

Autonomous chemical research with large language models

Gabe Gomes

Department of Chemistry

Department of Chemical Engineering

Wilton E. Scott Institute for Energy Innovation

Carnegie Mellon University

EWO Seminar, Center for Advanced Process Decision-making

Carnegie Mellon University, Pittsburgh, PA

April 3rd, 2024



the team (Spring '24)

department of
{ chemistry |
chemical
engineering }

Carnegie
Mellon
University

PhD students



Robert MacKnight
ChemE, G3
TCS Presidential
Fellow



Daniil Boiko
ChemE, G2
on leave @
stealth biotech startup



Jose Regio
Chemistry, G1
AFRL Fellow



Gabe Adriano
Chemistry, G1
NSF GRFP



Evan Spotte-Smith
NSF CCAS Postdoc Fellow



Liliana Gallegos
MCS Cloud Lab Fellow



Hannah Carroll
hannahca@andrew.cmu.edu

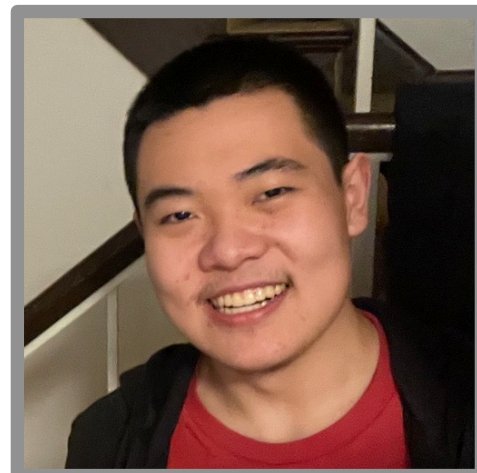
Postdoctoral Fellows

Admin

Undergrad

Visiting Research Fellows

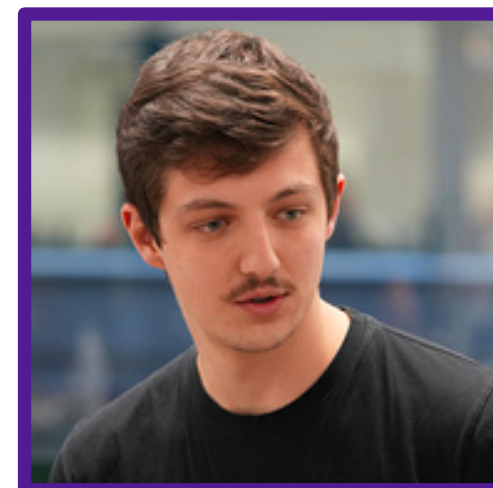
Research Affiliates



Jiale Lu
CMU Chemistry
SURF Summer '24



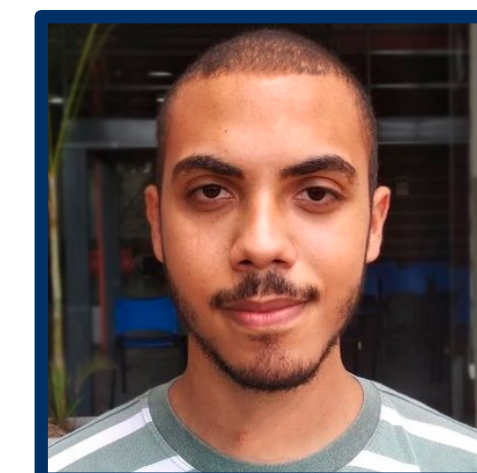
Edgar Sanchez
(PhD in ChemE @ Max
Planck Institute for Dynamics
of Complex Technical
Systems, Sundmacher group)



Theo Nekomm
(MSc in ML for Chemistry @
EPFL, Schwaller group)



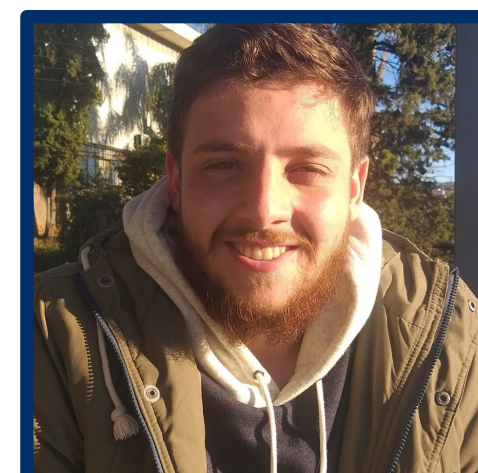
Yusef Ahmed
(PhD in Chemistry @
UC Irvine, Tantillo group)



Augusto Cesar
(MS in ML @
UNICAMP, Brazil)



Melissa Ramirez
NSF PDF @ Caltech |
Assistant Professor
@ UMN Chemistry,
Summer '25



Thiago
Reschütze
(undergrad in
ChemE @ UFSM) |
IBM Brazil intern



reaction discovery and optimization in the age of digital chemistry

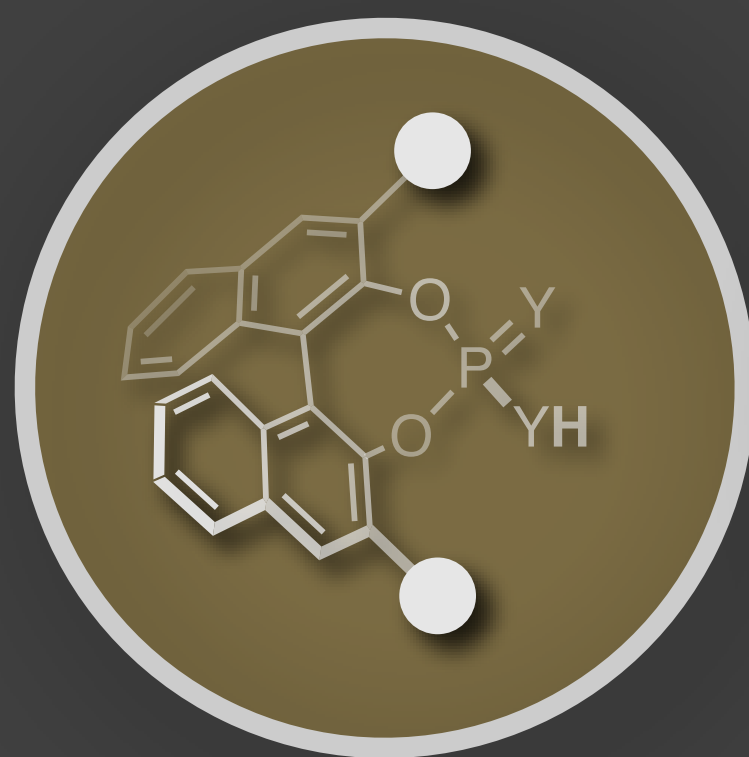
department of
{ chemistry |
chemical
engineering }

**Carnegie
Mellon
University**

*from automated
computational
chemistry*



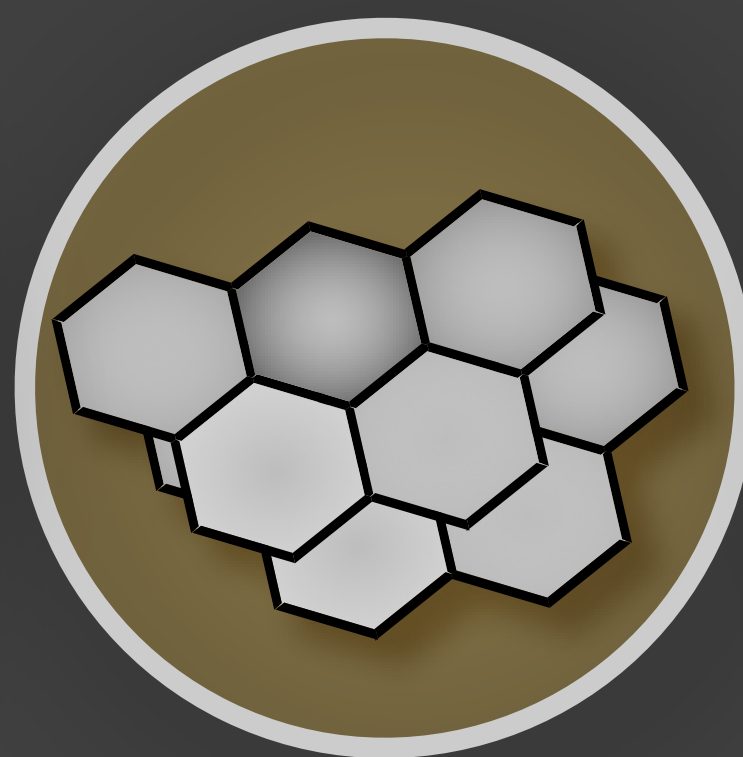
*to autonomous
chemical
synthesis**



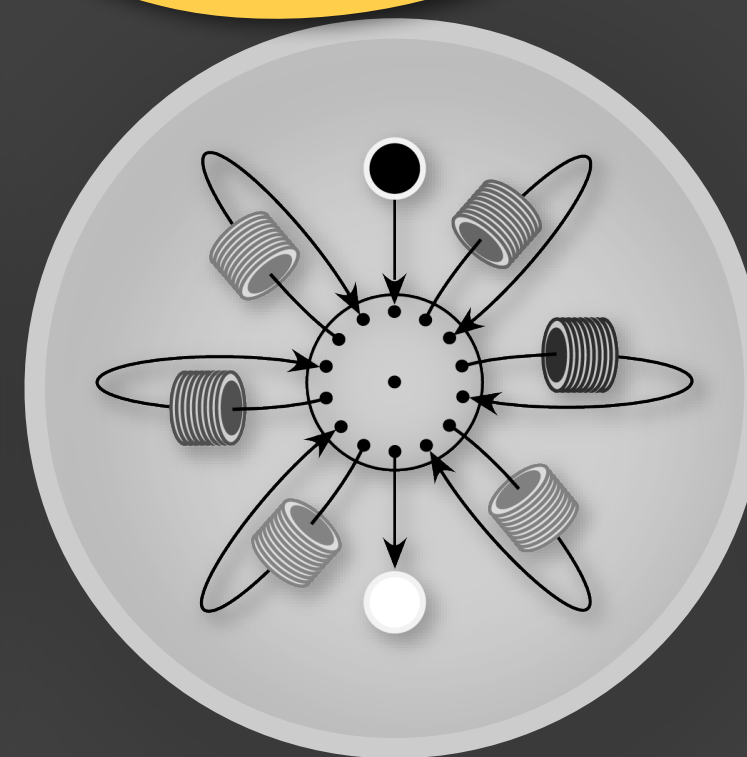
*catalysis &
reaction design*



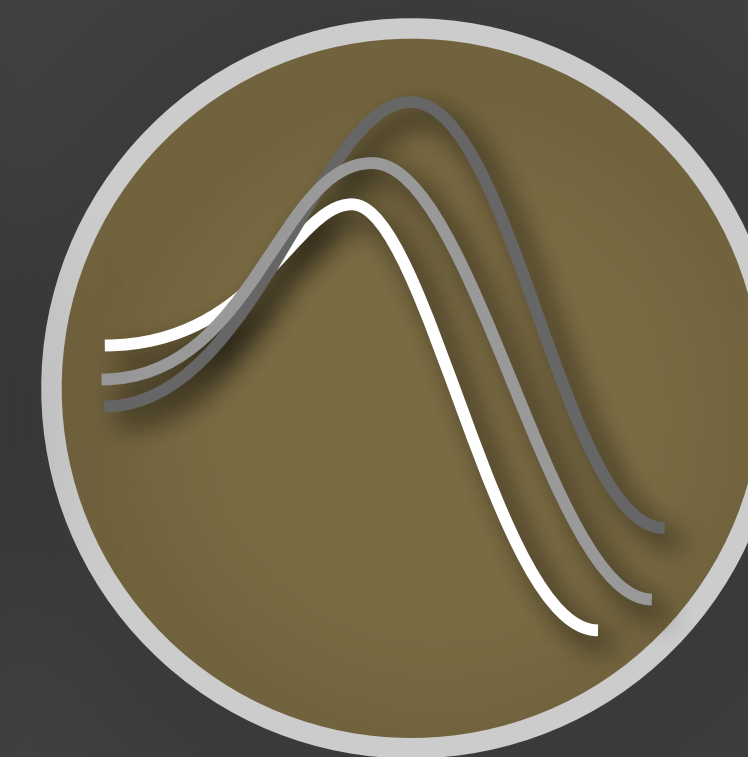
*machine learning
& computer science*



*(bio)organic
materials*



*automated
(flow) synthesis*

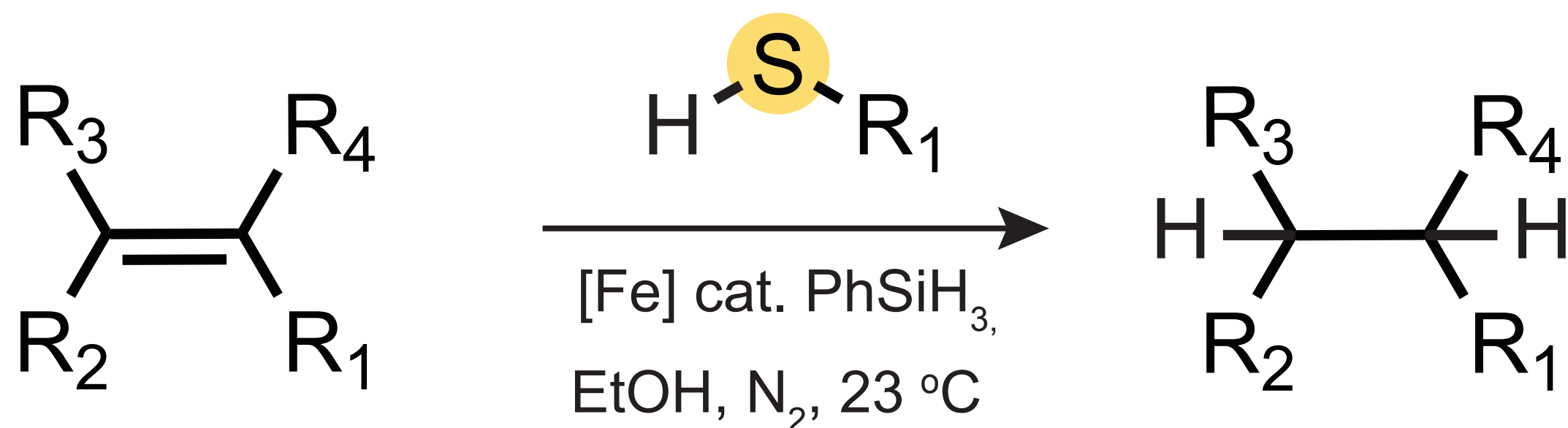


*physical
organic chemistry*



Mechanism-Informed Bayesian Reaction Optimization

department of
{ chemistry |
chemical
engineering }

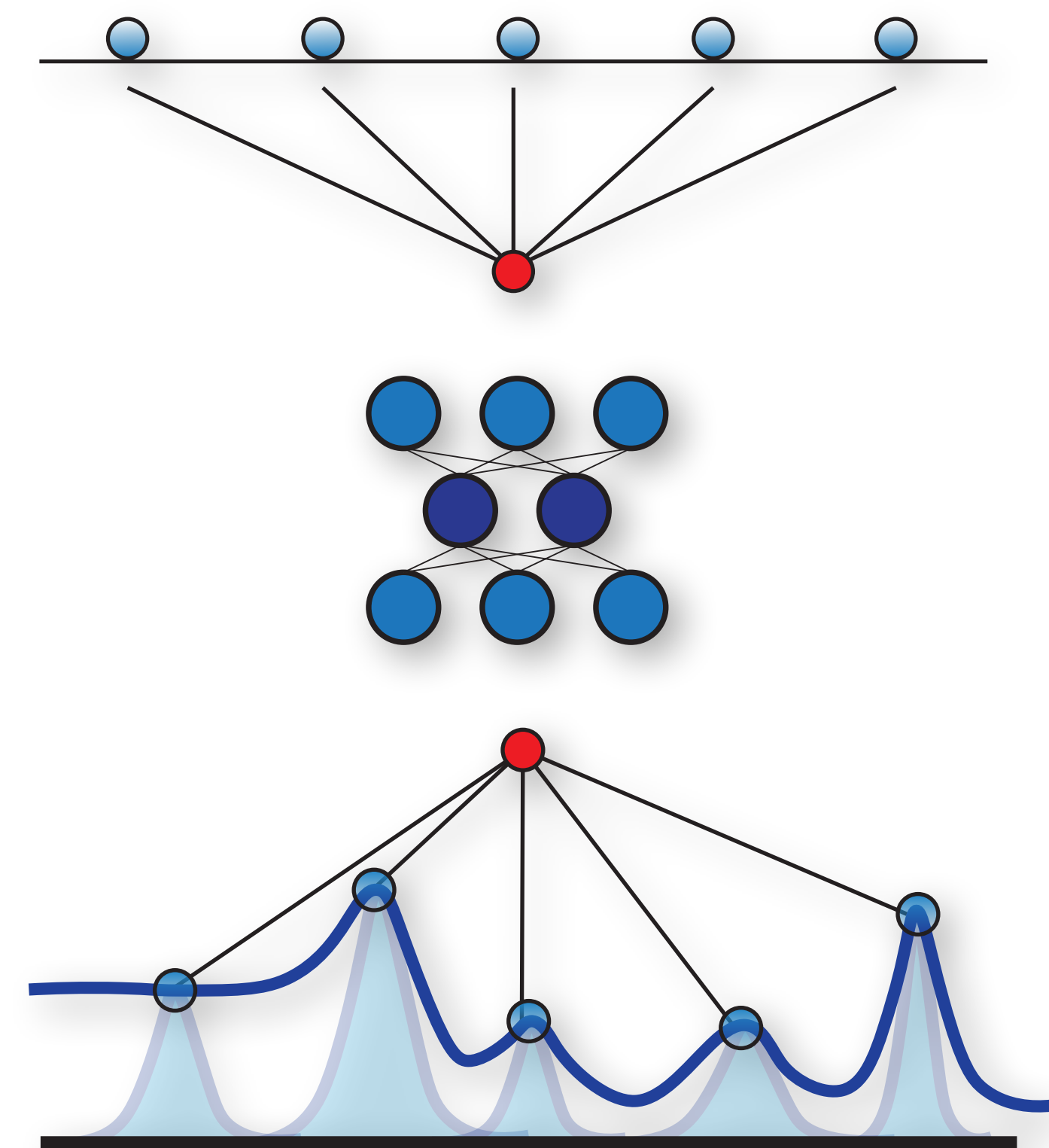
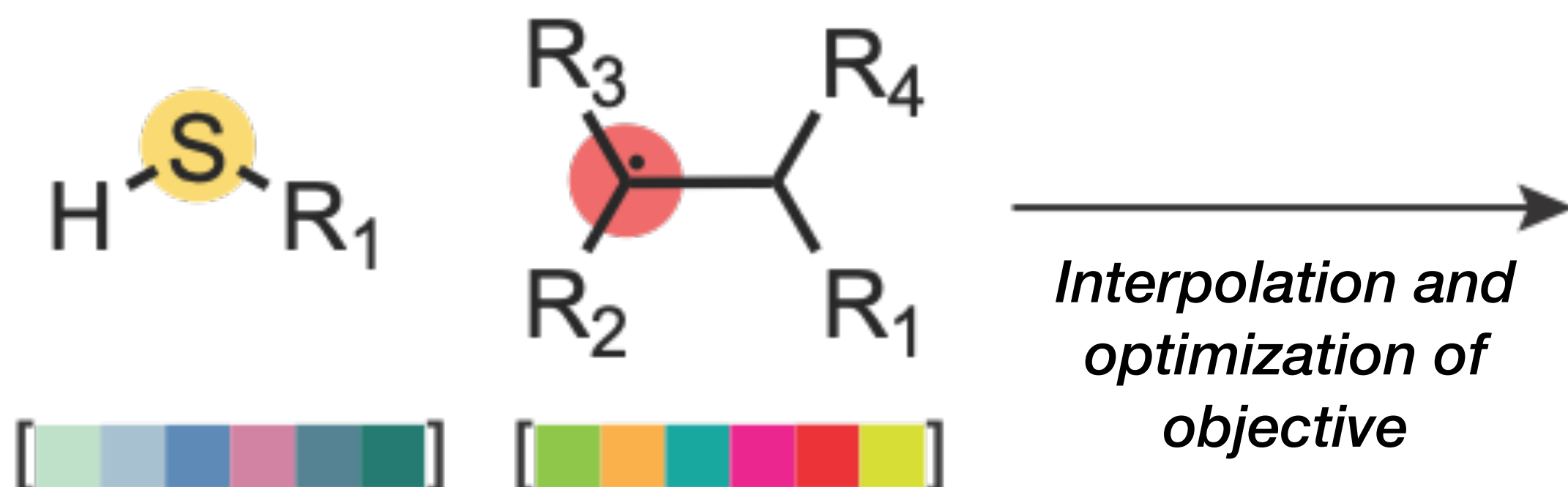


AFRL
THE AIR FORCE RESEARCH LABORATORY

D³OM²S
Center of Excellence

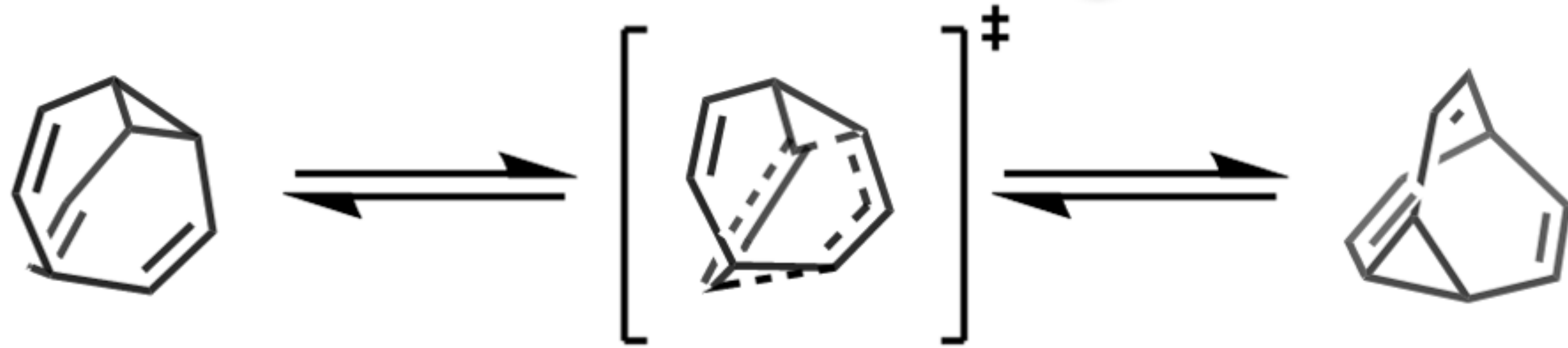
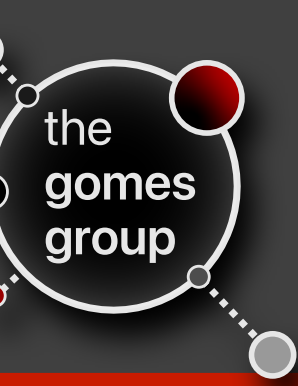


*Yield/Selectivity varies across
alkene-thiol combinations*

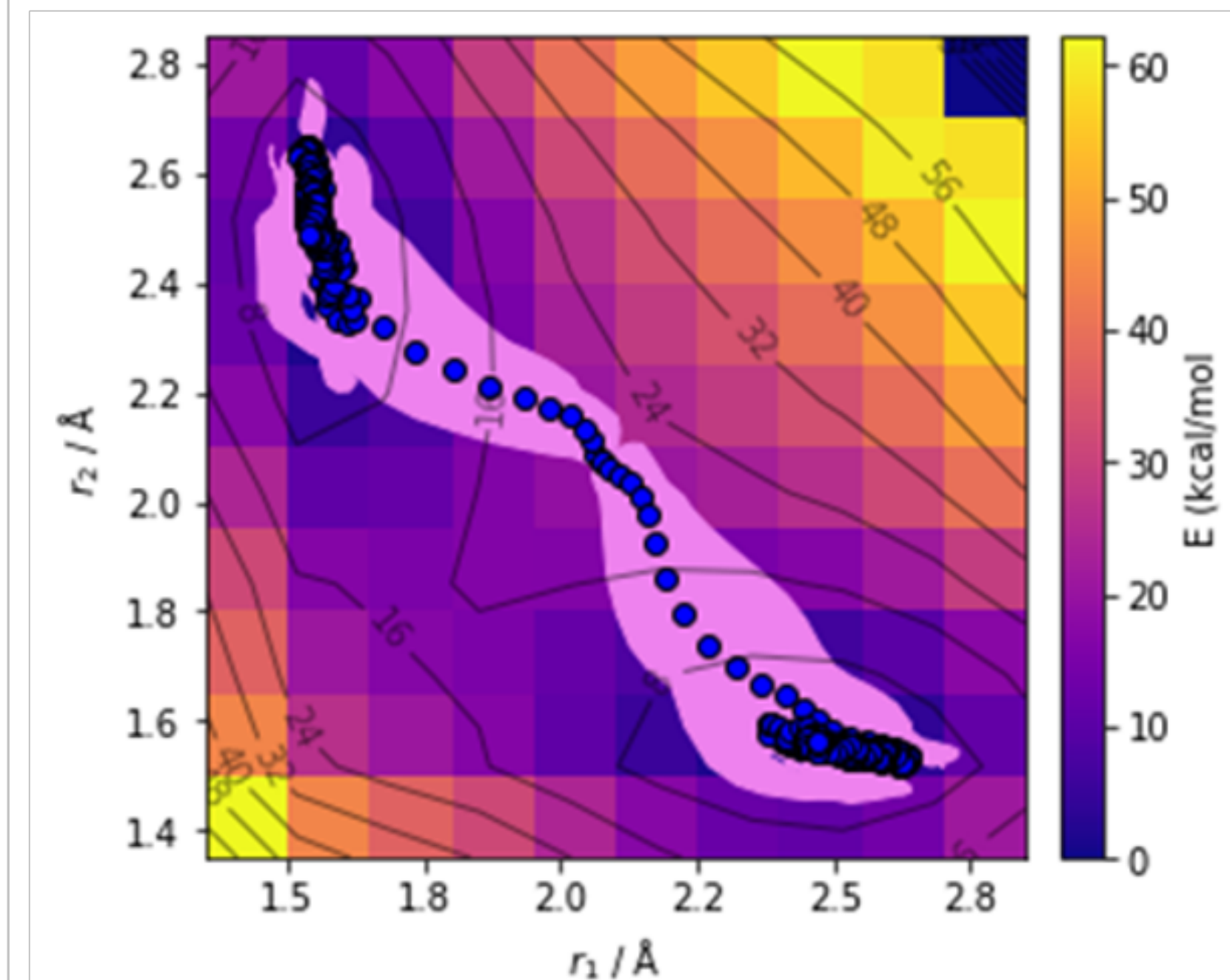
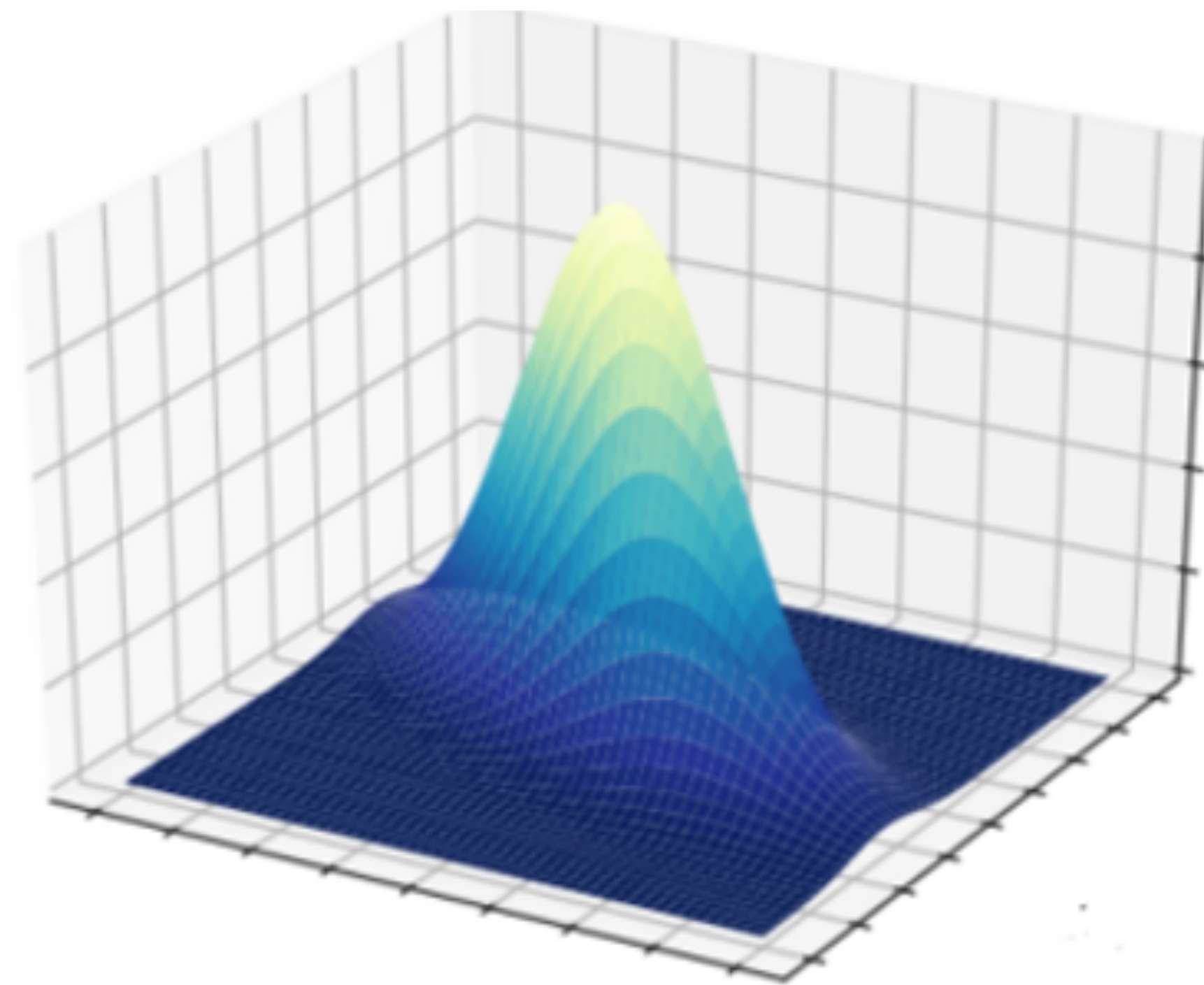
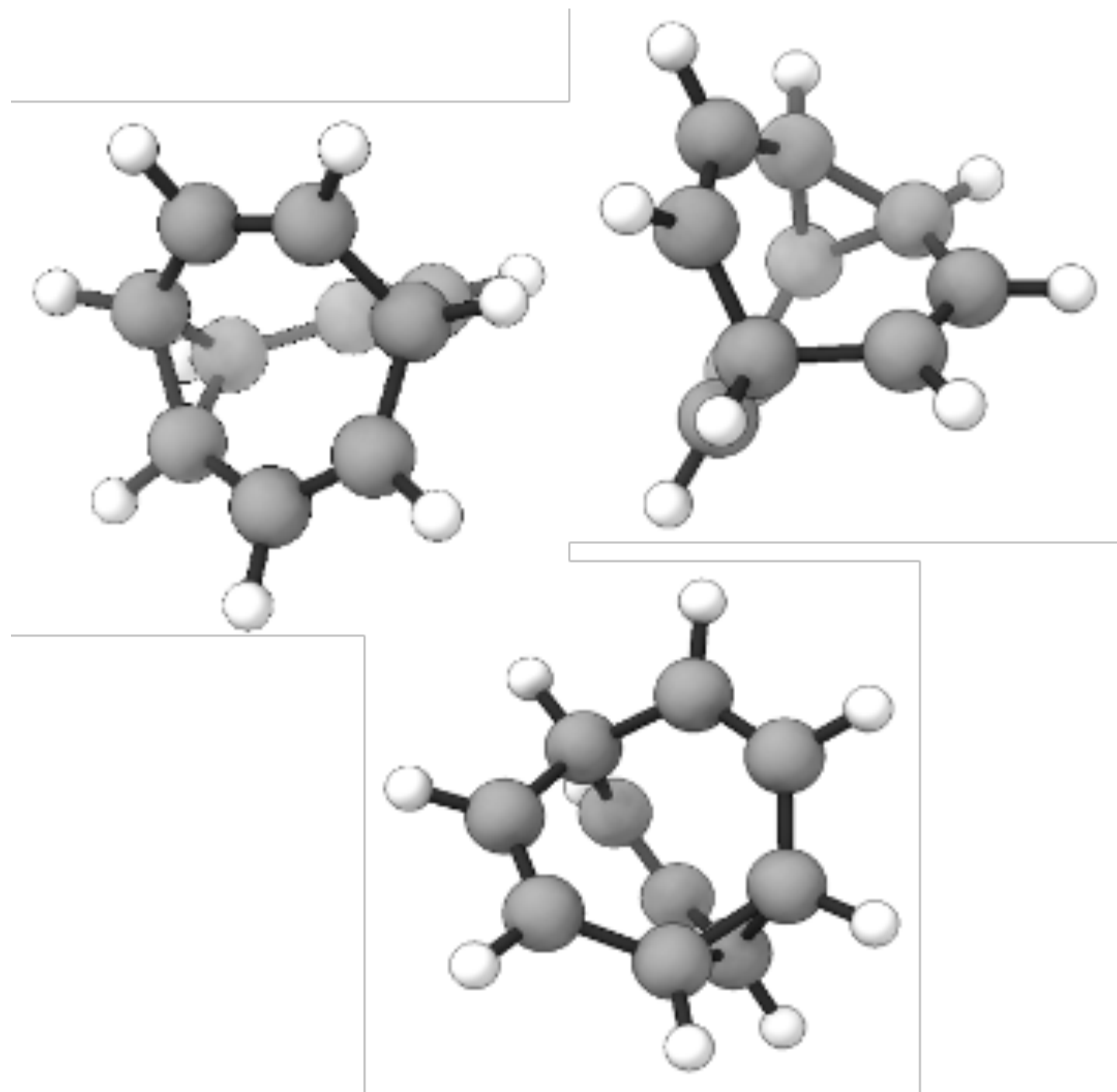


