UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

LEVERHULME TRUST

>_Alchemy
AI FOR CHEMISTRY

GPSS 2025

Accelerating Chemical Discovery with Knowledge Guided Bayesian Optimisation

Xenofon Evangelopoulos

Computational/AI Lead
Leverhulme Research Centre for Functional Materials

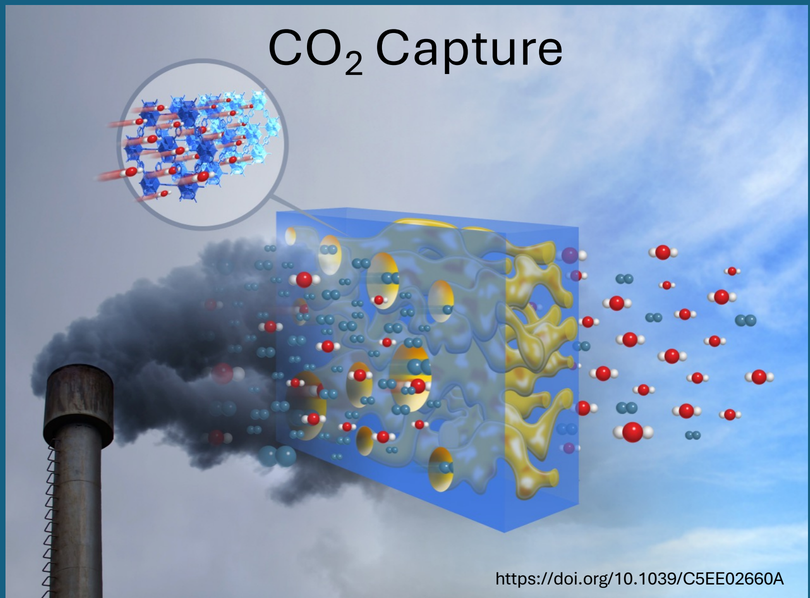
UNIVERSITY OF LIVERPOOL
MATERIALS INNOVATION FACTORY

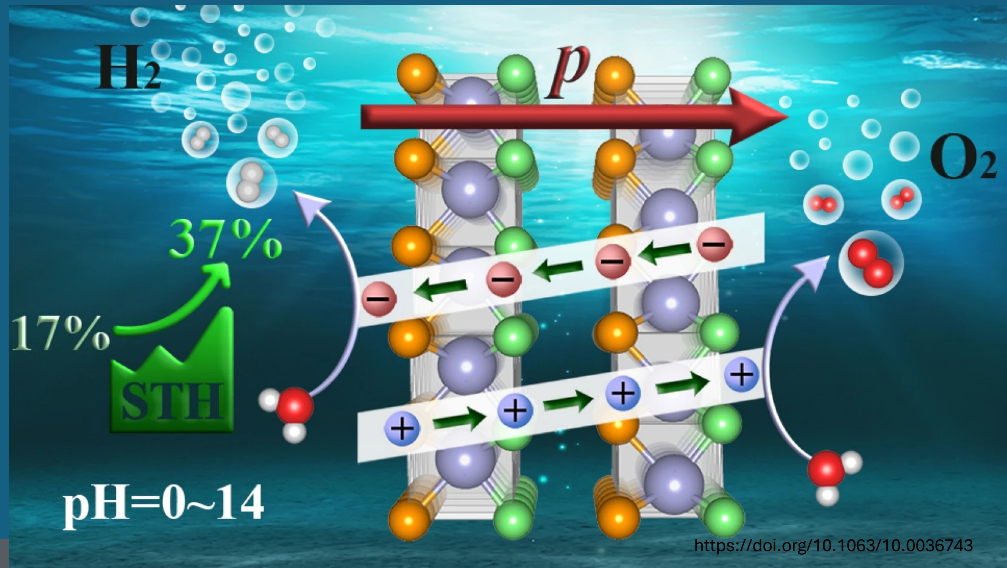
CLIMATE CHANGE

UNIVERSITY OF LIVERPOOL

MATERIALS INNOVATION FACTORY

FUNCTIONAL MATERIALS





UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

MATERIALS MANUFACTURING TIMESCALE

Idea/Design → Synthesis → Testing → Manufacturing

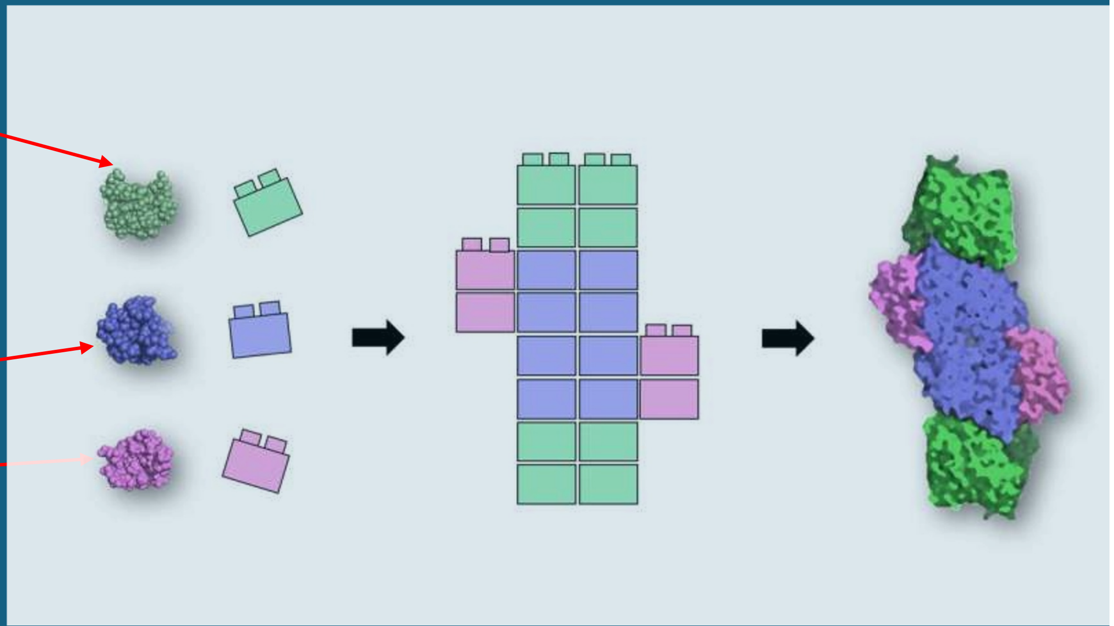
> 10 Years

> £10M

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

UNIVERSE OF MOLECULES

