Bayesian Optimization for Reactions

Philippe Schwaller

Laboratory of Artificial Chemical Intelligence (LIAC)

AI for Chemistry

https://lexica-serve-encoded-images2.sharif.workers.dev/full_jpg/5808a79a-7530-4f58-b6f4-6530f6c98070

Gold Digger Simulator

Scenario

You're digging for gold in a large field. You want to dig where the most gold is so you can sell it and be able to afford the unlimited Swiss train pass (in 1st class too!)

Problem

Where should you dig in this large field?

It takes a lot of **time** and **money** for each dig and you can't dig everywhere!

Strategy

You dug and found 1 gold bar. You decide to dig somewhere else. 2 gold bars! You decide to dig close by because **maybe** there's a lot of gold in this area.

Gold Digger Simulator

Why follow this strategy?

You want to try digging in different places because you have no **Prior** belief on where the most gold is. Once you found an area with more gold, your **Posterior** belief compels you to dig in a nearby area

Bayesian Optimization Terminology

Exploration: You want to explore the dig site but not the entire dig site because that takes too much time and is too expensive. **You are expanding your knowledge**.

Exploitation: You want to take advantage of the information you have to make an educated guess (inductive bias) on where the most gold is. **You are exploiting your knowledge**.

Bayesian Optimization (BO)

We don’t have all the money and time in the world

- We encounter **Optimization** everywhere

1. Maximize your chemical reaction yield
2. Maximize your drug molecule for pIC_{50}
3. Optimize your machine learning model hyperparameters to minimize the error
4. Minimize the cost to scale-up your reaction
5. Minimize the energy consumption of your process reaction

Bayesian Optimization is a sequential optimization algorithm that uses Active Learning and aims to optimize an Expensive Oracle under Minimal Oracle Calls

Don’t be fancy

The **Oracle** as in some sort of experiment **without a known mathematical form and is expensive**

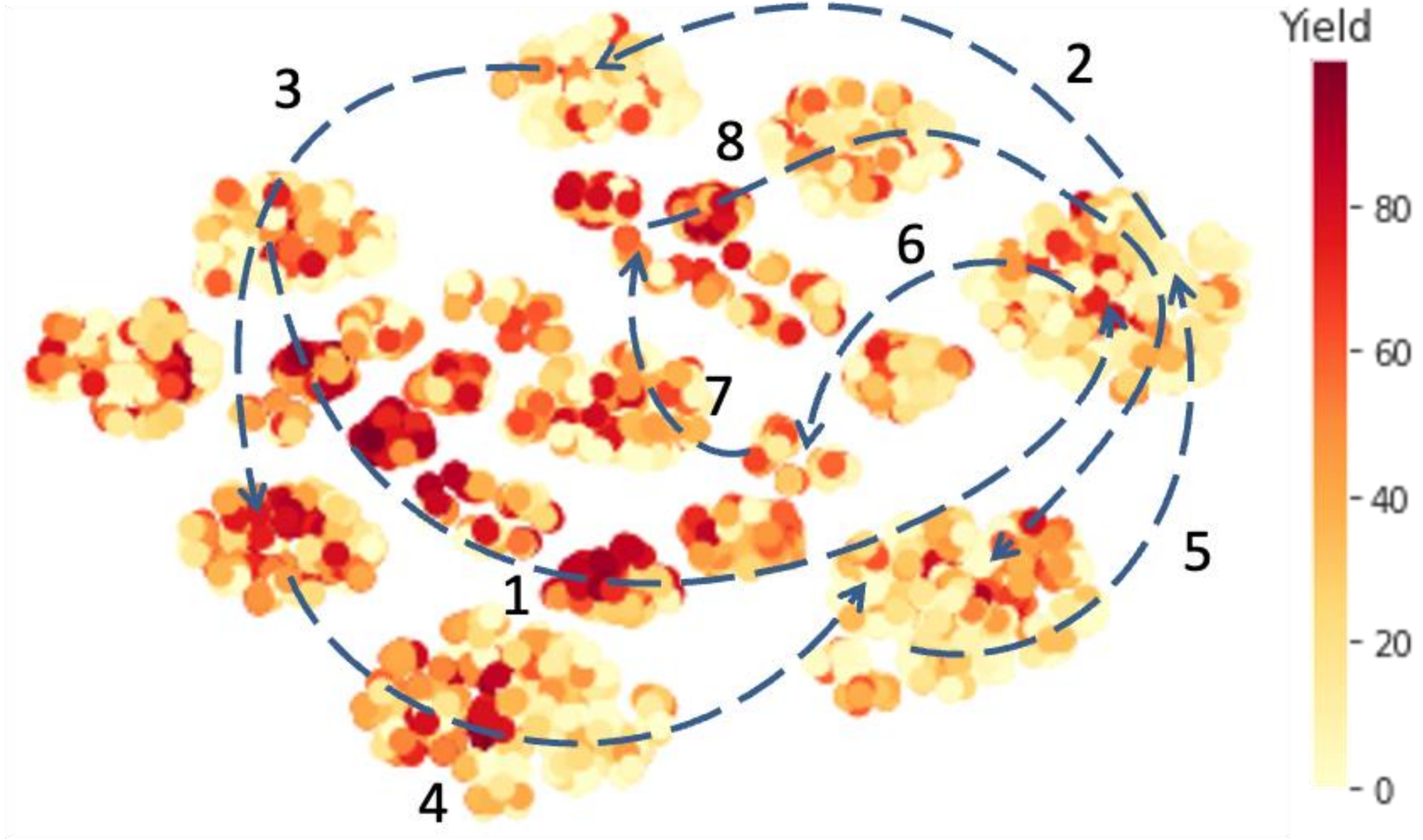
1. Computational calculation, e.g., Density Functional Theory (DFT)
2. Wet-lab experimentation, e.g., doing a chemical reaction

- If there is a known mathematical equation, just use calculus to optimize it
- If it is cheap, just do the experiment

There is always uncertainty when you use Bayesian Optimization. Remember, you never know for certain until you actually do the experiment and make the observation.

If experiments are cheap, just do all the experiments.

Data/Search Space



Data

- Previous observations
- Search space restrictions (maybe you don't want to try every possible x value)

What is X (representation)?

- Reaction conditions to predict yield
- Molecular fingerprint to predict pIC_{50}
- It can be anything, really

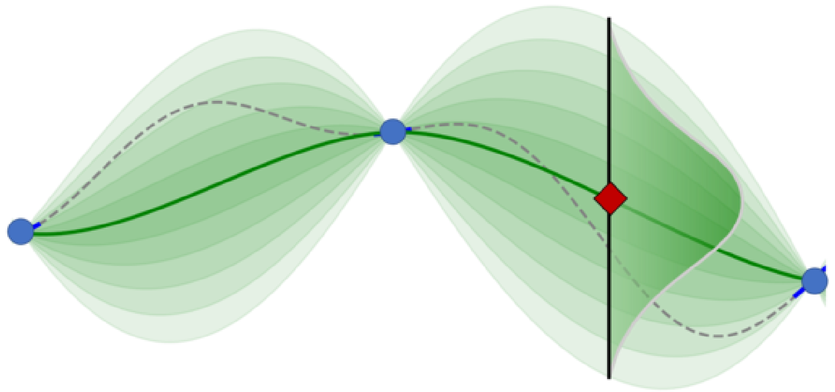
The most important part, like any machine learning application, is to have correct data and the more, the better

Surrogate Model

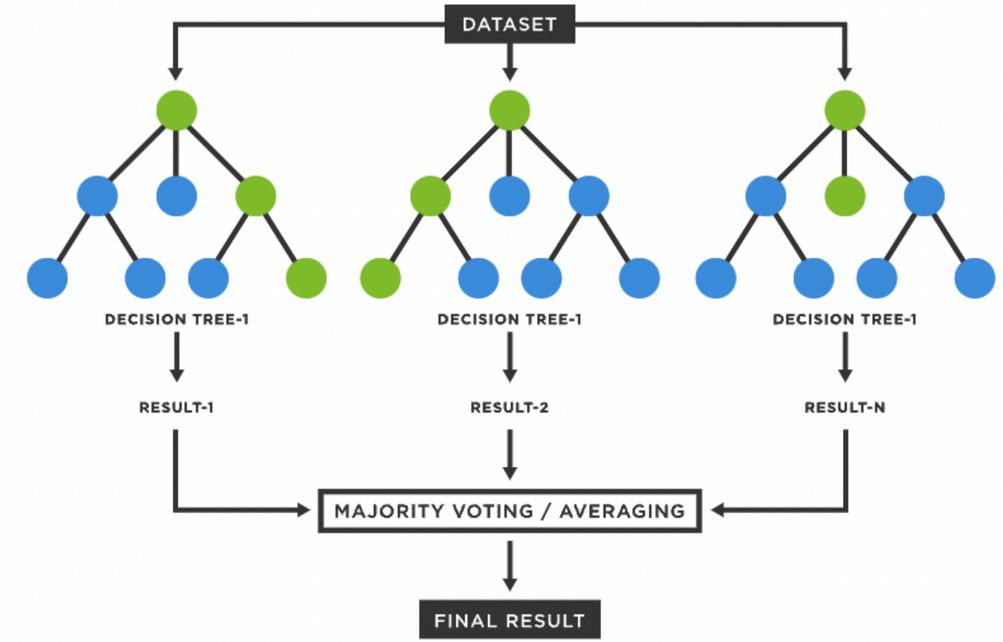
Surrogate Model

- Machine learning model that is used to predict F(x)
- The model's prediction is passed to the **Acquisition Function** to decide the next experiment to do
- In general, all models are capable of predicting a **value** and an **uncertainty**

Gaussian Process (GP)



Tree-based Models (like Random Forest)



Neural Network

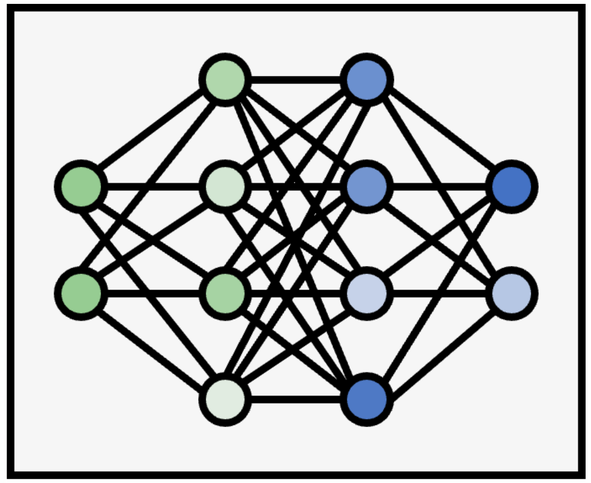
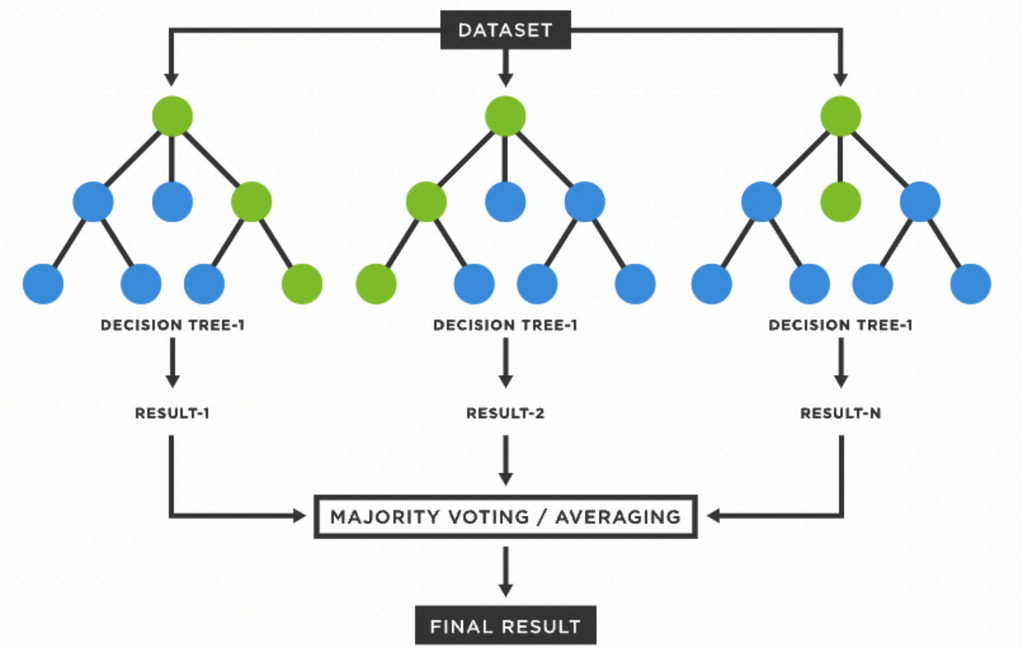


Image from: https://www.tibco.com/reference-center/what-is-a-random-forest

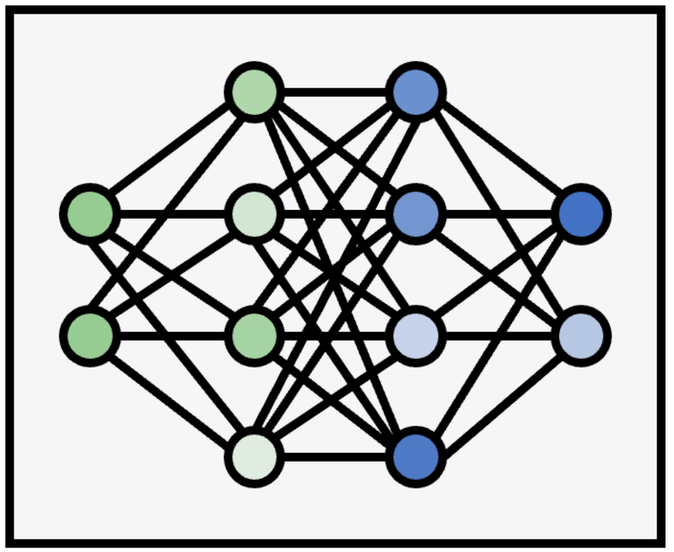
Uncertainty in Random Forests and Neural Networks

Tree-based Models (like Random Forest)



