizing composition and processing parameters to maximize alloy yield strength or identifying synthesis pathways that maximize efficacy of HIV drugs (Figure 1). Hackathons help people to connect, gain skills, and flesh out new ideas. In the words of Michelle Duke, the “Hackathon Queen”:

A hackathon is a short competition where people work together in teams to solve problems and challenges by coming up with solutions and ideas.

The goal of the AC BO Hackathon was to leverage the expertise of a diverse, global community to advance the development and application of BO techniques for solving critical challenges in the physical sciences. The hackathon also aimed to foster collaboration and knowledge sharing among participants from different backgrounds, including academia, national laboratories, government agencies, and private industry. The event attracted 139 active participants from 39 teams, representing 52 academic institutions, 3 national labs, and 14 companies. Likewise, the participants were located in 59 cities, 19 countries, and 4 continents (Figure 2). A full list of projects, including links to the corresponding GitHub repositories, submission video, and social media post are provided in Table I.

II. Hackathon Details and Setup

This hackathon was in large part inspired by Jablonka *et al.* [2] and is one of several materials-focused hackathons in the past [3, 4].

Participants were provided with various resources to prepare for the hackathon – this included GitHub classroom assignments with automated feedback, application- and theory-focused videos and tutorials, Python refresher materials, a list of tools to consider using during the hackathon, project submission instructions, and authorship criteria (Figure 3).

One of the unique aspects of this event is that it was hosted in Gather Town, a sort of union between traditional video conferencing software and retro arcade-style avatars and virtual spaces (Figure 4). Participants create a custom avatar and maneuver in a two-dimensional space. The videos and audio of other participants appear and become audible when nearby, and fade out when far away, simulating an in-person experience. At the beginning of the hackathon, all participants gathered to listen to keynotes in realtime, which were broadcasted via YouTube live and embedded into the Gather Town space. The videos were then made available on the hackathon website, which collectively garnered approximately 1600 views within two months. After the keynotes, teams were assigned tables in breakout rooms, each with a whiteboard. Individual tables were assigned as “private spaces” which isolated the shared audio and video within each space. This had a number of advantages for collaboration within and across teams.

The hackathon concluded with a project showcase accompanied by crowdsourced judging within a “poster room” (Figure 5).

Community judging occurred via Gavel, an automated pairwise comparison judging system. Use of this system helped to improve fairness, scalability, and accuracy by having judges compare projects relative to each other rather than assigning subjective numerical scores (for example, “rate your pain on a scale from 1 to 10, where 10 is the worst possible pain you can imagine”). The approach of pairwise judging reduces bias, allows for handling large competitions efficiently, and produces high-quality rankings using statistical models with dynamic assignments to judges to maximize information gain. It has been successfully used at HackMIT and other events to streamline judging and enhance transparency and credibility.

The corresponding Gavel web app was hosted on Heroku according to directions in the Gavel repository. Individualized links were distributed to judges via email using Gavel’s SendGrid integration. Collectively, 35 judges cast 319 votes.

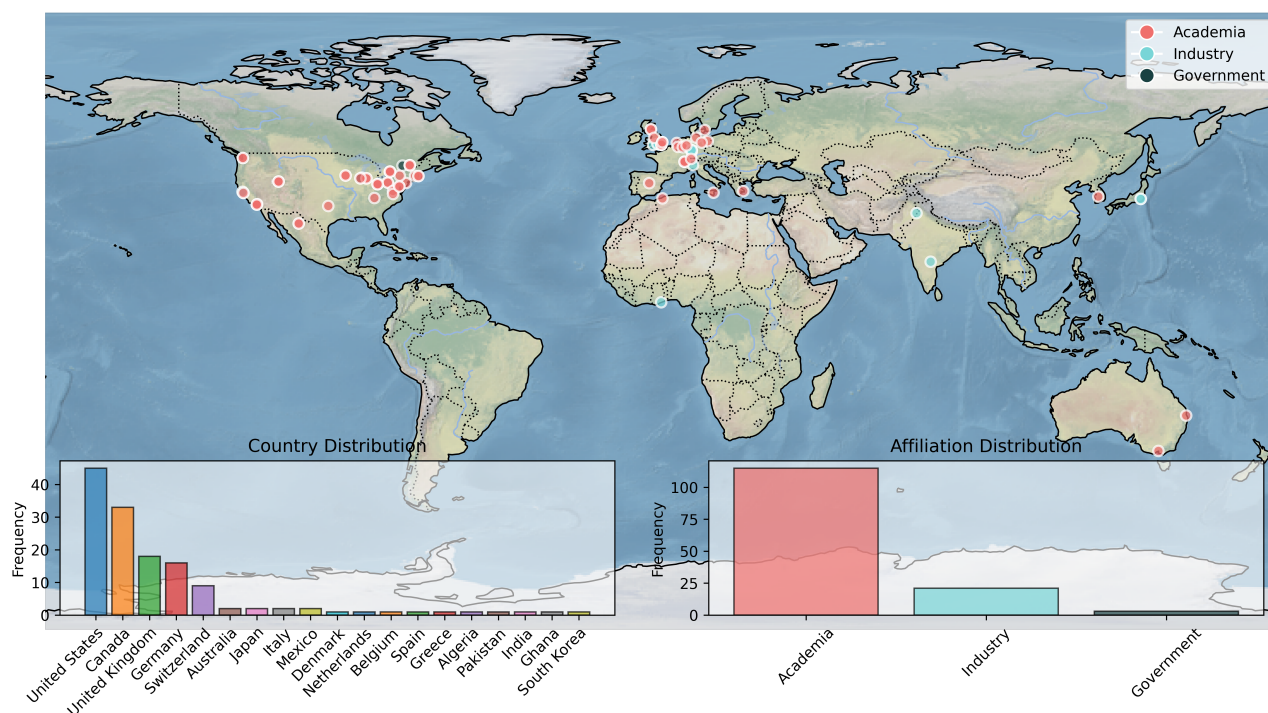


Figure 2. Demographic distributions of the hackathon participants and their affiliations. The hackathon supported 139 participants across 69 academic, industry, and government organizations located in 59 cities, 19 countries, and 4 continents.

BO Hackathon with Acceleration Consortium

Registration | Agenda | Projects | Submission | Videos & slides | Resources | FAQ | About

Home / Hackathon resources

Hackathon resources

Getting Started

- [Slack](#) - Join the hackathon's Slack channel to connect, ask questions, and form teams
- [Project proposals](#) - team leaders should submit their project proposals using the instructions [here](#)
- [Gather](#) - Add the Gather Town event to your calendar
- [Accelerated discovery forum](#) - create an account to join the discussion forums
- [Bayesian optimization](#) - Watch the Accelerate '23 training workshop on Bayesian optimization below

[What are Marie's objectives?](#)

[A Gentle Introduction to Bayesian Optimization](#)

Orientation Modules

Please complete the following orientation assignments in preparation for the hackathon to familiarize the tools and concepts you'll need. You will need to [create a GitHub account](#) to access these resources. If you are already familiar with git and GitHub, you may skip directly to "Intro to GitHub Classroom". When prompted, please select an identifier from the list that's given. In addition to these orientation modules, we also recommend that you familiarize yourself with [Markdown syntax](#) if this is new to you.

- [Intro to Git and GitHub](#)
- [Intro to GitHub Classroom](#)
- [Python Refresher](#)
- [Simple Bayesian optimization script](#)

For those looking for a refresher on Python programming or to implement a simple BO example, see the following GitHub Classroom assignments:

[What is GitHub?](#)

Bayesian optimization

[Bayes Rule by 3Blue1Brown](#) - explanation of the fundamental concept behind Bayesian methods

[This is Bayes' rule](#)

[A Visual Exploration of Gaussian Processes](#) - An explanation of a commonly used predictive model in Bayesian optimization with visualizations and mathematical rigor

[Exploring Bayesian Optimization](#) - An overview of active learning and Bayesian optimization with visualizations and mathematical rigor

[Bayesian Optimization Book](#) - For those that REALLY want to understand BO, this is a self-contained, comprehensive deep dive that builds up basic principles from "scratch" and delves into advanced topics. The book is free and open-source.

Python

Programming expertise is not required, but at least beginner Python programming experience is recommended for participation in code focused projects of the hackathon. For those looking for a brief, interactive refresher on Python programming, see the [GitHub Classroom assignment](#) from the first section on this page. For those without prior Python experience, we recommend you complete an introductory Python course in preparation for the hackathon. Some resources are listed below:

- [Python for Everybody by University of Michigan](#)
- [Learn Python 3 by Codecademy](#)
- [Intro to Python](#)
- [Python Beginners Guide for Materials Engineers](#)
- [Python Beginners Guide for Programmers by Python Software Foundation](#)

If you have no prior programming experience, you may wish to start with the [Python Beginners Guide for Non-programmers by Python Software Foundation](#).

[100 SECONDS OF PYTHON](#)

[What is Python?](#)

BO Tools

Use of the tools listed on this page is not a requirement. A diverse set of packages and implementations is encouraged. Likewise, multiple teams using the same package is not a problem, in part because implementations can remain private during the course of the hackathon. If you'd like to see a specific tool listed here, please navigate to the "improve this page" link at the bottom of the page and open a pull request. See also the [Acceleration Consortium's curated list of optimization tools](#).