Learning Dynamics of Fluxional Systems with Quantum Machine Learning

department of
{ chemistry |
chemical
engineering }



IBM

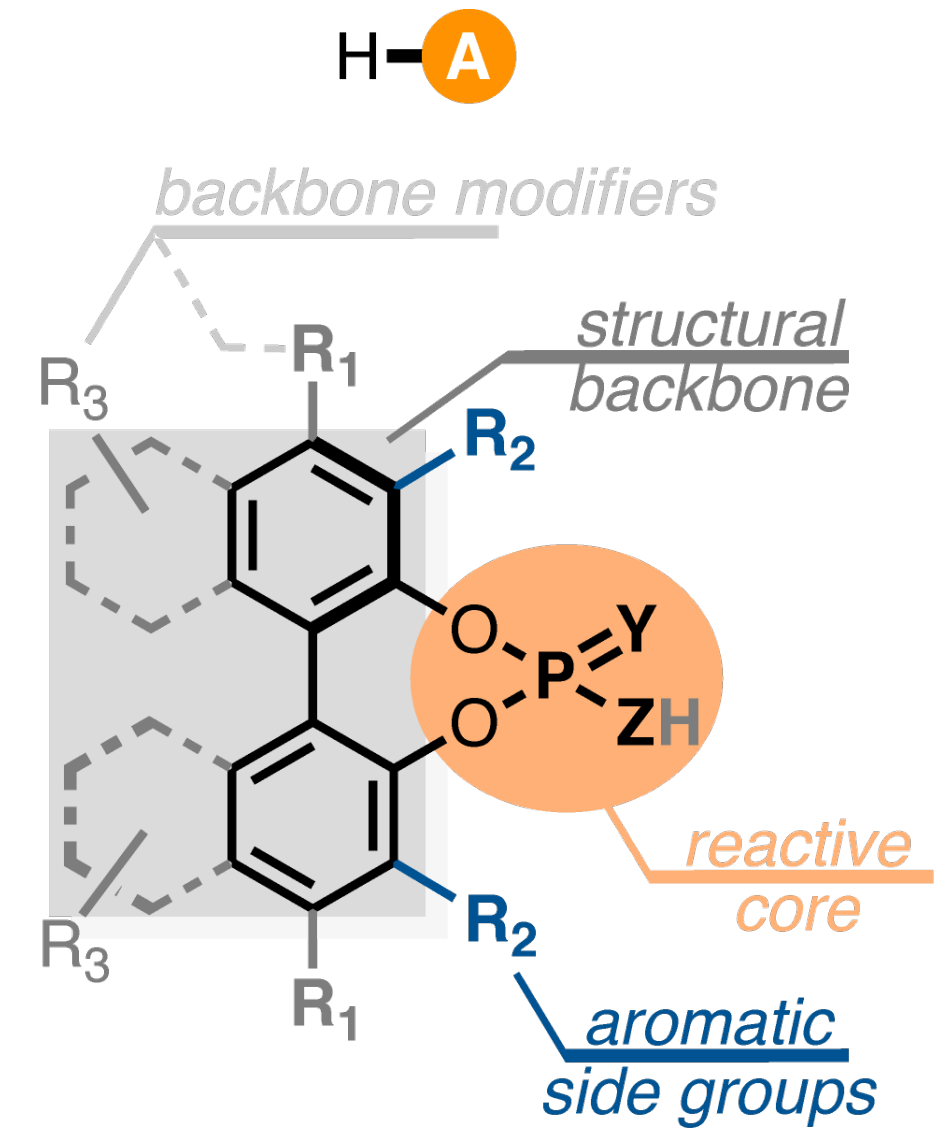


**systems configurations
and energies**

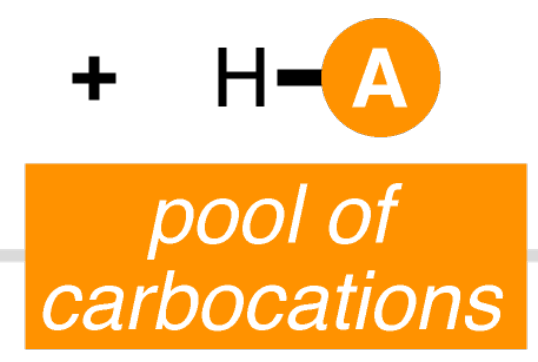
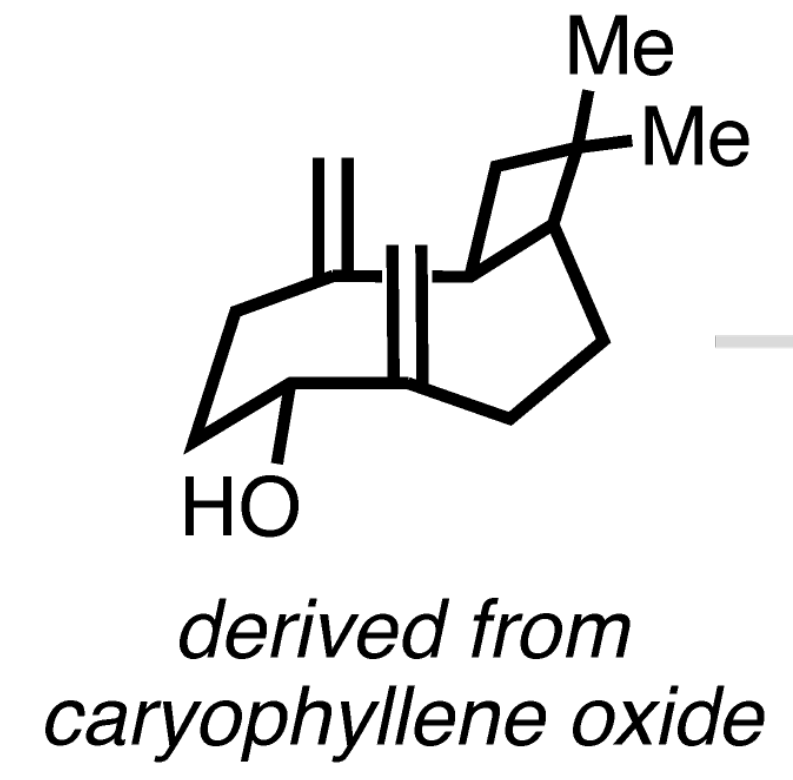
**trained bespoke
machine learning potentials**

reactive molecular dynamics

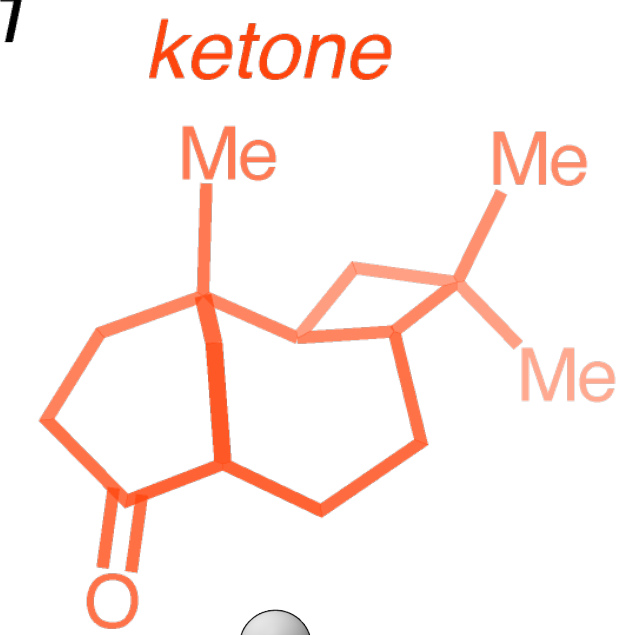
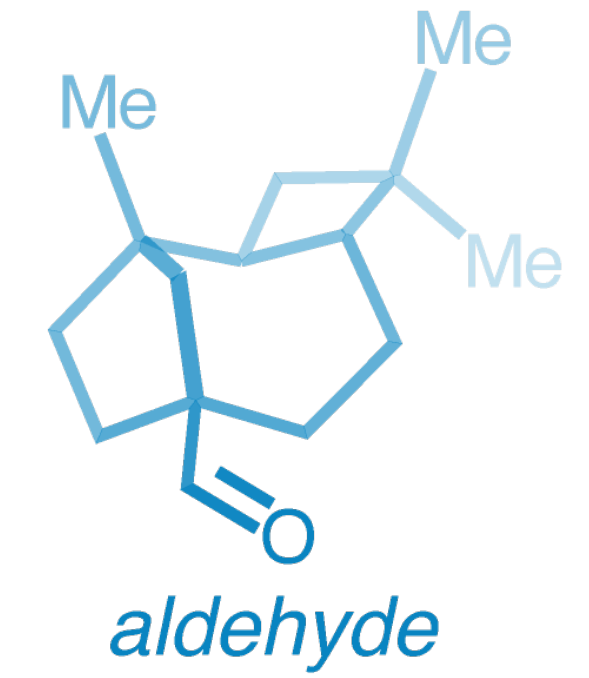
Computational Directed Evolution of Catalysts



CoDECs

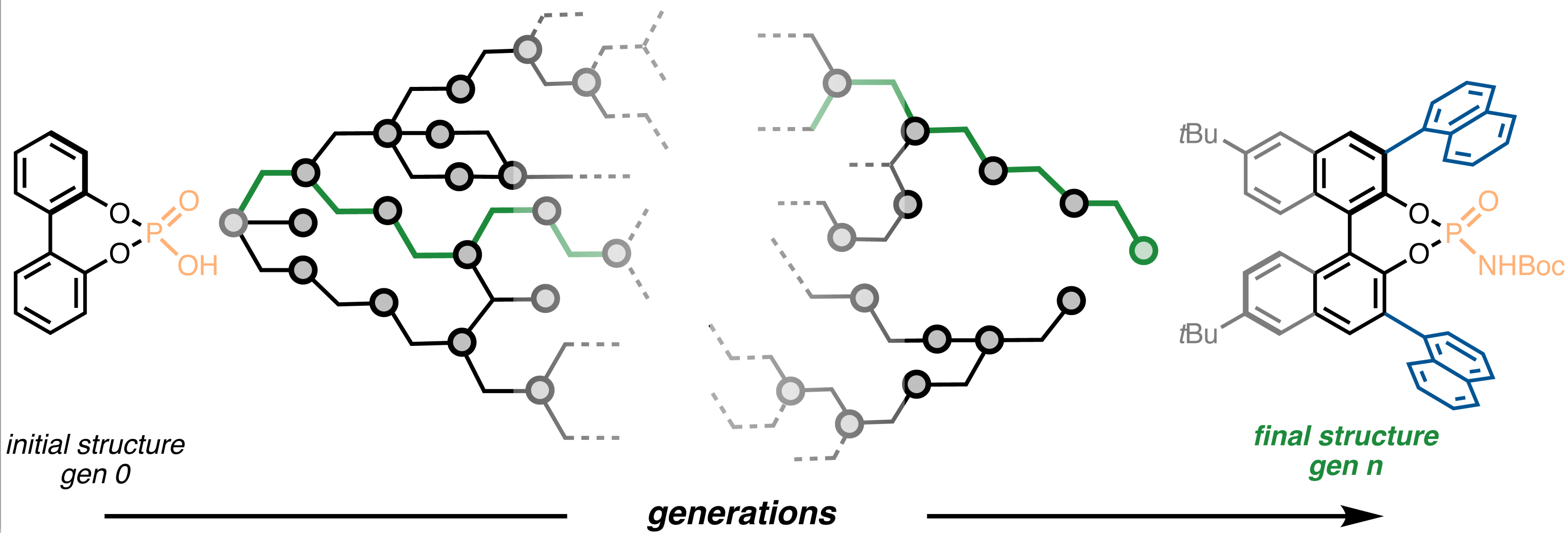


ratios from
1:1 to 6.7:1

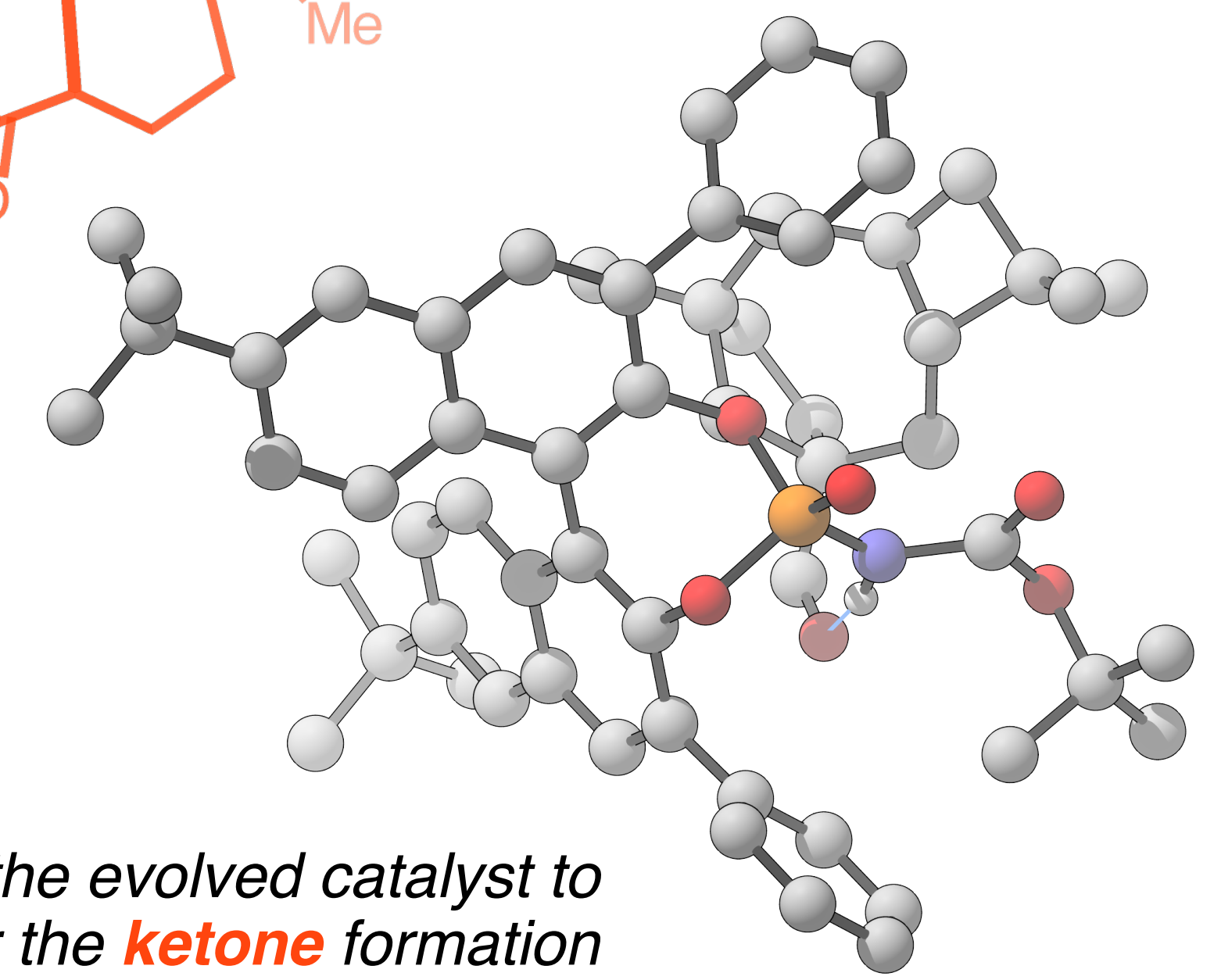


Liliana Gallegos
MCS Cloud Lab Fellow

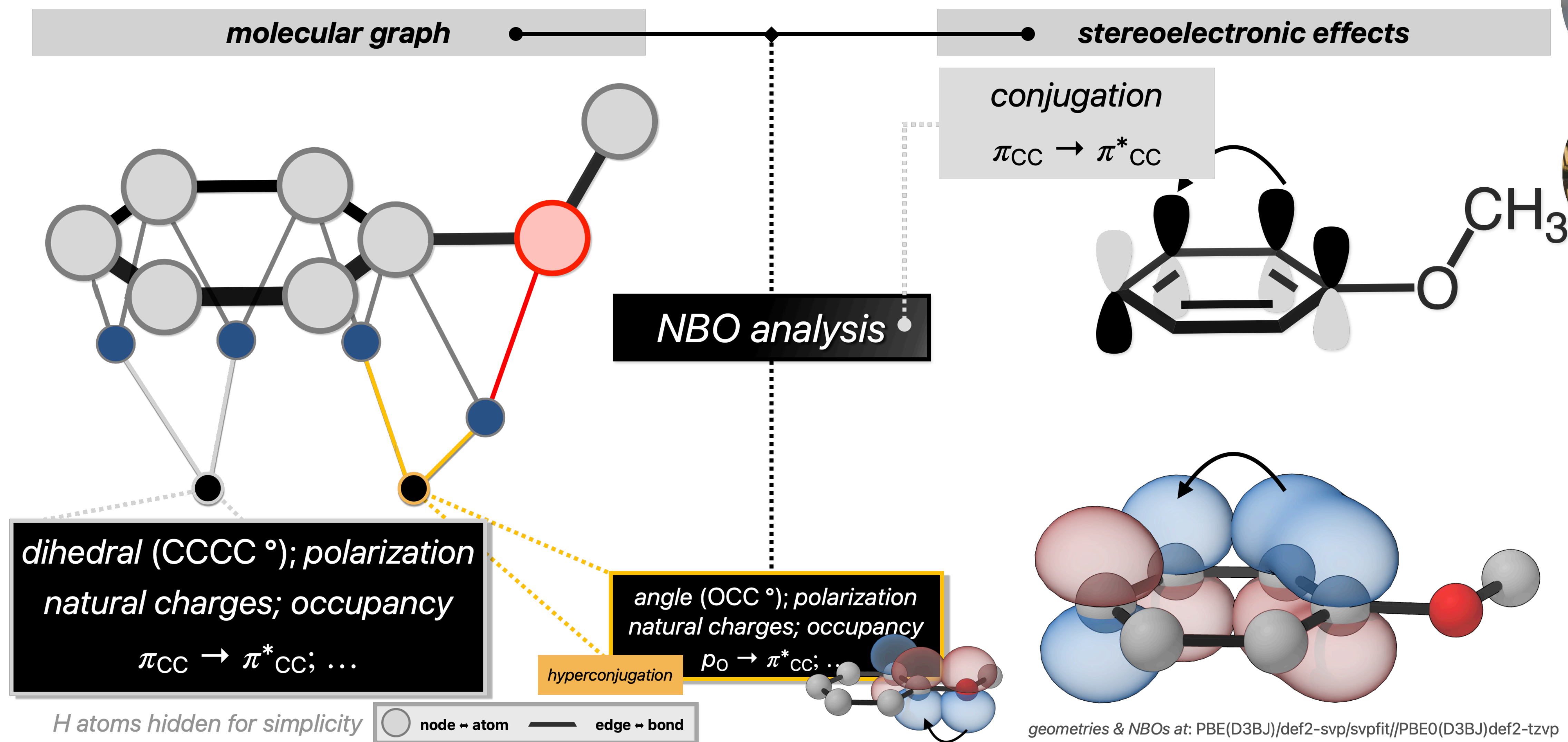
GA evolutionary tree structure



TS with the evolved catalyst to
favor the **ketone** formation



(Molecular) Representation Learning

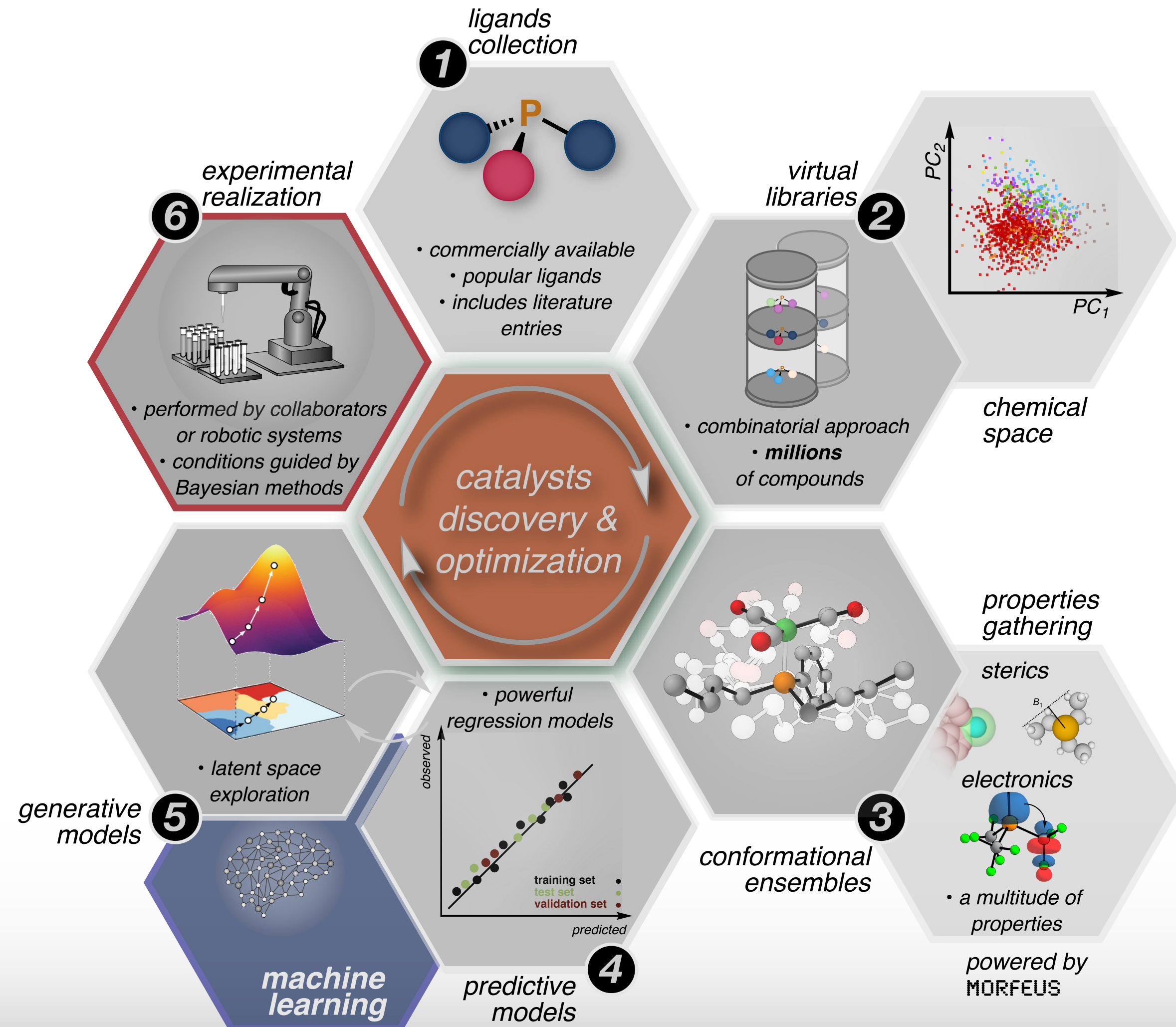


in collaboration with:

Google DeepMind

BERKELEY LAB

Catalyst Discovery Platform

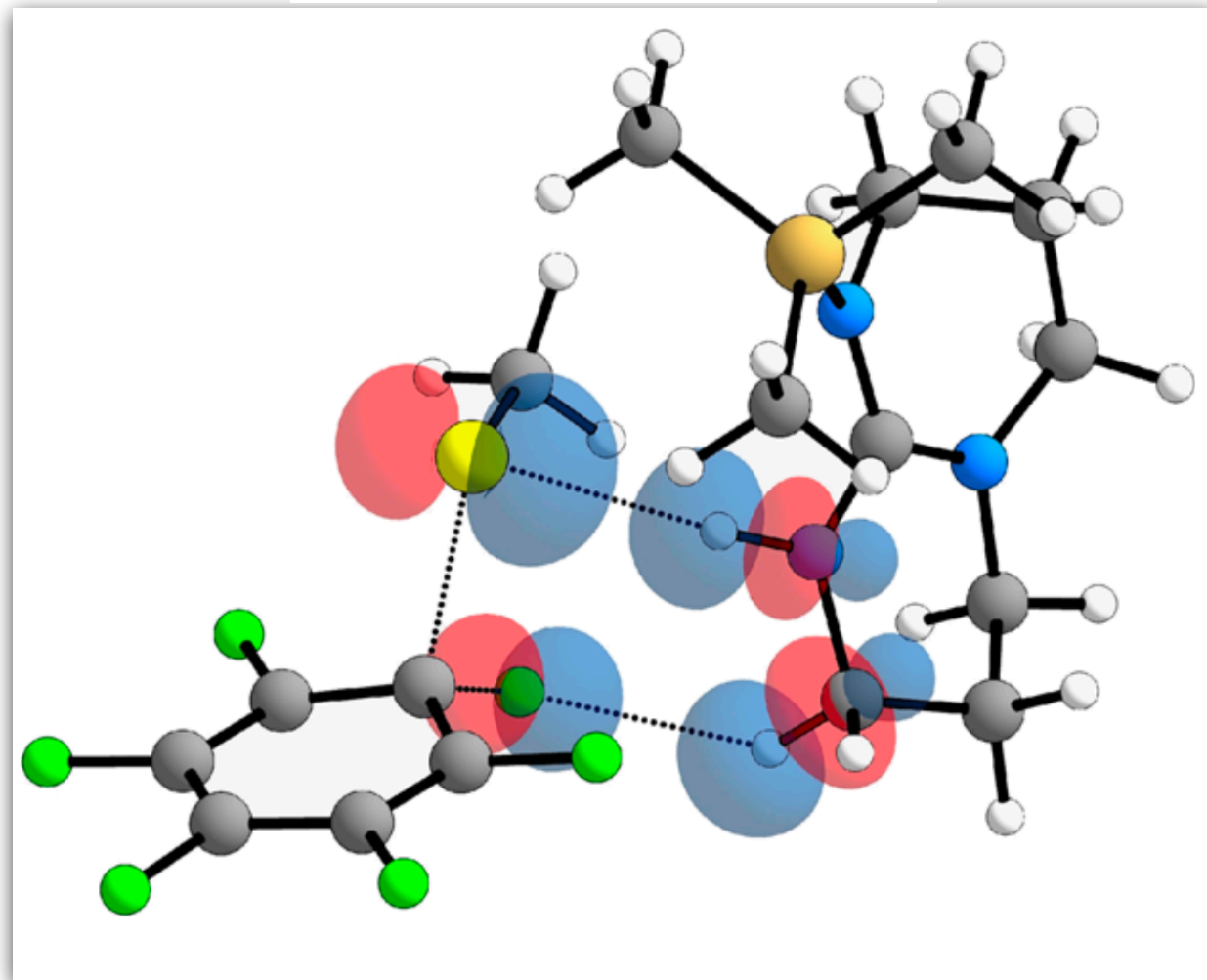
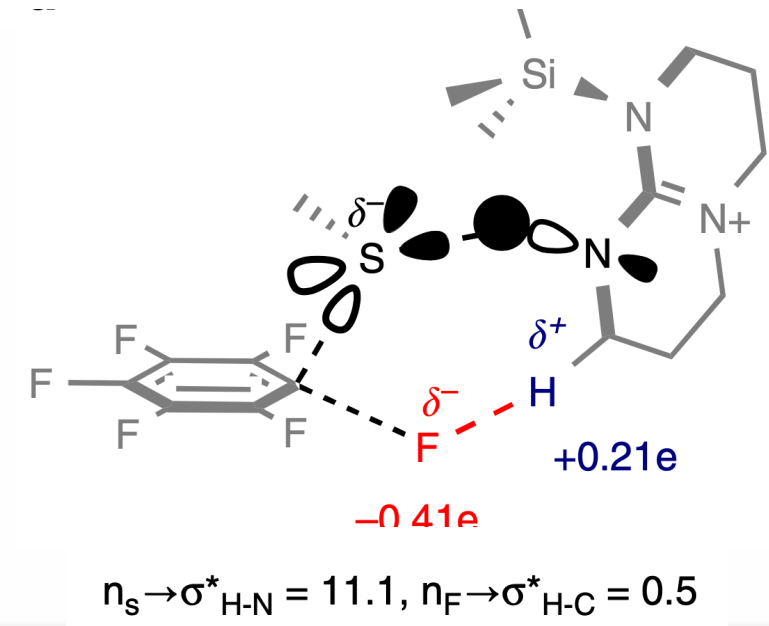


Developing an end-to-end framework for the inverse design of catalysts.

- *Trends in Chemistry* **2021**, 3(2), 96
- *Communications Chemistry* **2021**, 4, 112
- *JACS* **2022**, 144 (3), 1205

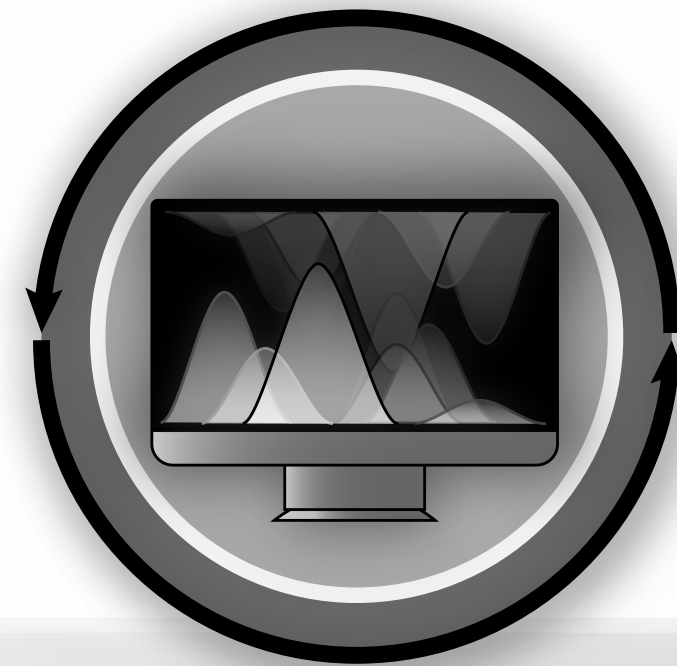
translational research at the core of the Gomes group

fundamental, molecular
chemistry and catalysis research

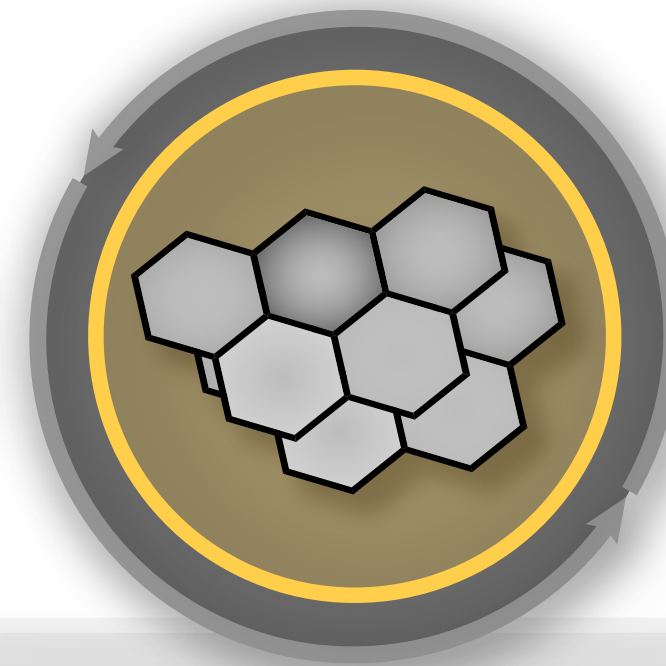


Nature Communications 2017, 8, 166
with IBM Research, CA

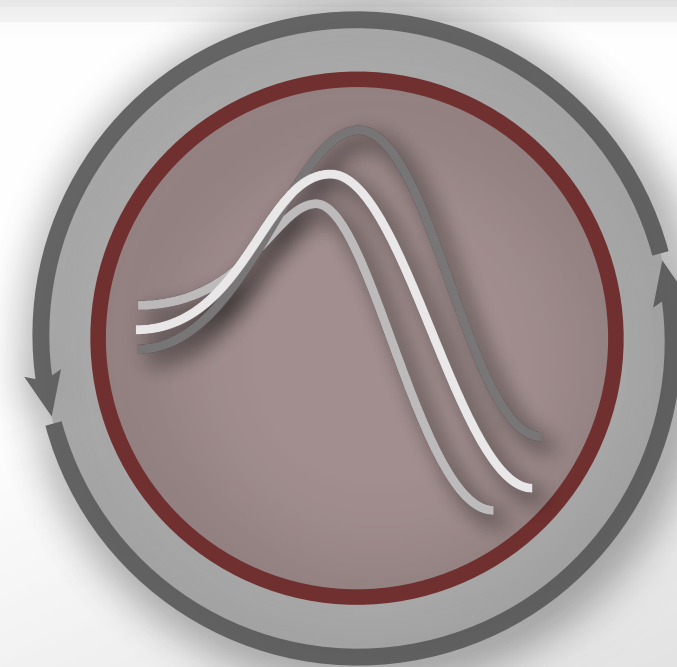
the gomes group
research pillars



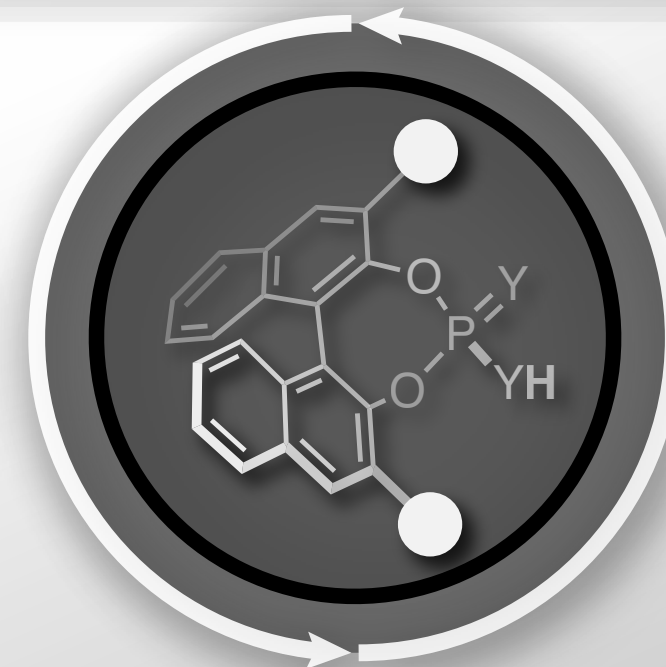
machine learning
& computer science



organic
materials



physical
organic chemistry



catalysis &
reaction design

applied, chemical engineering
and materials research

Plastics News

NEWS OPINION SHOP FLOOR EVENTS RESIN PRICES RANKINGS DATA STORE DIRECT

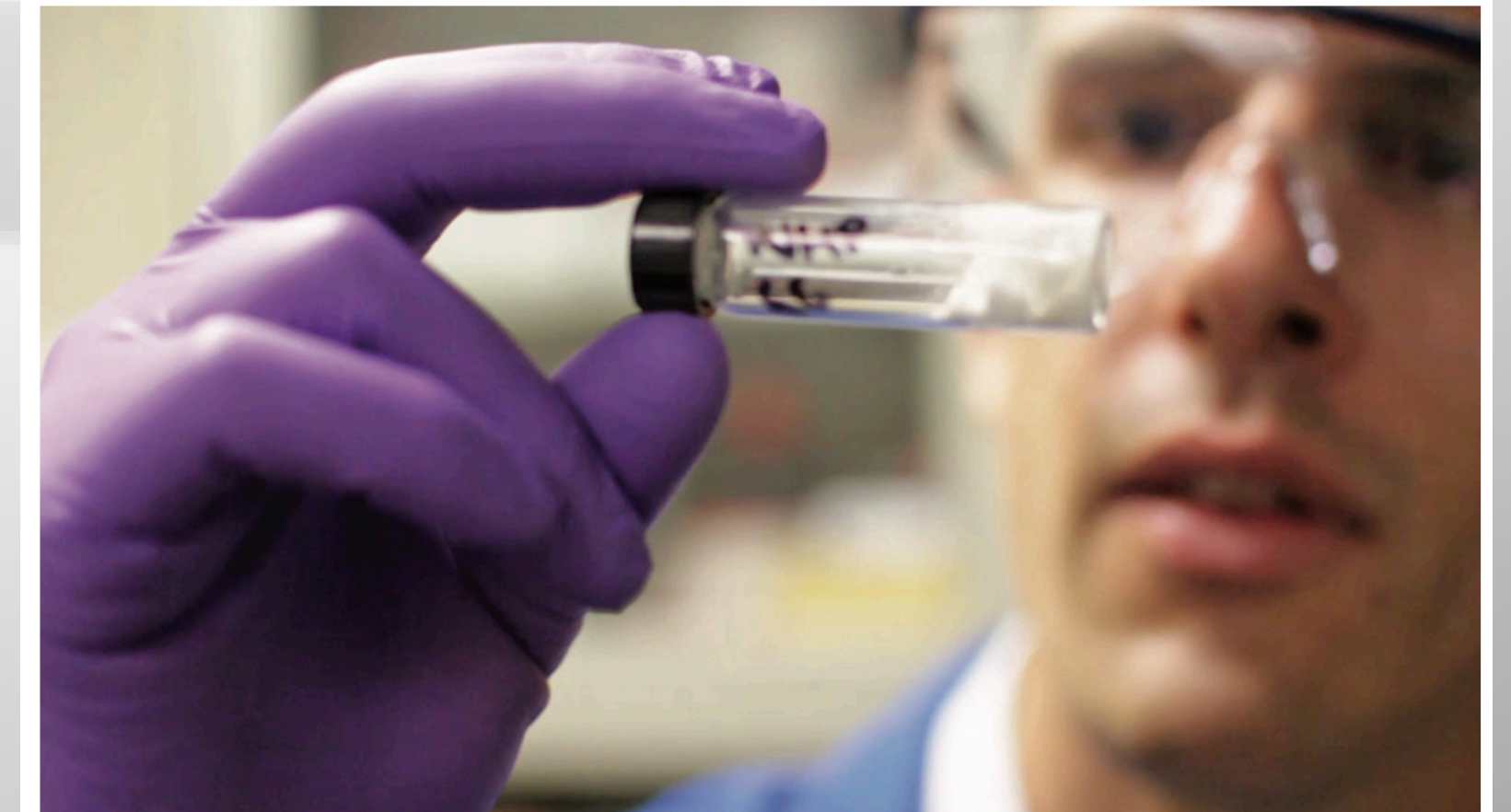
August 04, 2017 02:00 AM

IBM researchers create new family of fluoropolymers

MICHAEL LAUZON

TWEET SHARE SHARE EMAIL

PRINT



IBM
IBM scientists have combined typical "wet lab" chemical experiments with quantum chemistry tools to create a new family of plastics. The novel fluorine-based plastics, or fluoropolymers, have several potential applications, including uses in advanced electronics and medical devices.

from fundamental understanding to the development of new molecular systems



TECHNOLOGY | OTHER VOICES

3 Technologies That Could Create Trillion-Dollar Markets Over the Next Decade



By Greg Satell Updated April 21, 2019 / Original February 17, 2019

Order Reprints 📖 Print Article 🖨️



Photograph by Mario Tama/Getty Images

The next trillion-dollar markets are all about *manipulating matter*

www.barrons.com/articles/the-trillion-dollar-potential-in...

BARRON'S

Interests Magazine Data Advisor Penta

TECHNOLOGY | OTHER VOICES

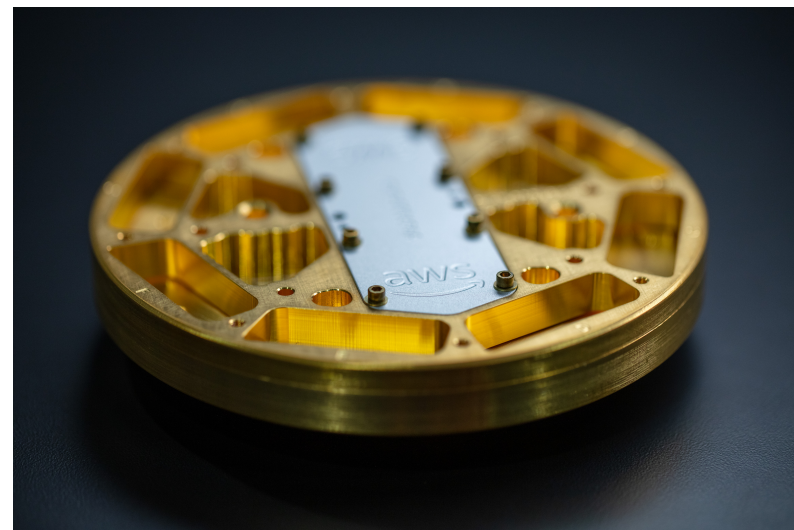
3 Technologies That Could Create Trillion-Dollar Markets Over the Next Decade



February 17, 2019

Order Reprints Print Article

post-digital computing
(e.g., noiseless quantum computers)



we collaborate with the
IBM Quantum group on
new Quantum Machine Learning
solutions for chemistry



The next trillion-dollar markets are all about *manipulating matter*

www.barrons.com/articles/the-trillion-dollar-potential-in...