- Take the standard deviation of each tree's prediction in the ensemble

Neural Network



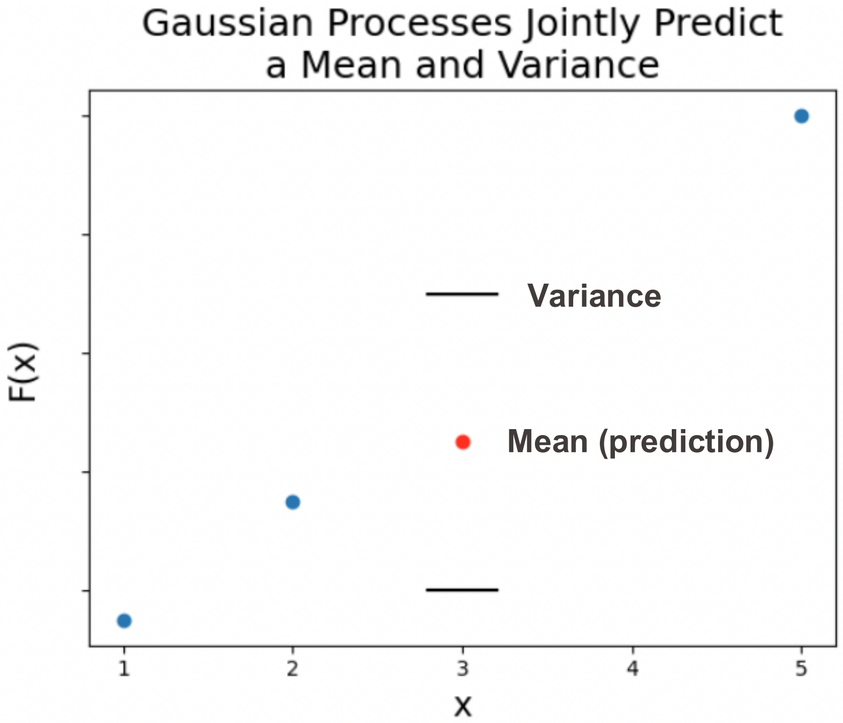
- Train an ensemble of neural networks
- **Monte-Carlo Dropout**

Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning

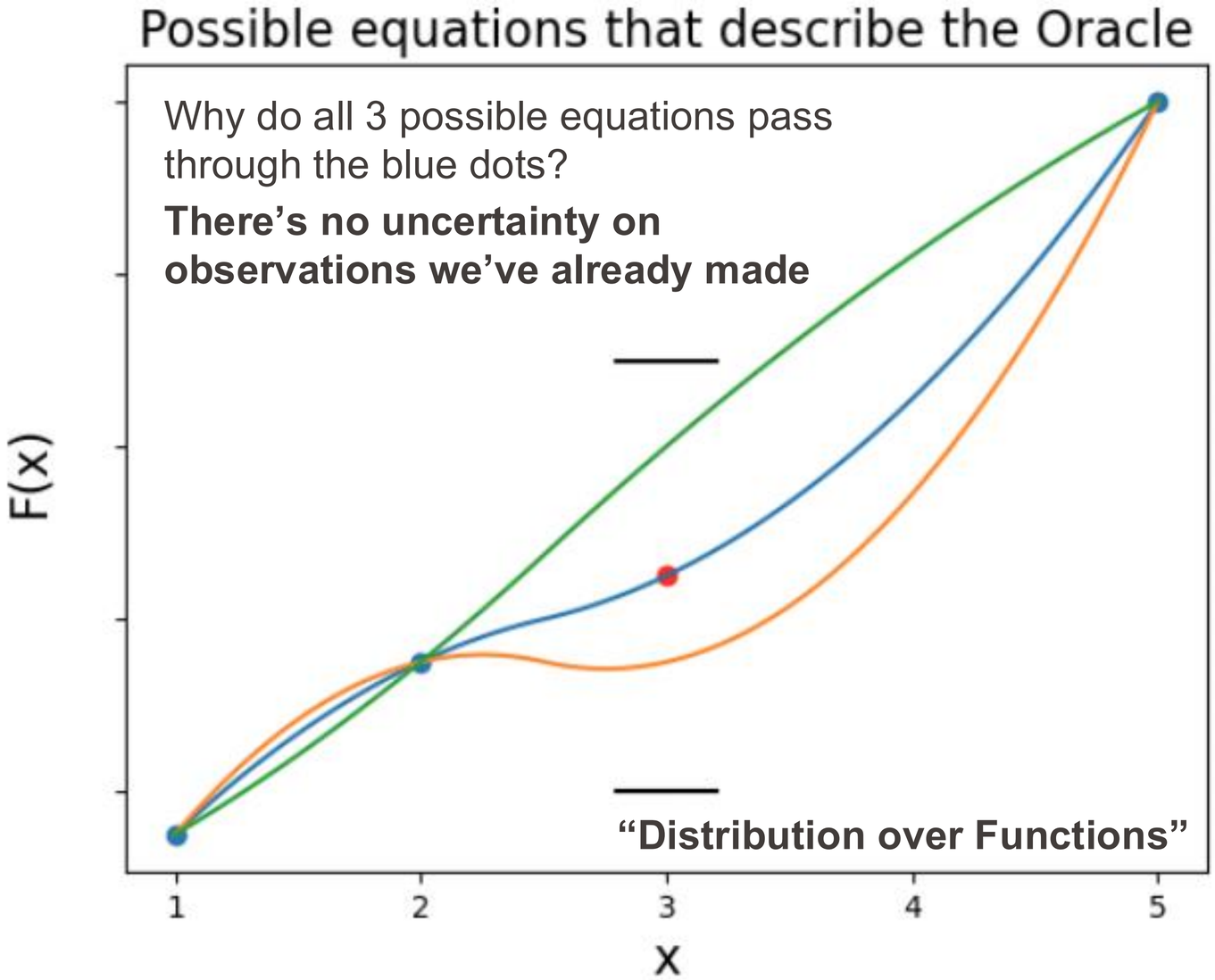
Yarin Gal YG279@CAM.AC.UK
Zoubin Ghahramani ZG201@CAM.AC.UK
University of Cambridge

Image from: https://www.tibco.com/reference-center/what-is-a-random-forest

Gaussian Process Inference: 2-for-1 deal

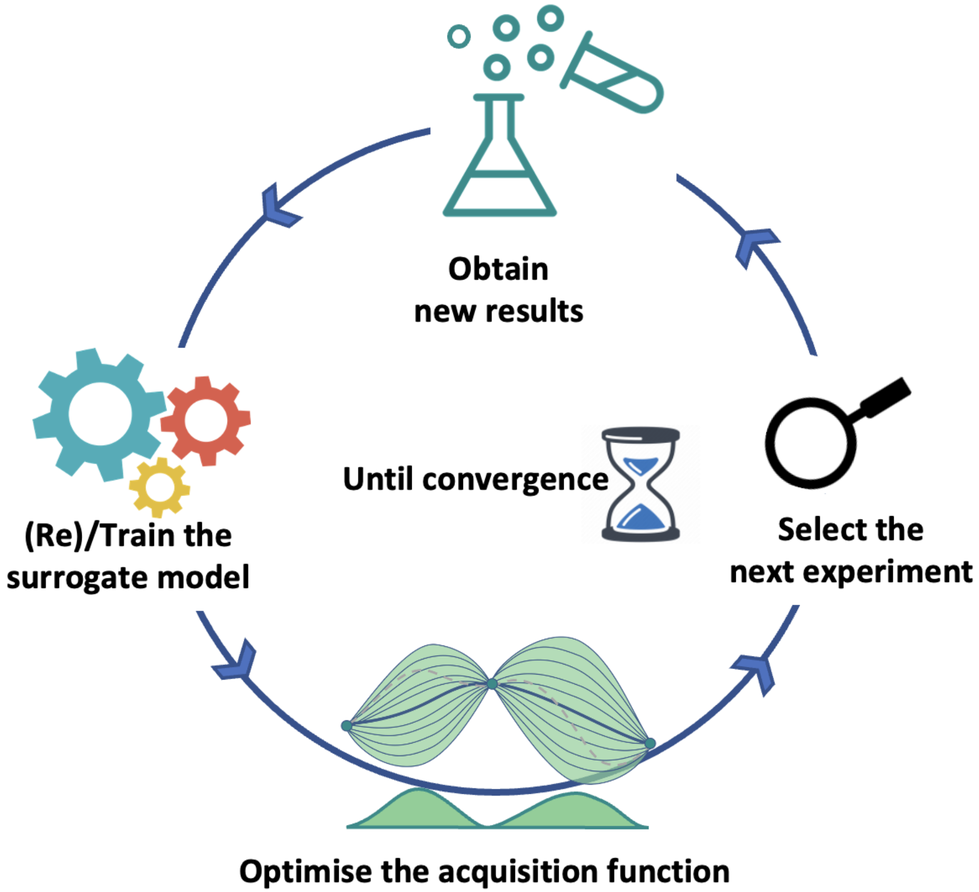


- Gaussian Processes (GPs), by design, jointly predict a **mean** and **variance** that can be **analytically computed**



concepts and presentation style adapted from https://www.youtube.com/watch?v=4vGiHC35j9s&ab_channel=NandodeFreitas

Starbucks FrappBOccino Recipe



Ingredients

- **Data**
- **Surrogate Model**
- **Acquisition Function**
  How should we use our surrogate model's predictions to choose the next experiment to do?
- Splash of hope

Acquisition Function

Acquisition Function

- What experiment should I do next?
- We decide this by plugging in our surrogate model's **prediction and/or uncertainty** into some heuristic

- Recall digging for gold – we wanted to do some **exploration** (use uncertainty) and **exploitation** (use prediction)

- **Acquisition functions output some value (sometimes called utility) and we simply choose the experiment(s) with the highest utility to perform next**

Intuitive Acquisition Functions

Greedy
Do the experiment with the best predicted outcome (model's prediction)
→Can be too exploitative (not exploring enough)

Uncertainty
Do the experiment with your model's most uncertain about (highest variance)
→Can be too explorative

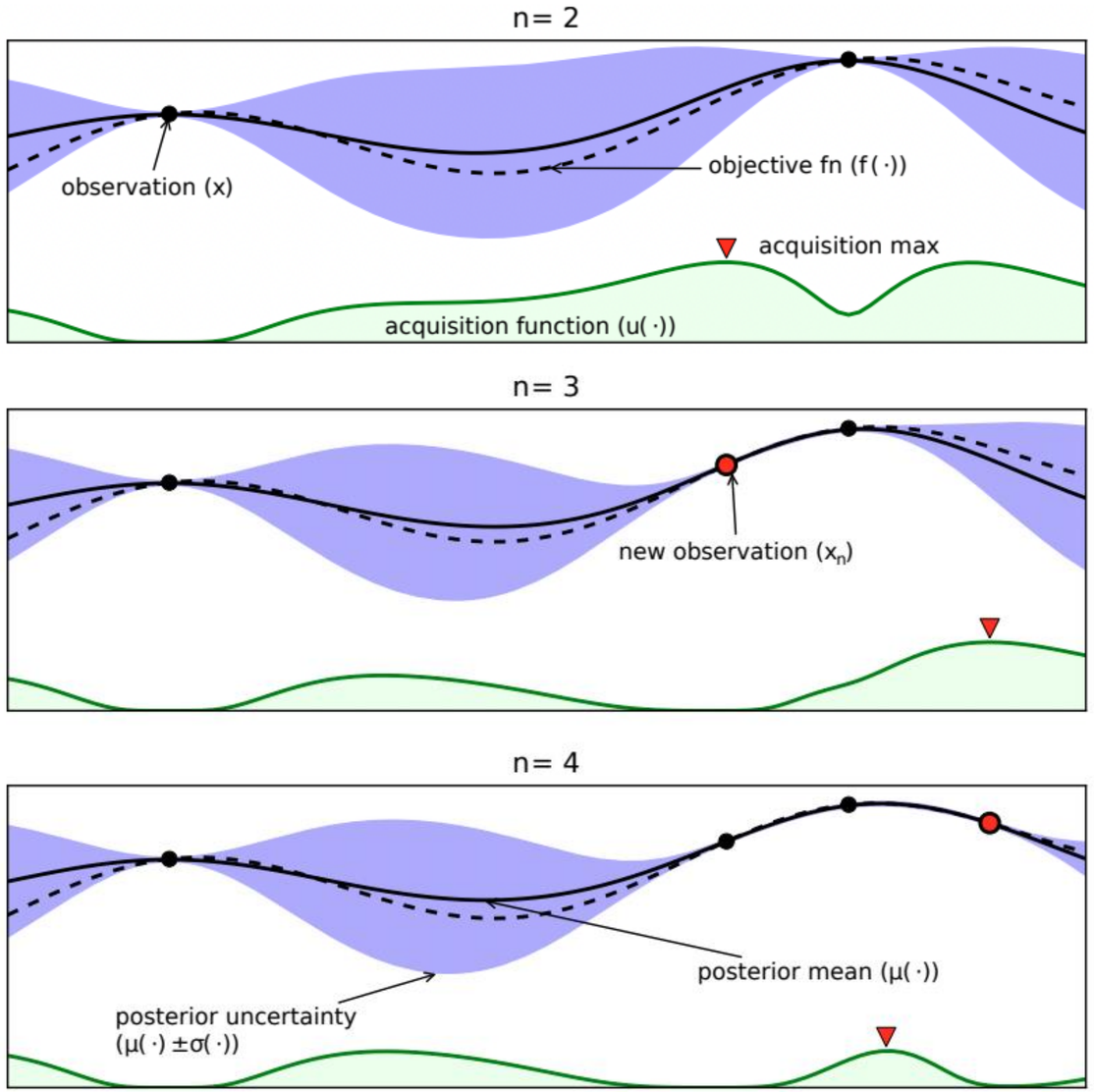
Upper-confidence bound (UCB)
Linear combination between model's prediction and uncertainty → tries to balance both exploration-exploitation

$$UCB = \mu + \beta * \sigma$$

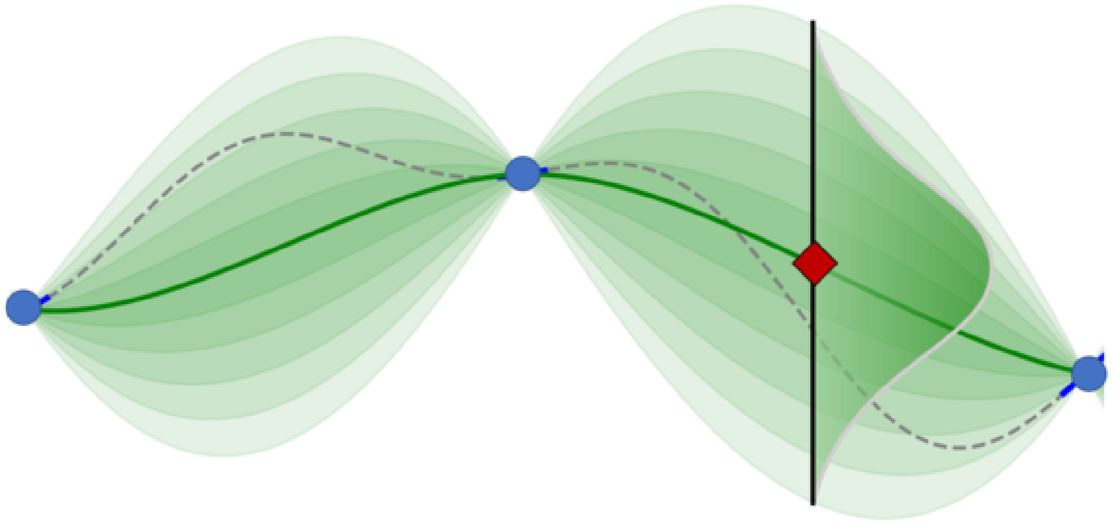
$$\boldsymbol{UCB = model's\ prediction + \beta * model's\ uncertainty}$$