BayE

[BayE](#)

[Bayesian Back End](#)

This hackathon is co-organized by Merck KGaA. You may consider using an internal Merck tool that was recently open-sourced, called the [Bayesian optimization BackEnd \(BayE\)](#). This tool focuses on real world experimental campaigns in the physical sciences with support for integrating domain knowledge through custom parameter encodings and built in chemical encodings. In addition, it allows for custom surrogate models, mixed variable types, and transfer learning.

Figure 3. A snapshot of resources listed on the hackathon webpage such as hackathon orientation, intro to BO, and a Python refresher assignment. These resources prepared participants to maximize their time during the two-day synchronous portion of the hackathon and helped level the playing field for participants with varied backgrounds and skill levels. Image is blurred per preprint server policy.

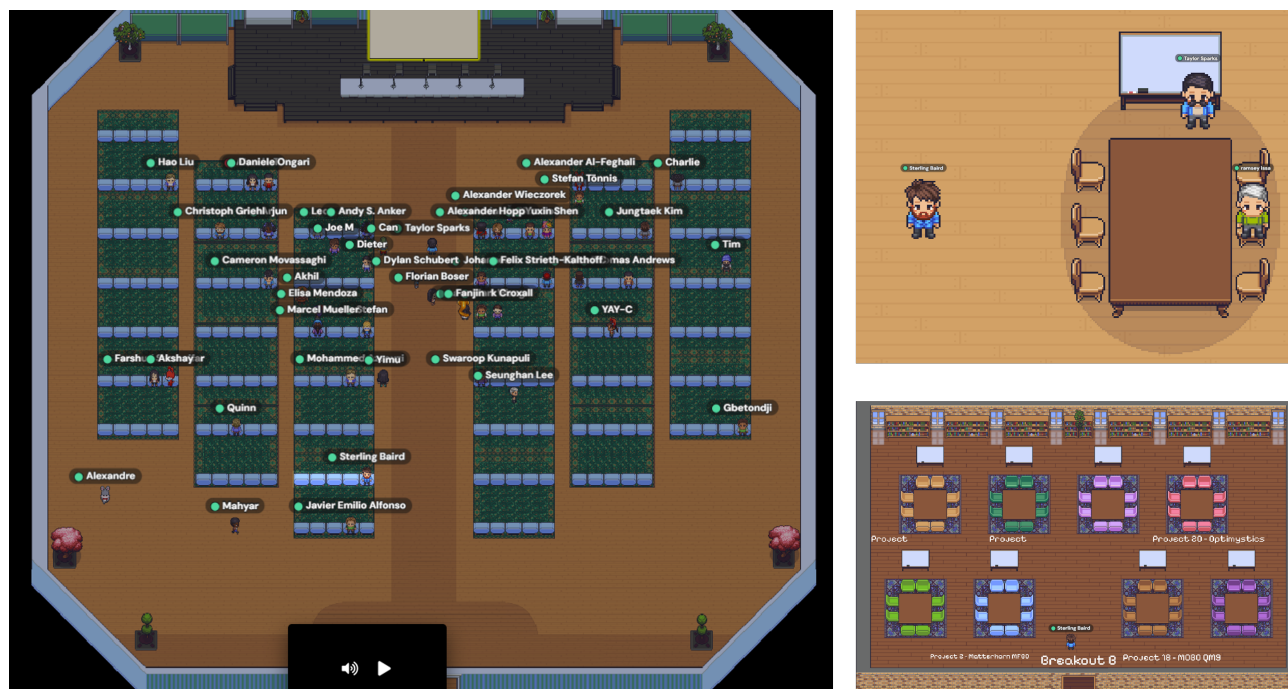







































































































Figure 4. Gather town keynote room (left), custom avatars (top-right), and an example of a breakout room for teams (bottom-right). Keynotes were broadcasted in realtime to participants via an embedded YouTube livestream. Use of Gather Town helped level the playing field for teams who were in physically separate locations and made it easier for facilitators and other teams to have more natural “check-ins” with other projects.



Figure 5. The synchronous portion of the hackathon concluded with a poster session and community judging. One participant noted that “it almost felt like a real poster session.”

Table I: List of projects with links to GitHub, Social Media, and Video.
Project pages on the hackathon website are available at <https://ac-bo-hackathon.github.io/projects/>.

Proj. #	Project Name	Links
1	Multi-objective Benchmarking of Dragonfly against BoTorch	  
2	Long-run Behaviour of Multi-fidelity Bayesian Optimisation	  
3	Take Your Time - Improving Optimization Performance Through Greater Investment in ACQF Optimizer Runtime	 
4	SimpleGPT-BO, Simplified GPT-Powered Bayesian Optimization	  
5	Comparing Bayesian Optimization Methods Across Multiple Hyperparameters Against Simulated “Human” Decision-making	 
6	Multi-Objective Bayesian Optimization for Transparent Electromagnetic Interference Shielding with Thin-Film Structures	  
7	BayBE One More Time - Exploring Corrosion Inhibitors for Materials Design	  
8	BO for Drug Discovery-What is the role of molecular representation?	  
9	Optimal MOF Selection for CO2 capture using Thompson sampling	 
10	Navigating the black box of zeolite synthesis with Bayesian Optimization	  
11	BlendDS - An intuitive specification of the design space for blends of components	  
12	Robust GPs for Sustainable Concrete via Bayesian Optimization	  
13	Interpretability of Bayesian Optimisation Campaigns	 
15	Adaptive Batch Sizes for Bayesian Optimization of Reaction Yield	 
16	BOPE-GPT, Preference Exploration with the curious AI chemist	  
17	Comparative Analysis of Acquisition Functions in Bayesian Optimization for Drug Discovery	  
18	Investigation of Multi-Objective Bayesian Optimization of QM9 Dataset	  
20	Closed loop optimization of hydrogel formulations using dynamic light scattering	  
21	Benchmarking Molecular Descriptors with Actively Identified Subsets (MolDAIS)	 
22	Chemical Similarity-Informed Earth Mover’s Distance Kernel Bayesian Optimization for Predicting the Properties of Molecules and Molecular Mixtures	 
23	Reliable Surrogate Models of Noisy Data	
24	ScattBO Benchmark - Bayesian optimisation for materials discovery	  
25	Bayesian Optimized De Novo Drug Design for Selective Kinase Targeting	  
26	Multiple-Context Bayesian Optimization	  
27	How does initial warm-up data influence Bayesian optimization in low-data experimental settings?	  
28	The Impact of Dataset Size on Bayesian Optimization, Insights from the QM9 Dataset	 
30	Active learning for voltammetry waveform design	  
31	A tutorial on ask/tell mode for Ax	  
32	Efficient Protein Mutagenesis using Bayesian Optimization	  
33	Bayesian Optimization for Hyperspectral Co-heritability Search	  
34	Streamlining Material Discovery - Bayesian Optimization in Thermal Fluid Mixtures	 
35	Tutorial for GAUCHE - A Library for Gaussian Processes in Chemistry	  
36	Scalable Nonmyopic Bayesian Optimization in Dynamic Cost Settings	  
37	The Effects of Post-Modelling Performance Metric Computation on the Efficiency of Bayesian Optimizers	  
38	Bayesian methods in symbolic regression	 
39	Divide and Conquer - Local Gaussian Processes to design Covalent Organic Frameworks for Methane Deliverable Capacity	  
40	Optimizing Chemical Reaction Conditions with Multi-Agent Systems Using LLM and BO	  
41	Retrieval Augmented Bayesian Optimization	  
43	Bayesian Optimization Awesome List	 
44	Rank-based Bayesian Optimization	  
45	Bayesian Optimization for generality	 

Continued on the next page

Table I: List of projects with links to GitHub, Social Media, and Video. Project pages on the hackathon website are available at <https://ac-bo-hackathon.github.io/projects/>.

Proj. #	Project Name	Links
---------	--------------	-------

III. Projects' Key Findings

This section provides a comprehensive summary and highlights the key findings from all project submissions. To streamline the evaluation process, all YouTube video submissions were transcribed and analyzed using an AI agent powered by Anthropic's *claude-3-5-sonnet-20240620* with a temperature of 0.3, ensuring consistent and accurate processing. Each transcript was then manually edited to ensure a high-quality manuscript compilation. This automated approach enhances efficiency while maintaining a structured and objective assessment of the submissions.

Project 1: Multi-objective Benchmarking of Dragonfly against BoTorch

This research project compared the performance of two BO packages, Dragonfly [5] and BoTorch [6], in a multi-objective optimization context. The researchers optimized the Branin-Currin test function with added noise, conducting 20 trials of BO over different random seeds. Each trial used a batch size of 1 and consisted of 20 iterations, evaluating 20 candidates per iteration. The results showed that BoTorch's acquisition functions, particularly the q-noisy expected hypervolume improvement (qNEHVI) [7], outperformed Dragonfly's acquisition function. The qNEHVI acquisition function achieved the lowest log hypervolume difference on average across the trials, while Dragonfly's acquisition function struggled to surpass a randomly generated set of candidates. Visualization of the average Pareto front produced by different acquisition functions further demonstrated BoTorch's superiority, with its acquisition functions covering a larger portion of the Pareto front compared to Dragonfly. The researchers concluded that BoTorch outperformed Dragonfly within the scope of their study, but noted the need for further investigation using different batch sizes, input and output constraints, and test functions simulating chemical reactions to apply the findings to material design problems.

