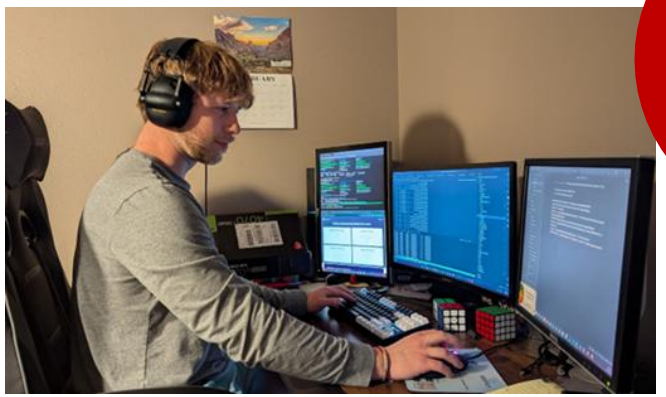
Natural Language to LAMMPS: LLMs as interfaces between researchers and scientific software

Ethan Holbrook, Juan C. Verduzco, and Alejandro Strachan

Day in the life of a computational materials scientist

Computational Materials Scientist



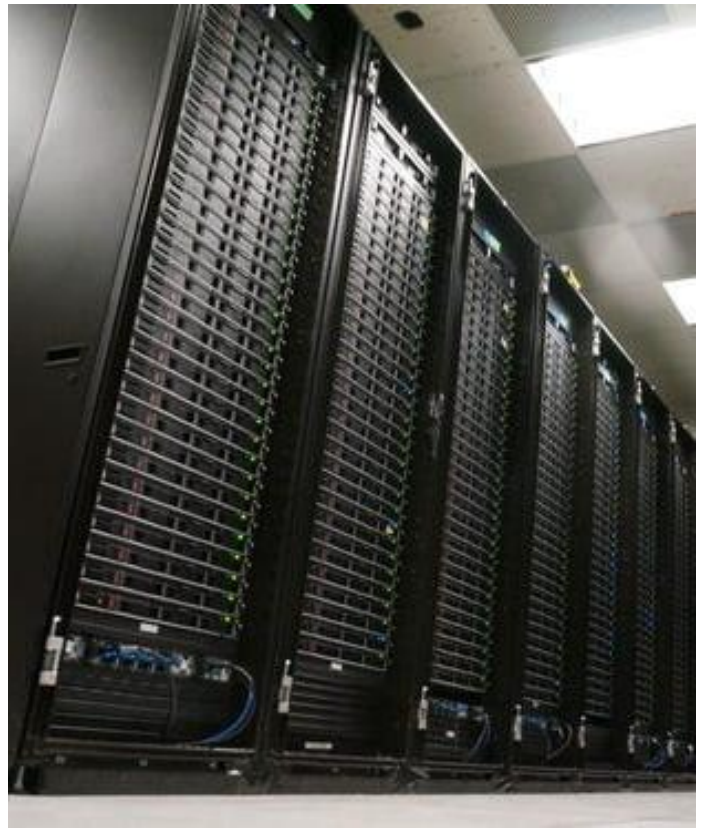
Computational Software

VASP

QUANTUMESPRESSO

LAMMPS

Supercomputer (At Purdue!)



Data

Domain-specific languages

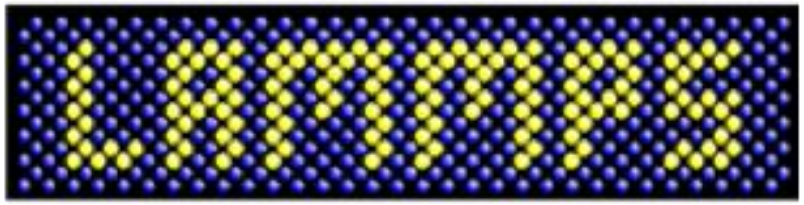
```
&CONTROL
    calculation = 'scf',
    prefix = 'silicon',
    outdir = './'
    pseudo_dir = '../pseudos/'

&SYSTEM
    ibrav =  2,
    celldm(1) = 10.26,
    nat =  2,
    ntyp = 1,
    ecutwfc = 30
    nbnd = 8

&ELECTRONS
    mixing_beta = 0.6

ATOMIC_SPECIES
  Si 28.086 Si.pz-vbc.UPF

ATOMIC_POSITIONS
  Si 0.0 0.0 0.0
  Si 0.25 0.25 0.25

K_POINTS
  1 1 1 0 0 0
```

```
SYSTEM = Rhodium surface calculation
ISTART =      0
ICHARG =      2

ENCUT  = 300.00 eV
ALGO   =  Normal
NELM   = 60
EDIFF  = 1E-06
SIGMA  =    0.05
ISMEAR =  0

EDIFFG =  -1E-02
NSW    =  20
IBRION =  2
POTIM  =  .5
KPAR   =    4
NCORE  =    4
LREAL  =    A
```

```
units                lj
atom_style           atomic

lattice                fcc 0.8442
region                 box block 0 10 0 10 0 10
create_box           1 box
create_atoms           1 box
mass                   1 1.0

velocity           all create 3.0 87287 loop geom

pair_style           lj/cut 2.5
pair_coeff           1 1 1.0 1.0 2.5

neighbor           0.3 bin
neigh_modify           every 20 delay 0 check no

fix                    1 all nve
```