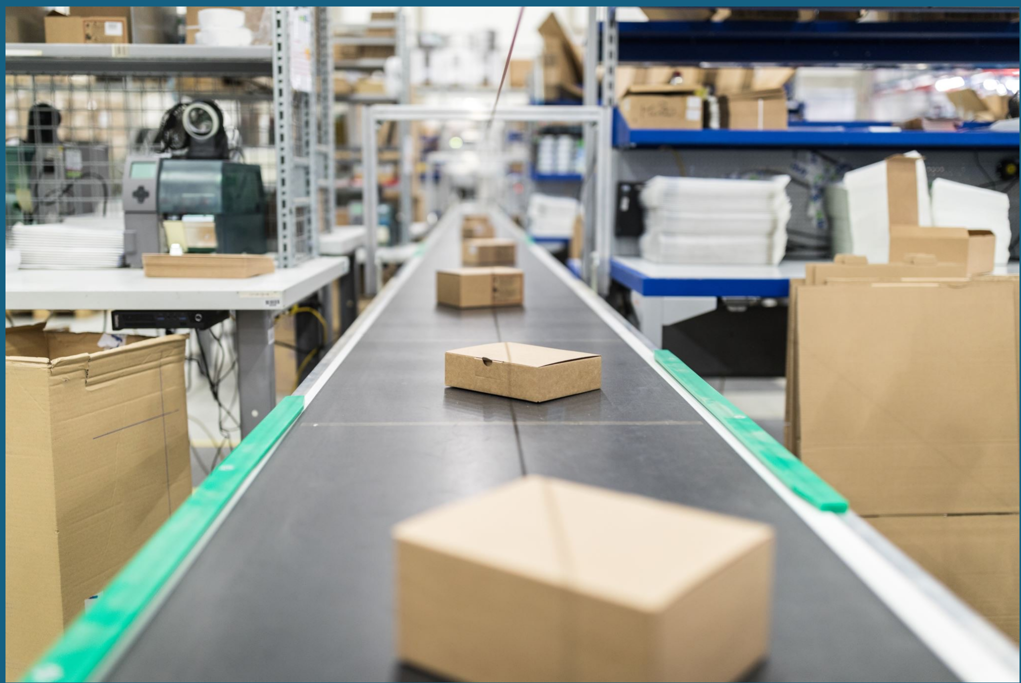
Building/Lego Blocks

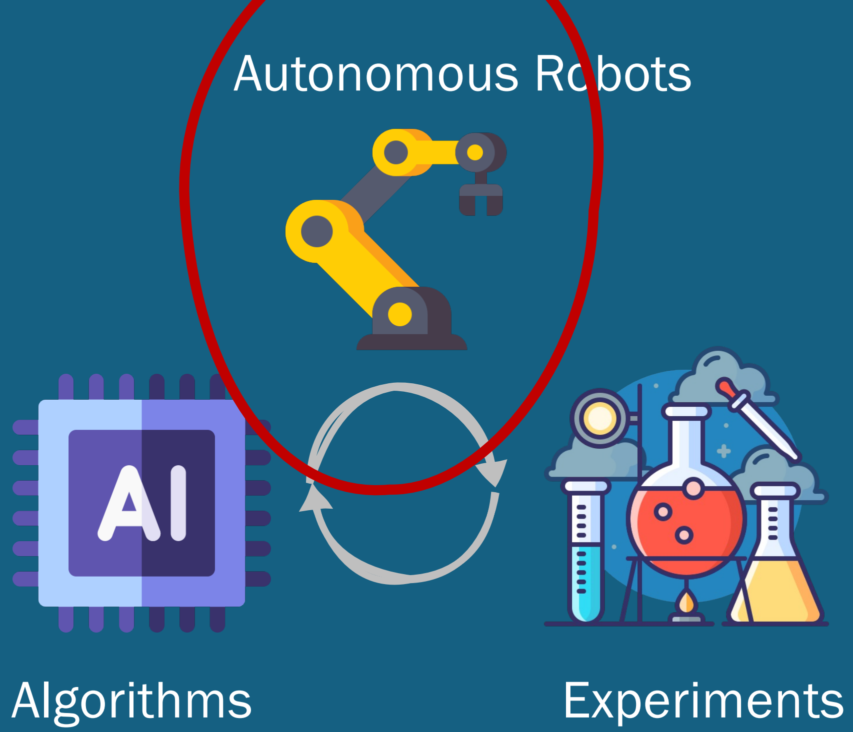


 # Chemical Compounds = # Stars in the Universe x 2

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

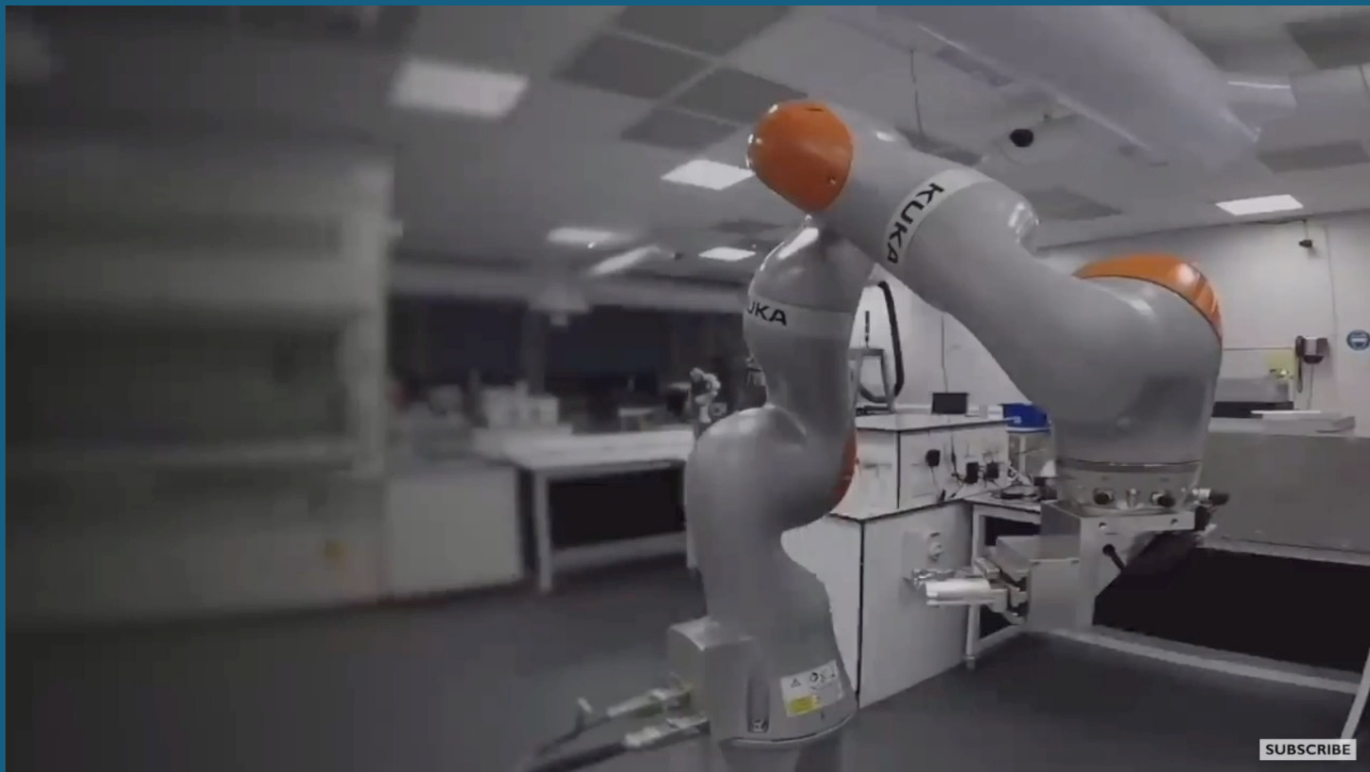
TECHNOLOGY INNOVATION





MATERIALS INNOVATION FACTORY

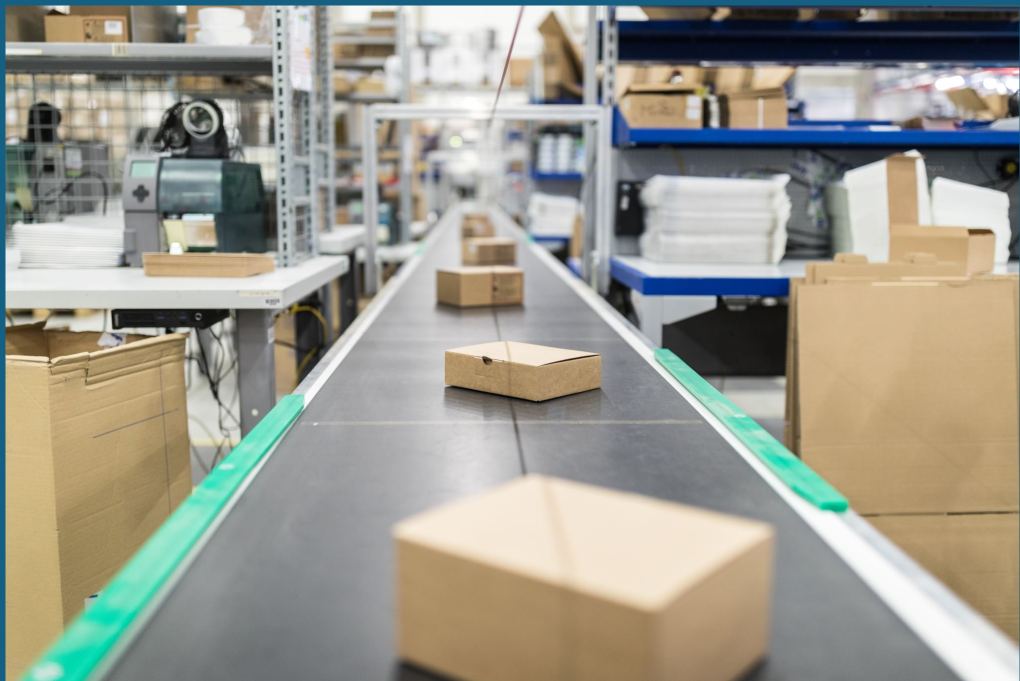
MOBILE ROBOTIC CHEMIST

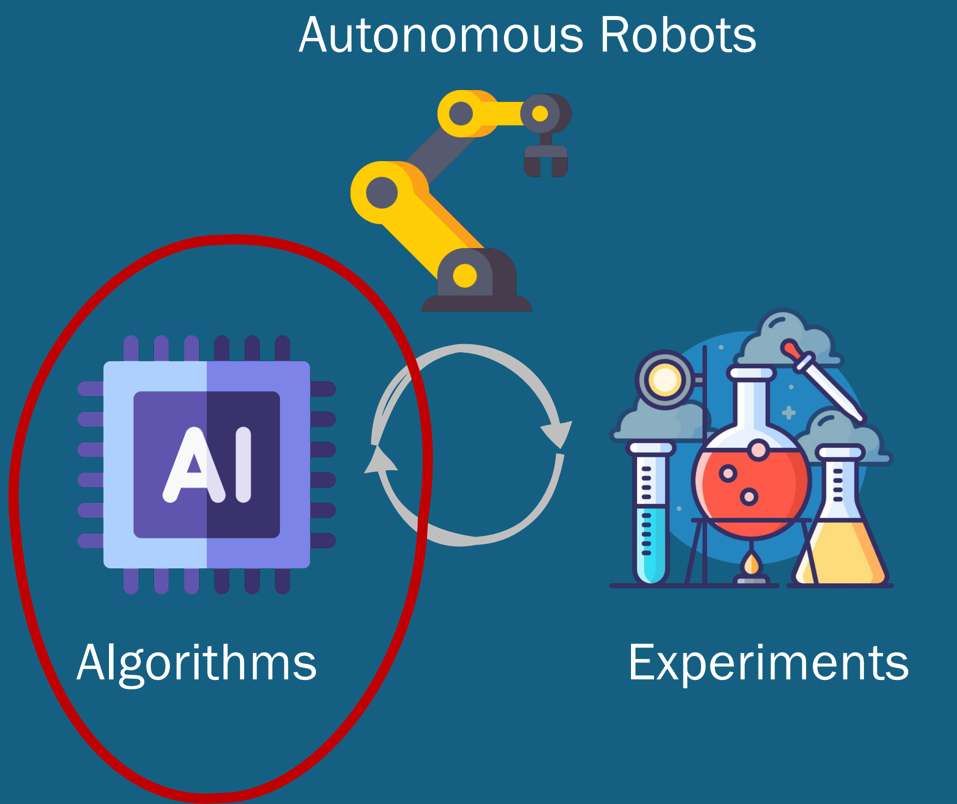


BBC News Feature, July 2020: www.youtube.com/watch?v=WBwZp5Bg2L8

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

TECHNOLOGY INNOVATION

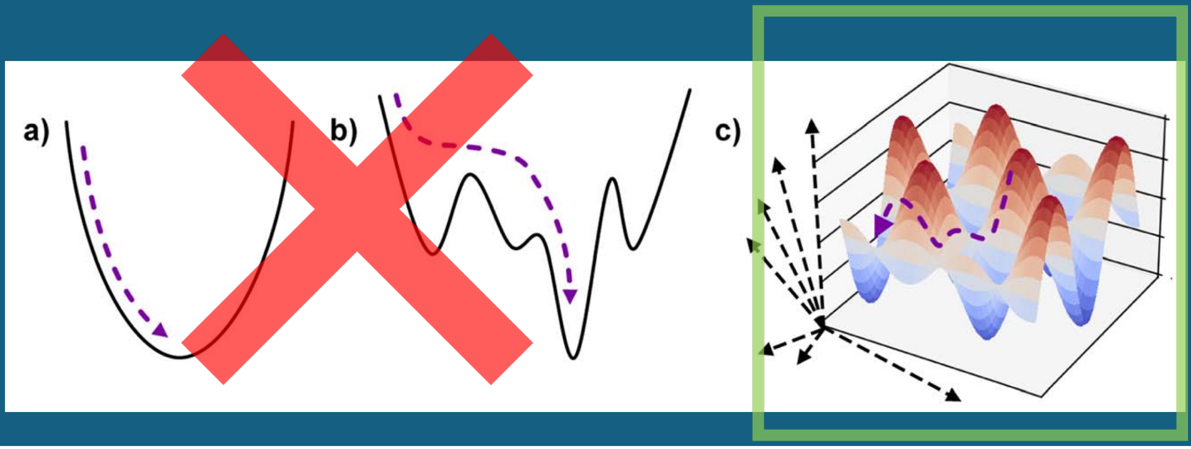




UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

CHEMICAL DISCOVERY – OPTIMISATION

No analytic formula → Hydrogen production | Reaction yield | Organic solid-state lasers