Project 2: Long-run Behaviour of Multi-fidelity Bayesian Optimisation

This research project investigates the long-term performance of multifidelity Bayesian optimization (MFBO) methods [8] compared to single-fidelity Bayesian optimization (SFBO) methods. The study focuses on scenarios where two fidelities are available: an accurate but expensive target fidelity that fully represents the objective, and a less accurate but cheaper low fidelity. The researchers observed that while MFBO methods initially

outperform SFBO up to a budget of approximately 20, they begin to underperform beyond this point. This behavior is particularly evident with linear kernel MFBO methods, where the performance gap becomes more pronounced after a cost of around 60. The project aims to evaluate the factors contributing to the long-term issues of MFBO methods. Specifically, the researchers plan to investigate the impact of kernel choice, frequency of low fidelity queries, and acquisition function selection on MFBO performance. By analyzing these aspects, the study seeks to identify potential improvements for MFBO methods to enhance their applicability in real-life optimization tasks. The ultimate goal is to develop MFBO approaches that can maintain their performance advantage over SFBO methods in the long run, leveraging the cost-effectiveness of low fidelity queries while achieving superior optimization results.

Project 3: Take Your Time - Improving Optimization Performance Through Greater Investment in ACQF Optimizer Runtime

This research project explores an enhancement to standard BO campaigns by introducing a random seed approach to improve optimization outcomes. The team from the University of Utah's Material Science and Engineering department investigated the variability in BO campaigns when starting from the same initial data points. The key innovation in this work is a custom approach to calculating the acquisition function. Instead of computing it once per iteration, as is standard in most BO methods [9, 10], the researchers calculate the acquisition function multiple times using different random seeds. They then select the highest-performing result at each iteration. This method, termed "random retries Optimizer", showed significant improvements in optimization performance, particularly in avoiding local optima. The researchers demonstrated that this approach consistently achieves near-optimal performance among the possible campaigns stemming from the initial data points. While this method requires additional computational time for the multiple acquisition function calculations, the performance gains appear to outweigh the increased computational cost. The team is continuing to investigate questions such as the trade-off between additional compute time and performance improvements, as well as how the method performs across different types of optimization problems, including simpler ones compared to the challenging example presented.

Table II. Ranked Projects with Team Names and Prize Distribution. To avoid incentivizing single-person teams and very large teams, both per-person and per-team limits were imposed (e.g., teams of 4 or more would have the max per-team amount divided equally rather than receive the max per-person amount).

Rank	Proj. #	Team Name	Project Name	Prize* (CAD)
1st	#23	Noisy Nerds	Reliable Surrogate Models of Noisy Data	300 (1000 max)
2nd	#34	BOMS Prob	Streamlining Material Discovery - Bayesian Optimization in Thermal Fluid Mixtures	150 (500 max)
3rd	#7	Surface Science Syndicate	BayBE One More Time - Exploring Corrosion Inhibitors for Materials Design	75 (250 max)
4th	#5	KLM	Comparing Bayesian Optimization Methods ... Against Simulated "Human" Decision-making	40 (125 max)
5th	#8	Molecular Representation	BO for Drug Discovery - What is the role of molecular representation?	40 (125 max)
6th	#9	PME No Hikari	Optimizing The CO2 Adsorption Capacity of Metal-Organic Frameworks Using Thompson Sampling	40 (125 max)
7th	#11	BlendDS	BlendDS - An intuitive specification of the design space for blends of components	40 (125 max)
8th	#30	SERO Opt	Active learning for voltammetry waveform design	40 (125 max)
9th	#43	General Optimizers	Bayesian Optimization for Generality	40 (125 max)
10th	#3	Sparks Group	Take Your Time - Measuring Optimization Performance as a Function of ACQF Optimizer Runtime	40 (125 max)

Table III. Project Topics for the Hackathon. See [the submission page](#) for more details.

Topic	Description
1 Apply Algorithms	Choose an algorithm and apply it to a hackathon benchmark task
2 Develop Benchmarks	Develop a new benchmark and add it to a suite of benchmarks
3 Create Tutorials	Create "gentle introduction" tutorials for advanced optimization topics
4 Propose Tasks	Propose materials tasks that <i>can</i> and <i>should</i> be tackled with BO
5 General	Other projects related to BO for the physical sciences

Project 4: SimpleGPT-BO, Simplified GPT-Powered Bayesian Optimization

SimpleGPT-BO is a ChatGPT-integrated tool that makes Bayesian Optimization accessible without installation requirements. The system provides guided optimization through multiple acquisition functions (PI, EI, UCB, TPE) and supports multi-objective optimization via weighted single objective transformation. Users access the tool directly through ChatGPT's interface or via provided URLs, benefiting from pre-generated instructions and global availability. While designed for straightforward optimization tasks rather than complex analyses, the open-source project (MIT License) welcomes community contributions to expand functionality and appli-

cations. The tool aims to democratize Bayesian Optimization by eliminating technical barriers and providing step-by-step guidance for users across all experience levels.

Project 5: Comparing Bayesian Optimization Methods Across Multiple Hyperparameters Against Simulated "Human" Decision-making

This research project investigated the effectiveness of BO methods that mimic human experimentalist approaches in materials science [11], comparing them to traditional BO techniques and random search baselines [10, 12]. The study focused on three key axes: model complexity, number of features, and acquisition

function. The researchers found that a simplified “human experimentalist” approach, using automated feature engineering with a linear model, three features, and an exploitative acquisition function, was competitive with more complex random forest models in terms of enhancement factor and acceleration factor. Both the simplified approach and random forest models outperformed random search. Interestingly, for linear models, increasing complexity through additional features or more exploratory acquisition functions led to decreased performance. The main conclusion was that low-dimensional, interpretable models were comparable to traditional BO methods for the datasets and hyperparameter regimes studied. The researchers suggest future work could involve comparing their results to actual human performance and improving their methodology using theoretical approaches.

Project 6: Multi-Objective Bayesian Optimization for Transparent Electromagnetic Interference Shielding with Thin-Film Structures

This research project focuses on applying multi-objective BO to design transparent electromagnetic interference (EMI) shielding using thin film structures [13]. The primary objectives are to maximize both transmittance and shielding effectiveness simultaneously. The optimization problem involves selecting materials and thicknesses for each layer of the thin film structure, with a search space comprising 12 material choices (Ag, Al, Al₂O₃, Cr, Ni, Pd, Si₃N₄, SiO₂, Ti, TiN, TiO₂, W) and thickness ranges from 5 to 20 nm. The researchers implemented multi-objective BO using random scalarization for the acquisition function via BayesO package [14]. Gaussian process regression and expected improvement were employed as part of the BO framework. The results demonstrate that the multi-objective BO approach effectively identified the Pareto frontier for the two objectives of transmittance and shielding effectiveness. This suggests that the method is successful in finding optimal trade-offs between transparency and EMI shielding performance for thin film structures, which could have applications in areas such as spacecraft windows where both properties are crucial.

Project 7: BayBE One More Time - Exploring Corrosion Inhibitors for Materials Design

This research project explored the use of BO for efficiently identifying effective corrosion inhibitors in materials design. The researchers utilized the BO implementation from the BayBE package [15] to evaluate its performance on multiple experimental datasets [16] involving different alloys, primarily aluminum-based. The study found that the choice of molecular descriptor encoding significantly impacted the optimization performance, especially for larger datasets like A1000 and A2024. The SMILES-based approach with Mordred encoding [17] outperformed the other methods for these datasets, while random sampling underperformed. Inter-

estingly, for the A775 dataset, which had a sparse distribution of reported efficiencies and fewer data points, this trend was not observed. The researchers also explored transfer learning capabilities, using information from the A2024 alloy dataset to inform the optimization process for the A1000 dataset. This transfer learning approach outperformed the uninformed method after only 15 iterations, demonstrating its potential for accelerating the discovery of effective corrosion inhibitors in experimental settings.

Project 8: BO for Drug Discovery-What is the role of molecular representation?

This research project investigated the impact of molecular featurization methods on BO performance for guiding molecular experiments. The team explored alternatives to the commonly used connectivity-based fingerprints, aiming to determine which featurization techniques yield the best results in BO for molecular discovery. The study found that without modifying the features, MACCS [18] and RDKit [19] featurization methods outperformed the default options presented in previous publications. To address the high dimensionality of molecular features, the researchers employed specialized Gaussian processes and explored random forest surrogates. However, the random forest approach did not improve performance and required more computational resources. Principal Component Analysis (PCA) was successfully used to enable Gaussian processes to handle large molecular featurizations. Notably, physicochemical featurizations like RDKit and Mordred [17] outperformed the previous benchmark winner, MolDQN [20]. The team’s findings were consistent with benchmarks in the BO library but suggested that RDKit could produce even better results than Mordred. Overall, the research highlights the importance of careful featurization selection in BO for molecular discovery, with RDKit emerging as a particularly effective option.