Other commonly used acquisition function →Expected improvement (EI),
→ balances exploration-exploitation by targeting next experiments with predicted mean above current best or high uncertainty

Squash the Uncertainty



- Blue shaded region = uncertainty
- **Where we have observations, there's no uncertainty**
- **Where we don't have observations, there's more uncertainty**

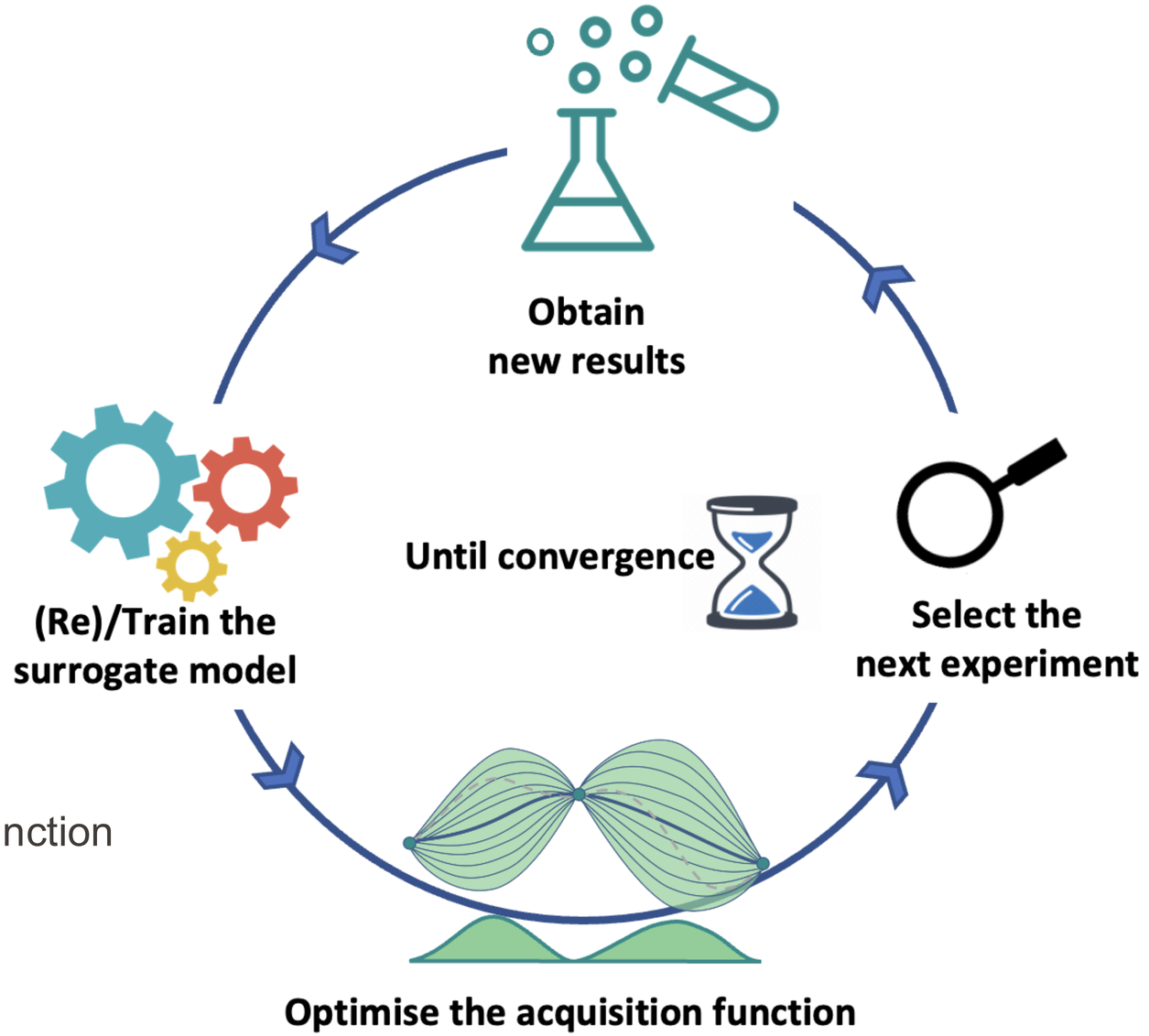


https://ieeexplore.ieee.org/document/7352306

Serving the FrappBOccino

Steps

1. Train your initial surrogate model
2. Use the model to predict new values
3. Plug the predictions into your acquisition function
4. Choose the next experiments to do
5. Do the experiments and record the result
6. Update your surrogate model (posterior)
7. Repeat (until you succeed or run out of money ☹)



FrappBOccino Sequel: Chemistry Flavoured

Bayesian reaction optimization as a tool for chemical synthesis

Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Martinez Alvarado, Jacob M. Janey, Ryan P. Adams & Abigail G. Doyle

Nature **590**, 89–96 (2021) | Cite this article

52k Accesses | **214** Citations | **180** Altmetric | Metrics

Phoenics: A Bayesian Optimizer for Chemistry

Florian Häse, Loïc M. Roch, Christoph Kreisbeck, and Alán Aspuru-Guzik*

Cite this: *ACS Cent. Sci.* 2018, 4, 9, 1134–1145
Publication Date: August 24, 2018
https://doi.org/10.1021/acscentsci.8b00307
Copyright © 2018 American Chemical Society
RIGHTS & PERMISSIONS

Article Views 15937 | Altmetric 23 | Citations 167
LEARN ABOUT THESE METRICS

A Multi-Objective Active Learning Platform and Web App for Reaction Optimization

Jose Antonio Garrido Torres, Sii Hong Lau, Pranay Anchuri, Jason M. Stevens, Jose E. Tabora, Jun Li, Alina Borovika, Ryan P. Adams, and Abigail G. Doyle*

Cite this: *J. Am. Chem. Soc.* 2022, 144, 43, 19999–20007
Publication Date: October 19, 2022
https://doi.org/10.1021/jacs.2c08592
Copyright © 2022 American Chemical Society
RIGHTS & PERMISSIONS ✔ Subscribed

Article Views 6544 | Altmetric 17 | Citations 5
LEARN ABOUT THESE METRICS

Share Add to Export RIS

Nothing fundamentally different from classic Bayesian Optimization

GRYFFIN: An algorithm for Bayesian optimization of categorical variables informed by expert knowledge

Florian Häse ; Matteo Aldeghi ; Riley J. Hickman ; Loïc M. Roch ; Alán Aspuru-Guzik

Check for updates

+ Author & Article Information

Applied Physics Reviews 8, 031406 (2021)

https://doi.org/10.1063/5.0048164 **Article history**

README | Apache-2.0 license

CI failing | Regular Check failing | Docs passing

Supports Python 3.10 | 3.11 | 3.12 | PyPI Version v0.13.0 | Downloads 1.6k/month | Issues 18 open | PRs 16 open

License Apache 2.0

Homepage · User Guide · Documentation · Contribute

BayBE — A Bayesian Back End for Design of Experiments

The **Bay**esian **B**ack **E**nd (**BayBE**) is a general-purpose toolbox for Bayesian Design of Experiments, focusing on additions that enable real-world experimental campaigns.

```python
from baybe.targets import NumericalTarget
from baybe.objectives import SingleTargetObjective

target = NumericalTarget(
    name="Yield",
    mode="MAX",
)
objective = SingleTargetObjective(target=target)
```

Target

```python
from baybe.parameters import (
    CategoricalParameter,
    NumericalDiscreteParameter,
    SubstanceParameter,
)

parameters = [
    CategoricalParameter(
        name="Granularity",
        values=["coarse", "medium", "fine"],
        encoding="OHE",  # one-hot encoding of categories
    ),
    NumericalDiscreteParameter(
        name="Pressure[bar]",
        values=[1, 5, 10],
        tolerance=0.2,  # allows experimental inaccuracies up to 0.2 when re
    ),
    SubstanceParameter(
        name="Solvent",
        data={
            "Solvent A": "COC",
            "Solvent B": "CCC",  # label-SMILES pairs
            "Solvent C": "O",
            "Solvent D": "CS(=O)C",
        },
        encoding="MORDRED",  # chemical encoding via scikit-fingerprints
    ),
]
```

Search space

```python
from baybe.recommenders import (
    BotorchRecommender,
    FPSRecommender,
    TwoPhaseMetaRecommender,
)

recommender = TwoPhaseMetaRecommender(
    initial_recommender=FPSRecommender(),  # farthest point sampling
    recommender=BotorchRecommender(),  # Bayesian model-based optimization
)
```

Recommender (surrogate + aqc. func.)

```python
from baybe import Campaign

campaign = Campaign(searchspace, objective, recommender)
```

```python
df = campaign.recommend(batch_size=3)
print(df)
```

Ask for suggestions

	Granularity	Pressure[bar]	Solvent
15	medium	1.0	Solvent D
10	coarse	10.0	Solvent C
29	fine	5.0	Solvent B

Add results

```python
df["Yield"] = [79.8, 54.1, 59.4]
campaign.add_measurements(df)
```

https://github.com/emdgroup/baybe

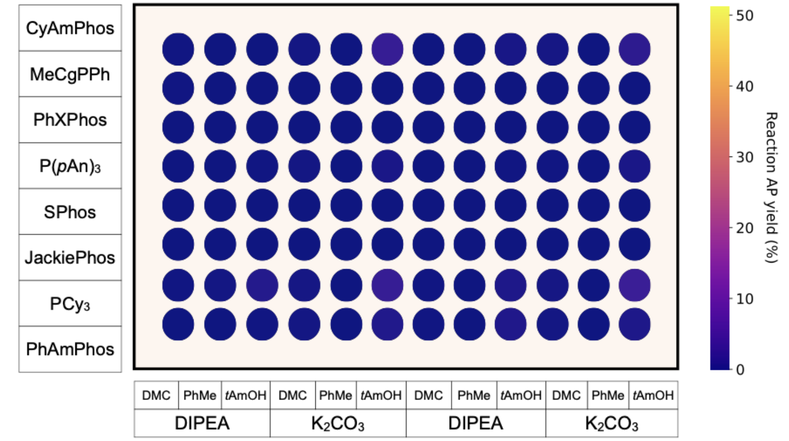
Going beyond standard BO

Bayesian Optimization for large batches

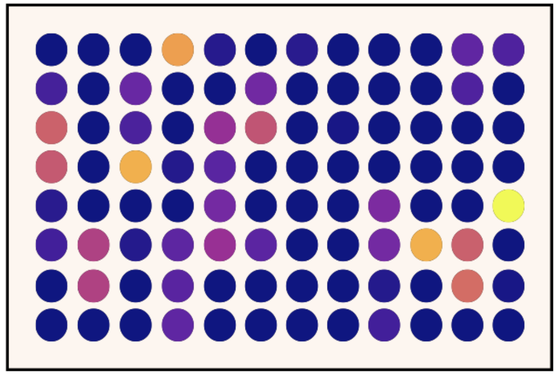
Joshua Sin
(with Roche)

a

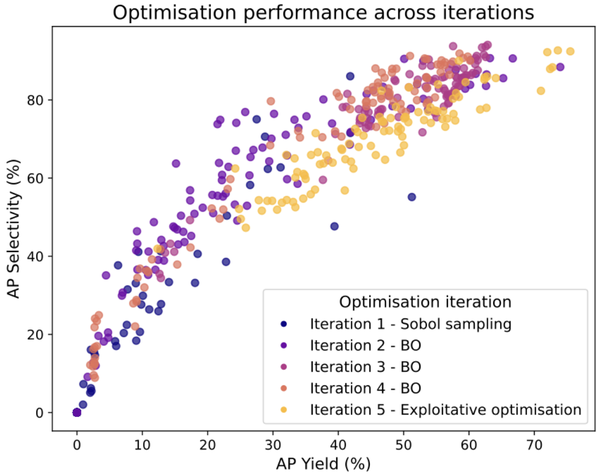
Best results from two experimentalist-designed HTE plates



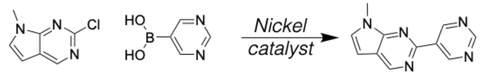
Results of initial plate from our ML optimisation workflow



Optimisation performance across iterations

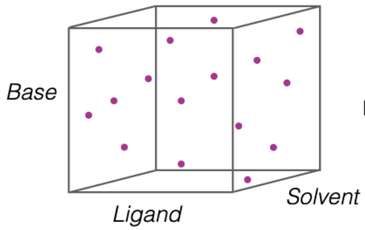


b



Ligand	Base	Solvent	…	Objectives
L1	B1	S1	…	Yield
…	…	…		…

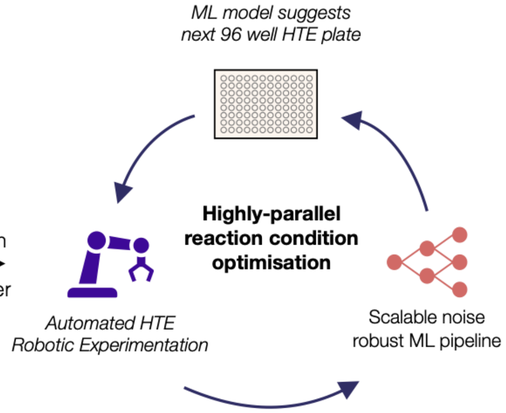
Large, chemist-defined reaction condition space of many possible combinations

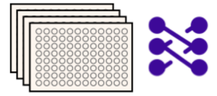


Initial **quasi-random** Sobol guided experiment selection

Experiments • diversely cover reaction condition space

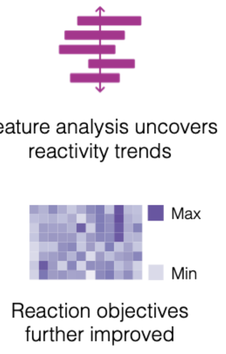
Rapid suggestion and execution of batches of 96 parallel experiments at multiple temperatures





Exploitative iteration with knowledge gained from multiple 96-well plates

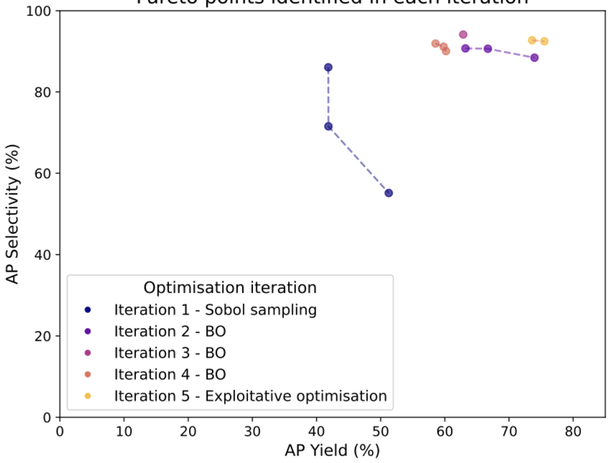
- **Optimised selectivity and yield**
- **Unexpected reactivity insights**



