Extracting and processing multimodal data from literature to guide robotic experiments

Intro to AI-driven Science Seminar

Katerina Vriza, Henry Chan, Jie Xu, Maria Chan

02.20.2024

Image generated by DALL-E3

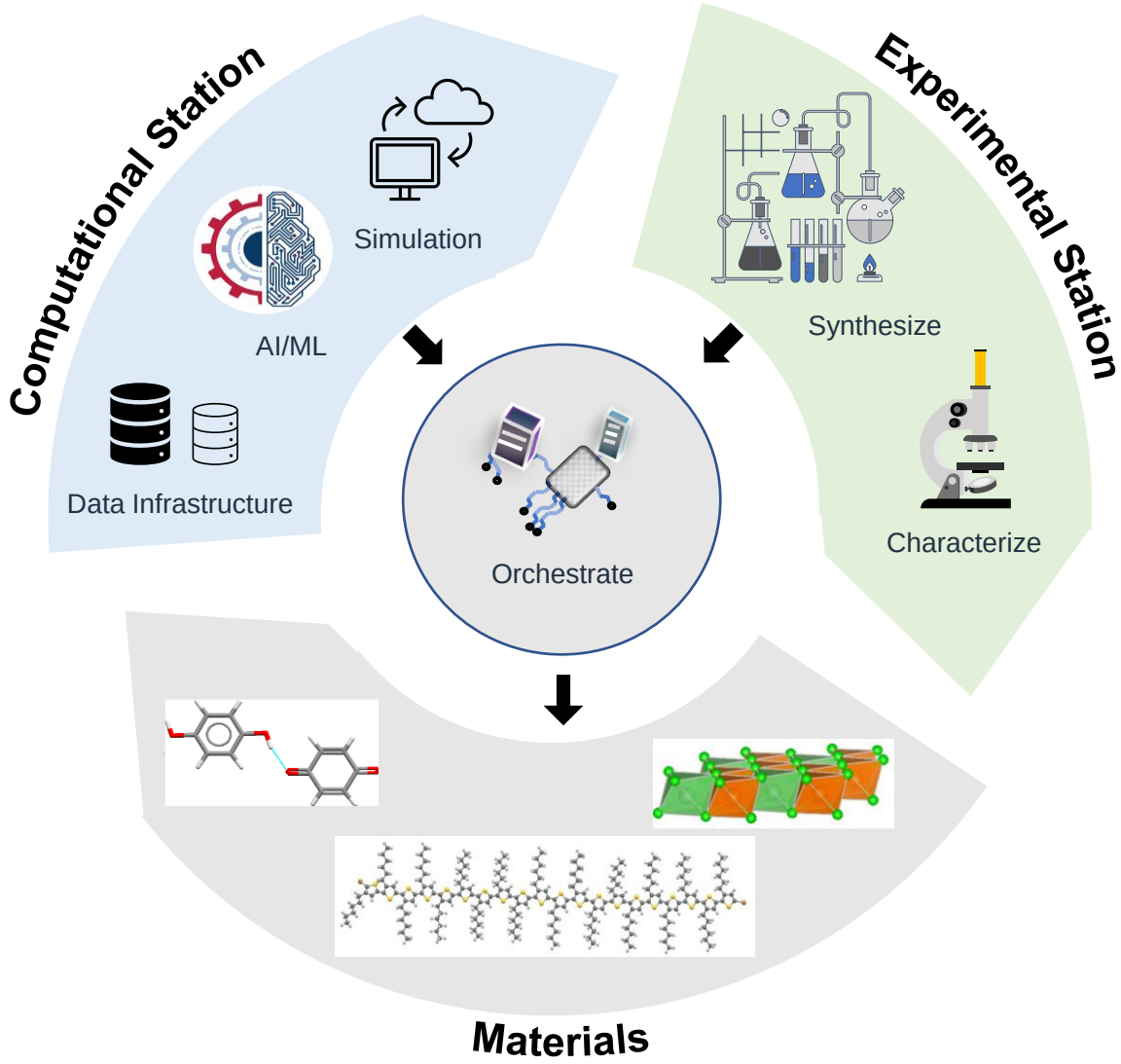
Argonne NATIONAL LABORATORY

Motivation: Accelerate materials discovery

We want new materials, more sustainable, more efficient and we want them fast

Problems:

- Limited fraction of data in machine-readable format
- Data are manually curated.



Problems:

- Trial and error approaches
- Lack of reproducibility

Possible solutions:

- Automated data extraction tools
- Natural language models for information extraction

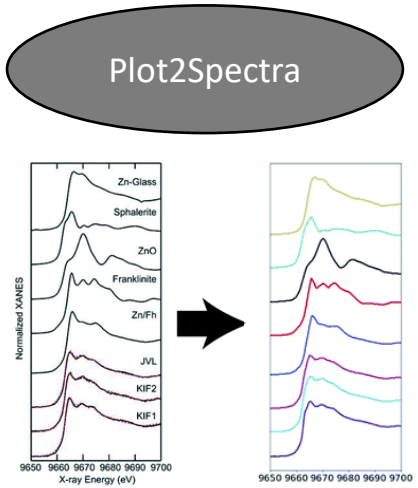
Possible solutions:

- Data driven experimentation
- High-throughput experiments
- Autonomous laboratories

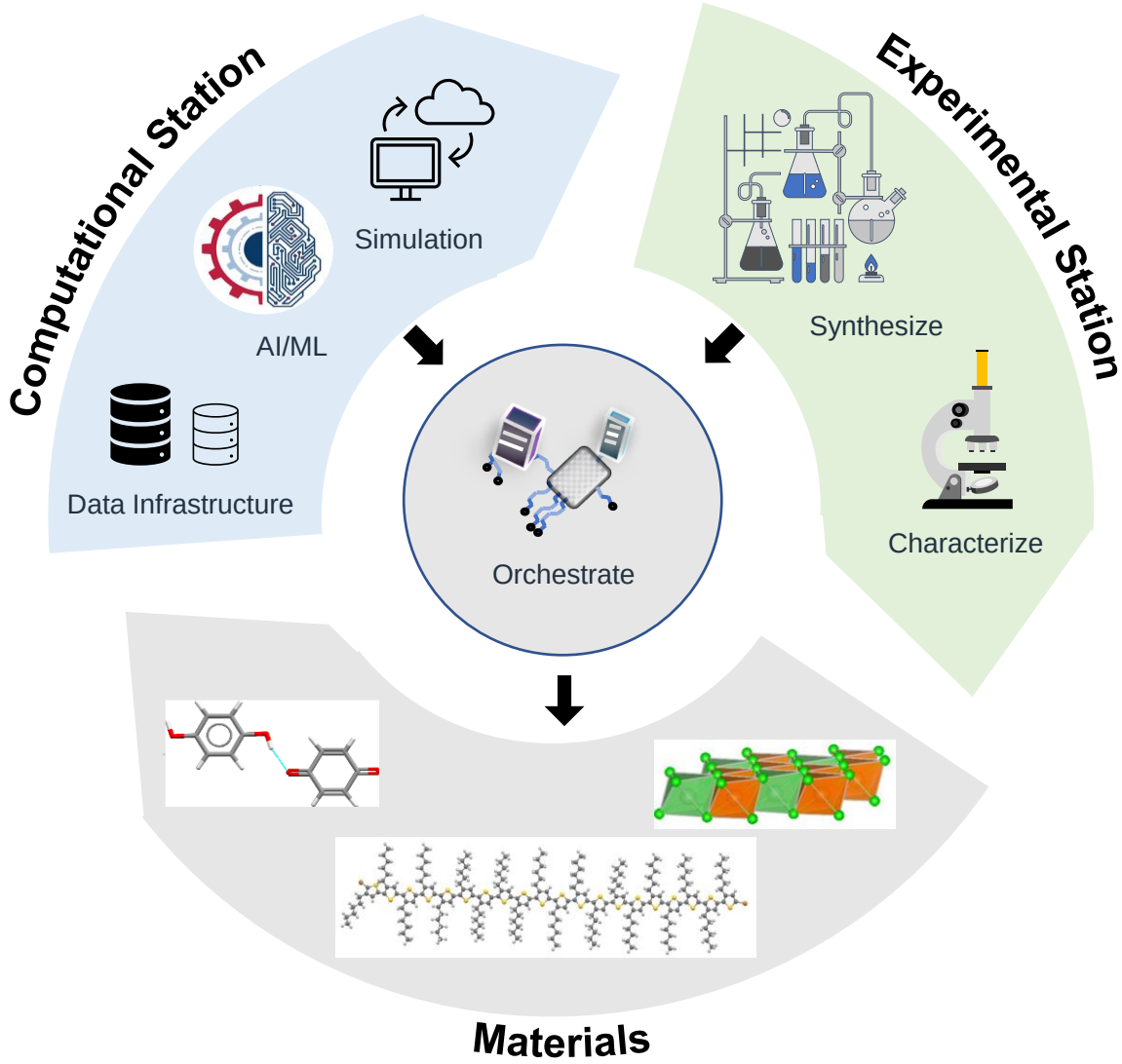
Argonne NATIONAL LABORATORY

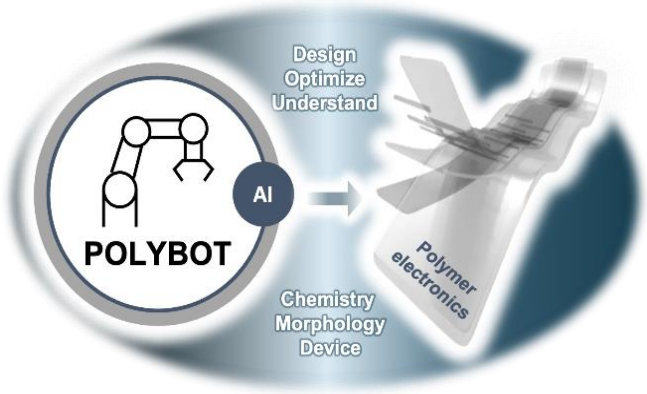
Motivation: Accelerate materials discovery

