



Northwestern  
University

# ET-AL: Entropy Targeted Active Learning for Materials Data Bias Mitigation

Henry Zhang<sup>1</sup>, Wayne Chen<sup>1</sup>, James Rondinelli<sup>2</sup>, Wei Chen<sup>1</sup>

<sup>1</sup>Mechanical Engineering, Northwestern University

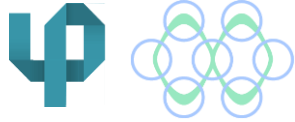
<sup>2</sup>Materials Science & Engineering, Northwestern University

# Materials Informatics, Data, and Bias

## Data sources for materials informatics



(1) Literature



(2) Databases

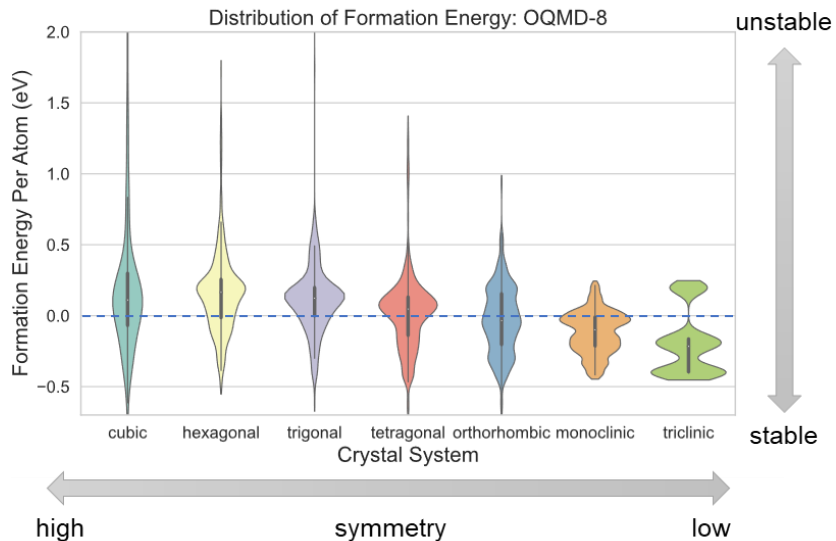


(3) Experiments/Computation

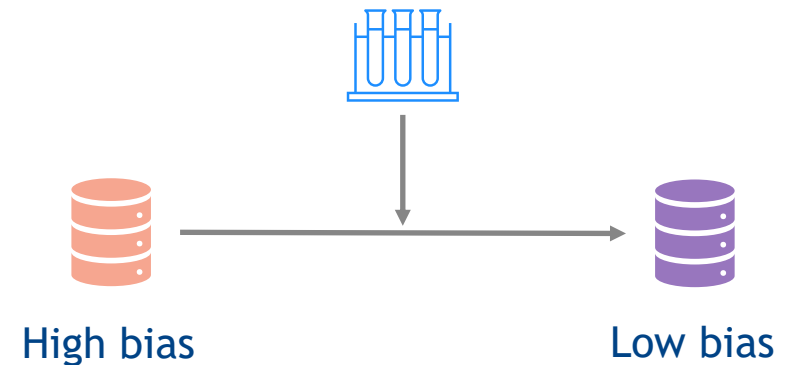
Biased data → problematic machine learning (ML) models

A framework for bias mitigation

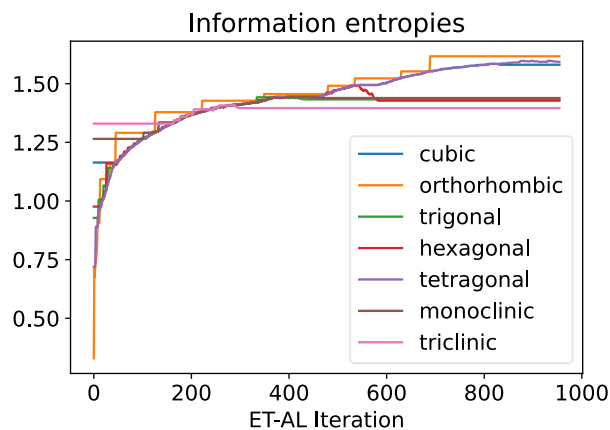
- Select underrepresented materials
- Add to database