BARRON'S

Interests Magazine Data Advisor Penta

TECHNOLOGY | OTHER VOICES

3 Technologies That Could Create Trillion-Dollar Markets Over the Next Decade



February 17, 2020

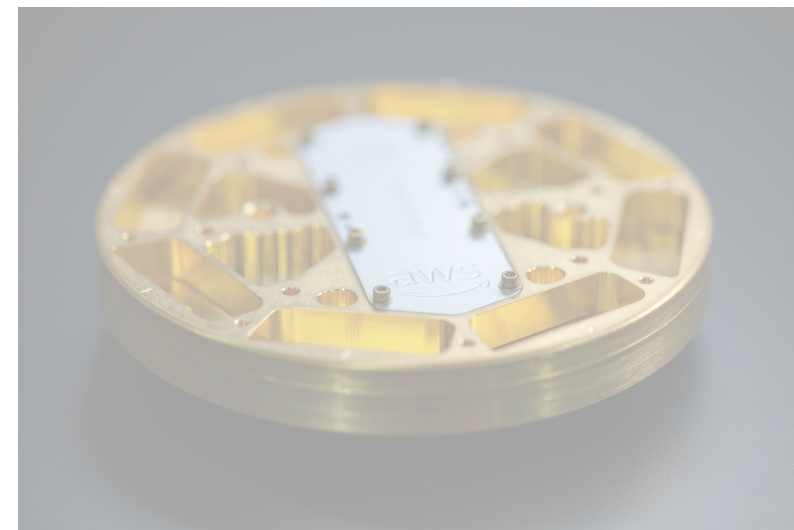
Reprints Print Article

post-digital computing

(e.g., noiseless quantum computers)

precise gene editing

(e.g., CRISPR)



we collaborate with the IBM Quantum group on new Quantum Machine Learning solutions for chemistry

The Nobel Prize in Chemistry 2020



© Nobel Prize Outreach. Photo: Bernhard Ludewig
Emmanuelle Charpentier
Prize share: 1/2



© Nobel Prize Outreach. Photo: Brittany Hosea-Small
Jennifer A. Doudna
Prize share: 1/2

The Nobel Prize in Chemistry 2020 was awarded jointly to Emmanuelle Charpentier and Jennifer A. Doudna "for the development of a method for genome editing."

The next trillion-dollar markets are all about *manipulating matter*

www.barrons.com/articles/the-trillion-dollar-potential-in...

BARRON'S

Interests Magazine Data Advisor Penta

TECHNOLOGY | OTHER VOICES

3 Technologies That Could Create Trillion-Dollar Markets Over the Next Decade



post-digital computing

(e.g., noiseless quantum computers)

February 17, 2020

precise gene editing

(e.g., CRISPR)

Reprints Print Article

materials science

(e.g., everything*)



we collaborate with the IBM Quantum group on new Quantum Machine Learning solutions for chemistry

© Nobel Prize Outreach. Photo: Bernhard Ludewig
Emmanuelle Charpentier
Prize share: 1/2

© Nobel Prize Outreach. Photo: Brittany Hosea-Small
Jennifer A. Doudna
Prize share: 1/2

The Nobel Prize in Chemistry 2020 was awarded jointly to Emmanuelle Charpentier and Jennifer A. Doudna "for the development of a method for genome editing."

*: *materials, drugs, energy, transistors, devices...*

Let's *invent* the technologies of *the future*

*it takes ~10 years and
>\$10M to develop
a new material
materials science
(e.g., everything*)*



Amazon Quantum Solutions Lab

Collaborate with experts to accelerate the development of quantum solutions

© Nobel Prize Outreach. Photo:
Bernhard Ludewig
Emmanuelle
Charpentier
Prize share: 1/2

© Nobel Prize Outreach. Photo:
Brittany Hosea-Small
Jennifer A. Doudna
Prize share: 1/2

The Nobel Prize in Chemistry 2020 was awarded jointly to Emmanuelle Charpentier and Jennifer A. Doudna "for the development of a method for genome editing."

**: materials, drugs, energy, transistors, devices...*

how do we make materials*?

materials science
(e.g., everything*)



Amazon Quantum Solutions Lab

Collaborate with experts to accelerate the development of quantum solutions

© Nobel Prize Outreach. Photo:
Bernhard Ludwig
Emmanuelle
Charpentier
Prize share: 1/2

© Nobel Prize Outreach. Photo:
Brittany Hosea-Small
Jennifer A. Doudna
Prize share: 1/2

The Nobel Prize in Chemistry 2020 was awarded jointly to Emmanuelle Charpentier and Jennifer A. Doudna "for the development of a method for genome editing."

**: materials, drugs, energy, transistors, devices...*

Let's invent the technologies of the future

*how do we
make materials*?*

*how do we
make molecules?*

Amazon

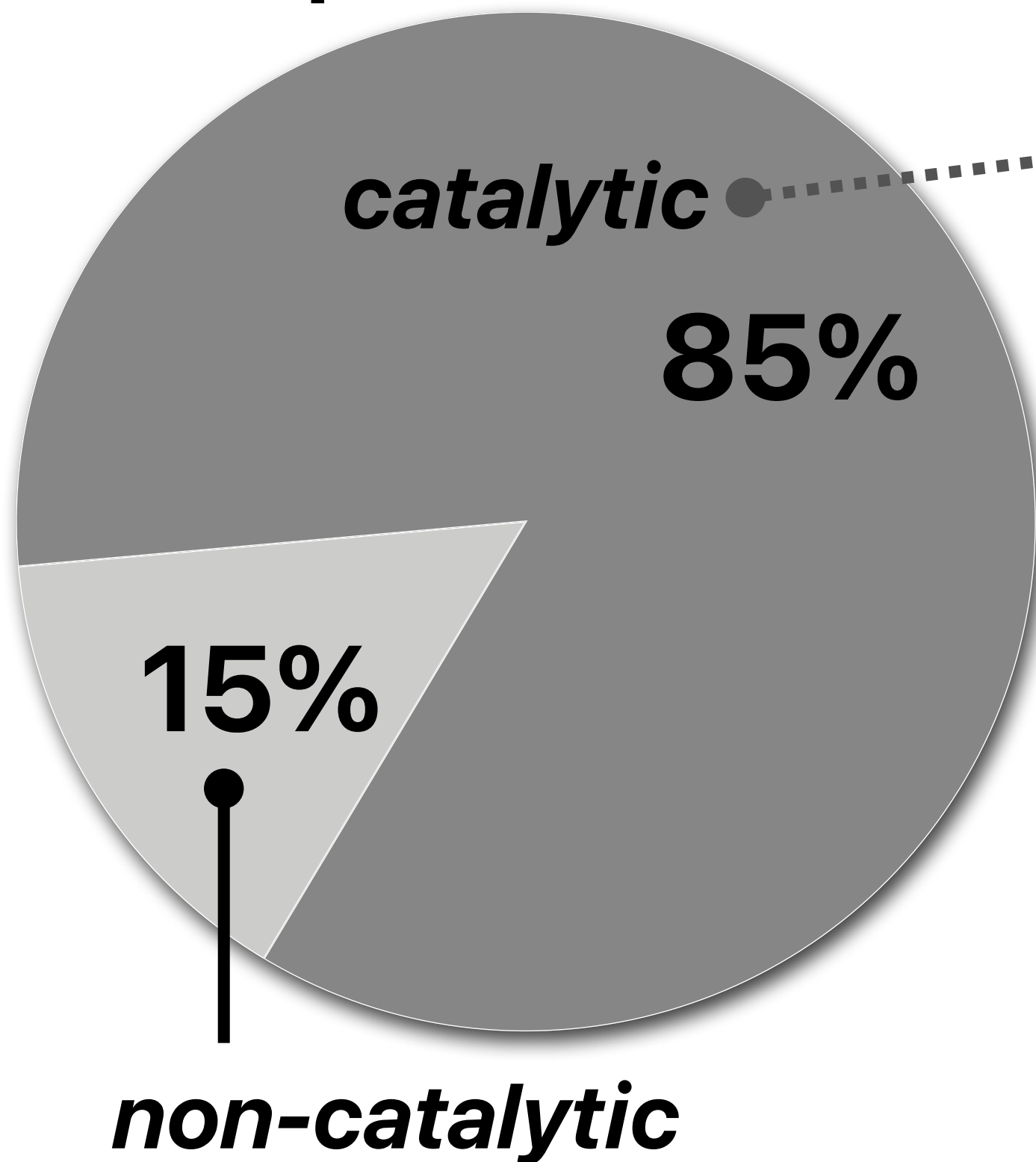
Collaborate with experts to accelerate the development of quantum solutions

Doudna "for the development of a method for genome editing."

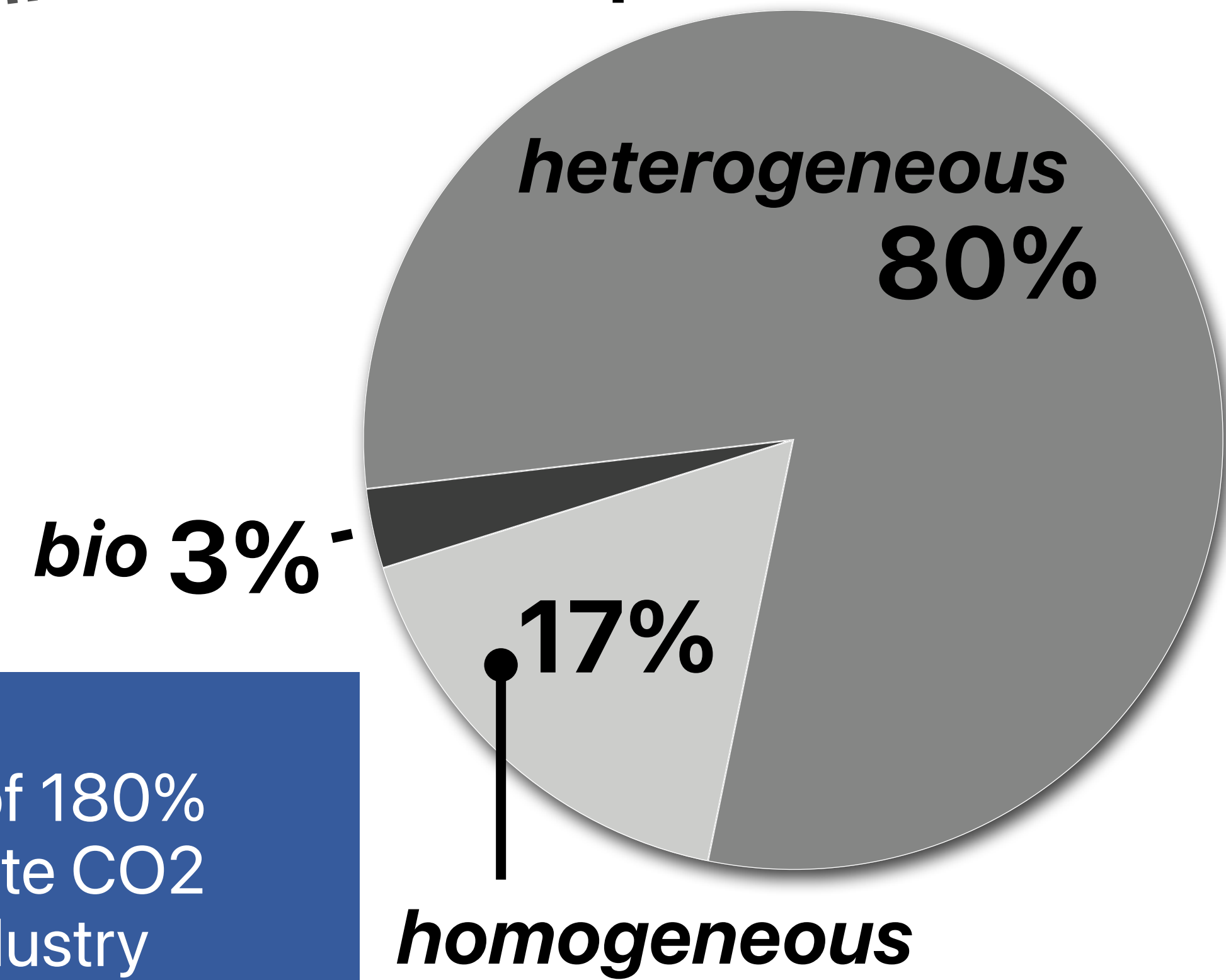
**: materials, drugs, energy, transistors, devices...*

Catalysis is a massive part of the chemical enterprise

all industrial processes



catalytic processes



by 2050 [2]:

- projected increase in demand of 180%
- a total 30% reduction of absolute CO₂ emissions in the chemical industry

- ~25% of the **global human energy consumption** is used for producing chemicals [1]
- **chemical industry** accounts for ~7% of the **global anthropogenic greenhouse gas emissions** [2]

[1]. Thomas J.M. "Summarizing comments on the discussion and a prospectus for urgent future action." *Philos. Trans. R. Soc. Lond. A.* **2016**, 374, 20150226

[2]. Levi P.G.; Cullen J.M. "Mapping global flows of chemicals: from fossil fuel feedstocks to chemical products." *Environ. Sci. Technol.* **2018**, 52, 1725

[3]. Bhaduri S.; Mukesh D. "Chemical industry and homogeneous catalysis." *Homogeneous Catalysis.* John Wiley & Sons, **2014**, 1-21

[4]. National Research Council "Impact of Advances in Computing and Communications Technologies on Chemical Science and Technology: Report of a Workshop." *National Academies Press*, **1999**

all industrial
processes

catalytic
processes

what do we need?

- to **develop new catalysts** for better industrial processes [3] 👍
- from conception to the discovery of new catalytic reactions, it can take several months to years [4] 👎👎👎

- ~25% of the **global human energy consumption** is used for producing chemicals [1]
- **chemical industry** accounts for ~7% of the **global anthropogenic greenhouse gas emissions** [2]

[1]. Thomas J.M. "Summarizing comments on the discussion and a prospectus for urgent future action." *Philos. Trans. R. Soc. Lond. A.* **2016**, 374, 20150226

[2]. Levi P.G.; Cullen J.M. "Mapping global flows of chemicals: from fossil fuel feedstocks to chemical products." *Environ. Sci. Technol.* **2018**, 52, 1725

[3]. Bhaduri S.; Mukesh D. "Chemical industry and homogeneous catalysis." *Homogeneous Catalysis. John Wiley & Sons*, **2014**, 1-21

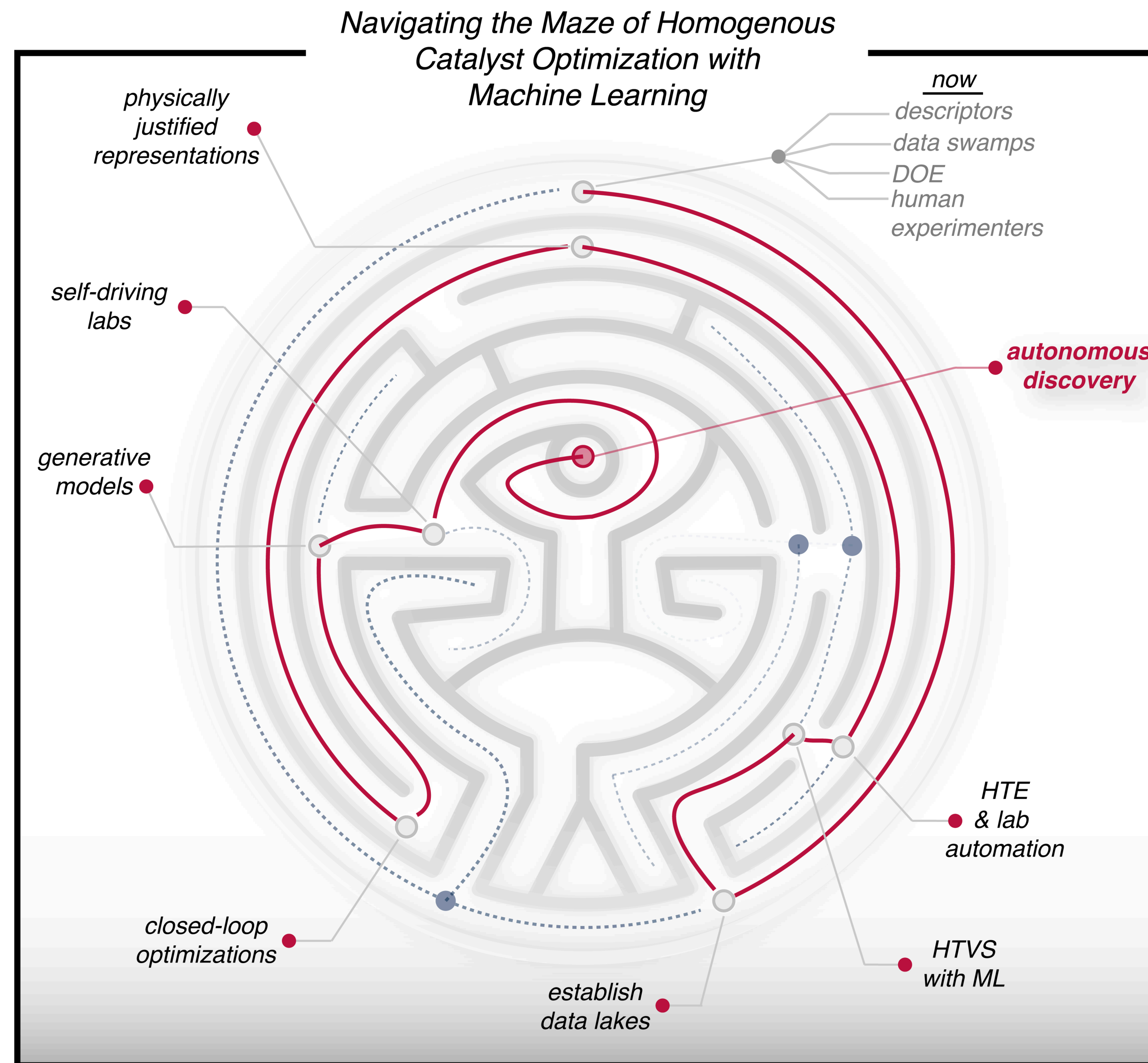
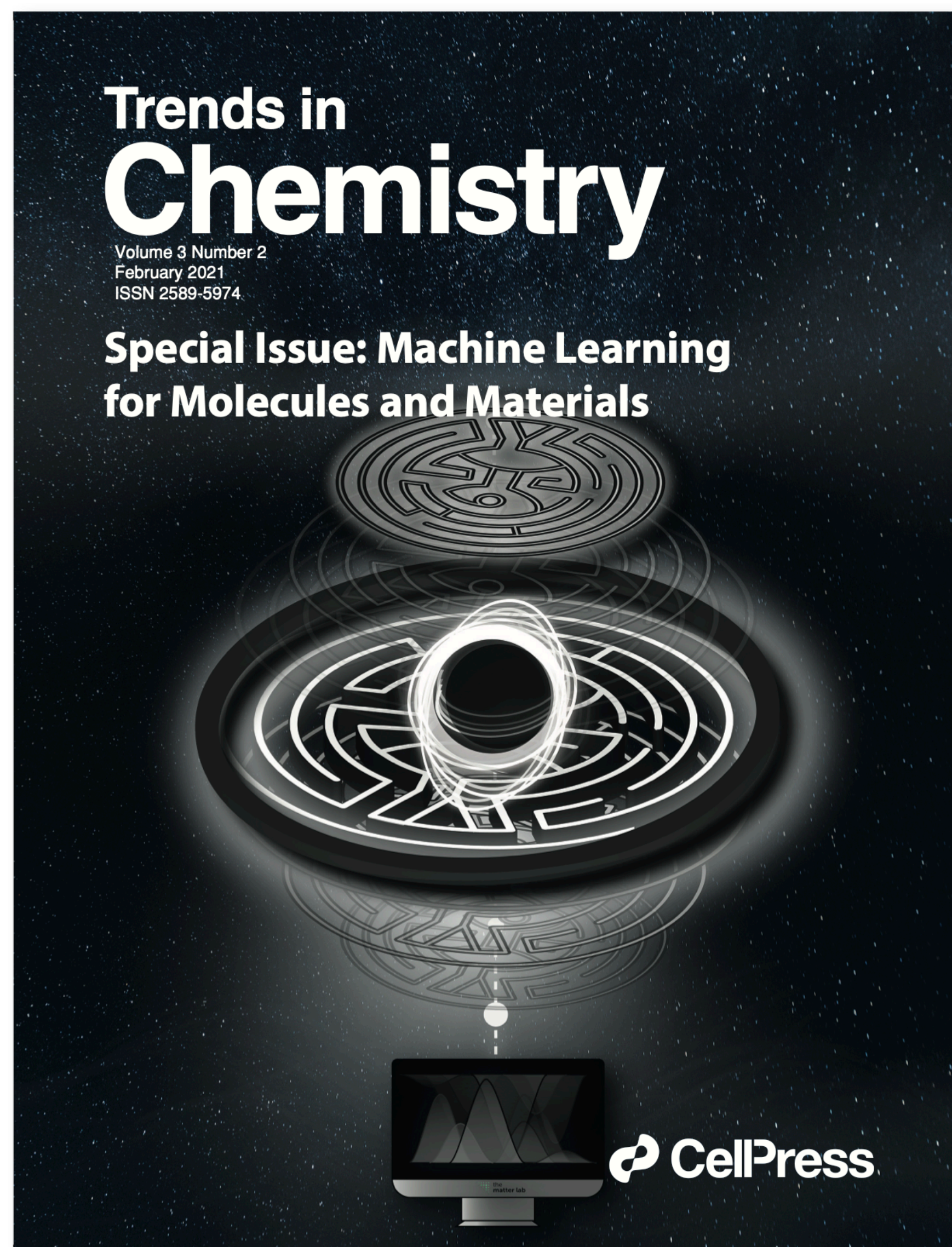
[4]. National Research Council "Impact of Advances in Computing and Communications Technologies on Chemical Science and Technology: Report of a Workshop." *National Academies Press*, **1999**

*computational sciences,
AI, and automation are
helping us accelerate
these timelines...*

*but **not fast enough!***

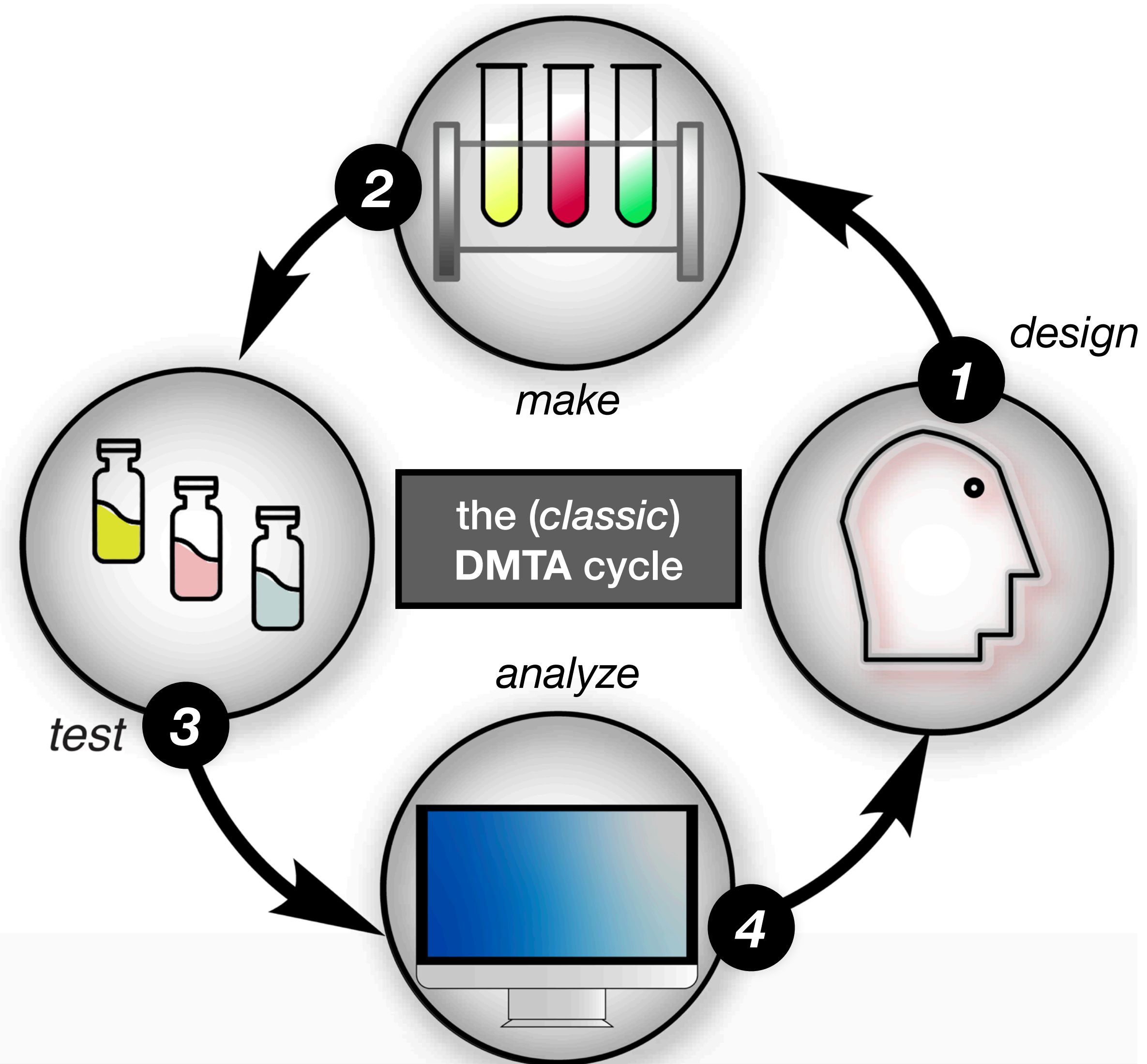
[1]. Th
[2]. Le
[3]. Bh

designing catalysts with ML: not a straight path



(almost) all technologies that humanity created followed this path

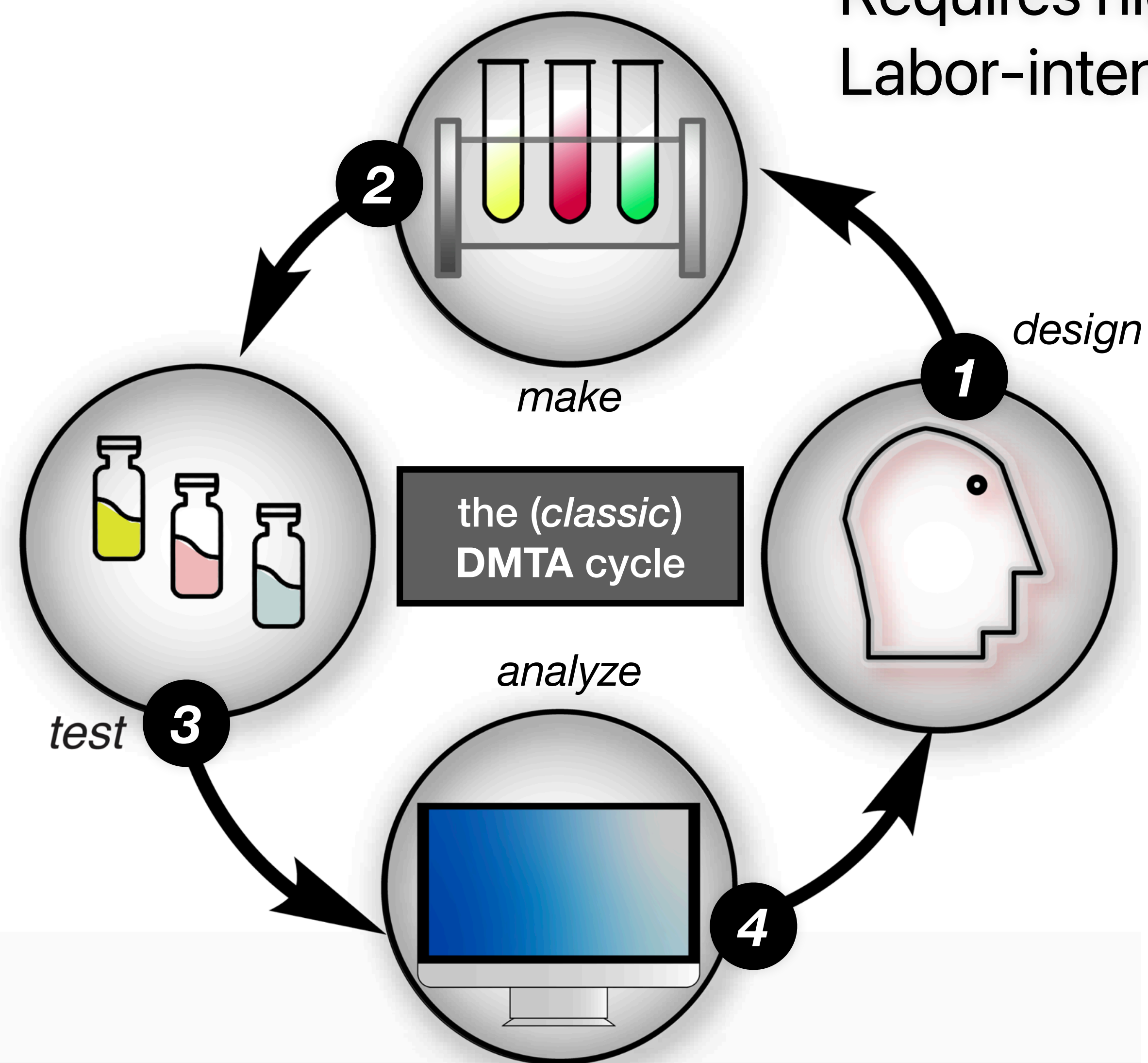
Centuries-old approach
to scientific research



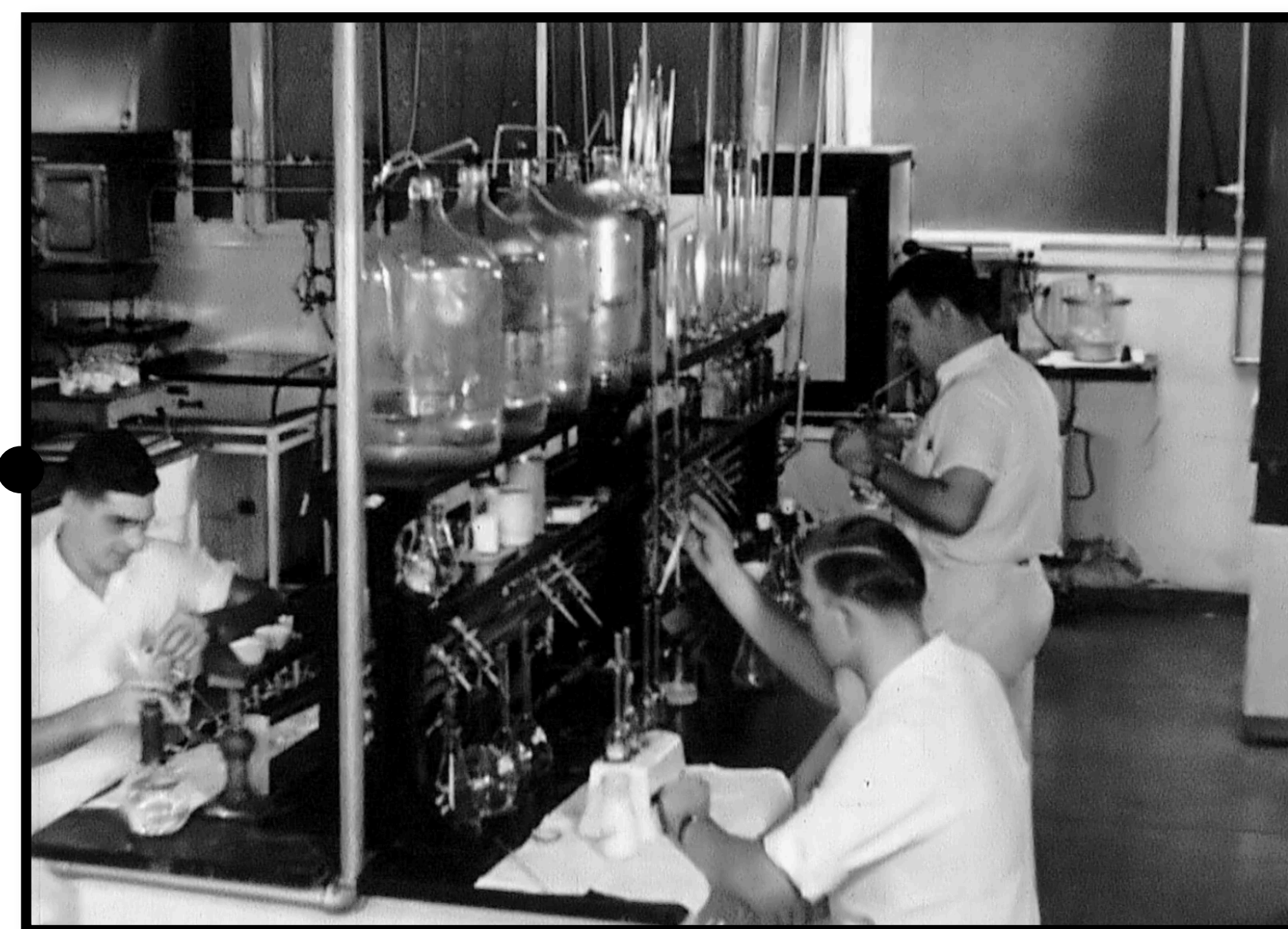
(almost) all technologies that humanity created followed this path

Centuries-old approach
to scientific research

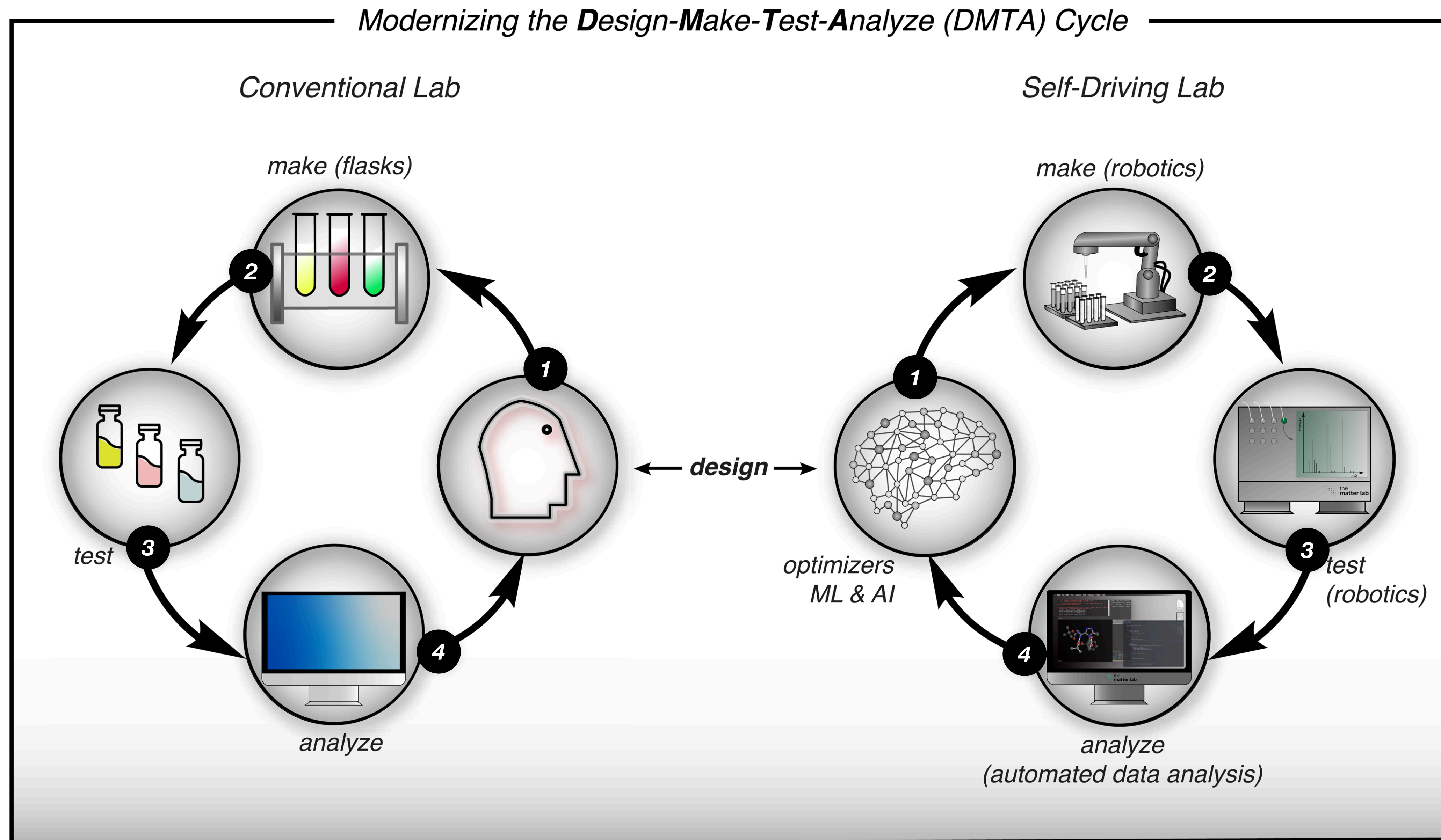
Extremely time-consuming
Requires high qualification
Labor-intensive



over 100
years



towards self-driving labs

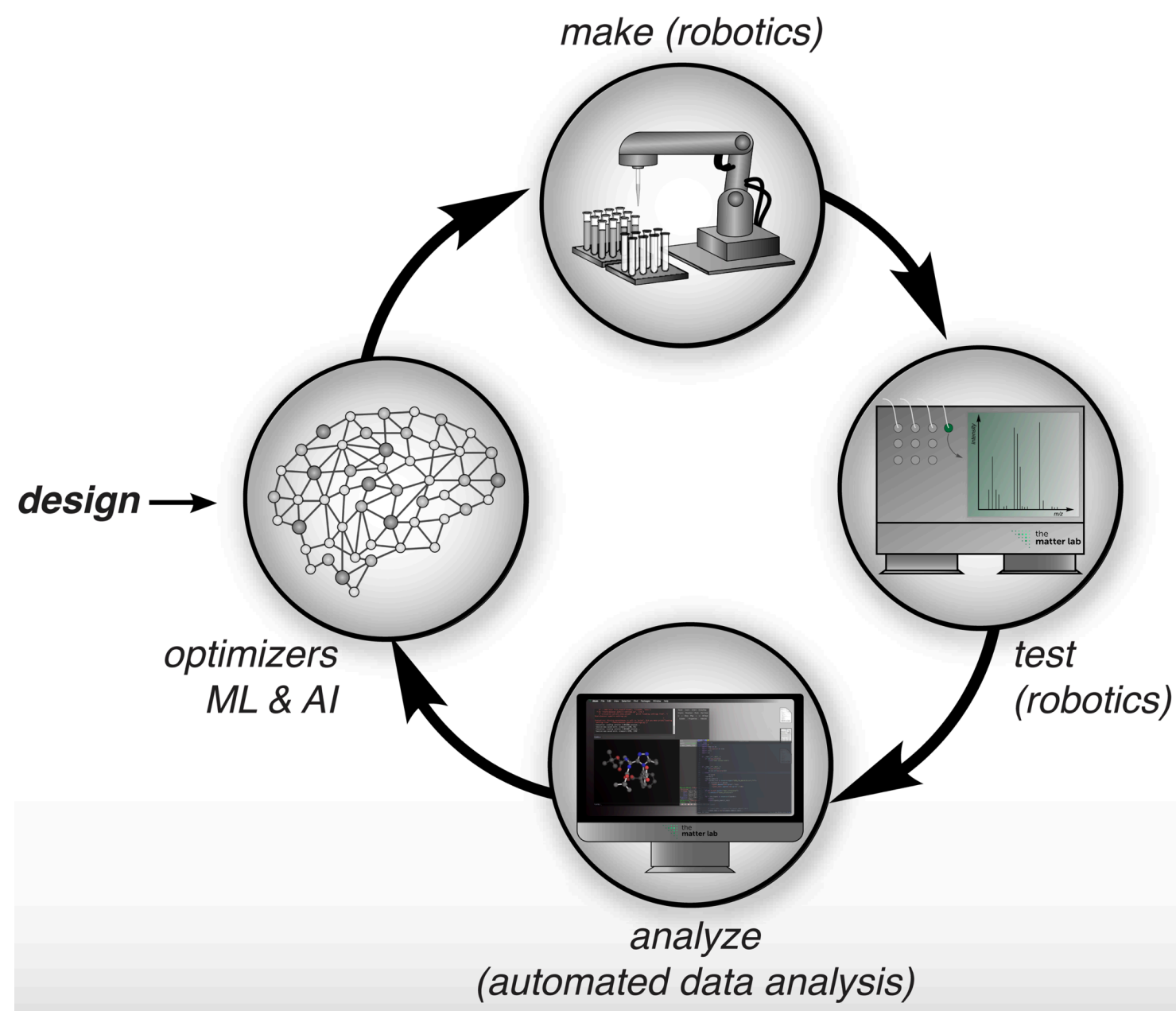


The CMU Cloud Lab will open its doors later this year

- Central code-based software platform
- Automated instrumentation + technicians
- Everything traceable
- >200 instrument types
- Synthesis, purification, experimentation, characterization
- Based on existing ECL facility