Schwenker *et al*,
https://arxiv.org/abs/2103.10631



Jiang *et al*,
Digital Discovery, 2022, 1,719

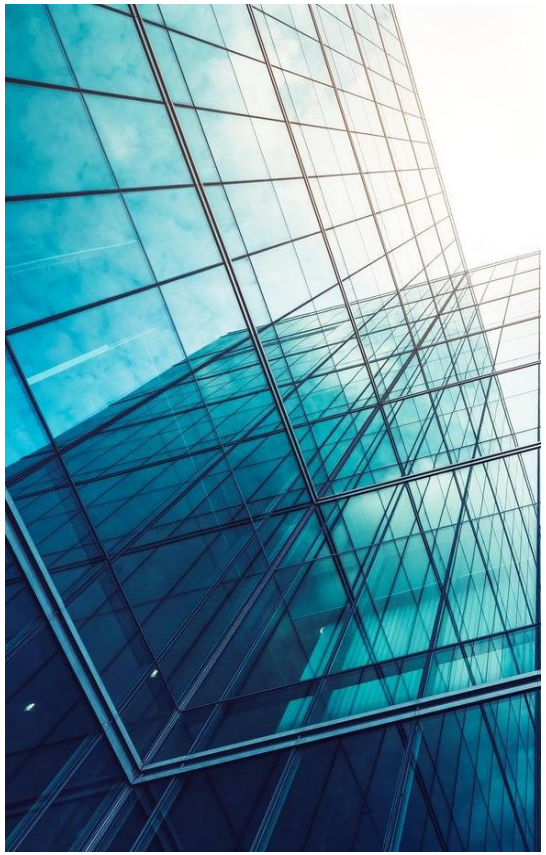


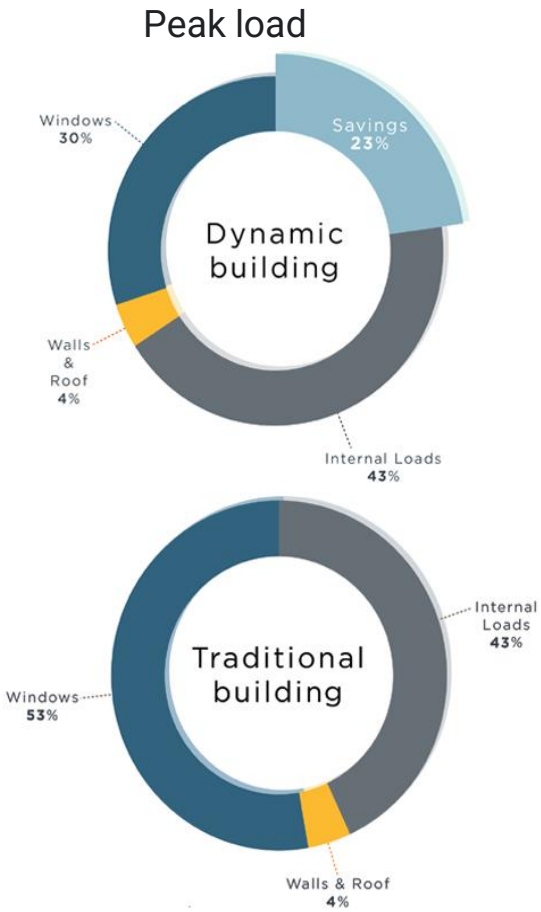


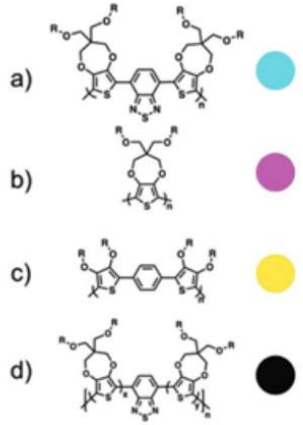
Vriza A, Chan H , Xu J,
accepted in Chemistry of Materials

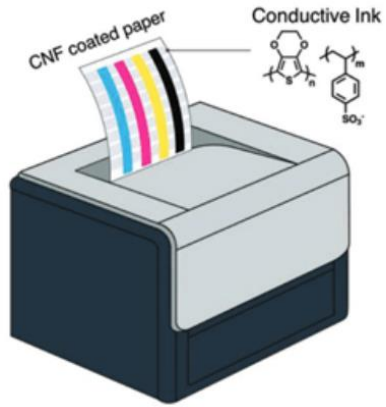
Argonne NATIONAL LABORATORY

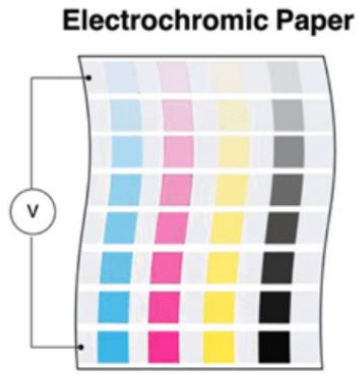
Autonomous synthesis of polymer electronics







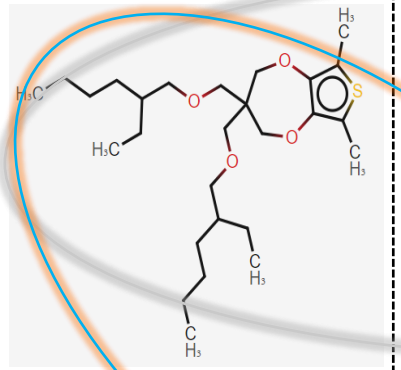




- Smart windows automatically change transparency when a voltage is applied.
- We need chemical recipes for the thin-film material in order to enhance the reliability of the window.

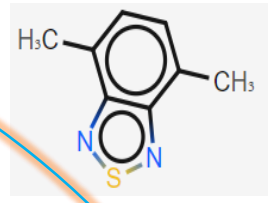
Autonomous synthesis of electrochromic polymers

