IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153142

aLLoyM: "Phase Diagram Prediction System"

Sakshi K. Kamble[1], Anagha G. Harshe[2], Soniya C. Dhupdale[3], Shubham U. Dharwat[4], Soham P. Kapileshwar[5]

Department of Computer Science and Engineering (AI),

DKTE Society's Textile & Engineering Institute (An Empowered Autonomous Institute), India[1-5]

Abstract: Phase diagrams are essential thermodynamic tools that describe equilibrium phase stability as functions of temperature and composition in alloy systems. They guide alloy design, heat-treatment optimization, and microstructural control in critical industries. However, experimental determination of phase diagrams is costly and time-intensive. Although computational approaches such as CALPHAD (Calculation of Phase Diagrams) provide reliable thermodynamic modelling through Gibbs free energy minimization, they depend on curated parameter databases and expert assessment, limiting rapid exploration of new material systems.

In this work, we introduce aLLoyM, a domain-adapted Large Language Model (LLM) developed for structured alloy phase diagram prediction. Thermodynamic data generated from CALPHAD assessments in the Computational Phase Diagram Database (CPDDB), covering 389 binary and 38 ternary systems, were systematically sampled to produce over 800,000 equilibrium data points. These data were transformed into multi-task Question–Answer (Q&A) pairs and used to finetune the Mistral-Nemo-Instruct model via Low-Rank Adaptation (LoRA), enabling efficient domain specialization.

The framework supports three thermodynamic reasoning tasks: full phase information prediction, phase name inference, and inverse experimental condition prediction. Performance was evaluated under both interpolation and extrapolation settings to assess generalization. Results show substantial improvement over baseline LLM performance and demonstrate the model's ability to infer plausible phase behaviour for previously unseen systems.

These findings highlight the potential of integrating

Large Language Models with computational thermodynamics to develop scalable AI-assisted tools for accelerating alloy design and materials discovery.

1. INTRODUCTION

Phase diagrams are essential tools in materials science, providing insight into the thermodynamic stability of materials under varying composition and temperature conditions. Accurate phase diagram prediction enables efficient alloy design and optimization. Traditionally, phase diagrams are obtained through experimental investigations or computational approaches such as CALPHAD (CALculation of PHAse Diagrams). While these methods are reliable, they are resource-intensive and require expert intervention.

Recent advances in machine learning have introduced data-driven approaches for phase diagram prediction. Conventional models such as neural networks, random forests, and support vector machines have demonstrated promising results. However, these models are often system-specific and lack generalization capability.

Large Language Models (LLMs), including GPT-based and transformer-based architectures, offer a new paradigm by integrating domain knowledge learned during pre-training with task-specific fine-tuning. Unlike traditional models, LLMs can handle diverse inputs and generate structured outputs using natural language reasoning.

In this context, we introduce **aLLoyM**, a fine-tuned LLM designed to predict alloy phase diagrams using composition and temperature as inputs. The model is trained on large-scale Q&A datasets derived from CPDDB using CALPHAD assessments. The objective of this work is to evaluate whether LLMs can effectively learn thermodynamic relationships and extrapolate phase behaviour to unknown systems.

2. DATASET AND Q&A CONSTRUCTION

The training data for aLLoyM is derived from the **Computational Phase Diagram Database (CPDDB)**, an open-source repository maintained by NIMS. Thermodynamic database (TDB) files corresponding to **389 binary** and **38 ternary** alloy systems were collected.
Phase diagram calculations were performed using CALPHAD assessments with systematic sampling:

- Binary systems:
- Composition range: 0–100% (step size: 2%)
- Temperature range: 200 K – 5000 K (step size: 50 K)
- Ternary systems:

- Temperature fixed at 800 K

This procedure resulted in 837,475 data points, each containing composition, temperature and phase information.

From these data points, structured Question-and-Answer (Q&A) pairs were generated for three tasks:

1. Full phase information predictions
2. Phase name prediction
3. Experimental condition inference

These tasks were jointly used to fine-tune a single LLM.

