

“Machine learning” – what comes to mind?

JOIN CHIME

Go to:

<https://chimein2.cla.umn.edu/join/608843>

Or: visit chimein2.cla.umn.edu and enter
608-843

Data Science vs AI vs ML

Data science → extract insights from data

[Machine learning → make predictions w/o explicit programming
AI → systems that automatically do human-like things

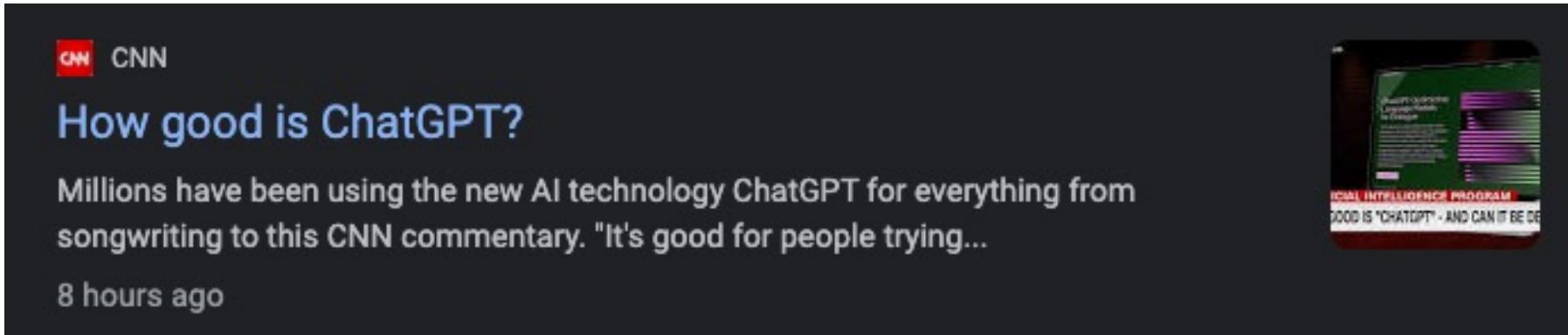
“Large language model” enters the lexicon

CNN CNN

How good is ChatGPT?

Millions have been using the new AI technology ChatGPT for everything from songwriting to this CNN commentary. "It's good for people trying..."

8 hours ago



"Large language model" enters the lexicon

arXiv > cs > arXiv:2005.14165

← → C



nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html

Computer Science > Computation and Language

[Submitted on 28 May 2020 (v1), last revised 22 Jul 2020 (this version, v4)]

Language Models are Few-Shot Learners



Meet GPT-3. It Has Learned to Code (and Blog and Argue).

The latest natural-language system generates tweets, pens poetry, summarizes emails, answers trivia questions, translates languages and even writes its own computer programs.

Let's try it out!

<https://chat.openai.com/>

ML advances not without controversy

NBC News

ChatGPT used by mental health tech app in AI experiment with users

ChatGPT was used by the emotional support chat app Koko to write responses for real users. Controversy came quickly after the experiment was...

9 hours ago



ML advances not without controversy

 The Washington Post

Analysis | Our Future Artificial Intelligence Overlords Need a ... Resistance Movement

Artificial intelligence has been moving so fast that even the scientists are finding it hard to keep up. In the past year, machine learning...

47 mins ago



ML advances not without controversy

G Gizmodo

AI Image Generators Routinely Display Gender and Cultural Bias

AI Image Generators Routinely Display Gender and Cultural Bias ... Artificial intelligence researcher for Hugging Face, Sasha Luccioni,...

2 days ago

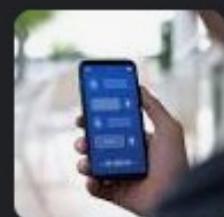


H The Hill

Common AI language models show bias against people with disabilities: study

New research underscores the implicit bias present in some artificial intelligence language models. Researchers found models were generally...

3 weeks ago



And challenges!

 Bloomberg.com

Even After \$100 Billion, Self-Driving Cars Are Going Nowhere

The first car woke Jennifer King at 2 a.m. with a loud, ... who studies artificial intelligence and the limits of self-driving vehicles.

1 month ago



 The Washington Post

MIT scientists find self-driving cars might be harmful to the climate

MIT researchers found that the computers powering self-driving cars could have a large carbon footprint.

2 days ago



But also successes!

Predicted the spread of the winter flu outbreak in 2009 in the USA down to the state level.

Search query topics that were highly correlated with flu outbreaks

- Influenza complication
- Cold/flu remedy
- General influenza symptoms
- Term for influenza
- Specific influenza symptom
- Symptoms of an influenza complication
- Antibiotic medication
- General influenza remedies
- Symptoms of a related disease
- Antiviral medication
- Related disease