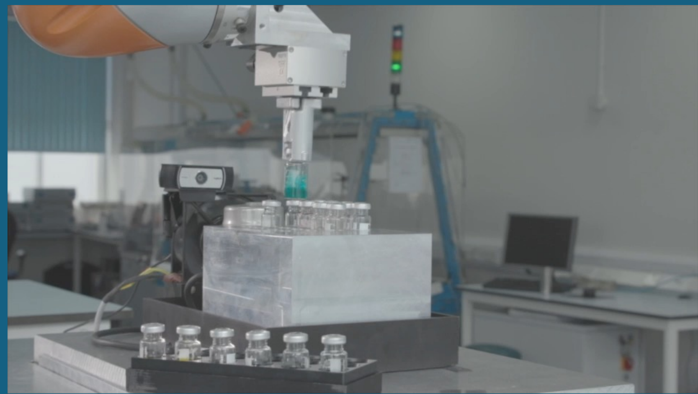


- Non-convex
- Not continuous
- High-dimensional

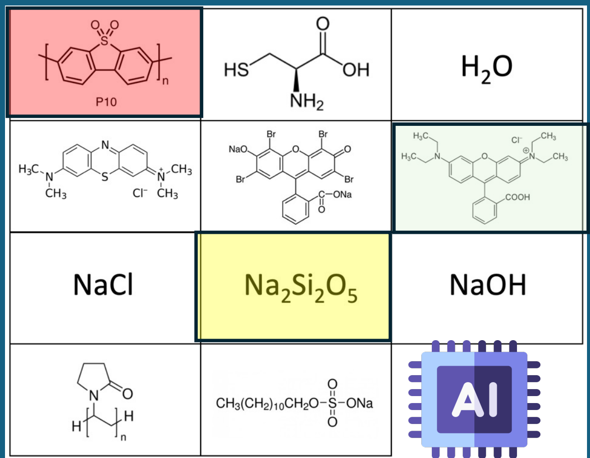
→ Active Learning

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

H_2 PRODUCTION OPTIMISATION



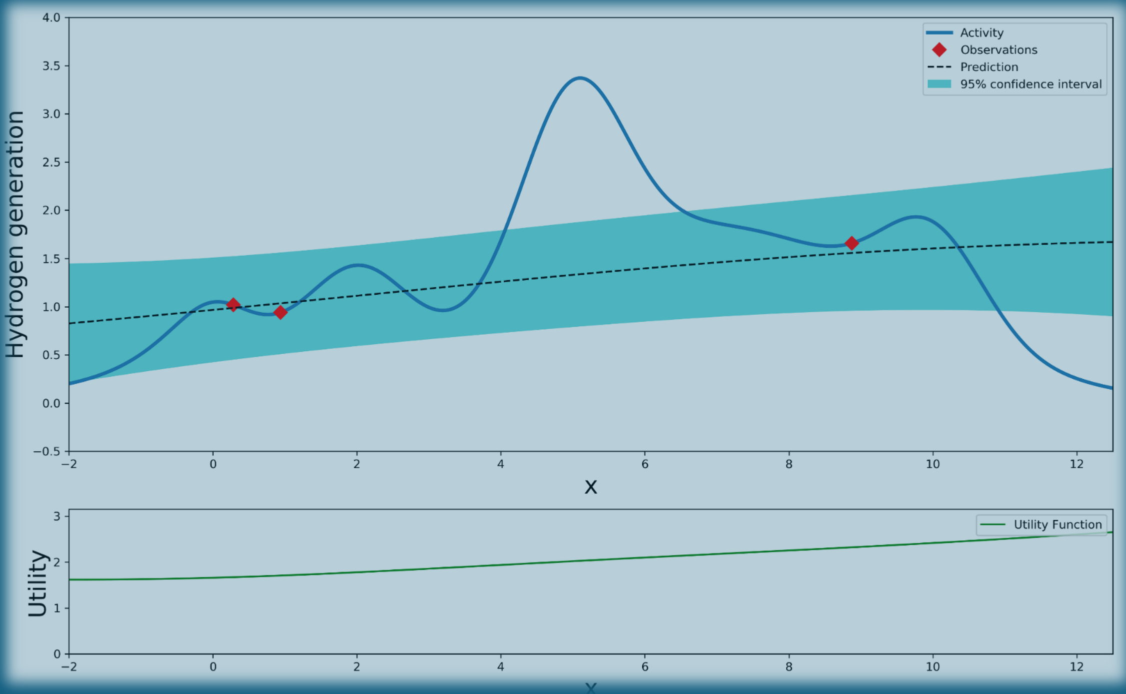
11 catalyst formulation components
→ 98.4 M possible combinations



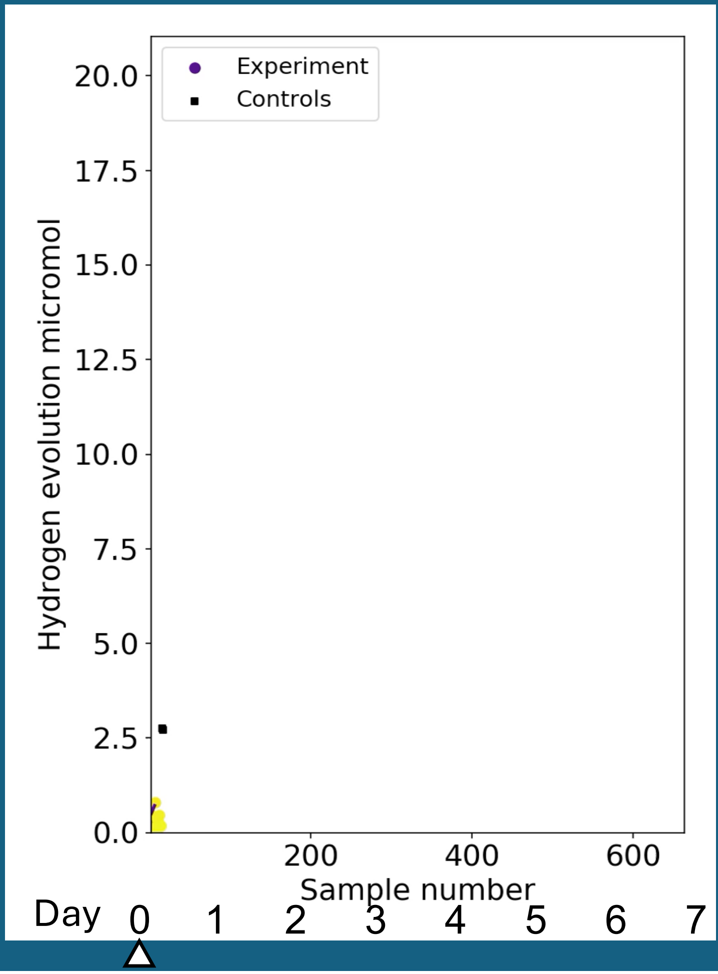
GP-BASED BAYESIAN OPTIMISER

How it works

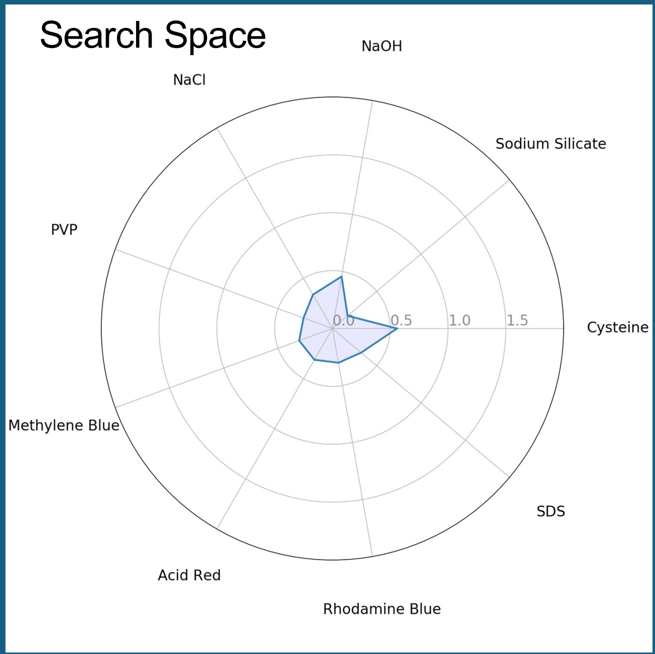
1. Prescribe a prior belief (Gaussian)
2. Calculate the posterior probability
3. Use an acquisition function based on the posterior
4. Sample the acquisition function according to the batch size



BAYESIAN OPTIMISER

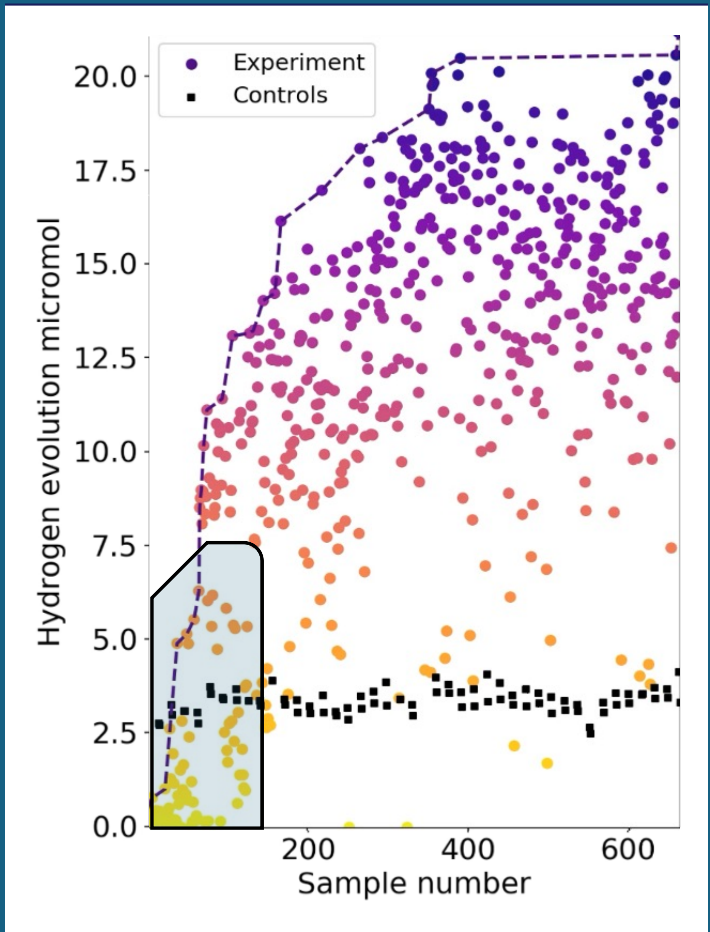


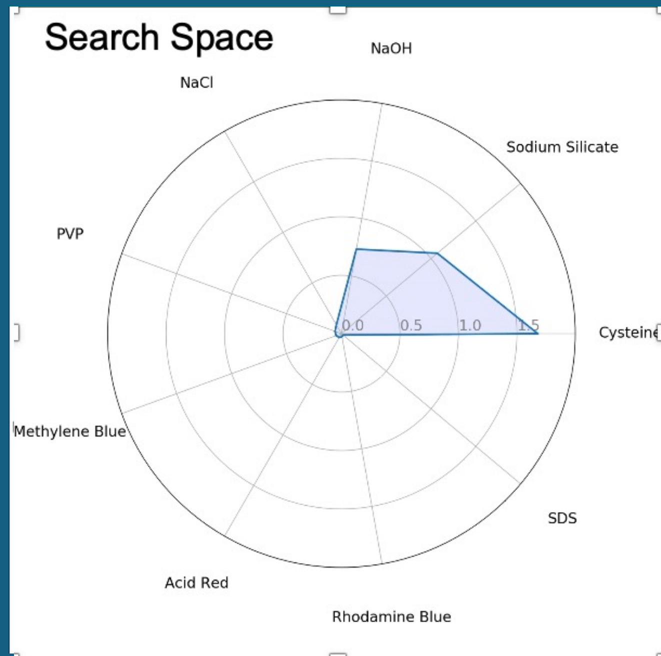
Did a PhD (in experimental load), within a week!



