

Material Informatics
& Mechanics Project

Comprehensive LLM to explain Corrosive properties of High-Entropy Alloy

Presented by Team 4

Ashwin – CB.AI.U4AID23029

Rithvik – CB.AI.U4AID23032

Aksay – CB.AI.U4AID23003

Deep – CB.AI.U4AID23004

Motivation

Once the project is completed, and implemented on a wider scale of the material informatics industry:



- The search for properties of HEA will be eased.
- HEA will also find a higher demand in the market with unique implementations.
- The materials industry will continue to expand with an increase in implementation.
- The chemical stability of various metals and HEAs will come to light.
- The project idea can be applied to other fields of search.

Problem Statement

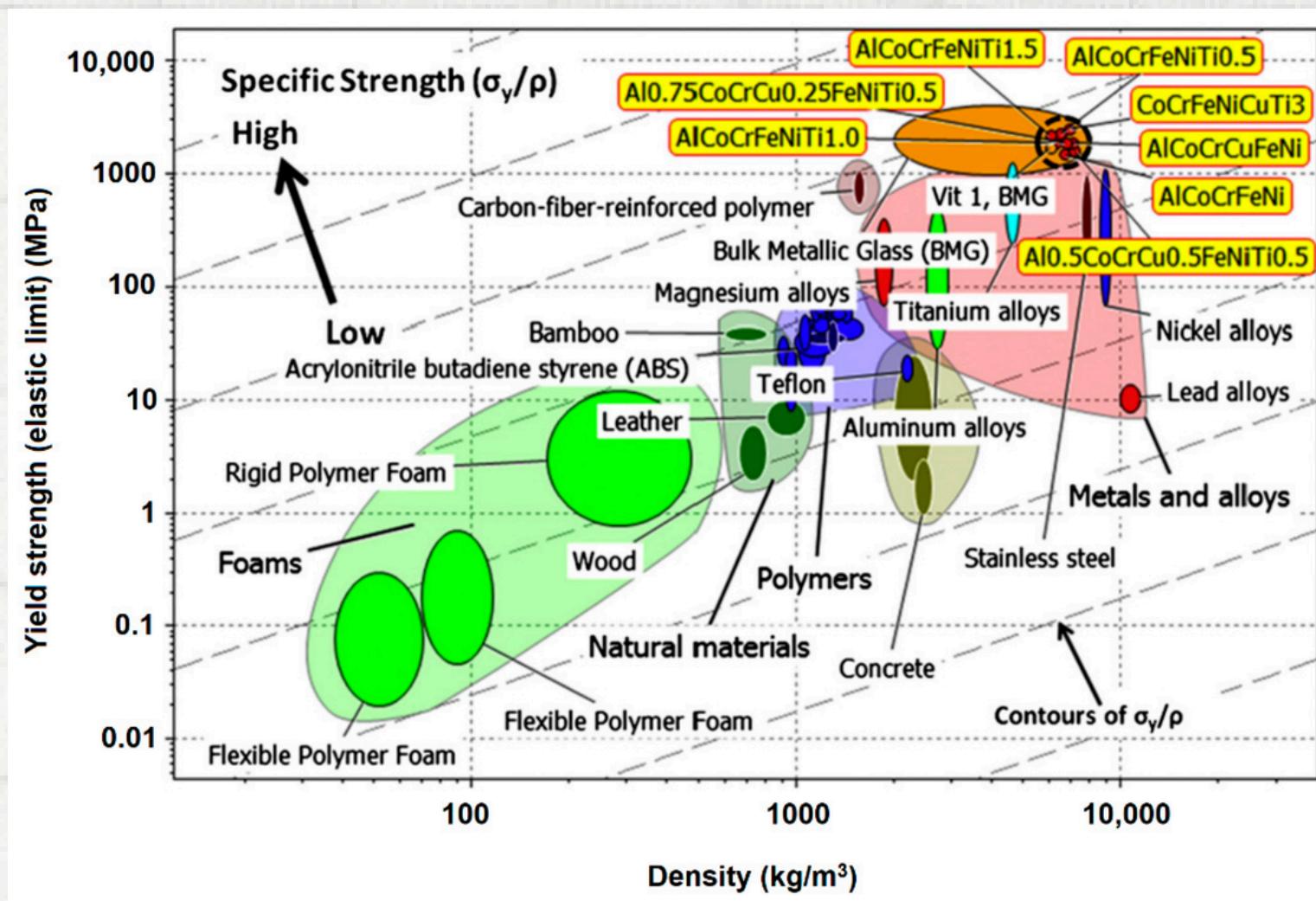
Despite HEA being a rather new topic, it has a huge data set even for a specific property of corrosion.

