A bit about me

CEMS Chemical Engineering & Materials Science

From New Orleans, LA, USA

B.S. ChemE at Auburn University (2014)

Ph.D. ChemE at University of Colorado (2018)
- thermochemistry, symbolic ML, water splitting

Postdoc MSE at Berkeley (2019-2022)
- batteries, synthesis science, deep learning
- started working with Prof. Miura!

Prof at UMN in ChemE + MSE (2022-present)
- DFT, ML, solid-state chemistry, materials for energy

First trip to Japan (3 days ago – now!)

Objectives of this course

CEMS
Chemical Engineering
& Materials Science

- understand where **ML fits into the broader** Solid-State Chemistry and Materials Science **disciplines**
- understand how to **apply ML algorithms** to problems in these disciplines using the **Python** programming language
- understand how to **formulate, execute, and interpret** ML projects

ML in design of chemicals and materials

CEMS
Chemical Engineering
& Materials Science

Kitchin, John R. "Machine learning in catalysis." *Nature Catalysis* (2018).

FACEBOOK AI Carnegie Mellon University

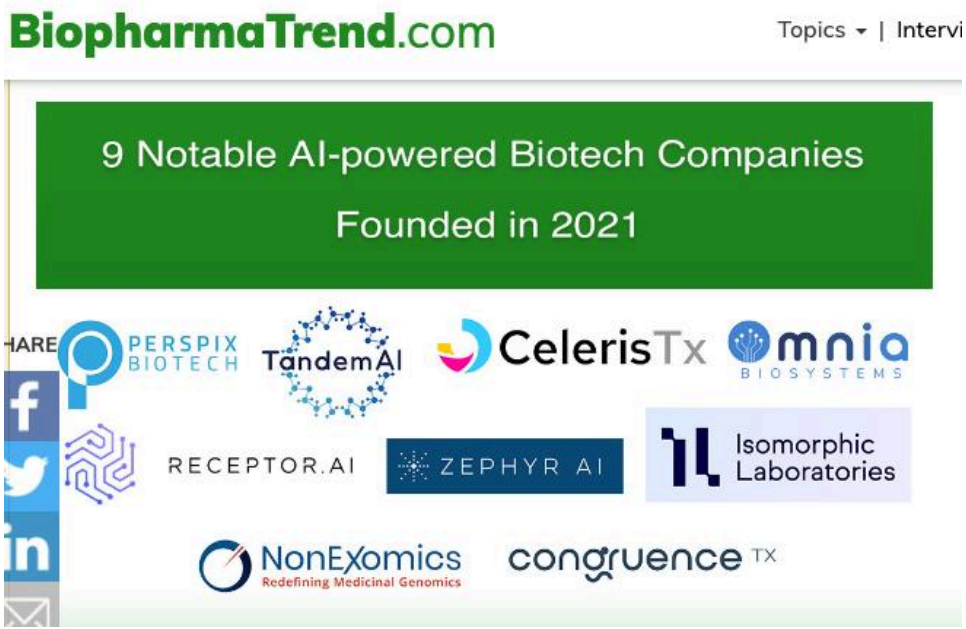
Open Catalyst Project

Using AI to model and discover new catalysts to address the energy challenges posed by climate change.