MATERIALS INNOVATION FACTORY

BAYESIAN OPTIMISER

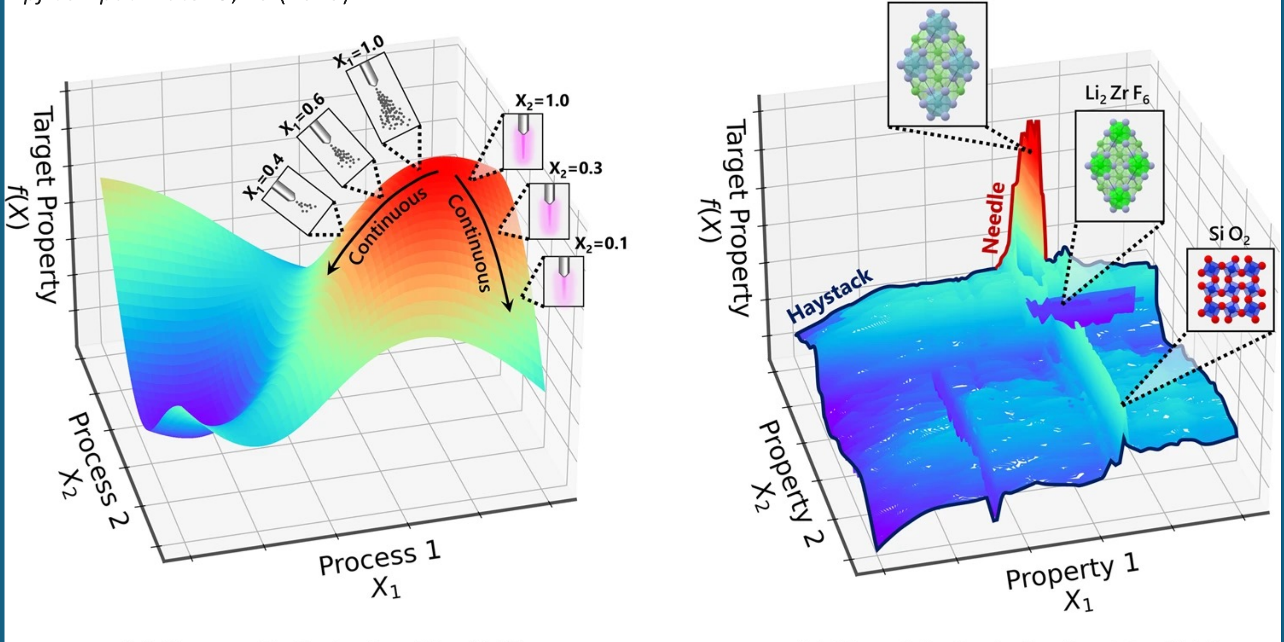




UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

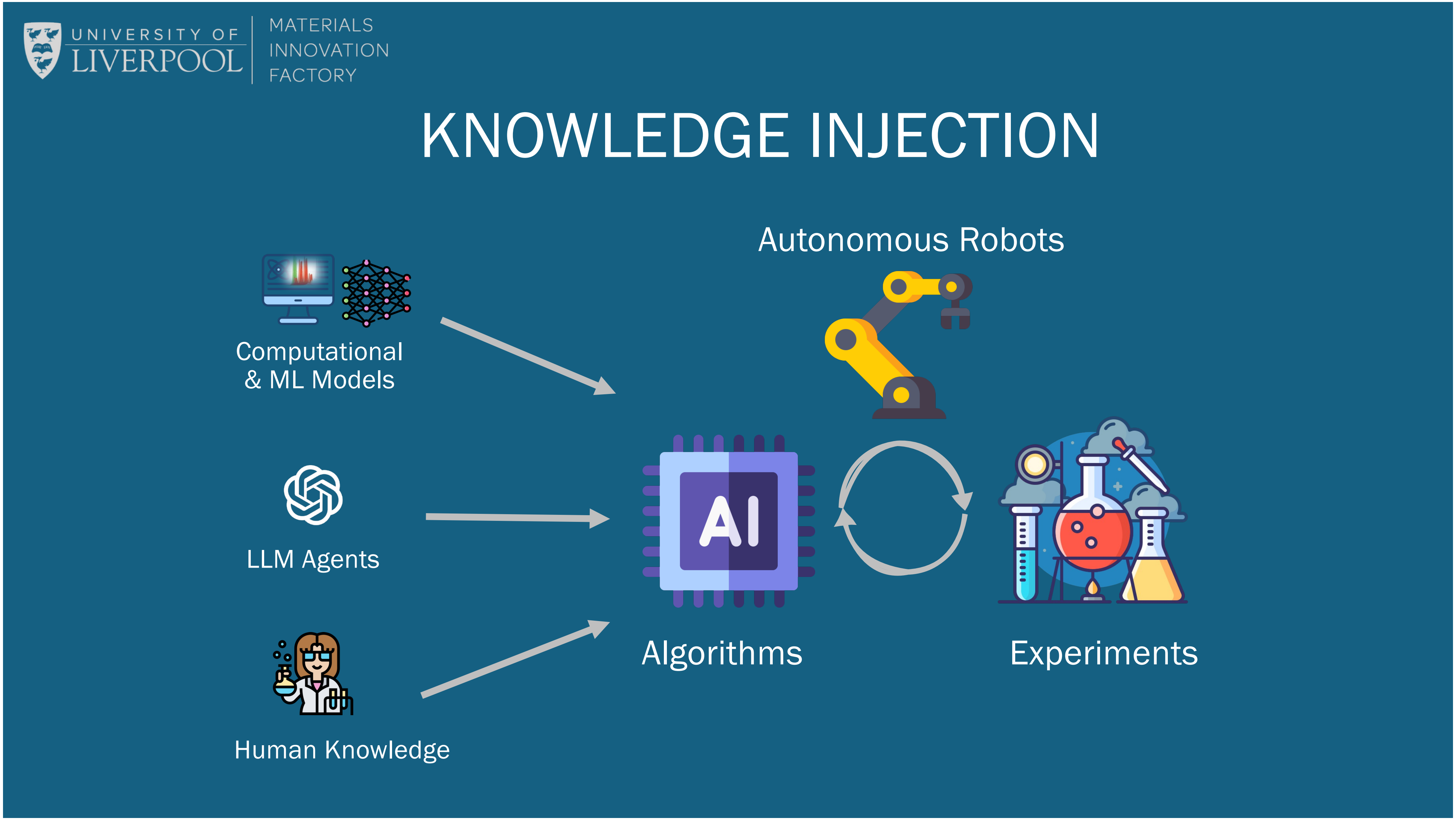
LOOKING FOR NEEDLES IN A HAYSTACK

npj Comput Mater **9**, 79 (2023)



(a) Process Optimization Manifold (b) Materials Optimization Manifold

Smoother Representation of Exploration Space → Focus on Fruitful Regions → Faster Navigation


UNIVERSITY OF LIVERPOOL
MATERIALS INNOVATION FACTORY
KNOWLEDGE INJECTION
Autonomous Robots
Computational & ML Models
LLM Agents
Human Knowledge
AI
Algorithms
Experiments

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

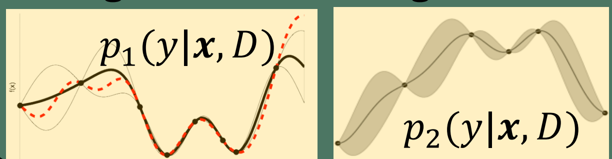
ECAI '23

ML-ENRICHED RESPONSE SURFACES

Sequential Multi-Model-based Optimisation

Dataset

Surrogate Model 1 **Surrogate Model 2**



Acquisition Function

$$\boldsymbol{x}_i = argmax_{x\in X}\ a(\boldsymbol{x}, p_1(y|\boldsymbol{x}, D), p_2(y|\boldsymbol{x}, D))$$

$$\{\boldsymbol{x}_i, y_i\} \cup Dataset$$

dynamic weight

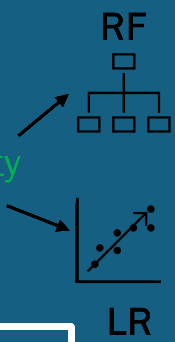
$$\mathbf{x}^\star = \arg\max_{\mathbf{x}\in\mathcal{X}} \alpha(\mathbf{x}, \mathcal{D}) + \gamma\xi(\mathbf{x}, \mathcal{D})$$

BO acquisition

domain knowledge acquisition

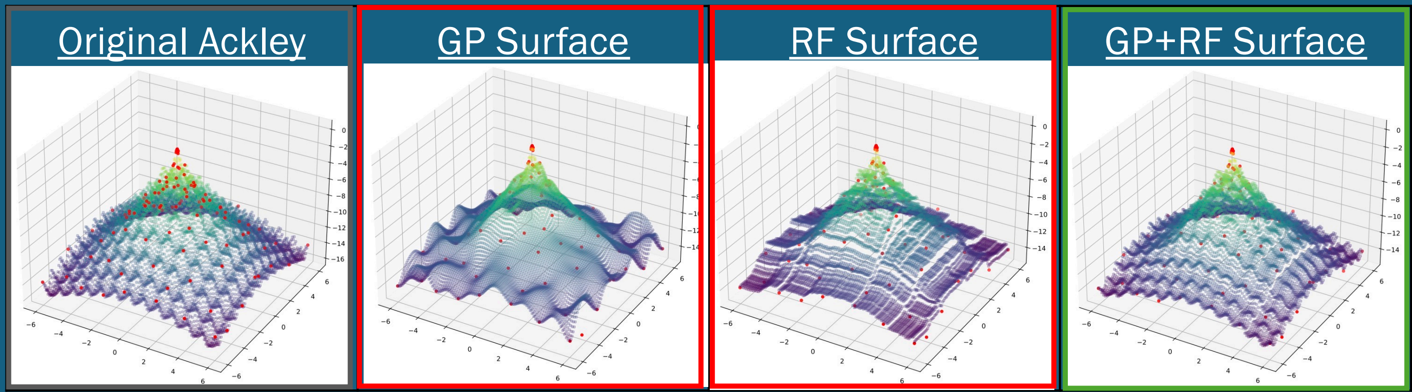
$$\mu(\mathbf{x}) + \kappa\sigma(\mathbf{x})$$

mean and uncertainty of ML model

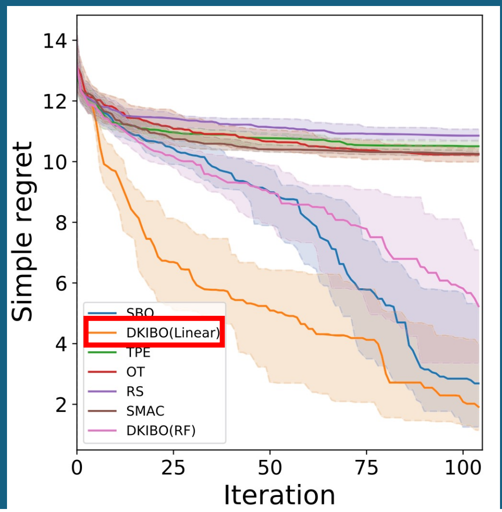


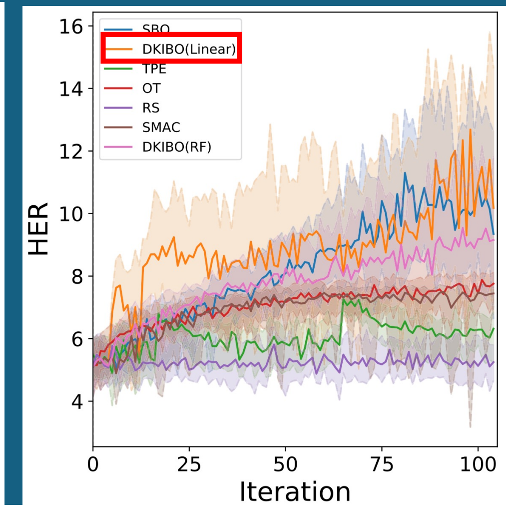
Different ML models capture different underlying structure of the function landscape