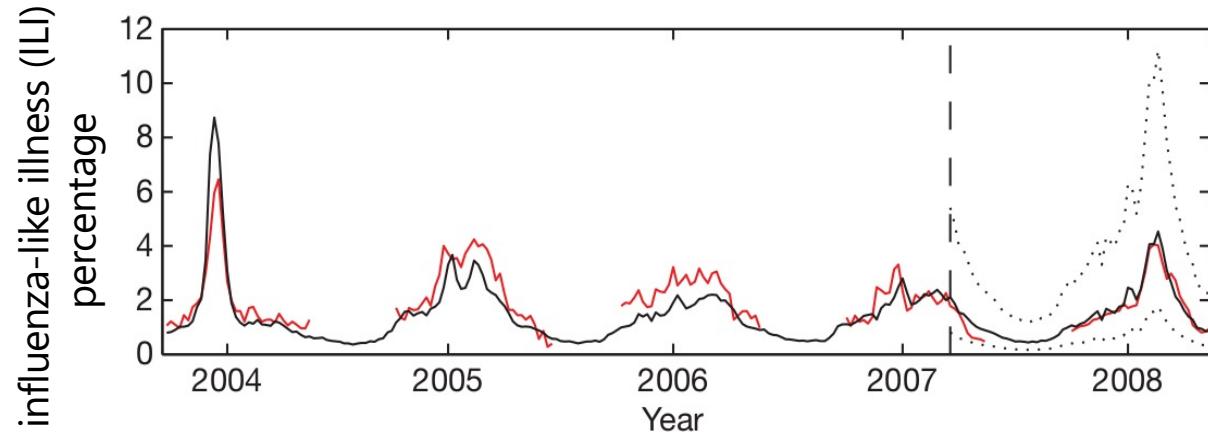
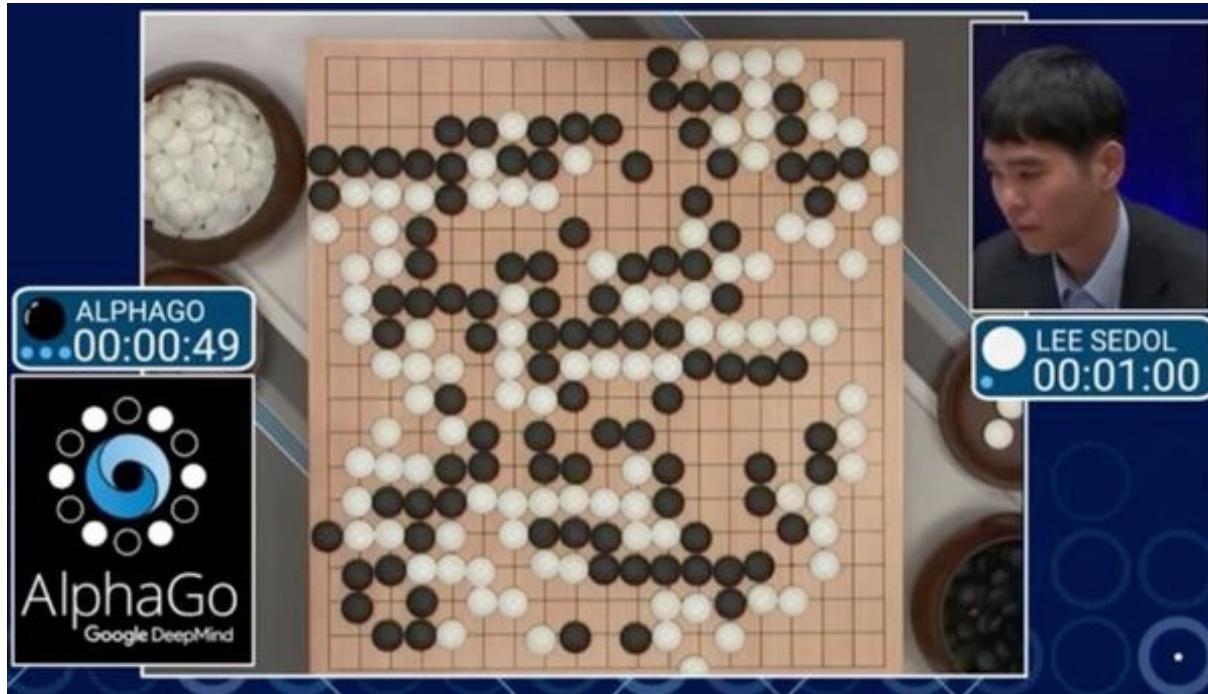


Figure 2 | A comparison of model estimates for the mid-Atlantic region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated.

Google took 50 million of the most common searches and compared them to the CDC data on the spread of the winter flu from 2003 to 2008. Ginsberg *et al.* Nature (2009).

But also successes!



<https://deepmind.com/research/alphago/>; doi:10.1038/nature16961 (also on Netflix!)

First ML for games → Samuel, Arthur L. "Some studies in machine learning using the game of checkers." *IBM Journal of research and development* 3.3 (1959): 210-229.

But also successes!

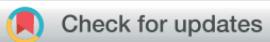


But also successes!

PERSPECTIVE

[View Article Online](#)

[View Journal](#) | [View Issue](#)



Cite this: *Digital Discovery*, 2022, 1, 79

Natural language processing models that automate programming will transform chemistry research and teaching†

Glen M. Hocky *^a and Andrew D. White *^b

```
"""
Compute the dissociation curve of H2 using the pyscf library
"""
```