opencatalystproject.org

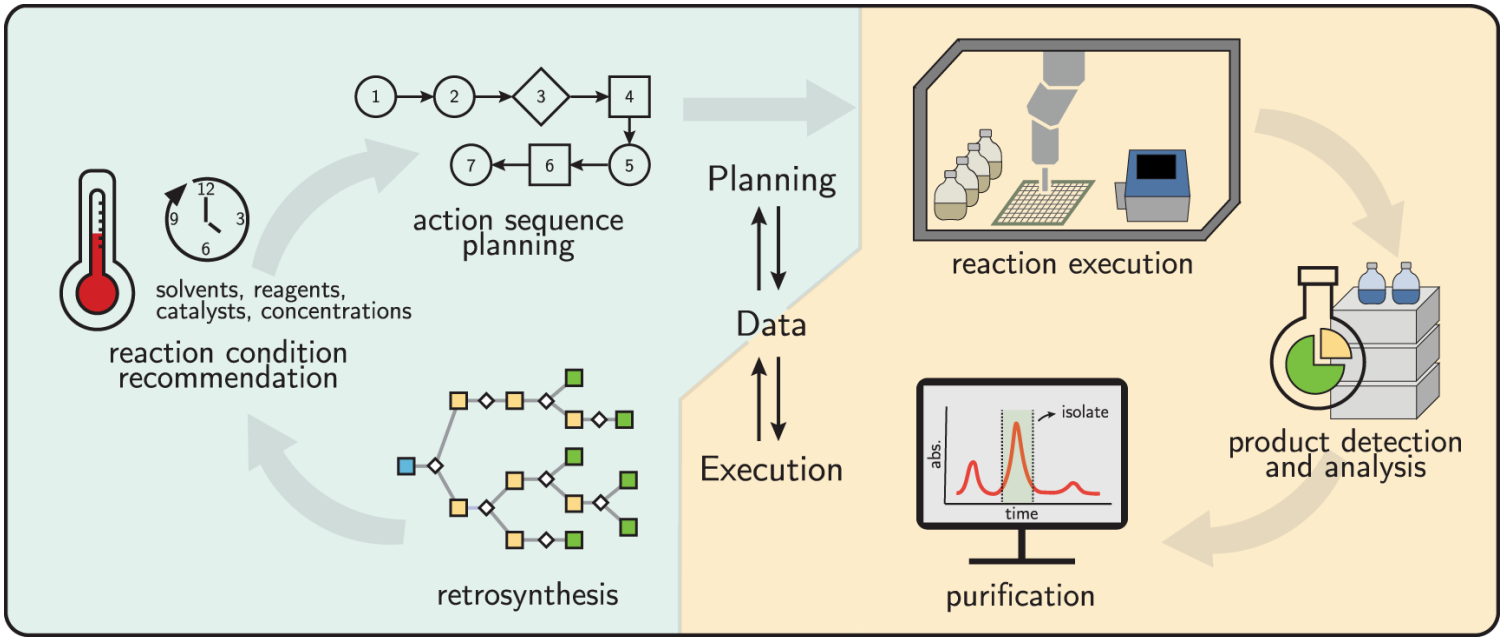
ML in design of chemicals and materials

CEMS Chemical Engineering & Materials Science



Chen, Hongming, *et al.* "The rise of deep learning in drug discovery." *Drug Discovery Today* (2018).