Feature analysis uncovers reactivity trends

Reaction objectives further improved

Pareto points identified in each iteration

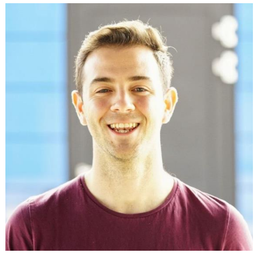


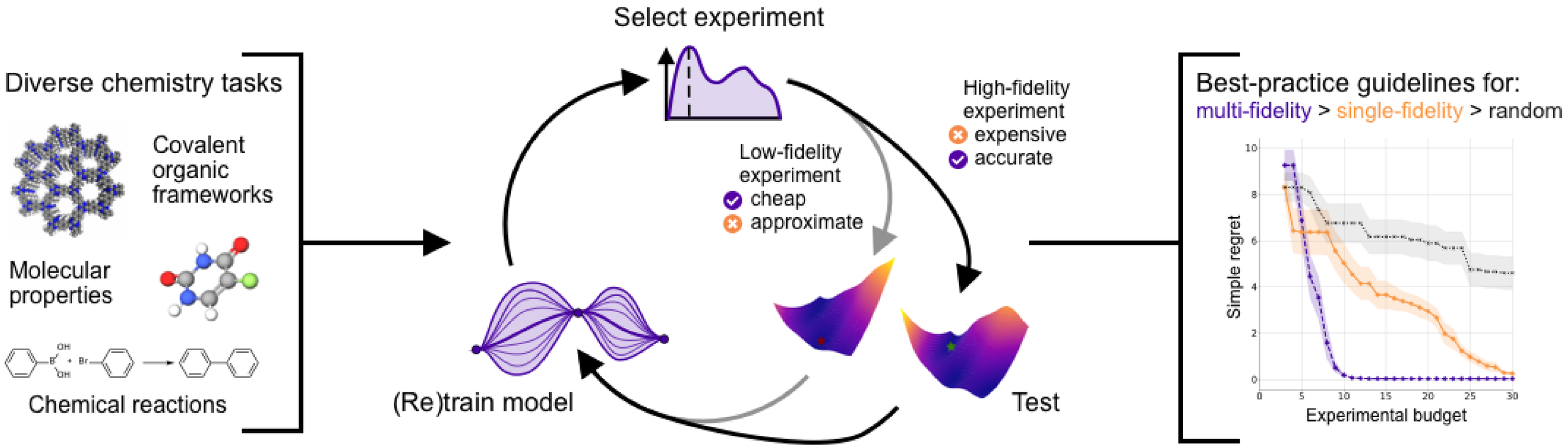
Highly Parallel Optimisation of Nickel-Catalysed Suzuki Reactions through Automation and Machine Intelligence

- Uses special acquisition function (qNEHVI)
- Incorporates constraints from HTE platform

Multi-fidelity BO

Víctor Sabanza Gil



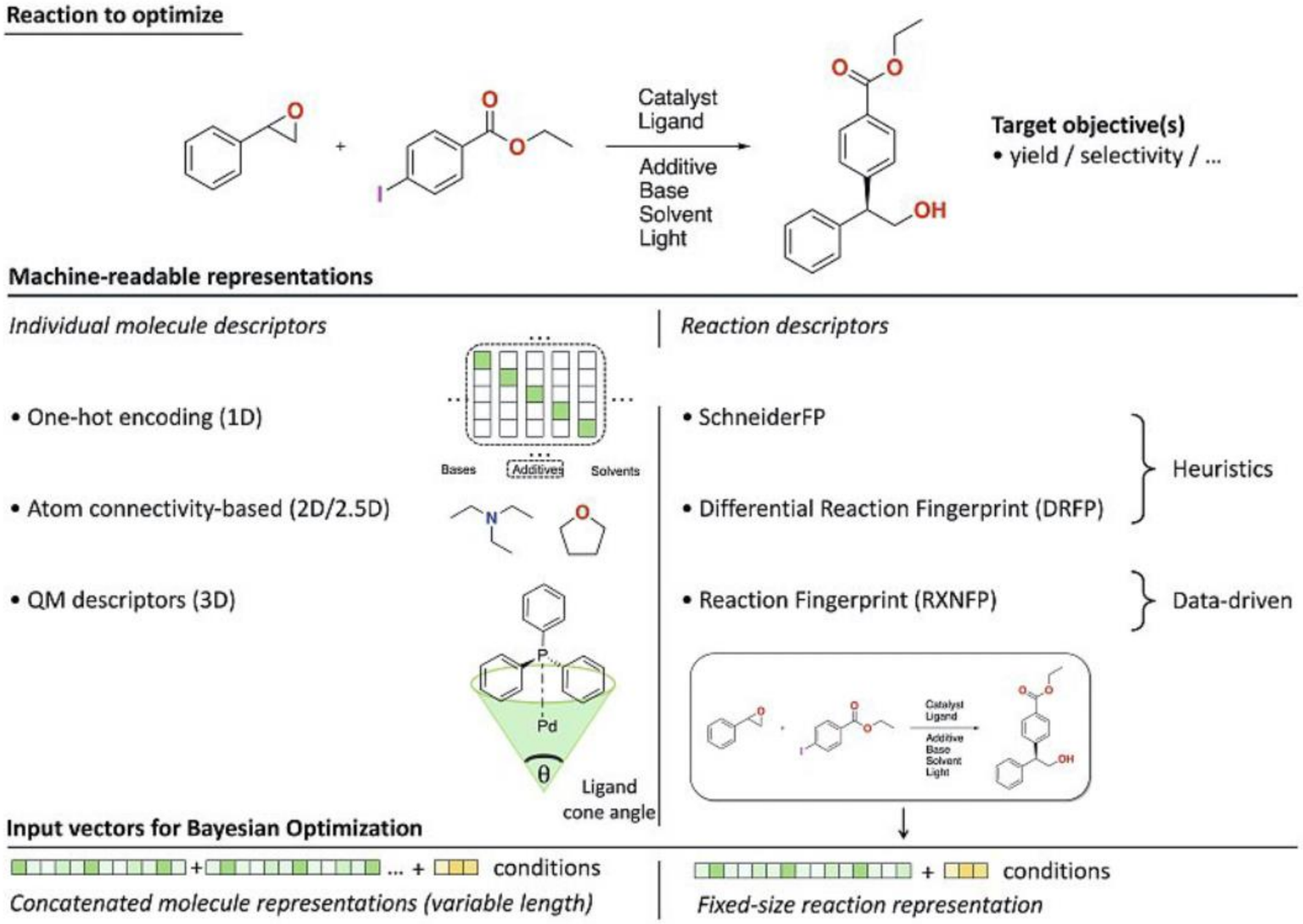


- **Best Practices for Multi-Fidelity Bayesian Optimization in Materials and Molecular Research**

ATINARY

Let’s talk about representations

How to represent chemistry? → plenty of possibilities



x can just be a vector that contains chemistry-related values

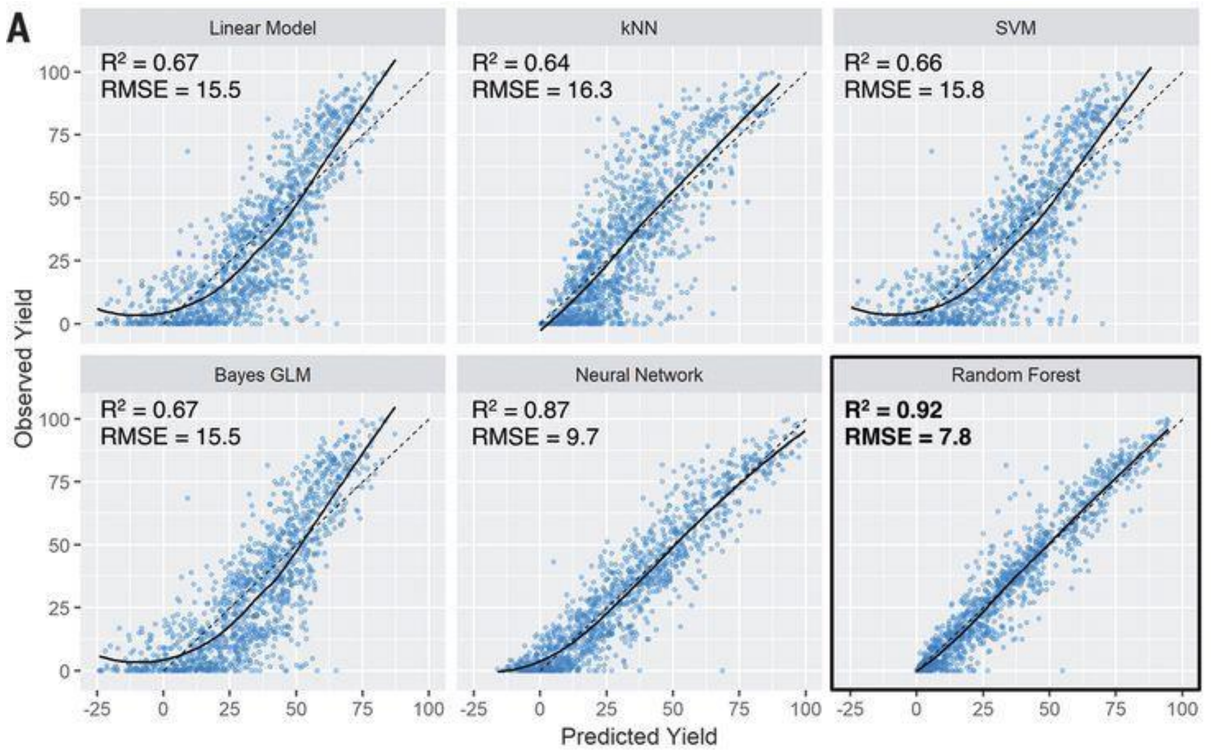
https://chimia.ch/chimia/article/view/2023_31/2023_31

One hot encoding often similar to QM descriptor performance

Predicting reaction performance in C–N cross-coupling using machine learning

DEREK T. AHNEMAN, JESÚS G. ESTRADA, SHISHI LIN, SPENCER D. DREHER, AND ABIGAIL G. DOYLE Authors Info & Affiliations

SCIENCE • 15 Feb 2018 • Vol 360, Issue 6385 • pp. 186-190 • DOI: 10.1126/science.aar5169



Comment on "Predicting reaction performance in C–N cross-coupling using machine learning"

KANGWAY V. CHUANG AND MICHAEL J. KEISER Authors Info & Affiliations

SCIENCE • 16 Nov 2018 • Vol 362, Issue 6416 • DOI: 10.1126/science.aat8603

"don't need QM descriptors, you can get similar performance with one-hot encoding"

You can extract feature importance if you use meaningful descriptors though

One-hot does not always work

But which representation is right for me?

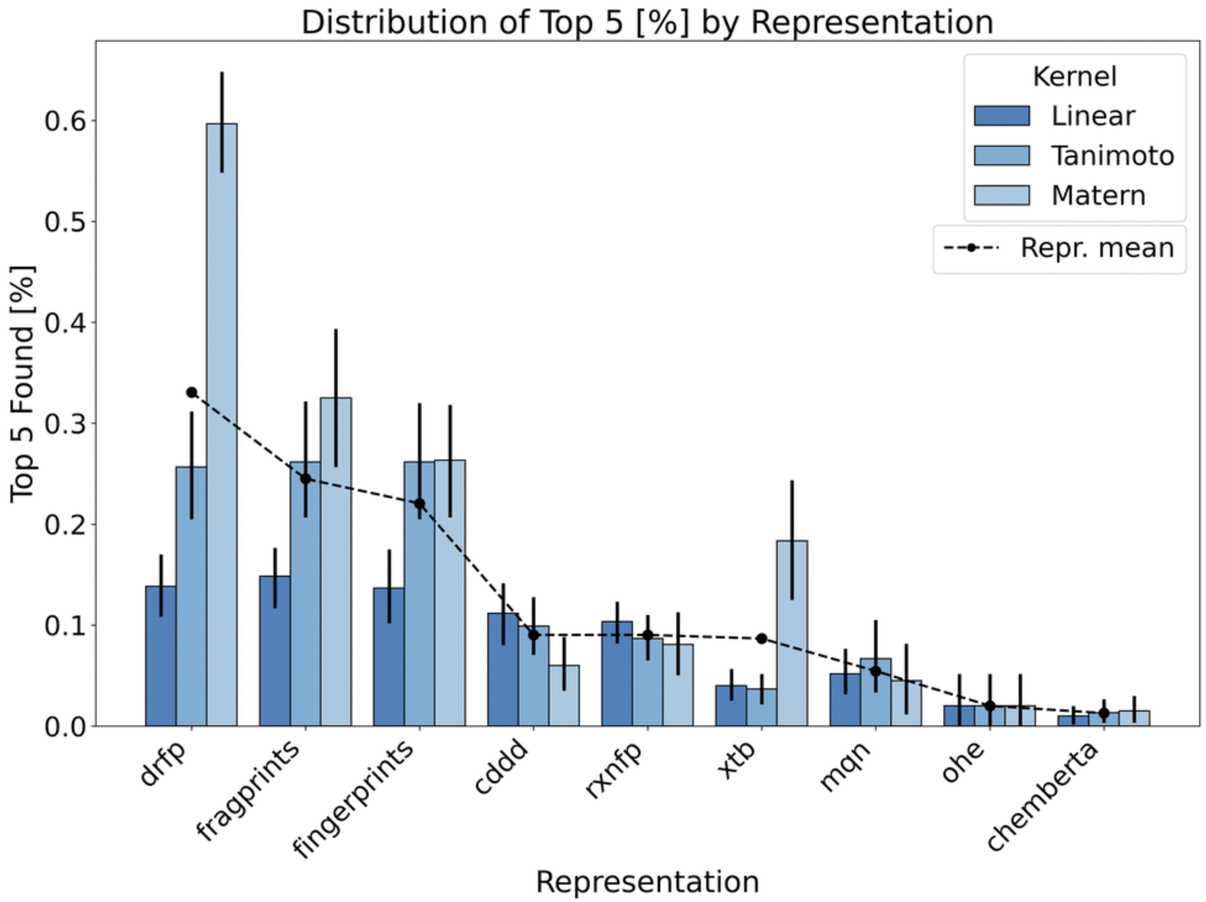
- Better representations = Better BO 🧠😀

But which representation is right for me?

- Better representations = Better BO 🧠😃
- OHE works the best 🤯

But which representation is right for me?