Monomer A

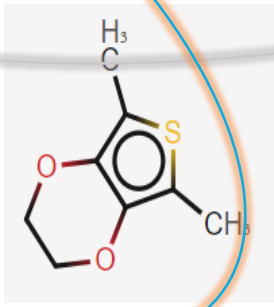


ProDOT

Monomer B

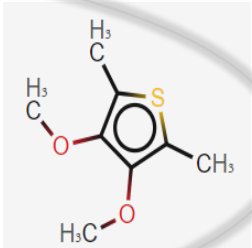


BTD

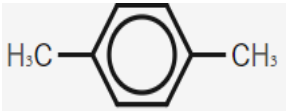


EDOT

Monomer C

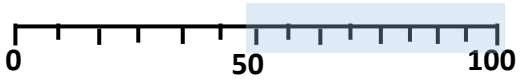


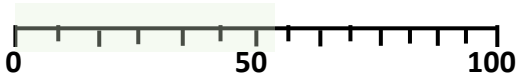
DMOT

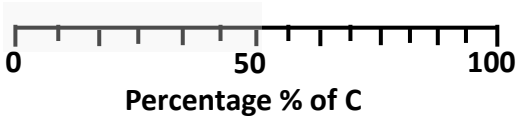


Ben

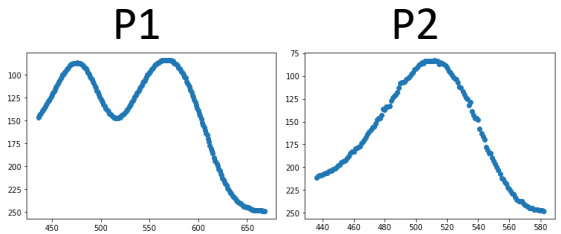
Select combinations

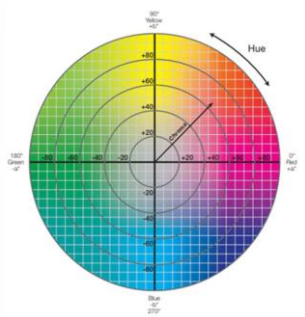






Select stoichiometries

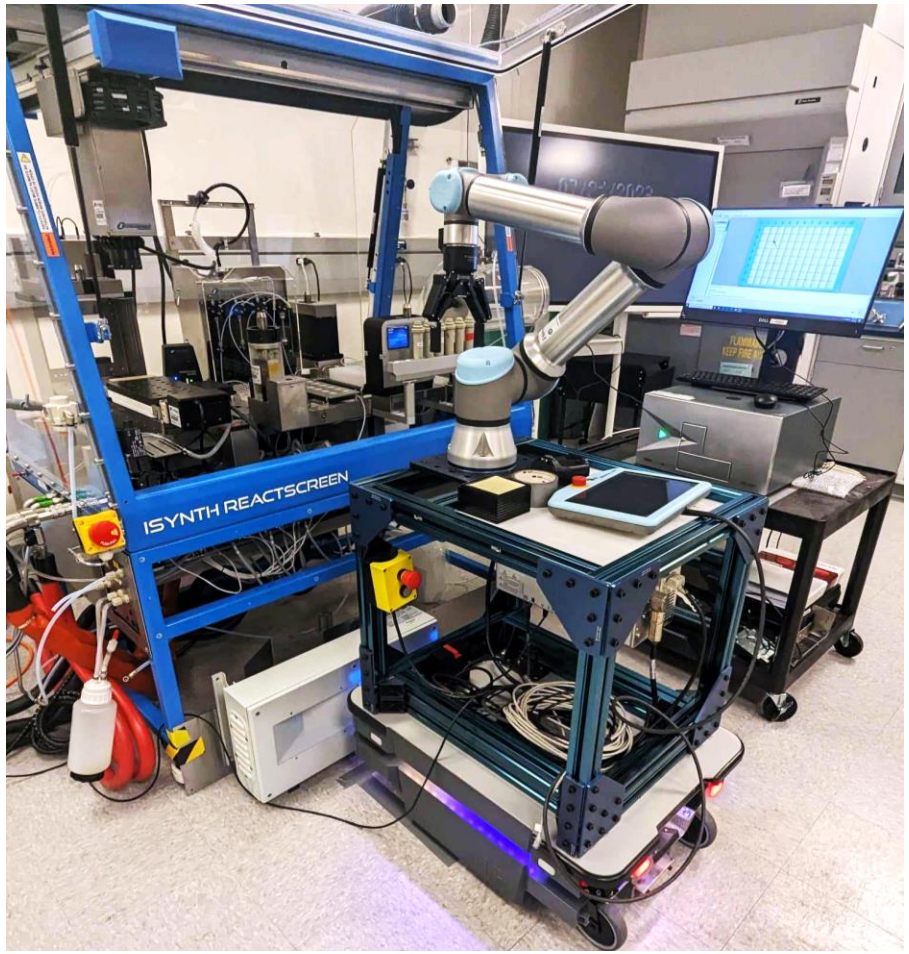




Characterize

Argonne NATIONAL LABORATORY

Autonomous synthesis of polymer electronics



1. Extract and analyse data from existing literature

2. Train ML models

3. Establish a robot network to run autonomous experiments

Autonomous synthesis of electrochromic polymers

1

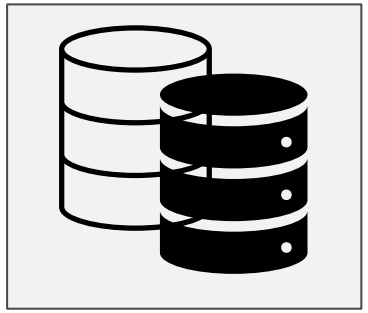
Start from a related query in a selected journal

{ **journal**: nature,

query: electrochromic polymers }

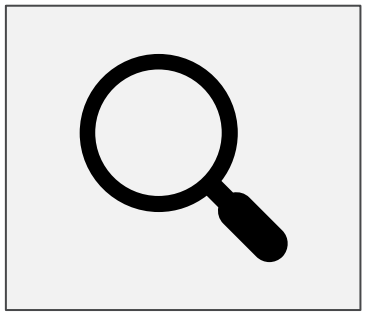
2

Create database with all the images and related keywords



3

Look up
- absorption spectra figures
- structure images
- text describing the ECPs



4

Dataset
Table with polymers and color coordinates

Polymer reference	L	a	b
EDOT			
PRODOT			
DMP			

Autonomous synthesis of electrochromic polymers

Journal Scrapping (**EXSCLAIM!**)

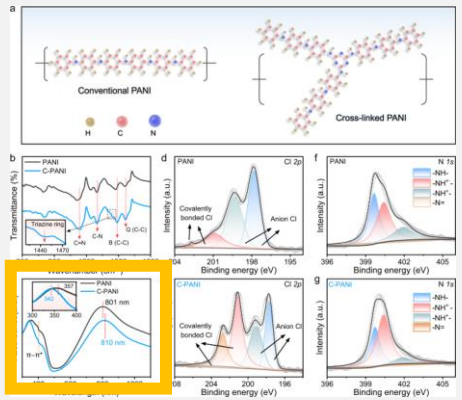


Fig 1. a Schematic synthesis of conventional PANI and C-PANI. b FTIR of conventional PANI and C-PANI. **c UV–Vis spectra of the conventional PANI and C-PANI samples. d…..**

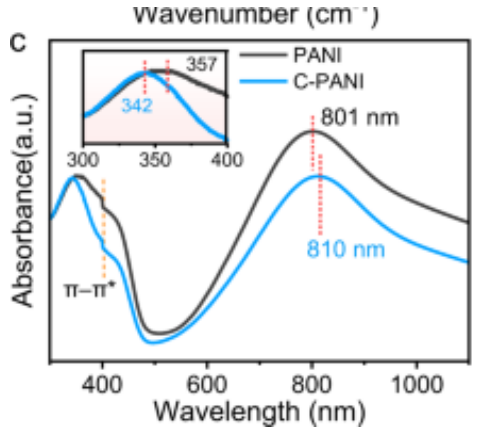
Caption Separator and context extraction (**GPT3.5**)

Caption c: UV–Vis spectra of the conventional PANI and C-PANI samples.

Keywords:

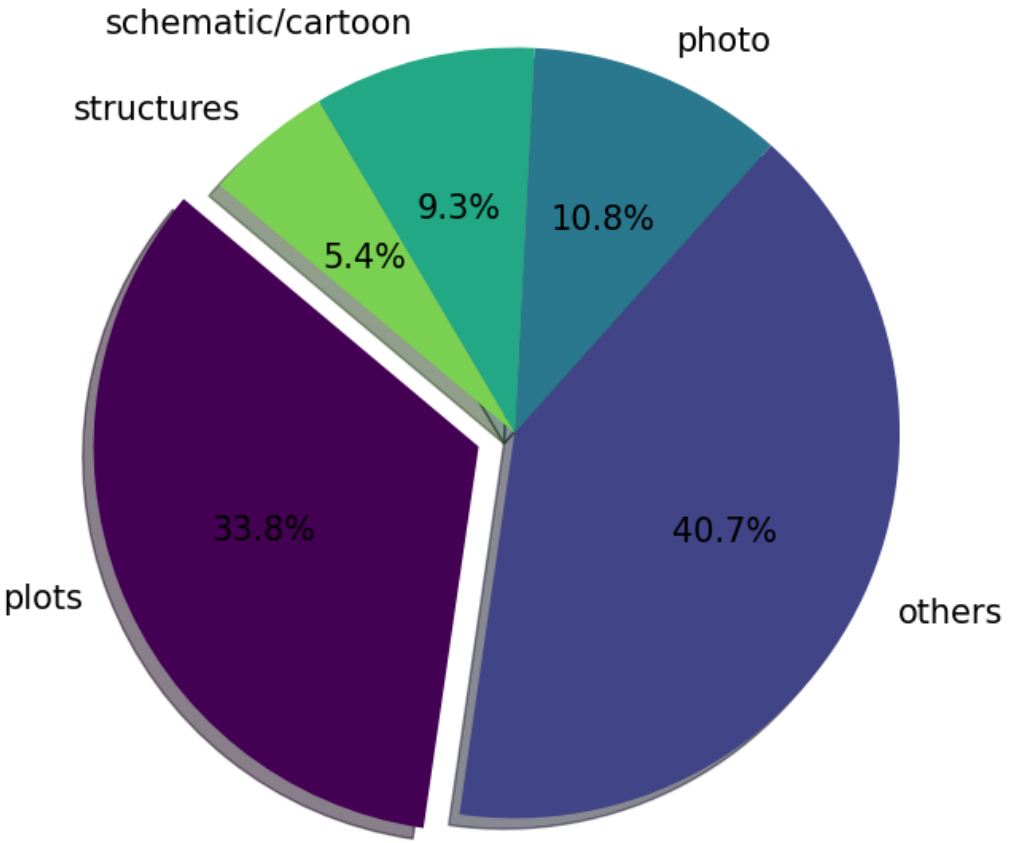
```
"UV-Vis spectra",
"conventional PANI",
"C-PANI samples"
```