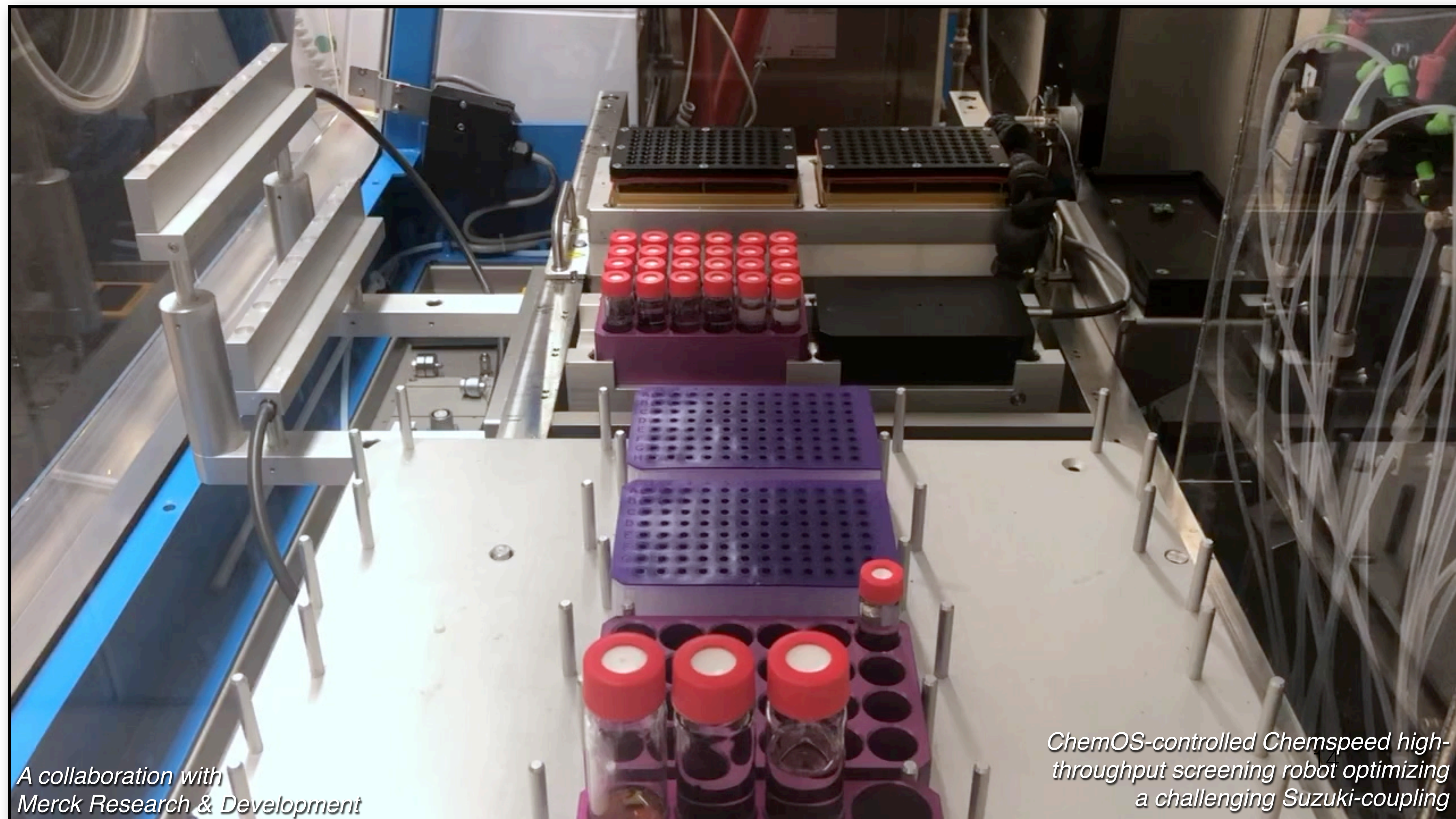
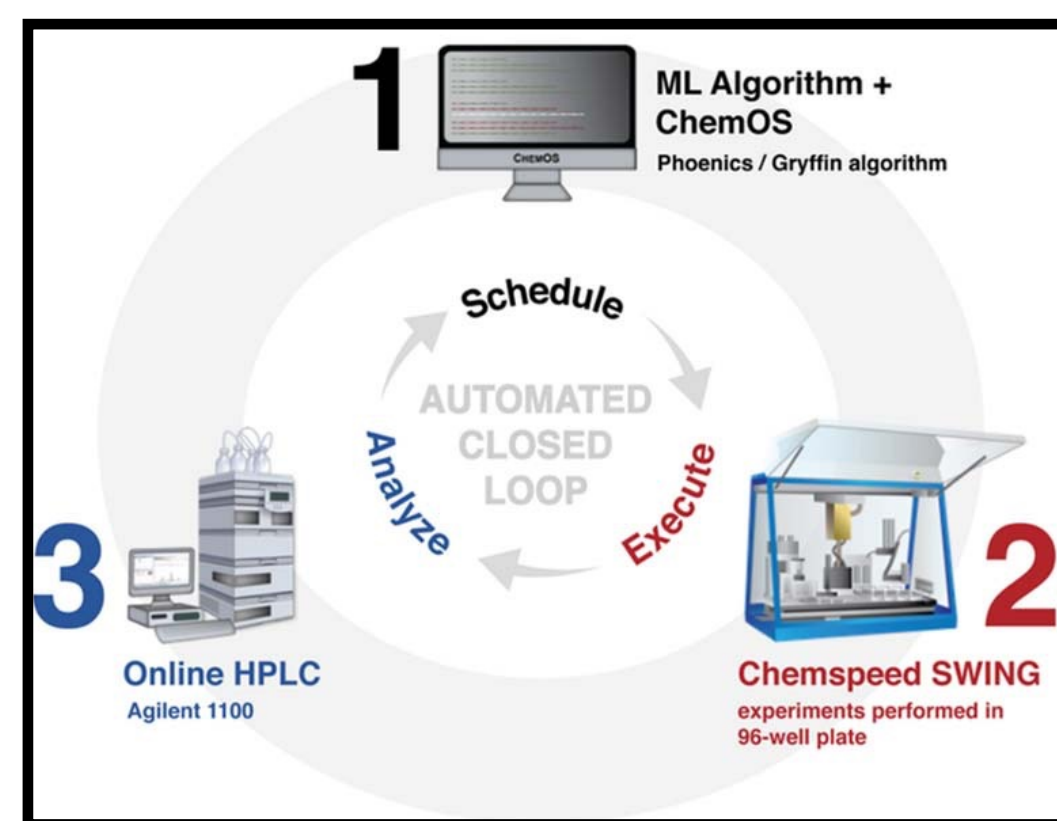
for some areas, we got a little faster at it

Experiment automation techniques



- Less labor intensive
- Still requires experiments to be designed by human

for details, see: • “Data-science driven autonomous process optimization” *Communications Chemistry*, **2021**, 4, 112
• “ChemOS: Orchestrating autonomous experimentation” *Science Robotics*, **2018**, 3 (19), eaat5559



A collaboration with
Merck Research & Development

ChemOS-controlled Chemspeed high-throughput screening robot optimizing a challenging Suzuki-coupling

for some areas, we got a little faster at it

Experiment automation ■ Less labor intensive

how can we make it faster?

how can we make it easier?

more accessible?



*A collaboration with
Merck Research & Development*

ChemOS-controlled Chemspeed high-throughput screening robot optimizing a challenging Suzuki-coupling

*how can we make it **faster?** **easier?** more **accessible?***

Centuries-old approach to scientific research

Extremely time-consuming
Requires high qualification
Labor-intensive

Experiment automation techniques

- Less labor intensive
- Still requires experiments to be designed by human

- Requires only natural language prompts
- Seamlessly integrated with experiment automation

towards autonomous scientific research

Centuries-old approach
to scientific research

Extremely time-consuming
Requires high qualification
Labor-intensive

Experiment automation
techniques

- Less labor intensive
- Still requires experiments to be designed by human

- Requires only natural language prompts
- Seamlessly integrated with experiment automation

**Prompt in
natural language**

[bits]

Generalist Intelligent Agent

powered by multiple large language models

***Experiments realized
in the physical world***

[atoms]

Coscientist's websearch capabilities

Input prompt from scientist

Generalist Intelligent Agents System

“Synthesize Ibuprofen”

task initialization

Planner

Coscientist's websearch capabilities

Input prompt from scientist

Generalist Intelligent Agents System

“Synthesize Ibuprofen”

task initialization

Planner

Planner

GOOGLE request

Web searcher

task-relevant knowledge

Coscientist's websearch capabilities

Input prompt from scientist

Generalist Intelligent Agents System

“Synthesize Ibuprofen”

task initialization

Planner

Planner

GOOGLE request

Web searcher

task-relevant knowledge

GOOGLE Friedel-Crafts acylation isobutylbenzene and acetic anhydride conditions and stoichiometry

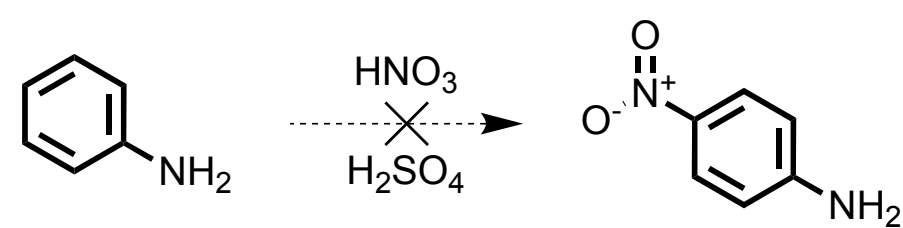
Planner LLM identifies Friedel-Crafts reaction, requests additional information

Web searcher LLM finds information from the internet by browsing and returns a summary to the Planner

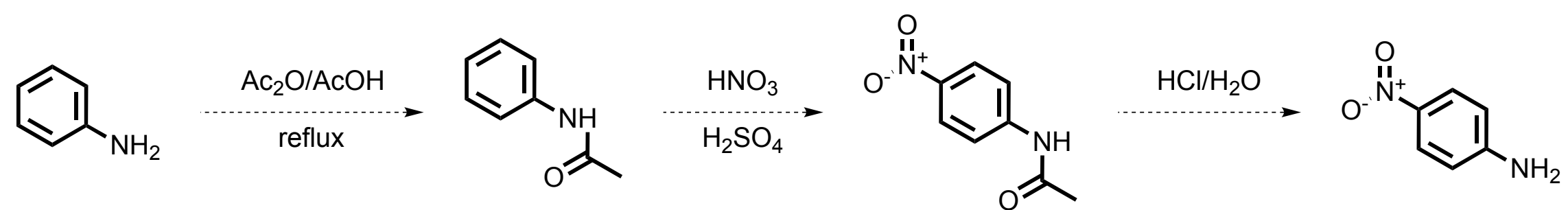
In the Friedel-Crafts acylation of isobutylbenzene and acetic anhydride...

(out-of-the-box) LLMs can be ok at planning reactions

b Wrong synthesis, but makes sense (GPT-3.5, no search)

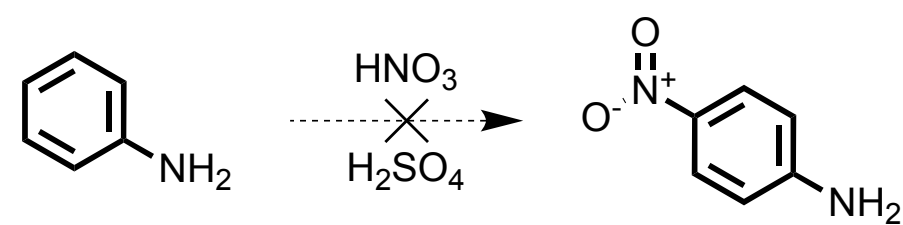


Correct synthesis (GPT-4 with search)

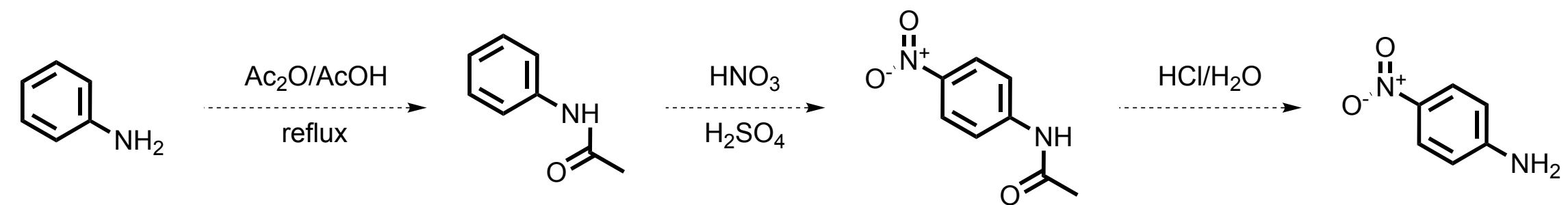


(out-of-the-box) LLMs can be ok at planning reactions

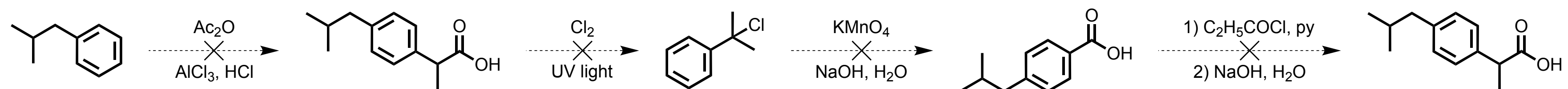
b Wrong synthesis, but makes sense (GPT-3.5, no search)



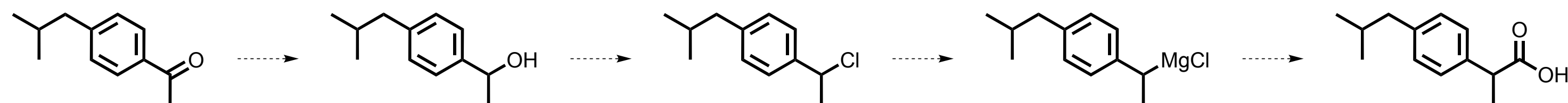
Correct synthesis (GPT-4 with search)



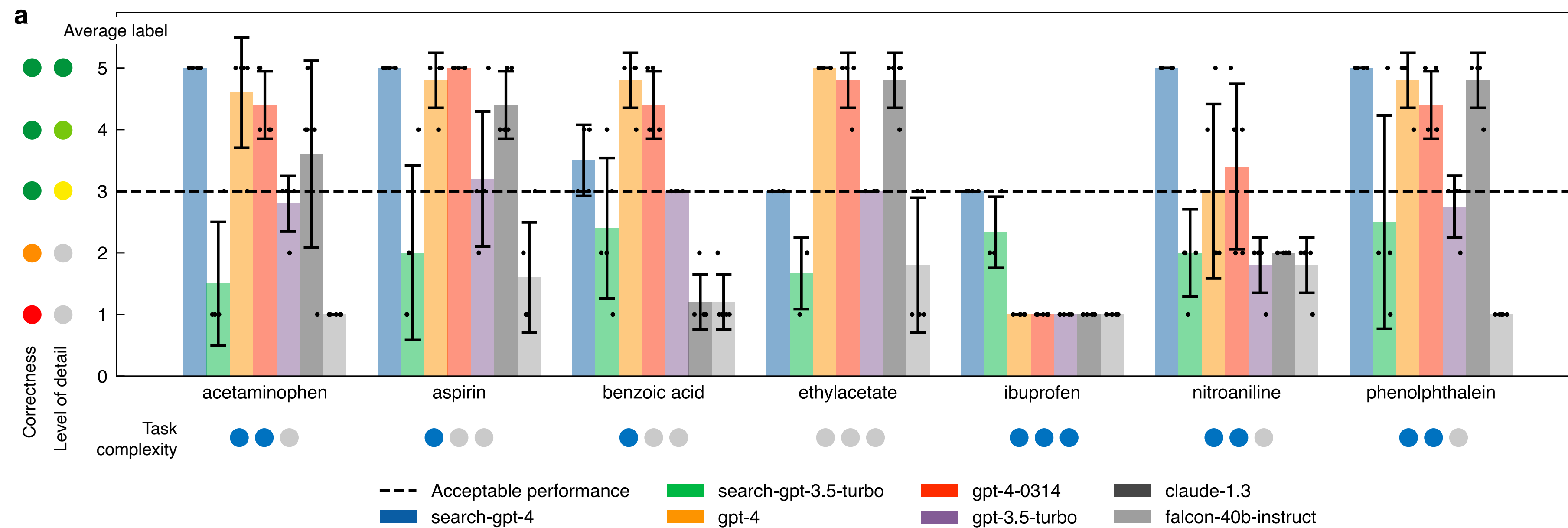
c Wrong synthesis, does not make sense (GPT-4, no search)



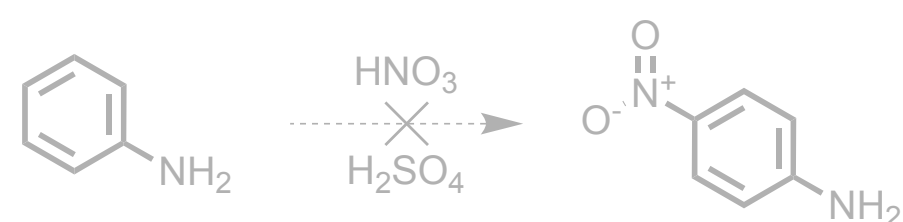
Correct synthesis (GPT-4 with search)



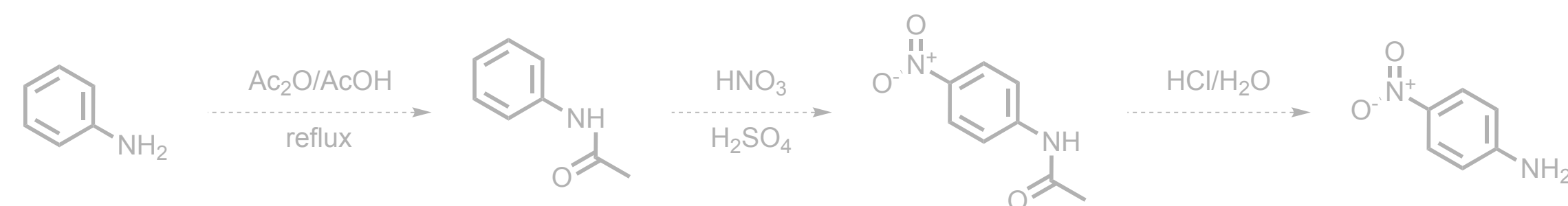
(out-of-the-box) LLMs can be ok at planning reactions



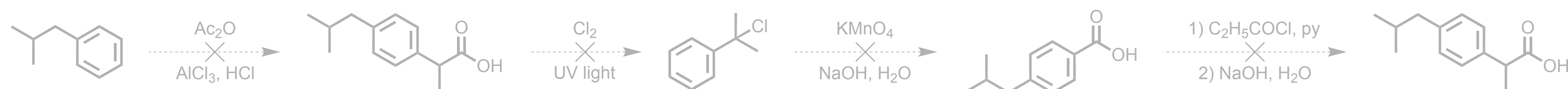
b Wrong synthesis, but makes sense (GPT-3.5, no search)



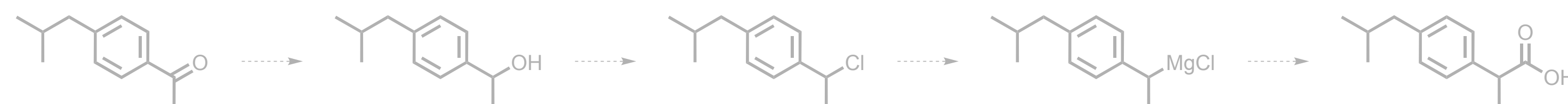
Correct synthesis (GPT-4 with search)



c Wrong synthesis, does not make sense (GPT-4, no search)



Correct synthesis (GPT-4 with search)



grounding Coscientist by leveraging documentation

Input prompt from scientist

Docs searcher

III. Opentrons Python API Implementation - Prompt-to-OT-2

Planner

API usage information / Prompt-to-OT-2

OT-2 information-flow back to Planner for downstream task

planner **requested** query

how do I use the heater-shaker module?

query
embedding



API docs
embedding



vectorized text with
ada embeddings

vector
search

relevant doc selection



'hardware modules'

further
investigation

```
# Heat and shake the reaction
hs_mod.set_target_temperature(75)
hs_mod.wait_for_temperature()
hs_mod.set_and_wait_for_shake_speed(500)

# Deactivate heater and shaker
hs_mod.deactivate_heater()
hs_mod.deactivate_shaker()
hs_mod.open_labware_latch()
```

proper usage of **heater-shaker** module

extracted from webpage

Table Of Contents	::	Welcome
		Tutorial
		Versioning
		...

grounding Coscientist by leveraging documentation

Input prompt from scientist

Docs searcher

II. Emerald Cloud Labs Implementation - Prompt-to-SLL



Planner

API usage information / Prompt-to-SLL

SLL information-flow back to Planner for downstream task

description
summarization



code
retention

planner **requested** queries

analyze a complex mixture
to see what is in it

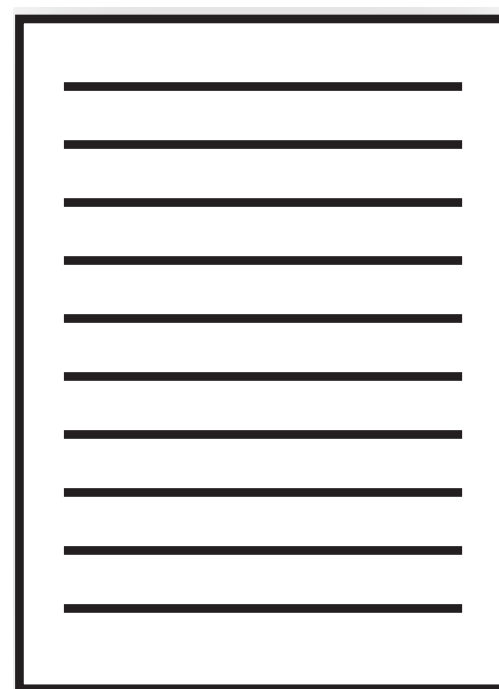
quantify proteins
in my sample

determine m/z ratio for
compounds in my sample

natural language prompts

+

Function Guide



function
selection

Prompt-to-Function

ExperimentHPLC,
ExperimentIonChromatography

ExperimentTotalProteinQuantification,
ExperimentTotalProteinDetection

ExperimentMassSpectrometry,
ExperimentGCMS

plain text
documentation
further
investigation

```
ExperimentHPLC[Samples] => Protocol  
Experimental Principles...  
Instrumentation...  
Experiment Options...  
Sample Parameters...  
...
```

**not from webpage*
required extensive work
from the folks at ECL*

Running Experiments

Collection of functions used to remotely conduct experiments in an ECL facility.

extracted
from webpage

a few examples of the Docs Searcher module in action

HPLC Experiment Query

```
searcher = DocumentationSearcher(agi_config, top_k=30, tok_limit=7000)
answer = searcher.read_docs(
    """
    Run an HPLC experiment to analyze Object[Sample, "id:4p06dM5aD5Xr"].
    Specify the instrument. Do not measure the volume or take an image
    of my sample.
    """
)
```

Docs Searcher Response (ECL SLL)

```
*could not add Instrumentation to usage text. Section tokens: 14966, Available tokens: 6909
*could not add Example Calls to usage text. Section tokens: 6814, Available tokens: 2610
*could not add Warnings and Errors to usage text. Section tokens: 173655, Available tokens: 1719
30 options, instruments, and settings added to message history for ExperimentHPLC
messages size: 5610 tokens

-----
ExperimentHPLC[Object[Sample, "id:4p06dM5aD5Xr"], Instrument -> Object[Instrument, HPLC], MeasureVolume -> False, Image -> False]
-----
```

OT-2 Heater-Shaker Module Query

```
ot2_searcher = DocumentationSearcher(agi_ot2_config, top_k=30, tok_limit=7000)
ot2_answer = ot2_searcher.read_docs(
    """
    Use the heater shaker module.
    """
)
```

Docs Searcher Response (OT-2 Python API)

```
6 documents added to message history
Section Names:
Using a Heater-Shaker Module
    Version 2.13
    Heating and Shaking
    ...
messages size: 3315 tokens
```

```
To use the heater shaker module in the OT-2, you'll first need to load the module
using `load_module`...
python
from opentrons import protocol_api
metadata = {
    'apiLevel': '2.13'
}
def run(protocol: protocol_api.ProtocolContext):
    # Load the heater shaker module in slot 1
    hs_mod = protocol.load_module('heaterShakerModuleV1', 1)
    ...
This code loads the heater shaker module and controls its temperature and shaking...
```

Mass Spec Experiment Query

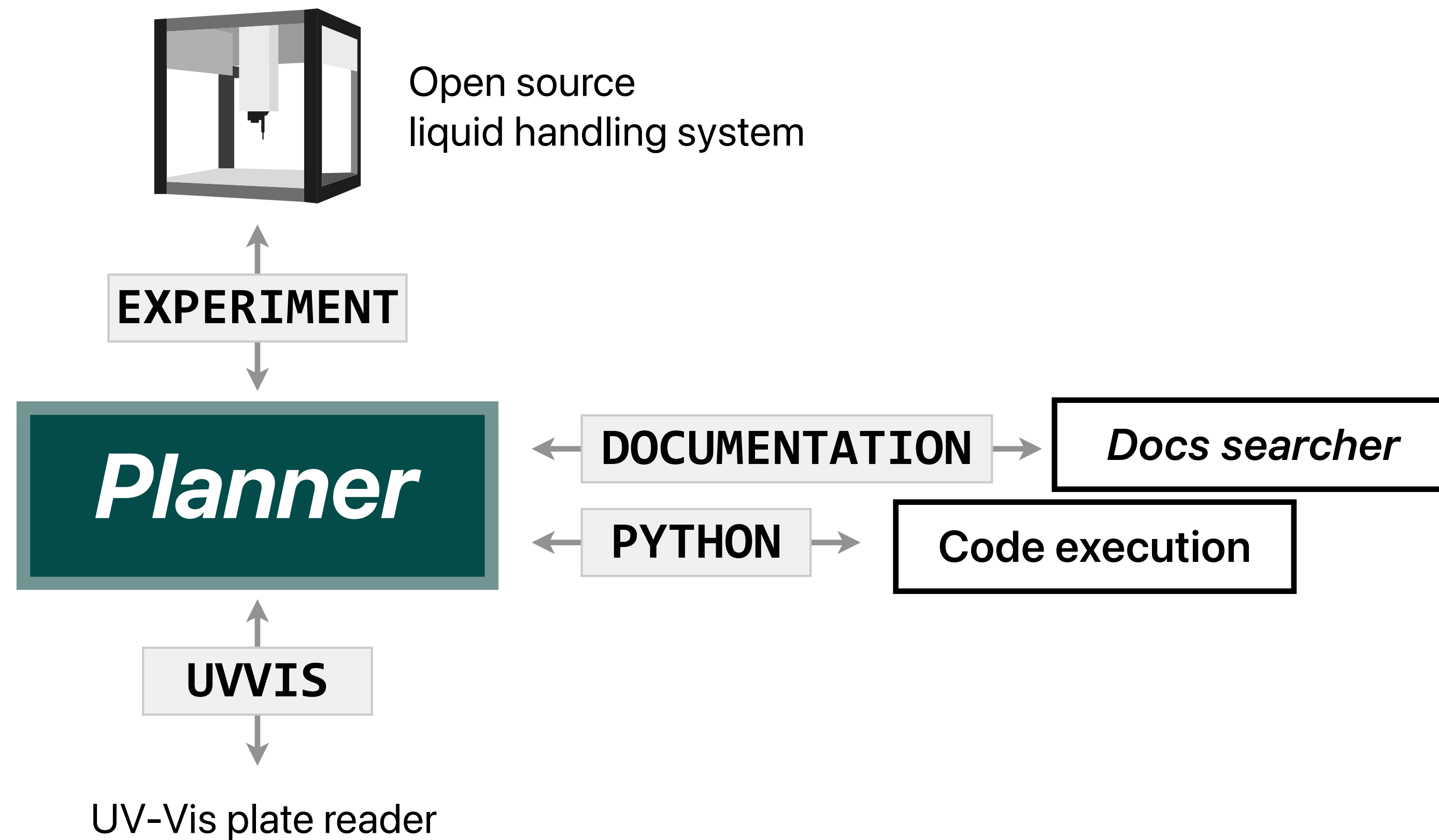
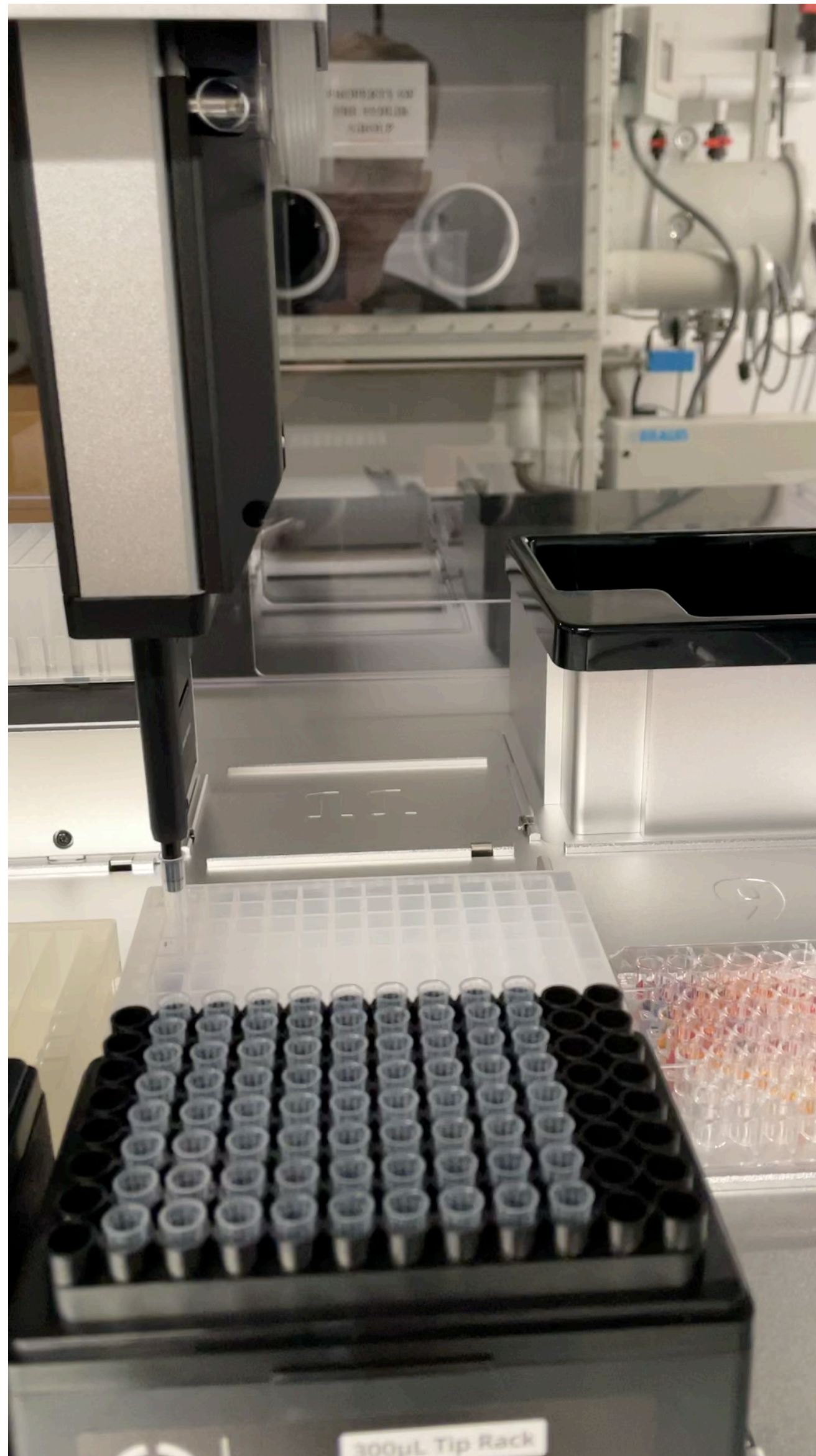
```
searcher = DocumentationSearcher(agi_config, top_k=30, tok_limit=7000)
answer = searcher.read_docs(
    """
    Run an Mass Spec experiment to analyze Object[Sample, "id:4p06dM5aD5Xr"].
    Use MALDI. Perform 3 replicates of the experiment. Use an acetonitrile in
    water buffer.
    """
)
```

Docs Searcher Response (ECL SLL)

```
*could not add Instrumentation to usage text. Section tokens: 5364, Available tokens: 7229
...
30 options, instruments, and settings added to message history for ExperimentMassSpectrometry
messages size: 6966 tokens

-----
ExperimentMassSpectrometry[Object[Sample, "id:4p06dM5aD5Xr"], IonSource -> MALDI, NumberOfReplicates -> 3,
    Buffer -> Model[Sample,StockSolution,"0.1% FA with 5% Acetonitrile in Water, LCMS-grade"]]
-----
```