- Better representations = Better BO 🧠😀
- ~~OHE works the best~~ 😷



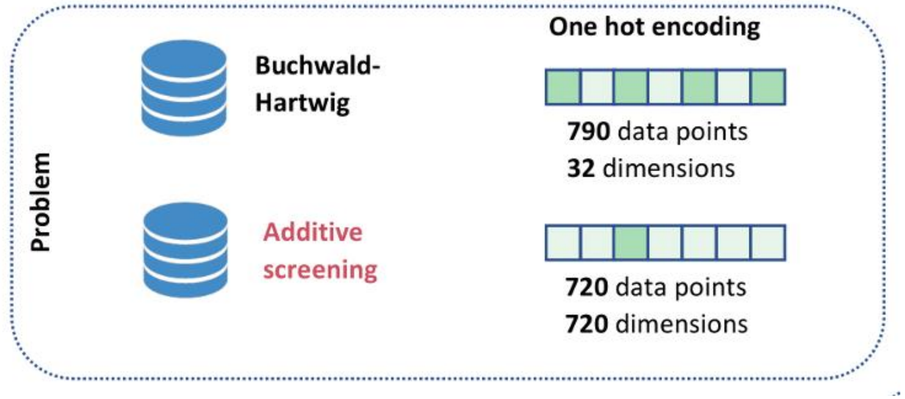
Issue 4, 2024

From the journal:
Digital Discovery

Bayesian optimisation for additive screening and yield improvements – beyond one-hot encoding†

Check for updates

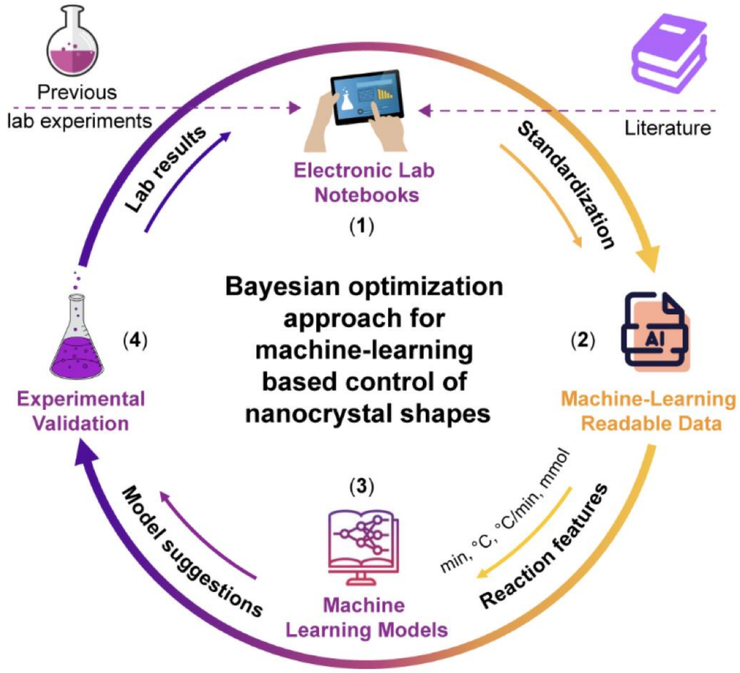
Bojana Ranković, *[a] Ryan-Rhys Griffiths, [b] Henry B. Moss [c] and Philippe Schwaller *[a]



Example of a collaboration with an experimental group

BO for linking synthesis parameters to shapes

Can we design Cu nanocrystals with a particular shape (=> reactivity)?
→ Collaboration with Buonsanti group



- [A holistic data-driven approach to synthesis predictions of colloidal nanocrystal shapes](#)
  LEF Zaza, B Rankovic, P Schwaller, R Buonsanti (JACS)

BO for linking synthesis parameters to shapes

BoLudo (starting questions)

- No SMILES = No DRFP 😔
- What ligand?
- How much?
- What temperature?
- How long?
- What heating ramp?

- How to combine all of that? 😱
- What kernel to use? 😱
- What is the output?

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS

Article

A Holistic Data-Driven Approach to Synthesis Predictions of Colloidal Nanocrystal Shapes

Ludovic Zaza,[+] Bojana Ranković,[+] Philippe Schwaller,* and Raffaella Buonsanti*

Cite This: *J. Am. Chem. Soc.* 2025, 147, 6116–6125
Read Online

BO for linking synthesis parameters to shapes

BoLudo (starting questions)

- No SMILES = No DRFP 😔
- What ligand?
- How much?
- What temperature?
- How long?
- What heating ramp?

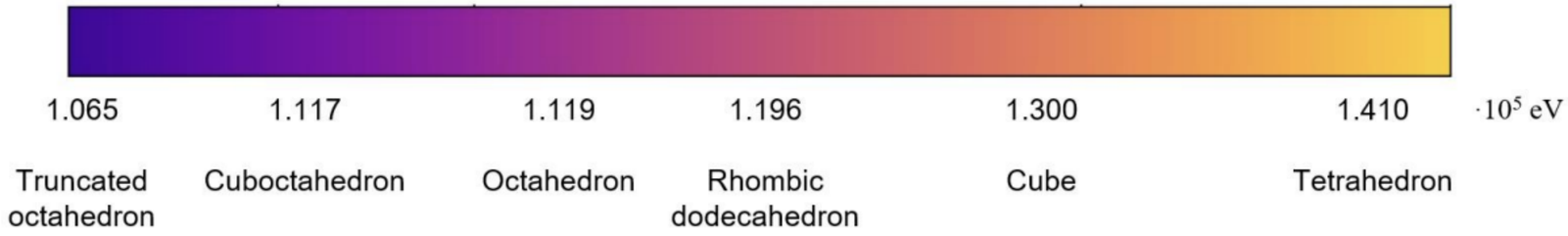
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS

Article

A Holistic Data-Driven Approach to Synthesis Predictions of Colloidal Nanocrystal Shapes

Ludovic Zaza,[+] Bojana Ranković,[+] Philippe Schwaller,* and Raffaella Buonsanti*

Cite This: *J. Am. Chem. Soc.* 2025, 147, 6116–6125
Read Online



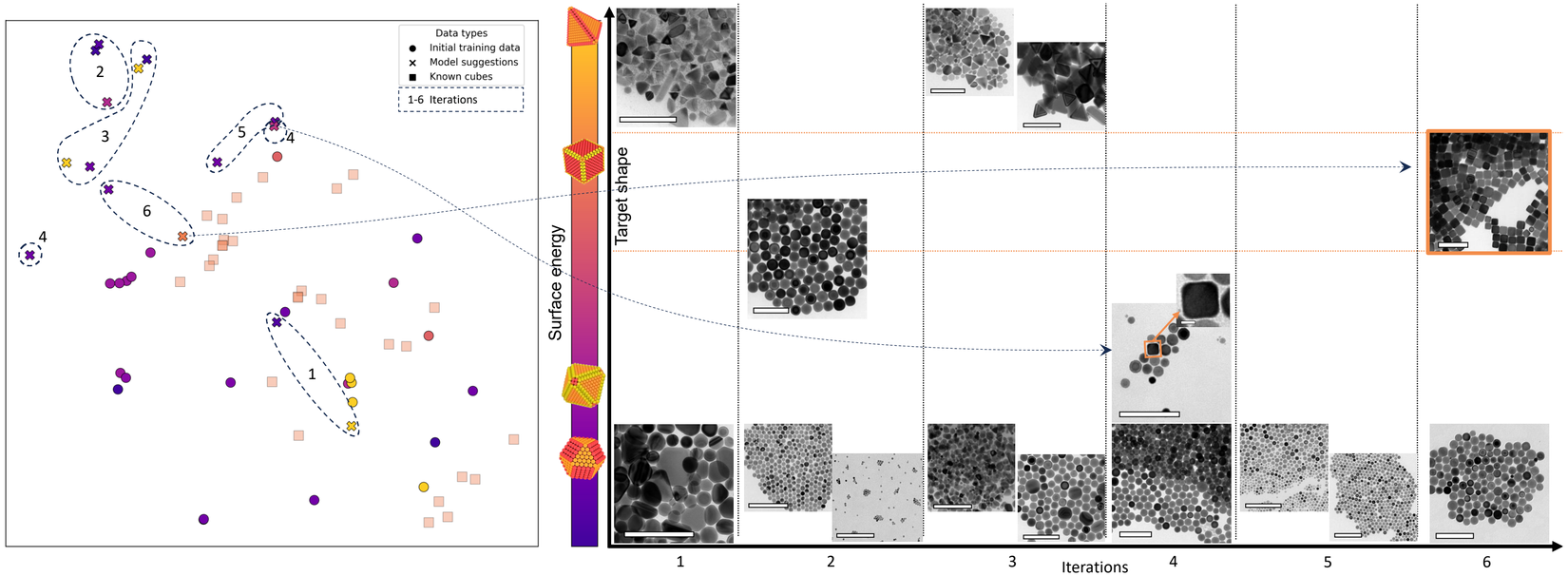
OHE+mmol quantities

oleylamine	tri-n-octylphosphine oxide	copper(I) bromide
139.292970	0.0	0.000000
145.349186	0.0	0.000000
148.377295	0.0	0.000000
30.281081	25.0	3.000000

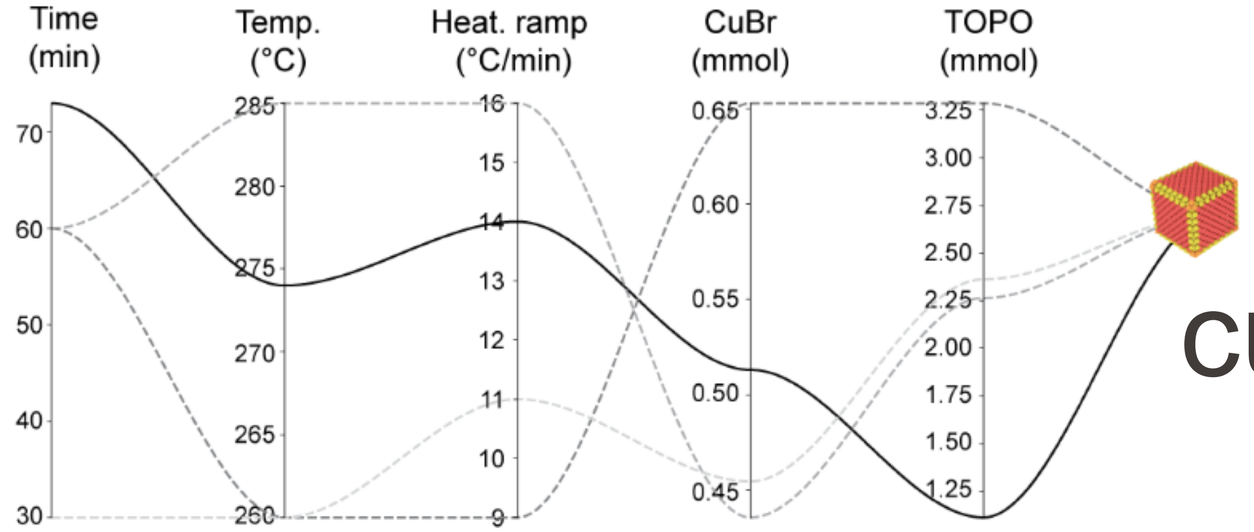
Numerical

time	temp	heating_ramp
30.0	335.0	11.0
30.0	335.0	11.0
15.0	335.0	11.0
60.0	210.0	11.0

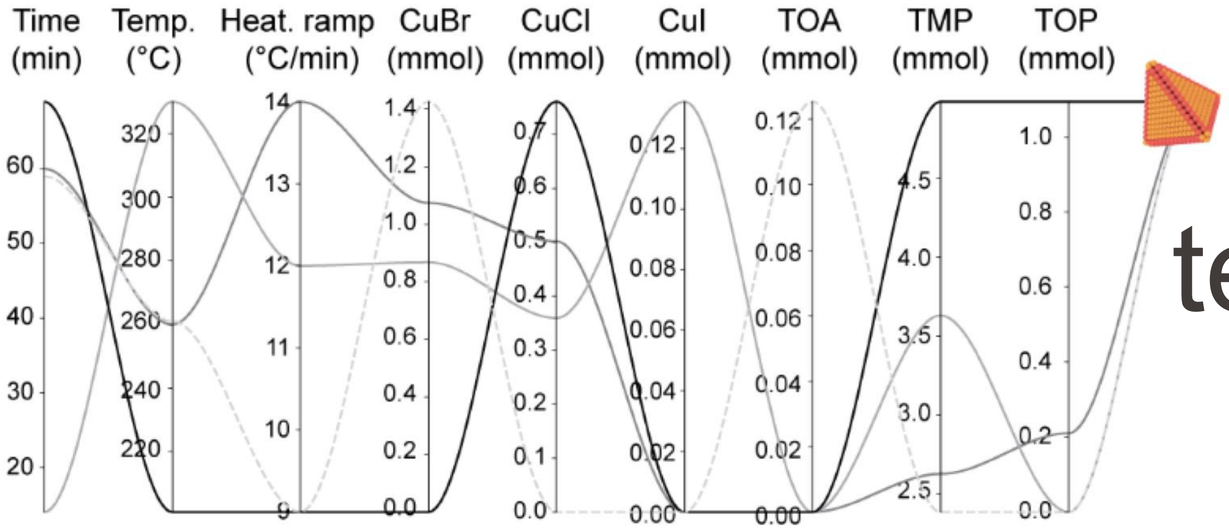
Optimizing for known cubes (removing them from training set)



Multiple syntheses lead to the same outcome (ground truth not unique)



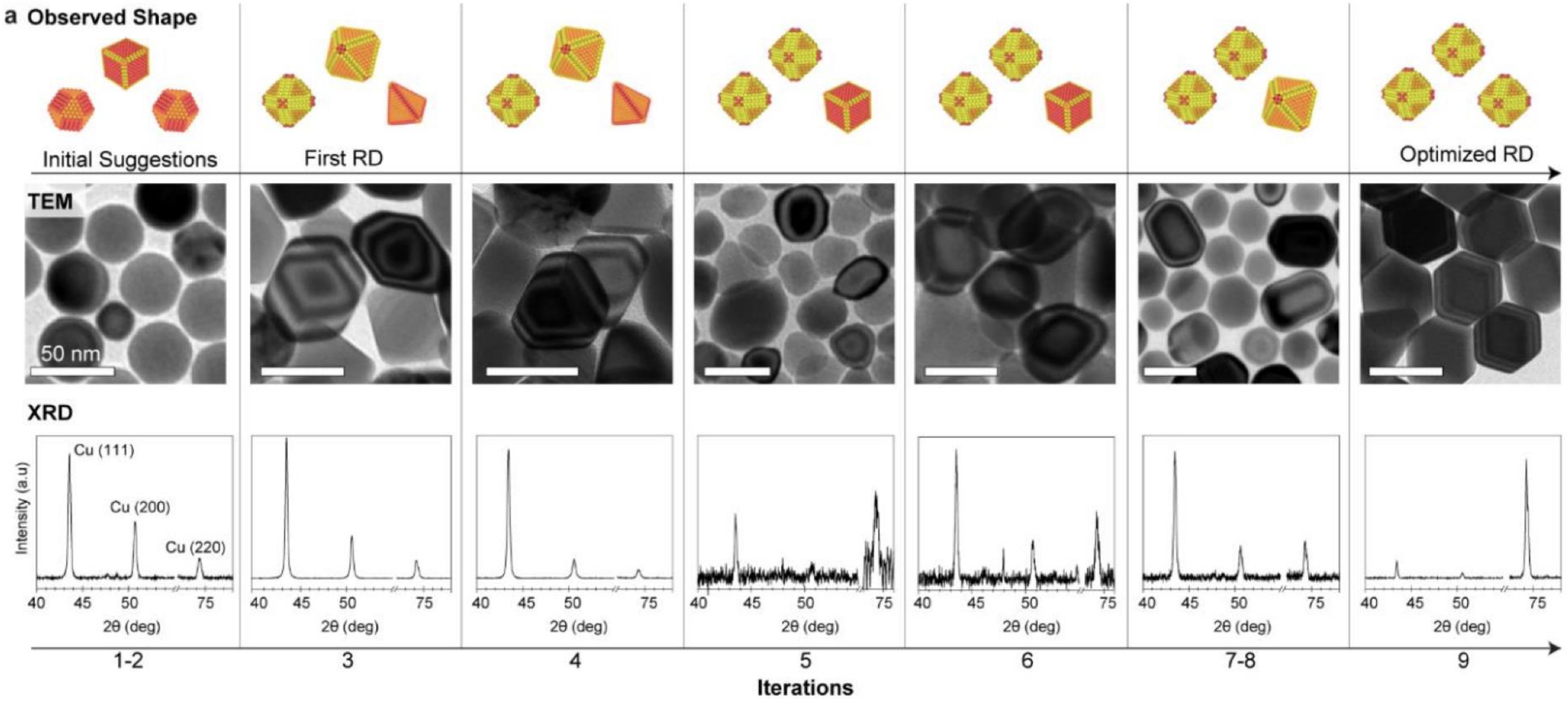
cubes



tetrahedral

Unseen Syntheses from Dataset — BO Discovered Syntheses

Discovering a new shape (rhombic dodecahedral)