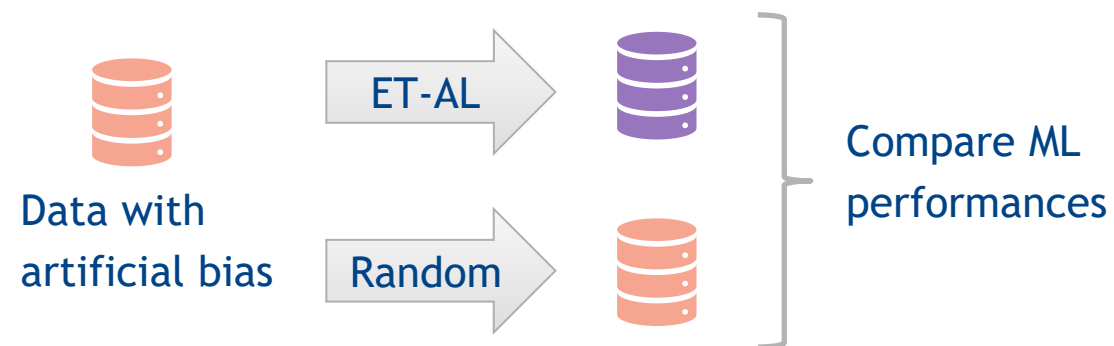
Example: the structure-stability bias in the Open Quantum Materials Database (OQMD)



# Lower Bias, Better ML Models

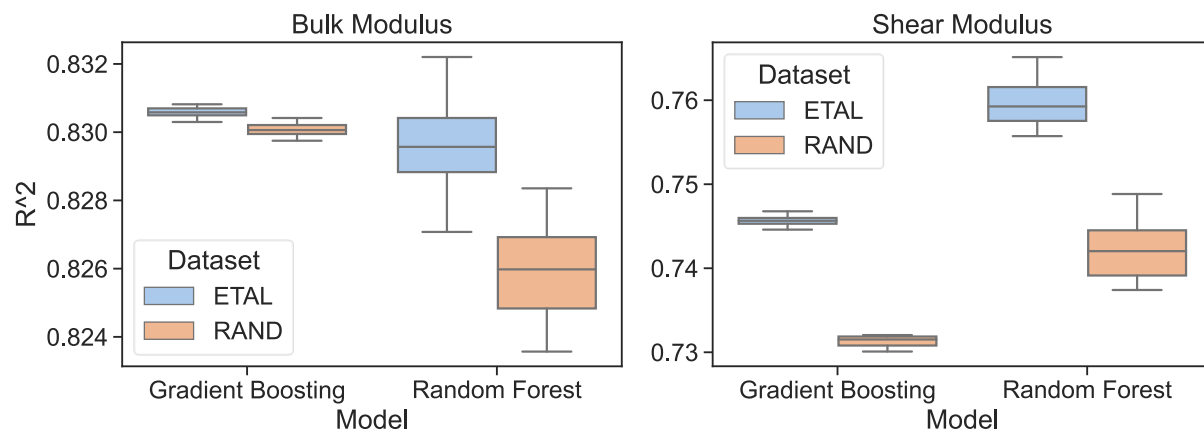


ET-AL effectively reduces bias



Experiments for demonstration on Jarvis dataset

## Reducing bias benefits ML for materials properties



## Applications

- Improve dataset quality
- Guide the construction of databases

**Thank you & welcome to our poster!**

Henry Zhang, Wayne Chen, James Rondinelli, Wei Chen  
November 8, 2022