Figure separator (**EXSCLAIM!**)

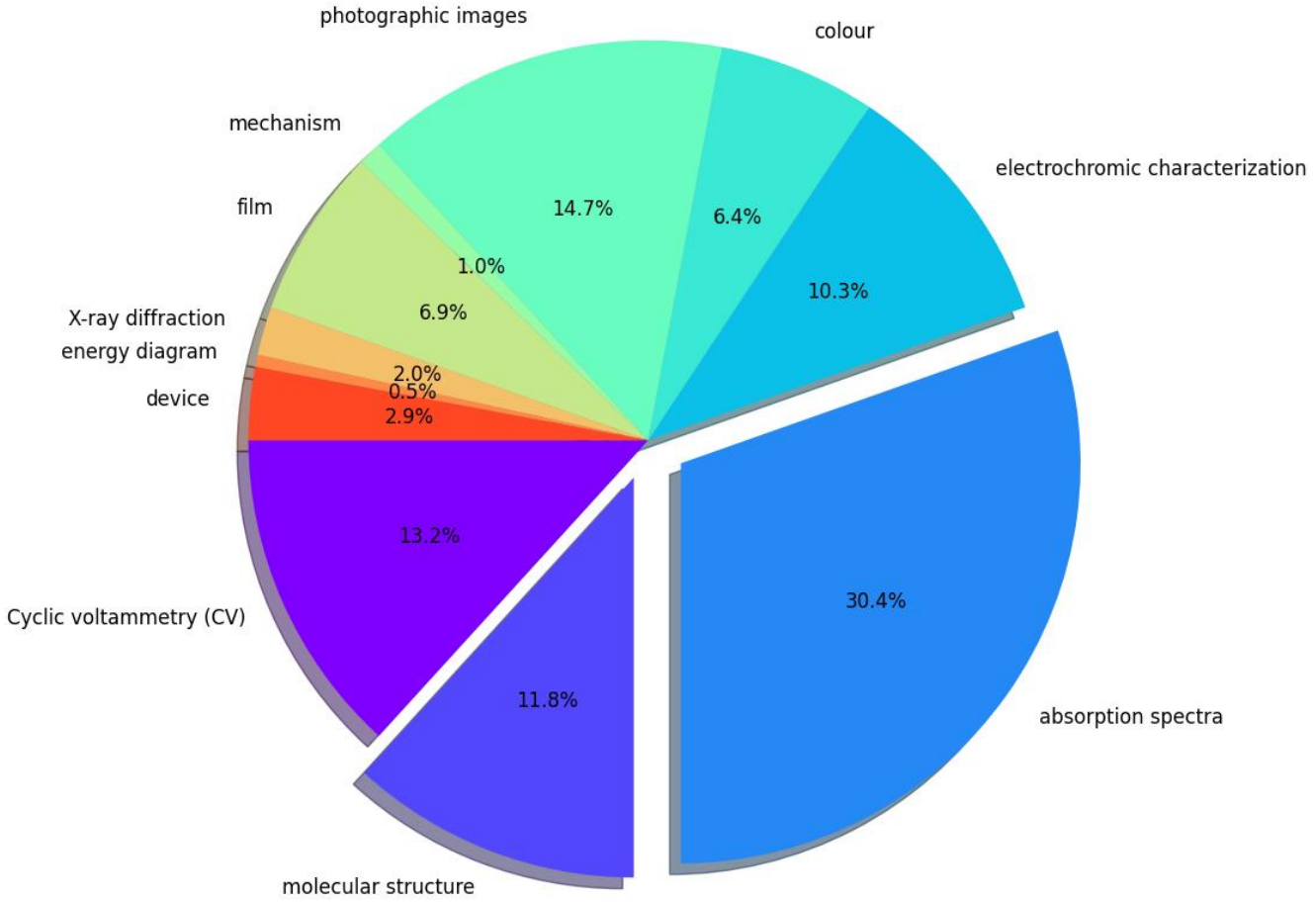


Autonomous synthesis of electrochromic polymers

Image categories extracted for electrochromic polymers discovery



General categories

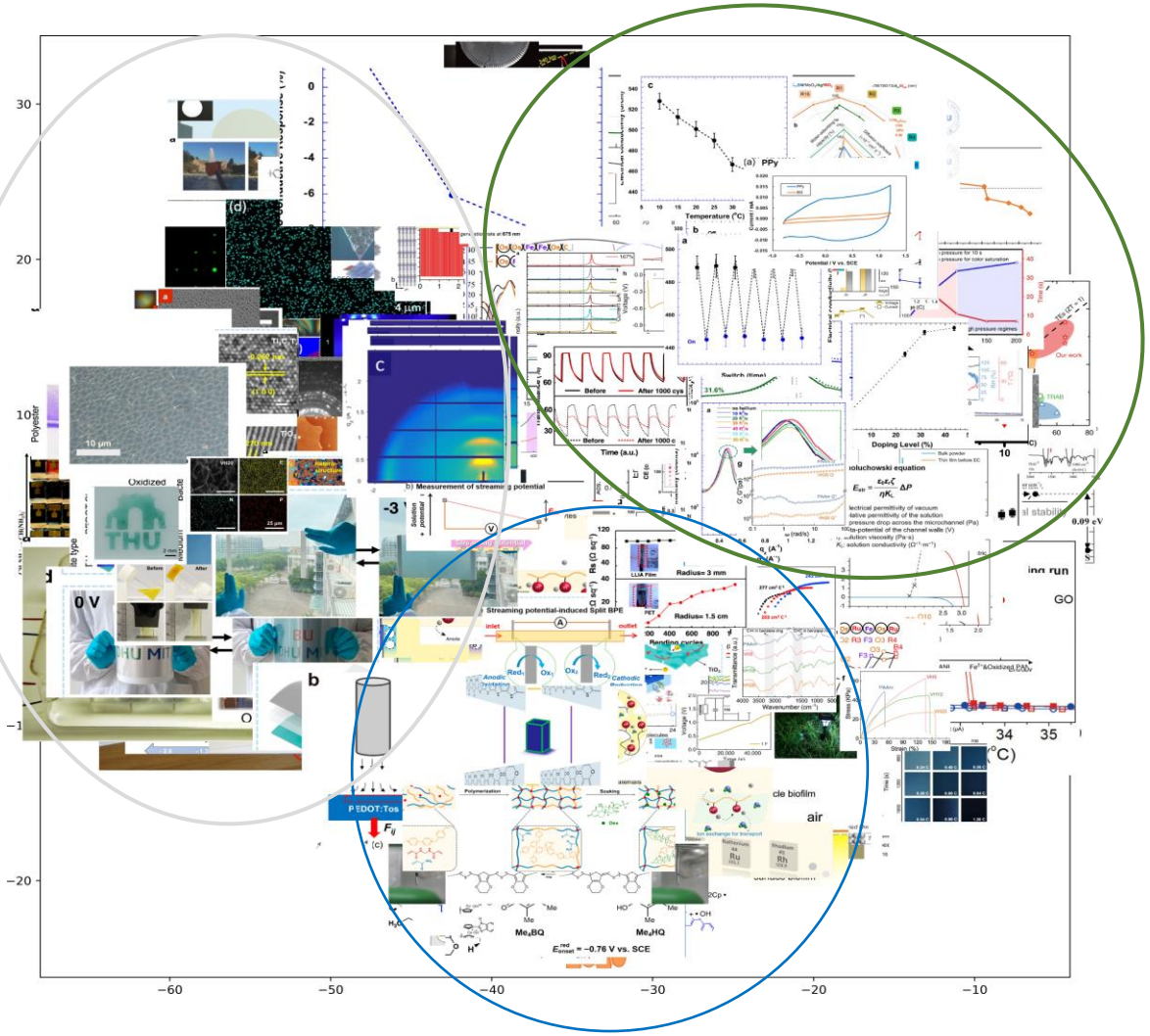


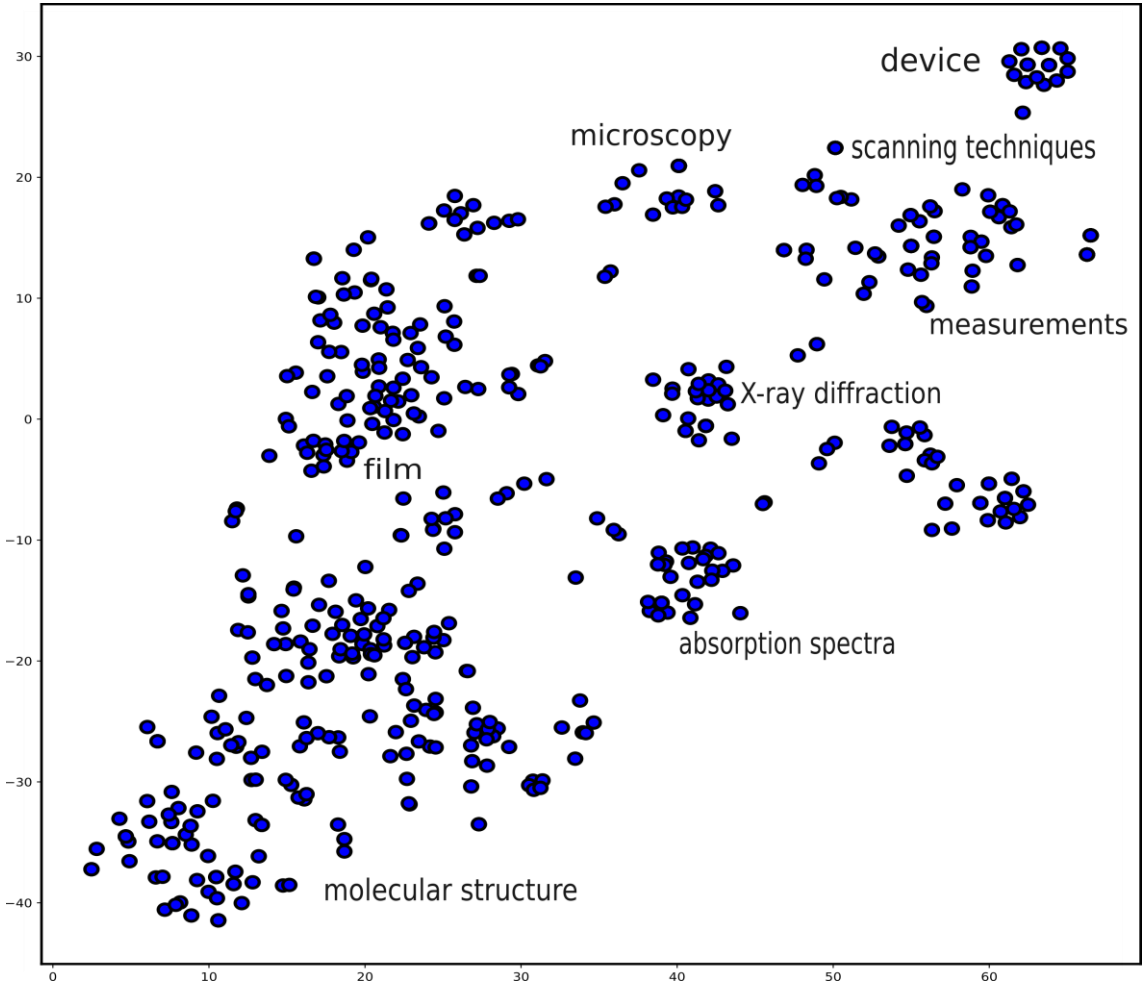