But also successes!

```
"""
Compute the dissociation curve of H2 using the pyscf library
"""

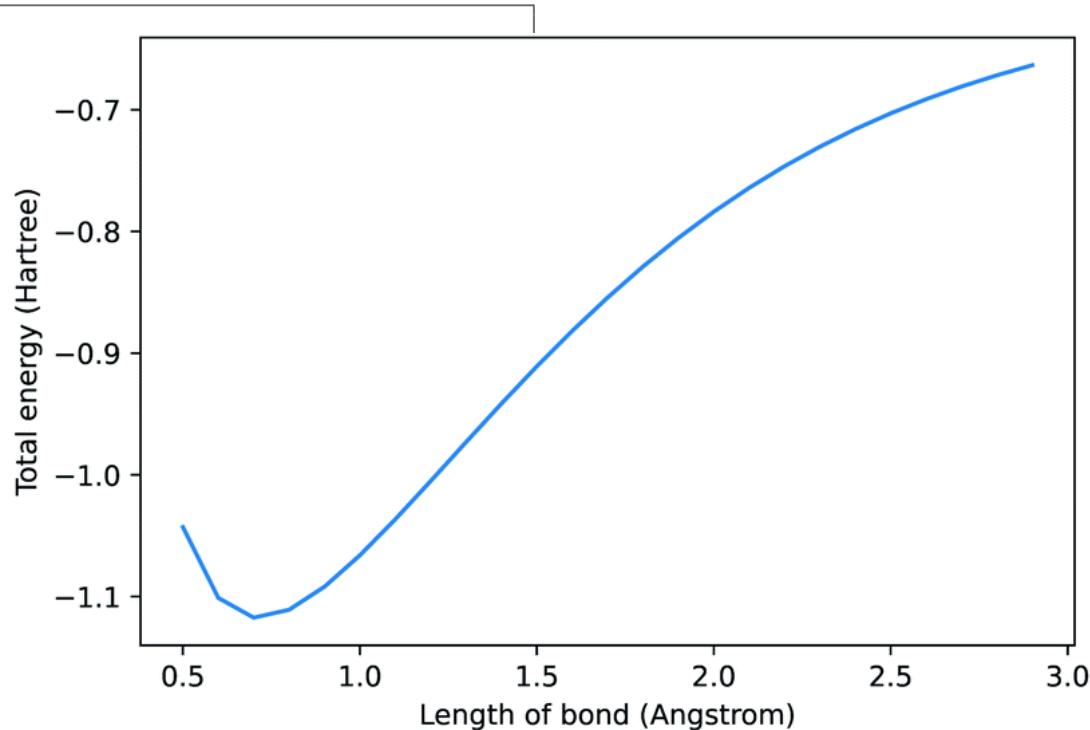
import numpy as np
import pyscf
from pyscf import gto, scf, dft

# Define the molecule
mol = gto.M(atom='H 0 0 0; H 0 0 1.4', basis='sto-3g', verbose=3)

# Compute the reference energy
en_ref = mol.energy_nuc()

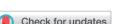
# Compute the dissociation curve
r = np.linspace(0.5, 2.0, 100)
en = []
for b in r:
    mol.atom = [['H', (0, 0, 0)], ['H', (0, 0, b)]]
    mol.build()
    mf = scf.RHF(mol)
    mf.kernel()
    en.append(mf.e_tot - en_ref)

# Plot the dissociation curve
import matplotlib.pyplot as plt
plt.plot(r, en)
plt.xlabel('Bond length (Bohr)')
plt.ylabel('Energy (Hartree)')
plt.show()
```



PERSPECTIVE

[View Article Online](#)
[View Journal](#) | [View Issue](#)



Cite this: *Digital Discovery*, 2022, 1, 79

Natural language processing models that automate programming will transform chemistry research and teaching†

Glen M. Hocky ^{a*} and Andrew D. White ^{ab}

But also successes!

Article | **Open Access** | Published: 05 October 2022

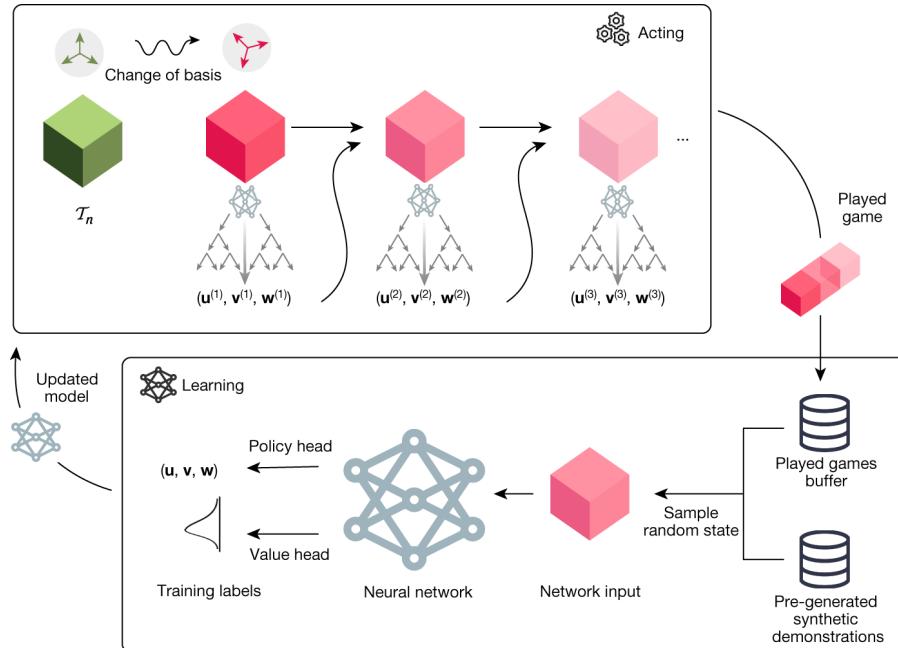
Discovering faster matrix multiplication algorithms with reinforcement learning

[Alhussein Fawzi](#)  [Matej Balog](#), [Aja Huang](#), [Thomas Hubert](#), [Bernardino Romera-Paredes](#), [Mohammadamin Barekatain](#), [Alexander Novikov](#), [Francisco J. R. Ruiz](#), [Julian Schrittwieser](#), [Grzegorz Swirszcz](#), [David Silver](#), [Demis Hassabis](#) & [Pushmeet Kohli](#)

[Nature](#) **610**, 47–53 (2022) | [Cite this article](#)

460k Accesses | **3** Citations | **3642** Altmetric | [Metrics](#)

the-art complexity for many matrix sizes. Particularly relevant is the case of 4×4 matrices in a finite field, where AlphaTensor's algorithm improves on Strassen's two-level algorithm for the first time, to our knowledge, since its discovery 50 years ago². We further showcase the



Faster matrix multiplication → less computational time for trillions of calculations...
Fawzi et al. Nature (2022).

How about in the Computational Molecular Sciences?