QUANTUMESPRESSO

VASP

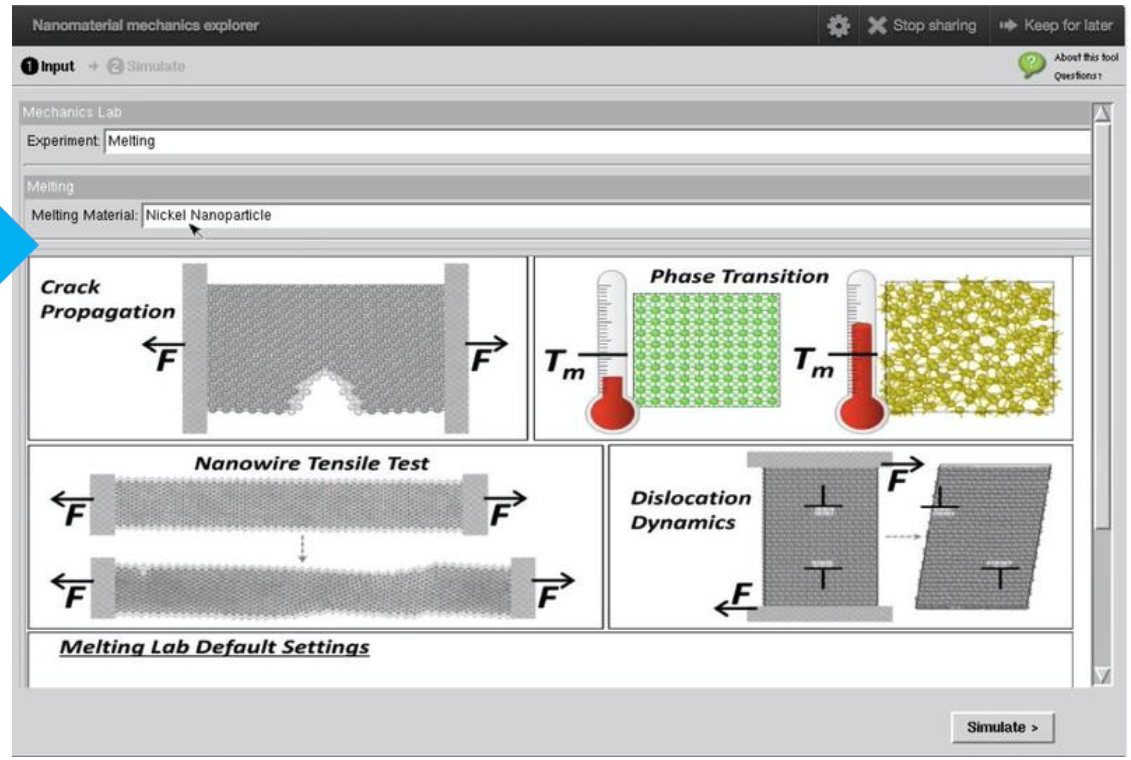


Current approaches

Projects — vi .//Semicond/MoTe2Crack/Crack/monolayer-MoTe2-hex.in — 104×45

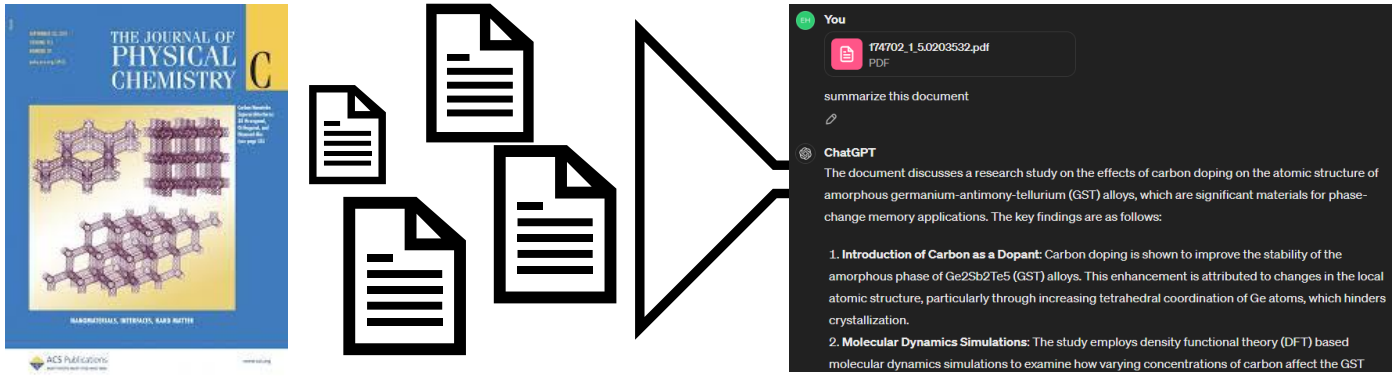
```
units           real
atom_style      full
boundary        p p s

box             tilt large
region          box prism 0 3.5562 0 3.0797 0 9.0000 -1.7781 0.0 0.0
create_box      2 box
lattice         custom 1.0 a1 3.5562 0.0 0.0 a2 1.7781 3.0797 0.0 a3 0.0 0.0 9.0 &
                basis 0.0 0.0 0.0
create_atoms    1 single 0.0 0.0 0.5 units lattice
create_atoms    2 single 0.3333 0.3333 0.7003 units lattice
create_atoms    2 single 0.3333 0.3333 0.2997 units lattice

mass            1 97.9055000
mass            2 129.9067000

replicate       30 30 1

pair_style      reax/c lmp_control checkqeq yes
pair_coeff      * * ffield-MoTeCu Mo Te

timestep        0.5
thermo_style    custom etotal pe ke temp evdwl press pxx pyy pzz pxy pxz pyz lx ly lz
thermo_modify   line multi flush yes
thermo          10

variable        ti equal 300.0

dump            1 all custom 100 traj.dump id type x y z q

fix             0 all qeq/reax 1 0.0 10.0 1.0e-6 reax/c
minimize        1.0e-8 1.0e-8 100 1000

velocity        all create ${ti} 4928459 mom yes rot yes dist gaussian
fix             1 all nvt/sllod temp ${ti} ${ti} 100.0
fix             2 all deform 1 x erate 0.000002 remap v units box
run             100000
~
~
```

LAMMPS-GUI



LLMs have shown a lot of potential

Summarize information from publications



Code assistant

```python
def count_character_frequency(input_string):
    frequency = {}
    for char in input_string:
        if char in frequency:
            frequency[char] += 1
        else:
            frequency[char] = 1
count_character_frequency("Hello, World!")
```

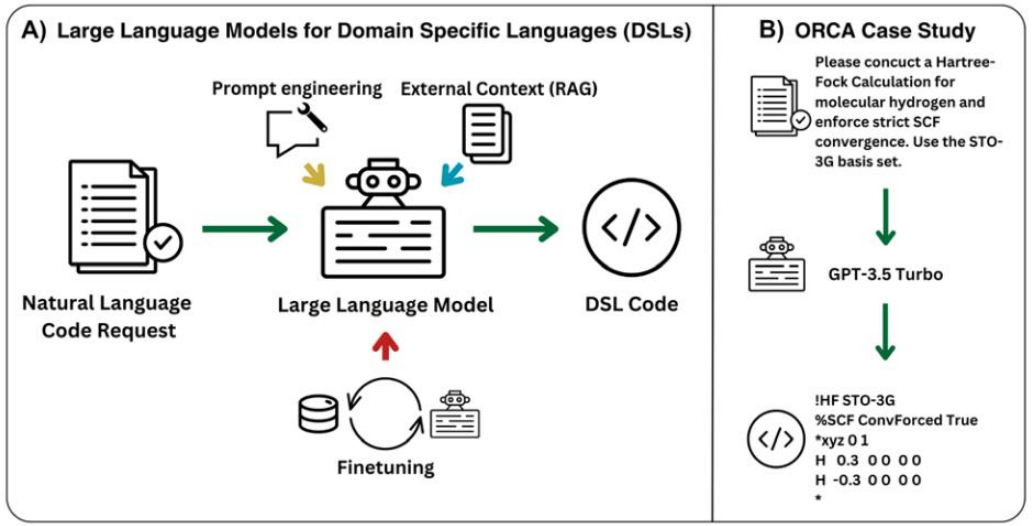
Extract AI-ready data from publications

https://www.who.int/southeastasia/publications

A new possibility… software interfaces

Current efforts have seen success in other SW, such as:

ORCA: Jacobs, P. F. & Pollice, R. Developing large language models for quantum chemistry simulation input generation. *Digit. Discov.* (2025)

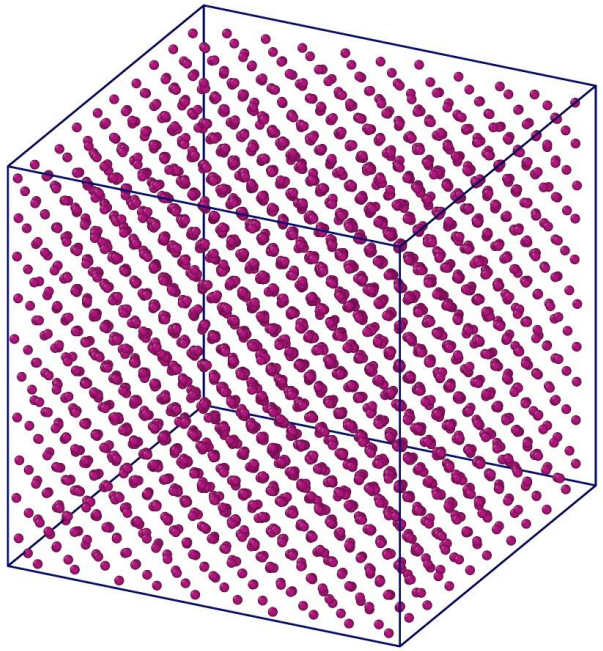


Interface between the researcher & tools

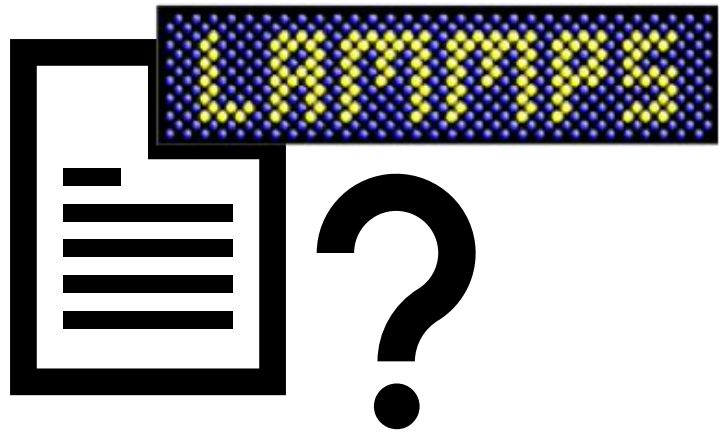
Example: LLMs as an AI research assistant

Prompt 2

We characterized the melting of a bulk Ni sample using molecular dynamics with LAMMPS. The initial condition was obtained by replicating the Ni unit cell 10 times in each direction. Initial velocities were drawn from the Maxwell-Boltzmann distribution at 600 K. The system was heated from 300 K to 2500 K continuously, at a rate of 10 K per ps under isothermal and isobaric conditions at 1 atm. Interactions were described using an embedded atom model developed by Mishin et al. in 1999 [1] obtained from OpenKIM.org. [1] EAM potential (LAMMPS cubic hermite tabulation) for Ni developed by Mishin et al. (1999) v005. OpenKIM; 2018. doi:10.25950/a88dfc37.



GPT-4

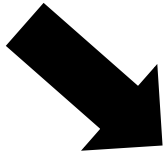


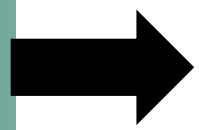
Verduzco, Holbrook, & AS. arXiv:2310.11458. 2023 Oct 4. https://doi.org/10.48550/arXiv.2310.11458

Script generation performance

Prompt 2

We characterized the melting of a bulk Ni sample using molecular dynamics with LAMMPS. The initial condition was obtained by replicating the Ni unit cell 10 times in each direction. Initial velocities were drawn from the Maxwell-Boltzmann distribution at 600 K. The system was heated from 300 K to 2500 K continuously, at a rate of 10 K per ps under isothermal and isobaric conditions at 1 atm. Interactions were described using an embedded atom model developed by Mishin et al. in 1999 [1] obtained from OpenKIM.org. [1] EAM potential (LAMMPS cubic hermite tabulation) for Ni developed by Mishin et al. (1999) v005. OpenKIM; 2018. doi:10.25950/a88dfc37.





Complexity →

Metric	Prompt 1	Prompt 2	Prompt 3
Script Validity	10/10	8/10	0/10
Simulation Accuracy	9/10	4/10	0/10
Consistency	0.87	0.85	0.81
Usefulness	0.88	0.76	0.81