Project 9: Optimal MOF Selection for CO₂ capture using Thompson sampling

This research project focuses on optimizing the selection of Metal-Organic Frameworks (MOFs) for carbon capture applications using BO, specifically Thompson Sampling. MOFs are nanoporous materials with high potential for carbon capture [21, 22], but their synthesis and characterization are expensive and time-consuming [23]. The goal was to develop an efficient method for identifying high-performing MOF candidates using a small dataset. The team employed a Gaussian Process model trained on the CRAFTED dataset [24], using both RACs features and geometric features of MOFs, with CO₂ uptake as the output. Thompson Sampling was used as the acquisition function to select the next best sample based on the posterior distribution, inherently performing a Bayesian update. The results showed that Thompson Sampling was twice as efficient as random sampling in identifying high-performing MOF candidates. This

approach is notable for being the first application of Thompson Sampling to MOF candidate selection, and it requires no hyperparameter tuning, making it easily transferable and cost-efficient. The potential impact of this method is significant, as it could accelerate MOF design for direct air capture applications, with a goal of absorbing 30 million tons of CO₂, and could also be applied to post-combustion capture scenarios.

Project 10: Navigating the black box of zeolite synthesis with Bayesian Optimization

This research project explores the application of BO to zeolite synthesis, a complex process with significant industrial importance. Zeolites are crystalline materials composed of interconnected silicate and aluminate tetrahedra, widely used as absorbents and catalysts [25]. The synthesis of zeolites involves multiple parameters, including silicon and aluminum sources, organic molecules, water, temperature, and time. The process aims to achieve specific properties such as high crystallinity, large external surface area, and particular silicon-to-aluminum ratios, while also considering economic factors like synthesis temperature, duration, and ingredient concentrations [26]. The project presents a GitHub repository containing an introductory text on zeolites and their synthesis, along with a Jupyter notebook demonstrating BO applications. The notebook is divided into two main sections: the first optimizes an analytical dummy function using zeolite synthesis parameters, exploring various aspects of BO including continuous, categorical, and mixed variable types, parameter constraints, and single and multiple objectives. The second section applies BO to propose new experiments based on existing literature data. This approach aims to accelerate the traditionally time-consuming trial-and-error process of zeolite synthesis optimization, potentially reducing costs and improving efficiency in industrial applications.

Project 11: BlendDS - An intuitive specification of the design space for blends of components

This project focuses on developing an interface for BO that bridges the gap between domain experts (i.e. chemists and materials scientists) and machine learning algorithms. The system allows scientists to specify experimental parameters and constraints in natural language, which is then translated into a structured dictionary format by a large language model. This dictionary is subsequently converted into a Python object that can be used for design of experiments and optimization tasks. The key feature of this interface is its ability to generate diverse trial designs for efficient sampling of the experimental space. It incorporates dimensionality reduction techniques to visualize the design space, enabling users to select the most diverse and informative trials. This approach aims to maximize the information gained from each experiment, potentially reducing the number of trials needed to reach optimal results. The project is open-source, inviting contributions and integration with

existing BO frameworks.

Project 12: Robust GPs for Sustainable Concrete via Bayesian Optimization

This project introduces Robust Gaussian Processes (Robust GPs) [27] to the concrete sciences experiments with Bayesian optimization campaigns. The research addresses the common issue of outliers in concrete science experiments by applying Robust GPs to account for these anomalies. The project provides a tutorial and accompanying code in a GitHub repository, demonstrating how to implement Robust GPs in this specific context. The study utilizes a dataset provided by Meta [28], which includes input variables and a target variable representing concrete strength. Initial data analysis reveals potential outliers in the strength variable, with values between 2 and 2,000. The researchers then apply Robust GP to this dataset, achieving good predictive performance. The project culminates in demonstrating how to obtain Bayesian recommendations using the Robust GP as a surrogate model, providing a practical approach for optimizing concrete mixtures while accounting for experimental outliers.

Project 13: Interpretability of Bayesian Optimisation Campaigns

This research project focused on developing novel methods for interpreting Bayesian Optimization campaigns by incorporating the temporal component often neglected in end-of-campaign analyses. The study utilized data from a self-driving lab experiment aimed at optimizing the conductivity of a coating. The researchers recreated the Gaussian Process (GP) model employed in the original experiment, training it on subsets of data to simulate mid-campaign conditions. Two primary methods were investigated: First, cross-sections of the GP model were taken at various stages of the campaign, holding all but one feature constant. This approach revealed that closer nozzle distances produced better results, aligning with prior beliefs and serving as a potential mid-campaign sense check to validate model alignment with previous experimentation. The second method involved predicting the error of the next sample based on the model trained on data collected up to that point. This analysis showed that performance improvement plateaued around the time the optimal sample was produced, suggesting its potential use as an early stopping criterion if error is monitored throughout the campaign. The research demonstrates the value of incorporating temporal analysis in Bayesian Optimization campaigns, offering insights that could enhance decision-making during experiments. The cross-sectional approach provides a means to validate model behavior against prior knowledge, while the error prediction method could inform stopping criteria, potentially improving efficiency in optimization processes. These techniques offer promising avenues for real-time interpretation and guidance in Bayesian Optimization experiments,

particularly in self-driving lab contexts.

Project 15: Adaptive Batch Sizes for Bayesian Optimization of Reaction Yield

This research project investigates the optimization of batch sizes in BO for chemical reaction yield optimization. The study focuses on determining the optimal batch size (Q) at each step of the BO process, considering both smaller and larger batch sizes. The researchers examine how Q impacts the cost of retraining models, such as large language models, and how it relates to additional experimentation overhead. The project findings indicate that the optimal Q depends on a trade-off between model retraining time and additional experimentation overhead. For a fixed batch size, the researchers observed two minima at $Q=3$ and $Q=7$, reflecting the relative importance of Q in different scenarios. With very large batch sizes, the retraining cost had minimal impact on the overall experimental overhead, while experimentation overhead had a significant influence. The study also explored adaptive batch sizing strategies, decreasing batch sizes when performance was good and increasing them as the surrogate model became more trustworthy. The researchers conclude that batch sizes have a crucial effect on BO efficiency in real-world settings, and adaptive batch sizes could effectively balance the trade-off between model retraining and batch sampling effectiveness.

Project 16: BOPE-GPT, Preference Exploration with the curious AI chemist

This research project focused on optimizing the Fischer-Tropsch synthesis process [29], which converts syngas into biofuel hydrocarbons. The optimization problem on the dataset [30, 31] involved four input variables (space time, syngas ratio, temperature, and pressure) and four output variables (carbon conversion, selectivity, methane to paraffins, and olefins). The researchers proposed a novel approach using preferential BO [32] with large language models (LLMs) to prioritize outputs based on stakeholder preferences. The methodology employed the typical BO process powered by BoTorch, using Expected Utility as the acquisition function. However, the innovation lay in using an LLM (specifically GPT-4) to perform pairwise comparisons for output prioritization. The results showed that outputs exhibited well-defined optima trade-offs and some degree of monotonicity, which was important for preferential BO. The researchers tested three objective scenarios: optimizing all outputs equally, maximizing only CO conversion, and maximizing three objectives while minimizing olefins. For the first two objectives, the LLM-based preferential BO performed similarly to traditional utility function-based methods. However, for the third, more complex objective, the LLM approach performed worse than the utility function method and only slightly better than random exploration, suggesting limitations in the LLM's ability to handle complex optimization scenarios. The project also included the development of an interactive

application for visualizing BO results and running live optimizations.

Project 17: Comparative Analysis of Acquisition Functions in Bayesian Optimization for Drug Discovery

This project investigates the application of BO techniques directly on molecular fingerprints for drug discovery, focusing on comparing different surrogate models, acquisition functions and small, diverse, unbalanced, and noisy datasets [33]. The researchers used the LD50 dataset from PTDC with SMILES encodings [34, 35], which were transformed into ECFP fingerprints [36], creating a 2048-dimensional feature space. Two surrogate models were examined: Gaussian Processes (GP) and Random Forests (RF). The results showed that in this high-dimensional space, Random Forests with uncertainties evaluated as variance between different trees achieved expected performance across all acquisition functions, significantly outperforming random selection. In contrast, Gaussian Processes failed to perform well, likely due to the challenges posed by the high-dimensional space (22048 possible combinations) and the relatively small dataset of only 7,000 data points. The researchers concluded that machine learning models, particularly Random Forests with various acquisition functions, perform well when dealing with high-dimensional molecular fingerprint spaces, demonstrating their potential for drug discovery applications with certain selection biases.

Project 18: Investigation of Multi-Objective Bayesian Optimization of QM9 Dataset

In this work, the team explores the application of multi-objective Bayesian optimization (MoBO) techniques to molecular property optimization using the QM9 dataset [37]. The project implements and compares state-of-the-art MoBO algorithms—including Expected Hypervolume Improvement (EHVI) [38] and ParEGO—to [39] efficiently navigate the trade-offs between multiple molecular objectives such as HOMO-LUMO gap, molecular weight, and dipole moment. Leveraging molecular feature encodings and Gaussian process surrogate models, the optimization framework identifies candidate molecules that balance competing objectives in chemical design. This study highlights the potential of MoBO to accelerate lead discovery by providing a principled, sample-efficient method for exploring high-dimensional molecular spaces.