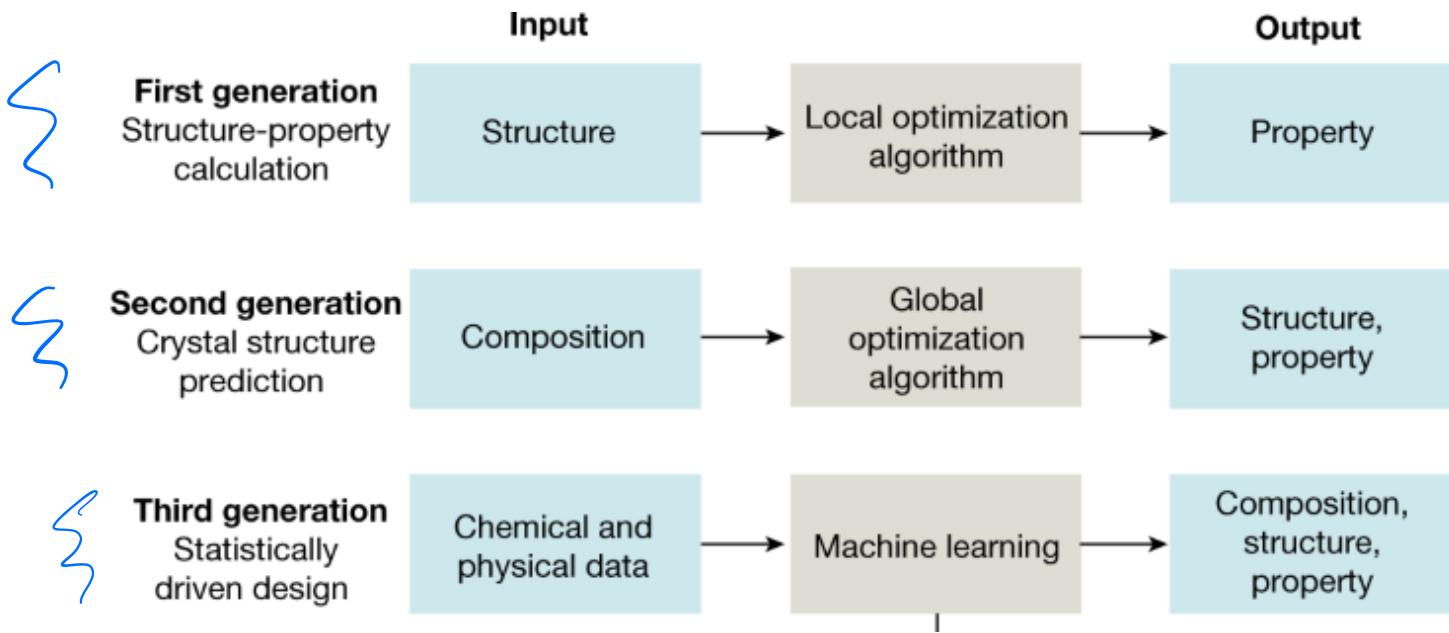
JOIN CHIME

Go to:

<https://chimein2.cla.umn.edu/join/608843>

Or: visit chimein2.cla.umn.edu and enter
608-843

ML in Computational Molecular Sciences



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

ML in Computational Molecular Sciences



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 Computational Molecular Sciences