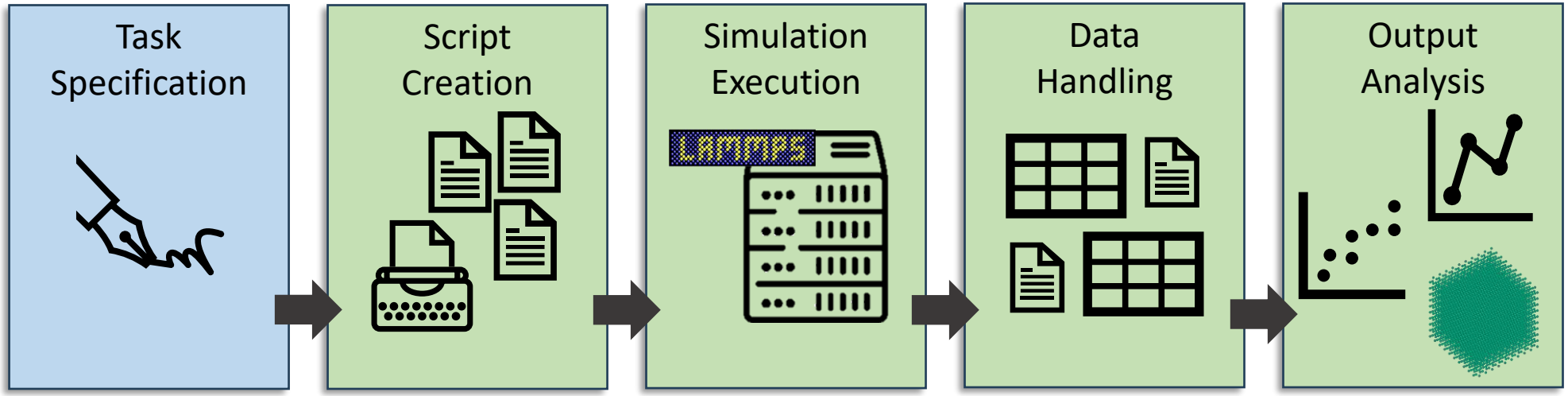
In our experiments, it worked … sometimes

```
1  units metal
2  dimension 3
3  boundary p p p
4
5  atom_style atomic
6  lattice fcc 3.52
7
8  region box block 0 10 0 10 0 10  units lattice
9  create_box 1 box
10 create_atoms 1 box
11
12 replicate 10 10 10
13
14 pair_style eam/fs
15 pair_coeff * *  <potential file name>  Ni
16 mass 1 58.69
17
18 velocity all create 600  <random seed>  mom yes  rot yes  dist gaussian
19
20 fix 1 all npt temp 300.0 2500.0 0.1 iso  1 1 1
21
22 thermo  <thermo steps>
23 thermo_style  <custom step temp pe etotal press vol>
24
25 timestep 0.001
26 run 500000
27 unfix 1
28
29 dump, print, write_data, ...
```

Annotated example of a LAMMPS input script from Prompt 2.
Always added correct commands. *Often added correct commands that match default values.*
Values that were generated with different options in some scripts. *Incorrect commands.*

Verduzco, Holbrook, & AS. arXiv:2310.11458. 2023 Oct 4. https://doi.org/10.48550/arXiv.2310.11458

Can LLMs be a general interface?



jupyter

Edit App

Submit a tick

2. The first time you enter an API key, type the API key in the box and press "Store/Fetch Key". The second time, you can simply press the button to retrieve your key.
3. Match the potential file or structure file you upload to the name in the generated script.
4. For thermodynamic simulation outputs, download the yaml file.
5. For trajectories or other simulation files, download the zip folder.
6. Found a bug? Email holbrooe@purdue.edu

Your alias

Unique run identifier

OpenAI API Key

Store/Fetch Key

[No Title]

