

# MNT-CURN Summer Research 2024

## Exploration of Large Language Models for Media Analysis

Kat Nykiel, Dr. Alejandro Strachan

Purdue University

December 15, 2023

# The Strachan Lab at Purdue

Our group uses **simulations** and **machine learning** to study materials at the atomic scale

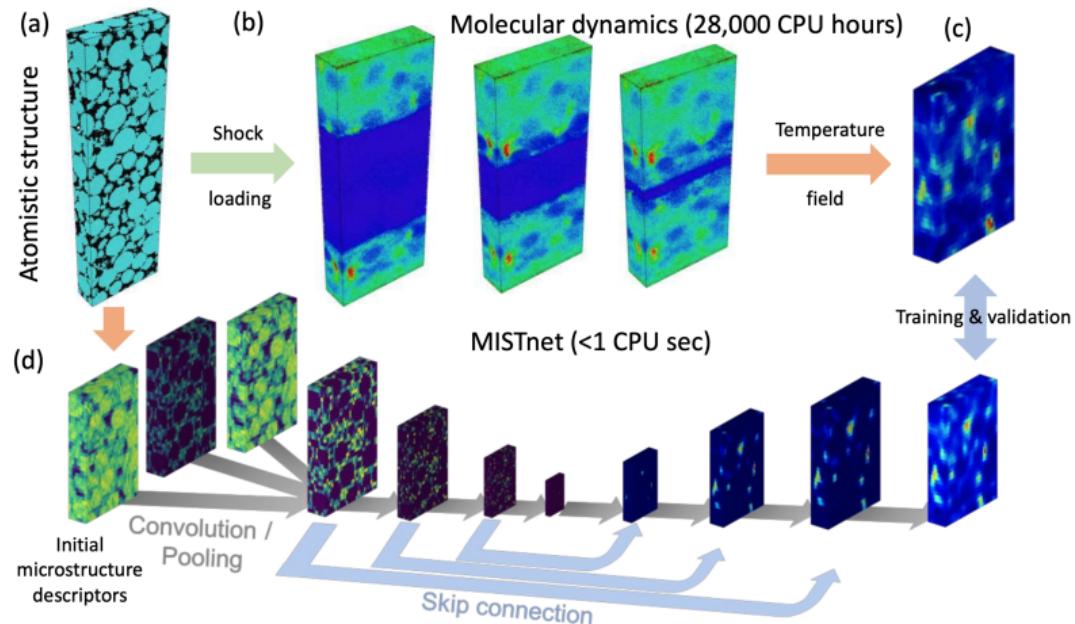


Figure 1: Machine learning can significantly lower computational cost [1]

# Large Language Models (LLMs) as Research Tools

How can we use LLMs to accelerate research in materials science?

```
units metal
dimension 3
boundary p p p

atom_style atomic
lattice fcc 4.08

region box 0 5 0 5 0 5 units lattice
create_box 1 box
create_atoms 1 box

replicate 5 5 5

pair_style <eam/fs> or <eam/alloy>
pair_coeff * * <potential file name> Al
mass i 26.9815

neighbor 2.0 bin
neigh_modify delay 10

velocity all create 300 <random seed> mom yes rot yes dist <uniform> or <gaussian>

fix 1 all npt temp 300 300 0.1 iso 1 1 1
thermo <thermo steps>
thermo_style <custom step temp pe etotal press vol>

timestep 0.001
run 500000
unfix 1

dump, print, write.data, ...
```

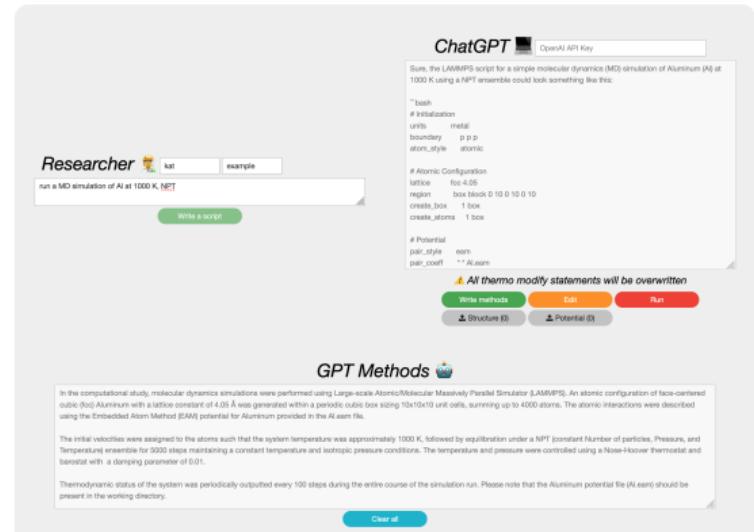


Figure 3: GPT to write methodology [3]

Figure 2: GPT for simulation design [2]

# Summer 2024 MNT-CURN Project: Content Analysis with LLMs

Research Question:

How do LLMs compare against statistical natural language processing methods for extracting information from articles?