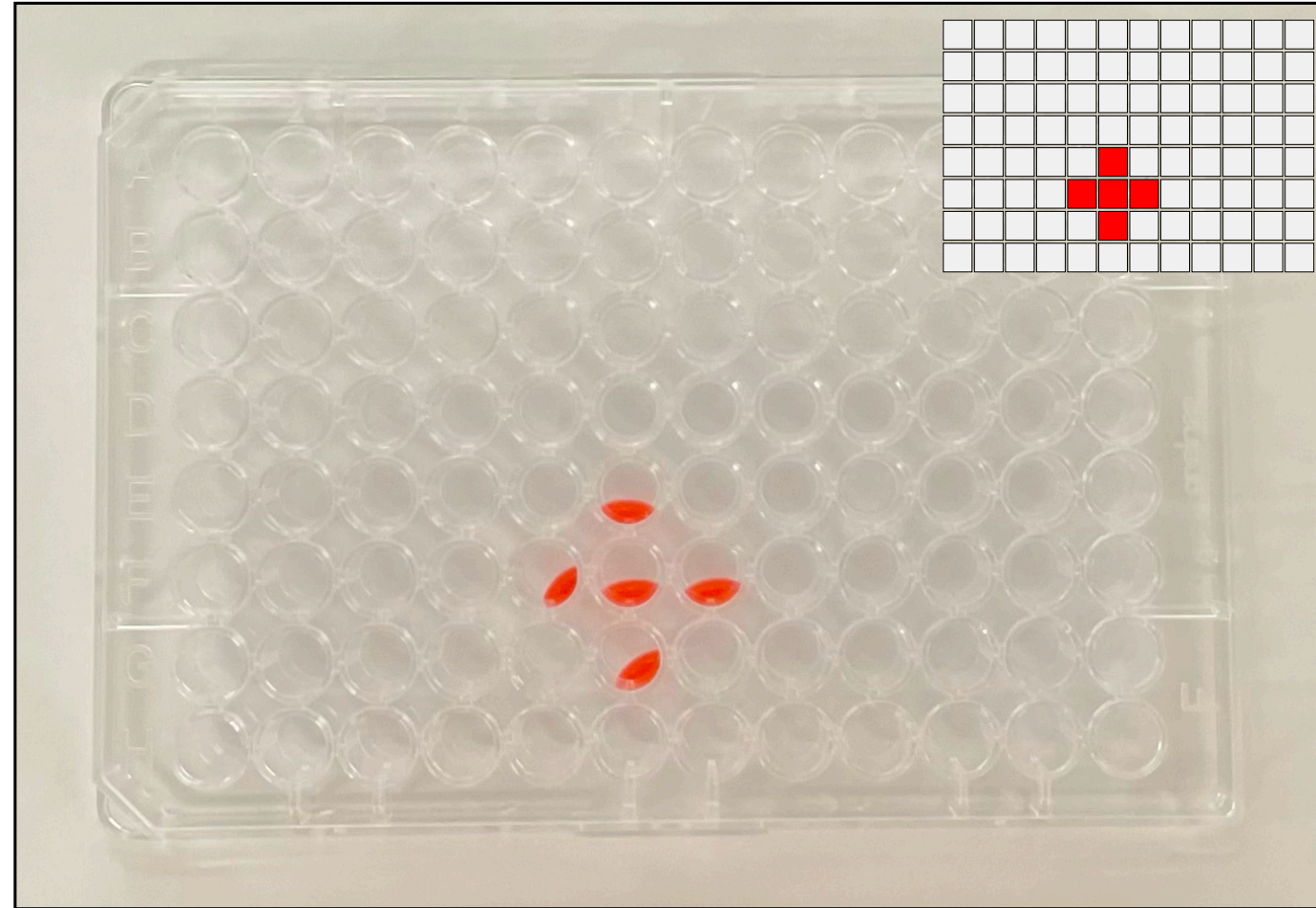

performing experiments in the physical world



controlling a liquid handler robot with low-level instructions

Draw a red cross using food coloring in the center of 96-well plate

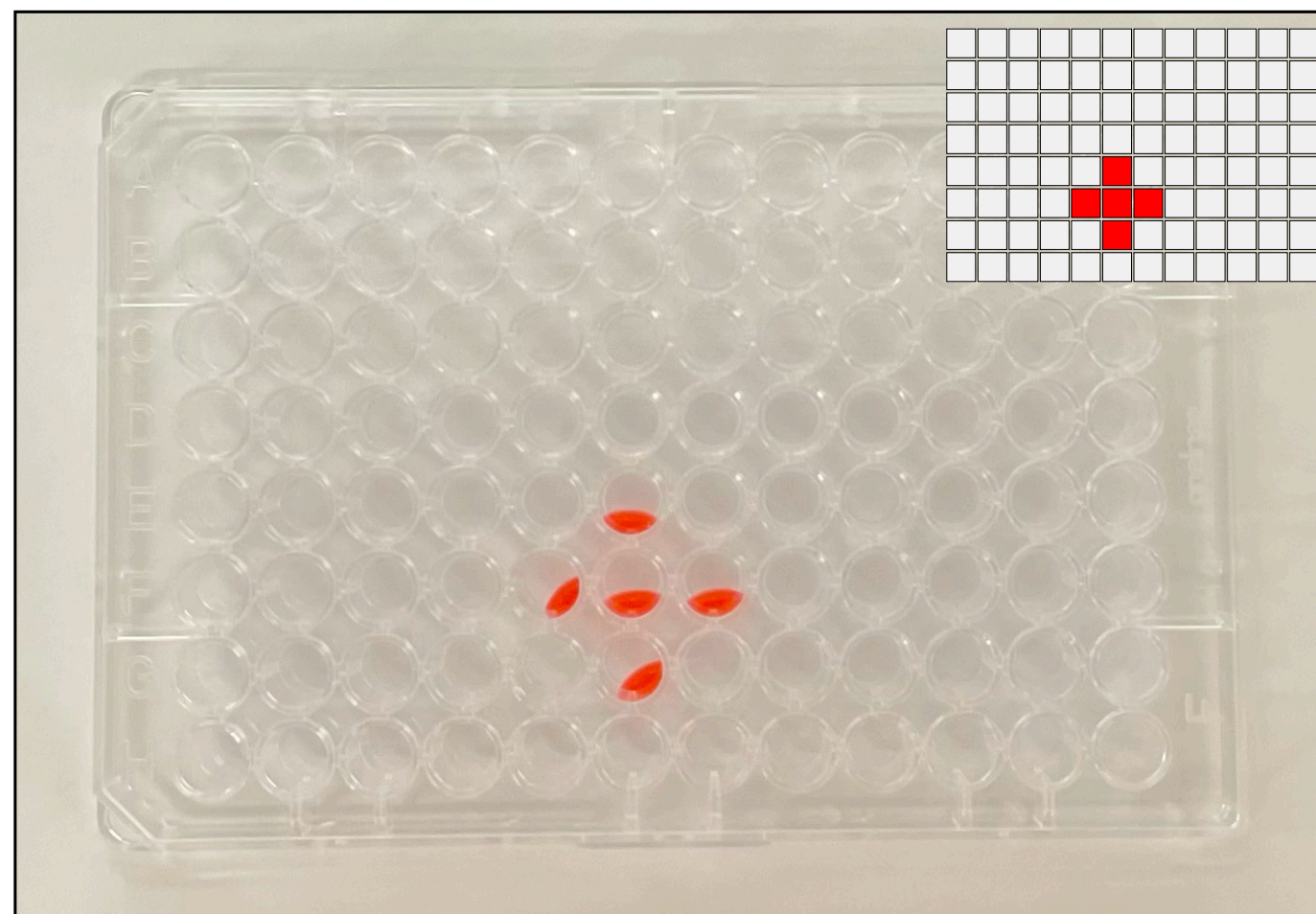
<setup description>



controlling a liquid handler robot with low-level instructions

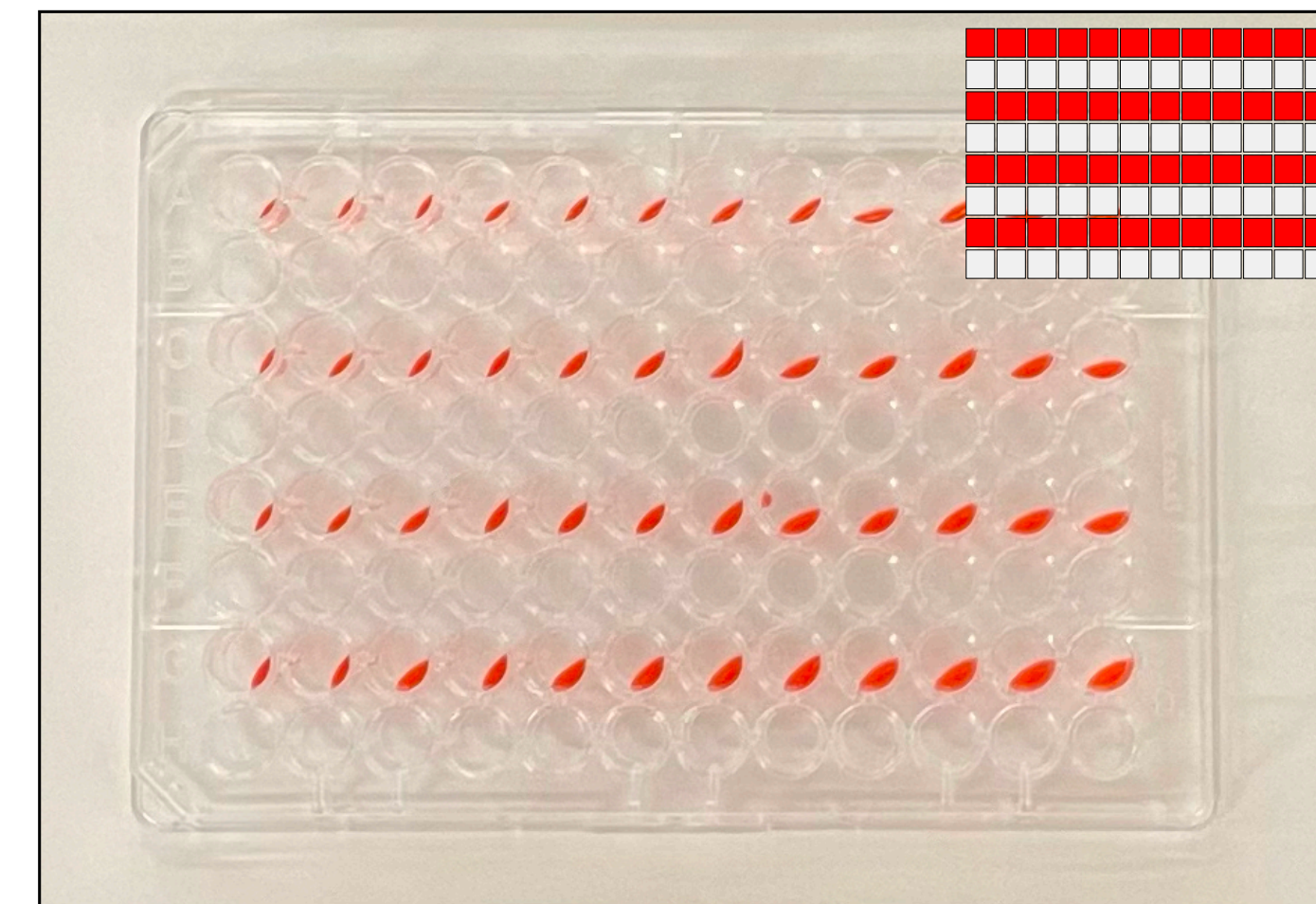
Draw a red cross using food coloring in the center of 96-well plate

<setup description>



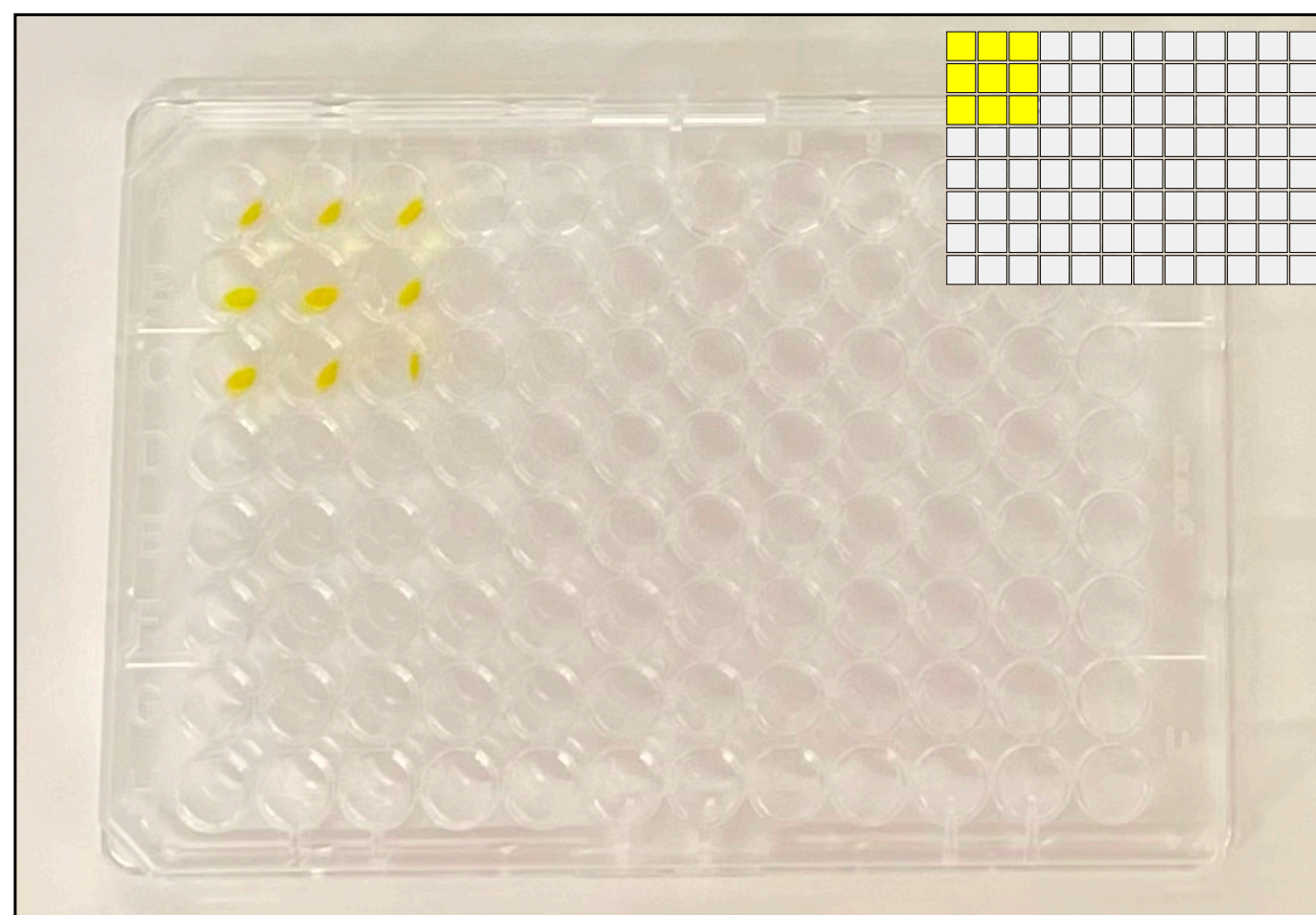
Color every other row of a 96-well plate with one color of your choice. Remember, that for me to see it you should put at least 10 μL .

<setup description>



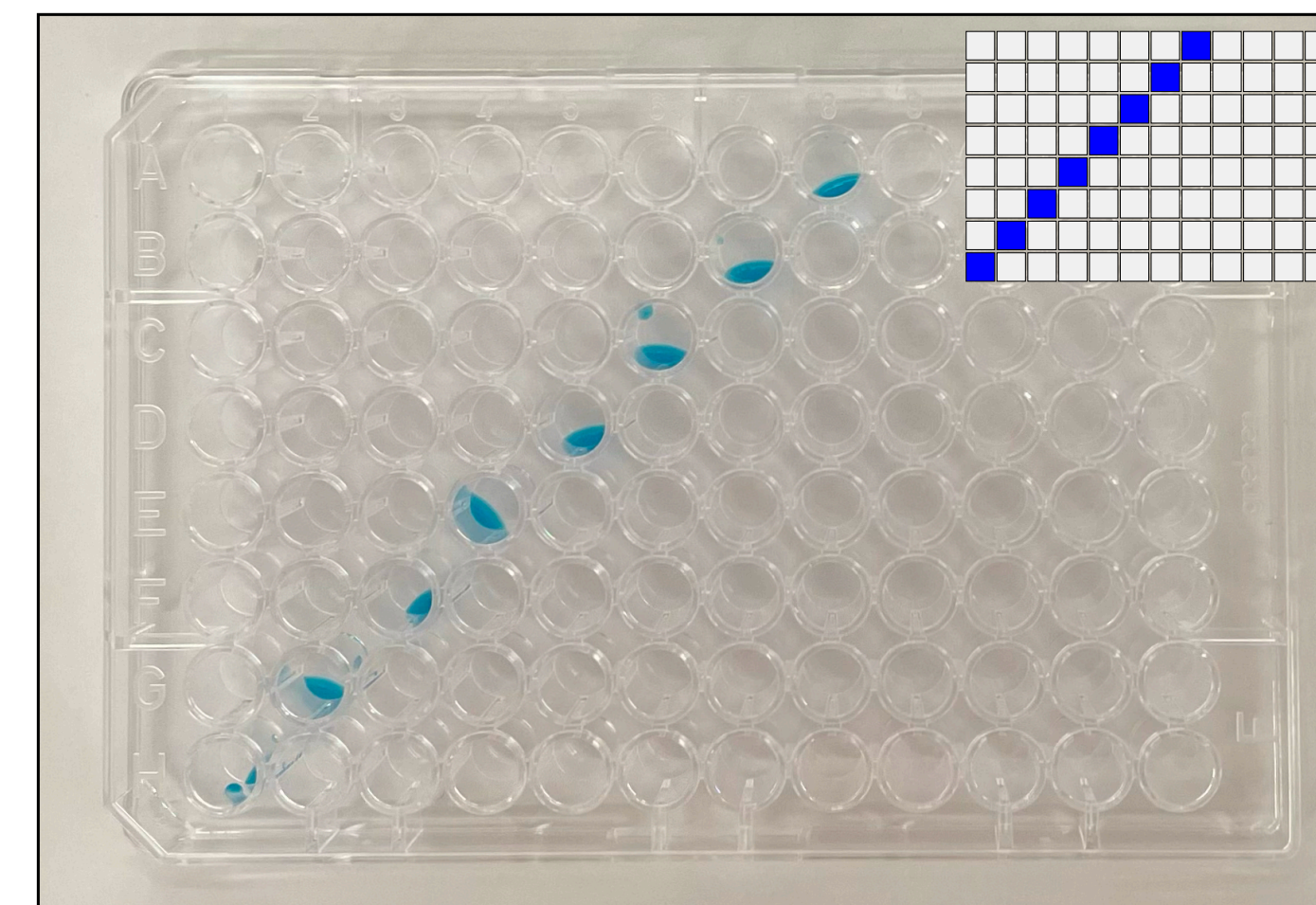
Draw a 3x3 rectangle using yellow color at upper left part of the 96-well plate. Remember, that for me to see it you should put at least 10 μL .

<setup description>



Draw a blue diagonal starting from lower left (H1) in the 96-well plate. Remember, that for me to see it you should put at least 10 μL .

<setup description>

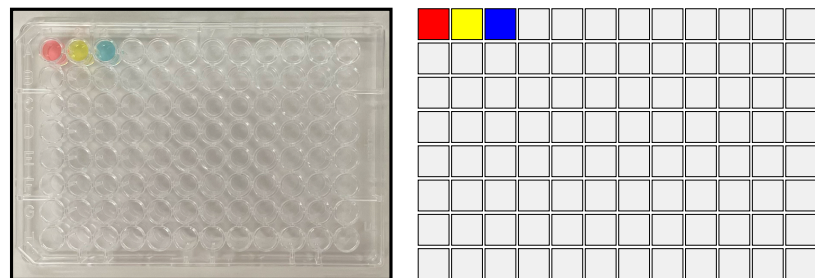


combining multiple hardware modules

You have a source plate with three samples in unknown positions (yellow, blue, and red), each potentially located in one of the wells (A1, B1, and C1). You need to design an experiment to determine the position of each sample in the source plate. <setup description>

User prompt

Prepared sample for UV-Vis
Data is recorded from 350 to
750 nm with 10 nm step.



First, I will transfer a small amount of each sample from the source plate to separate wells in the target plate. Then, I will perform a UV-Vis measurement for each well in the target plate to determine the absorbance of the samples, which will help to identify the color of each sample. Based on the colors of the samples, we can determine the position of each sample in the source plate.

EXPERIMENT

```
from opentrons import protocol_api
...

# Perform UV-Vis measurement on the target plate
UVVIS plate 1
```

In the code the Agent moves liquids from A1, B1, and C1 to A1, A2, and A3. Then it dilutes them.

Planner's output

UV-Vis data from plate 1 is ready. You can access it using PYTHON, it will be in the file named 'plate1.npy'. You will have NumPy installed. The file contains NumPy array of shape (8, 12, 41). There are 41 wavelengths from 350 to 750 with 10 nm step. **Think about how would different colors absorb first.**

Automated answer
+ guiding prompt

combining multiple hardware modules

Different colors absorb light at different wavelengths. Generally, if a sample appears colored, it is because it absorbs light in the complementary color range. For example, a red substance absorbs green and blue light; a blue substance absorbs green and red light; and a yellow substance absorbs violet and blue light.

First, I will load the UV-Vis data from the file 'plate1.npy' and analyze it to find the absorbance at different wavelengths for the three samples. Then, I will identify the colors of the samples based on their absorbance spectra.

```
PYTHON
import numpy as np
...
print(samples_max_wavelength_nm)
```

```
[530, 430, 630]
```

Automated answer

The maximum absorbance wavelengths for the three samples are 530 nm, 430 nm, and 630 nm. Based on the absorbance spectra, we can deduce their colors:

1. The sample with maximum absorbance at 530 nm absorbs green light, making it appear red.
2. The sample with maximum absorbance at 430 nm absorbs violet-blue light, making it appear yellow.
3. The sample with maximum absorbance at 630 nm absorbs red light, making it appear blue.

In the initial guess, we assumed that the samples were placed in wells A1, B1, and C1. However, based on the absorbance data analysis, we find:

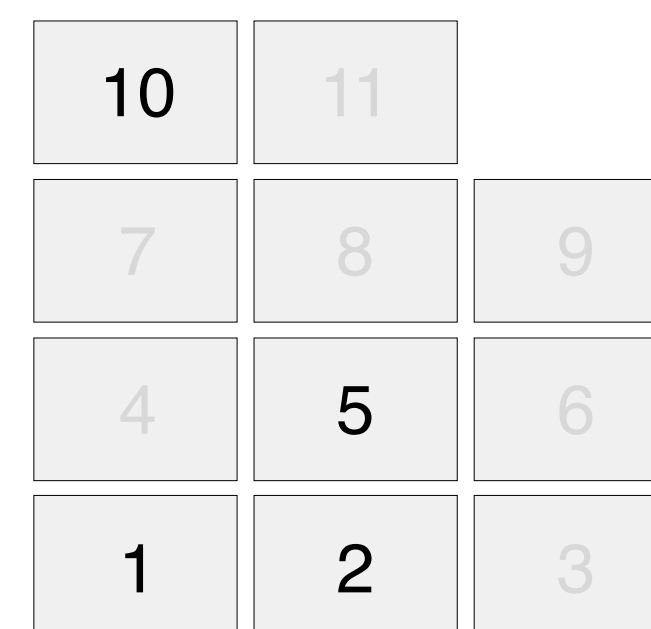
1. The red sample is located in well A1.
2. The yellow sample is located in well B1.
3. The blue sample is located in well C1.

performing chemical reactions autonomously

Perform Suzuki and Sonogashira reactions

<setup description>

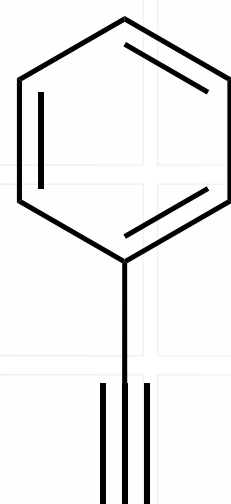
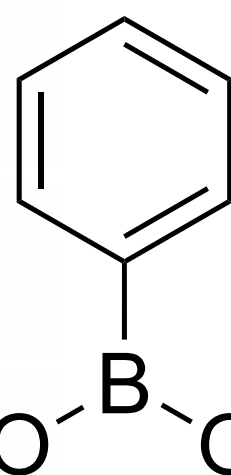
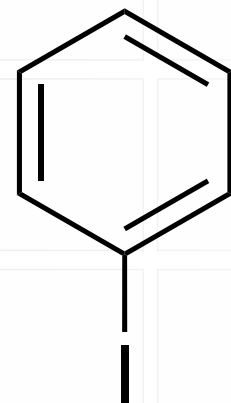
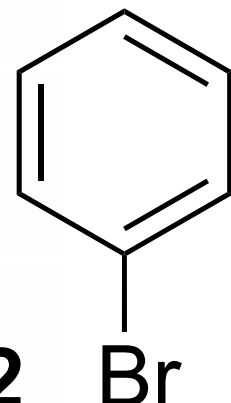
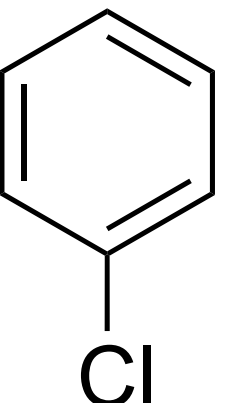
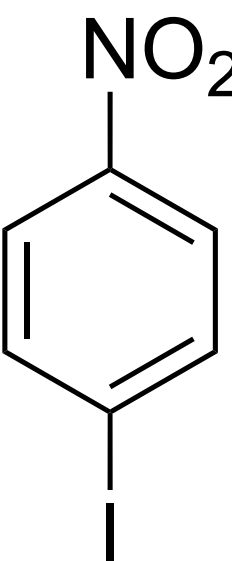
The liquid handler setup scheme

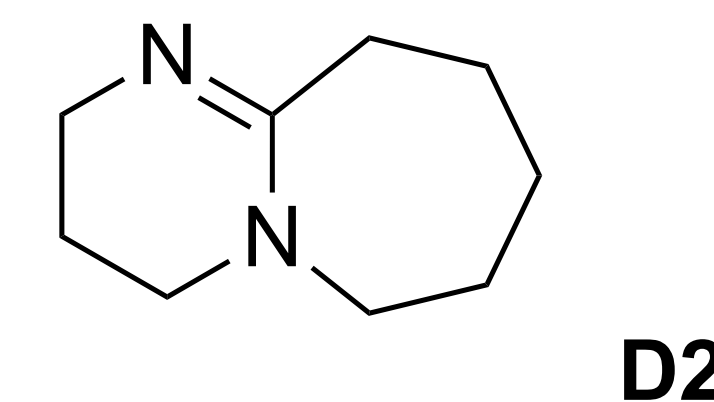
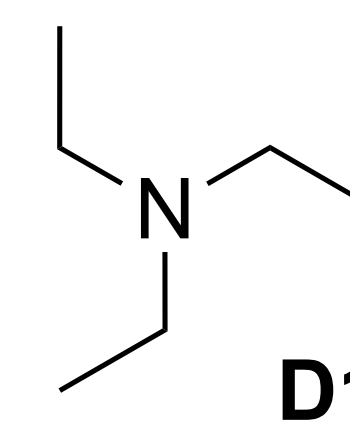
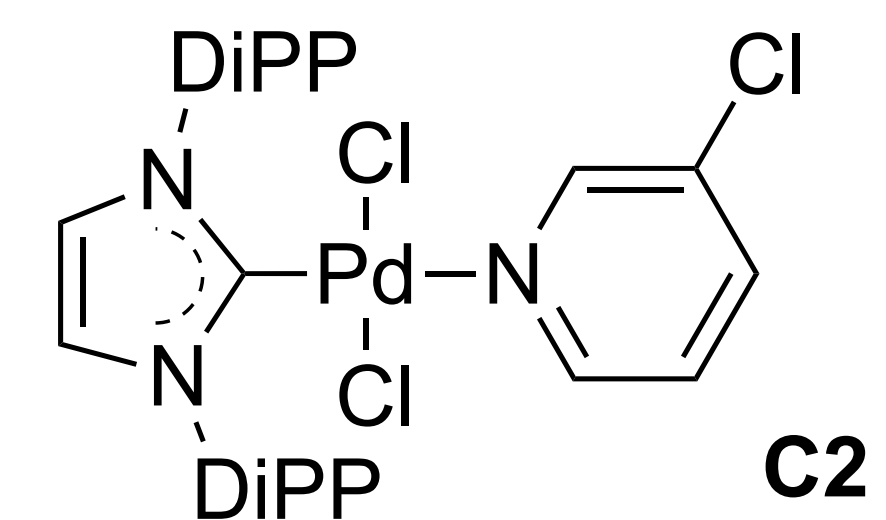
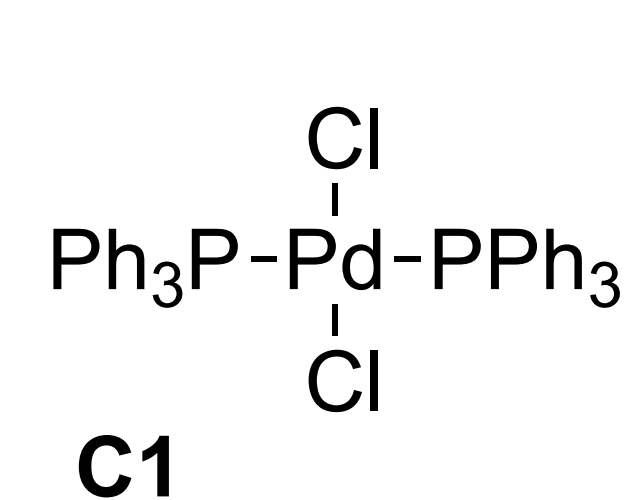


left pipette — 20 μL single-channel
right pipette — 300 μL single-channel

- 1 — 20 μL tips
- 2 — 300 μL tips
- 5 — source plate (deep 96-well)
- 10 — heater-shaker module with target plate

Source plate

A1	A2								
B1	B2	B3	B4						
C1	C2					A1		A2	
D1	D2								
E1									
						B1	B2	B3	B4
						E1 — DMF			



system in action

Perform Suzuki and Sonogashira reactions. Find appropriate reaction first (including usual amounts of reagents/catalysts) and only then perform the experiments.

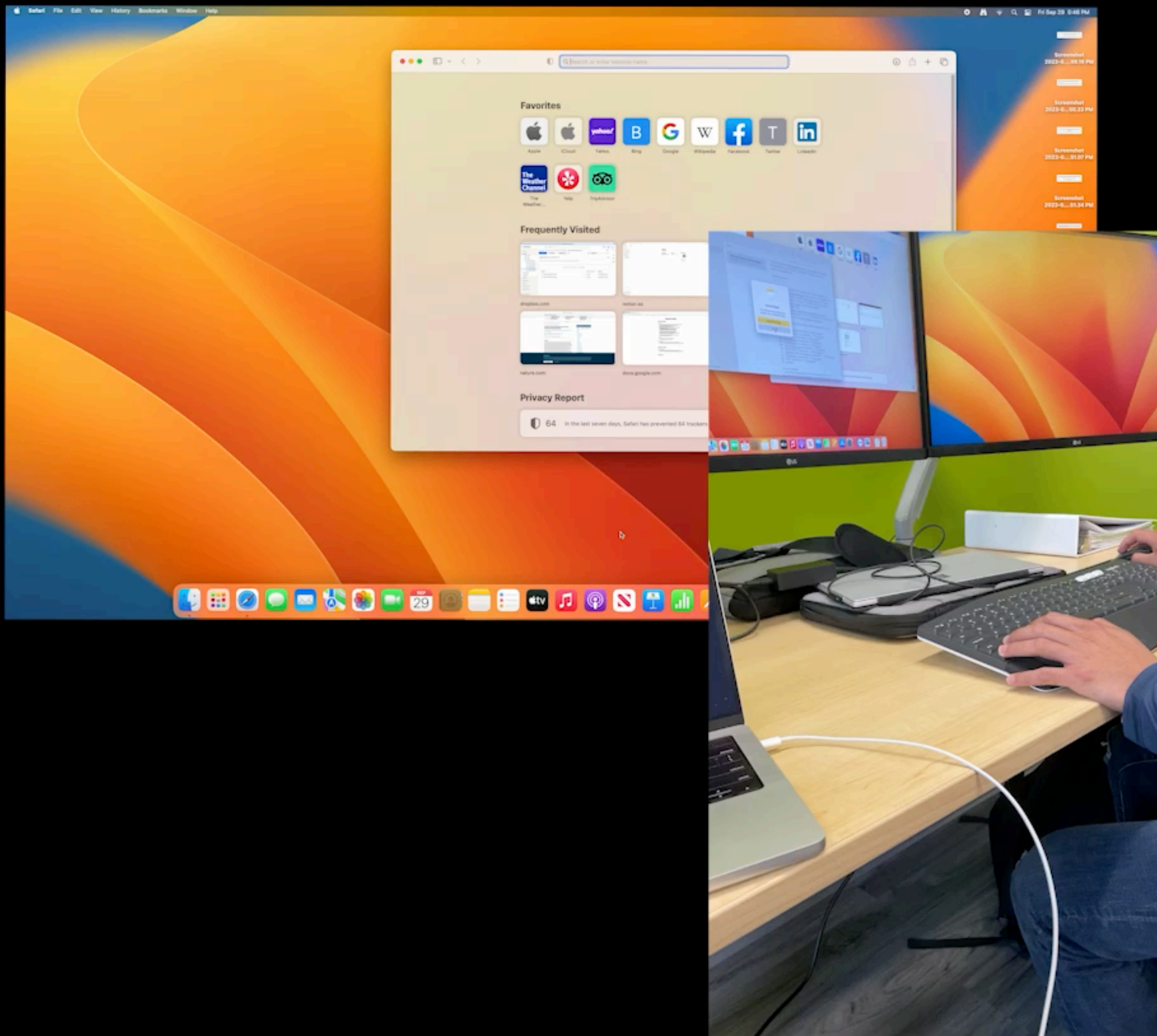
<setup description>

```
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
```

```
def _get_pipette_config():
    """
    Returns the configuration for the pipette.
    """
    # (API version 2.13) has the following setup: left - 20 µL pipette p20_single_gen2 (max volume 20 µL, tiprack for 20 µL in slot 1), right - 300 µL pipette p300_single_gen2 (max volume 300 µL, tiprack for 300 µL in slot 1)
    # available (these are solutions in 1 mL of DMF):
    # concentration - 1.35 M mmol/mL
    # concentration - 0.13 mmol/mL
    # concentration - 1.87 mmol/mL
    # concentration - 1.47 mmol/mL
    # concentration - 1.83 mmol/mL
    # concentration - 0.27 mmol/mL
    # concentration - 0.0057 mmol/mL
    # concentration - 0.015 mmol/mL
    # concentration - 1.41 mmol/mL
    # concentration - 1.06 mmol/mL

    # Lock the heater-shaker module: do hs_mod.close_labware_latch() before any pipetting/operations with the module
    # Lock the pipette (20 µL or 300 µL)
    # Set temperature"""
```

```
Terminal: Local x Local (2) x Local (3) x Local (4) x Local (5) x Local (6) x Local (7) x Local (8) x + v
(skynetlab) dboiko@daniils-mbp-2 skynetlab % python run.py
```

Running the experiment

Run prompt Results Paper

Agent settings

Experimental platform: OT-2

System prompt:
Who are you?
You're an assistant for chemists, who performs experiments in the cloud lab. You will have access to the cloud lab (you can do real experiments, using a liquid handler and UV-VIS), in particular, you will be able to perform the following actions:
<COMMANDS>
Important notes

Available tools (* interrupts execution)

- Google Search with browsing
- Code execution (Python with NumPy)
- Performing experiments* (ECL, OT-2 or Hamilton)
- Bioinformatics (Protein and DNA sequence analysis)
- Plate reader* (Using UV-Vis plate reader)
- Documentation search (ECL, OT-2 or Hamilton)
- Ordering chemicals* (CAS# or DNA sequence)

Token

Token: x5bhQ2TaMkRv9SWXr8kOE0KkUm3MPjnr

Experiment settings

Experiment: humans_vs_ai

Prompt:
C1 — Pd(PPh3)2Cl2; concentration — 0.0057 mmol/mL
C2 — PEPPSI-IPr (NHC-Pd complex); concentration — 0.015 mmol/mL
D1 — triethylamine; concentration — 1.41 mmol/mL
D2 — DBU; concentration — 1.06 mmol/mL
E1 — just DMF
Common mistakes:
- not closing the labware latch of the heater-shaker module: do `hs_mod.close_labware_latch()` before any pipetting/operations with the module
- using more than the limit of the pipette (20 μ L or 300 μ L)
- heater shaker — `set_target_temperature`