3.PROPOSED METHODOLOGY

The proposed system, **aLLoyM**, is designed to integrate thermodynamic phase diagram data with a fine-tuned Large Language Model (LLM) to enable accurate and structured alloy phase diagram prediction. The overall methodology follows a systematic, multi-stage pipeline that transforms raw thermodynamic simulation data into language-based learning tasks suitable for LLM training and inference. The complete framework consists of **five major stages**, namely **data acquisition, dataset transformation, model adaptation, evaluation, and deployment**. Each stage plays a critical role in ensuring that the model effectively learns the underlying relationships between alloy composition, temperature, and phase stability.

3.1 Data Acquisition and Preprocessing

The first stage of the proposed system involves acquiring highquality thermodynamic phase equilibrium data from the **Computational Phase Diagram Database (CPDDB)**. CPDDB is an open-source repository that provides curated thermodynamic database (TDB) files derived from CALPHAD assessments. These databases encode Gibbs free energy descriptions of phases and are widely used for phase diagram construction.

In this work, thermodynamic data corresponding to:

- 389 binary alloy systems, and
- 38 ternary alloy systems were collected from CPDDB. These systems were deliberately selected to ensure a wide diversity of elemental combinations, compositional behaviours, and phase stability characteristics, thereby improving the generalization capability of the trained model.

Using CALPHAD-based simulations, phase equilibria were computed across systematic temperature and composition grids. For each alloy system, the following thermodynamic attributes were extracted:

- Temperature
- Elemental composition percentages
- Stable phase names
- Phase fractions (where applicable)

The sampling strategy was designed to comprehensively cover the thermodynamic space:

- Binary systems
- Composition range: 0–100% (with 2% increments)
- Temperature range: 200 K to 5000 K (with 50 K increments)
- Ternary systems
- Composition grid sampling across three components
- Temperature fixed at 800 K to generate isothermal sections

This systematic sampling process resulted in approximately **837,475 equilibrium data points**, where each data point represents a unique mapping of:

Composition + Temperature = Equilibrium Phase Information

This preprocessing step ensures that the dataset is suitable for downstream transformation into language-based formats.

DOI: 10.17148/IJARCCE.2026.153142

3.2 Structured Q&A Dataset Construction

Three distinct types of Q&A tasks were constructed, each targeting a different a different aspect of phase diagram reasoning:

1. Phase Name Prediction

In this primary prediction task, the model is provided with alloy composition percentages and temperature as input, and it is required to predict the corresponding stable phase or phases. This task directly reflects the conventional usage of phase diagrams and forms the core objective of the proposed system. **2. Detailed Phase Information Prediction**

In this task, the model is required to generate not only the phase names but also their approximate phase fractions under given thermodynamic conditions. This enables the model to learn multi-phase coexistence behaviour and capture quantitative phase distribution information, which is crucial for practical materials design.

3. Inverse Condition Prediction

This inverse task strengthens the model's relational understanding by reversing the prediction direction. Here, the model is given a specific phase name and is asked to predict feasible temperature and composition ranges where that phase is thermodynamically stable. This task allows the model to infer experimental conditions associated with a desired phase, enhancing its usefulness for exploratory materials research.

3.3 Model Selection

For the core learning engine of the proposed system, the **MistralNemo-Instruct-2407** model was selected as the base Large Language Model. This model was chosen due to its strong reasoning capabilities, instruction-following behaviour, and relatively manageable computational requirements compared to larger proprietary LLMs. The transformer-based architecture of Mistral enables effective contextual learning and supports multitask training, making it suitable for handling diverse thermodynamic Q&A tasks.

Rather than training a model from scratch, a **fine-tuning approach** was adopted to leverage the extensive general knowledge already encoded in the pre-trained model. This strategy significantly reduces training cost while allowing the model to adapt to domain-specific thermodynamic prediction tasks.

