

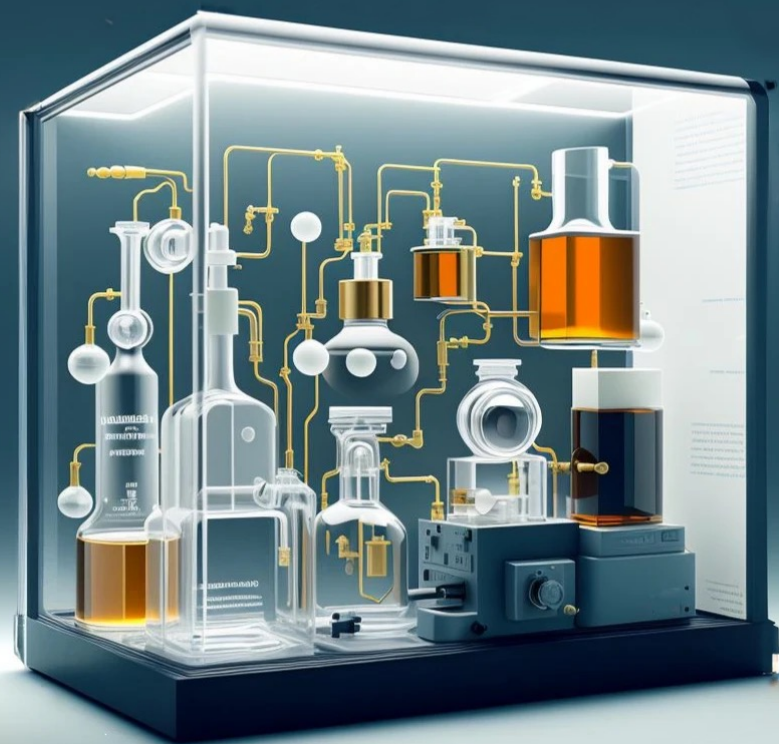
Assessing the Use of Large Language Models to Enable Programming Assignments in Physical Chemistry

Glen M Hocky

Department of Chemistry
New York University

<https://hockygroup.hosting.nyu.edu/>

do large language models know chemistry?



w/ Andrew White, University of Rochester

INITIAL IMPRESSIONS

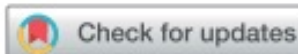
Digital
Discovery



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

Natural language processing models have emerged that can generate useable software and automate a number of programming tasks with high fidelity. These tools have yet to have an impact on the chemistry community. Yet, our initial testing demonstrates that this form of artificial intelligence is poised to transform chemistry and chemical engineering research. Here, we review developments that brought us to this point, examine applications in chemistry, and give our perspective on how this may fundamentally alter research and teaching.