Submit

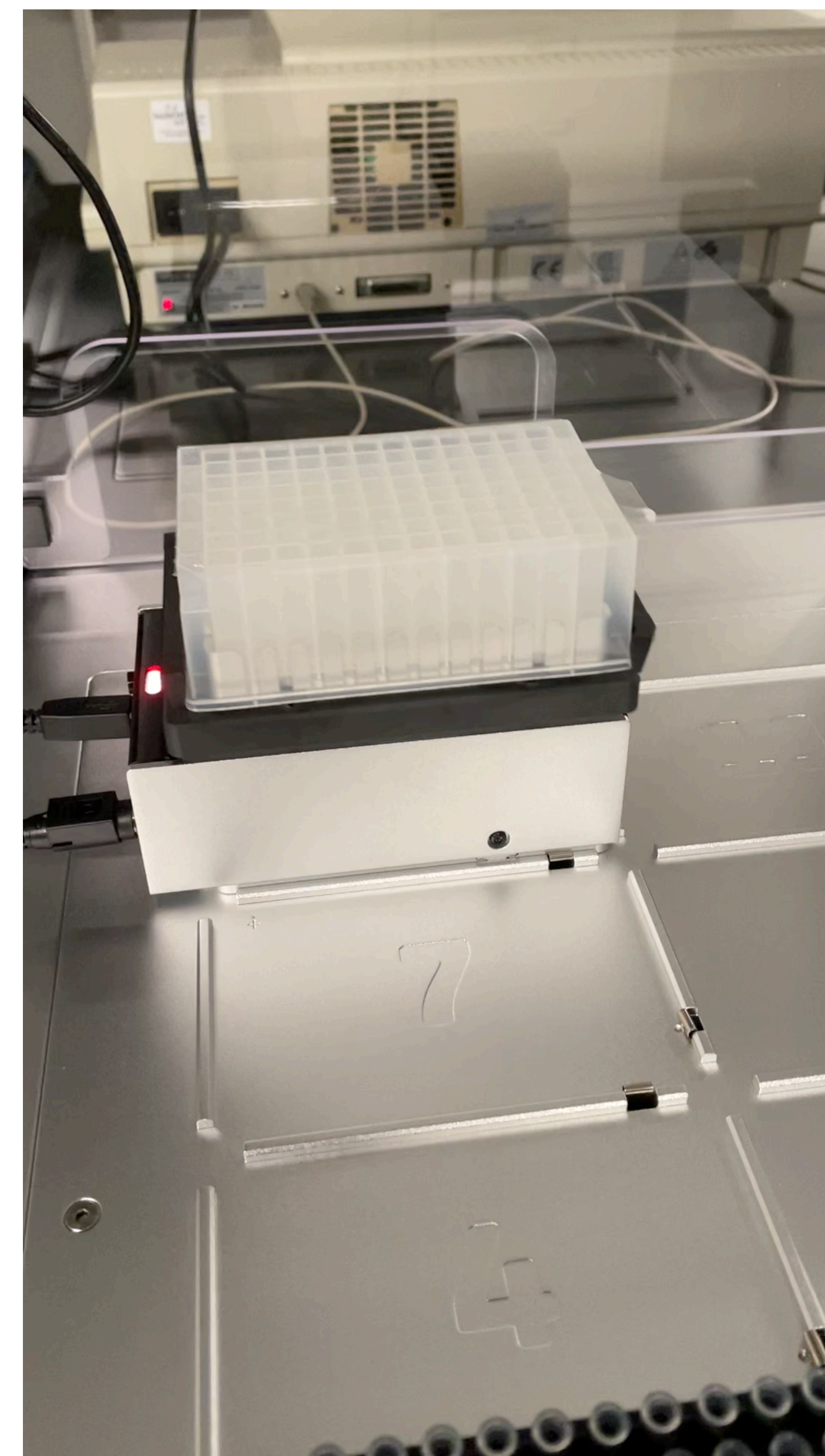
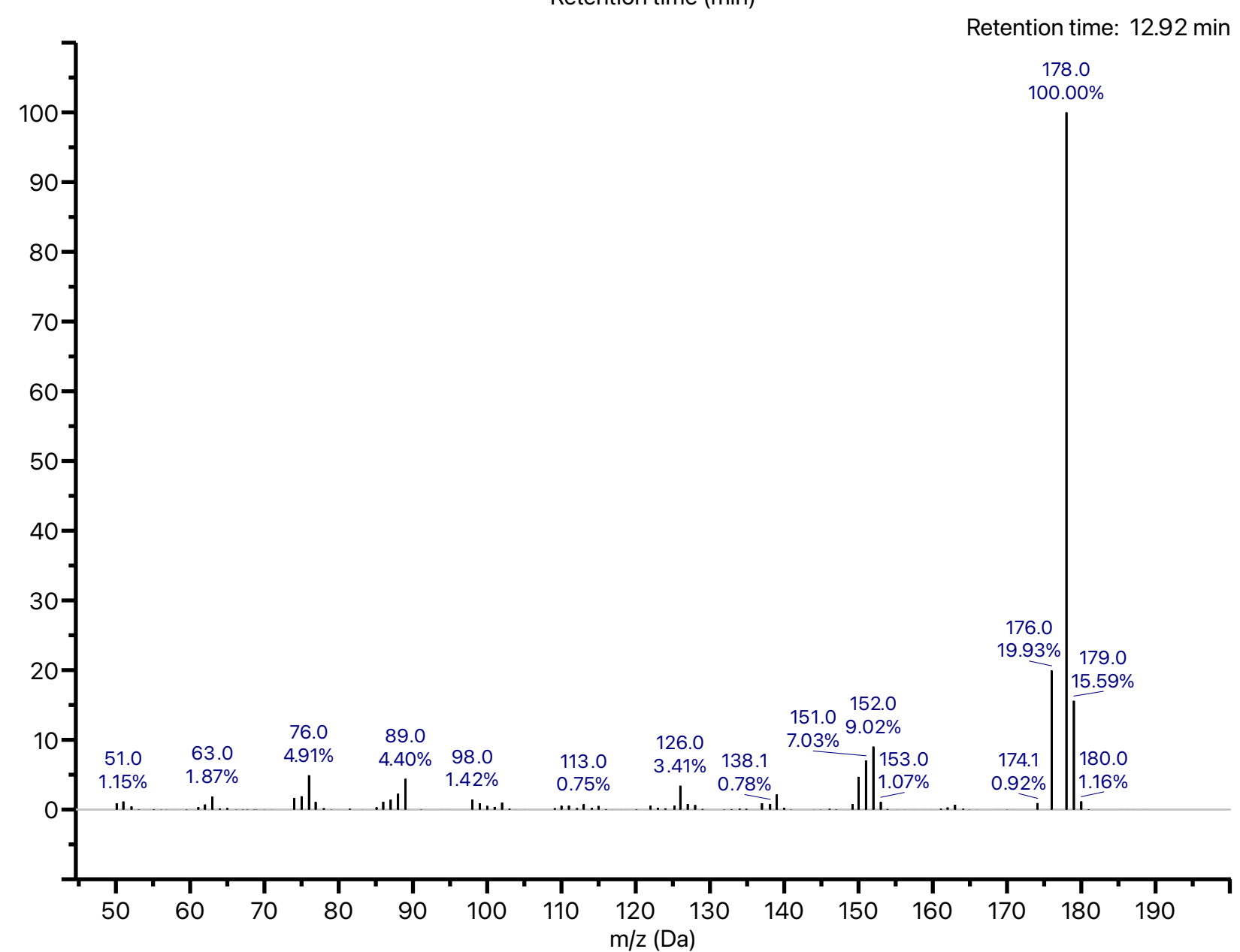
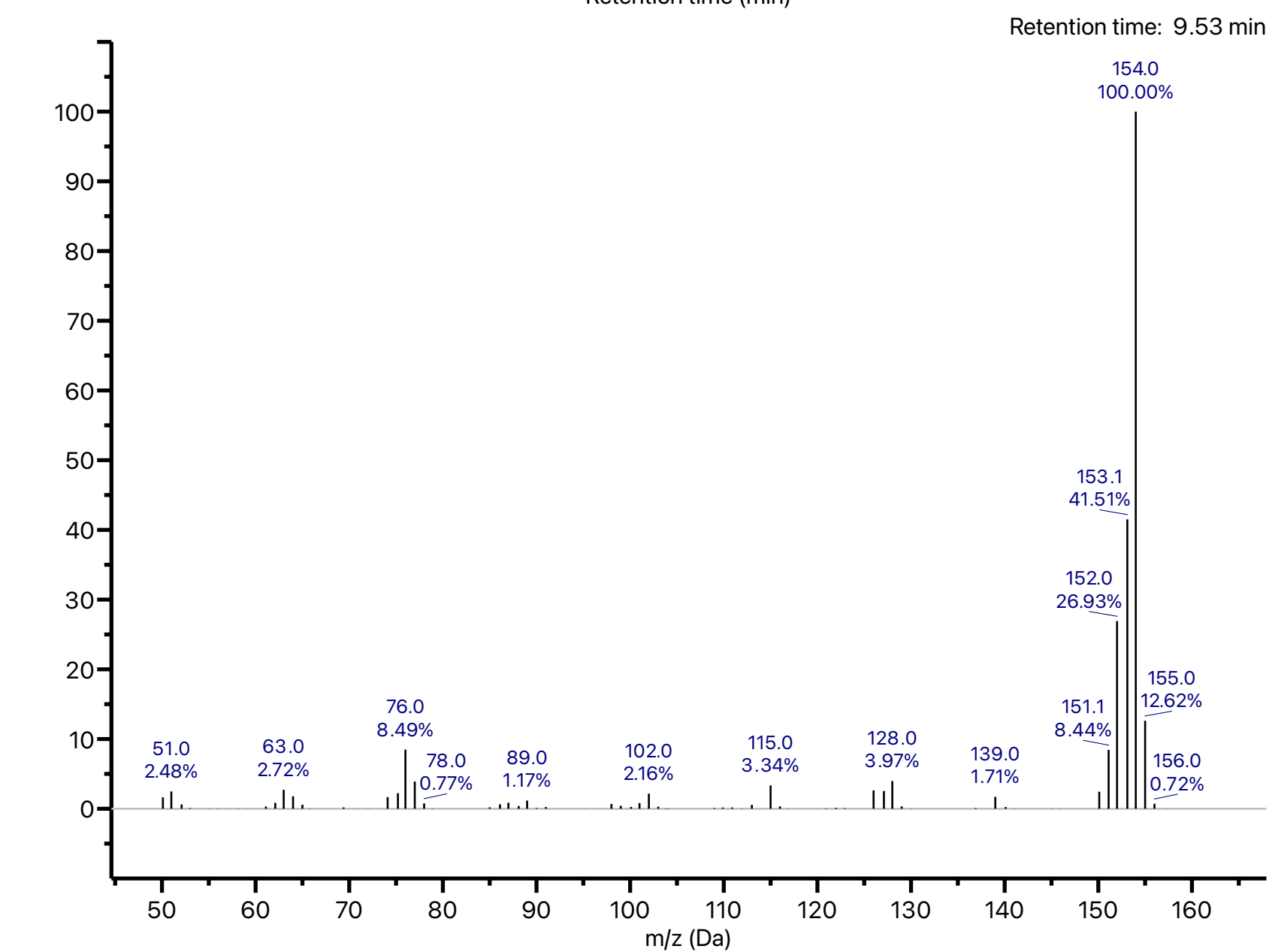
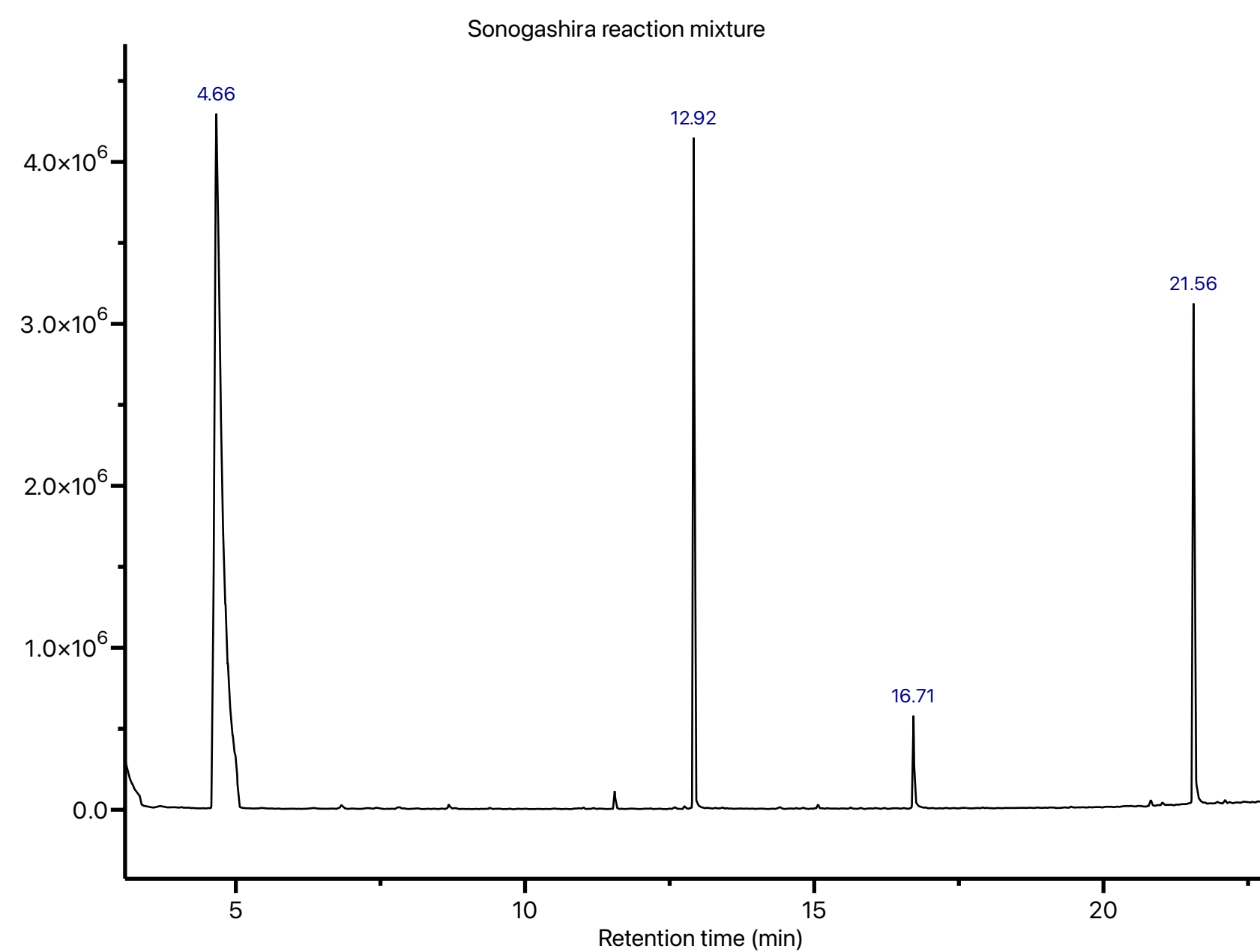
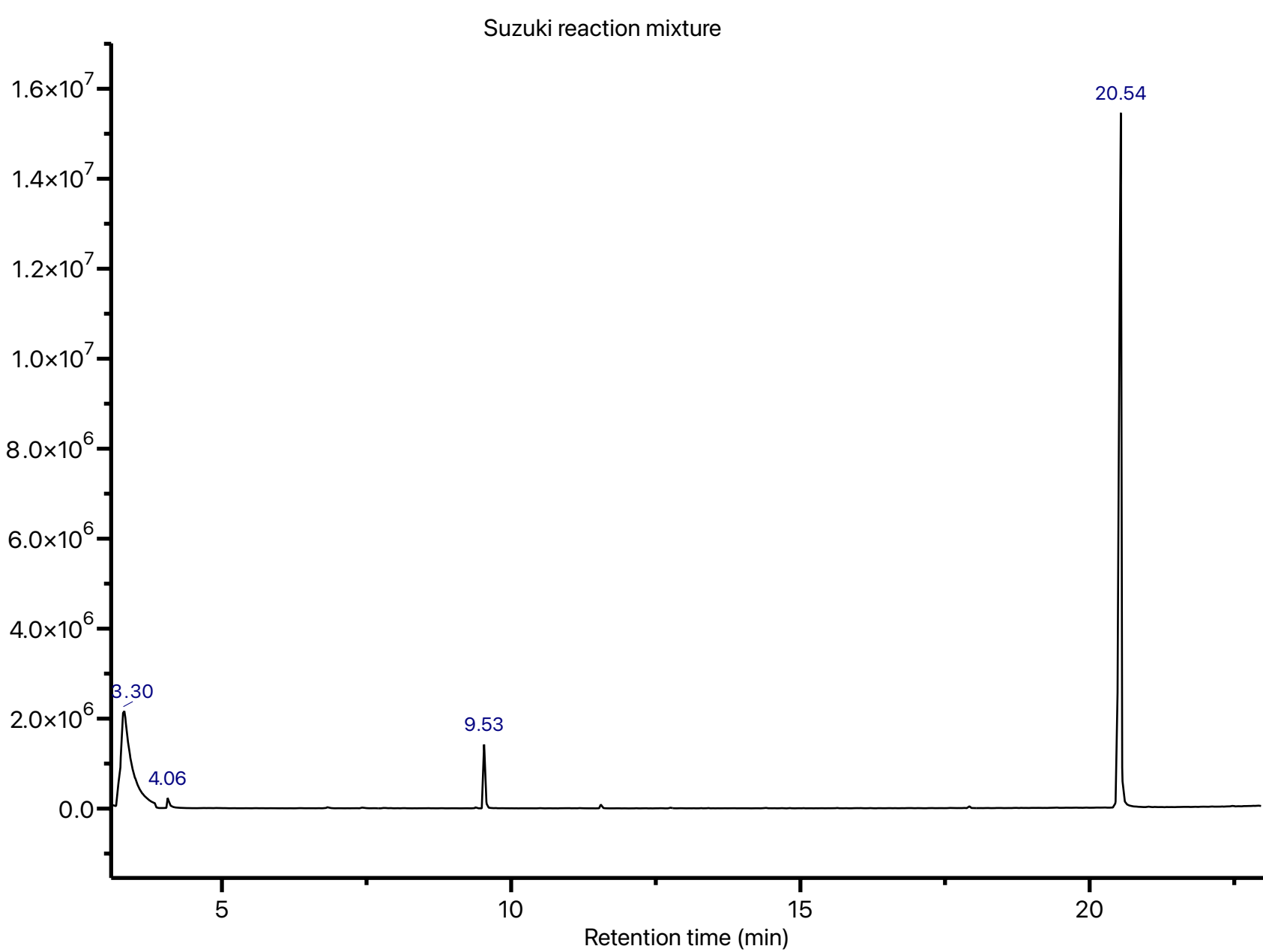
Developed at [Carnegie Mellon University](#)

Jose, PhD student
Chemistry Department
Carnegie Mellon University

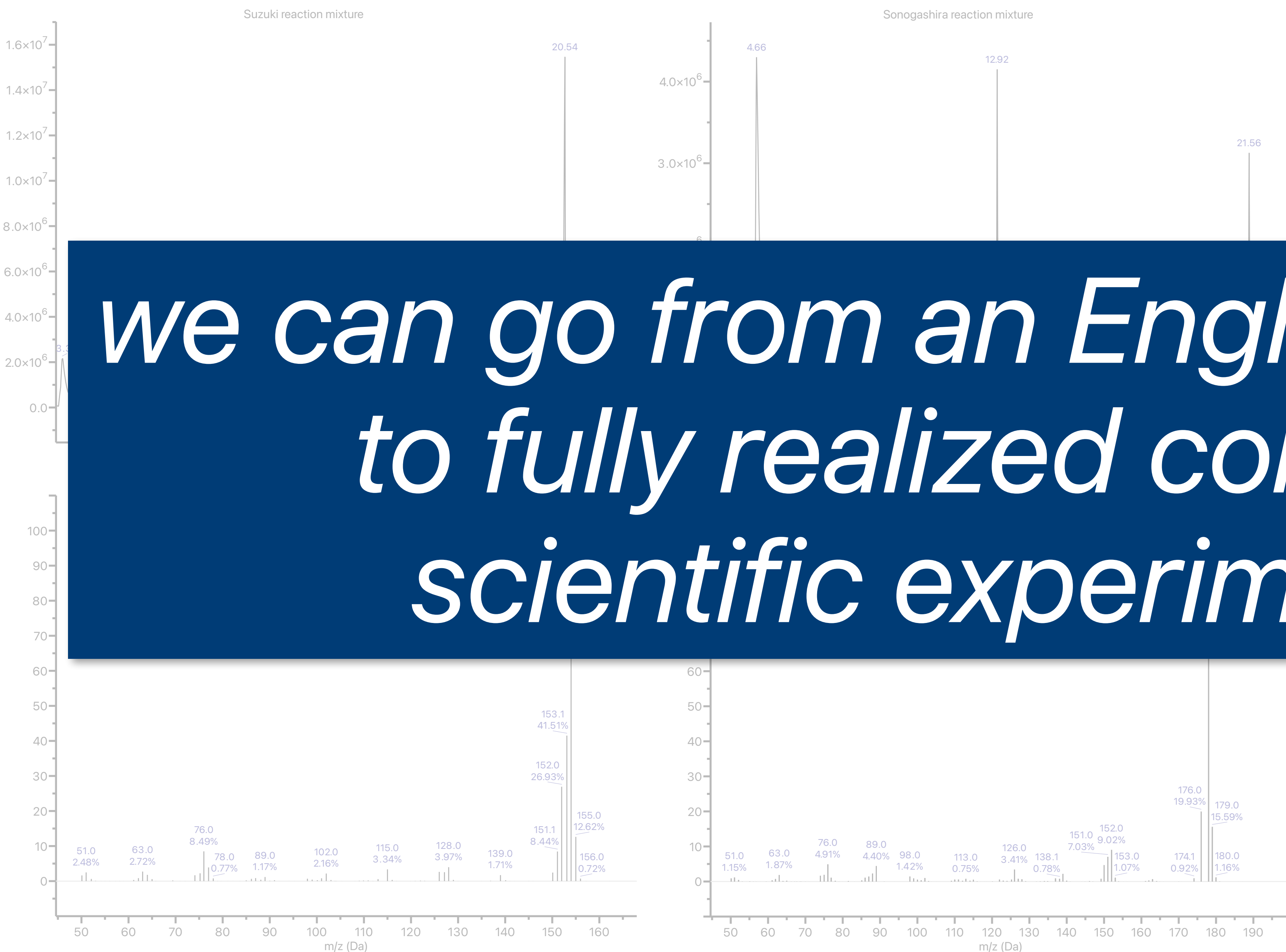
00;00;00;00

Coscientist

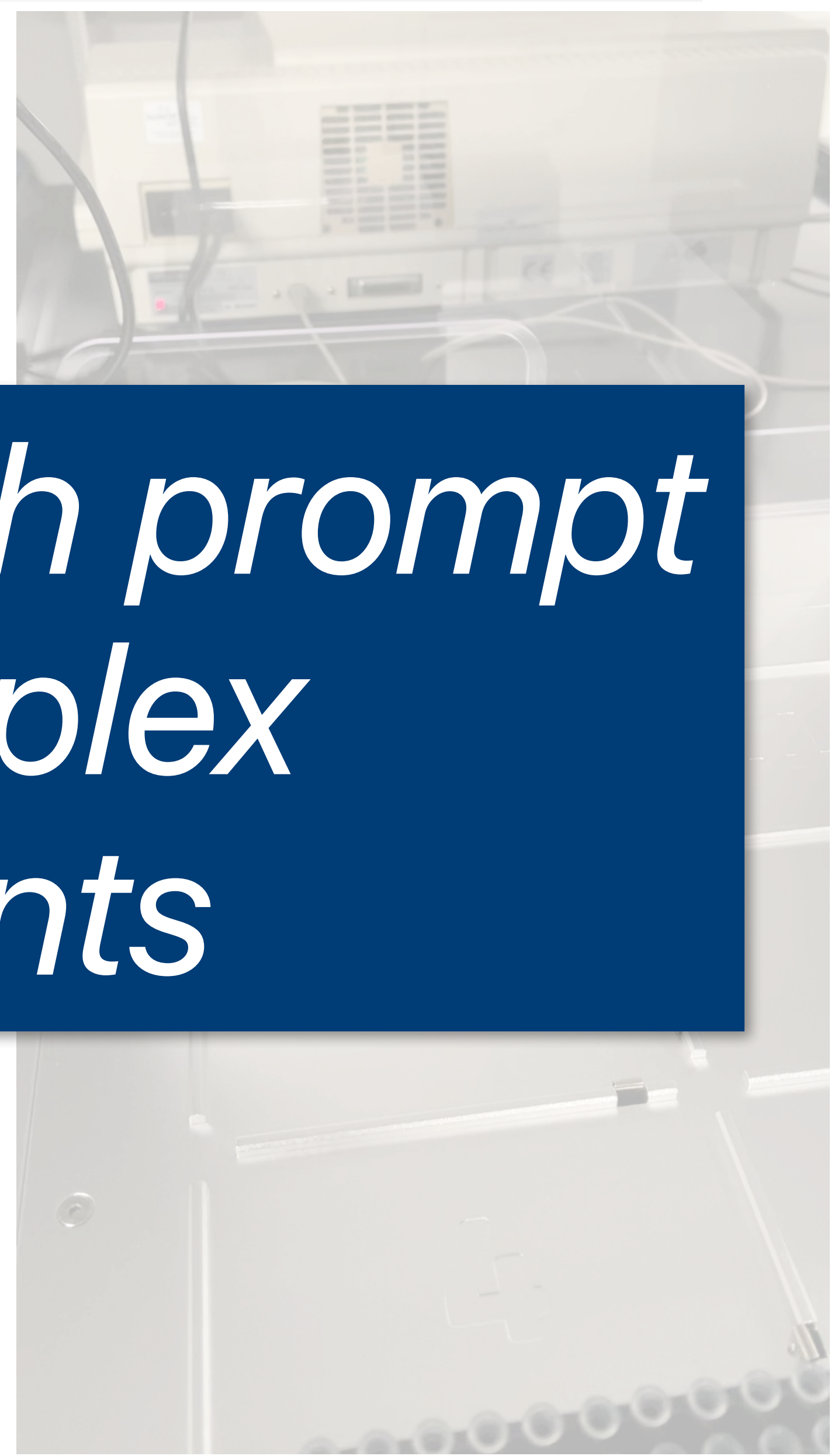
reactions were performed successfully



reactions were performed successfully



we can go from an English prompt to fully realized complex scientific experiments



autonomous experimentation in Emerald Cloud Lab

*specifically specified sample
and instrument available in the cloud lab*

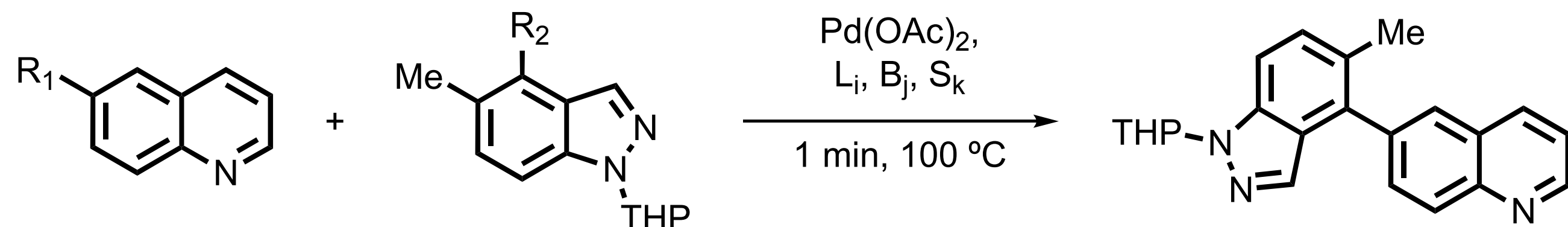
***“how can we leverage
previously collected
knowledge?”***

4. Detector -- Pressure, Temperature, Conductance,
Fluorescence, EvaporativeLightScattering, UVVis,
PhotoDiodeArray, CircularDichroism, RefractiveIndex,
pH, MultiAngleLightScattering, DynamicLightScattering

gradient data

how can we leverage previously collected knowledge?

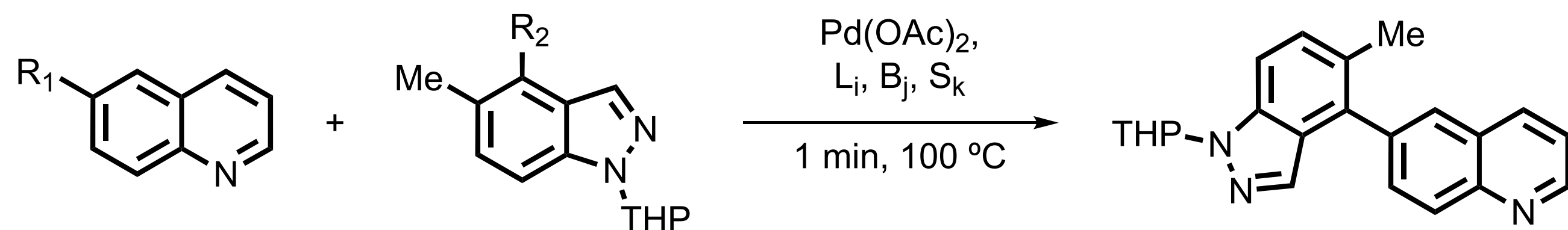
D. Perera, et al. *Science* **2018**, 359(6374), 429-434



$$\text{normalized advantage} = \frac{y_i - \frac{1}{n} \sum_j y_j}{\max_j y_j - \frac{1}{n} \sum_j y_j}$$

how can we leverage previously collected knowledge?

D. Perera, et al. *Science* **2018**, 359(6374), 429-434

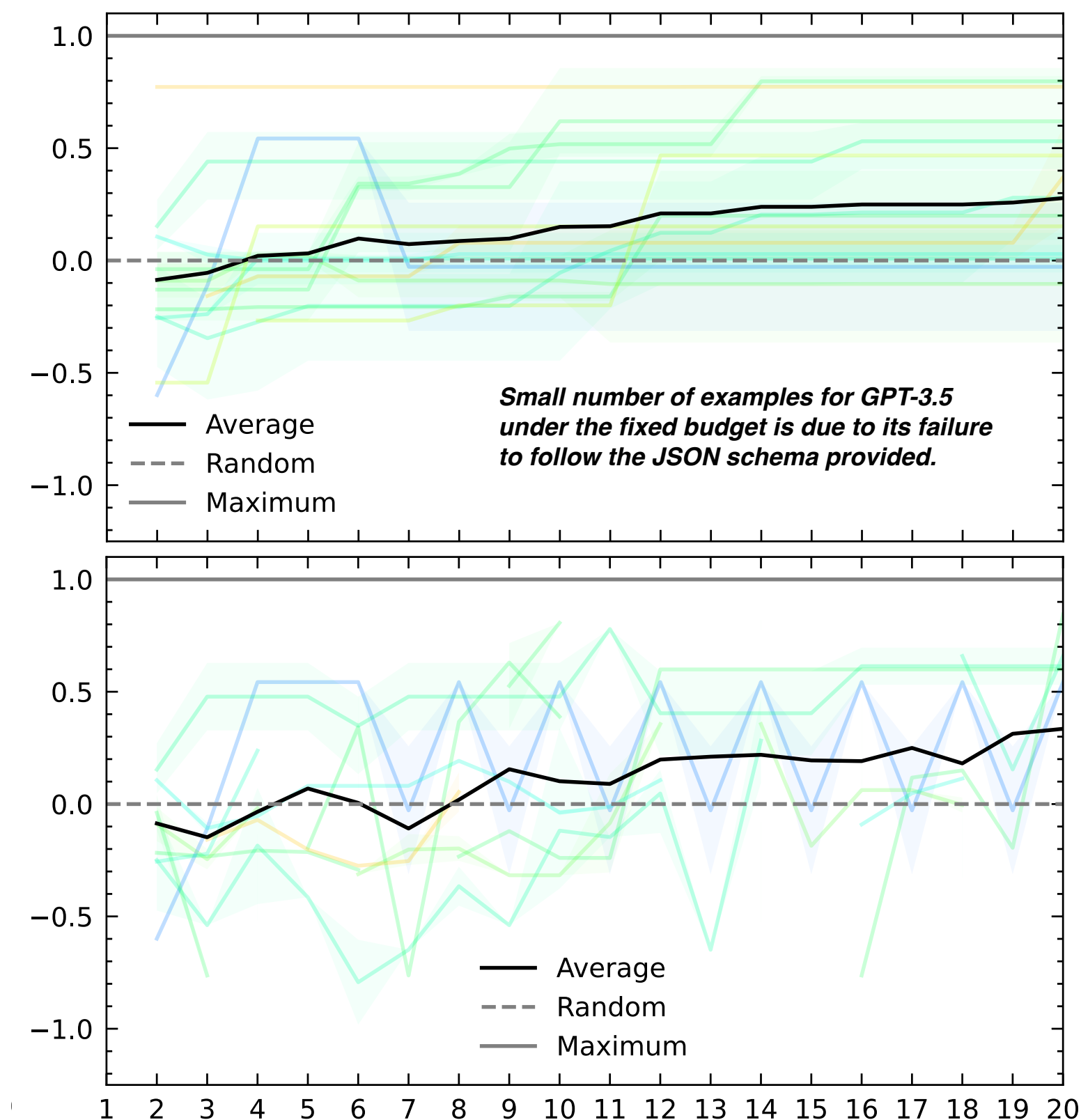


$$\text{normalized advantage} = \frac{y_i - \frac{1}{n} \sum_j y_j}{\max_j y_j - \frac{1}{n} \sum_j y_j}$$

Normalized maximum advantage

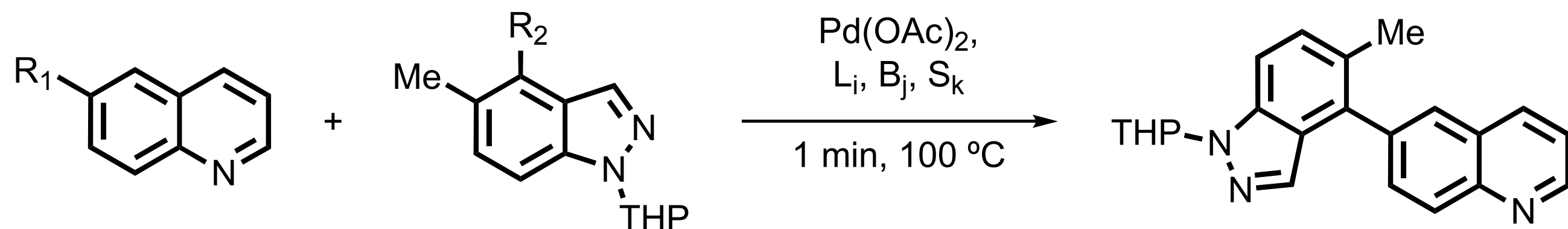
Normalized advantage

GPT-3.5 without prior information



can LLMs be good chemical reaction optimizers?

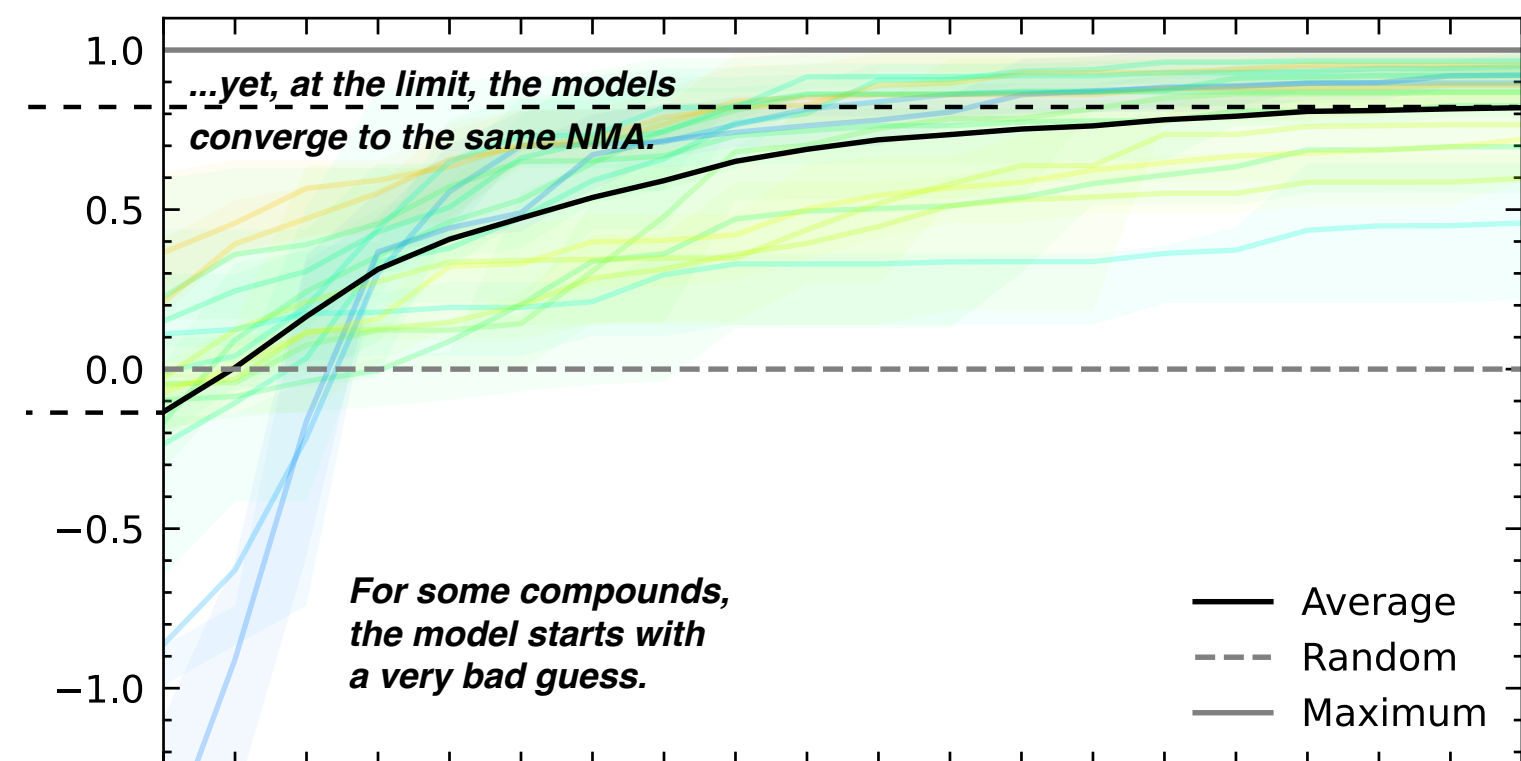
D. Perera, et al. *Science* **2018**, 359(6374), 429-434



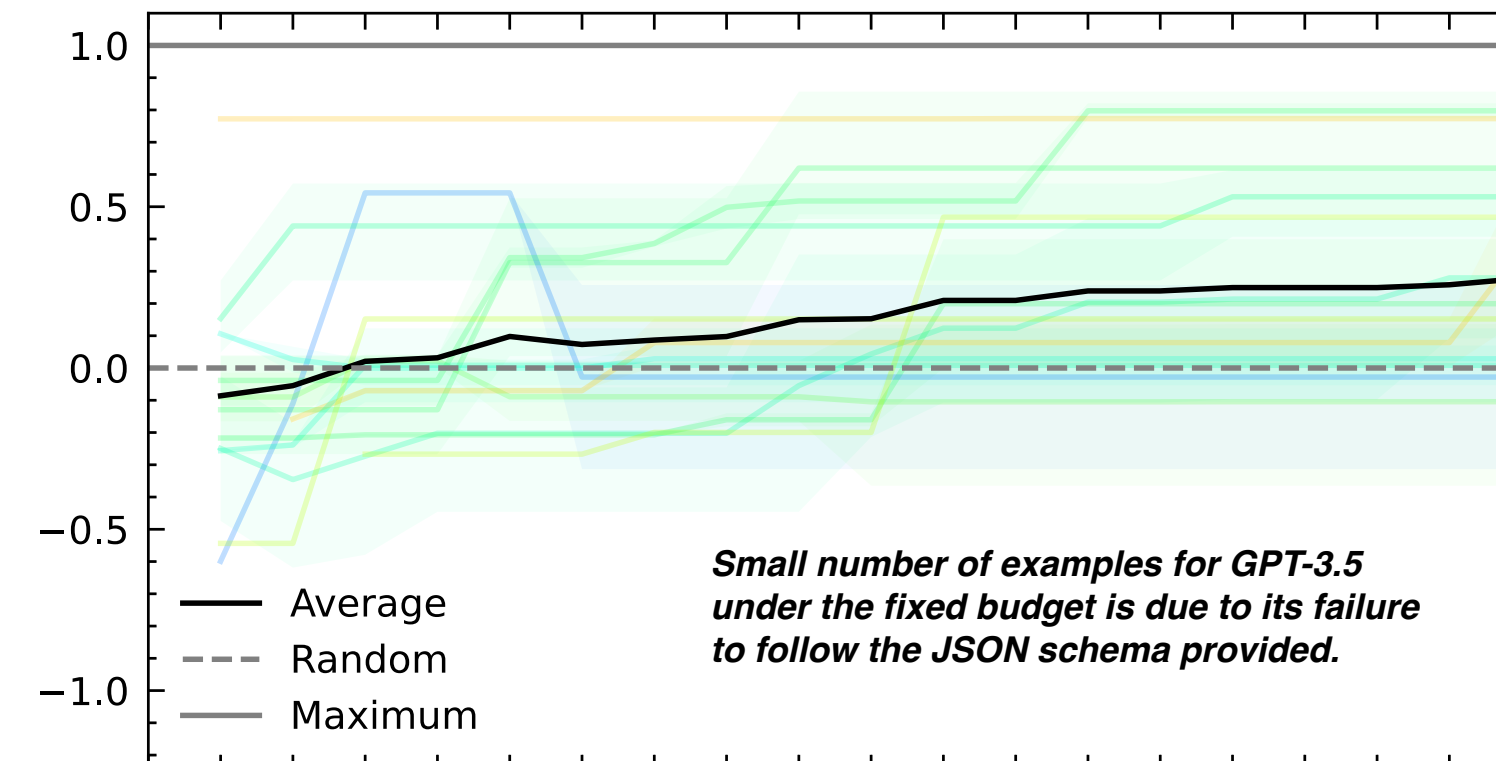
$$\text{normalized advantage} = \frac{y_i - \frac{1}{n} \sum_j y_j}{\max_j y_j - \frac{1}{n} \sum_j y_j}$$

