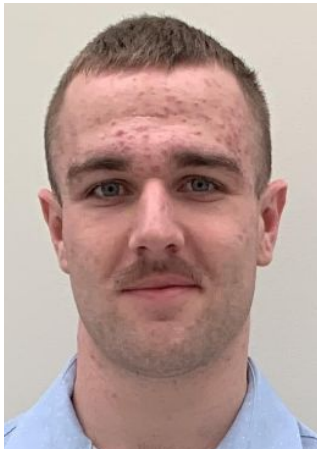
ECD 422

Machine learning enabled database for predictive computational design of perovskite materials for solar cells

BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Meet the Team

Advisor: Prof Mengen Wang



Team Lead: Alex Kinman

Database Engineer: Sanjitha Bhaskar

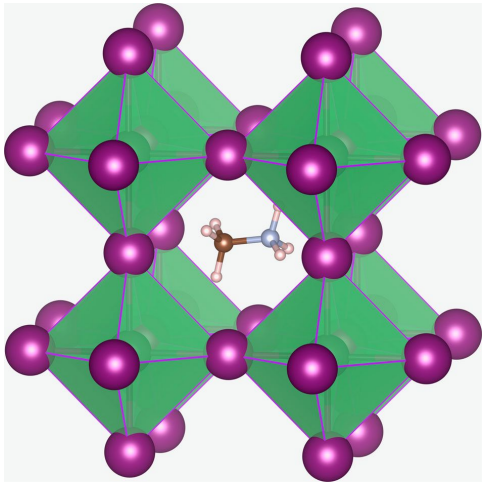
User Interface Engineer: Venkat Gutta

Project Summary

- Use machine learning for advanced development of perovskite solar cells
- Employ ML algorithm to predict outcome of first-principles calculations
- Propose optimal material compositions and synthesizing conditions for perovskites while avoiding traditionally high time and cost

Operational Context

Perovskite Material



Additional Data

- Elemental properties
- Formation environment
- Cell defect calculations
- Results of DFT calculations