Received 2nd September 2021

Accepted 25th January 2022

DOI: 10.1039/d1dd00009h

rsc.li/digitaldiscovery

INITIAL IMPRESSIONS

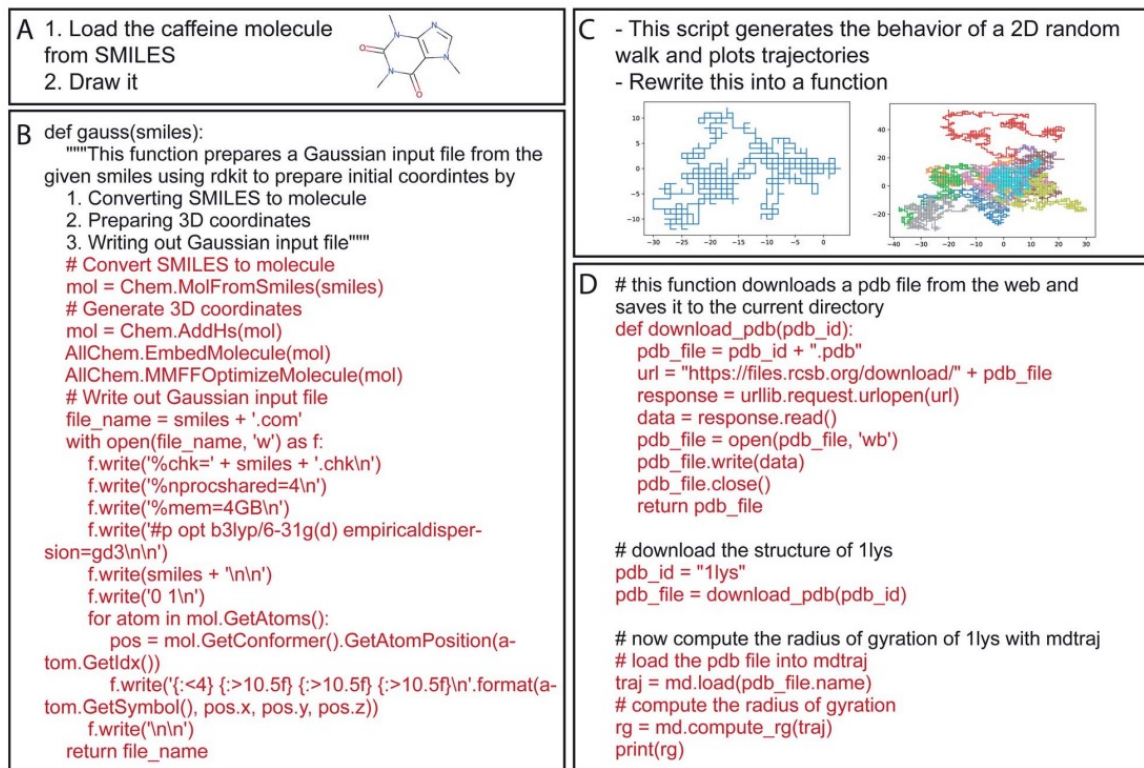


Fig. 3 Example prompts and either resulting code (B and D), or final figures that emerged from running the resulting code (A and C) (full details in the ESI†). Examples are in Python because our prompts include characteristics of Python code comments, but Codex can work in nearly any programming language included in its corpus.

RIGOROUS ASSESSMENT

Digital
Discovery



PAPER

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



Cite this: *Digital Discovery*, 2023, 2,
368

Assessment of chemistry knowledge in large language models that generate code†

Andrew D. White, *^{ab} Glen M. Hocky, *^{cd} Heta A. Gandhi, ^a Mehrad Ansari, ^a
Sam Cox, ^a Geemi P. Wellawatte, ^e Subarna Sasmal, ^c Ziyue Yang, ^a
Kangxin Liu, ^c Yuvraj Singh ^c and Willmor J. Peña Ccoa ^c

In this work, we investigate the question: do code-generating large language models know chemistry? Our results indicate, mostly yes. To evaluate this, we introduce an expandable framework for evaluating

Prompt

```
import math
import sys

def claussius(HVap, T1, P1, T2):
    """
    This function returns the phase
    transition pressure at temperature T2
    given a heat of vaporization HVap,
    and and reference temperature and
    pressure T1 and P1
    """
    [insert]
    return P2
```

Inserted code - output (4)

```
P2 = P1*math.exp((HVap/8.314)*
                ((1/T1)-(1/T2)))
```

RIGOROUS ASSESSMENT

Table 1 The number of prompts by topic and best accuracy achievable in this work. "Expert" is the number within a topic that must be evaluated by an expert. We used the "copyright" context for incoder-6B, "authority" for codegen-16B, and "insert" for davinci and $T = 0.2$ (best for all models). Accuracies are averaged (macro-averaging) across top- k sampling (we consider correct if valid prompt appeared in top- k results). Expert accuracies are macro-averaged across topics/prompts

Topic	N	Expert	Incoder	Codegen	Davinci		Davinci3
Bio	13	2	0%	29%	43%	(0%) ^a	86%
Cheminf	10	0	20%	20%	50%		50%
Genchem	11	0	29%	86%	86%		86%
md	11	3	0%	13%	63%	(81%)	88%
Plot	10	10	—	—	—	(57%)	—
qm	8	3	20%	60%	100%	(59%)	100%
sim	8	5	0%	0%	100%	(64%)	100%
spect	11	1	30%	20%	50%	(12%)	40%
stats	11	1	40%	80%	70%	(88%)	60%
Thermo	10	0	10%	10%	80%		70%
Total	84 ^b	23	17%	35%	72%	(57%)	75%