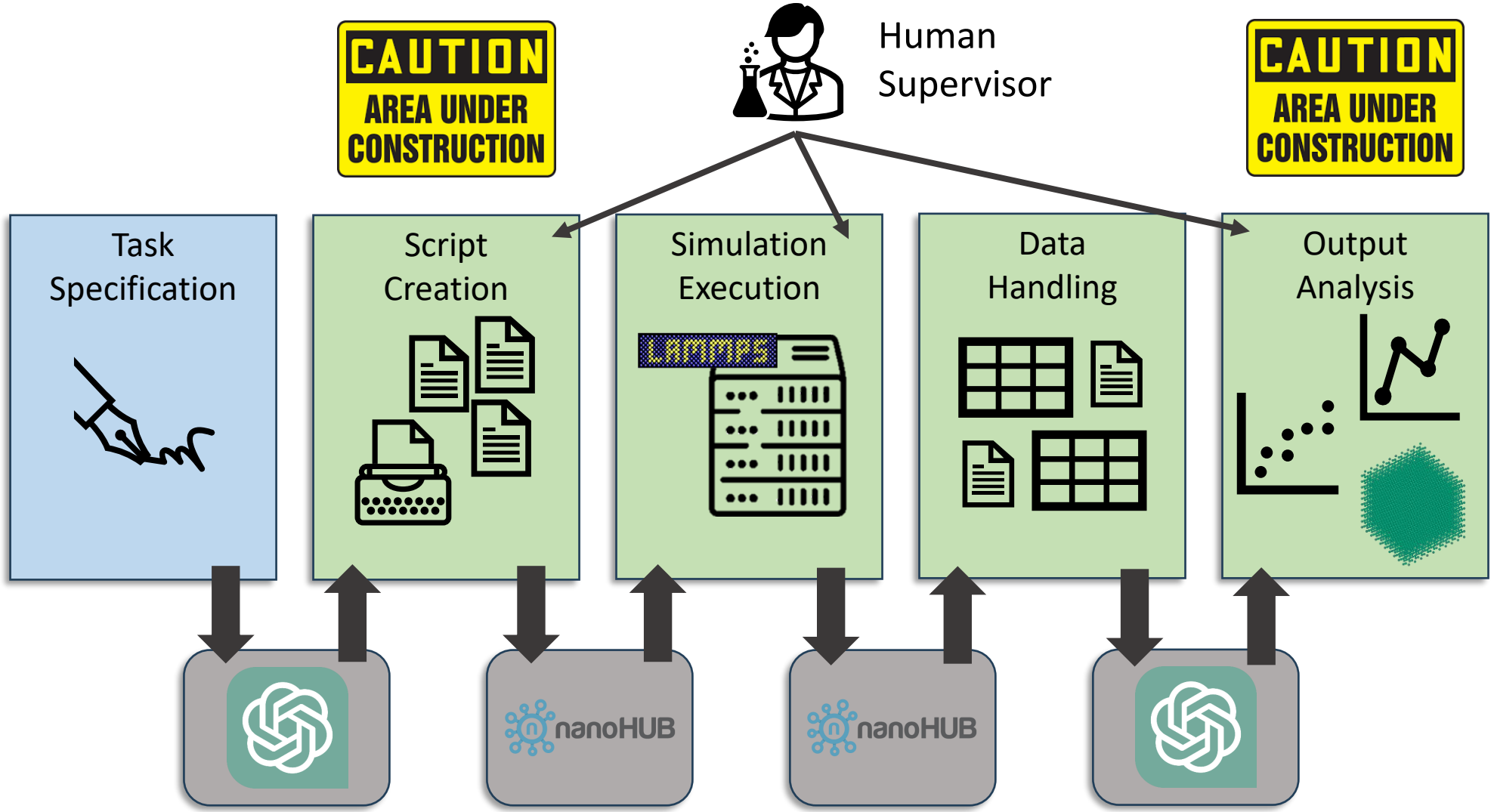
Researcher

Write the description of a molecular dynamics simulation for GPT-4 to generate

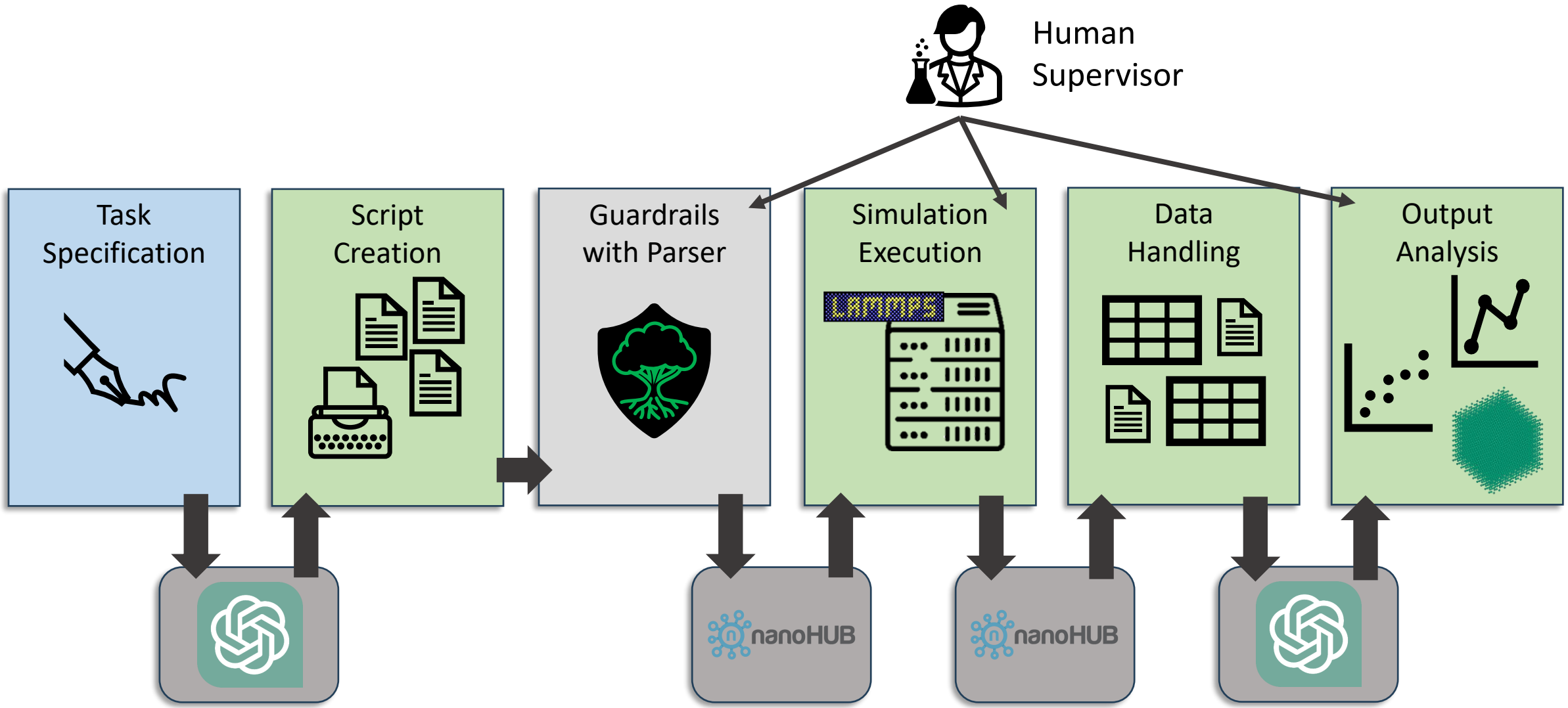
GPT-4

GPT-written LAMMPS script

Can LLMs be a general interface?