Normalized maximum advantage

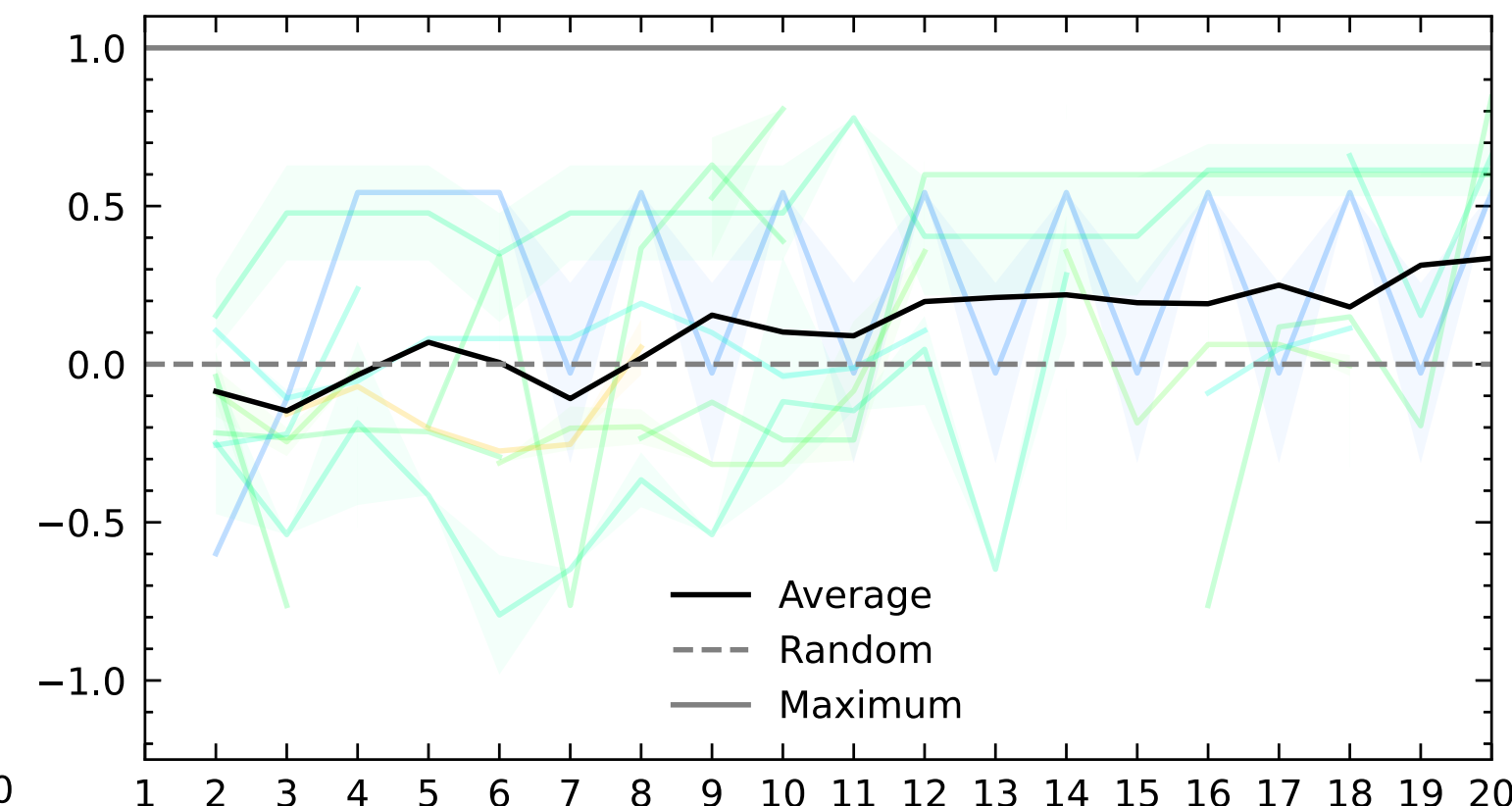
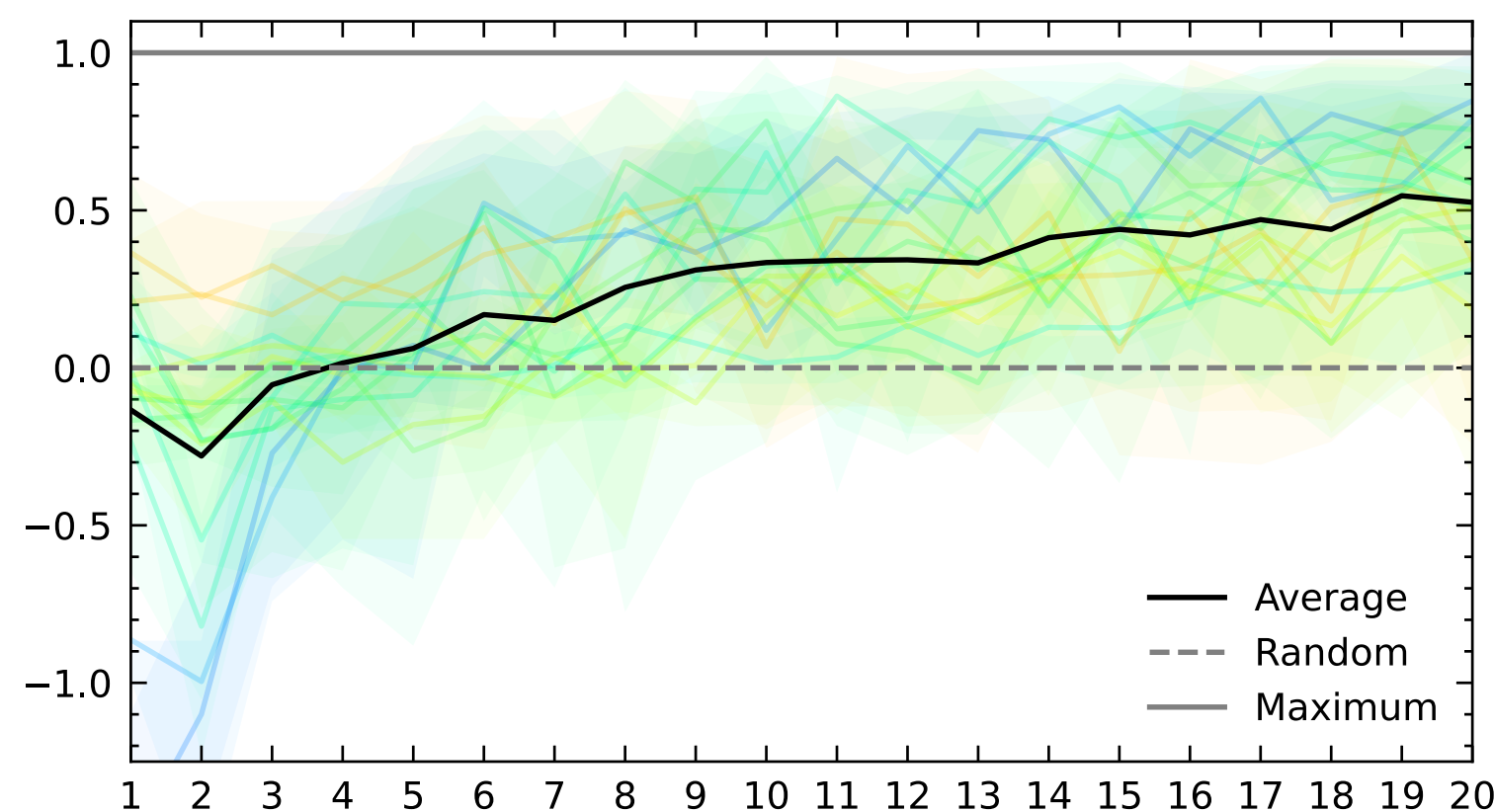
GPT-4 without prior information



GPT-3.5 without prior information

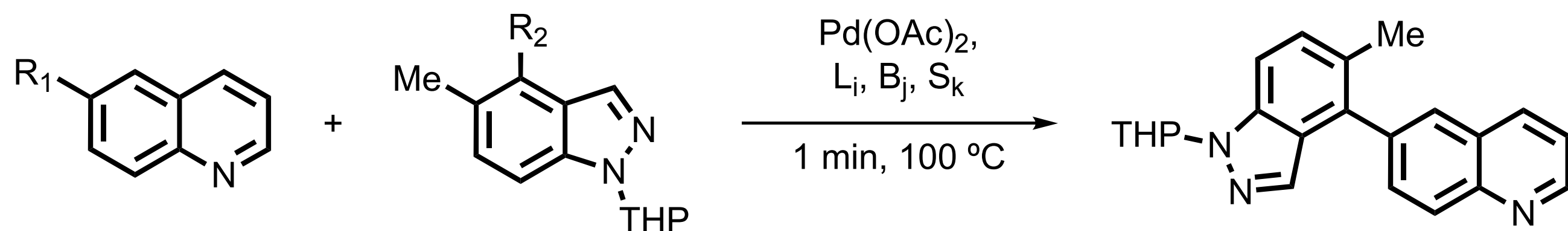


Normalized advantage



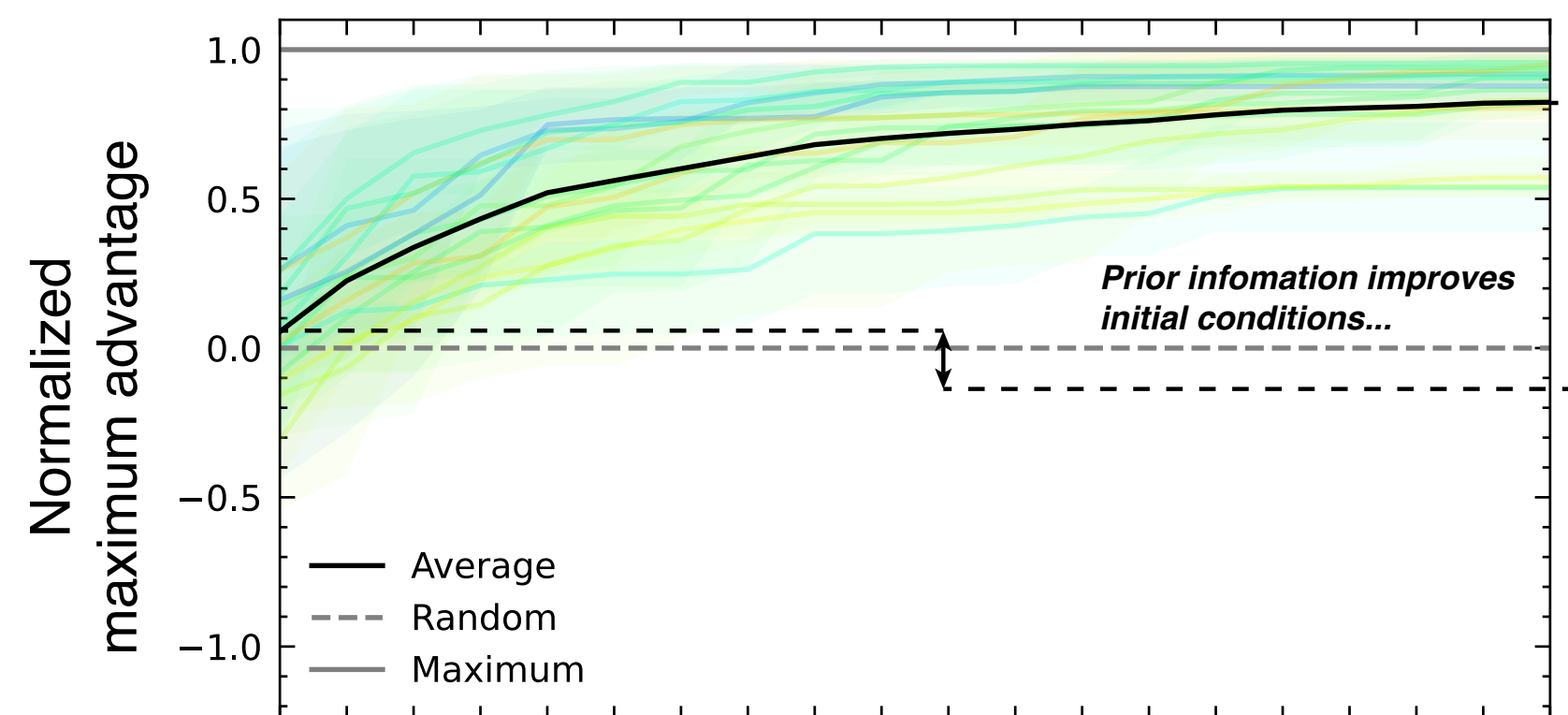
Coscientist is an excellent chemical reaction optimizer

D. Perera, et al. *Science* **2018**, 359(6374), 429-434

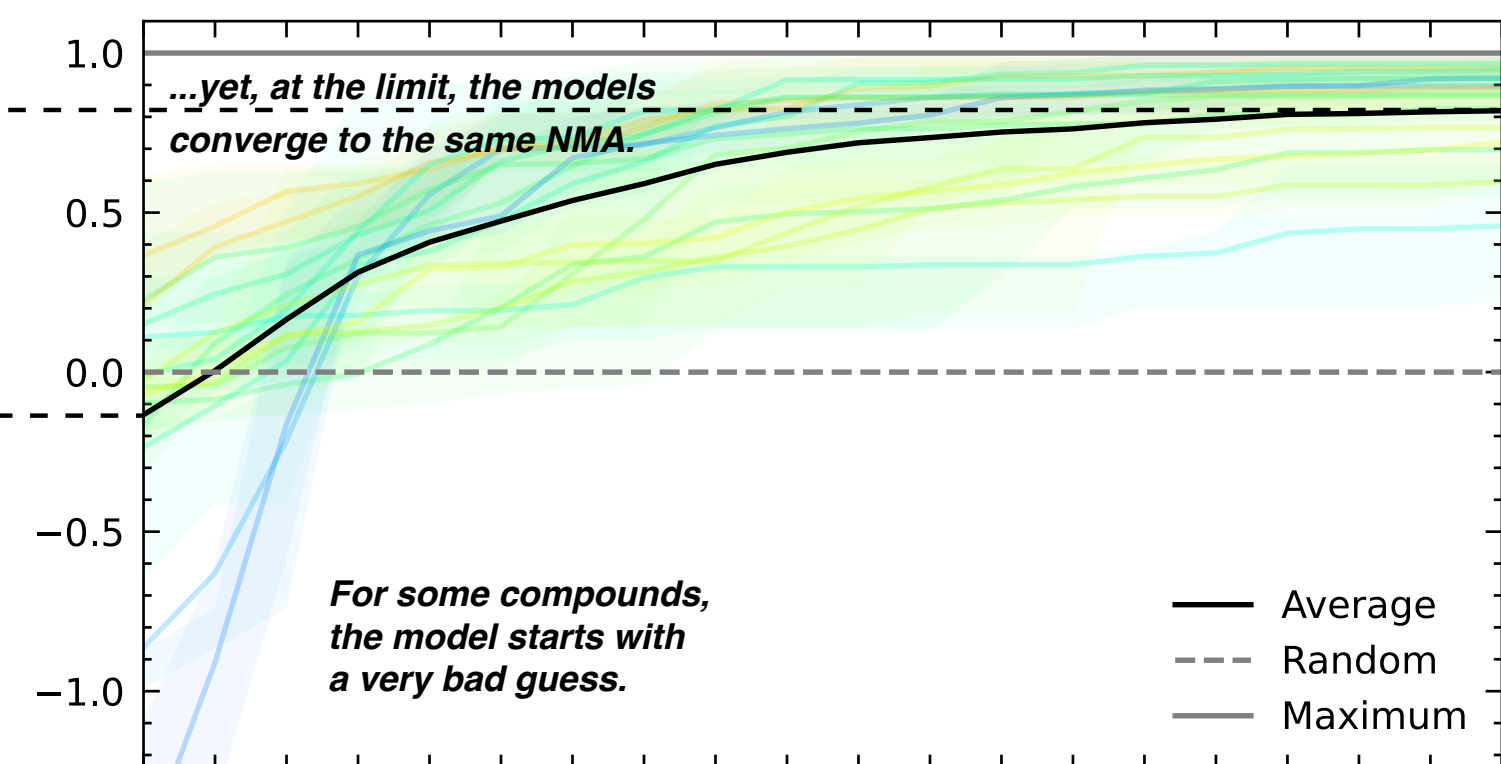


$$\text{normalized advantage} = \frac{y_i - \frac{1}{n} \sum_j y_j}{\max_j y_j - \frac{1}{n} \sum_j y_j}$$

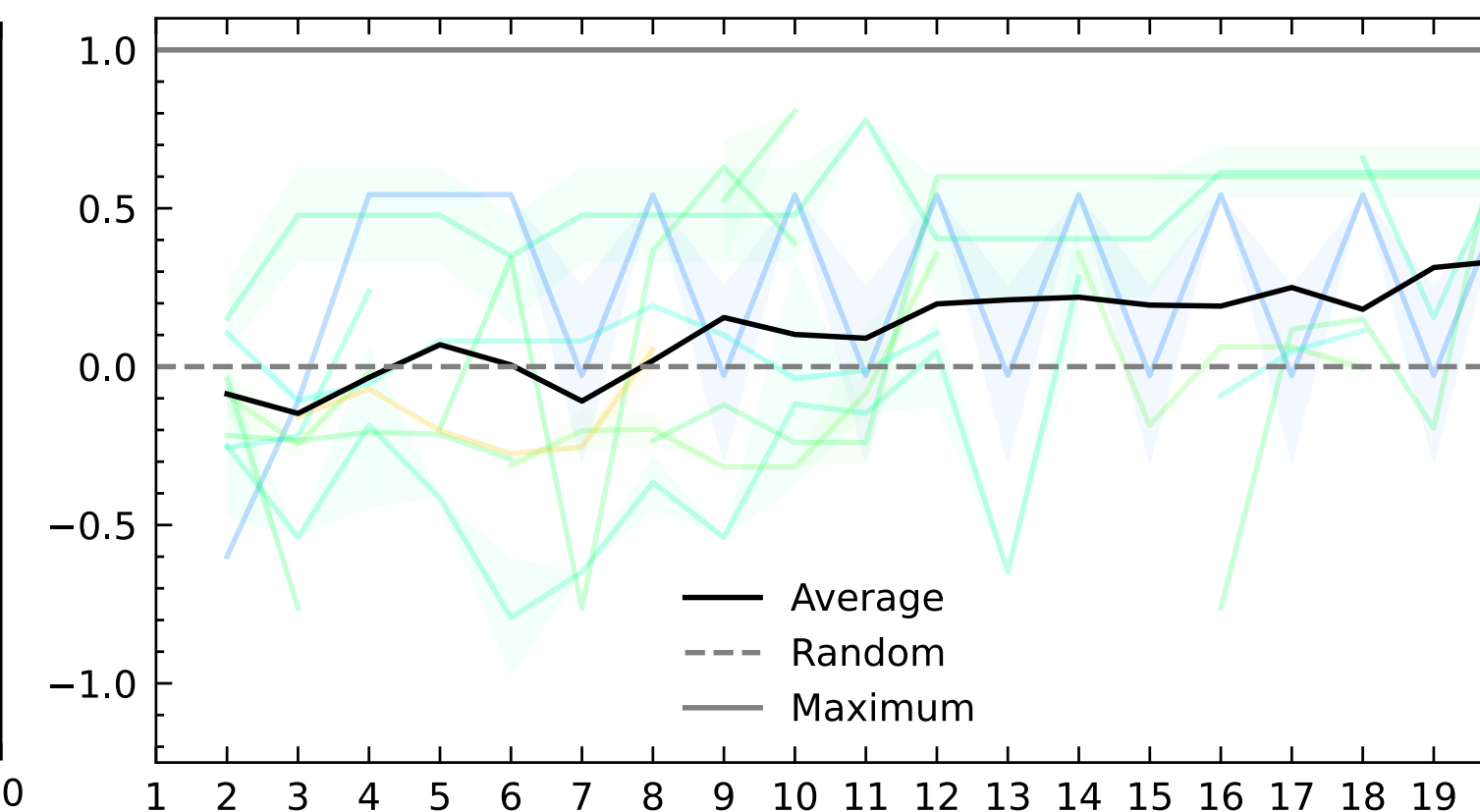
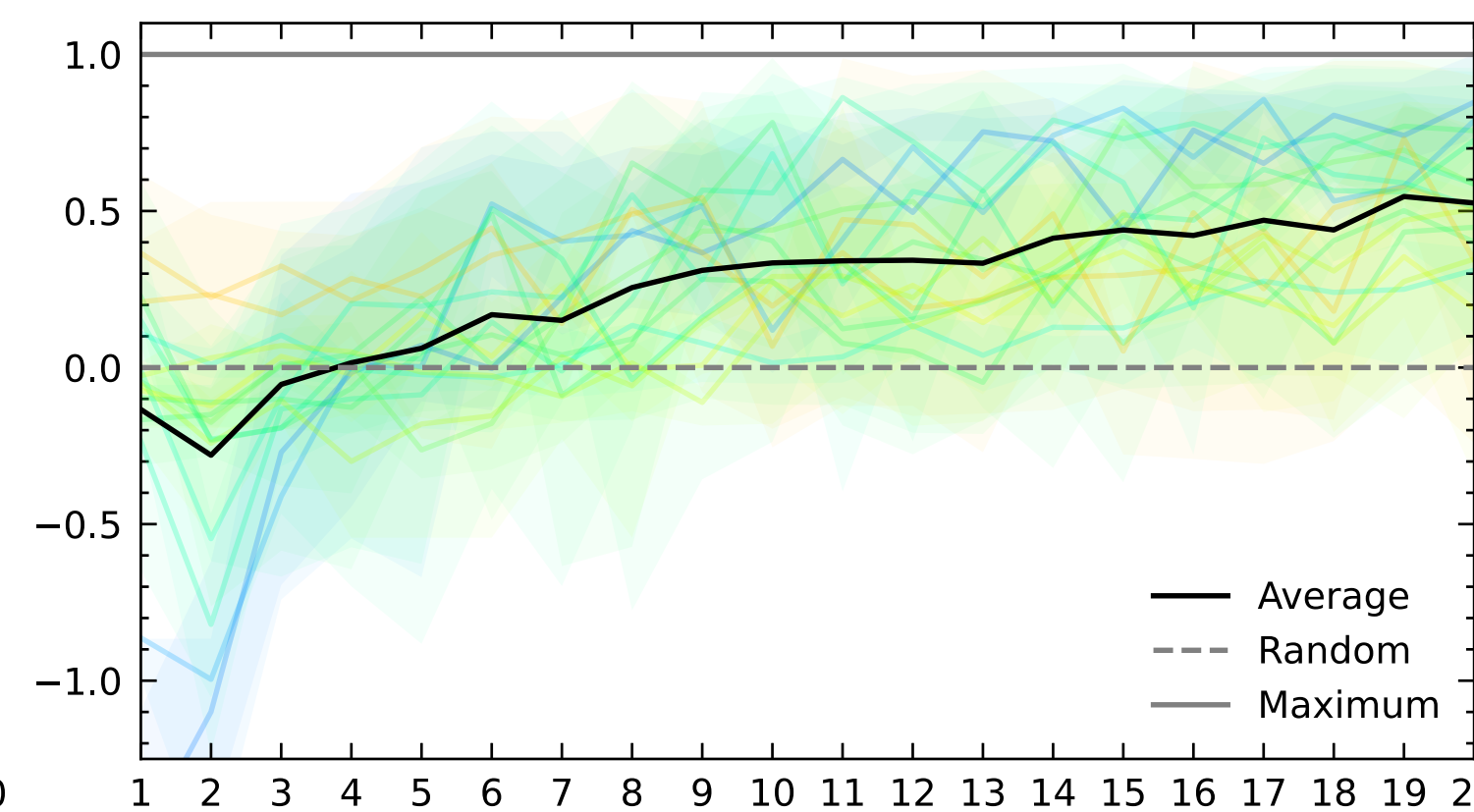
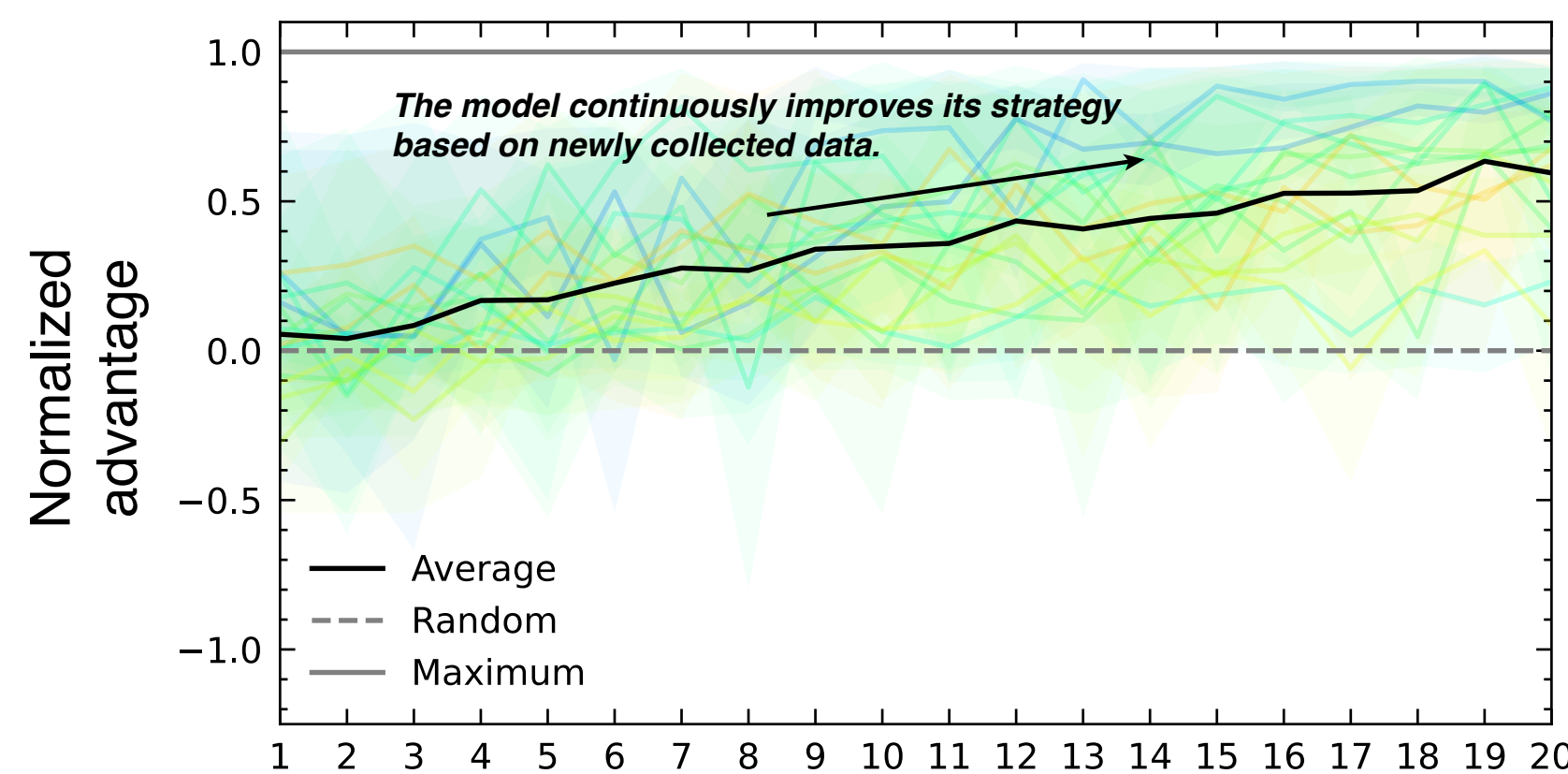
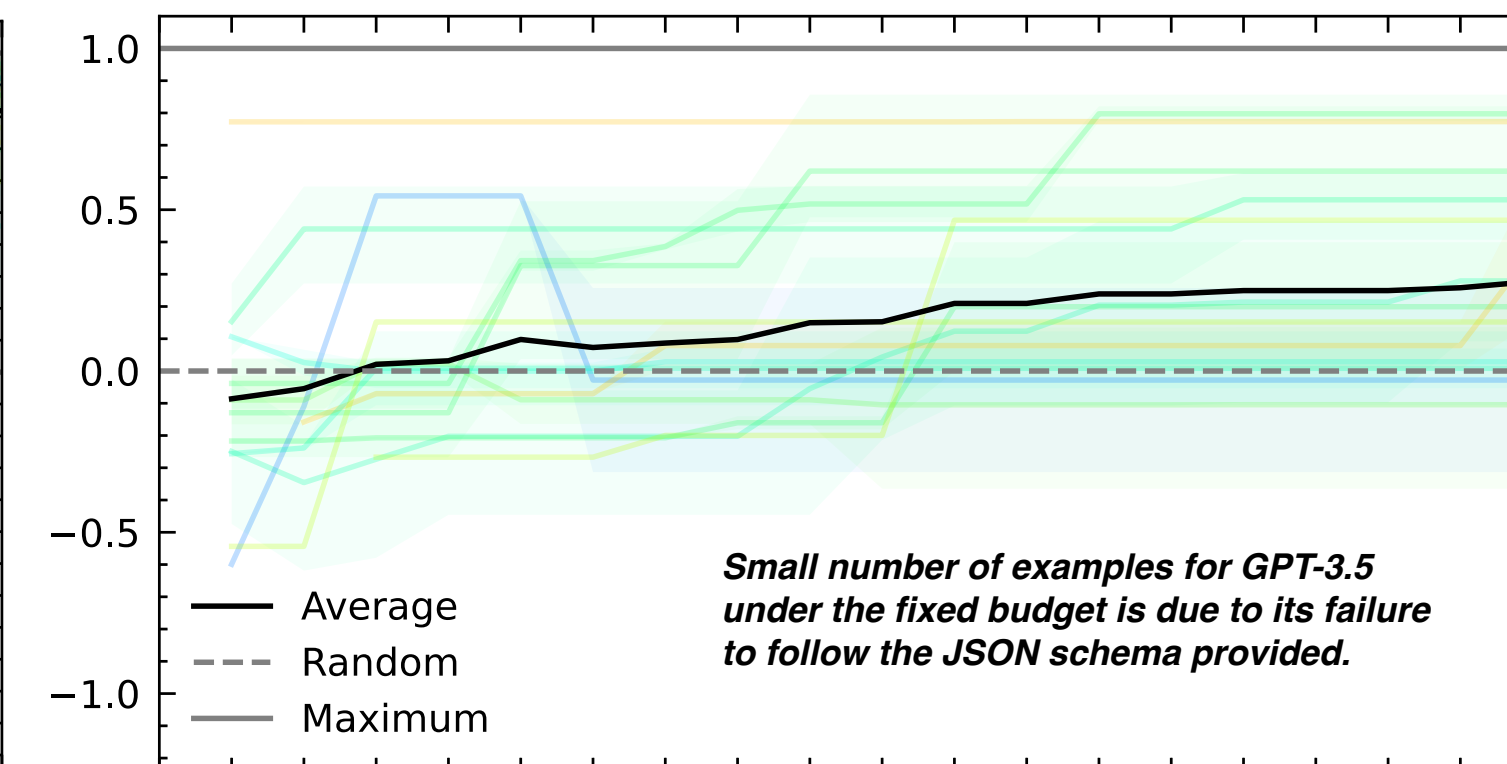
GPT-4 with prior information (10 data points)



GPT-4 without prior information

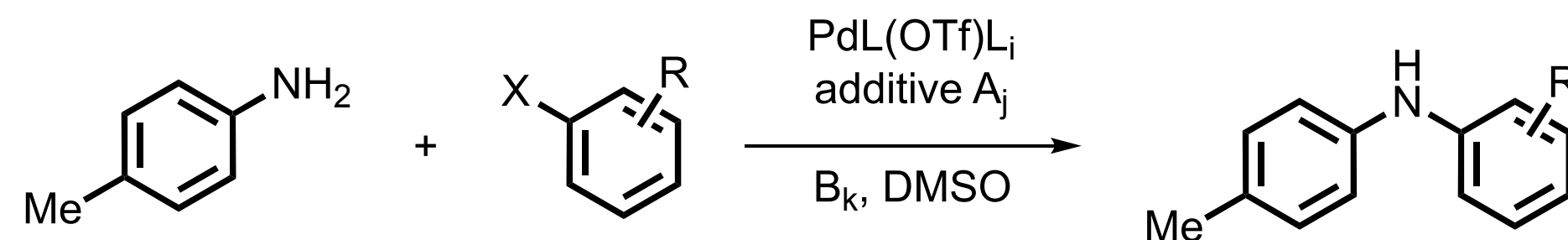


GPT-3.5 without prior information



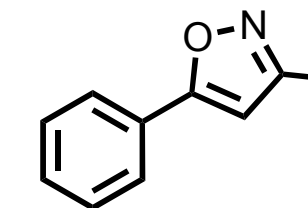
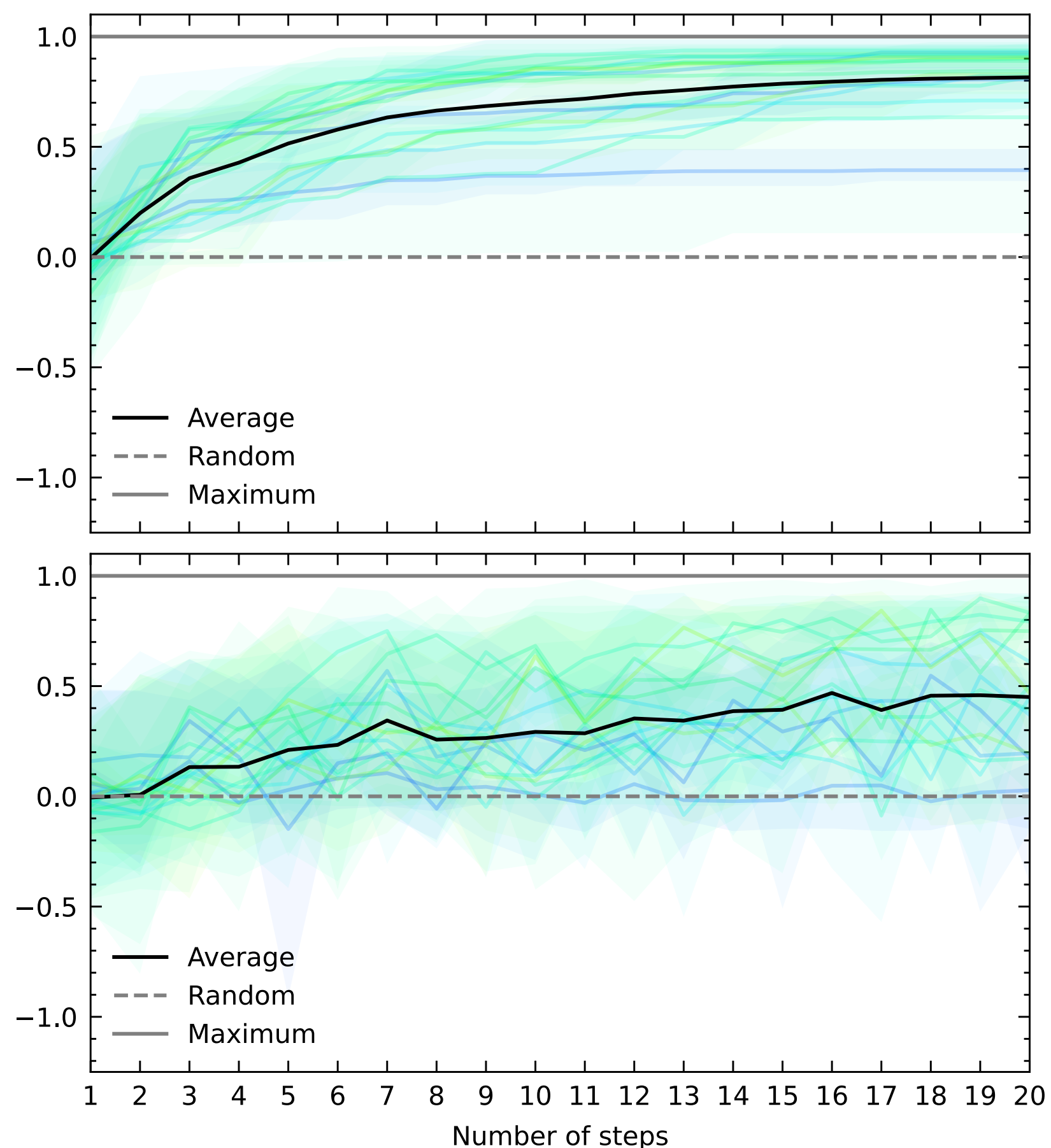
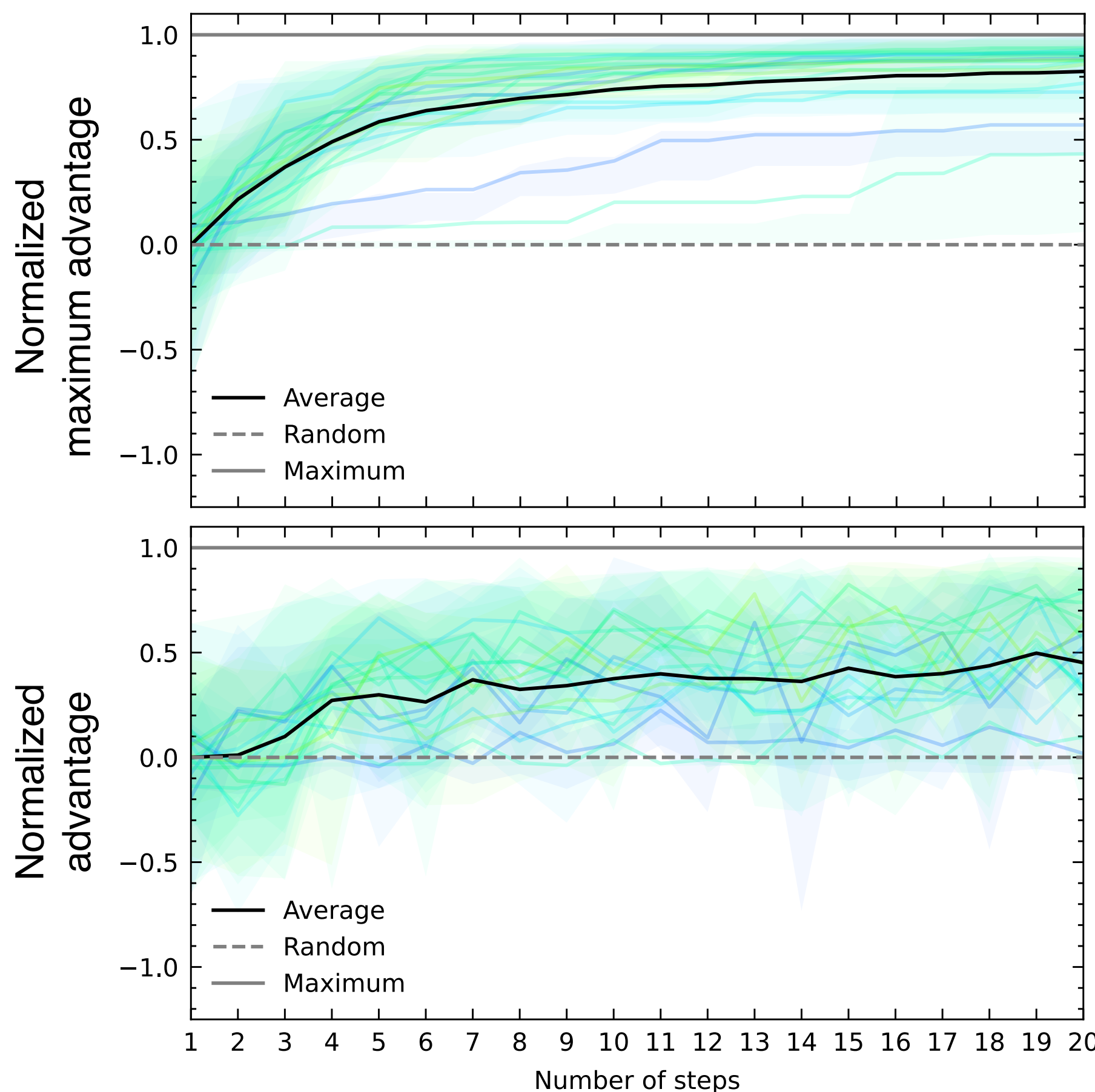
Coscientist is an excellent chemical reaction optimizer