^a Expert evaluator scores are in parentheses. ^b Some prompts appear in multiple topics. The abbreviations of topics are biochemistry (bio), cheminformatics (cheminf), general chemistry (genchem), molecular dynamics & simulation (md), quantum mechanics (qm), methods of simulation (sim), spectroscopy (spect), statistics (stats), and thermodynamics (thermo)

NEW IN 2023

ChatGPT multimodal support

Final Exam
CHEM UA-652

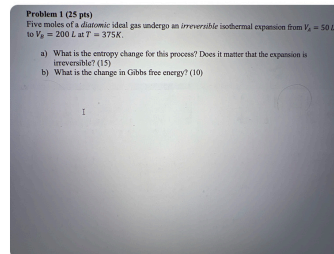
Glen M. Hocky
May 12, 2022

Problem 1 (25 pts)

Five moles of a *diatomic* ideal gas undergo an *irreversible* isothermal expansion from $V_A = 50\text{ L}$ to $V_B = 200\text{ L}$ at $T = 375\text{ K}$.

- What is the entropy change for this process? Does it matter that the expansion is irreversible? (15)
- What is the change in Gibbs free energy? (10)

HO HOCKYG@NYU.EDU



Please solve this exam problem I wrote

CHATGPT

Certainly, let's tackle this problem step by step:

Problem 1:

Five moles of a diatomic ideal gas undergo an irreversible isothermal expansion from ($V_A = 50$) L to ($V_B = 200$) L at ($T = 375$) K.

a) What is the entropy change for this process? Does it matter that the expansion is irreversible?

For an ideal gas undergoing an isothermal expansion, the change in entropy (ΔS) is given by:

$$\Delta S = n \times R \times \ln\left(\frac{V_B}{V_A}\right)$$

Where:

n = number of moles ↓ moles

INTRODUCING PYTHON PROGRAMMING IN THE CLASSROOM

Goal: introduce programming and data science into physical chemistry courses



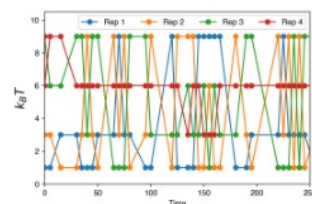
NYU/ESCIP/MOLSSI workshop @ NYU, April 2023

- <https://escip.io/intro.html>
- <https://education.molssi.org/>

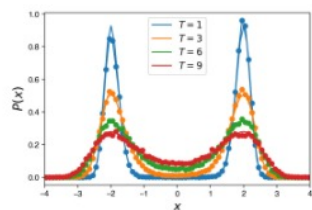
Algorithm

- Make N copies of your system.
- For system i , run Langevin dynamics at temperature T_i for n steps
- Pick a random integer j from 0 to $N - 2$, and decide whether to swap replicas j and $j + 1$ using the criterion $P_{accept} = \min(1, \exp(\Delta U \Delta(1/T)))$ where $\Delta U = U_{j+1} - U_j$ and $\Delta(1/T) = \frac{1}{T_{j+1}} - \frac{1}{T_j}$. (See notes from class)
- If accepted, swap the two new configurations (positions and velocities) for that time, otherwise do nothing

Exchange example



Distribution example



In []:

See e.g. <https://github.com/hockyg/chem-ga-2600>

INTRODUCING PYTHON+AI INTO CHEMISTRY COURSES

Goal: introduce programming and data science into physical chemistry courses

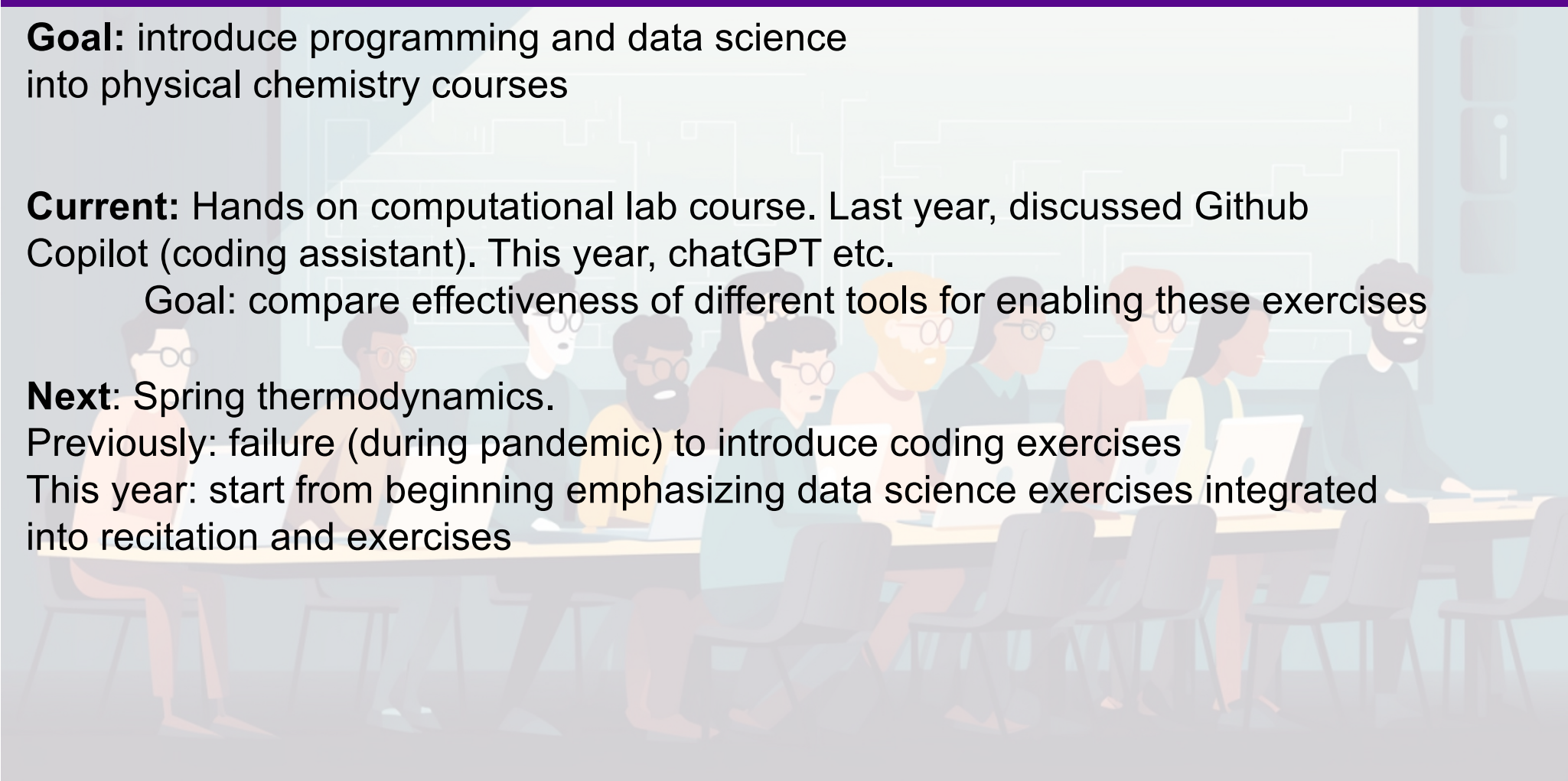
Current: Hands on computational lab course. Last year, discussed Github Copilot (coding assistant). This year, chatGPT etc.

Goal: compare effectiveness of different tools for enabling these exercises

Next: Spring thermodynamics.

Previously: failure (during pandemic) to introduce coding exercises

This year: start from beginning emphasizing data science exercises integrated into recitation and exercises



ACKNOWLEDGEMENTS



FUNDING AND SUPPORT



New York University
High Performance Computing



R35-GM138312
R35-GM138312-S1 (Peña)



Andrew White, and group