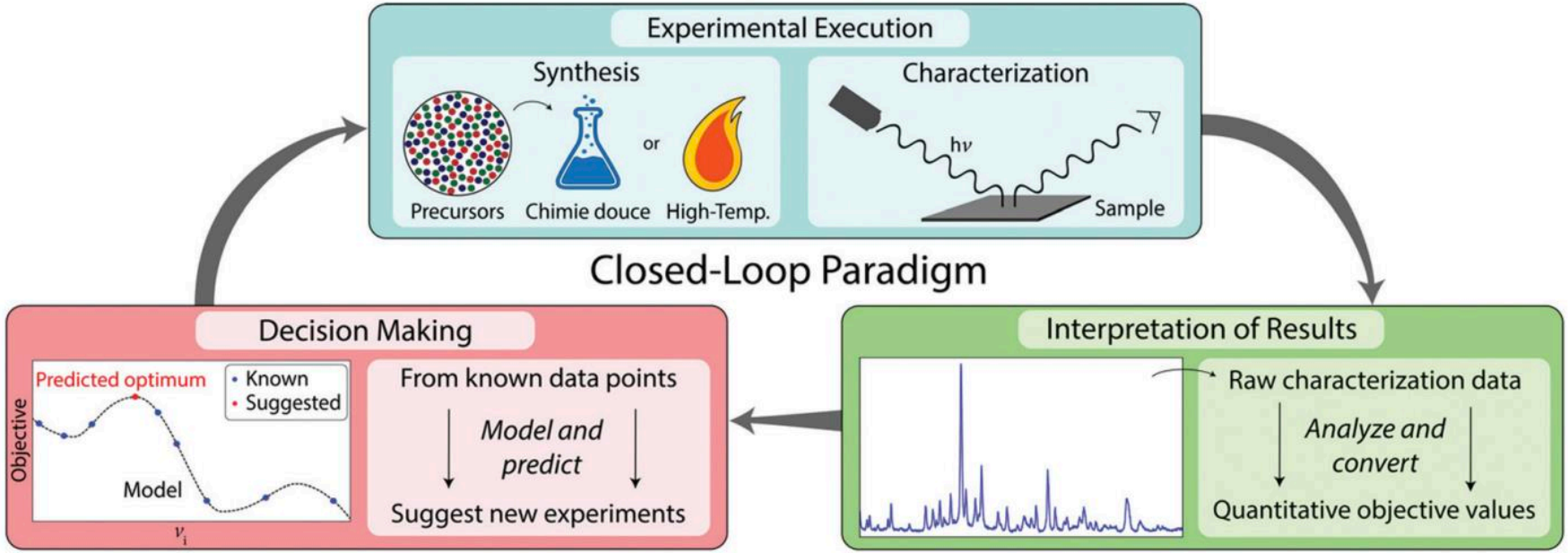
ML in design of chemicals and materials



Gao, Raghavan, Coley. "Autonomous platforms for data-driven organic synthesis." *Nature Communications* (2022).