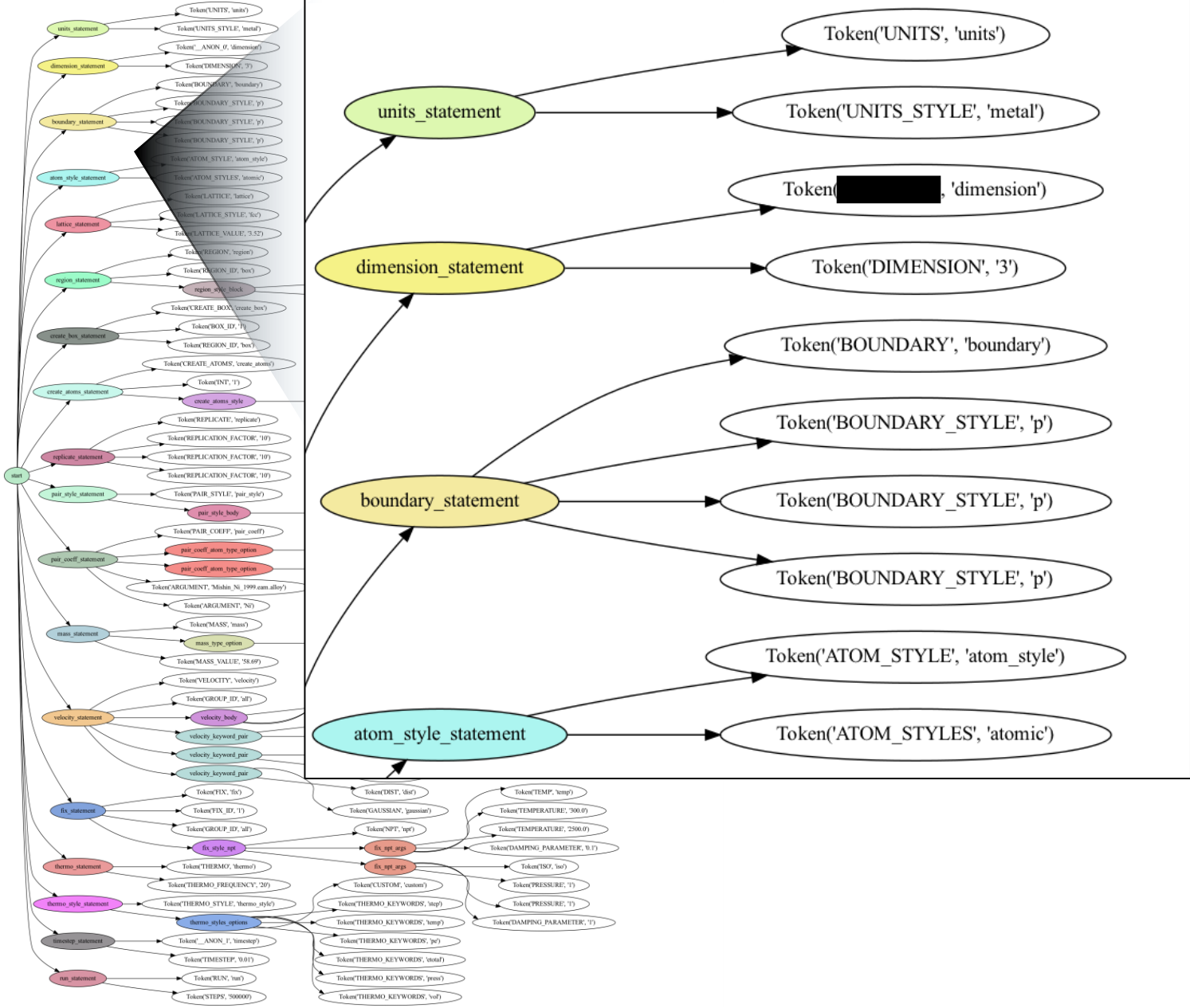


Towards LLMs as a general interface



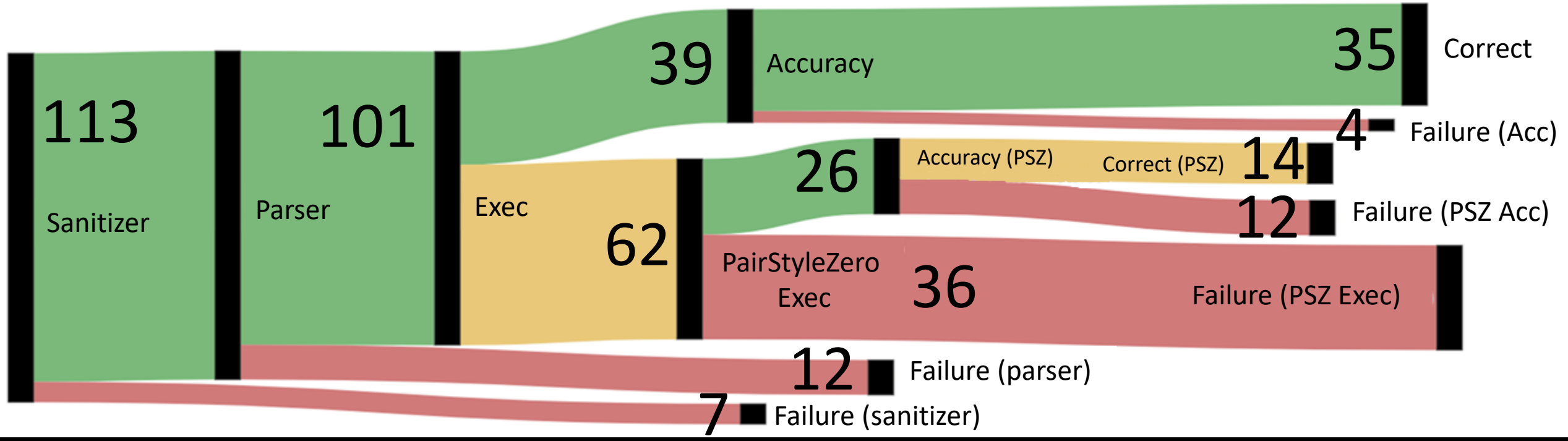
Abstract Syntax Trees

- We are developing a parser for LAMMPS input scripts
- Enables syntax error detection



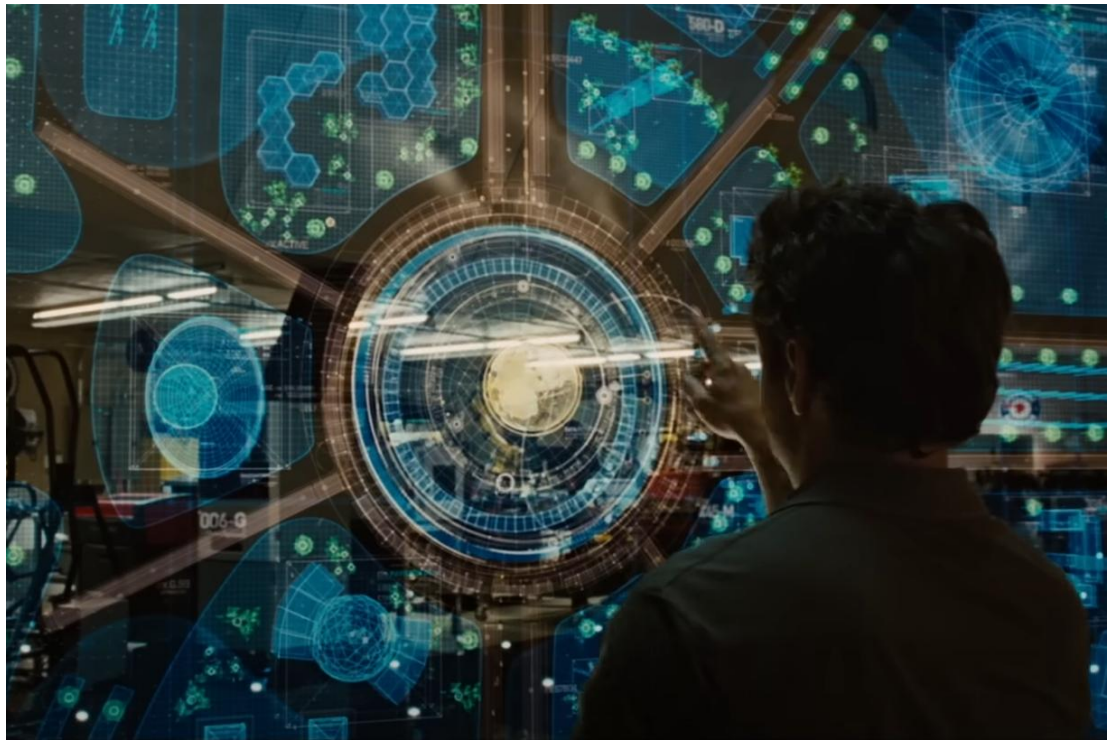
LAMMPS prompting results

- **4 models:** GPT-4o, GPT-o3, GPT-4.1, Claude Opus 4
  - **3 Prompts:** Simple, Medium, Complex
- Careful of misleading word-matching
- Region/group specifications, units, and pair style most prevalent issues