ML-ENRICHED RESPONSE SURFACES



H_2 Production Optimisation

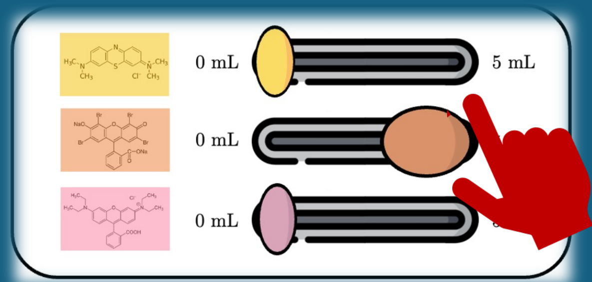




UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

BO DOES NOT UNDERSTAND CHEMISTRY

- Serendipity
- Gut feeling
- Experience
- Creativity
- Wisdom
- Context



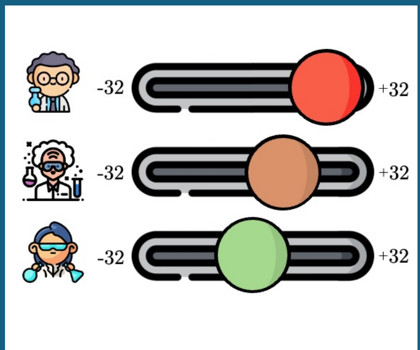
- Dyes will have a positive effect
- AR87 was successful with other related photocatalysts [Wang et al., 2018]

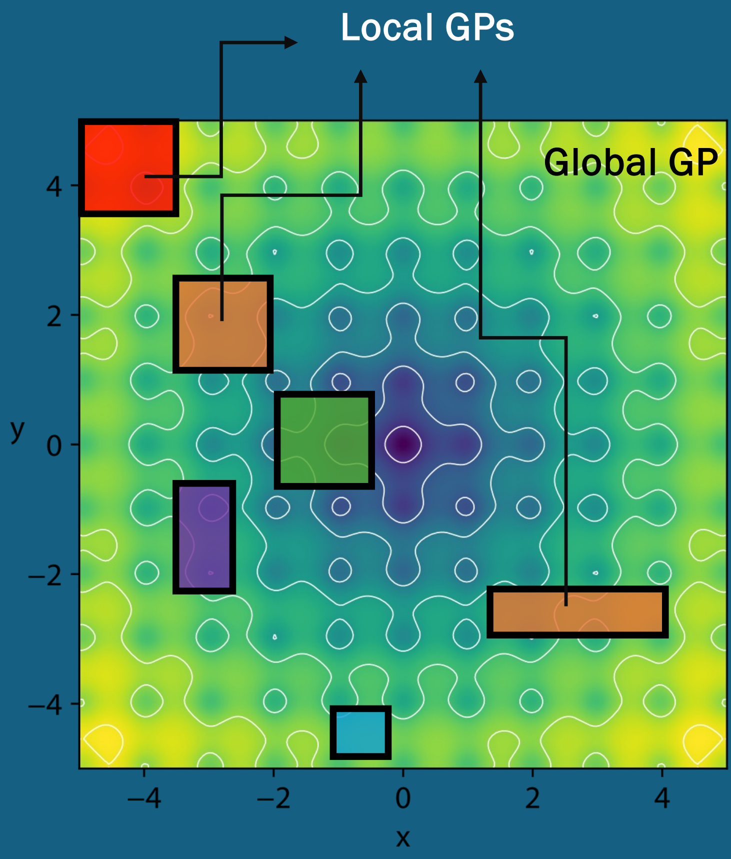
UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

IJCAI ‘24

HYPOTHESIS BAYESIAN OPTIMISATION

‘Tapping’ into promising regions in the parameter space





Hypotheses Subspaces $\mathcal{H}$

▼

‘Soft’ box constraints

f is approximated in each hypothesis subspace by a local GP model:

$$\phi_j \sim \mathcal{N}(\mu_j(x), k_j(x, x'))$$

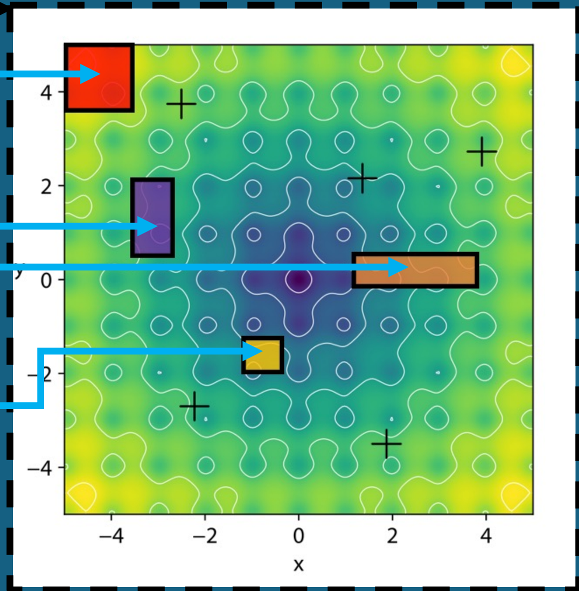
UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

BI-LEVEL OPTIMISATION

Key Concept: Use of expert hypotheses to generate improved seed points.

Upper Level: Global search using vanilla BO to find the maximum of the target function

Lower Level: Local search within hypothesis subspaces to find the promising seeds.



$$x^* = \underset{x\in\mathcal{X}}{argmax}\, \alpha(x, \mathcal{D}\cup\{(s_t, f(s_t))\}_{t=1}^{T}) \qquad \text{(Upper)}$$

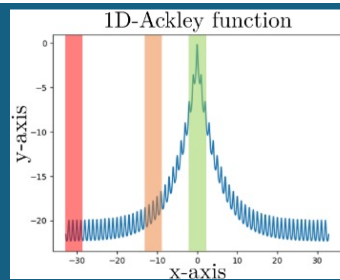
s.t

$$\{s_t\}_{t=1}^{T} \in \underset{s\in\bigcup_{j=1}^{J}\mathcal{H}_j}{argmax}\{\underset{s\in\mathcal{H}_j}{max}\, \alpha_{\phi_j}(s,\{\mathcal{D}\cup\{x^*, f(x^*)\}\}\cap\mathcal{H}_j)\}_{j=1}^{J}. \qquad \text{(Lower)}$$

UNIVERSITY OF LIVERPOOL

MATERIALS INNOVATION FACTORY

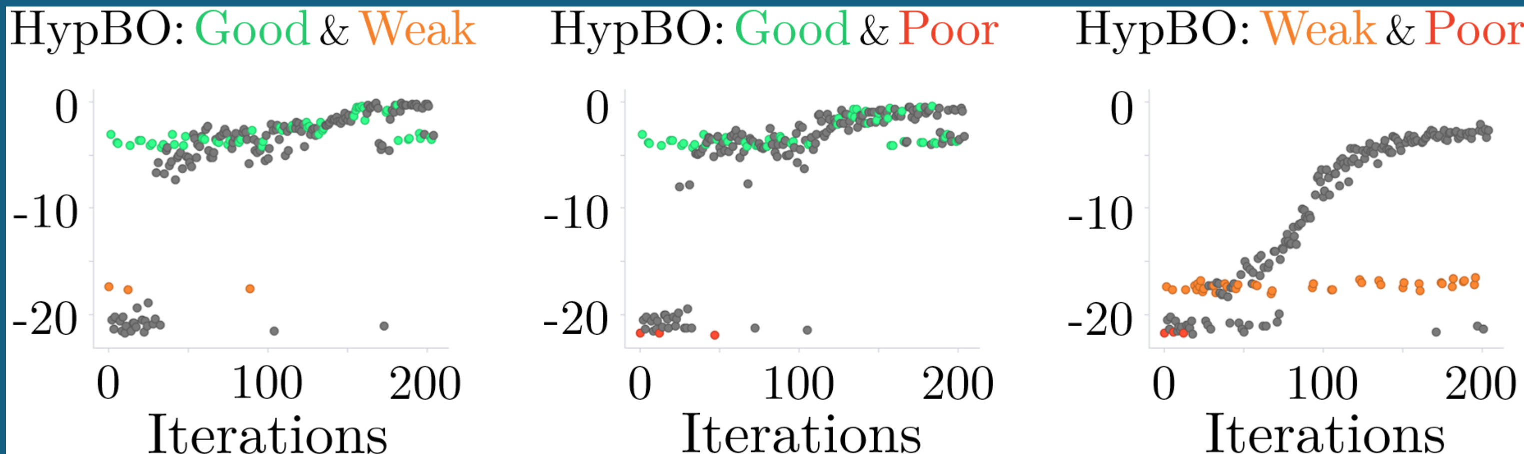
MATERIALS INNOVATION FACTORY



HypBO

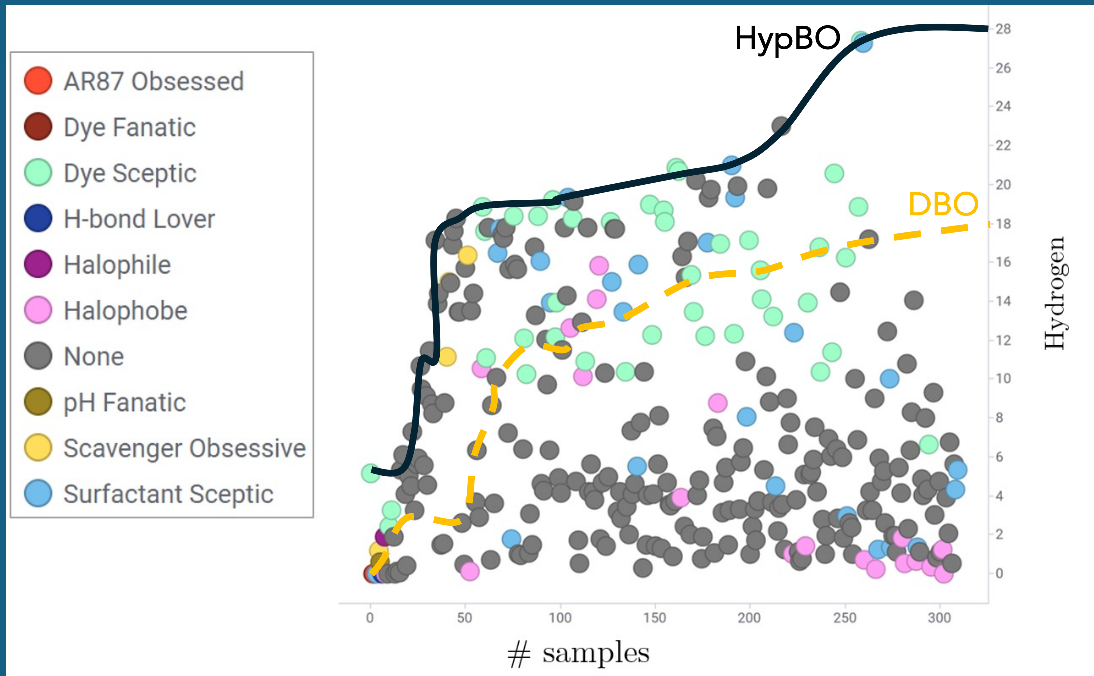
- HypBO on $Ackley_{d9}$ with three different mixtures of hypotheses of various qualities.
- Colored sample points came from the hypotheses (lower level) while the grey ones came from the upper level.