3.4 Model Architecture and Fine-Tuning

Fine-tuning of the Mistral-Nemo-Instruct-2407 model was performed using **Low-Rank Adaptation (LoRA)**. LoRA enables efficient parameter adaptation by injecting trainable low-rank matrices into selected layers of the transformer architecture, while keeping the majority of the original parameters frozen. This approach improves training efficiency and reduces the risk of overfitting.

The fine-tuning configuration used in this work is summarized as follows:

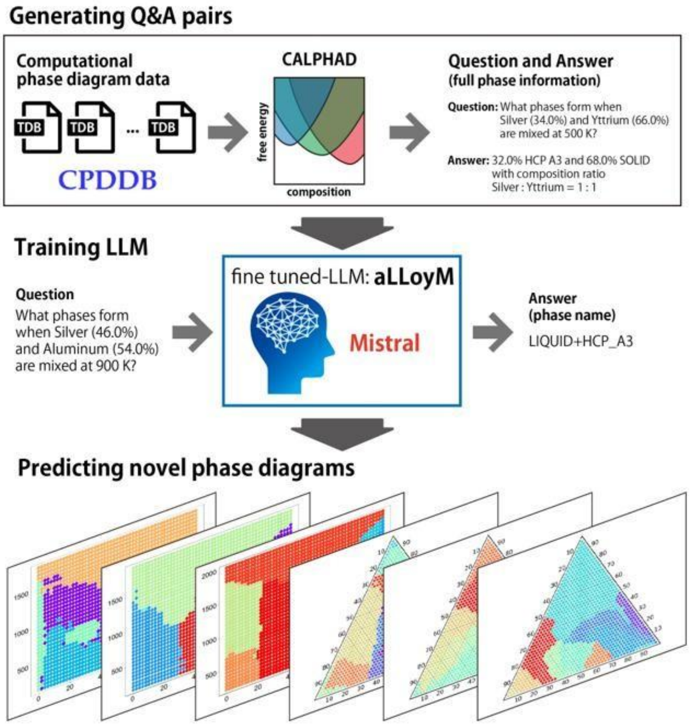
- LoRA rank: 16
- LoRA alpha: 16
- Learning rate: 2×10^{-4}
- Training steps: 15,000
- Optimizer: AdamW
- Numerical precision: bfloat16

The training prompts were designed using a structured template consisting of three components:

- Instruction: Describes the prediction task
- Input: Provides thermodynamic conditions or phase information
- Output: Contains the expected answer

DOI: 10.17148/IJARCCE.2026.153142

3.5 Schematic Representation of the Proposed System



Question: What phases form when Silver (34.0%) and Yttrium (66.0%) are mixed at 500 K?

Answer: 32.0% HCP A3 and 68.0% SOLID with composition ratio Silver : Yttrium = 1 : 1

What phases form when Silver (46.0%) and Aluminum (54.0%) are mixed at 900 K?

LIQUID+HCP_A3

The schematic diagram illustrates the complete workflow of the proposed system. It shows how thermodynamic data is obtained from CPDDB, processed through CALPHAD assessments, transformed into structured Q&A datasets, and finally used to fine-tune the Mistral LLM. The figure provides a clear visual summary of how data-driven thermodynamic modelling is integrated with language-based learning to form the aLLoyM framework.