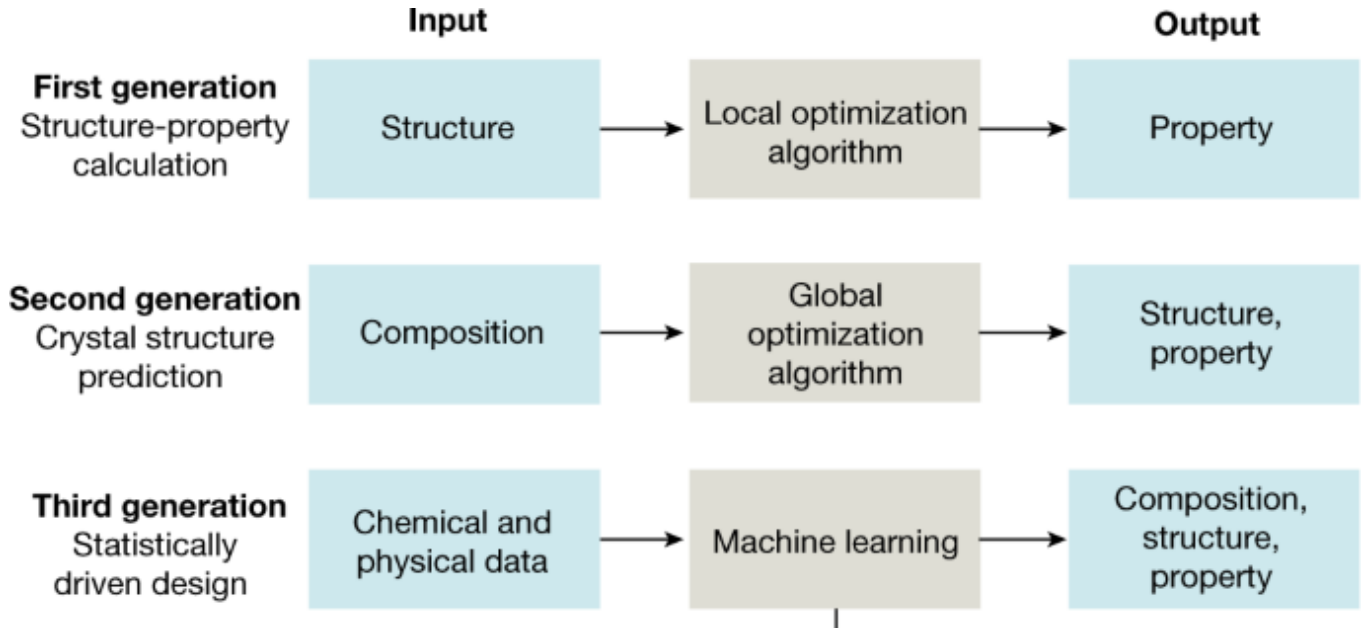
ML in design of chemicals and materials

CEMS
Chemical Engineering
& Materials Science



Szymanski et al. "Toward autonomous design and synthesis of novel inorganic materials." *Materials Horizons* (2021).

ML in design of chemicals and materials



Butler et al. "Machine learning for molecules and materials science." *Nature* (2021).

What is the objective with ML?

CEMS
Chemical Engineering & Materials Science

ML helps us make sense of data quickly

eg predicting outcome of chemical reaction

$A + B \rightarrow ?\ A_2B + B \ldots$

No ML ⇒ use intuition or chemical knowledge

ML ⇒ learn from history to predict outcome

Defining some basic terminology

true prediction = model (features)

$y \approx \hat{y} = f(\vec{X})$ $\quad \vec{X} = \begin{bmatrix} T \\ P \end{bmatrix}$

eg yield $= f(T, P) = 3.1T + 2.4P$

* learned during training

1) Make observations
(features, outcome)
$(T_1, P_1) \rightarrow y_1$
$(T_2, P_2) \rightarrow y_2$

2) Train a model, f to learn $\vec{X} \longleftrightarrow y$

Challenges of ML for science?

CEMS
Chemical Engineering
& Materials Science

Given what you know now, discuss w/ one another some challenges of applying ML for science and engineering applications

Challenges of ML for science?

CEMS Chemical Engineering & Materials Science

Letter | Published: 11 September 2019

Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis

Xiwen Jia, Allyson Lynch, Yuheng Huang, Matthew Danielson, Immaculate Lang'at, Alexander Milder, Aaron E. Ruby, Hao Wang, Sorelle A. Friedler ✉, Alexander J. Norquist ✉ & Joshua Schrier ✉

Nature **573**, 251–255 (2019) | Cite this article

Some reactants are popular, but they aren't as good as random reactants!

Challenges of ML for science?

CEMS
Chemical Engineering
& Materials Science

Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling

Wiktor Beker, Rafał Roszak, Agnieszka Wołos, Nicholas H. Angello, Vandana Rathore, Martin D. Burke*, and Bartosz A. Grzybowski*

Cite this: *J. Am. Chem. Soc.* 2022, 144, 11, 4819–4827
Publication Date: March 8, 2022

Article Views	Altmetric	Citations	Share	Add
26644	83	9		

Similar story in organic chemistry…

Challenges of ML for science?