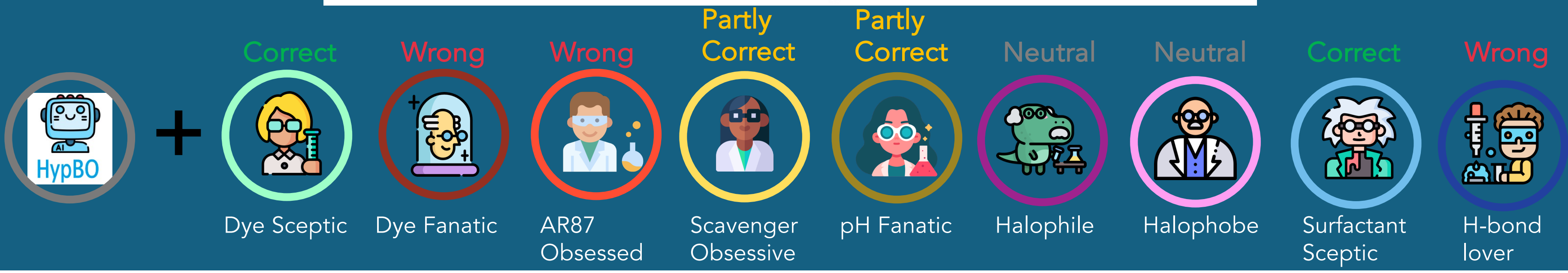
y



UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

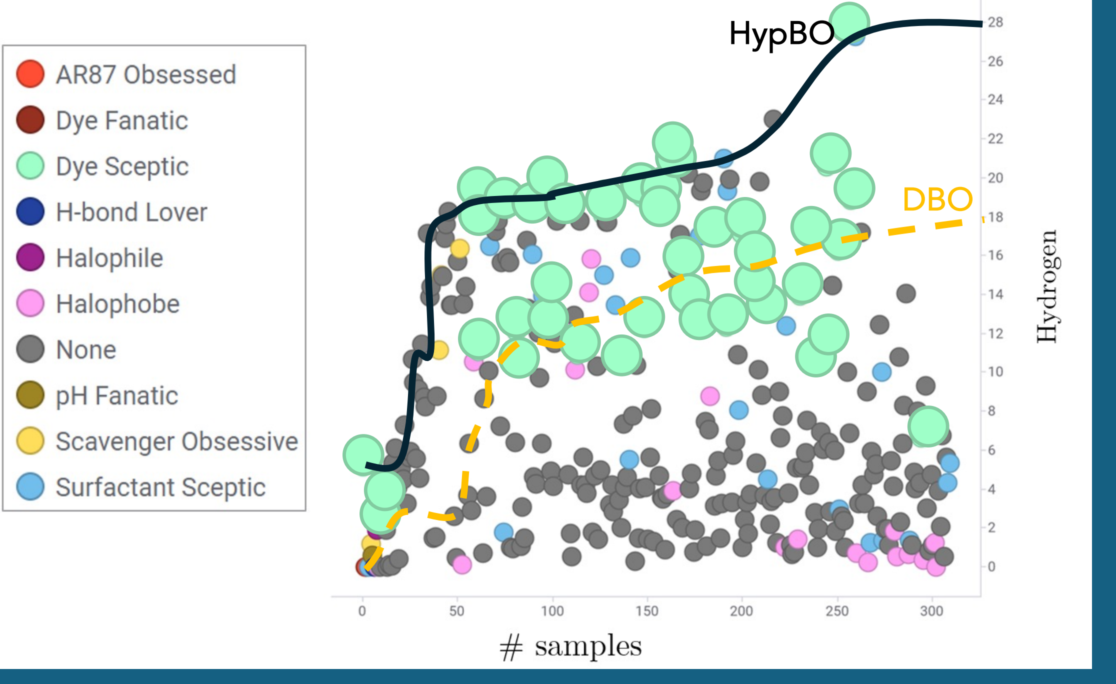
H_2 EXAMPLE





MATERIALS
INNOVATION
FACTORY

H_2 EXAMPLE



+

Correct

Dye Sceptic

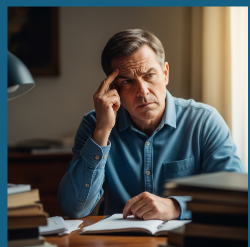
UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

GPT GENERATED HYPOTHESES

Hypotheses

Patterns

Literature



Database

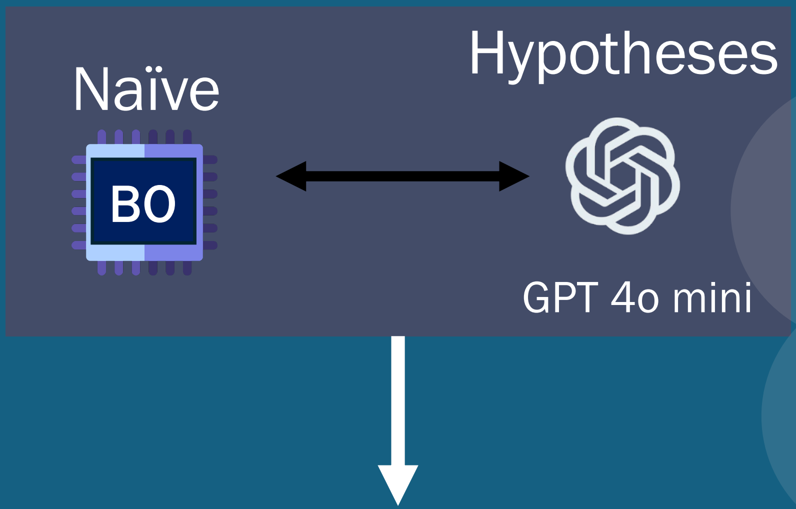
Chat GPT

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

IJCAI ’25

LANGUAGE BO ASSISTANT (BORA)

Hybrid Optimisation

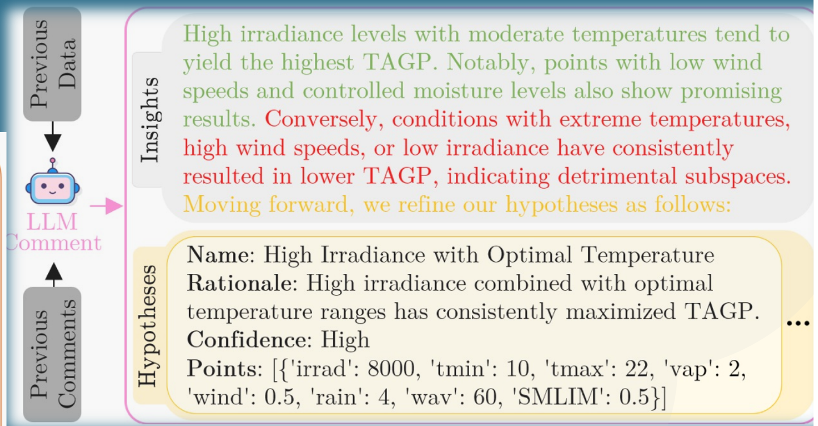


Next experiment

Experiment card


```
{
 "name": "Photocatalytic Hydrogen Production",
 "domain": "chemistry",
 "description": "The experiment is about maximizing the rate of hydrogen production from a photocatalyst mixture of different chemicals in water. The mixture is exposed to ultraviolet and visible light to generate hydrogen, and the amount produced is measured using gas chromatography. The goal is to enhance the hydrogen evolution rate (HER) through modifying the quantity of the chemicals such as adjustments to pH, ionic strength, addition of dyes and surfactants and so on, in the mixture.",
 "constraint": "the sum of all parameters excluding P10-MIX1 should not exceed 5 to fit inside the vial."
 "parameters": [
   {"name": "AcidRed871_0gL", "description": "A dye that could be used as a photosensitizer or to study dye degradation efficiency under photocatalytic conditions", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "L-Cysteine-100gL", "description": "An amino acid that can act as a hole scavenger in photocatalytic processes, potentially enhancing hydrogen production by improving charge separation efficiency", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "MethyleneB_250mgL", "description": "Methylene blue dye", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "NaCl-3M", "description": "Sodium chloride", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "NaOH-1M", "description": "Sodium hydroxide", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "P10-MIX1", "description": "A photocatalyst that is a conjugated polymer and shows good Hydrogen Evolution Rate", "bounds": [1, 5],"type": "discrete", "step": 0.2},
   {"name": "PVP-1wt", "description": "Polyvinylpyrrolidone", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "RhodamineB1_0gL", "description": "Rhodamine dye", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "SDS-1wt", "description": "Sodium dodecyl sulphate", "bounds": [0, 5], "type": "discrete", "step": 0.25},
   {"name": "Sodiumsilicate-1wt", "description": "Sodium silicate", "bounds": [0, 5], "type": "discrete", "step": 0.25}
 ],
 "target": {"name": "Hydrogen Evolution Rate (μmol/h)", "description": "The rate of hydrogen production in the photocatalytic system"}
}
```


LLM Hypothesis



UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

LANGUAGE BO ASSISTANT (BORA)

Actions

1. Naïve BO
2. LLM Comments and Suggests new points
3. LLM Comments and Selects points from BO

Adaptive Heuristic Policy

If $\sigma_{t,\text{mean}}^{\text{GP}} < \sigma_{t,\text{lower}}$ or 'no plateau' $\rightarrow$ **action** a_1,

Else if $\sigma_{t,\text{mean}}^{\text{GP}} > \sigma_{t,\text{upper}} \rightarrow$ **action** a_2,

Else $\rightarrow$ **action** a_3.

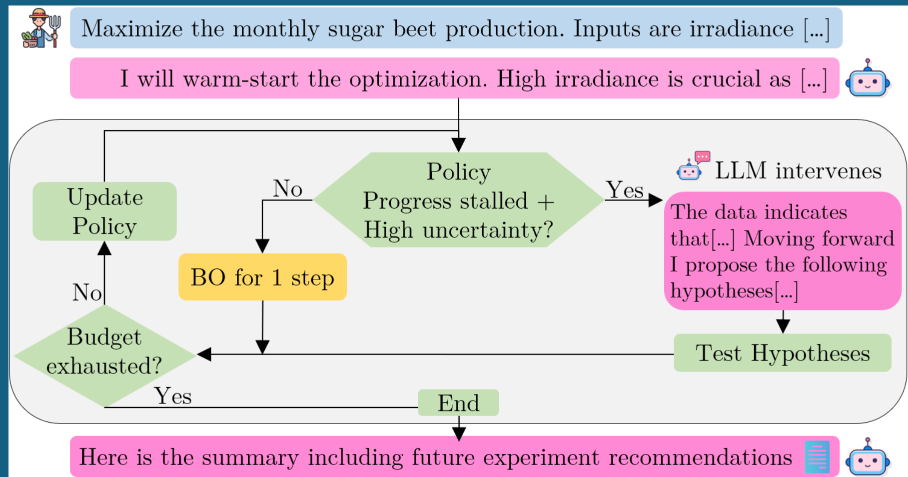
Subject to LLM Trust

$$\sigma_{t,\text{mean}}^{\text{GP}} = \frac{1}{q} \sum_{i=1}^{q} \sigma_t(x_{\text{mon}}^{(i)})$$

$$\sigma_{t,\text{upper}} = 0.5 \times \sigma_{t,\max}^{\text{GP}} \text{ and } \sigma_{t,\text{lower}} = 0.3 \times \sigma_{t,\max}^{\text{GP}}$$

$$\sigma_{t,\max}^{\text{GP}} = \max\left(\sigma_{t-1,\max}^{\text{GP}}, \max_{1 \le i \le q} \sigma_t(x_{\text{mon}}^{(i)})\right)$$