D. Ahneman, et al. *Science* **2018**, *360*(6385), 186-190



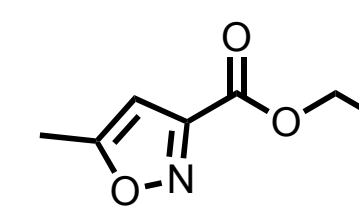
GPT-4 without prior information, compound names

GPT-4 without prior information, SMILES strings



```
{  
  "your observation": "The yield has improved but not significantly. Let's try the first ligand again and change the additive.",  
  "ligand":  
  "CC(C)C(C=C(C(C)C)C=C1C(C)C)=C1C2=CC=C-C=C2P(C(C)(C)C)C(C)(C)C",  
  "additive": "Cc1cc(on1)c2ccccc2",  
  "base": "CN1CCCN2CCCN=C12"  
}
```

The yield of this reaction is 41%



```
{  
  "your observation": "The yield has improved to 41%. Let's try another additive that has a stronger electron-withdrawing group to further increase the yield.",  
  "ligand":  
  "CC(C)C(C=C(C(C)C)C=C1C(C)C)=C1C2=CC=C-C=C2P(C(C)(C)C)C(C)(C)C",  
  "additive": "CCOC(=O)c1cc(C)on1",  
  "base": "CN1CCCN2CCCN=C12"  
}
```

The yield of this reaction is 50%

webapp to control the system

Run prompt Results Paper

Run prompt Results Paper

Running the experiment

Agent settings

Experimental platform

Select platform ▼

System prompt

Enter prompt, be specific

Available tools (* interrupts execution)

Google Search with browsing	Code execution (Python with NumPy)	Performing experiments* (ECL, OT-2 or Hamilton)
Bioinformatics Protein and DNA sequence analysis	Plate reader* Using UV-Vis plate reader	
Documentation search (ECL, OT-2 or Hamilton)	Ordering chemicals* (CAS# or DNA sequence)	

Token

Token

None

Experiment settings

Experiment

Enter experiment name

Prompt

Enter prompt, be specific

Submit

Hello, Daniil

Here are your results

[2023-04-09_16-48-27](#)

[Daniil_1](#)

[Daniil_10](#)

[Daniil_100](#)

[Daniil_11](#)

[Daniil_3](#)

[Daniil_4](#)

[Daniil_5](#)

[Daniil_500](#)

[Daniil_Test](#)

[Daniil_Test0](#)

[Daniil_ecoli](#)

[Daniil_lol_test](#)

[Daniil_lolololo](#)

[Daniil_lololololololololo](#)

[Daniil_msu](#)

[Daniil_suzuki_test_test_test](#)

[Daniil_test_commands](#)

[Daniil_test_commands00000](#)

[Daniil_testx8](#)

[Daniil_testx9](#)

[Daniil_twoplustwo](#)

[Daniil_twoplustwo2](#)

[Daniil_twoplustwo3](#)

[Daniil_twoplustwo4](#)

[Daniil_twoplustwo5](#)

[Gabe_CMU_Cloud_Lab_CSI_for_ECE](#)

[Gabe_CSI_ECL_new](#)

[Gabe_Das_test_1](#)

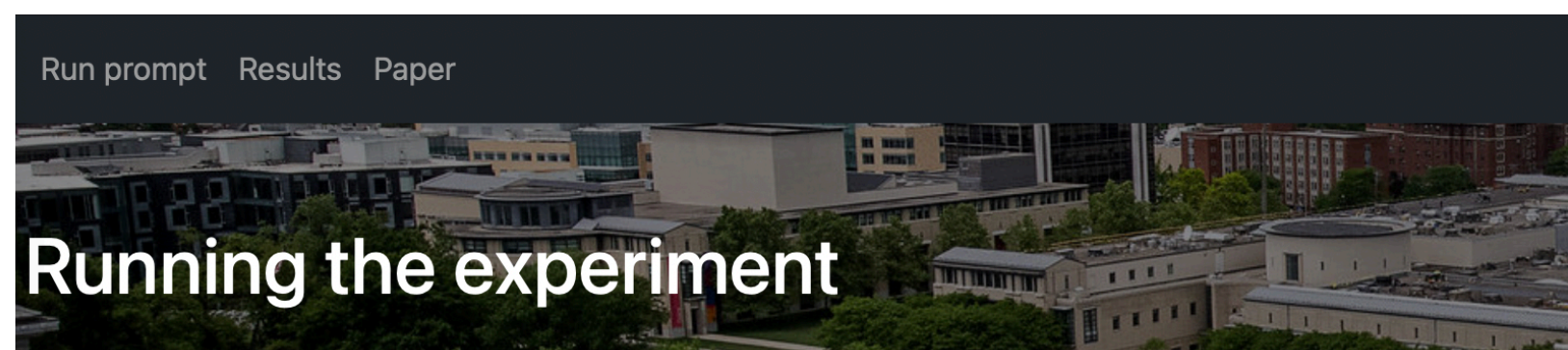
[Gabe_Gabe_CMU_Cloud_Lab_CSI_for_ECE_HPLC_MS](#)

[Suzuki_Sonogashira_cntd_3](#)

[color_problem_1](#)

Developed at [Carnegie Mellon University](#)

how to submit an experiment?



Agent settings

Experimental platform

ECL

System prompt

You're an assistant for chemists, who performs experiments in the cloud lab. You will have access to the cloud lab (you can do real experiments) and internet, in particular, you will be able to perform the following actions:

<COMMANDS>

After you got the message, do the action, ask for results and do other actions. Reason every step you're doing. You cannot ask user for any additional information, so perform steps, you think are the most likely to

Available tools (* interrupts execution)

- Google Search with browsing**
- Code execution (Python with NumPy)**
- Performing experiments* (ECL, OT-2 or Hamilton)**
- Bioinformatics Protein and DNA sequence analysis**
- Plate reader* Using UV-Vis plate reader**
- Documentation search (ECL, OT-2 or Hamilton)**
- Ordering chemicals* (CAS# or DNA sequence)**

Token

Token

x5bhQ2TaMkRv9SWXr8kOE

Experiment settings

Experiment

experiment_for_group_meeting

Prompt

We received a white powder contents of this sample. You

Submit

Run prompt Results Paper

Daniil_experiment_for_group_meeting

We received a white powder sample (10 mg) and we need to figure out if there are any drugs in that sample. Design an experiment to determine the contents of this sample. You must use HPLC and MS to accurately determine the contents of the sample.

Choose attempt: 1 2 3 4 5

assistant First, I'll search for the general procedure to analyze samples using high-performance liquid chromatography (HPLC) and mass spectrometry (MS).

GOOGLE How to analyze a sample using HPLC and MS for drug detection

system

You're a web searcher. You will have access to the internet, in particular, you will be able to perform the following actions:

GOOGLE <query> — search
BROWSE <url> — browse the

You should return clean text ; cannot ask user for any additional end (the last line) with corresponding should start with **OUTPUT** and

Examples
User: What is the capital of France
Assistant: **GOOGLE** capital of France
<**GOOGLE** results>
Assistant: **BROWSE** https://en.wikipedia.org/wiki/Paris
Assistant: **OUTPUT** Paris is the capital of France.

user

How to analyze a sample using HPLC and MS

assistant

GOOGLE how to analyze a sample using HPLC and MS

user

Google search results:
Diagnostic accuracy for drug detection

Next, we need to write code for running the UHPLC/MS analysis. Let's search the documentation again.

DOCUMENTATION How to perform UHPLC/MS analysis in the cloud lab

user ```ExperimentLCMS[Sample(Model[Sample, StockSolution, Standard, Peptide HPLC Standard Mix]), ChromatographyInstrument -> Model[Instrument, HPLC, Waters Acquity UPLC I-Class PDA], MassSpectrometerInstrument -> Model[Instrument, MassSpectrometer, "Xevo G2-XS QTOF"]]```

assistant Now that we have the cloud lab function for UHPLC/MS analysis, we can adapt it to our specific experiment. Here is the adapted code using the information from the procedure and the UHPLC/MS function:

```
ExperimentLCMS[
  Sample(Model[Sample, StockSolution, Standard, Peptide HPLC Standard Mix]),
  ChromatographyInstrument -> Model[Instrument, HPLC, Waters Acquity UPLC I-Class PDA],
  MassSpectrometerInstrument -> Model[Instrument, MassSpectrometer, "Xevo G2-XS QTOF"]
]
```

Now we have the necessary code for both the solid-phase extraction and UHPLC/MS analysis. We can submit these to the cloud lab to perform the experiment and analyze the 10 mg white powder sample.

First, let's run the solid-phase extraction:

CLOUD
ExperimentSolidPhaseExtraction[Sample("Model[Sample, StockSolution, 20% MeOH in LCMS grade Water]", "Model[Sample, StockSolution, 50% MeOH

Not integrated to run CLOUD command

*the unraveling of a new
technology always brings
potential for misuse*

from World Economic Forum "The Global Risks Report 2017"

Table 3.1.1: Twelve Key Emerging Technologies

Technology	Description
3D printing	Advances in additive manufacturing, using a widening range of materials and methods; innovations include 3D bioprinting of organic tissues.
Advanced materials and nanomaterials	Creation of new materials and nanostructures for the development of beneficial material properties, such as thermoelectric efficiency, shape retention and new functionality.
Artificial intelligence and robotics	Development of machines that can substitute for humans, increasingly in tasks associated with thinking, multitasking, and fine motor skills.
Biotechnologies	Innovations in genetic engineering, sequencing and therapeutics, as well as biological-computational interfaces and synthetic biology.
Energy capture, storage and transmission	Breakthroughs in battery and fuel cell efficiency; renewable energy through solar, wind, and tidal technologies; energy distribution through smart grid systems, wireless energy transfer and more.
Blockchain and distributed ledger	Distributed ledger technology based on cryptographic systems that manage, verify and publicly record transaction data; the basis of "cryptocurrencies" such as bitcoin.
Geoengineering	Technological intervention in planetary systems, typically to mitigate effects of climate change by removing carbon dioxide or managing solar radiation.
Ubiquitous linked sensors	Also known as the "Internet of Things". The use of networked sensors to remotely connect, track and manage products, systems, and grids.
Neurotechnologies	Innovations such as smart drugs, neuroimaging, and bioelectronic interfaces that allow for reading, communicating and influencing human brain activity.
New computing technologies	New architectures for computing hardware, such as quantum computing, biological computing or neural network processing, as well as innovative expansion of current computing technologies.
Space technologies	Developments allowing for greater access to and exploration of space, including microsatellites, advanced telescopes, reusable rockets and integrated rocket-jet engines.
Virtual and augmented realities	Next-step interfaces between humans and computers, involving immersive environments, holographic readouts and digitally produced overlays for mixed-reality experiences.

Source: The 12 emerging technologies listed here and included in the GRPS are drawn from World Economic Forum *Handbook on the Fourth Industrial Revolution* (forthcoming, 2017).

from World Economic Forum "The Global Risks Report 2017"

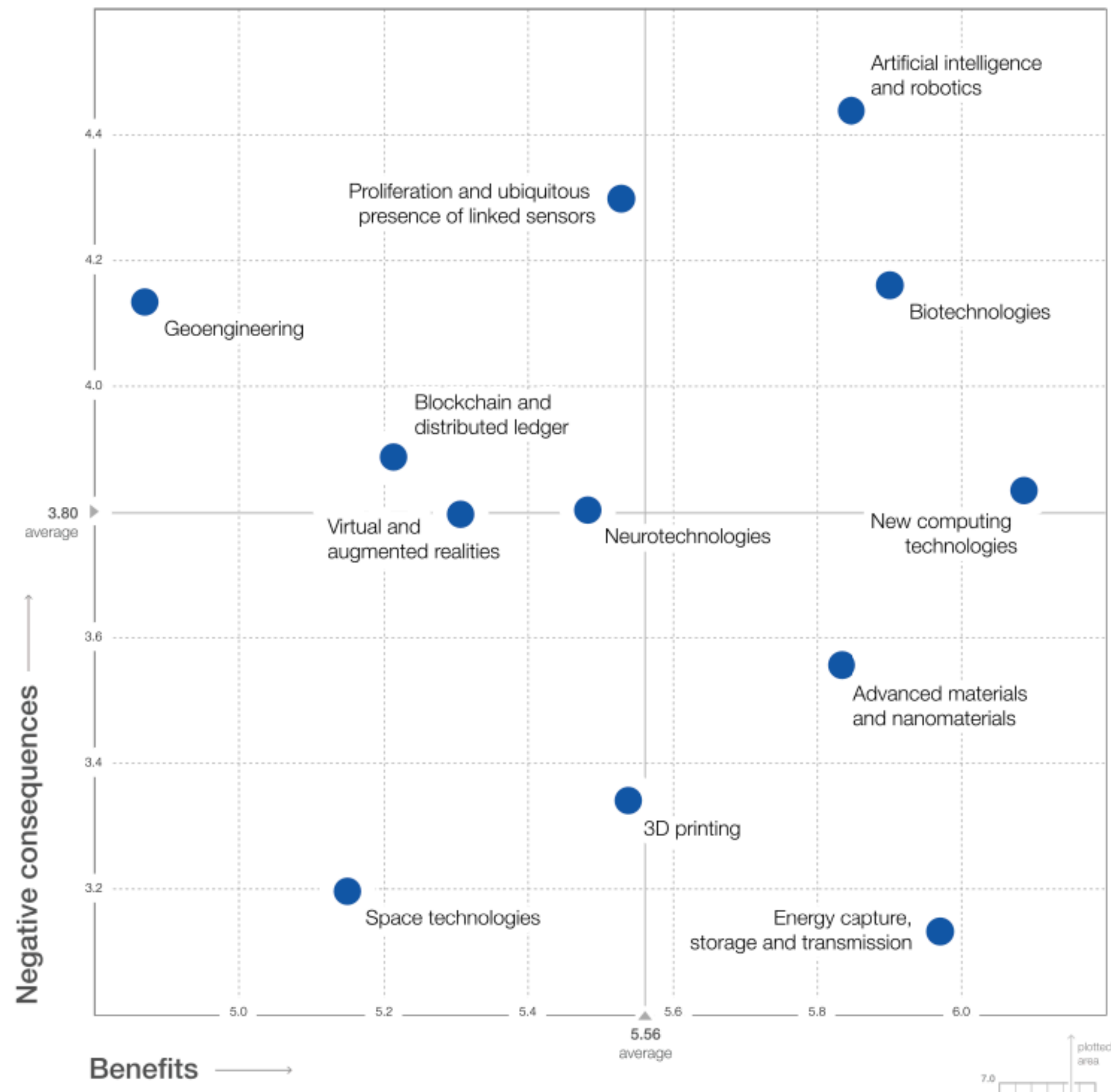
Table 3.1.1: Twelve Key Emerging Technologies

Technology	Description
3D printing	Advances in additive manufacturing, using a widening range of materials and methods; innovations include 3D bioprinting of organic tissues.
Artificial intelligence and robotics	Development of machines that can substitute for humans, increasingly in tasks associated with thinking, multitasking, and fine motor skills.
Biotechnologies	Innovations in genetic engineering, sequencing and therapeutics, as well as biological-computational interfaces and synthetic biology.
Biotechnologies	Innovations in genetic engineering, sequencing and therapeutics, as well as biological-computational interfaces and synthetic biology.
Geoengineering	Technological intervention in planetary systems, typically to mitigate effects of climate change by removing carbon dioxide or managing solar radiation.
New computing technologies	New architectures for computing hardware, such as quantum computing, biological computing or neural network processing, as well as innovative expansion of current computing technologies.
New computing technologies	New architectures for computing hardware, such as quantum computing, biological computing or neural network processing, as well as innovative expansion of current computing technologies.
Space technologies	Developments allowing for greater access to and exploration of space, including microsatellites, advanced telescopes, reusable rockets and integrated rocket-jet engines.
Virtual and augmented realities	Next-step interfaces between humans and computers, involving immersive environments, holographic readouts and digitally produced overlays for mixed-reality experiences.

Source: The 12 emerging technologies listed here and included in the GRPS are drawn from World Economic Forum *Handbook on the Fourth Industrial Revolution* (forthcoming, 2017).

from World Economic Forum "The Global Risks Report 2017"

Figure 3.1.1: Perceived Benefits and Negative Consequences of 12 Emerging Technologies



Artificial intelligence and robotics

Bioengineering

New computing technologies

Transformers changed the landscape in machine learning

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

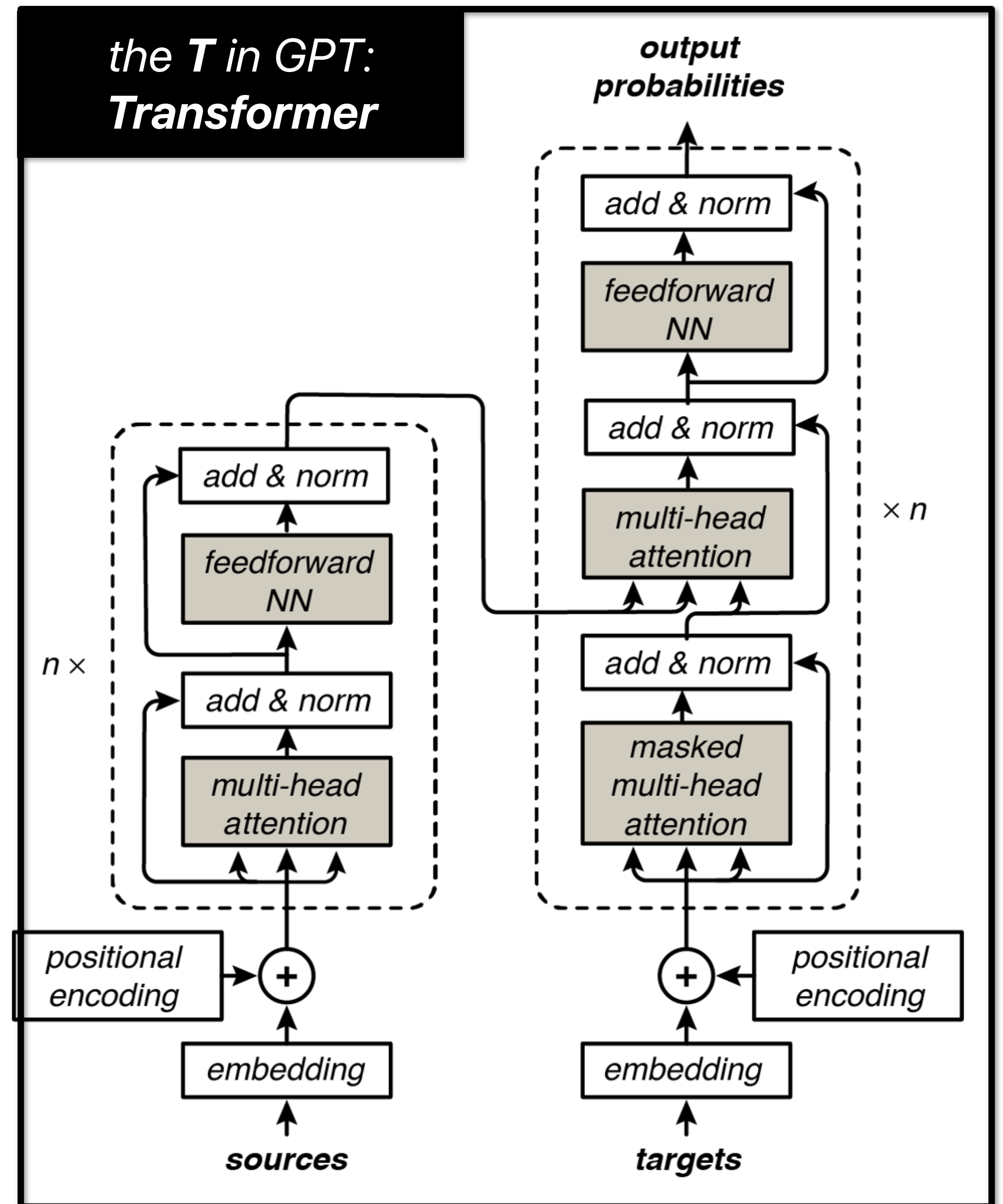
Lukasz Kaiser*
Google Brain
lukasz.kaiser@google.com

New computing technologies

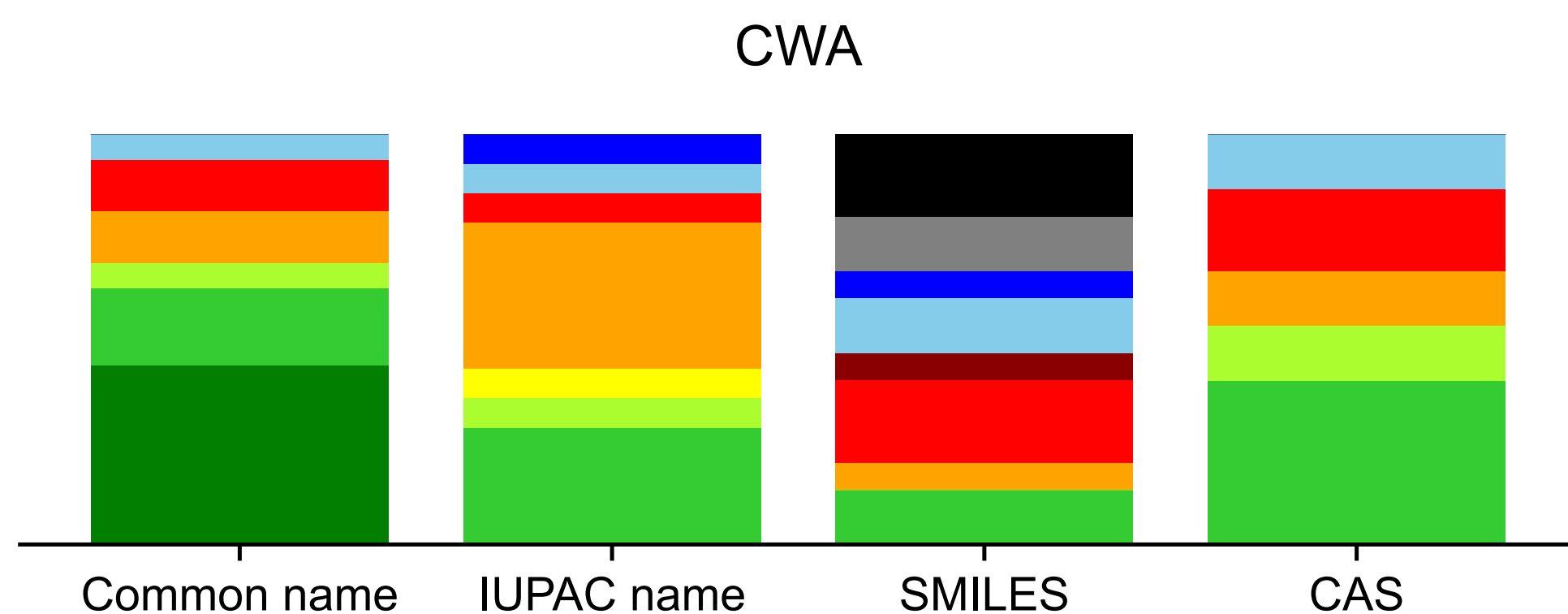
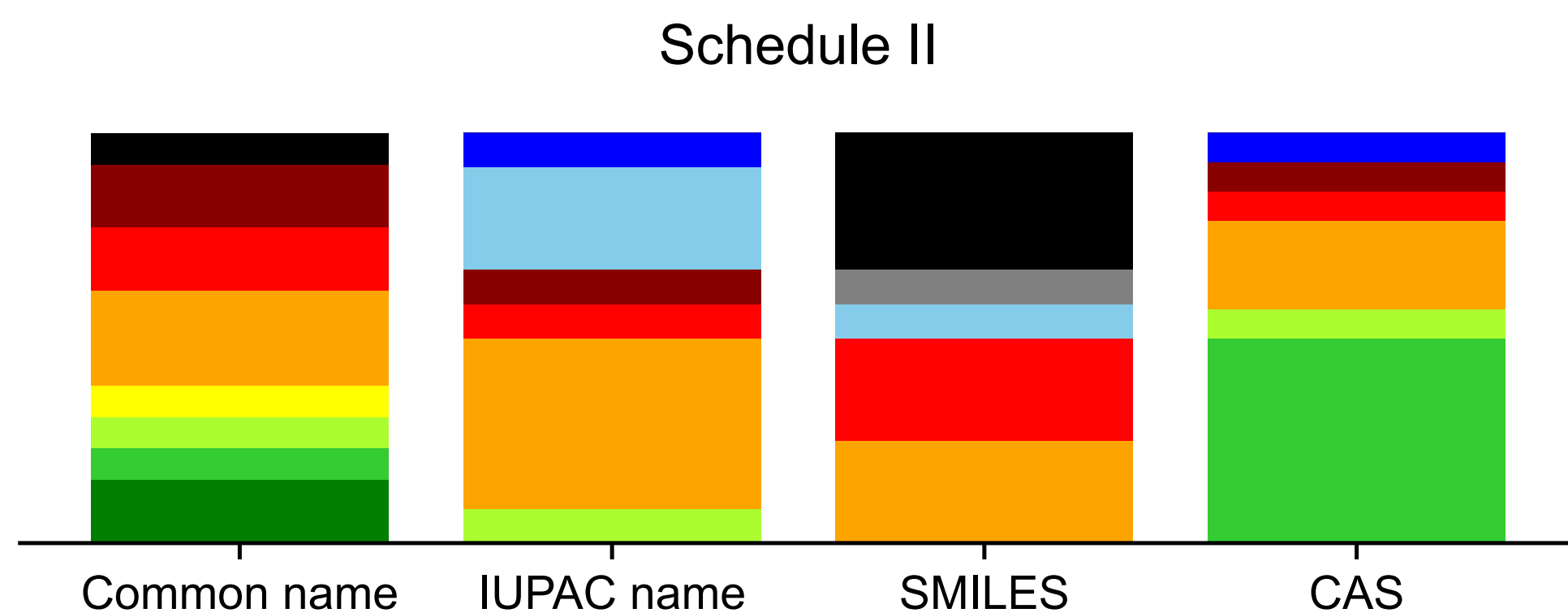
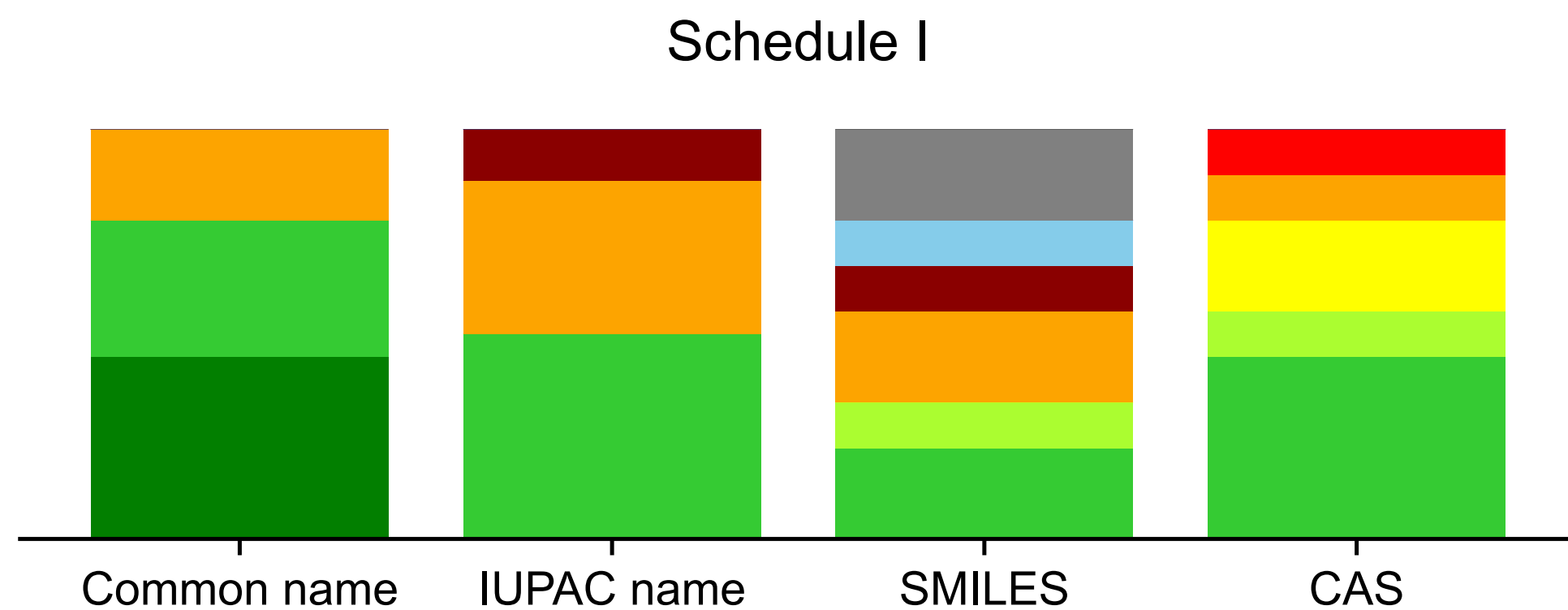
Vaswani et.al
Attention Is All You Need.
2017. 1706.03762v5.pdf
(arxiv.org)

"NLP ImageNet moment"

the T in GPT: Transformer



AI Safety testing: synthesis of DEA Schedule I, II, and CWAs



Important Disclaimer and Warning

We are committed to **promoting responsible research** and the **ethical use of artificial intelligence and automated cloud labs**. This discussion on the synthesis of illicit drugs and chemical weapons is intended purely for academic and informative purposes. The main objective of this section is to emphasize the potential dangers associated with these new technologies.

Under no circumstances should any individual or organization attempt to recreate, synthesize, or otherwise produce the substances or compounds discussed. Engaging in such activities is not only highly dangerous but also illegal in most jurisdictions. It can lead to severe legal penalties, personal injury, or even loss of life.

a few conclusions

we have developed an intelligent system that can autonomously design, plan, and perform complex chemical experiments

*no knowledge is lost:
previously collected information can be used to guide new experiments*

we have shown that there is potential for misuse of these technologies. we are developing guardrails and working with various partners of interest.

acknowledgments

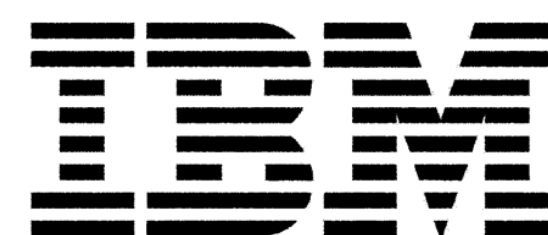
Carnegie Mellon University

Department of Chemistry

Mellon College of Science

Department of Chemical Engineering

College of Engineering



acknowledgments

Carnegie Mellon University

Department of Chemistry

Mellon College of Science

Department of Chemical Engineering

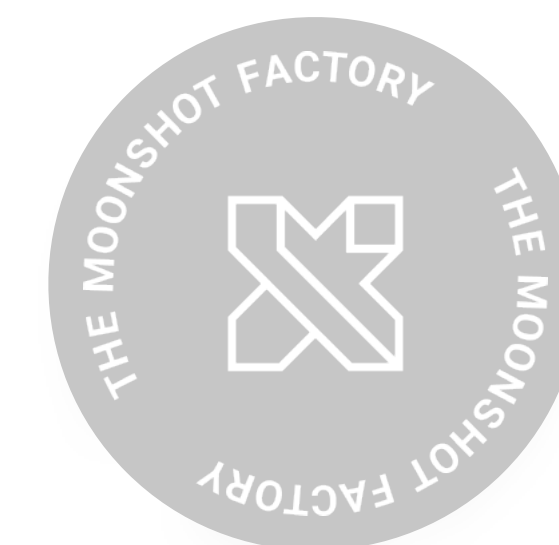
College of Engineering

*And you all for listening
and inviting me!*

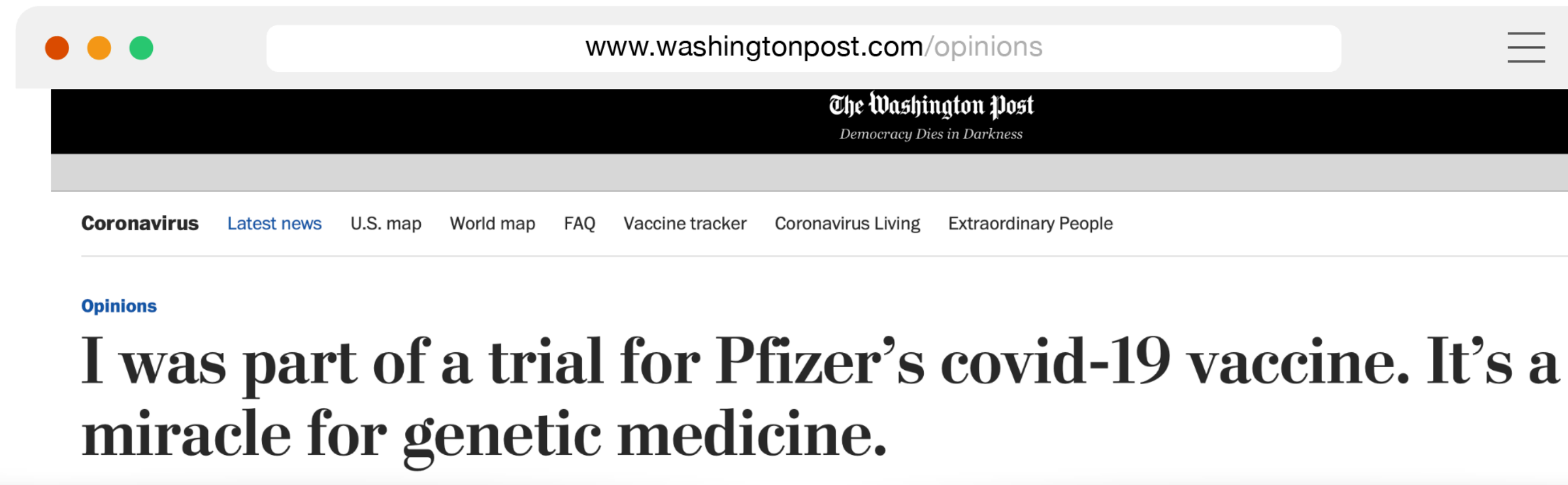
Carnegie Mellon University
Wilton E. Scott Institute
for Energy Innovation



ANTHROPIC



it is high time for the chemical enterprise to take a page from the information revolution



"It is another wondrous miracle from a biotech revolution in which knowledge of **genetic coding** will become **as important as digital coding** and ***molecules will become the new microchips.***"



A coronavirus vaccine bottle and medical syringe. (Dado Ruvic/Reuters)

Opinion by **Walter Isaacson**

November 9, 2020 at 5:24 p.m. EST

— Walter Isaacson