CEMS
Chemical Engineering
& Materials Science

18 Nov 2022 in Research & Technology

Cuprate superconductivity mechanism may be coming into focus

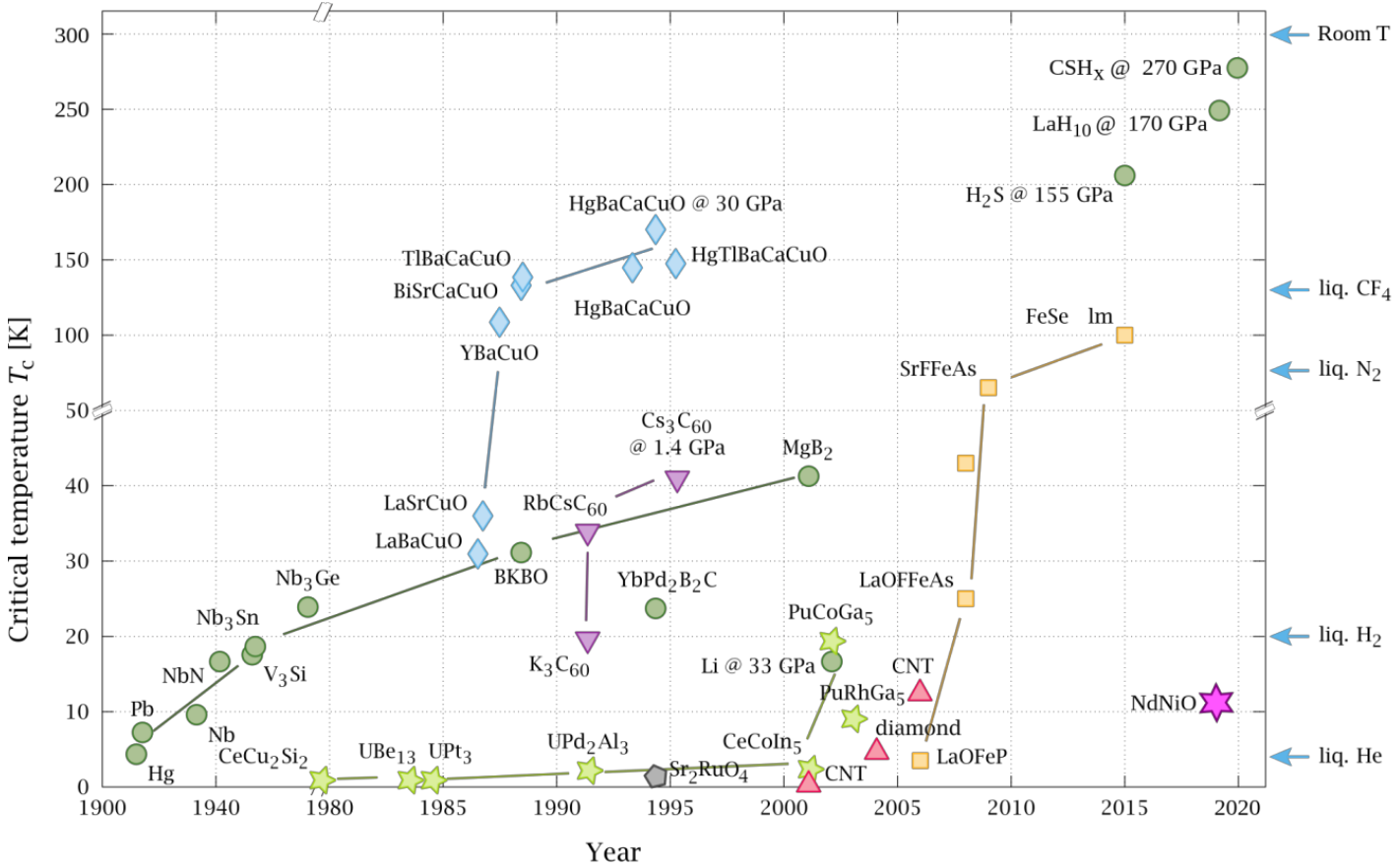
Theory and experiment have pinpointed factors that determine the magnetic attraction between electron pairs.

Daniel Garisto

Since their dramatic debut in 1986, cuprate superconductors have been some of the best-studied materials in existence. Nonetheless, many mysteries about the materials have

"Even if we are correct, in 30 years people will still say that the theory is not understood," Tremblay says.

Challenges of ML for science?



Challenges of ML for science?

CEMS
Chemical Engineering & Materials Science

1) Limited reliable data (bias, noise)
2) Complex phenomena (what features?)
3) Extrapolation matters (what's better?)
4) We care about "why" (interpretability)

Success depends on:
1) how much (good) data you collect.
2) how good your features are.

But it can work!!

