**Who**: Ankita Biswas, Kwaku Acheampong, Suvo, Huanhuan Zhao (hzhao31@vols.utk.edu), Abid Amin Naeem, Collin Kovacs  
**Where**: 2025 LLM Hackathon for Materials  
**What**: Project Statement and Description

### **Project Name**

**QSPHAgent** *(Qualitative Structure-to-Property Hypothesis)*

### **Project Statement**

For a given structure, the agent will generate a hypothesis explaining how certain structural features contribute to the specific characteristics of the material’s density of states (DOS).

### Idea

Given a composition and all corresponding materials entries from the Materials Project that include DOS data:

1. Generate textual descriptions from the structural and DOS data.

2. Convert these descriptions into a **vector database** for retrieval-augmented generation (RAG).

3. Use a **reflection agent** to hypothesize **how and why** particular structural characteristics are related to observed DOS features.

### Output Example

**User**:  
I have a material composition <Al, O> with structural features <textual description of the structure>. Hypothesize how these structural features might correlate with its possible DOS.

**Agent**:  
Based on the retrieved knowledge, here is a possible hypothesis on how the density of states may behave:

*“The tetrahedral coordination around Al atoms and the short Al–O bond lengths are likely to result in a wide band gap and sharp oxygen p-orbital peaks below the Fermi level...”*

### **Goal**

Develop a hypothesis-generating algorithm that links material structure to density of states (DOS) features at the Fermi level.

### **Motivation**

The DOS at the Fermi level critically influences key properties such as electrical conductivity, magnetism, and superconductivity. Rapid screening of DOS characteristics is essential for accelerating materials design and discovery.

**Step 1: Data Collector (Tool 1)(Collin)**• Pull material data (e.g., CIF, DOS, metadata) from the Materials Project API. For a given composition.  
• Store raw and enriched data for later processing.

**Step 2: Feature Extraction Scripts (Tool 2) (Huanhuan)**• Use Python to extract:  
○ Structural info (lattice, symmetry, Wyckoff, CN)  
○ Chemical features (composition, valence, etc.)

Make sure all the features relevant to the DOS are extracted, and also, the goal is to return the output as a textual description. - Use Robocrystallizer

**Step 3: DOS Interpreter Tool (Tool 3) - DOSspell - Ankita**• Analyze DOS files to generate short textual summaries:  
○ Band gap type  
○ Peak locations  
○ Possible metallic/semiconducting behavior

Make sure to return the output as a textual description.

**Step 4: Vector Database Builder (Tool4)**• Convert material data (or summaries) into embeddings.  
• Store them in a vector database for semantic retrieval by the LLM.

Combine the structure description and DOS description into a vector database.

**Step 5: Final: Reflection + Hypothesis Agent (Suvo )**• LLM receives:  
○ Structural features  
○ DOS summaries  
○ Retrieved material context (via RAG)  
• Then it generates hypotheses:

Project UI [If time permits]

·

A working Streamlit/Gradio demo app.

·

Integrated backend calls to the RAG index and fine-tuned LLM.

·

Clean presentation of results with citations and optional plots.

This final step pulls everything

together into a **usable product**. It shows the power of the pipeline — starting

from raw materials data, moving through ML + RAG + LLM fine-tuning, and ending

with an interactive tool that can answer questions about electronic DOS at the

Fermi level.