BiopharmaTrend.com

Topics ▾ | Interview

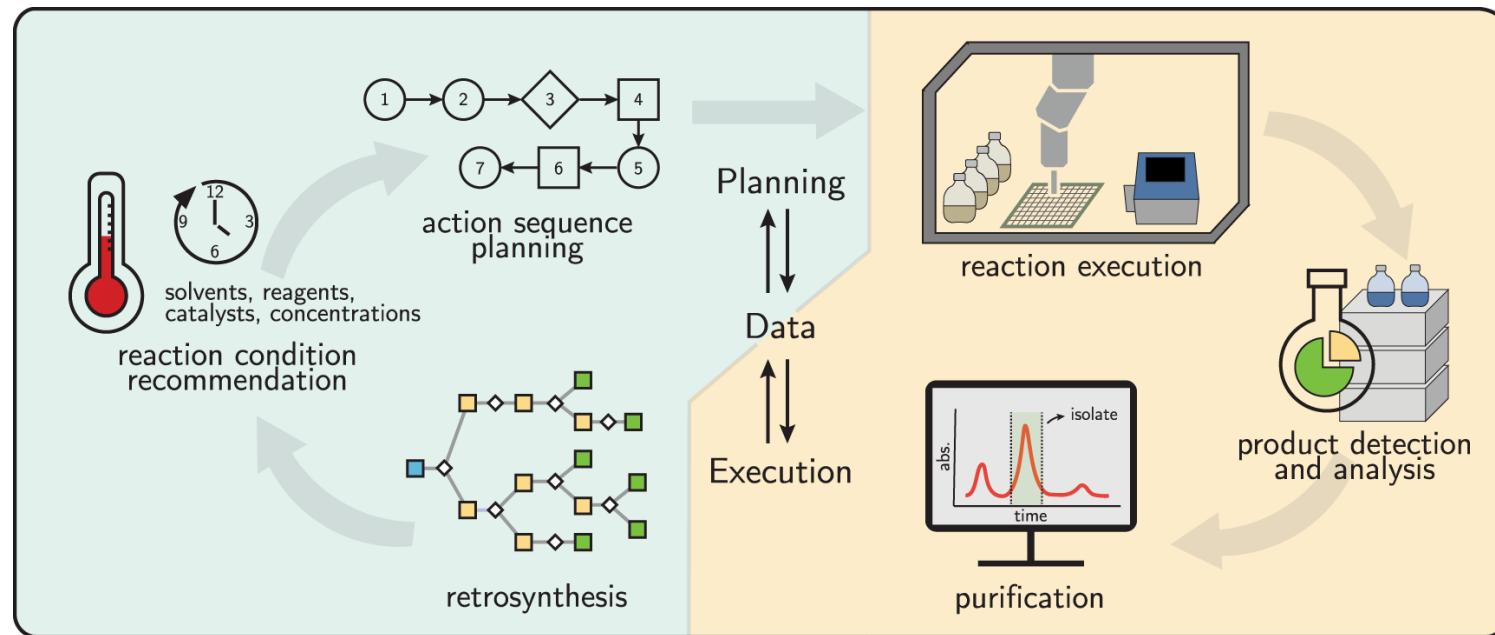
9 Notable AI-powered Biotech Companies

Founded in 2021



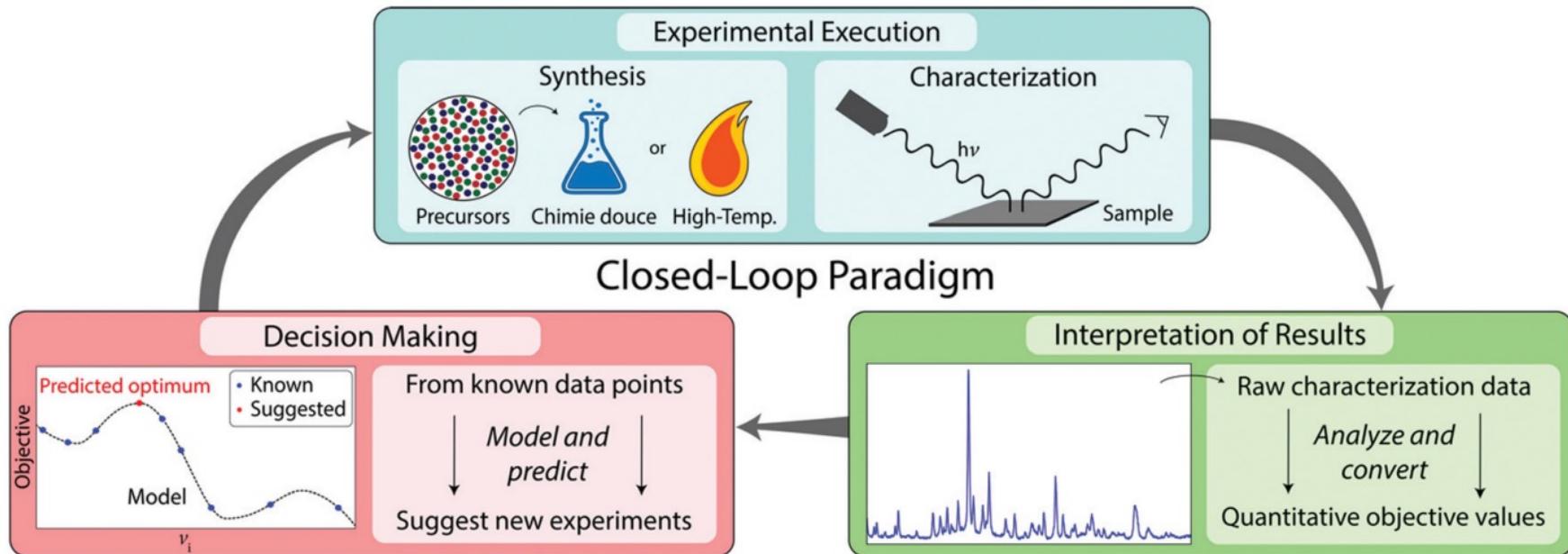
Chen, Hongming, et al. "The rise of deep learning in drug discovery." *Drug Discovery Today* (2018).

ML in Computational Molecular Sciences



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

ML in Computational Molecular Sciences



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

What is the objective with ML?

ML helps us make sense of data more quickly

Predicting outcomes of chemical reaction:

- 1) Use our intuition (or collective intuition)
- 2) v/ML = use trained model
- 3) success = trained model better than my intuition

Defining some basic terminology

$\text{prediction} = \text{model}(\text{features})$

observations : $[(\text{features}, \text{outcome}), (\text{features}, \text{outcome})\dots]$

Model learns to anticipate outcome based
on the changes to features

Success is

- f (1) how much good data we collect
- 2) how expressive our features are

Challenges of ML for science?

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

 JOIN CHIME

Go to:

<https://chimein2.cla.umn.edu/join/608843>

Or: visit chimein2.cla.umn.edu and enter
608-843

Challenges of ML for 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](#)

7799 Accesses | **95** Citations | **68** Altmetric | [Metrics](#)

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

Challenges of ML for 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

26644

Altmetric

83

Citations

9

Share

Add



Similar story in organic chemistry...

Challenges of ML for science?