Specific categories

Argonne NATIONAL LABORATORY

Autonomous synthesis of electrochromic polymers

text embeddings

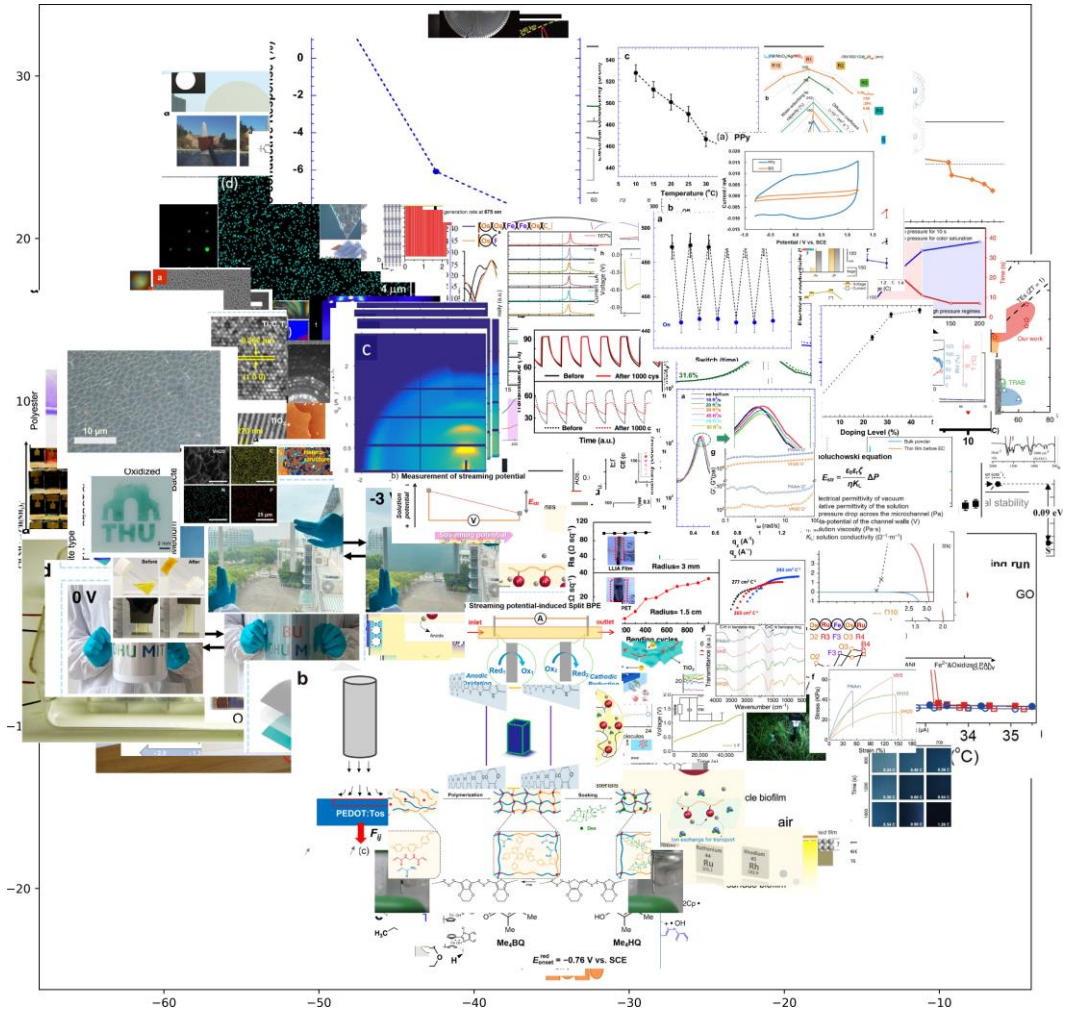




https://github.com/openai/CLIP

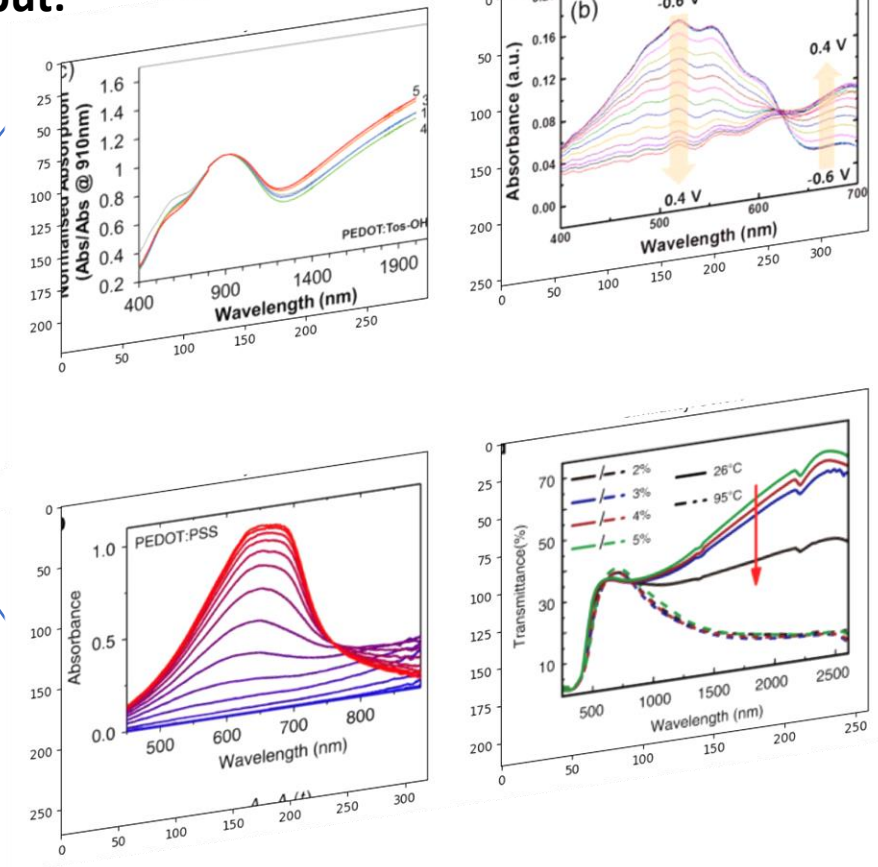
Autonomous synthesis of electrochromic polymers