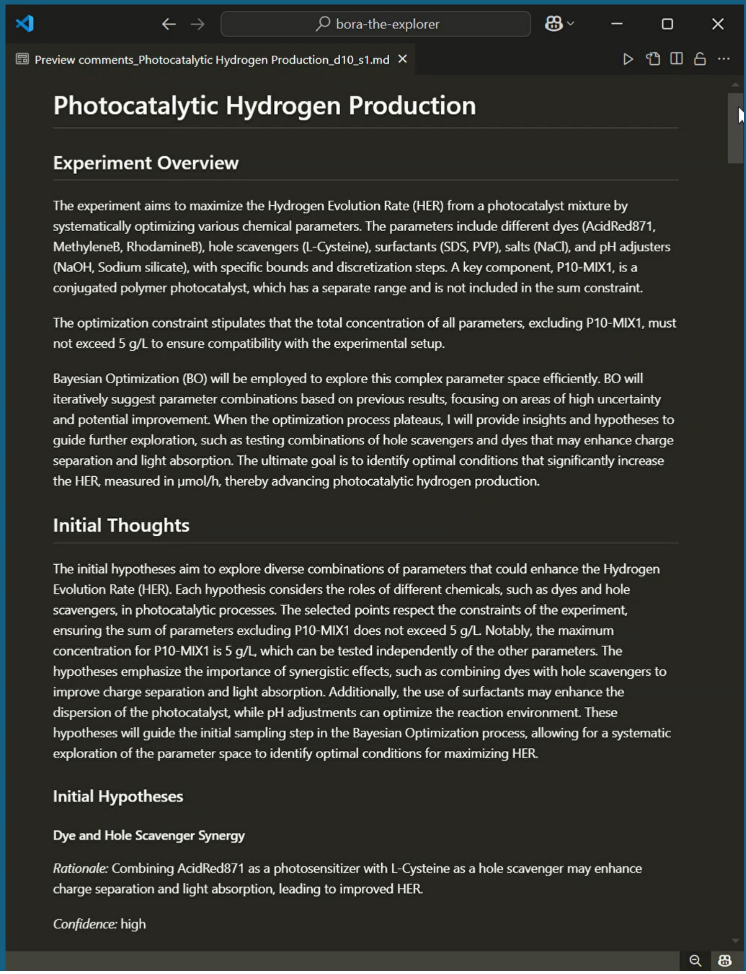
Workflow



UNIVERSITY OF LIVERPOOL

MATERIALS INNOVATION FACTORY

LANGUAGE BO ASSISTANT (BORA)


bora-the-explorer

Preview comments_Photocatalytic Hydrogen Production_d10_s1.md

Photocatalytic Hydrogen Production

Experiment Overview

The experiment aims to maximize the Hydrogen Evolution Rate (HER) from a photocatalyst mixture by systematically optimizing various chemical parameters. The parameters include different dyes (AcidRed871, MethyleneB, RhodamineB), hole scavengers (L-Cysteine), surfactants (SDS, PVP), salts (NaCl), and pH adjusters (NaOH, Sodium silicate), with specific bounds and discretization steps. A key component, P10-MIX1, is a conjugated polymer photocatalyst, which has a separate range and is not included in the sum constraint.

The optimization constraint stipulates that the total concentration of all parameters, excluding P10-MIX1, must not exceed 5 g/L to ensure compatibility with the experimental setup.

Bayesian Optimization (BO) will be employed to explore this complex parameter space efficiently. BO will iteratively suggest parameter combinations based on previous results, focusing on areas of high uncertainty and potential improvement. When the optimization process plateaus, I will provide insights and hypotheses to guide further exploration, such as testing combinations of hole scavengers and dyes that may enhance charge separation and light absorption. The ultimate goal is to identify optimal conditions that significantly increase the HER, measured in μmol/h, thereby advancing photocatalytic hydrogen production.

Initial Thoughts

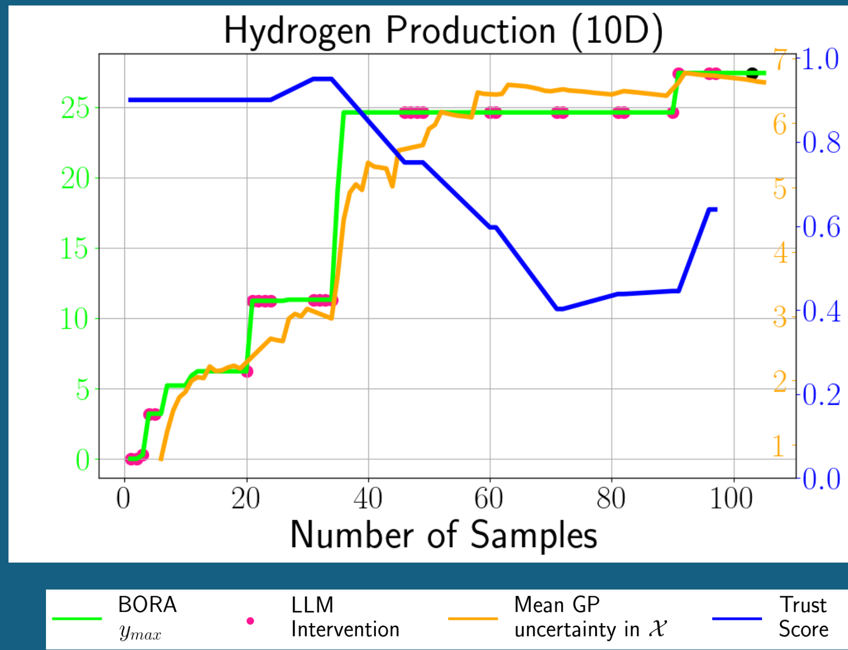
The initial hypotheses aim to explore diverse combinations of parameters that could enhance the Hydrogen Evolution Rate (HER). Each hypothesis considers the roles of different chemicals, such as dyes and hole scavengers, in photocatalytic processes. The selected points respect the constraints of the experiment, ensuring the sum of parameters excluding P10-MIX1 does not exceed 5 g/L. Notably, the maximum concentration for P10-MIX1 is 5 g/L, which can be tested independently of the other parameters. The hypotheses emphasize the importance of synergistic effects, such as combining dyes with hole scavengers to improve charge separation and light absorption. Additionally, the use of surfactants may enhance the dispersion of the photocatalyst, while pH adjustments can optimize the reaction environment. These hypotheses will guide the initial sampling step in the Bayesian Optimization process, allowing for a systematic exploration of the parameter space to identify optimal conditions for maximizing HER.

Initial Hypotheses

Dye and Hole Scavenger Synergy

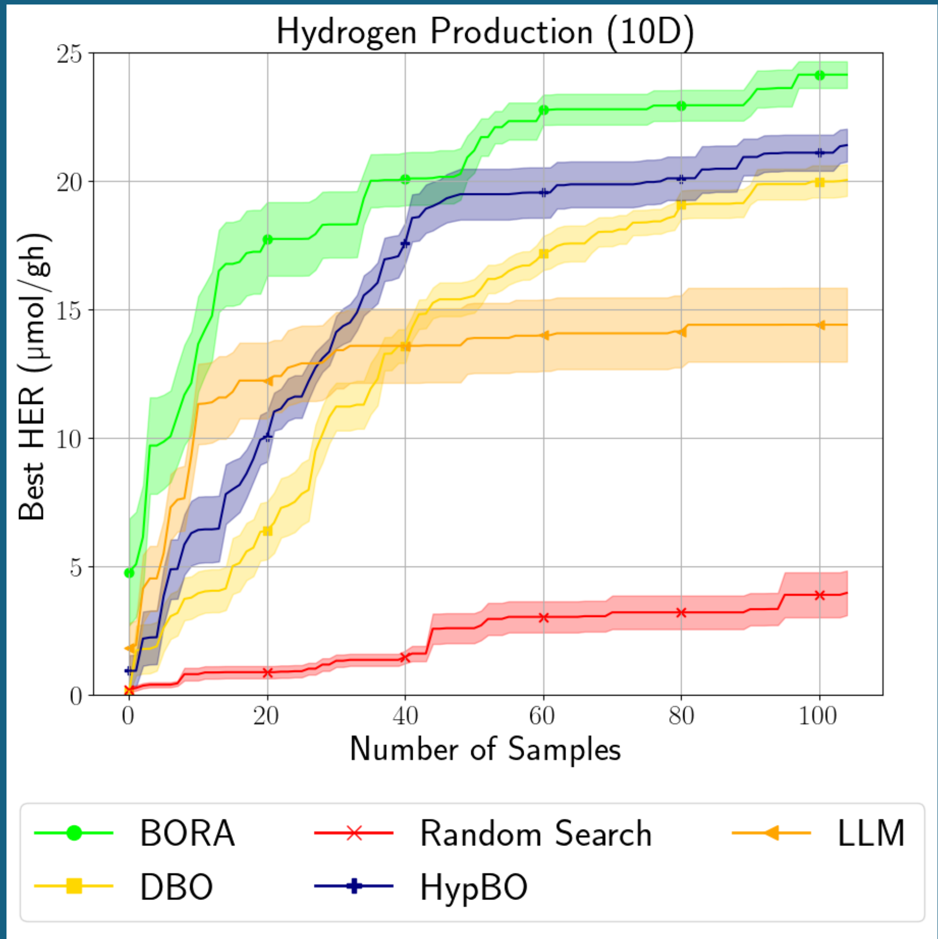
Rationale: Combining AcidRed871 as a photosensitizer with L-Cysteine as a hole scavenger may enhance charge separation and light absorption, leading to improved HER.

Confidence: high



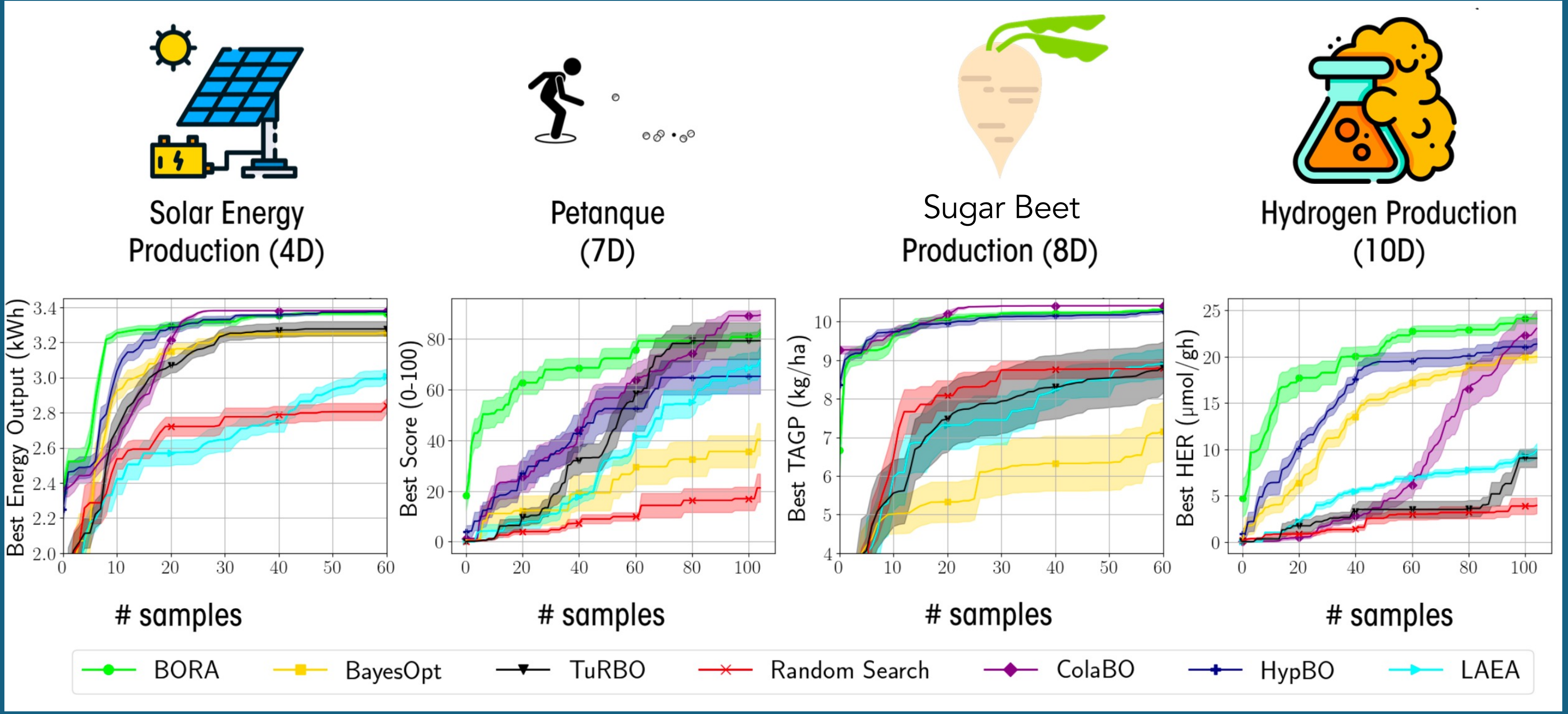
UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