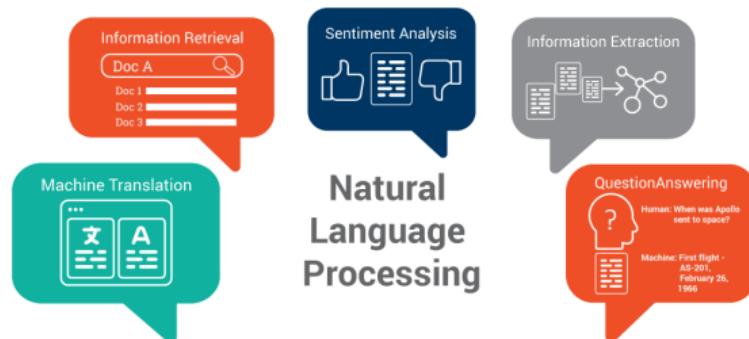


Figure 4: Common NLP tasks [4]



Figure 5: Prominent LLM architectures [5]

Possible topics include comparative media analysis, bias detection, and topic modeling

# Past MNT-CURN Project: News Analysis

Research Question:

What can we learn from the affiliations of experts quoted in news articles?

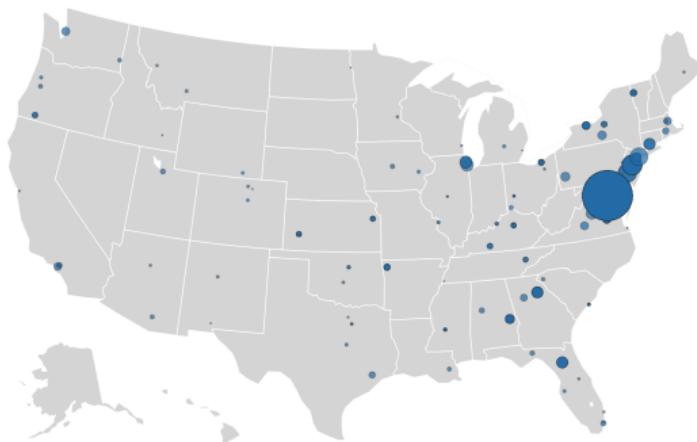


Figure 6: Regional affiliations in news articles

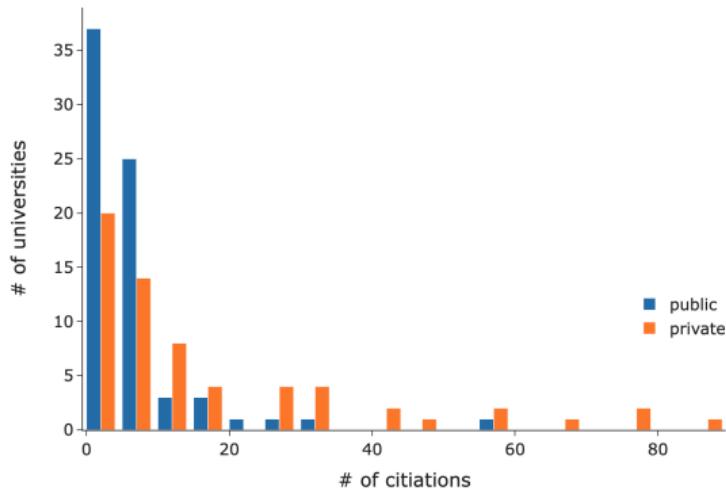


Figure 7: Institutional affiliations in news articles

# Project Details

## Prerequisites

This project requires introductory python coding skills. Familiarity with APIs, pandas, and scikit-learn is helpful, but not a requirement.

## Location

This program will be entirely online, with regular virtual meetings.

## Program Window

The start and end date of this program are flexible, and will be worked out between the mentor, faculty, and student. The tentative schedule is May - July 2024.

## How to Apply

Apply by reaching out to the graduate student mentor, Kat Nykiel ([knykiel@purdue.edu](mailto:knykiel@purdue.edu)), with your resume and a short summary of your interest in this project by February 1, 2024.

## References I

- [1] Chunyu Li et al. *Mapping Microstructure to Shock-Induced Temperature Fields Using Deep Learning*. Mar. 30, 2023. DOI: 10.48550/arXiv.2303.17345. arXiv: 2303.17345 [cond-mat]. URL: <http://arxiv.org/abs/2303.17345> (visited on 12/15/2023). preprint.
- [2] Juan C. Verduzco, Ethan Holbrook, and Alejandro Strachan. *GPT-4 as an Interface between Researchers and Computational Software: Improving Usability and Reproducibility*. Comment: 22 pages, 7 figures. Oct. 4, 2023. DOI: 10.48550/arXiv.2310.11458. arXiv: 2310.11458 [cond-mat]. URL: <http://arxiv.org/abs/2310.11458> (visited on 12/15/2023). preprint.
- [3] Juan Carlos Verduzco Gastelum et al. “Large Language Model Demonstration for LAMMPS”. In: (Nov. 5, 2023). URL: <https://nanohub.org/resources/llm4lammps> (visited on 12/15/2023).

## References II

- [4] Milena Yankova. *Top 5 Semantic Technology Trends to Look for in 2017*. Ontotext. Jan. 17, 2017. URL:  
<https://www.ontotext.com/blog/top-5-semantic-technology-trends-2017/> (visited on 12/15/2023).
- [5] *Top Five AI Resources for Large Language Models*. URL:  
<https://www.triggermesh.com/blog/top-5-ai-resources-for-l1lms> (visited on 12/15/2023).