Our Approach

We're designing a chatbot to easily navigate through the huge dataset and to understand its specific properties by answering questions.

High Entropy Alloy

- They are a new class of metallic materials with exceptional properties.
- They are composed of five or more elements in near-equal atomic percentages, unlike traditional alloys.
- The high degree of disorder in their atomic structure gives HEAs superior strength, corrosion resistance, and thermal stability.
- HEAs are being considered for various applications in industries such as aerospace, automotive, and energy.



Methodology

01 Extraction and Summarization of Papers

- **PAPER EXTRACTION (Scopus):**
 - Using Scopus, we downloaded papers relating to HEAs to pdf files.



ELSEVIER
Scopus

- **PAPERS SUMMARIZATION (Bart):**
 - Using Meta Ai's "Bart", we summarized each of these pdf files to .txt files.



02

Generating dataset for test & train

- **DATASET GENERATION(Llama 3):**

- The summarized papers are used to generate custom dataset using Llama 3.

```
{  
  "question": "What is the typical morphology of AlCoCrFeNiTax alloys?",  
  "answer": "The AlCoCrFeNiTax alloys had typical dendritic morphology."  
},
```



- **FINAL DATASET:**

- The custom dataset of each paper were concatenated to form the final dataset.

03

Training the LLM on the Dataset

- **UnSloth** makes fine-tuning of LLMs 2.2x faster and use 80% less VRAM.
- **PEFT** module contains Optimised LoRA algorithm.
- **LoRA**, using matrix decomposition, drastically reduces computational power.