H_2 PRODUCTION



UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

BORA AS A SCIENCE OPTIMISER



UNIVERSITY OF LIVERPOOL

MATERIALS INNOVATION FACTORY

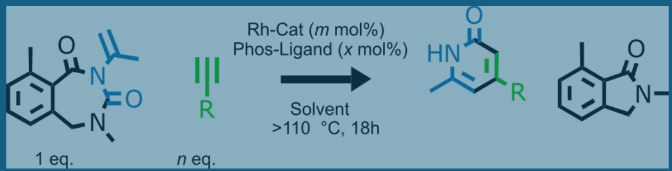
HOW TO AUTOMATE CHEMISTRY?

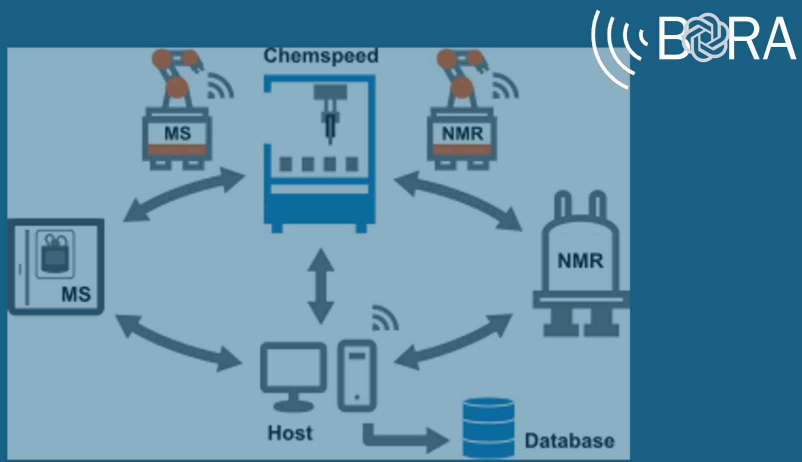
WHAT CHEMISTRY TO AUTOMATE?

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

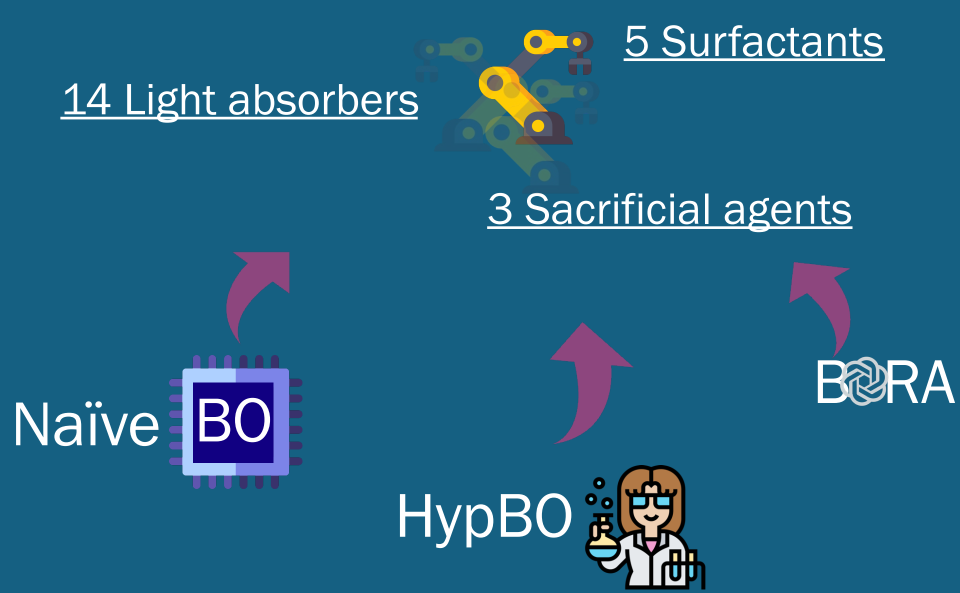
BORA AS A GENERAL TOOL

Reaction Generalisation



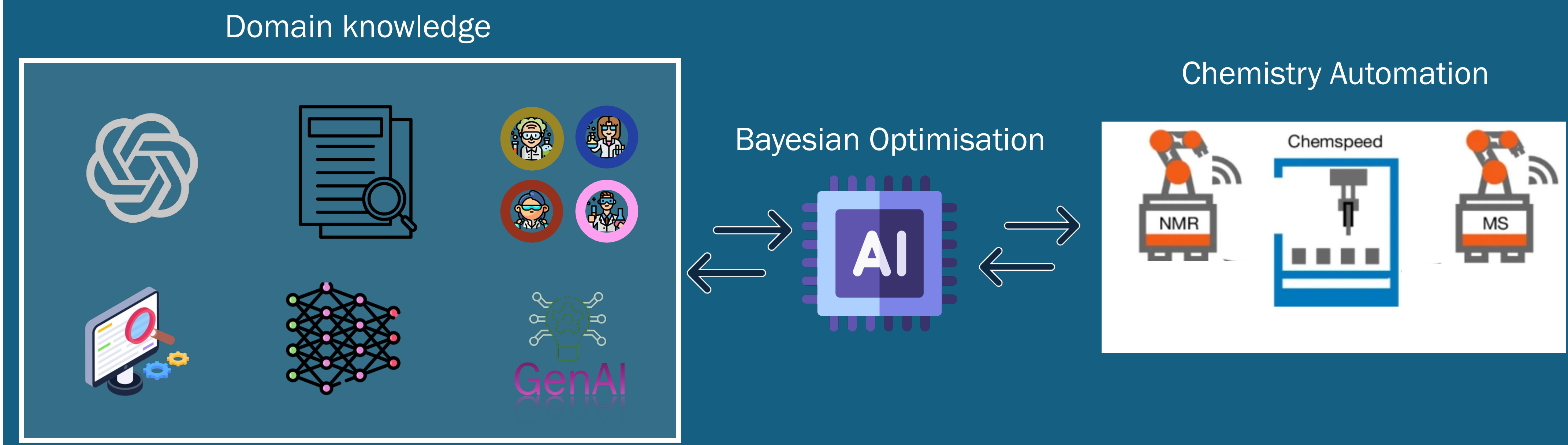


Hydrogen Production 22-D



UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

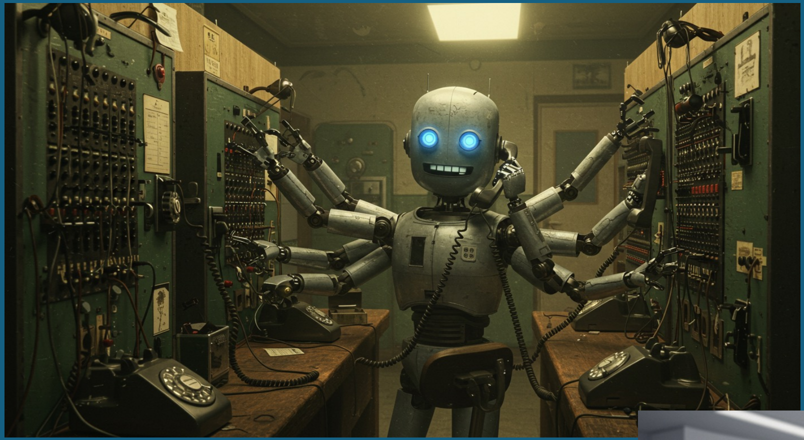
HYBRID INTELLIGENCE



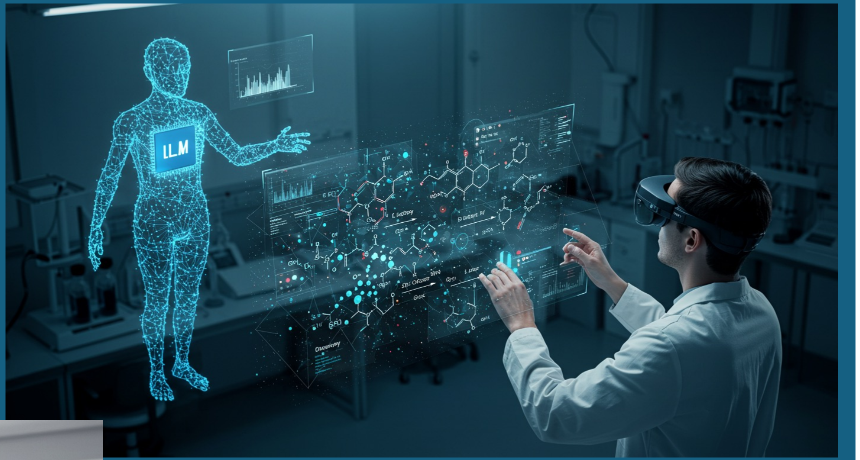
UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

OPEN CHALLENGES

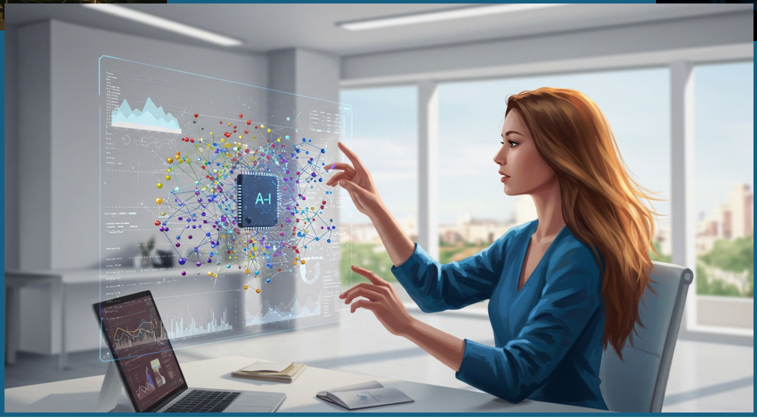
Multi-fidelity BO



LLM Chemistry Conceptualisation



HITL Data Visualisation



UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

ACKNOWLEDGEMENTS

Prof. Andy Cooper

Cooper group

LEVERHULME TRUST

UNIVERSITY OF LIVERPOOL | MATERIALS INNOVATION FACTORY

THANK YOU FOR LISTENING ☺

Get in touch:

Scholar

Linkedin

Xenofon.Evangelopoulos@liverpool.ac.uk