Image embeddings



Query = 'Absorption Spectra'

Output:

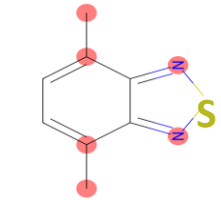


https://github.com/openai/CLIP

Argonne NATIONAL LABORATORY

Autonomous synthesis of electrochromic polymers

Monomer representation



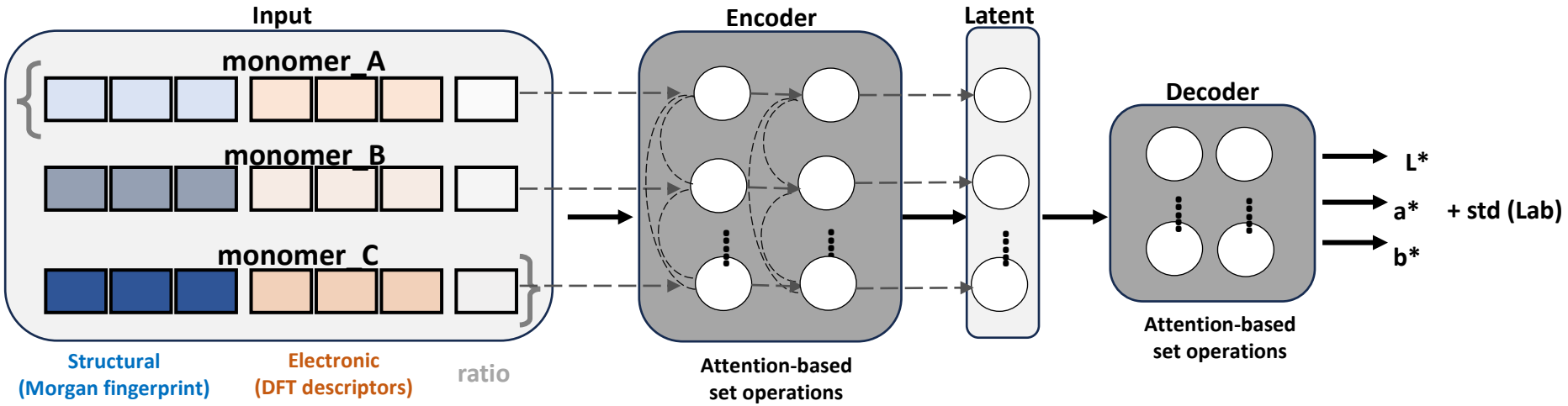
Morgan fingerprint

DFT descriptors

2D: n_heteroatoms, n_double_bonds
3D:'GeomDiameter', 'GeomRadius'