Conclusions

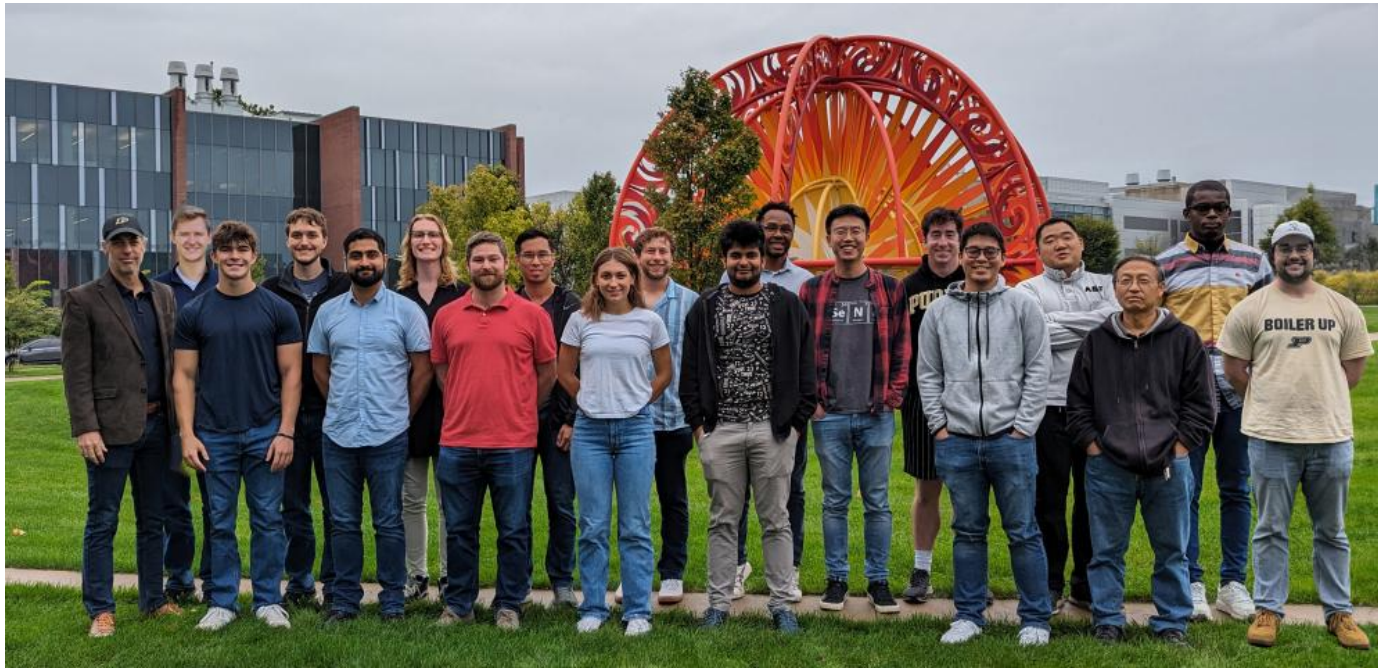
- LLMs are generally capable of generating valid LAMMPS input scripts
- LLMs at the interface of human and researcher shows promise for accelerating progress in computational materials science
- Developing tools for these domain-specific languages can help make these models better



Thanks for listening!

https://nanohub.org/tools/llm4lammps

- Alejandro Strachan's Group
- Email: holbrooe@purdue.edu
- Questions??



Simulation details

ethan_tms

run1

Example 2 (Nickel Temperature Ramp)

sk-kGym1BZJBdZRXvdforCmT3BlbkFJzcyHRCmX1D64nA

Store/Fetch Key

Researcher

We characterized the melting of a bulk Ni sample using molecular dynamics with LAMMPS. The initial condition was obtained by replicating the Ni unit cell 10 times in each direction. Initial velocities were drawn from the Maxwell-Boltzmann distribution at 600 K. The system was heated from 300 K to 2500 K continuously, at a rate of 10 K per ps under isothermal and isobaric conditions at 1 atm. Interactions were described using an embedded atom model developed by Mishin et al. in 1999 [1] obtained from OpenKIM.org. [1] EAM potential (LAMMPS cubic hermite tabulation) for Ni developed by Mishin et al. (1999) v005. OpenKIM; 2018. doi:10.25950/a88dfc37.

Generate LAMMPS script with G…

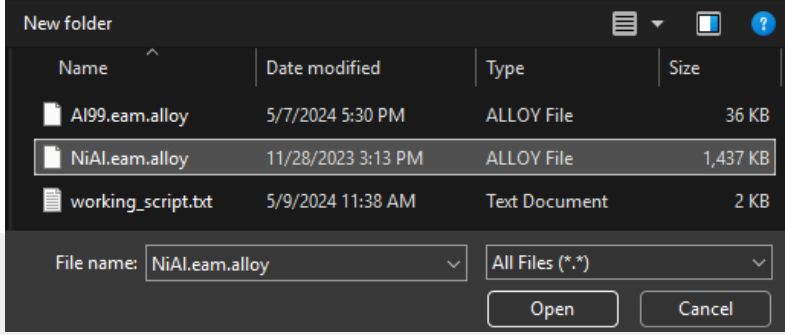
```
region mybox block 0 10 0 10 0 10
create_box 1 mybox
create_atoms 1 box

# Set potential for Ni using Mishin EAM potential from OpenKIM
pair_style eam/alloy
pair_coeff * * NiAl.eam.alloy Ni

# Define simulation parameters
mass 1 58.6934
```

⚠ All thermo modify statements will be modified to use YAML formatting

Structure (0) | Potential (1) | Supplementary (0)

Edit | Run

Add potential file



Researcher

We characterized the melting of a bulk Ni sample using molecular dynamics with LAMMPS. The initial condition was obtained by replicating the Ni unit cell 10 times in each direction. Initial velocities were drawn from the Maxwell-Boltzmann distribution at 600 K. The system was heated from 300 K to 2500 K continuously, at a rate of 10 K per ps under isothermal and isobaric conditions at 1 atm. Interactions were described using an embedded atom model developed by Mishin et al. in 1999 [1] obtained from OpenKIM.org. [1] EAM potential (LAMMPS cubic hermite tabulation) for Ni developed by Mishin et al. (1999) v005. OpenKIM; 2018. doi:10.25950/a88dfc37.