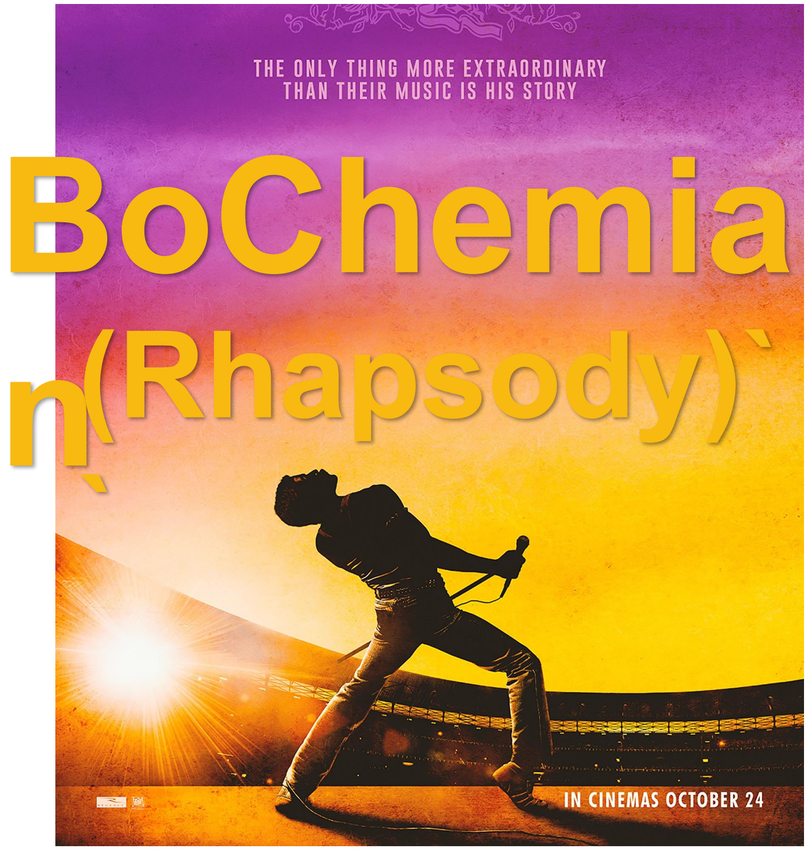
Back to representations…

Could we find a more holistic representation?

Why not represent all of it as text?

And use an LLM to encode the text to a vector?

And input that vector to a GP?



Starring: Bojana Ranković, Philippe Schwaller

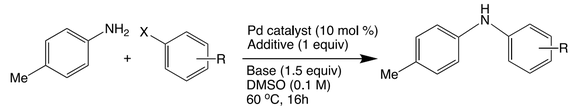
- **Flexible**

```
The reaction was prepared with:
temperature: {numerical_value}°C
solvent: {solvent_smile}
ligand: {ligand_smile}
```

- **Continuous** `Matérn kernel`

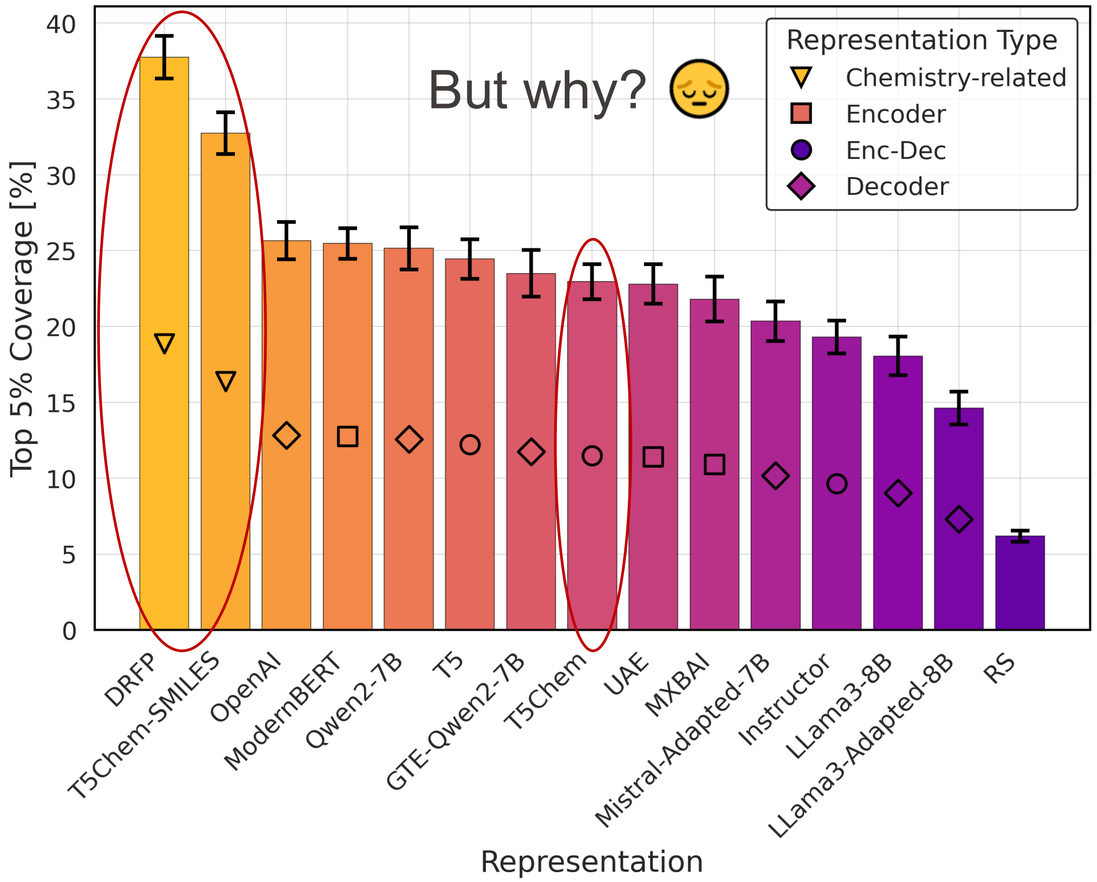
- **Unified representation** `categorical, numerical, smiles, iupac`

Fixed LLM embeddings as an input to the GP



BoChemian: Large Language Model Embeddings for Bayesian Optimization of Chemical Reactions

Bojana Ranković, Philippe Schwaller



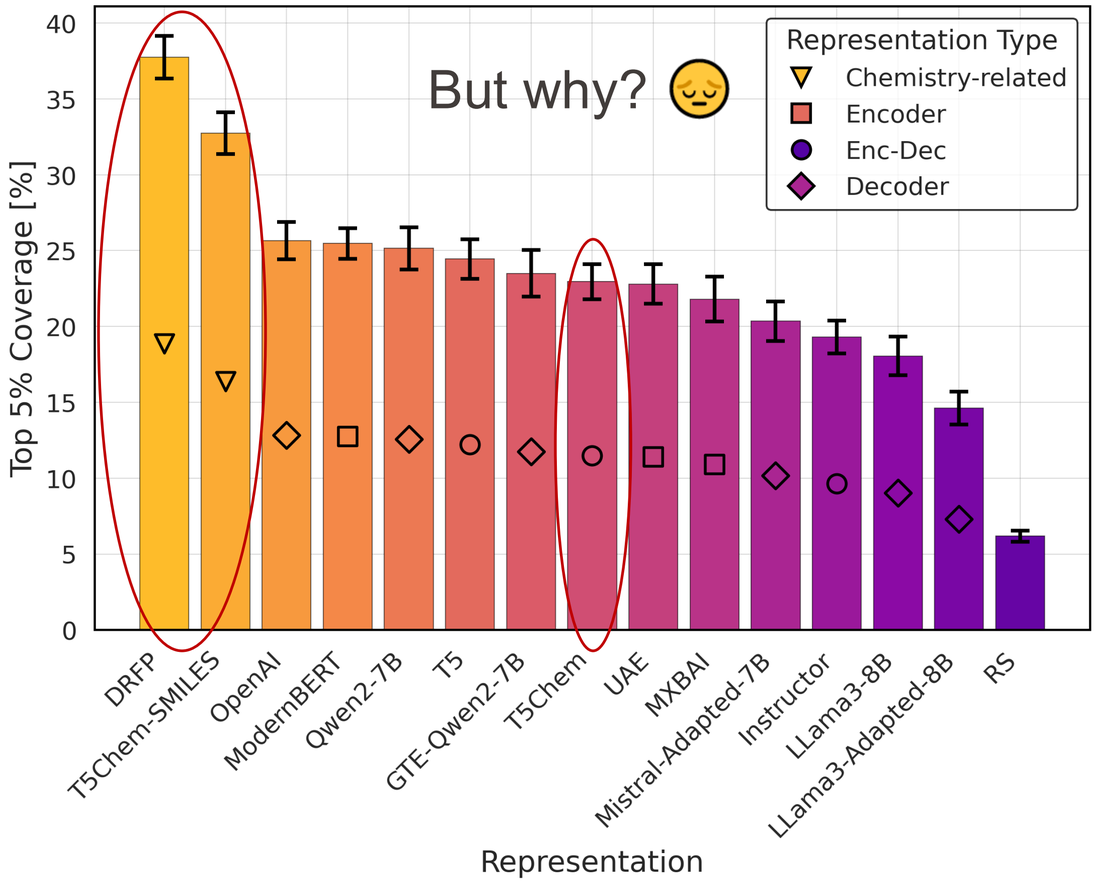
Fixed LLM embeddings as an input to the GP

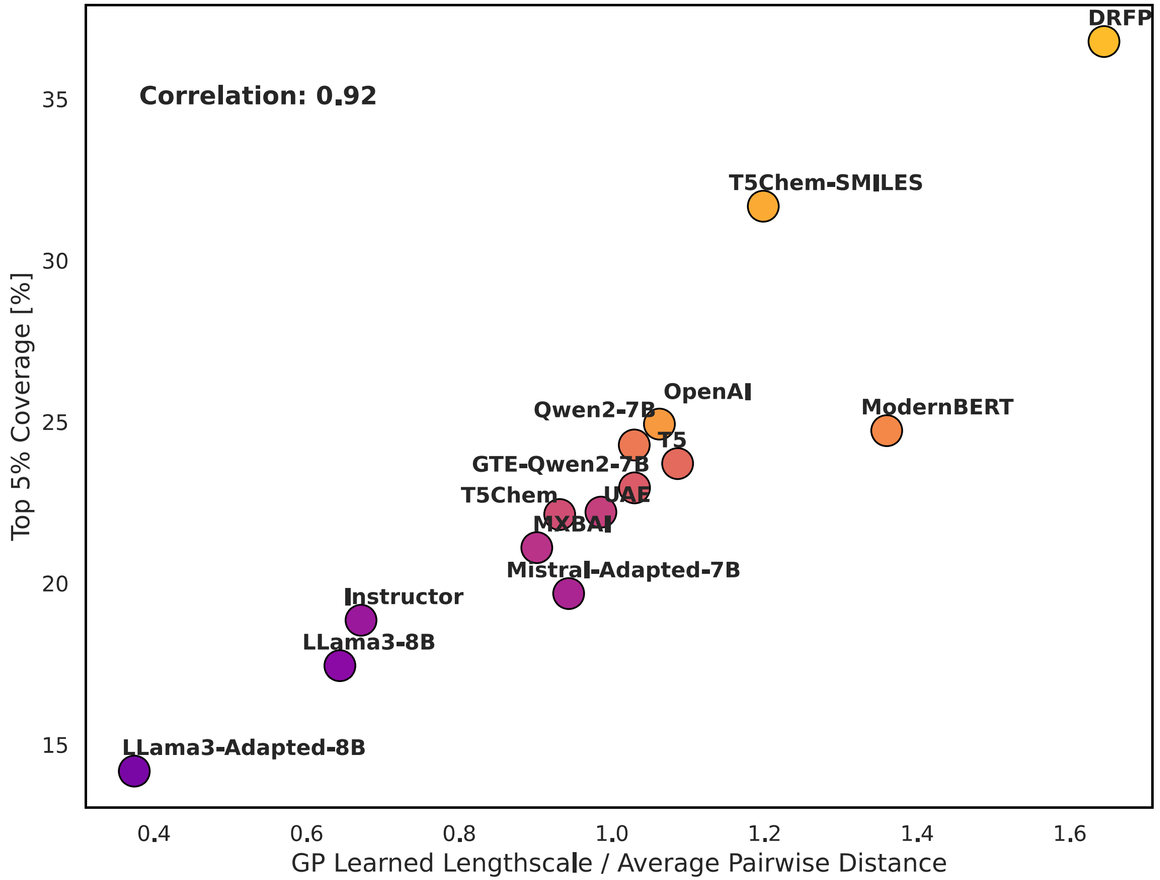
BoChemian: Large Language Model Embeddings for Bayesian Optimization of Chemical Reactions