CEMS
Chemical Engineering
& Materials Science

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

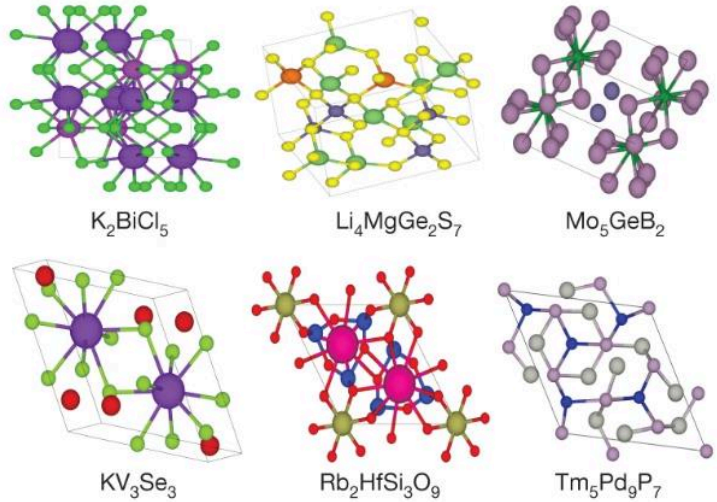
Article | Open access | Published: 29 November 2023

Scaling deep learning for materials discovery

Amil Merchant ✉, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon & Ekin Dogus Cubuk ✉ ← PhD Physics

Nature **624**, 80–85 (2023) | Cite this article

5 Citations | **718** Altmetric | Metrics



But it can work!!

CEMS
Chemical Engineering
& Materials Science

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > articles > article

Article | Open access | Published: 29 November 2023

An autonomous laboratory for the accelerated synthesis of novel materials

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel, Kristin Persson, Yan Zeng ✉ & Gerbrand Ceder ✉

Nature **624**, 86–91 (2023) | Cite this article

6 Citations | **996** Altmetric | Metrics

UMN CEMS postdoc!

PhD Materials Science

But it can work!!

CEMS
Chemical Engineering
& Materials Science

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

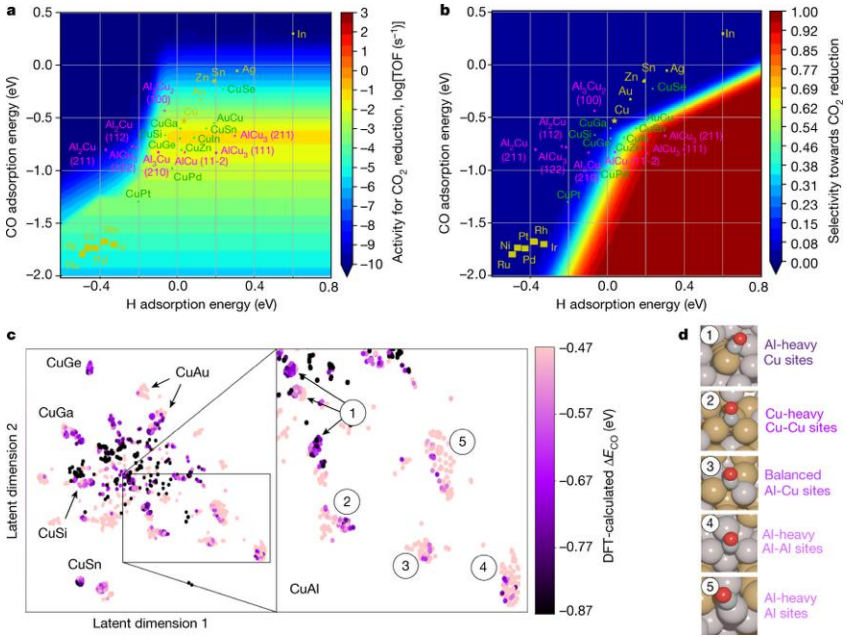
nature › articles › article

Article | Published: 13 May 2020

Accelerated discovery of CO_2 electrocatalysts using active machine learning

Miao Zhong, Kevin Tran, Yimeng Min, Chuanhao Wang, Ziyun Wang, Cao-Thang Dinh, Phil De Luna, Zongqian Yu, Armin Sedighian Rasouli, Peter Brodersen, Song Sun, Oleksandr Voznyy, Chih-Shan Tan, Mikhail Askerka, Fanglin Che, Min Liu, Ali Seifitokaldani, Yuanjie Pang, Shen-Chuan Lo, Alexander Ip, Zachary Ulissi ✉ & Edward H. Sargent ✉

Nature **581**, 178–183 (2020) | Cite this article



PhD Chemical Engineering

But it can work!!