Generate LAMMPS script with G…

GPT-4

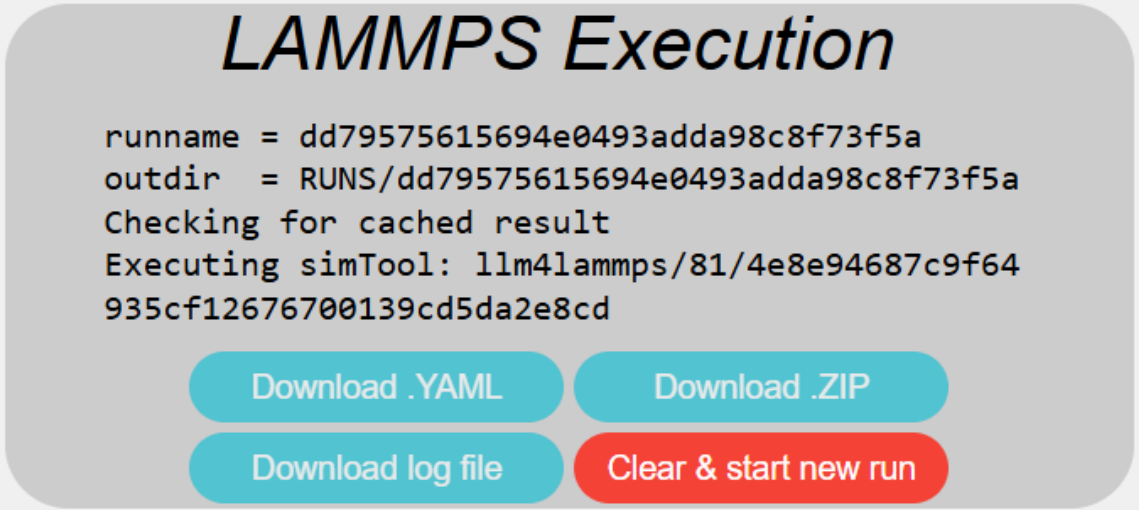
```
region mybox block 0 10 0 10 0 10
create_box 1 mybox
create_atoms 1 box

# Set potential for Ni using Mishin EAM potential from OpenKIM
pair_style eam/alloy
pair_coeff * * NiAl.eam.alloy Ni

# Define simulation parameters
mass 1 58.6934
```

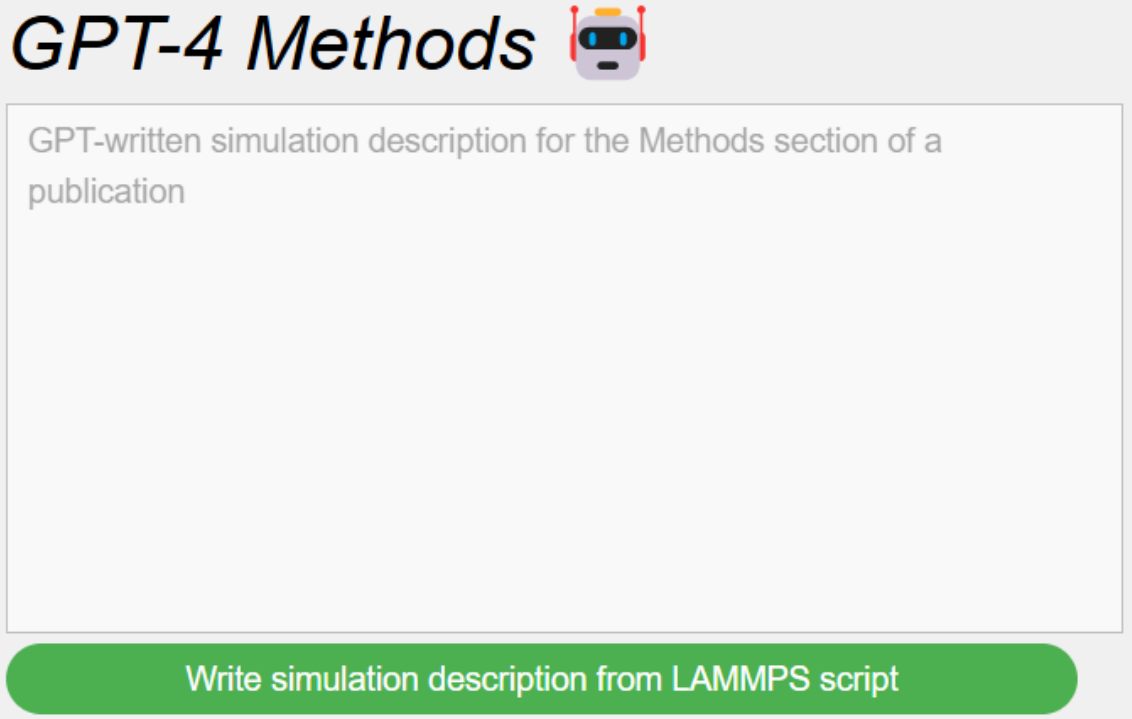
⚠ All thermo modify statements will be modified to use YAML formatting

Structure (0) | Potential (1) | Supplementary (0)

Edit | Run

Execution


LAMMPS Execution
runname = dd79575615694e0493adda98c8f73f5a
outdir = RUNS/dd79575615694e0493adda98c8f73f5a
Checking for cached result
Executing simTool: llm4lammps/81/4e8e94687c9f64
935cf12676700139cd5da2e8cd
Download .YAML
Download .ZIP
Download log file
Clear & start new run


GPT-4 Methods
GPT-written simulation description for the Methods section of a publication
Write simulation description from LAMMPS script

Methods

LAMMPS Execution

```
runname = dd79575615694e0493adda98c8f73f5a
outdir  = RUNS/dd79575615694e0493adda98c8f73f5a
Checking for cached result
Executing simTool: llm4lammps/81/4e8e94687c9f64
935cf12676700139cd5da2e8cd

Archiving SimTool Result
Starting push to cache: Thu 20 Mar 2025 04:13:0
6 AM UTC
Finished push to cache: Thu 20 Mar 2025 04:13:2
1 AM UTC
Fetching cached result
```

Download .YAML | Download .ZIP

Download log file | Clear & start new run

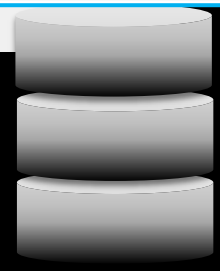
GPT-4 Methods

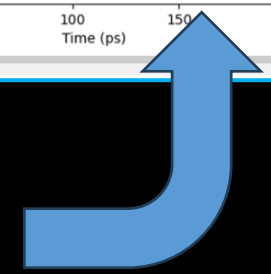
A face-centered cubic (FCC) lattice structure, representative of Ni, was initialized with a lattice constant of 3.52 Å. A simulation cell of dimensions 10x10x10 unit cells was constructed, resulting in a total of 4000 atoms. The metal unit style was used to ensure appropriate physical units for the simulation.

Initial atom velocities were assigned based on a Maxwell-Boltzmann distribution at a temperature of 600 K, with a random seed used to generate initial velocity distributions. An NPT ensemble was applied with isotropic pressure control to maintain a constant pressure of 1 atm

Write simulation description from LAMMPS script

LLMs as universal interfaces





Verduzco, Holbrook, & AS. arXiv:2310.11458. 2023 Oct 4. https://doi.org/10.48550/arXiv.23