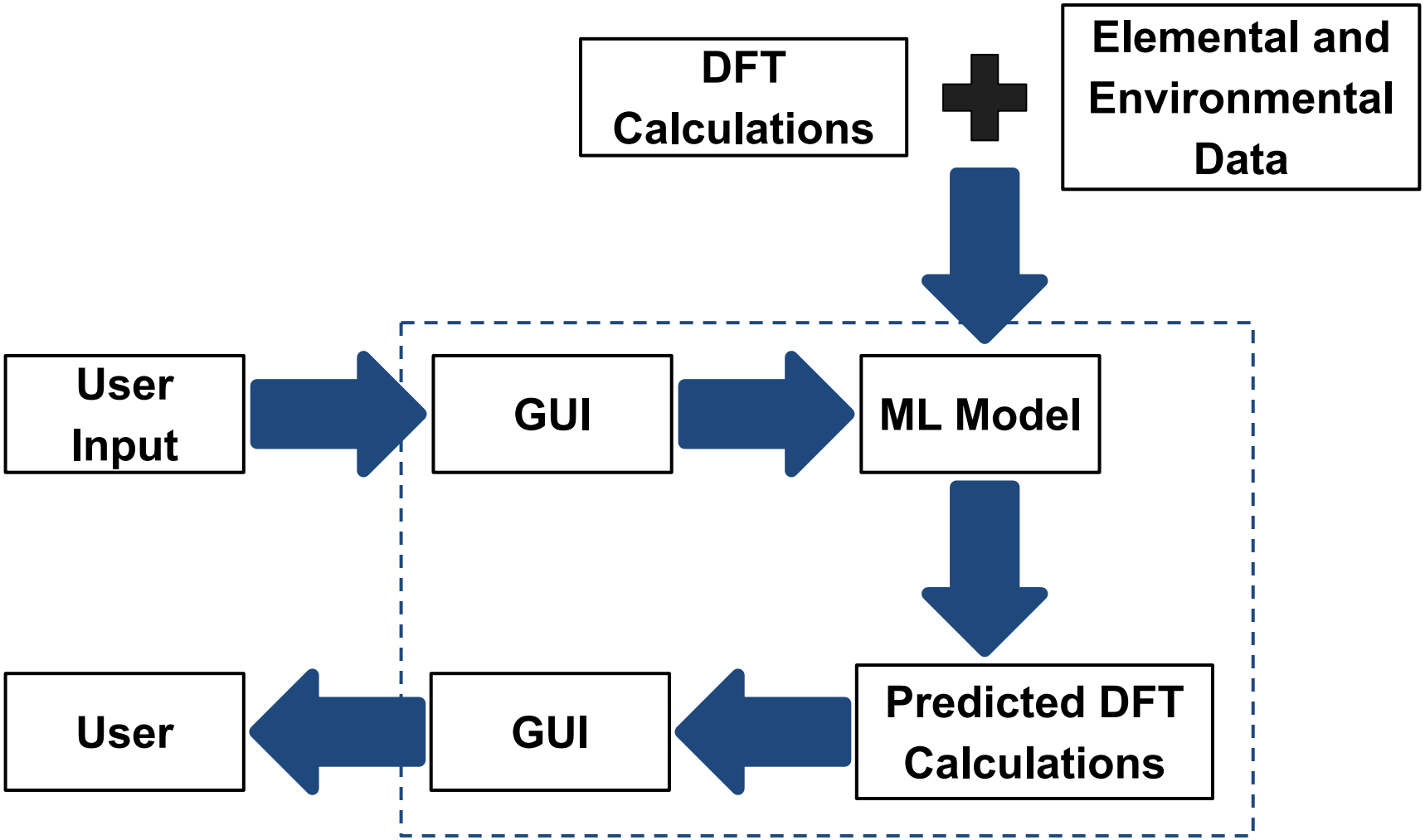
Density Functional Theory Calculations

- Costly
- Time intensive
- Requires high power computer

Machine Learning

- Requires data from DFT for training
- Nearly instant
- Low computational cost

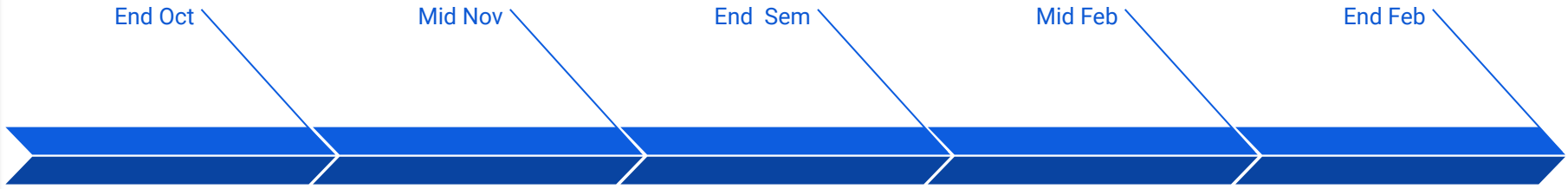
Operational Context



Specifications and Goals

- Shall use a machine learning algorithm to predict DFT calculations
- Shall create a GUI to display ML model output
- ML models shall be written in Python
- GUI may be written in Javascript or Python
- Should test more than one machine learning algorithm to determine best model
- May import the data into SQL Database
- Stretch goal: Should generate more data from DFT calculations to expand capabilities of ML model

Timeline



Train ML Models

Determine which ML models will be best for our application, begin training on existing data.

Test ML for Accuracy

Once models are trained, test and fine tune for accuracy then choose best model.

Develop GUI

Develop GUI to accept user inputs and output results using Google Sites.

Integrate ML Model into Website

Connect ML model to GUI using Google Sites.

Test ML and GUI Interaction

Ensure that the GUI is displaying correct results.

- Agile methodology

Engineering Tools

- VESTA
- VASP
- GitHub
- Google Colab
- SQL
- Anaconda
- Visual Studio code
- Google Sites

Current Status

- Three ML models chosen
  - Gaussian Process Regression
  - Neural Network
  - Random Forest Regression
- Data to begin training

Date	Summary	Cost	Total
-	Starting Budget	-	$1780
9/18/2023	Spiedie HPC subscription	$1,675	$105

Questions