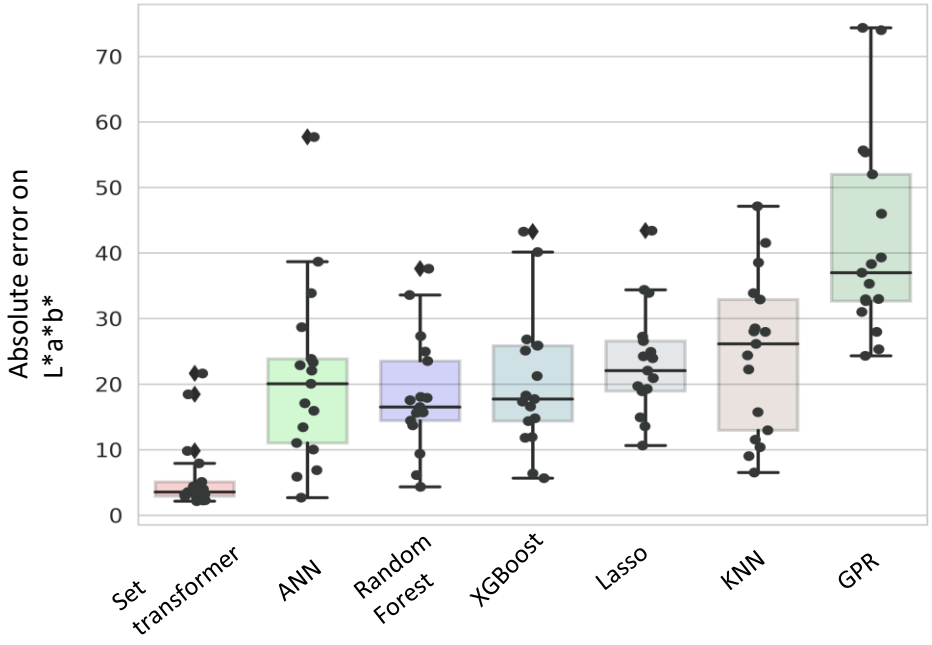
Mordred descriptors

Model architecture

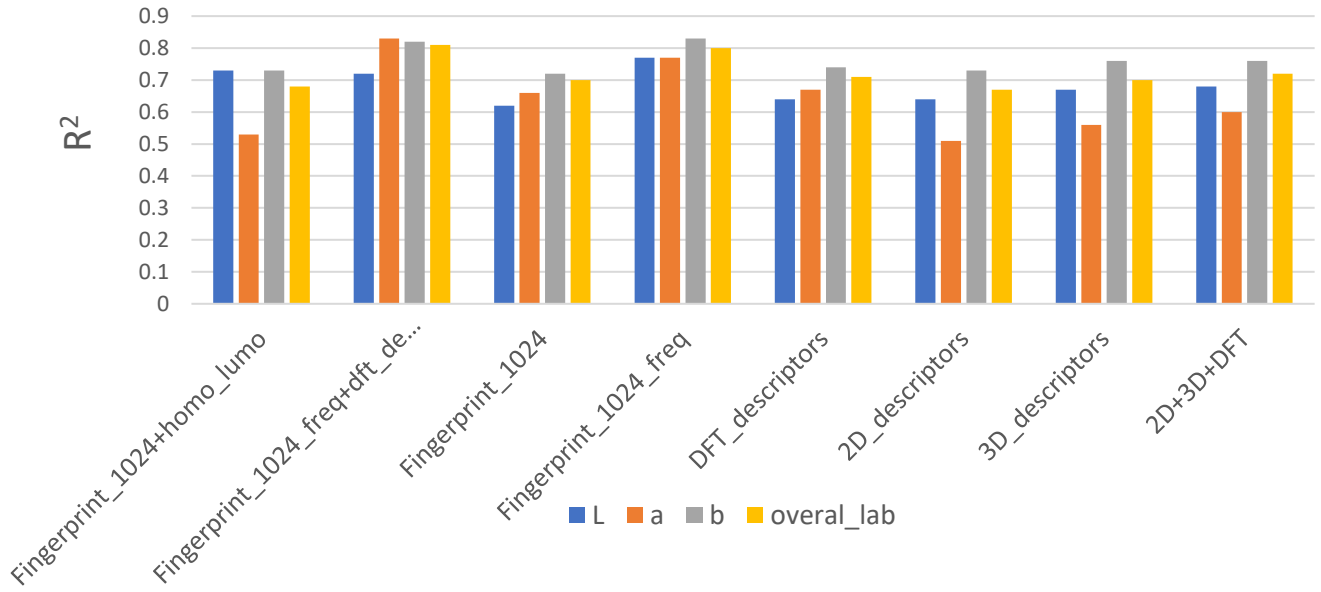


Autonomous synthesis of electrochromic polymers

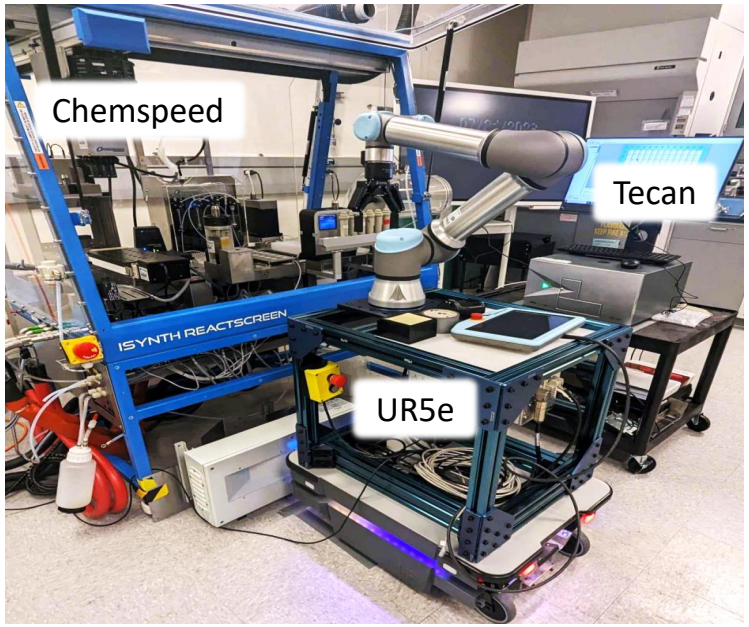
ML models comparison

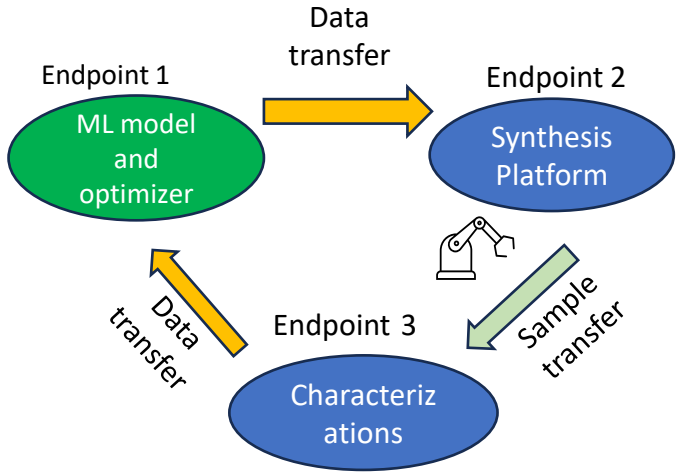


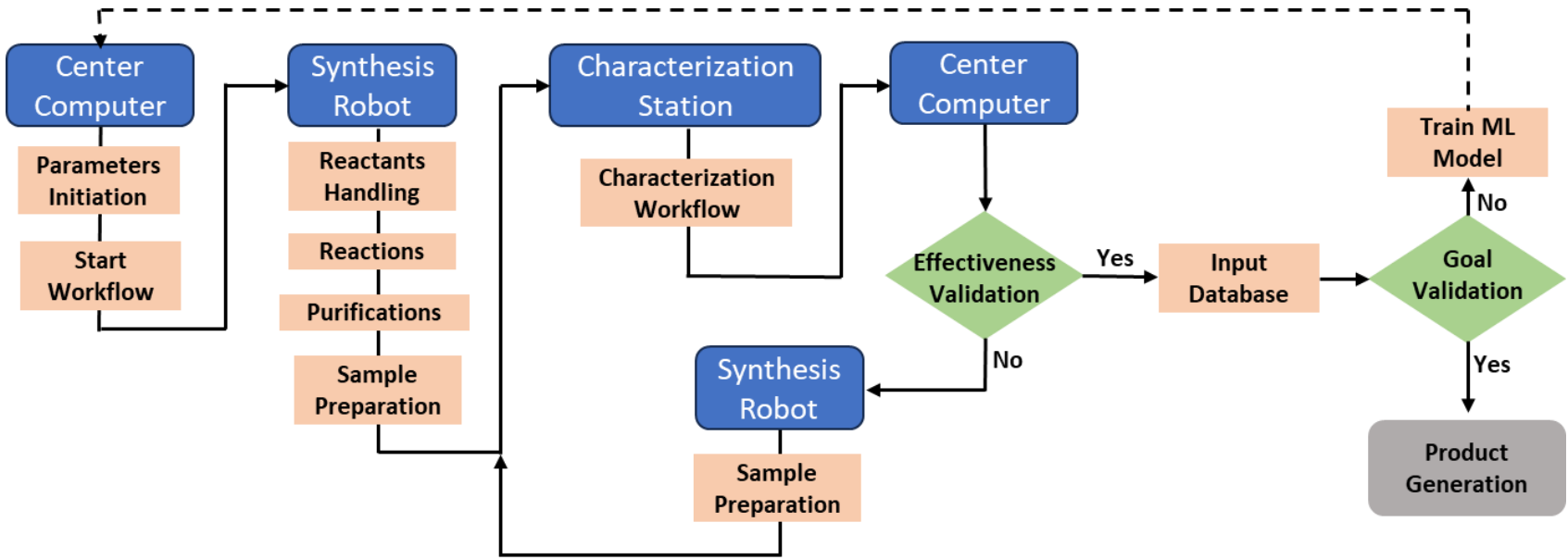
Molecular representation techniques comparison