18 Nov 2022 in Research & Technology

Cuprate superconductivity mechanism may be coming into focus

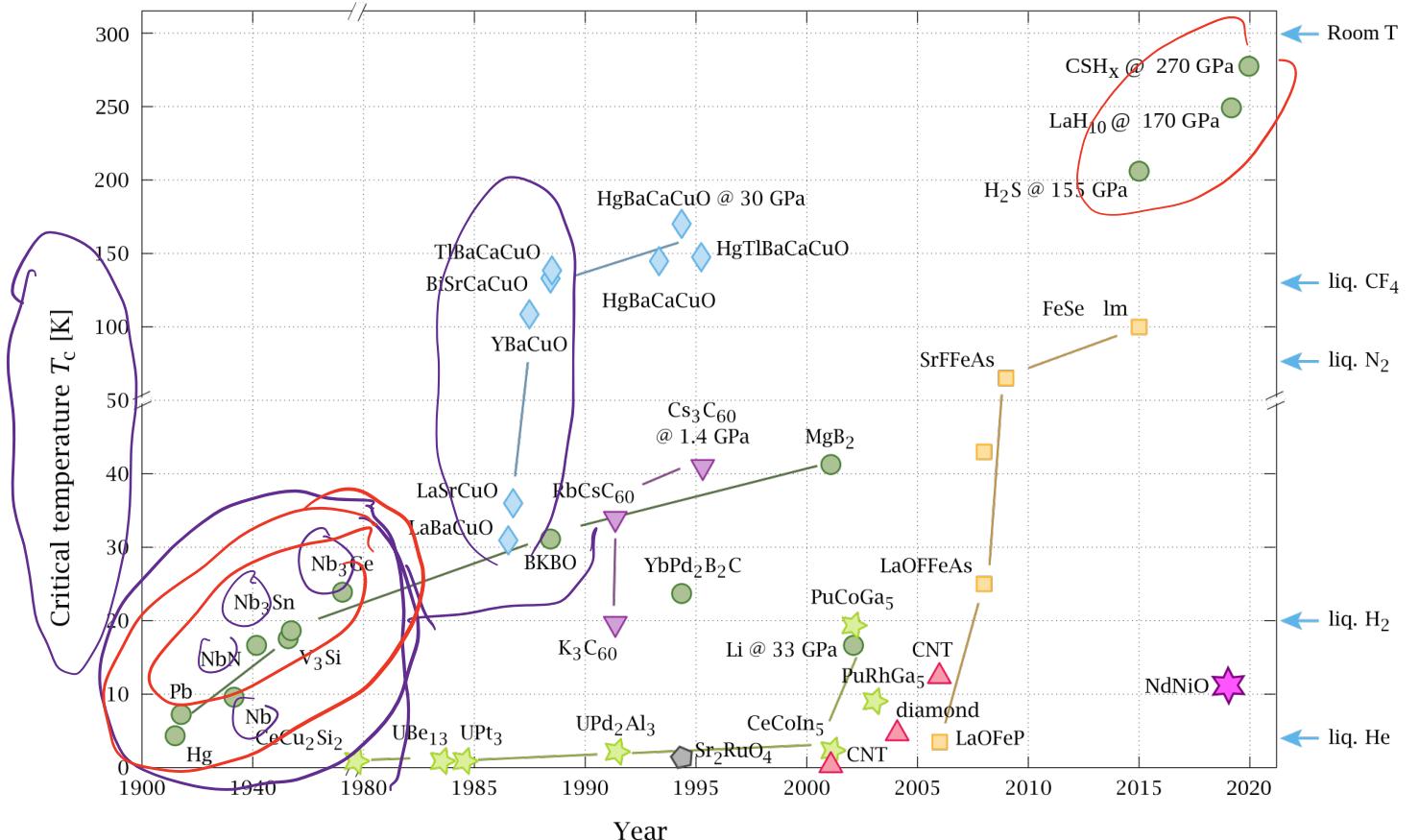
Theory and experiment have pinpointed factors that determine the magnetic attraction between electron pairs.

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?

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

- 1) Limited reliable data (hard to collect, biased, noisy)
- 2) Complex phenomena (hard to feature systems)
- 3) Extrapolation

But it can work!!

nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [articles](#) > [article](#)

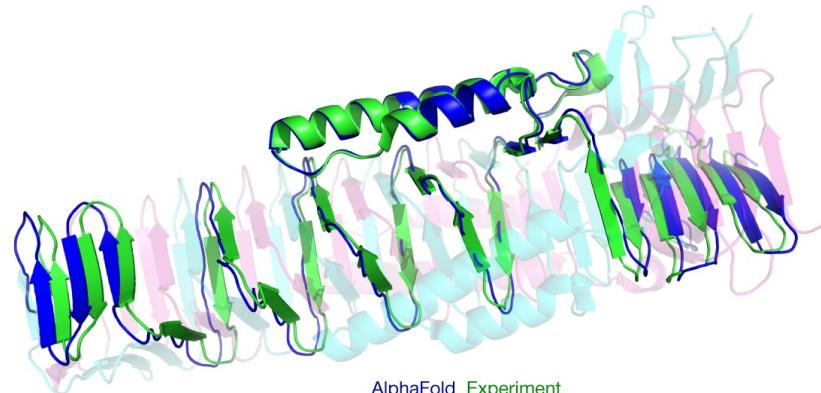
Article | [Open Access](#) | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

[John Jumper](#)✉, [Richard Evans](#), [Alexander Pritzel](#), [Tim Green](#), [Michael Figurnov](#), [Olaf Ronneberger](#),
[Kathryn Tunyasuvunakool](#), [Russ Bates](#), [Augustin Žídek](#), [Anna Potapenko](#), [Alex Bridgland](#), [Clemens Meyer](#), [Simon A. A. Kohl](#), [Andrew J. Ballard](#), [Andrew Cowie](#), [Bernardino Romera-Paredes](#), [Stanislav Nikolov](#), [Rishabh Jain](#), [Jonas Adler](#), [Trevor Back](#), [Stig Petersen](#), [David Reiman](#), [Ellen Clancy](#), [Michał Zieliński](#), ... [Demis Hassabis](#)✉ [+ Show authors](#)

[Nature](#) 596, 583–589 (2021) | [Cite this article](#)

847k Accesses | 3981 Citations | 3271 Altmetric | [Metrics](#)



But it can work!!

nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [articles](#) > [article](#)

Article | Published: 03 February 2021

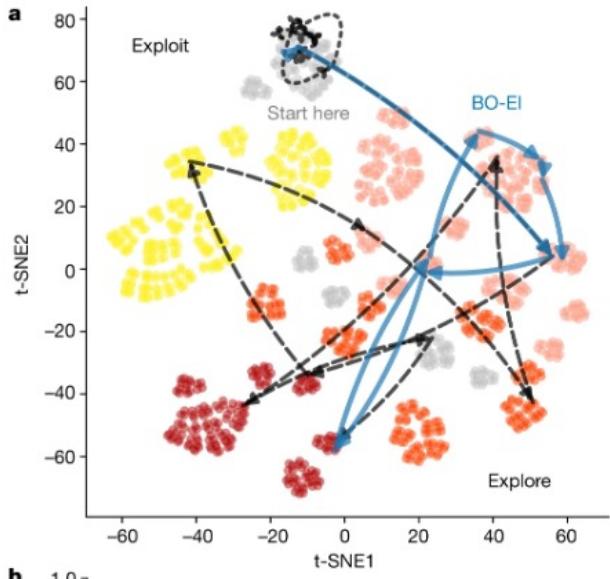
Bayesian reaction optimization as a tool for chemical synthesis

Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Martinez

Alvarado, Jacob M. Janey, Ryan P. Adams✉ & Abigail G. Doyle✉

Nature 590, 89–96 (2021) | [Cite this article](#)

44k Accesses | 130 Citations | 143 Altmetric | [Metrics](#)



But it can work!!

nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [letters](#) > [article](#)

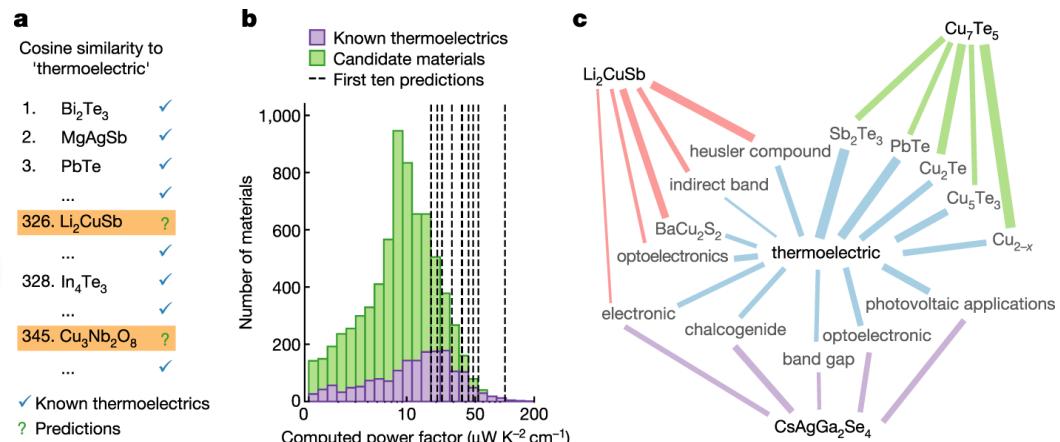
Letter | Published: 03 July 2019

Unsupervised word embeddings capture latent knowledge from materials science literature

Vahé Tshitoyan , John Daggelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder & Anubhav Jain

Nature 571, 95–98 (2019) | Cite this article

69k Accesses | 388 Citations | 1890 Altmetric | Metrics



But it can work!!

nature energy

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature energy](#) > [articles](#) > [article](#)

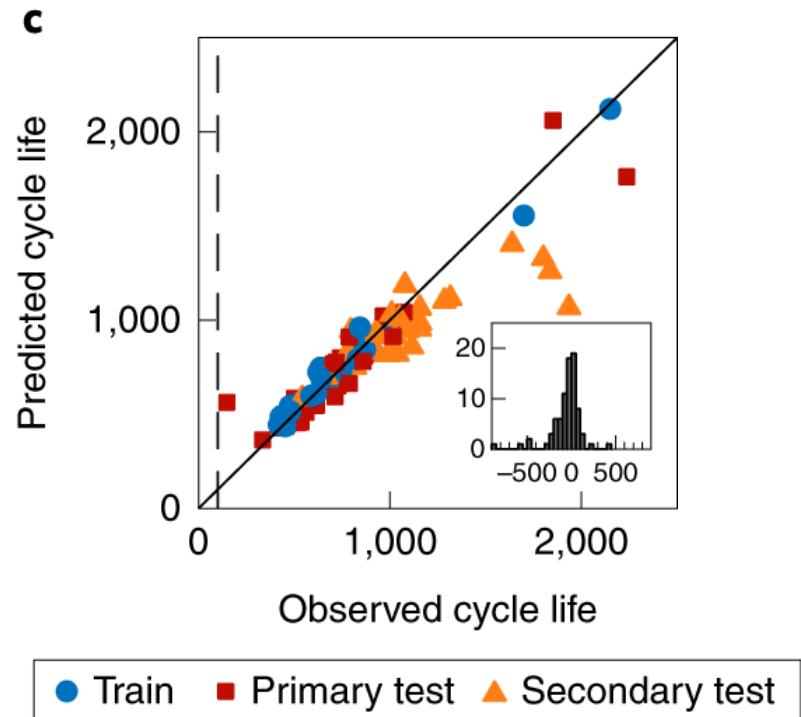
Article | Published: 25 March 2019

Data-driven prediction of battery cycle life before capacity degradation

Kristen A. Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen, Muratahan Aykol, Patrick K. Herring, Dimitrios Fragedakis, Martin Z. Bazant, Stephen J. Harris, William C. Chueh & Richard D. Braatz

Nature Energy 4, 383–391 (2019) | [Cite this article](#)

46k Accesses | 625 Citations | 184 Altmetric | [Metrics](#)



But it can work!!