Bojana Ranković, Philippe Schwaller

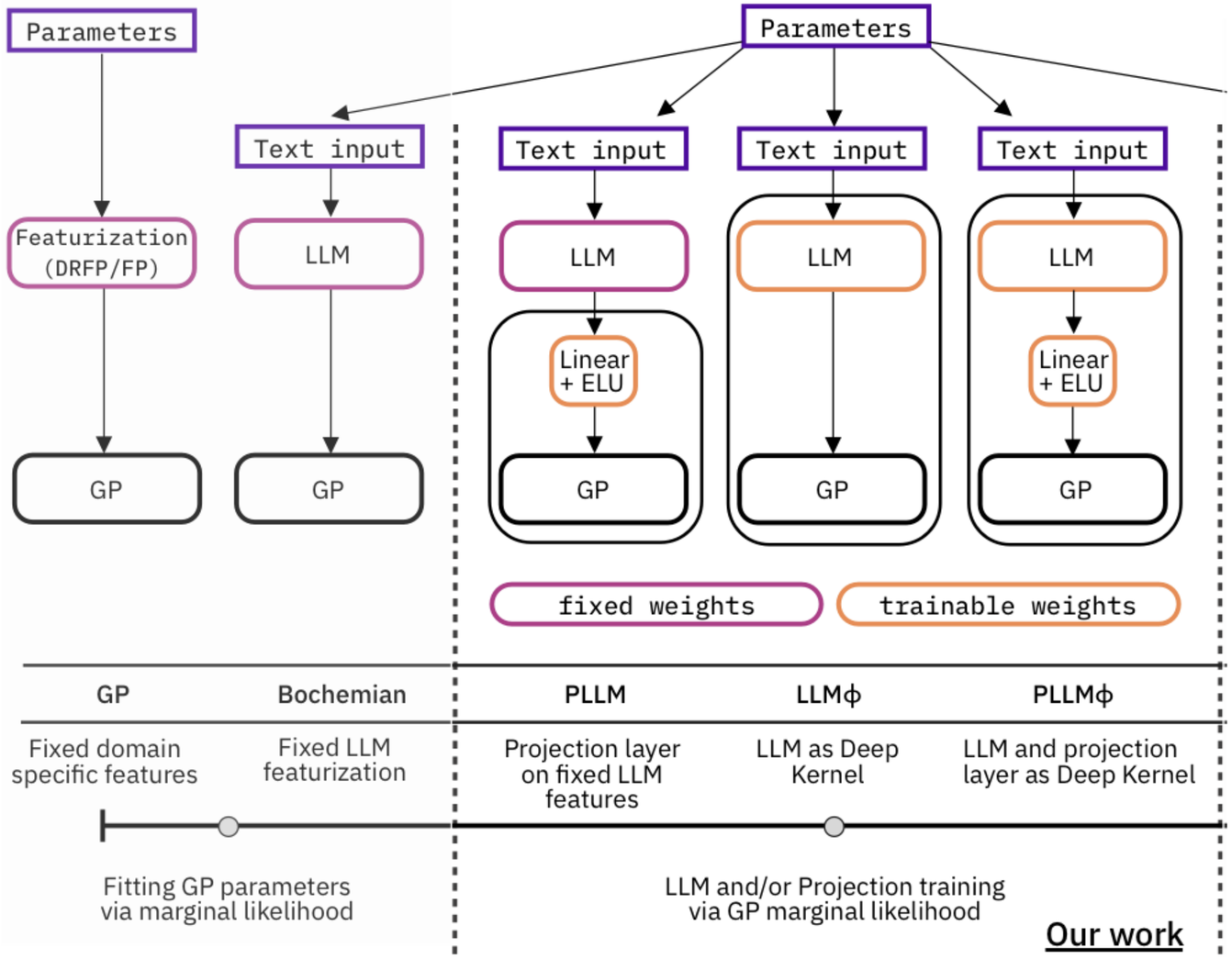
... it's about structure

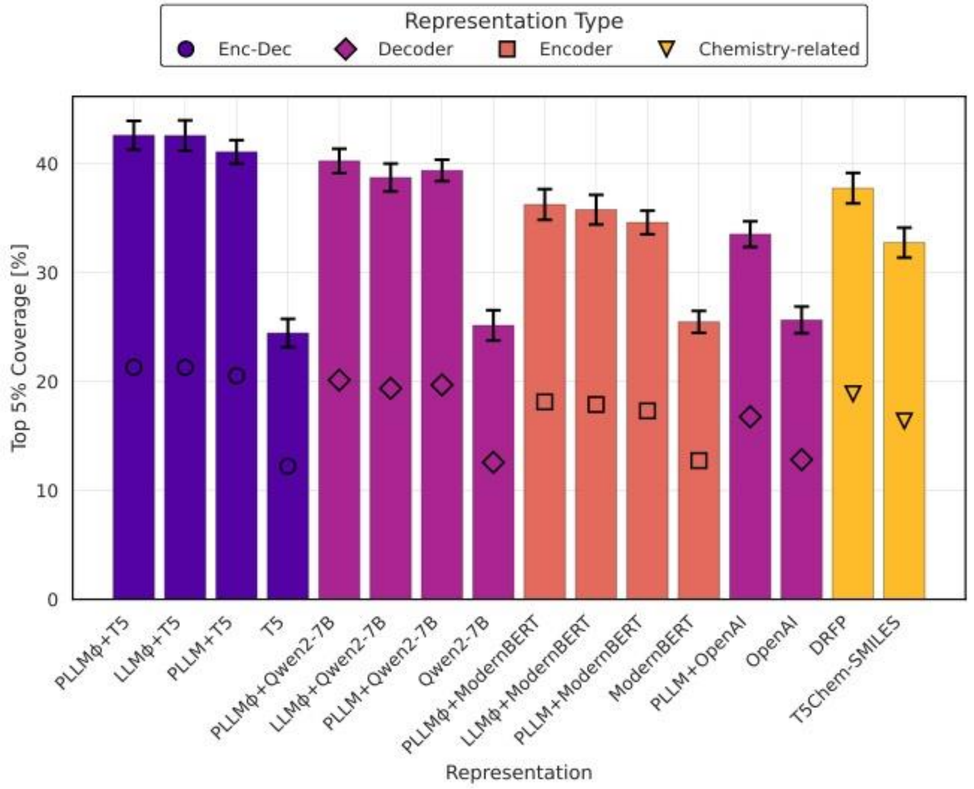
Why do some embeddings perform better 👇





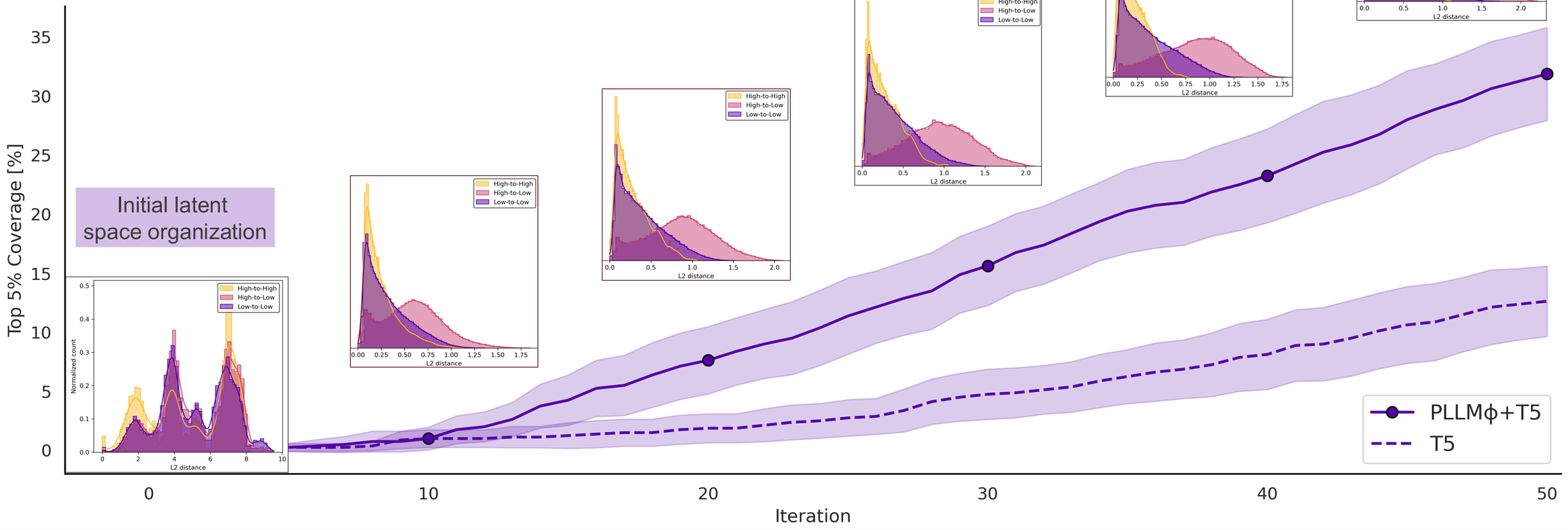
Training LLM embeddings jointly with Gaussian Process





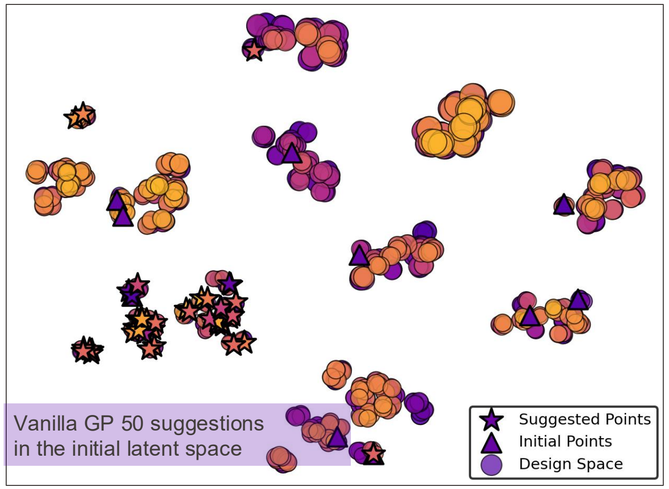
Implicit metric learning with LLM-based Deep Kernel Gaussian process surrogates

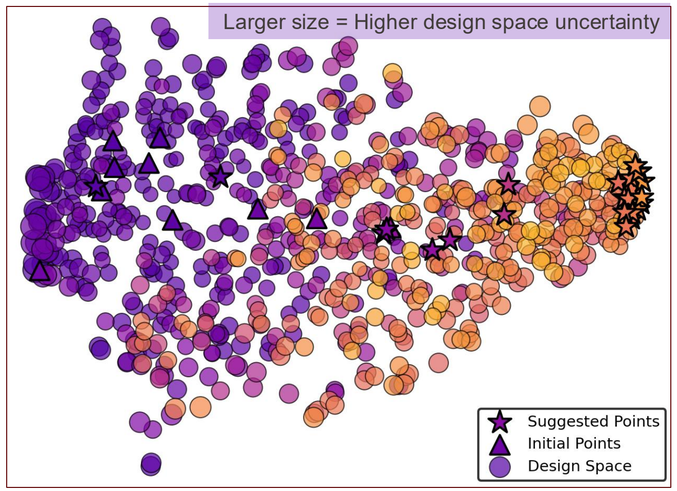
Implicit contrastive learning effect in the latent space

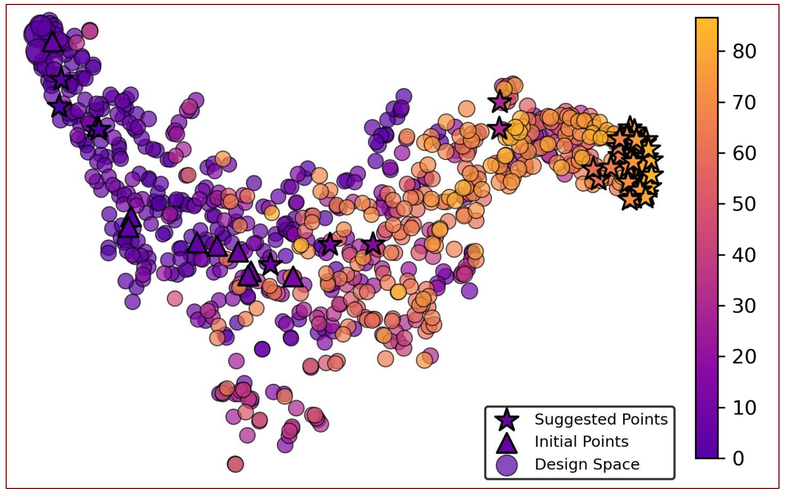


Embedding space reorganisation during optimization

Latent space and suggested points during the 50 iterations of optimization







Iteration: 0 Iteration: 25 Iteration: 50

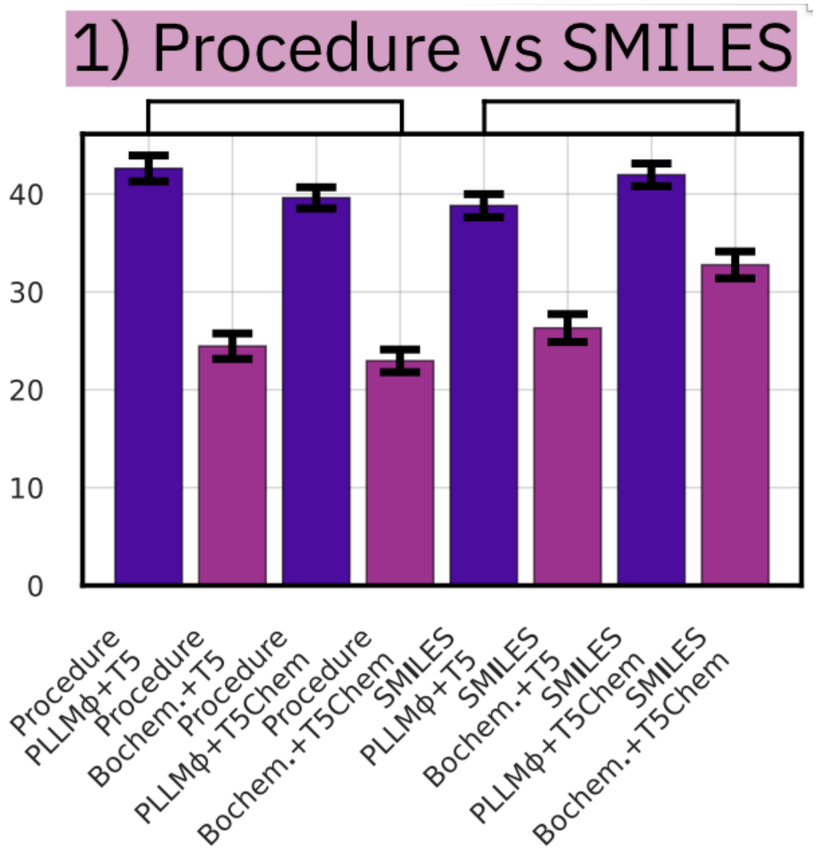
Extensive evaluation of LLM-based Deep Kernel in Chemistry

Reaction SMILES

```
CCc1ccc(I)cc1...COC(=O)
c1ccno1>>CCc1ccc(Nc2ccc
(C)cc2)cc1
```