Project 20: Closed loop optimization of hydrogel formulations using dynamic light scattering

This research project proposes a self-driving lab approach to optimize hydrogel formulations using BO and automated characterization techniques. The study aims to address the challenge of creating hydrogels with specific properties for applications such as cell culture, tissue engineering, drug delivery, and agriculture. The proposed workflow utilizes an open-source liquid handling

robot to mix hydrogel formulations in a 96-well plate, followed by cross-linking using various methods (LED array, heating module, or time). The key innovation is the use of Dynamic Light Scattering (DLS) [40] for automated characterization of gelation and viscoelastic properties, which can measure gel stiffness up to 10 kilopascals. This DLS method is implemented using a plate reader specifically designed for 96-well plates, enabling high-throughput analysis. The resulting property data is then fed into a Bayesian Optimizer to determine the next set of formulations and processing parameters to test. This approach aims to efficiently explore the complex relationship between hydrogel formulation parameters and their resulting properties, potentially accelerating the development of custom hydrogels for specific applications in biological research and improving data availability and reliability in the field.

Project 21: Benchmarking Molecular Descriptors with Actively Identified Subsets (MolDAIS)

This research presents a novel approach called MOLDES (Molecular Descriptors with Actively Identified Subspaces) for molecular property optimization. The method addresses the challenge of optimizing molecules in high-dimensional spaces by using molecular descriptors - sets of rotationally and translationally invariant calculations performed on molecular graphs - coupled with active subspace identification. MOLDES employs a sparse axis-aligned subspace Gaussian Process prior, which actively learns an encoding while performing Bayesian optimization. Recent works [41, 42] are increasingly turning towards active encoding of molecular feature spaces. The researchers evaluated MOLDES on three case studies: experimental lipophilicity (4,200 compounds), log P optimization benchmark (250,000 molecules), and power conversion efficiency from the Harvard Clean Energy Project (30,000 compounds). In all cases, MOLDES demonstrated superior performance compared to other optimizers, particularly in larger datasets. For the log P optimization, MOLDES consistently found the optimal molecule within 100 iterations. The method also showed strong performance in constrained optimization problems, often achieving the best-case scenario and maintaining a favorable worst-case scenario compared to other methods. Overall, MOLDES proved efficient in identifying high-performing molecules in low-data regimes, offering a promising approach for molecular property optimization tasks.

Project 22: Chemical Similarity-Informed Earth Mover's Distance Kernel Bayesian Optimization for Predicting the Properties of Molecules and Molecular Mixtures

This research project focuses on developing chemical similarity-informed distance functions and kernels for explainable Bayesian optimization, specifically targeting the prediction of properties for molecular mixtures. The researchers propose a novel approach that bypasses the

need for embedding vectors by directly providing pairwise distances between data points in the kernel function of a Gaussian Process (GP) model [43]. The project introduces the Earth Mover's Distance (EMD) kernel [44] into the GP framework to calculate pairwise distances between mixtures based on individual component distances. This method was tested for predicting yields of binary reactant mixtures, demonstrating high chemical resolution in mixture analysis. The results show that the EMD kernel achieves accurate yield predictions with narrow distributions for both high and low-yield cases, indicating improved performance in distinguishing between different mixture compositions. By incorporating smooth distance metrics, the researchers successfully extended Bayesian optimization techniques from pure components to molecular mixtures, potentially enhancing the efficiency and interpretability of materials property prediction in complex chemical systems.

Project 24: ScattBO Benchmark - Bayesian optimisation for materials discovery

This project presents ScattBO, a Python-based benchmark that simulates a self-driving laboratory (SDL) for materials discovery. A self-driving laboratory is an autonomous platform that conducts machine learning-selected experiments to achieve a user-defined objective, such as synthesizing a specific material [45]. The benchmark addresses the challenge that such SDLs can be expensive to run, making intelligent experimental planning essential, while only a few people have access to real SDLs for materials discovery. ScattBO provides an *in silico* simulation of an SDL where, based on synthesis parameters, the benchmark 'synthesizes' a structure, calculates the scattering pattern [46], and compares it to the target structure's scattering pattern. The benchmark acknowledges that scattering data may not be sufficient to conclusively validate that the target material has been synthesized [47], but can include other types of data as long as they can be simulated. This makes it currently challenging to benchmark Bayesian optimization algorithms for experimental planning tasks in SDLs, and ScattBO fills this gap by providing an accessible simulation environment.

Project 25: Bayesian Optimized De Novo Drug Design for Selective Kinase Targeting

This project focused on incorporating Bayesian optimization to guide de novo drug design, specifically targeting growth factor receptors for cancer therapeutics. The team built upon the DOCKSTRING paper, Python library, and dataset [48], using a Gaussian process with a Matérn kernel on Morgan fingerprint representations. They employed a graph genetic algorithm to generate SMILES strings guided by the Bayesian optimization output. The researchers explored both selective and promiscuous binding scenarios. For selective binding, they optimized for binding to FGFR1 while penalizing overbinding to other growth factor receptors relative to

their median. For promiscuous binding, they maximized the maximum binding affinity across multiple receptors. They found that a sigmoidal penalty function was more effective than simple absolute differences when optimizing against multiple proteins. The team also incorporated a drug-likeness measure (QED) [49] as a penalty in the optimization process, though its effect was limited. Due to time and resource constraints, the project was unable to extensively explore the chemical space or use more accurate binding affinity calculations beyond docking. The authors suggest that future work could incorporate known unknowns through an evasion process, further optimize selective binding, and compare different molecular representations.

Project 26: Multiple-Context Bayesian Optimization

This project focused on exploring multiple context Bayesian optimization using the BayBE code [15], a Bayesian backend package built on BoTorch with additional features. The primary aim was to investigate transfer learning capabilities by incorporating data from existing campaigns into new optimization tasks. The researchers examined both analytical functions (like the Hartmann 3D function) and real-world data (direct arylation reaction dataset) to assess the effectiveness of transfer learning in Bayesian optimization. They introduced noise, scaling, shifting, and negation to the analytical functions to simulate realistic scenarios. The results demonstrated that using a small percentage (1-25%) of existing data significantly improved optimization performance compared to the baseline without transfer learning. Interestingly, they found that using larger amounts of data (50-100%) did not necessarily lead to better results and in some cases performed worse than the baseline. The team also explored clustering the existing data and using cluster centroids as source data, which proved effective. For the real-world direct arylation dataset, they analyzed correlations between different reaction temperatures and observed that small amounts of transfer data (1-10%) substantially improved the Bayesian optimization campaign. Overall, the project highlighted the potential of transfer learning in Bayesian optimization while emphasizing that careful selection of the amount and type of transfer data is crucial for optimal performance.

Project 27: How does initial warm-up data influence Bayesian optimization in low-data experimental settings?

This research project investigated the influence of warm-up sampling methods and dataset sizes on property optimization in low data regimes, specifically focusing on molecular property prediction. The team used the QM9 dataset [50] and selected band gap as the optimization target. They compared two chemically-inspired sampling methods for the warm-up dataset: Morgan fingerprints and MolFormer language model fingerprints. The researchers also referenced the GDB-17 chemical universe database [51] in their background work. The researchers

performed dimensionality reduction on the fingerprints using PCA, projecting them into a 2D space for sampling. They conducted experiments to analyze how the warm-up dataset size affects optimization results. The most significant finding was the comparison between Morgan fingerprints and MolFormer fingerprints at a constant data regime of 50 data points. The results showed that MolFormer fingerprints substantially outperformed Morgan fingerprints, suggesting that pre-trained models on large chemical spaces can potentially improve model optimization rates. This study aims to initiate broader discussions on how dataset sizes and sampling methodologies impact final optimization tasks in molecular property prediction.

Project 28: The Impact of Dataset Size on Bayesian Optimization, Insights from the QM9 Dataset

The Chihuahuas team investigated the critical dataset size threshold required for effective Bayesian Optimization of Graph Convolutional Neural Networks (GCNNs) using the QM9 molecular dataset [37]. The research systematically examined hyperparameter optimization (learning rate, batch size, number of neurons) across dataset sizes ranging from 200 to 2000 molecules to determine minimum viable data requirements for reliable BO performance. Results demonstrated that BO successfully optimized GCNN hyperparameters even with small datasets (200 molecules), though performance improved with larger datasets, achieving lowest losses at 2000 molecules (0.001134 for learning rate optimization). The study revealed a trade-off between dataset size and computational efficiency, with larger datasets requiring more time to reach convergence despite better performance. The research provides a methodological framework for applying BO in data-scarce scenarios common in computational chemistry and materials science, demonstrating that machine learning optimization remains viable even with limited molecular data.

Project 30: Active learning for voltammetry waveform design

This research project focused on using Bayesian Optimization to optimize voltammetry waveforms for sensor performance testing. The researchers employed the scikit-optimize package [52] to implement a human-in-the-loop optimization process, allowing for pauses between optimization iterations to conduct laboratory experiments. The study encoded a four-step pulse waveform into a continuous parameter space for the optimization model. The process began with six randomly generated initial waveforms, which were tested in the lab to obtain performance metrics. These results were used to initialize a Gaussian process model. The optimization then proceeded in batches, with the model suggesting the next best waveforms to test based on previous results. The researchers conducted four total batches, updating the model after each laboratory testing phase. Progress plots were generated to visualize the improvement in sensor error across batches, demonstrating the effectiveness

of the approach in finding optimal waveform parameters. The study also explored the interactions between model parameters and the performance metric across multiple dimensions, providing insights into the complex relationship between waveform characteristics and sensor performance.