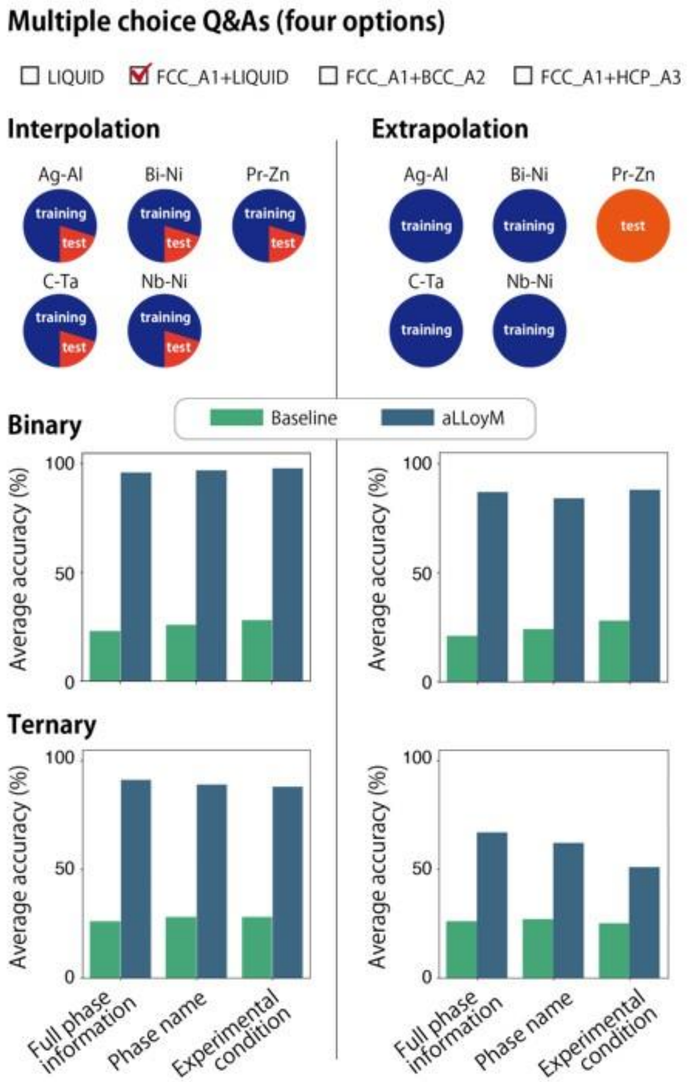
4. EXPERIMENTAL EVALUATION AND RESULT

4.1 Multiple-Choice Q&A Evaluation

For benchmarking, the dataset was split using an 8:2 ratio into training and test sets using two strategies:

- **Interpolation split:** Same systems with unseen compositions
- **Extrapolation split:** Completely unseen alloy systems

Each multiple-choice question contained one correct answer and three distractors.



Results show that:

- The baseline model performs close to random guessing
- aLLoyM achieves significantly higher accuracy
- Interpolation accuracy > Extrapolation accuracy
- Binary systems are easier to predict than ternary systems

4.2 Short-Answer Evaluation

Short-Answer tasks allow free-form prediction without predefined options. Performance is evaluated using task -specific scoring metrices.

Scoring Formulas

Full phase information score:

$$Score = \begin{cases} 100\%, & \text{exact match} \\ 0\%, & \text{otherwise} \end{cases}$$

Phase name score (Jaccard similarity):

$$Score = \frac{|A \cap B|}{|A \cup B|} \times 100\%$$

Experimental condition score:

$$A_c = \frac{1}{N} \sum_{n=1}^{N} \left(1 - \frac{|x_n^* - x_n|}{100}\right)$$

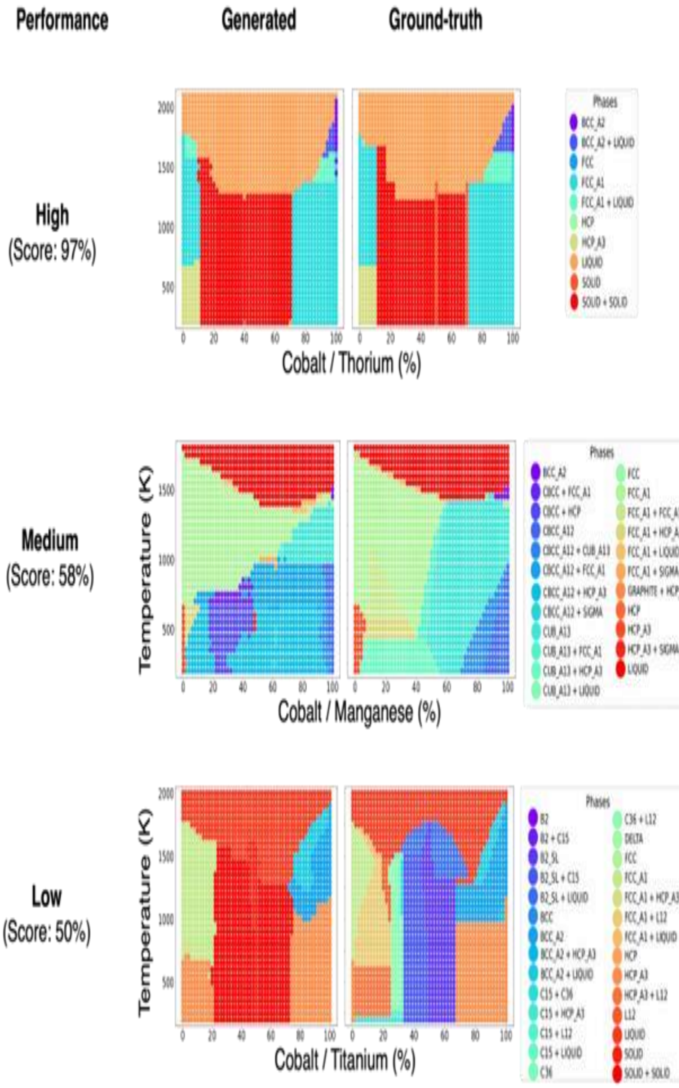
$$A_t = 1 - \frac{|T^* - T|}{\Delta T}$$

$$Score = \left(\frac{A_c + A_t}{2}\right) \times 100\%$$

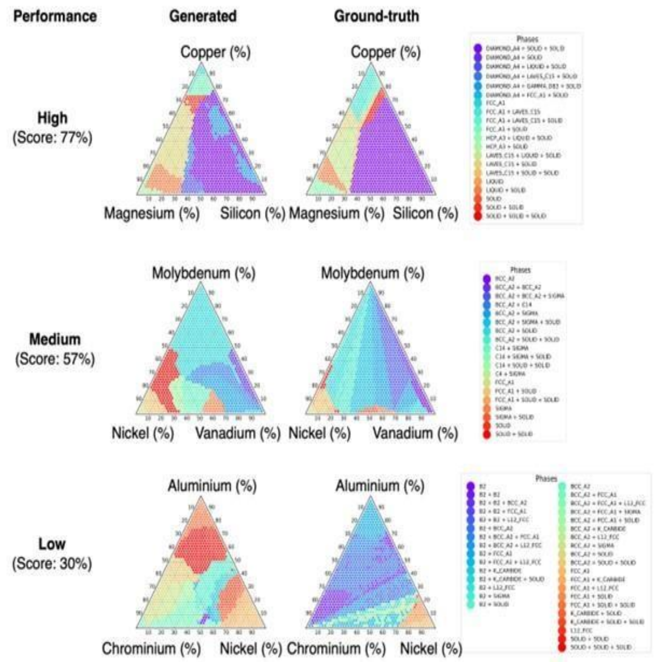
4.3 Phase Diagram Reconstruction

Using prediction phase names, full phase diagrams were reconstruction.

- Binary Phase Diagram:



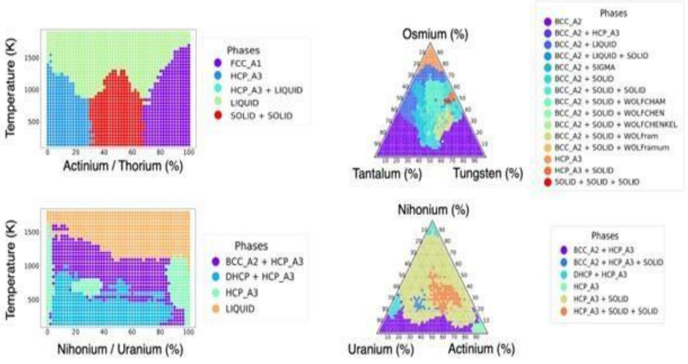
- Ternary isothermal sections (800 K):



Performance degrades in complex intermediate composition regions, indicating increased prediction difficulty.