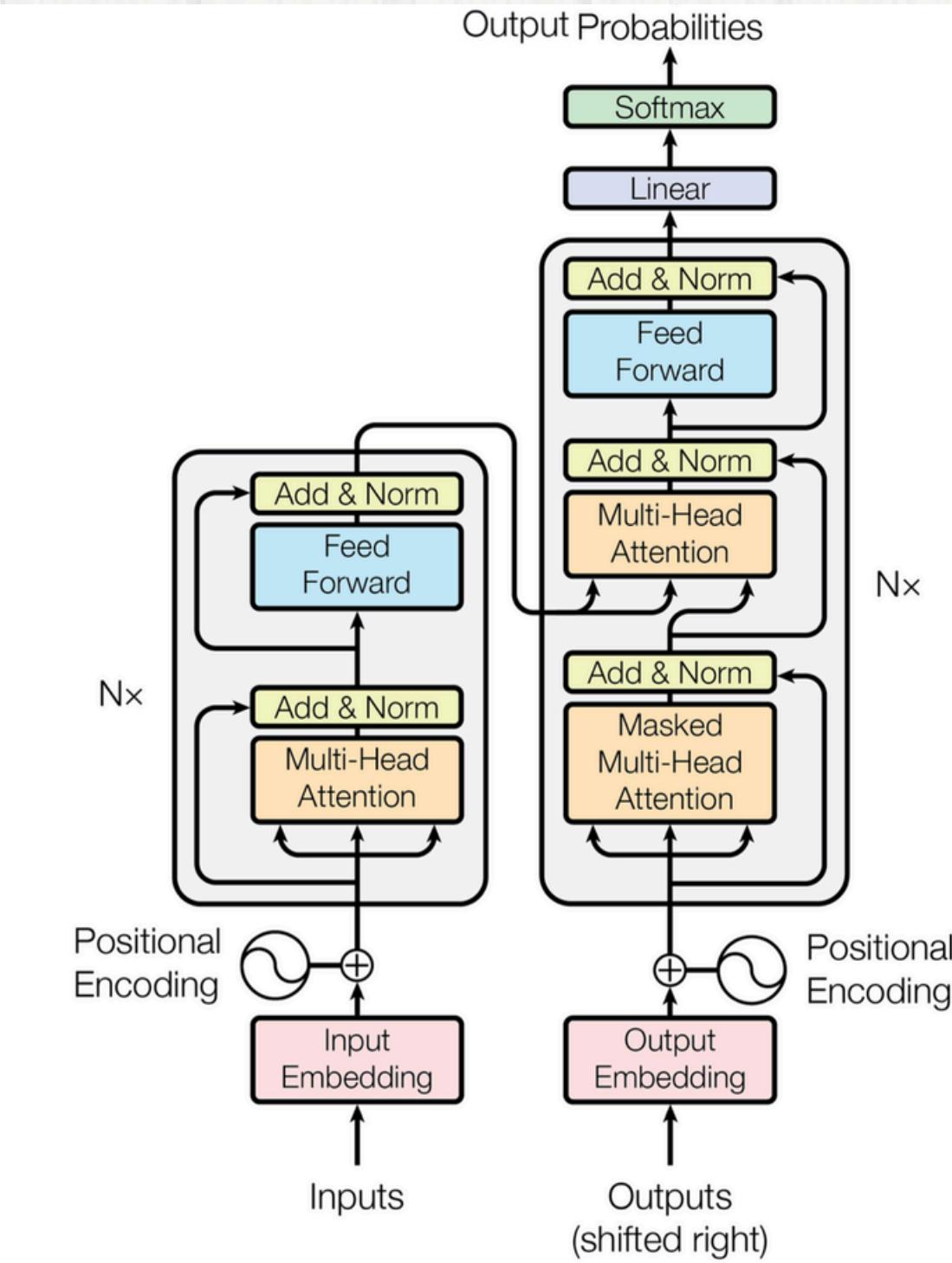


Figure 1: The Transformer - model architecture.