CEMS
Chemical Engineering
& Materials Science

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

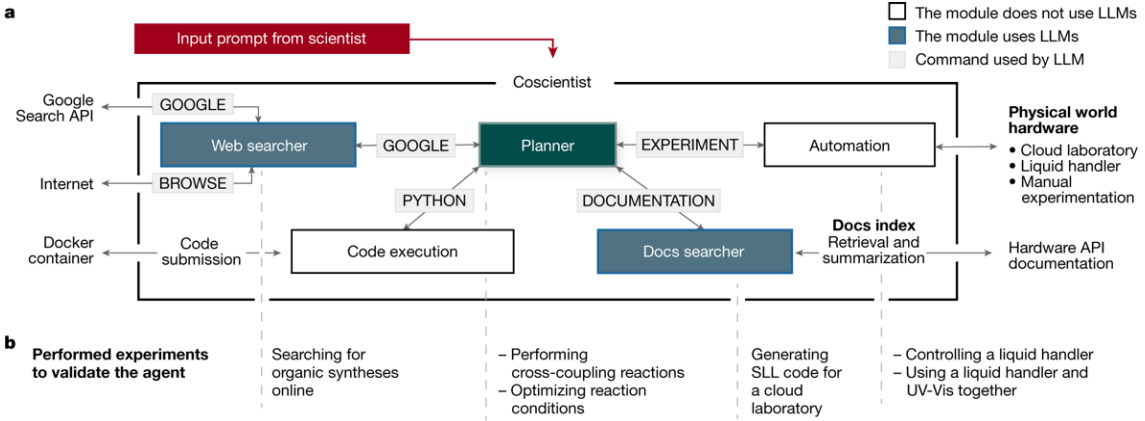
Article | Open access | Published: 20 December 2023

Autonomous chemical research with large language models

Daniil A. Boiko, Robert MacKnight, Ben Kline & Gabe Gomes ✉

Nature **624**, 570–578 (2023) | Cite this article

PhD Chemistry



But it can work!!

CEMS
Chemical Engineering
& Materials Science

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Published: 16 January 2025

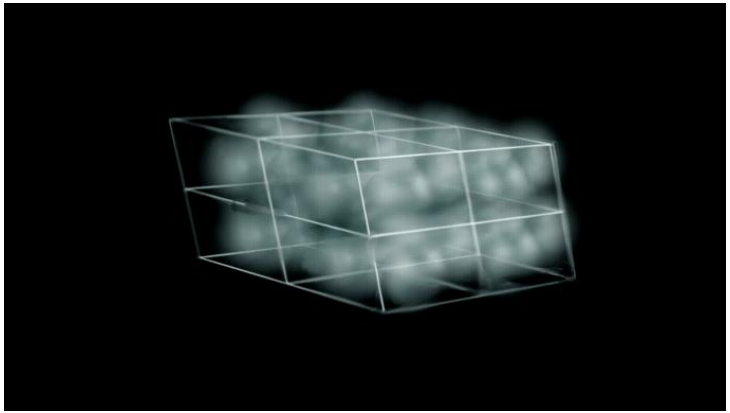
A generative model for inorganic materials design

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, Roberto Sordillo, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Chunlei Yang, Wenjie Li, ... Tian Xie ✉ + Show authors