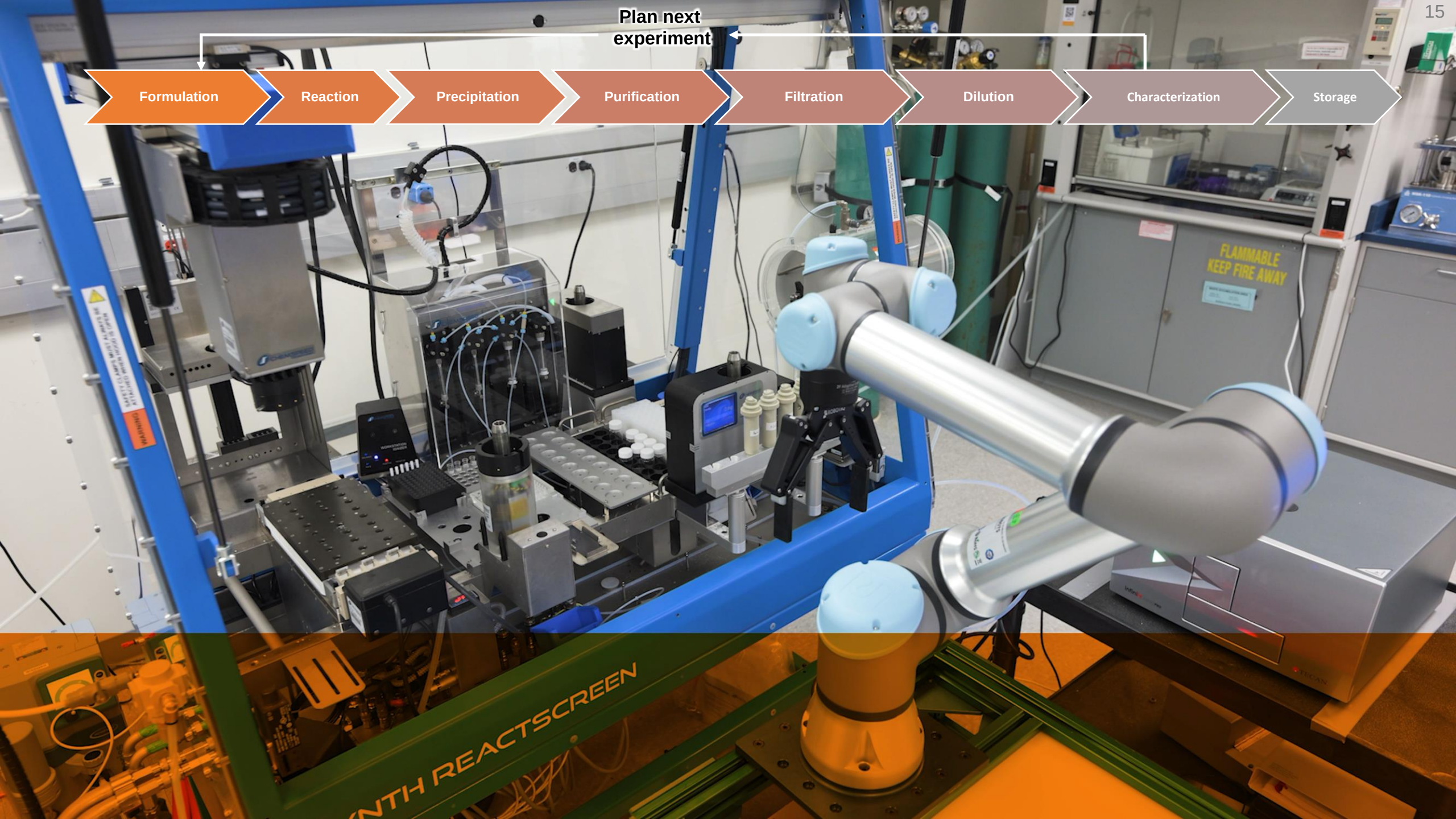


Autonomous synthesis of electrochromic polymers





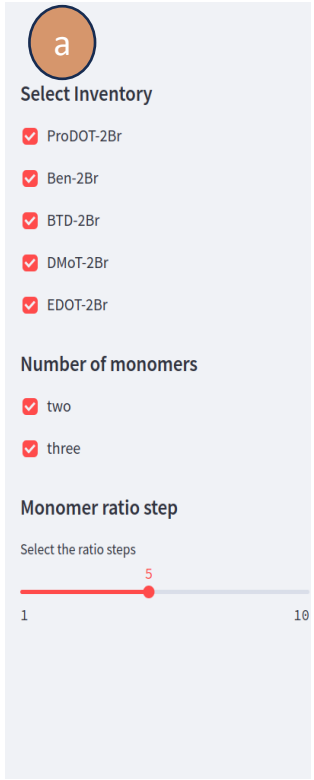


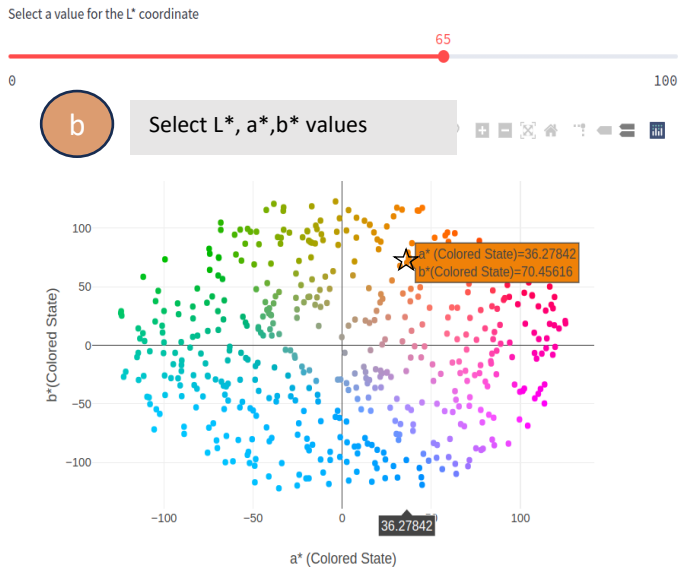

Plan next experiment
Formulation
Reaction
Precipitation
Purification
Filtration
Dilution
Characterization
Storage
FLAMMABLE
KEEP FIRE AWAY
NTH REACTSCREEN

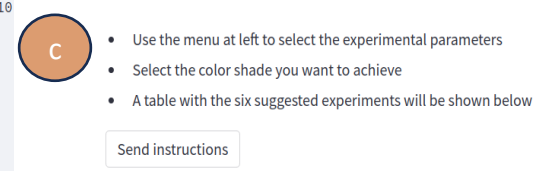
Autonomous synthesis of electrochromic polymers

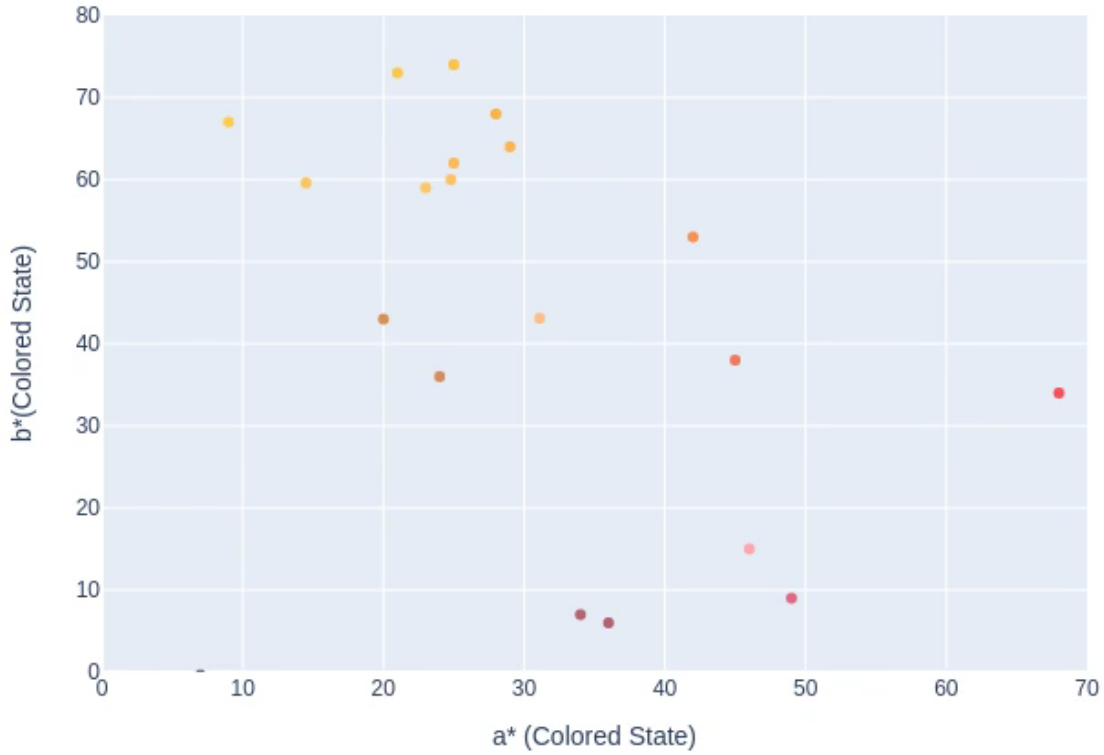
Importance | Database creation | Literature-aware ML | Robotic network | **Results**

- **SDL campaign**: Synthesize on demand ECPs with a selected color.
- **Objective**: Minimize color distance from user selection (ΔELab).
- Goal achieved within three experimental loops







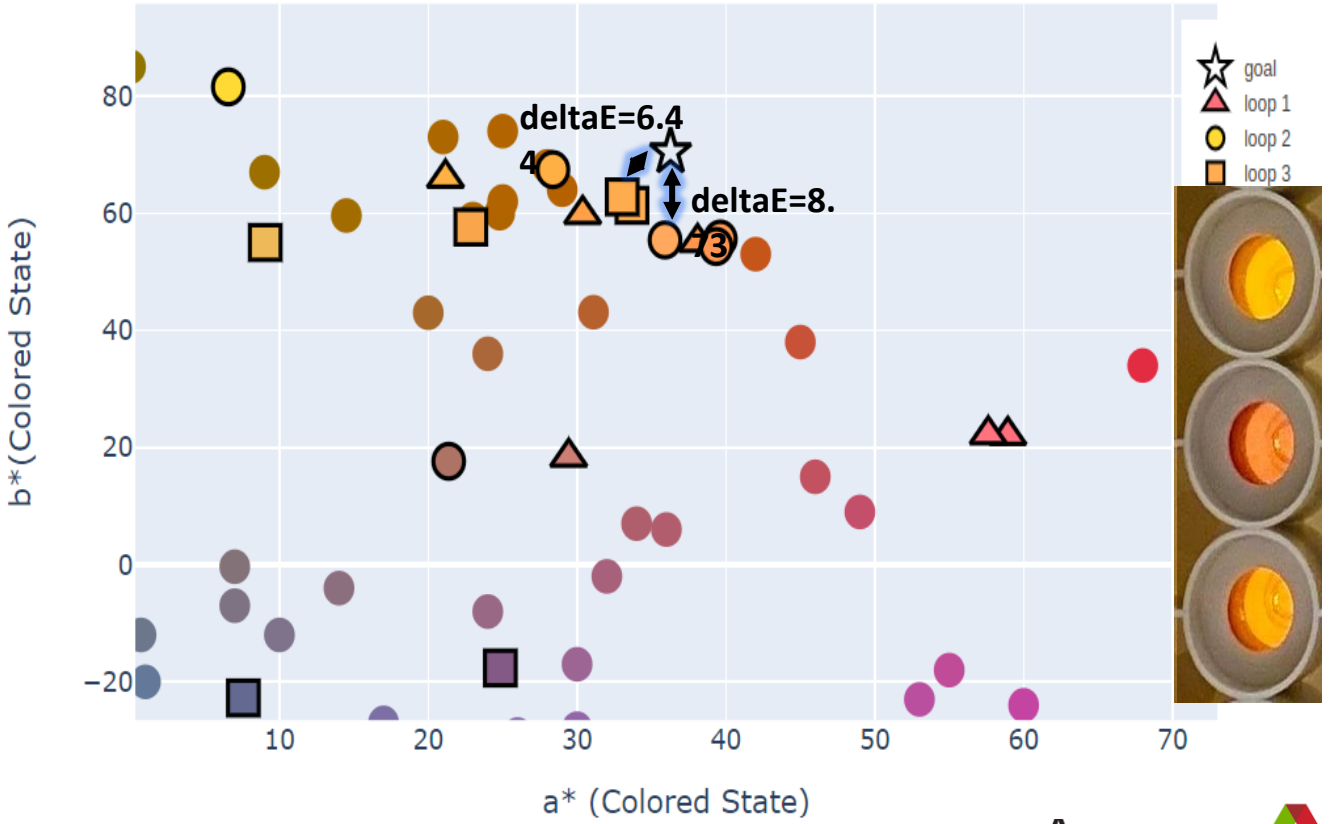


Argonne NATIONAL LABORATORY

Autonomous synthesis of electrochromic polymers

Importance → Database creation → Literature-aware ML → Robotic network → **Results**

- **SDL campaign**: Synthesize on demand ECPs with a selected color.
- **Objective**: Minimize color distance from user selection (ΔELab).
- Goal achieved within three experimental loops



Argonne NATIONAL LABORATORY