Project 31: A tutorial on ask/tell mode for Ax

This project focuses on utilizing the open-source package AX [53] for Bayesian optimization in an asynchronous, human-in-the-loop experimental setting, particularly applied to voltammetry. The researchers developed custom figures and explanations to elucidate the complex interactions within the AX package, emphasizing its user-friendly nature while acknowledging its potential complexity. The study explores various aspects of the optimization process, including strategy selection, surrogate model choice, and search space definition. It demonstrates two primary approaches: a “cold start” scenario, where optimization begins with no prior data, and a “warm start” scenario for situations with existing experimental data. For the cold start, the researchers showcase the use of sequence generators like Sobol sequences and random generators to initialize the optimization campaign. The project also highlights the package’s ability to handle real-world experimental constraints, such as marking broken electrodes to exclude them from future suggestions. In the warm start scenario, the researchers detail the process of translating existing experimental workflows into the AX framework, emphasizing the package’s automatic tracking of experiments, results, and parameters. The study also notes the inclusion of visualization tools within AX, though these are not described in detail.

Project 32: Efficient Protein Mutagenesis using Bayesian Optimization

This research project explored the application of Bayesian optimization to protein mutagenesis, specifically aiming to optimize protein binding affinity to fentanyl for potential use in biosensors. The team utilized a pre-trained BERT language model [54], which predicts ligand binding to proteins based on amino acid sequences and SMILES strings. The methodology involved using Bayesian optimization to iteratively select positions and amino acids for mutation, starting with a protein that already had affinity to fentanyl [55]. The researchers focused on residues within 5 angstroms of the ligand as potential mutation sites. The team compared their Bayesian optimization approach to a baseline using random mutations. They analyzed the resulting mutations using PyRosetta’s mutagenesis tool to identify potential new positive interactions and clashes that could alter protein conformation. The structural changes were further investigated using AlphaFold [56].

Project 33: Bayesian Optimization for Hyperspectral Co-heritability Search

This research project focuses on applying Bayesian Optimization (BO) to the field of crop genetics, specifically for selecting optimal proxy data to train genomic prediction models. The goal is to identify resilient crops that can withstand factors like pests and climate change. The study uses coheritability as a measure to determine the best proxy traits for desired traits, treating it as a black-box optimization problem. The researchers compared various BO methods using Gaussian Processes (GPs) against random search for four different desired traits. Surprisingly, they found that all BO methods performed similarly to random search. However, a pre-trained probabilistic model with a mixed multitask kernel outperformed standard BO approaches, identifying better proxy data more quickly. The study concludes that Gaussian Processes may not be the optimal choice for this particular problem, challenging the common use of Matérn kernels and GPs as the default in Bayesian Optimization. The researchers also criticize the widespread use of these methods without proper justification. As part of their contribution, the team has released their dataset [57] for further research in this area.

Project 35: Tutorial for GAUCHE - A Library for Gaussian Processes in Chemistry

This research project focuses on implementing input warping for Bayesian Optimization within the Gauche library [58], which was previously developed by the team and published at NeurIPS 2023. The primary innovation of Gauche is the introduction of Gaussian process (GP) kernels that enable modeling of discrete entities such as SMILES strings, graphs, and bit vectors, which are common representations in molecular sciences. The motivation behind using Gaussian processes for Bayesian Optimization is their suitability for automated tasks where fine-tuning for each problem is not feasible. GPs offer a good balance between performance and simplicity, with few trainable hyperparameters that can reliably converge on each iteration of the Bayesian Optimization loop. This makes them particularly attractive as surrogate models compared to more complex alternatives like deep neural networks, which might require careful monitoring during training at each iteration. The Gauche library extends the applicability of GPs to discrete input spaces, allowing for Bayesian Optimization over molecular representations. The project team has developed a range of tutorials and applications, including molecular property prediction, protein fitness prediction, and sparse GP regression, all available in the Gauche GitHub repository.

Project 36: Scalable Nonmyopic Bayesian Optimization in Dynamic Cost Settings

This research project focuses on scalable Bayesian optimization in dynamic settings, addressing limitations of previous approaches that rely on myopic acquisition func-

tions and assume fixed cost structures. The researchers introduce a novel method using non-myopic acquisition functions [59] that incorporate a look-ahead mechanism and dynamic cost functions. The project evaluates the proposed algorithm, named HBE, through two main experimental setups. First, they use synthetic functions across 14 different environments with varying dimensions to test scalability. Second, they apply the method to a real-world protein sequence design problem, aiming to maximize a protein score. The researchers compare their HBE algorithm against six other acquisition functions, including state-of-the-art methods. To enhance practicality, they integrate automatic hyperparameter tuning to reduce the number of optimization parameters. While specific results are not provided in the given context, the approach aims to overcome suboptimal resource allocation in dynamic cost experiments and improve upon existing Bayesian optimization techniques.

Project 37: The Effects of Post-Modelling Performance Metric Computation on the Efficiency of Bayesian Optimizers

This research project explored modifications to the Bayesian Optimization algorithm to enhance its efficiency in solving real-world problems. The key innovation was changing the order of objective calculation and model fitting. Instead of calculating all objectives first and then modeling each objective, the proposed method only models the essential observations (e.g., yield) and calculates the objectives after the modeling process. This approach was tested using a simulation of a real-world system based on kinetics discovered by Hornel, employing a plug flow reactor. The optimization aimed to maximize both space-time yield and E factor for the predicted yield. The results demonstrated significant time savings by computing objectives after modeling, primarily because only a single model for yield needed to be fitted instead of separate models for space-time yield and E factor. While computational performance improved, there was negligible impact on convergence speed. Interestingly, the research suggests that this new approach might potentially improve the efficiency of selection criteria by allowing the fitting of a prior distribution biased towards regions with more potential based on the objective function's structure. In summary, the study found that fitting objectives after modeling can lead to substantial increases in computational performance, with the possibility of also improving hypervolume convergence efficiency.

Project 38: Bayesian methods in symbolic regression

This research project focused on enhancing symbolic regression methods for physical science data sets using Bayesian optimization techniques. The researchers explored two main approaches to improve the performance of symbolic regression algorithms. First, they investigated the impact of assigning varying importance to individual data points under data-limited conditions. Using the Optuna framework, they implemented a tuning process

for data point importance, which resulted in improved predictive performance on a benchmark set compared to treating all points equally. This approach could be particularly useful when dealing with experimental data of uncertain quality. Second, they integrated prior knowledge from Wikipedia's known mathematical equations into the loss function, combining it with the Bayesian Information Criterion to statistically favor established symbolic structures. This method was implemented as a fitness function in the Symbolic Regression Julia library. While the researchers faced challenges in evaluating results due to numerical instability, both approaches showed promise in improving symbolic regression performance by incorporating uncertainty and prior information. The team suggests that future work should adhere more closely to the original paper's methodology, using the Bayesian information prior differences between equations as a probability metric for accepting mutational operations in the symbolic regression process.

Project 39: Divide and Conquer - Local Gaussian Processes to design Covalent Organic Frameworks for Methane Deliverable Capacity

This research project focuses on improving Bayesian Optimization (BO) for high-dimensional, large-scale datasets, specifically applied to the design of Covalent Organic Frameworks (COFs) for methane storage [60]. The researchers developed a novel approach combining unsupervised clustering with local Gaussian Process (GP) models to enhance BO efficiency in the high data regime. The method begins by using K-means clustering to partition the dataset into distinct clusters. A portion of data from each cluster is then sampled to train separate local GP models. An epsilon-greedy algorithm is employed to determine which GP to train next. The researchers applied this approach to a COF dataset [61] containing over 70,000 2D and 3D structures, assembled in silico from 666 organic linkers and four synthetic routes. The objective was to maximize methane storage performance, measured as deliverable capacity. The results demonstrated that their divide-and-conquer approach with local GP surrogates significantly outperformed a single GP model. While the single GP model reached a maximum deliverable capacity of 2.74 in 60 iterations, the proposed method surpassed this maximum within 5-10 iterations. Both methods retrained the GP model after 20 iterations and used 5% of the data for initial surrogate model training. This research highlights the potential of using local GP surrogates in combination with unsupervised clustering to perform more efficient Bayesian optimization in high-dimensional, large-scale datasets.

Project 40: Optimizing Chemical Reaction Conditions with Multi-Agent Systems Using LLM and BO

This research project focuses on optimizing Suzuki reaction conditions using Bayesian optimization enhanced

by a multi-agent large language model (LLM) system. The target dataset consists of Suzuki reactions, with the goal of optimizing reaction ligands, bases, and solvents. The researchers observed that while the target dataset contains unique combinations of conditions, many of the ligands are similar to those in existing reaction data points, suggesting potential for knowledge transfer. The developed system employs a multi-agent LLM approach comprising three domain expert agents and a reaction optimization agent. The process involves four steps: (1) the reaction optimization agent assigns tags to domain agents, (2) gathers feedback from them, (3) uses this feedback to warm-start initial data point generation and evaluation, and (4) acquires new data points based on existing data. In demo experiments, the multi-agent LLM system demonstrated improved performance, reaching target reaction yields with fewer iterations compared to traditional methods. Additionally, the multi-agent approach showed an ability to avoid local maxima traps, which is particularly valuable given the challenging nature of reaction optimization due to data sparsity. This research highlights the potential of integrating LLM-based multi-agent systems with Bayesian optimization techniques to enhance reaction condition optimization in chemistry.