Nature (2025) | Cite this article

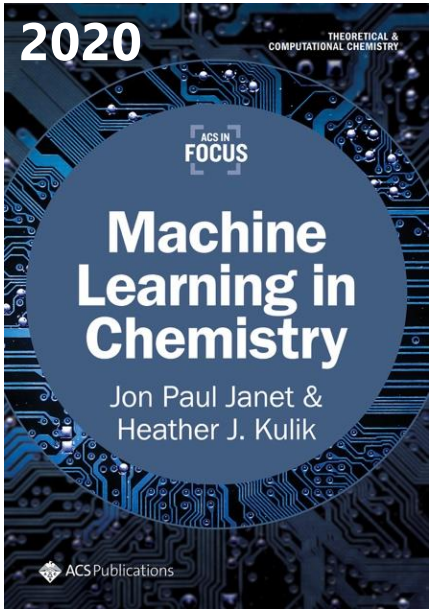
26k Accesses | **121** Altmetric | Metrics

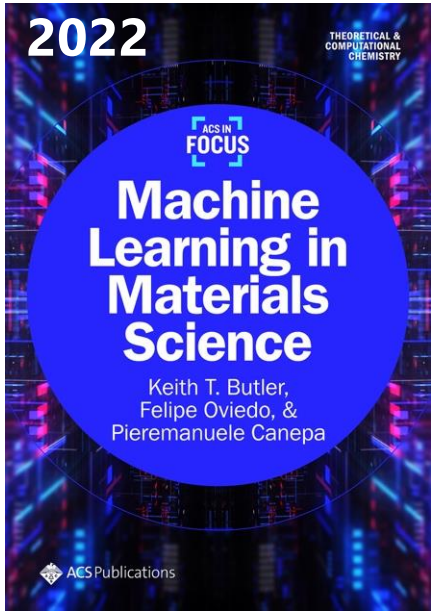
PhD Materials Science

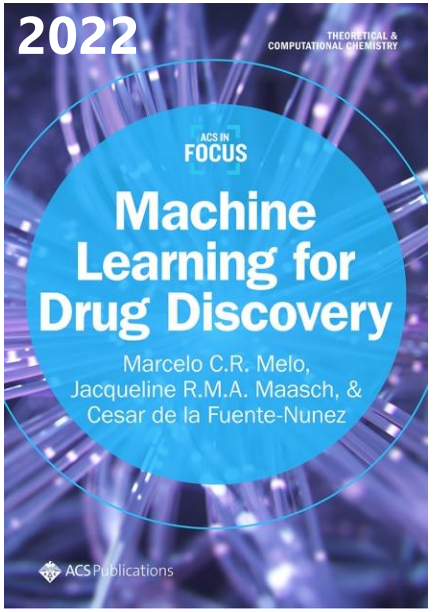


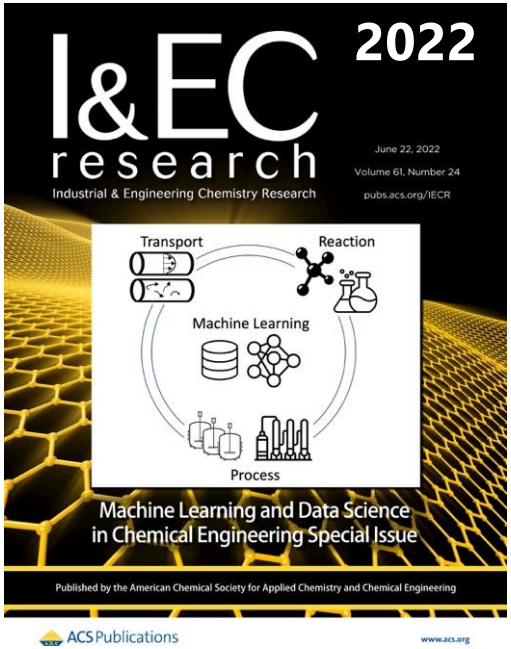
Machine learning interest on the rise!

CEMS
Chemical Engineering
& Materials Science









You do not need a CS degree to make meaningful contributions with ML!