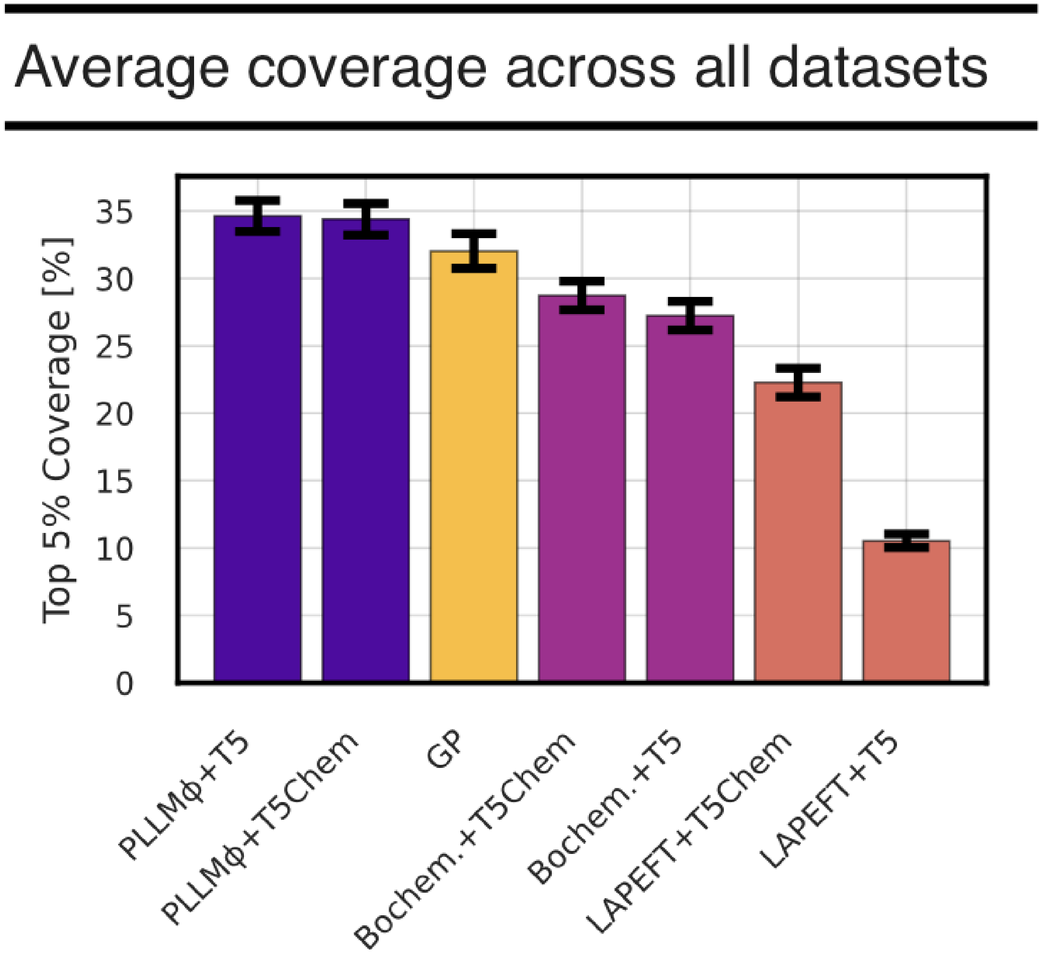
Simplified procedure

```
Reaction Setup
 - Reactant:
IC1=CC=C(CC)C=C1
 - Additive:
O=C(OC)C1=CC=NO1
        ...
```



Robust to input formatting ✅

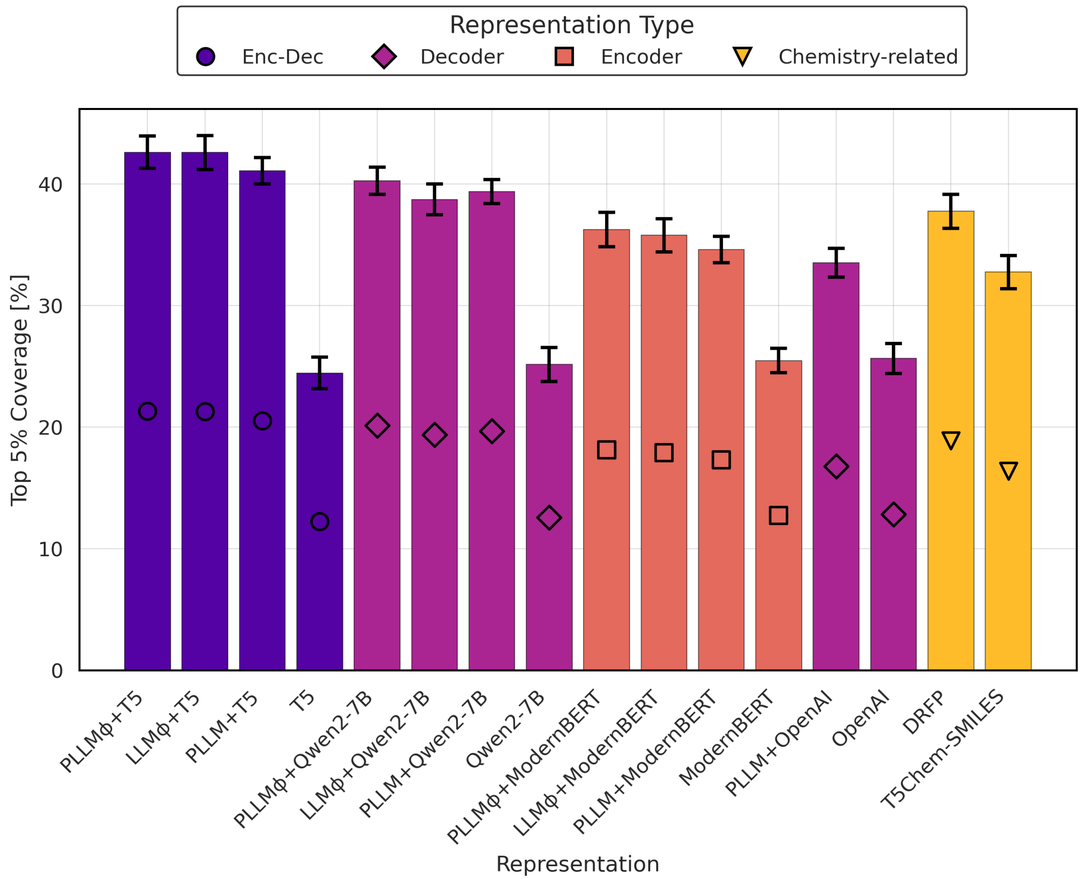
Extensive evaluation of LLM-based Deep Kernel in Chemistry



Robust to input formatting ✅

Robust to pretraining ✅

Extensive evaluation of LLM-based Deep Kernel in Chemistry

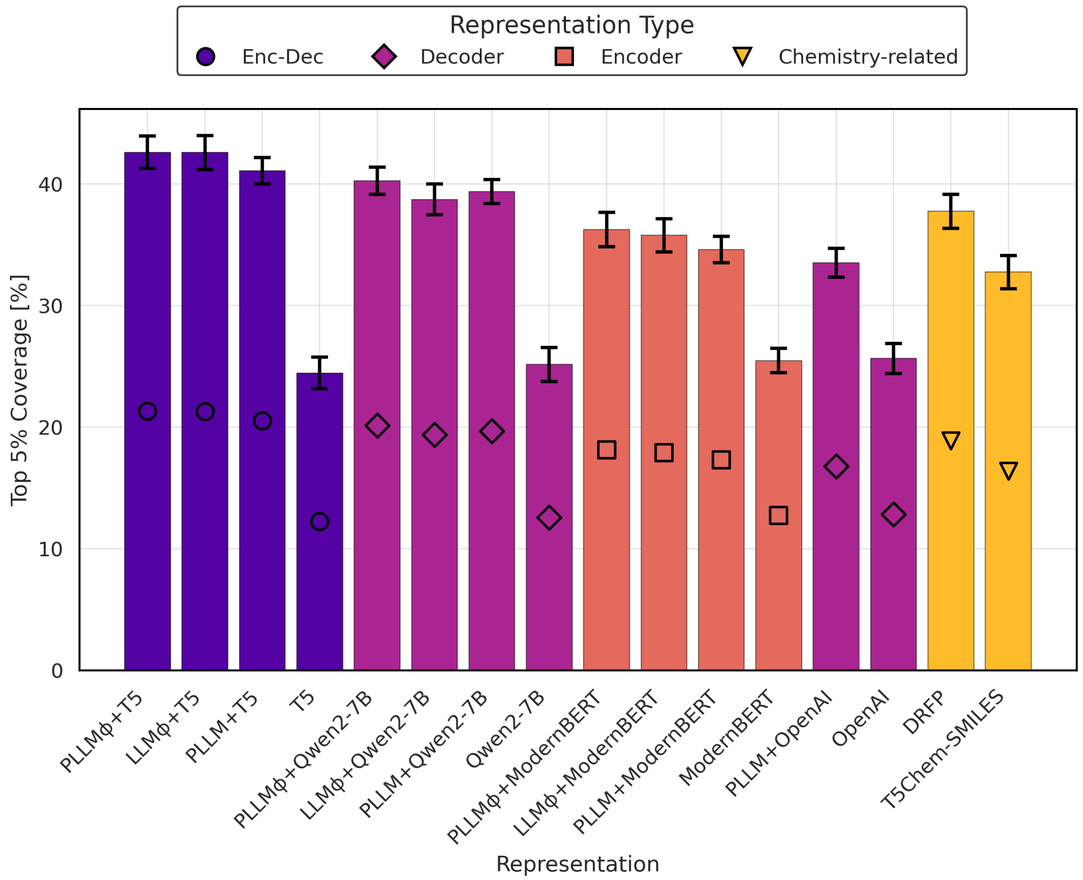


Robust to input formatting ✅

Robust to pretraining ✅

Robust to architecture ✅

Extensive evaluation of LLM-based Deep Kernel in Chemistry



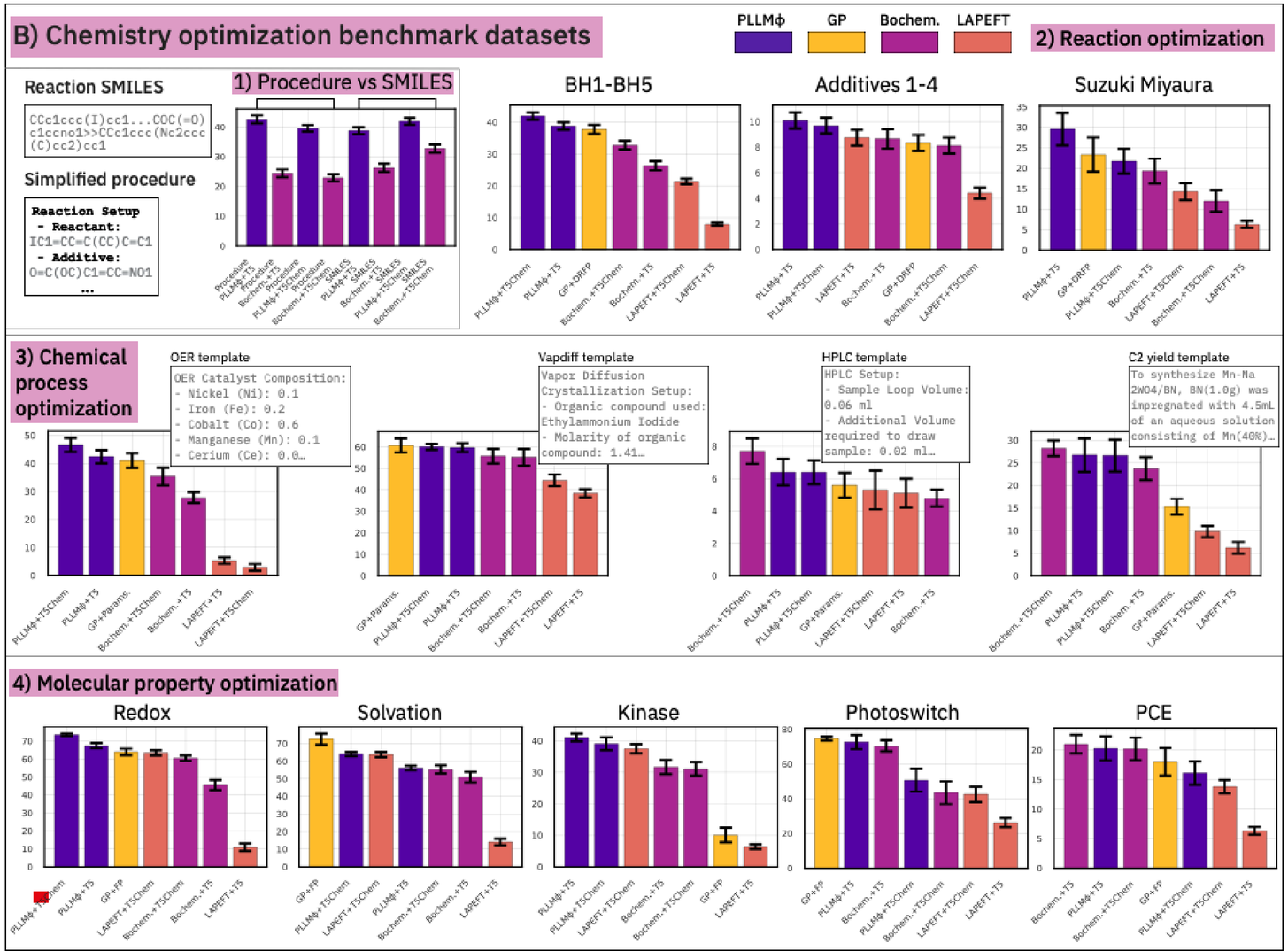
Robust to input formatting ✅

Robust to pretraining ✅

Robust to architecture ✅

Robust to hyperparametr. ✅

Extensive evaluation of LLM-based Deep Kernel in Chemistry



Robust to input formatting ✅
Robust to pretraining ✅
Robust to architecture ✅
Robust to hyperparametr. ✅
Robust to tasks ✅

19 datasets:
14 reaction/process optimization
5 molecular optimization

GOLLuM: Gaussian Process Optimized Large Language Models

6 Future work

This research underscores the potential of using BO for accelerating additive discovery in chemical reactions, paving the way for more efficient experimental design and optimisation in the field of chemistry. The reaction type and its unique chemical features influence the performance of specific chemical representations in the optimisation process. In addition, devising methods to evaluate the fit of different representations for distinct sets of reactions could enhance the optimisation process, leading to more accurate and reliable results.

Future research should focus on determining the optimal reaction representation, or possibly a dynamic combination of representations for employing bo on different reaction types while incorporating domain knowledge. For example, switching from one reaction representation to

Issue 4, 2024

Previous Article | Next Article

From the journal:
Digital Discovery

Bayesian optimisation for additive screening and yield improvements – beyond one-hot encoding†

Check for updates

Bojana Ranković, *a Ryan-Rhys Griffiths, b Henry B. Moss c and Philippe Schwaller *a

GOLLuM learns it 🎯

https://github.com/schwallergroup/gollum

Challenges

- **Can Bayesian optimization be practical in large design spaces?**

→ Imagine you want to optimize a chemical reaction

→ There are > 10^6 possible combinations of reagents

→ Can BO find a good solution under a practical budget?

- **Can we trust the uncertainty quantification of Gaussian Processes? Random Forests? Neural Networks?**

→ If the acquisition function uses uncertainty, this can be detrimental to overall performance

- **If I don't have any initial data, how can I choose the initial experiments to try? What are the implications on the overall BO trajectory?**

- **How can ML practitioners work with experimentalists best?**

→ If you apply BO and it suggests performing a reaction due to exploration and the outcome is bad, this is discouraging for the experimentalist but useful for the model

→ How do we navigate these situations?

Next week:

Dr. Edvin Fako on “How Computational Methods and Data Shape AI for Discovery”

I’ll be at the Bürgenstock conference.