Project 41: Retrieval Augmented Bayesian Optimization

This research project, called RAMBO, combines retrieval-augmented generation with Bayesian optimization to enhance the initial point selection for new optimization tasks. The system leverages existing data from literature or internal databases to identify optimal starting points for Bayesian optimization processes, particularly in the context of chemical reactions. The RAMBO pipeline begins by describing the design space of a reaction to be optimized and determining which reaction parameters should be explored first. It then queries available data to extract the most relevant data points and assembles initial suggestions for the user. The system was demonstrated using a Suzuki-Miura coupling reaction as an example, which involves an extensive combinatorial space of different parameters. RAMBO can describe the design space and limit it to available compounds, providing starting parameters to optimize the reaction. The project includes a demo interface where users can input their reaction of interest, and the React system in the backend queries and extracts relevant data to form a reply containing necessary conditions to initiate the Bayesian optimization process. Users can also explore the reasoning behind the system's suggestions by analyzing the data and reactions used to generate the final answer.

Project 43: Bayesian Optimization Awesome List

This awesome list curates a collection of software, tutorials, research papers, and other resources related to Bayesian optimization in materials science and chemistry.

Project 44: Rank-based Bayesian Optimization

This research project focused on using ranking models as surrogates in Bayesian optimization for materials discovery, specifically comparing ranking-based and conventional mean squared error (MSE) loss approaches. The key motivation was that in experimental campaigns, finding the best molecule is more important than accurately predicting absolute property values, inspired by work on molecular pool-based active learning [62]. The study employed a pairwise ranking loss (margin ranking loss) and a simple fully-connected multi-layer perceptron with three hidden layers and 100 nodes as the model architecture. Experiments were conducted on multiple datasets, including a solubility dataset (Delaney) and two datasets from Ali et al. with varying roughness [63]. Results showed that the ranking loss model consistently outperformed the MSE-based model and random baseline in Bayesian optimization, acquiring more top-performing candidates within fewer evaluations. Interestingly, all models performed better on the smoother dataset, contrary to expectations. The study also found that model performance did not always correlate with its effectiveness as a surrogate in Bayesian optimization, as evidenced by the MSE model performing worse than the random baseline on the rougher dataset. The research highlights the potential of ranking-based models in overcoming overfitting issues common in Bayesian optimization with limited data points, particularly in materials discovery applications.

Project 45: Bayesian Optimization for generality

This research introduces a novel approach to Bayesian Optimization (BO) focused on finding general parameters that perform well across multiple related tasks, a common challenge in fields like chemistry where optimal reaction conditions for various substrates are sought. The researchers formulate this problem within the BO framework, aiming to maximize a general response from a set of multiple related objective functions. The proposed algorithm operates in two steps: first selecting a value in the optimization domain X , then choosing an individual objective function to optimize for that value. This approach aims to find the most general values in X without exhaustively evaluating every individual surface, thus minimizing evaluation costs. To support further research in this area, the team adapted popular test surfaces (Dixon-Price, Branin, and Beale) to be compatible with their generality problem and made these benchmark problems available on Hugging Face. This contribution allows other researchers to test and develop new algorithms for generalized Bayesian Optimization, potentially advancing the field and its applications in natural sciences.

Acknowledgements

We express a huge thanks to Kevin Jablonka and Ben Blaiszik (latter a co-author of this paper) and others did phenomenal work paving the way with the LLMs for Chemistry and Materials Hackathon (2023). Much of the setup and many of the patterns of the Bayesian optimization hackathon was either directly replicated based off of or inspired by the LLM hackathon.

This work was undertaken thanks in part to funding provided to the University of Toronto's Acceleration Consortium from the Canada First Research Excellence Fund (CFREF-2022-00042). Tim Würger acknowledges Tiago L. P. Galvão for providing Al alloy data from the CODATA database. Can Özkan acknowledges the VIPCOAT project (Virtual Open Innovation Platform

for Active Protective Coatings Guided by Modelling and Optimisation) funded by the Horizon 2020 research and innovation programme of the European Union by grant agreement no. 952903. Ankur Kumar Gupta acknowledges support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05CH11231, FWP No. DAC-LBL-Long.

Finally, we express thanks to all participants in the hackathon. The hackathon was only possible and meaningful due to the time dedicated by many in the projects themselves.

References

- [1] S. G. Baird and T. D. Sparks, Building a "Hello World" for self-driving labs: The Closed-loop Spectroscopy Lab Light-mixing demo, *STAR Protocols* **4**, 102329 (2023).
- [2] K. M. Jablonka, Q. Ai, A. Al-Feghali, S. Badhwar, J. D. Bocarsly, A. M. Bran, S. Bringuier, L. C. Brinson, K. Choudhary, D. Circi, S. Cox, W. A. De Jong, M. L. Evans, N. Gastellu, J. Genzling, M. V. Gil, A. K. Gupta, Z. Hong, A. Imran, S. Kruschwitz, A. Labarre, J. Lála, T. Liu, S. Ma, S. Majumdar, G. W. Merz, N. Moitessier, E. Moubarak, B. Mouriño, B. Pelkie, M. Pieler, M. C. Ramos, B. Ranković, S. G. Rodrigues, J. N. Sanders, P. Schwaller, M. Schwarting, J. Shi, B. Smit, B. E. Smith, J. Van Herck, C. Völker, L. Ward, S. Warren, B. Weiser, S. Zhang, X. Zhang, G. A. Zia, A. Scourtas, K. J. Schmidt, I. Foster, A. D. White, and B. Blaiszik, 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon, *Digital Discovery* **2**, 1233 (2023).
- [3] G. Mulholland and B. Meredig, Hackathon aims to solve materials problems, *MRS Bulletin* **40**, 366 (2015).
- [4] T. D. Sparks, F. E. Curtis, D. C. Fredrickson, and N. A. Benedek, Insights and Innovations from the SSMCDAT 2023: Bridging Solid-State Materials Chemistry and Data Science, *Chemistry of Materials* **36**, 5293 (2024), publisher: American Chemical Society.
- [5] K. Kandasamy, K. R. Vysyaraju, W. Neiswanger, B. Paria, C. R. Collins, J. Schneider, B. Poczos, and E. P. Xing, Tuning hyperparameters without grad students: Scalable and robust bayesian optimisation with dragonfly, *Journal of Machine Learning Research* **21**, 1 (2020).
- [6] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization, in *Advances in Neural Information Processing Systems* **33** (2020).
- [7] S. Daulton, M. Balandat, and E. Bakshy, Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement, *Advances in Neural Information Processing Systems* **34**, 2187 (2021).
- [8] M. Poloczek, J. Wang, and P. Frazier, Multi-information source optimization, *Advances in neural information processing systems* **30** (2017).
- [9] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, Taking the human out of the loop: A review of bayesian optimization, *Proceedings of the IEEE* **104**, 148 (2015).
- [10] J. Snoek, H. Larochelle, and R. P. Adams, Practical bayesian optimization of machine learning algorithms, *Advances in neural information processing systems* **25** (2012).
- [11] E. S. Muckley, J. E. Saal, B. Meredig, C. S. Roper, and J. H. Martin, Interpretable models for extrapolation in scientific machine learning, *Digital Discovery* **2**, 1425 (2023).
- [12] J. Bergstra and Y. Bengio, Random search for hyperparameter optimization, *The journal of machine learning research* **13**, 281 (2012).
- [13] M. Li, M. J. McCourt, A. J. Galante, and P. W. Leu, Bayesian optimization of nanophotonic electromagnetic shielding with very high visible transparency, *Optics Express* **30**, 33182 (2022).
- [14] J. Kim and S. Choi, Bayeso: A bayesian optimization framework in python, *Journal of Open Source Software* **8**, 5320 (2023).
- [15] M. Fitzner, A. Šošić, A. Hopp, and A. Lee, BayBE – a bayesian back end for design of experiments, <https://github.com/emdgroup/baybe> (2022), accessed: 2024-02-22.
- [16] T. L. Galvão, I. Ferreira, A. Kuznetsova, G. Novell-Leruth, C. Song, C. Feiler, S. V. Lamaka, C. Rocha, F. Maia, M. L. Zheludkevich, *et al.*, Cordata: An open data management web application to select corrosion inhibitors, *npj Materials Degradation* **6**, 48 (2022).
- [17] H. Moriwaki, Y.-S. Tian, N. Kawashita, and T. Takagi, Mordred: a molecular descriptor calculator, *Journal of cheminformatics* **10**, 4 (2018).
- [18] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, Reoptimization of mdl keys for use in drug discovery, *Journal of chemical information and computer sciences* **42**, 1273 (2002).
- [19] G. Landrum, Rdkit documentation, Release **1**, 4 (2013).
- [20] Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, Optimization of molecules via deep reinforcement learning,