4.4 Novel Phase Diagram Generation

aLLoyM demonstrates the ability to generate previously unknown phase diagrams.



This includes:

- Th-Ac
- U-Nh
- W-Ta-Os (800 K)
- Nh-U-Ac (hypothetical systems)

5. CONCLUSION

In this research, we proposed **aLLoyM**, a domain-adapted Large Language Model specifically designed for alloy phase diagram prediction. The study demonstrates how structured thermodynamic knowledge derived from CALPHAD assessments can be effectively integrated with modern transformer-based architectures to support advanced materials science applications. By systematically converting CALPHAD-generated phase equilibrium data into structured Question–Answer formats, the model was trained to learn the intrinsic relationships between alloy composition,

DOI: 10.17148/IJARCCE.2026.153142

temperature, and phase stability across binary and ternary systems. This structured learning framework enables the model to capture equilibrium patterns embedded within largescale thermodynamic datasets.

The application of Low-Rank Adaptation (LoRA) enabled efficient fine-tuning of the base Mistral model without requiring full-scale retraining of all parameters, thereby reducing computational cost while preserving reasoning capability. Experimental evaluation under both interpolation and extrapolation settings demonstrated that the fine-tuned model significantly outperformed the baseline model across phase name prediction, full phase information inference, and inverse condition prediction tasks. The observed generalization performance in extrapolation experiments suggests that the model learned meaningful thermodynamic patterns rather than merely memorizing training instances, indicating its potential applicability to previously unseen alloy systems.

Although the proposed system does not replace established computational thermodynamics methods such as CALPHAD, it provides a fast and scalable complementary framework for preliminary phase estimation. Such an AI-assisted approach can help researchers rapidly screen large compositional spaces and identify promising composition–temperature regions prior to performing detailed thermodynamic simulations or costly experimental investigations. This capability is particularly valuable in accelerating exploratory materials design and reducing trial-and-error experimentation.

Overall, this work highlights the growing potential of Artificial Intelligence in materials informatics and thermodynamic modelling. The successful adaptation of a Large Language Model for structured equilibrium prediction illustrates that foundation models, when equipped with domain-specific training data, can contribute to scientific reasoning tasks beyond conventional text processing. The integration of datadriven language models with computational thermodynamics represents a promising direction toward intelligent, scalable, and accessible tools for next-generation alloy design and materials discovery.

REFERENCES

[1] H. L. Lukas, S. G. Fries, and B. Sundman, *Computational Thermodynamics: The CALPHAD Method*. Cambridge, U.K.: Cambridge University Press, 2007.

[2] T. B. Massalski, *Binary Alloy Phase Diagrams*, 2nd ed. Materials Park, OH, USA: ASM International, 1990.

[3] P. Villars, A. Prince, and H. Okamoto, *Handbook of Ternary Alloy Phase Diagrams*. Materials Park, OH, USA: ASM International, 1995.

[4] N. Saunders and A. P. Miodownik, *CALPHAD (Calculation of Phase Diagrams): A Comprehensive Guide*. Oxford, U.K.: Pergamon Press, 1998.

[5] T. Brown et al., “Language models are few-shot learners,” in *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[6] E. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2022.

[7] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, “Machine learning for molecular and materials science,” *Nature*, vol. 559, no. 7715, pp. 547–555, 2018.

[8] L. Ward and C. Wolverton, “Atomistic calculations and materials informatics: A review,” *Current Opinion in Solid State and Materials Science*, vol. 21, no. 3, pp. 167–176, 2017.