QLora

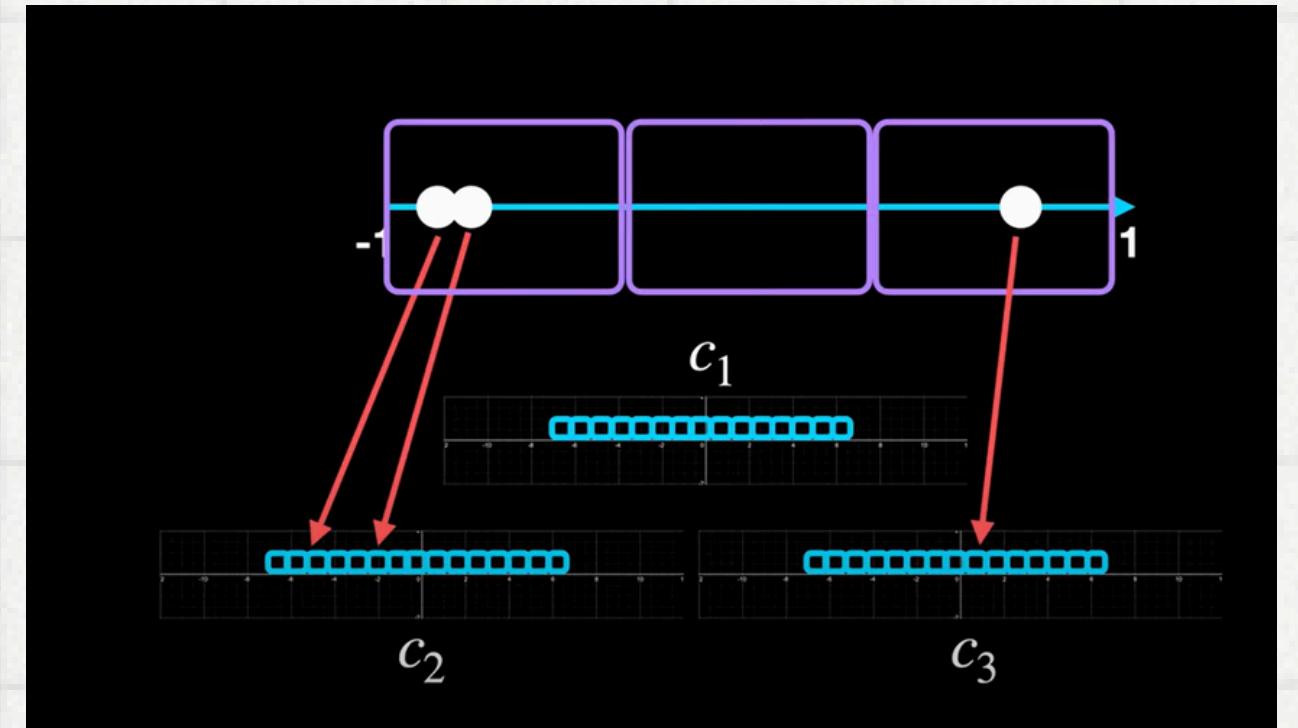
03

- **Low-Rank Adaptor (LoRA):**

- The weights are taken in Matrix format.
- The matrix is reduced based on its rank.

- **Quantization:**

- Weights are mapped to lower bits by blocks.
- They usually mapped from 32 bits to 4 bits.
- The Quantization produces Quantization constants which are again quantized called double quantization.