- Scientific reports **9**, 10752 (2019).
- [21] H. Furukawa, K. E. Cordova, M. O’Keeffe, and O. M. Yaghi, The chemistry and applications of metal-organic frameworks, *Science* **341**, 1230444 (2013).
 - [22] D. Y. Heo, H. H. Do, S. H. Ahn, and S. Y. Kim, Metal-organic framework materials for perovskite solar cells, *Polymers* **12**, 2061 (2020).
 - [23] D. DeSantis, J. A. Mason, B. D. James, C. Houchins, J. R. Long, and M. Veenstra, Techno-economic analysis of metal-organic frameworks for hydrogen and natural gas storage, *Energy & Fuels* **31**, 2024 (2017).
 - [24] F. L. Oliveira, C. Cleeton, R. Neumann Barros Ferreira, B. Luan, A. H. Farmahini, L. Sarkisov, and M. Steiner, Crafted: An exploratory database of simulated adsorption isotherms of metal-organic frameworks, *Scientific Data* **10**, 230 (2023).
 - [25] M. Dusselier and M. E. Davis, Small-pore zeolites: synthesis and catalysis, *Chemical reviews* **118**, 5265 (2018).
 - [26] A. J. Mallette, K. Shilpa, and J. D. Rimer, The current understanding of mechanistic pathways in zeolite crystallization, *Chemical Reviews* **124**, 3416 (2024).
 - [27] M. Altamirano, F.-X. Briol, and J. Knoblauch, Robust and conjugate gaussian process regression, *arXiv preprint arXiv:2311.00463* (2023).
 - [28] S. Ament, A. Witte, N. Garg, and J. Kusuma, Sustainable concrete via bayesian optimization, 2023, in *URL https://arxiv.org/abs/2310.18288. NeurIPS 2023 Workshop on Adaptive Experimentation in the Real World*.
 - [29] H. Mahmoudi, M. Mahmoudi, O. Doustdar, H. Jahangiri, A. Tsolakis, S. Gu, and M. LechWyszynski, A review of fischer tropesch synthesis process, mechanism, surface chemistry and catalyst formulation, *Biofuels Engineering* **2**, 11 (2017).
 - [30] G. Lozano-Blanco, J. W. Thybaut, K. Surla, P. Galtier, and G. B. Marin, Single-event microkinetic model for fischer- tropesch synthesis on iron-based catalysts, *Industrial & engineering chemistry research* **47**, 5879 (2008).
 - [31] A. Chakkingal, P. Janssens, J. Poissonnier, M. Virginie, A. Y. Khodakov, and J. W. Thybaut, Multi-output machine learning models for kinetic data evaluation: A fischer-tropsch synthesis case study, *Chemical Engineering Journal* **446**, 137186 (2022).
 - [32] Z. J. Lin, R. Astudillo, P. Frazier, and E. Bakshy, Preference exploration for efficient bayesian optimization with multiple outcomes, in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2022) pp. 4235–4258.
 - [33] H. Bellamy, A. A. Rehim, O. I. Orhobor, and R. King, Batched bayesian optimization for drug design in noisy environments, *Journal of Chemical Information and Modeling* **62**, 3970 (2022).
 - [34] K. Wu and G.-W. Wei, Quantitative toxicity prediction using topology based multitask deep neural networks, *Journal of chemical information and modeling* **58**, 520 (2018).
 - [35] D. Chen, K. Gao, D. D. Nguyen, X. Chen, Y. Jiang, G.-W. Wei, and F. Pan, Algebraic graph-assisted bidirectional transformers for molecular property prediction, *Nature communications* **12**, 3521 (2021).
 - [36] D. Rogers and M. Hahn, Extended-connectivity fingerprints, *Journal of chemical information and modeling* **50**, 742 (2010).
 - [37] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, *Scientific Data* **1**, 140022 (2014).
 - [38] K. Yang, M. Emmerich, A. Deutz, and T. Bäck, Multi-objective bayesian global optimization using expected hypervolume improvement gradient, *Swarm and evolutionary computation* **44**, 945 (2019).
 - [39] J. Knowles, Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, *IEEE transactions on evolutionary computation* **10**, 50 (2006).
 - [40] J. Stetefeld, S. A. McKenna, and T. R. Patel, Dynamic light scattering: a practical guide and applications in biomedical sciences, *Biophysical reviews* **8**, 409 (2016).
 - [41] F. Sorourifar, T. Banker, and J. A. Paulson, Accelerating Black-Box Molecular Property Optimization by Adaptively Learning Sparse Subspaces, *arXiv preprint arXiv:2401.01398* (2024).
 - [42] N. Maus, H. T. Jones, J. S. Moore, M. J. Kusner, J. Bradshaw, and J. R. Gardner, Local Latent Space Bayesian Optimization over Structured Inputs, *arXiv preprint arXiv:2201.11872* (2023).
 - [43] H. B. Moss and R.-R. Griffiths, Gaussian process molecule property prediction with flowmo, *arXiv preprint arXiv:2010.01118* (2020).
 - [44] C. J. Hargreaves, M. S. Dyer, M. W. Gaultois, V. A. Kurlin, and M. J. Rosseinsky, The earth mover’s distance as a metric for the space of inorganic compositions, *Chemistry of Materials* **32**, 10610 (2020).
 - [45] N. J. Szymanski, Y. Zeng, H. Huo, C. J. Bartel, H. Kim, A. Parija, N. Shoghi, S. Rigas, J. B. Neaton, G. Ceder, K. A. Persson, F. M. Toma, and J. M. Gregoire, An autonomous laboratory for the accelerated synthesis of novel materials, *Nature* **624**, 86 (2023).
 - [46] F. L. Johansen, A. S. Anker, T. L. Christiansen, L. S. Voss, P. Juhás, S. J. Billinge, and K. M. Jensen, A GPU-Accelerated Open-Source Python Package for Calculating Powder Diffraction, Small-Angle-, and Total Scattering with the Debye Scattering Equation, *Journal of Open Source Software* **9**, 6024 (2024).
 - [47] J. Leeman, A. A. Kunitsa, K. S. Vecchio, G. Ceder, and A. G. Kusne, Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis, *PRX Energy* **3**, 011002 (2024).
 - [48] M. García-Ortegón, G. N. C. Simm, A. J. Tripp, J. M. Hernández-Lobato, A. Bender, and S. Bacallado, Dockstring: Easy molecular docking yields better benchmarks for ligand design, *Journal of Chemical Information and Modeling* **62**, 3486 (2022).
 - [49] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, and A. L. Hopkins, Quantifying the Chemical Beauty of Drugs, *Nature Chemistry* **4**, 90 (2012).
 - [50] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, *Scientific Data* **1**, 140022 (2014).
 - [51] L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17, *J. Chem. Inf. Model.* **52**, 2864 (2012).
 - [52] T. Head, MechCoder, G. Louppe, I. Shcherbatyi, fcharas, D. Cournapeau, F. Pedregosa, and A. Mueller, *scikit-optimize/scikit-optimize: v0.9.0* (2021).
 - [53] E. Bakshy, Ax: A platform for adaptive experimentation, *GitHub repository* (2018).
 - [54] A. E. Blanchard, J. Gounley, D. Bhowmik, M. Chandra Shekar, I. Lyngaas, S. Gao, J. Yin, A. Tsaris,

- F. Wang, and J. Glaser, Language models for the prediction of sars-cov-2 inhibitors, *The International Journal of High Performance Computing Applications* **36**, 587 (2022).
- [55] M. J. Bick, P. J. Greisen, K. J. Morey, M. S. Antunes, D. La, B. Sankaran, L. Reymond, K. Johnsson, J. I. Medford, and D. Baker, Computational design of environmental sensors for the potent opioid fentanyl, *Elife* **6**, e28909 (2017).
- [56] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, Highly accurate protein structure prediction with alphafold, *nature* **596**, 583 (2021).
- [57] STAIR Lab, coh2: Conversational hf interaction dataset, <https://huggingface.co/datasets/stair-lab/coh2> (2025), accessed: 2025-06-15.
- [58] R.-R. Griffiths, L. Klärner, H. Moss, A. Ravuri, S. Truong, Y. Du, S. Stanton, *et al.*, Gauche: A library for gaussian processes in chemistry, in *Advances in Neural Information Processing Systems*, Vol. 36 (2024).
- [59] S. Jiang, D. Jiang, M. Balandat, B. Karrer, J. Gardner, and R. Garnett, Efficient nonmyopic bayesian optimization via one-shot multi-step trees, in *Advances in Neural Information Processing Systems*, Vol. 33 (2020) pp. 18039–18049.
- [60] A. Deshwal, C. M. Simon, and J. R. Doppa, Bayesian Optimization of Nanoporous Materials, *Molecular Systems Design & Engineering* **6**, 1066 (2021).
- [61] R. Mercado, R.-S. Fu, A. V. Yakutovich, L. Talirz, M. Haranczyk, and B. Smit, In silico design of 2d and 3d covalent organic frameworks for methane storage applications, *Chemistry of Materials* **30**, 5069 (2018).
- [62] D. E. Graff, E. I. Shakhnovich, and C. W. Coley, Accelerating high-throughput virtual screening through molecular pool-based active learning, *Chemical Science* **12**, 7866 (2021).
- [63] M. Aldeghi, F. Häse, R. J. Hickman, I. Tamblyn, and A. Aspuru-Guzik, Roughness of Molecular Property Landscapes and Its Impact on Modellability, *Journal of Chemical Information and Modeling* **62**, 4660 (2022).