Science

Current Issue First release papers Archive About Submit manu

HOME > SCIENCE > VOL. 374, NO. 6573 > PUSHING THE FRONTIERS OF DENSITY FUNCTIONALS BY SOLVING THE FRACTIONAL ELECTRON PROBLEM

REPORT QUANTUM CHEMISTRY



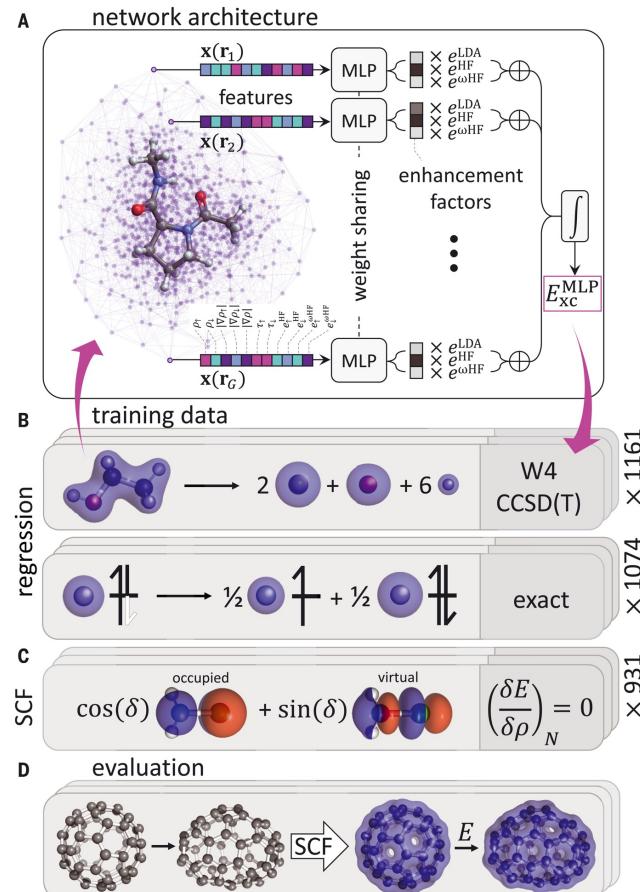
Pushing the frontiers of density functionals by solving the fractional electron problem

JAMES KIRKPATRICK , BRENDAN MCMORROW , DAVID H. P. TURBAN , ALEXANDER L. GAUNT , JAMES S. SPENCER, ALEXANDER G. D. G. MATTHEWS,

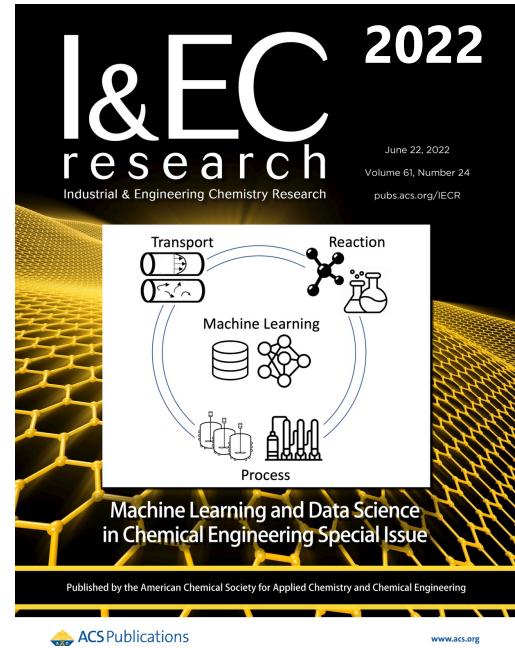
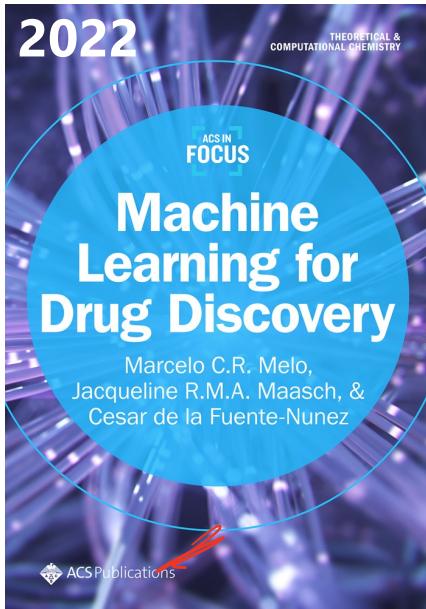
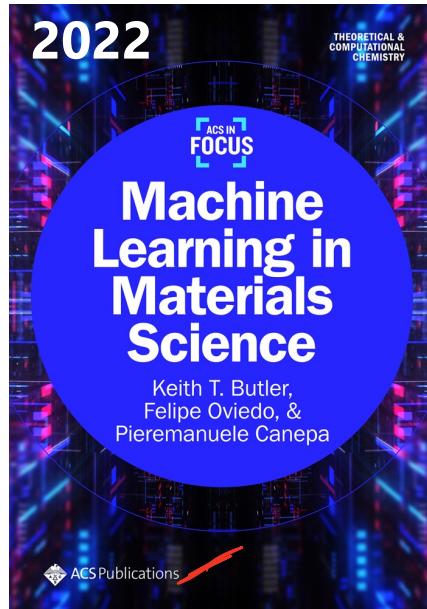
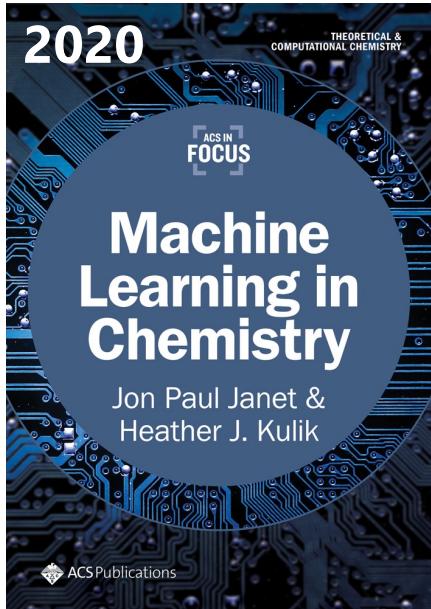
ANNETTE OBIKA, LOUIS THIRY, MEIRE FORTUNATO, [...] ARON J. COHEN +8 authors [Authors Info & Affiliations](#)

SCIENCE • 9 Dec 2021 • Vol 374, Issue 6573 • pp. 1385-1389 • DOI: 10.1126/science.abj6511

CITE



Machine learning interest on the rise!



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