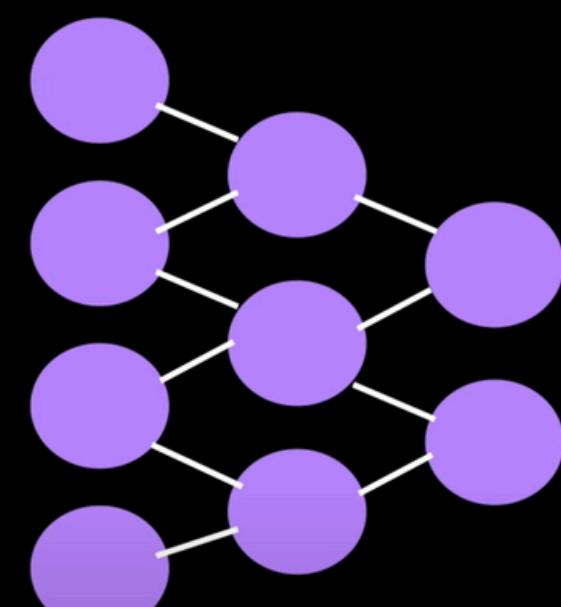


03

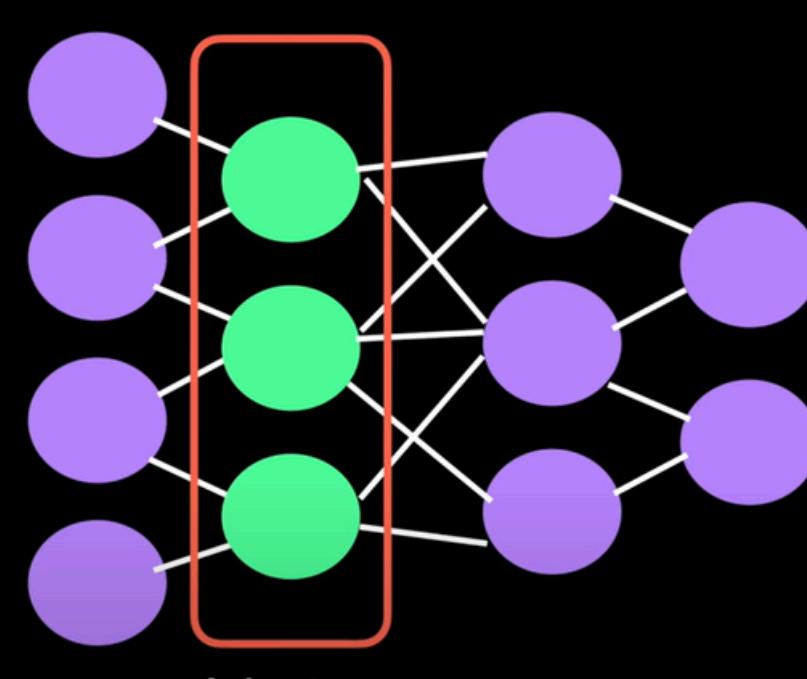
$$W_0 + \Delta W = W_0 + BA$$



Large Model

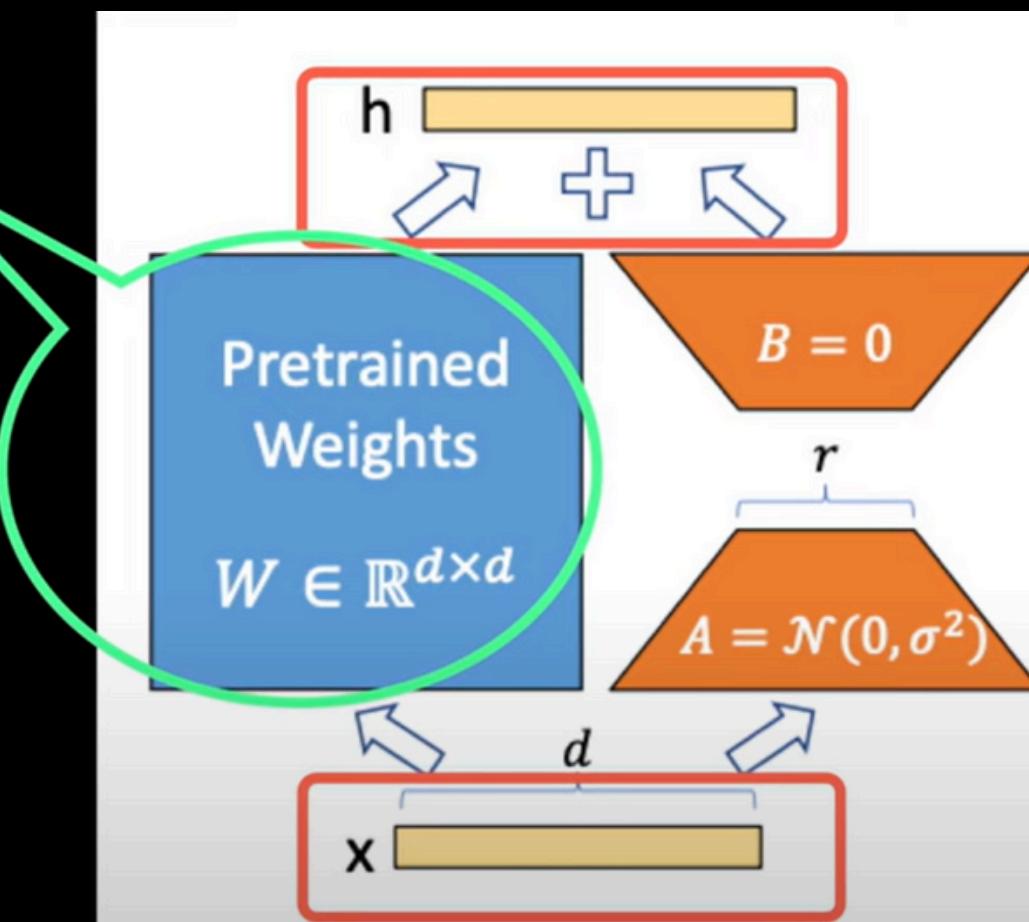


Pre-Trained
Model



Adapters

Frozen



Result

Instruction:

What is the significance of the TiC-TiB₂ system for HEA coatings?.

Input:

Response:

The TiC-TiB₂ system for HEA coatings is significant because it allows for the creation

Conclusion

- A dataset is created from papers directly.
- The model's LLM is trained on a question-answer dataset.
- Project uses bart, Llama 3, transformer models.
- QLORA and unsloth techniques are also utilized.

Future Scope



- There is significant potential for advancements and application of HEAs.
- Opportunities arise from industries ranging from aerospace and automotive to energy and infrastructure.
- Such LLM will also be in huge demand.
- The model can be used with various other datasets.
- It will help us to dive deeper into topics which has a lot of source to read from.

**Thank you
very much!**