Supplementary Information

Minh-Quyet Ha,[1] Dinh-Khiet Le,[1] Viet-Cuong Nguyen,[2] Hiori Kino,[3] Stefano Curtarolo,[4,5] and Hieu-Chi Dam[1,6,a)]

[1)] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

[2)] *HPC SYSTEMS Inc., 3-9-15 Kaigan, Minato, Tokyo 108-0022, Japan*

[3)] *Research Center for Materials Informatics, Department of Advanced Data Science, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

[4)] *Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA*

[5)] *Center for Extreme Materials, Duke University, Durham, NC 27708, USA*

[6)] *International Center for Synchrotron Radiation Innovation Smart (SRIS), Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*

(Dated: 1 December 2025)

[a)]Electronic mail: dam@jaist.ac.jp

I. TUNING HYPER-PARAMETERS OF THE PREDICTIVE MODELS

Because the datasets used in this study are derived from calculation-based prediction methods, we introduce a degree of uncertainty α into the mass function that models substitutability evidence from material datasets. For each dataset, a grid search is performed to determine the optimal α, which best reproduces the alloy labels by achieving the highest cross-validation score. The cross-validation scheme employs a 10-fold cross-validation repeated three times. The search space for α includes the values $[0.001, 0.005, 0.01, 0.05, 0.1, 0.5]$. The optimal α values for the datasets $\mathcal{D}_{0.9T_m}$, $\mathcal{D}_{1350\mathrm{K}}$, and $\mathcal{D}_{\mathrm{Mag}}$ are found to be 0.01, while for the dataset $\mathcal{D}_{T_C}$, the optimal value is 0.05.

In contrast, the degree of uncertainty β in the mass function representing substitutability evidence from domain knowledge is assigned a fixed value of $1/N_{domains}$, where $N_{domains}$ denotes the number of knowledge domains applied in the study. This ensures balanced weighting of evidence across multiple sources.

The computational efficiency of our framework varies across datasets, depending on both the number of data instances and knowledge sources. For the stability datasets $\mathcal{D}_{0.9Tm}$ and $\mathcal{D}_{1350\mathrm{K}}$, which contain 14,950 alloys, the runtime per fold of our hybrid framework is 3.5 hours. In contrast, for $\mathcal{D}_{\mathrm{Mag}}$ and $\mathcal{D}_{T_C}$, each containing 5,968 alloys, the runtime per fold is approximately 0.5 hours. To enhance computational efficiency, all experiments are implemented in parallel using Apache Spark and executed on a 32-CPU cluster.

Additionally, the grid search parameter tuning is also applied to other traditional models. The parameter grid for logistic regression is shown in Supplementary Table II.

II. COMBINING PIECES OF EVIDENCE USING DEMPSTER'S RULE OF COMBINATION

We assume that we can collect q pieces of evidence to compare a specific pair of element combinations, C_t and C_v. If no evidence is found, the mass function $m_{none}^{C_t,C_v}$ is initialized, which assigns a probability mass of 1 to subset $\{similar, dissimilar\}$. $m_{none}^{C_t,C_v}$ models the condition under which no information about the similarity (or dissimilarity) between C_t and C_v is available. Any two pieces of evidence a and b modeled by the corresponding mass func-

tions $m_a^{C_t,C_v}$ and $m_b^{C_t,C_v}$ can be combined using the Dempster rule[1] to assign the joint mass $m_{a,b}^{C_t,C_v}$ to each subset ω of Ω_{sim} (i.e. $\{similar\}$, $\{dissimilar\}$, or $\{similar, dissimilar\}$) as follows:

$$\begin{aligned} m_{a,b}^{C_t,C_v}(\omega) &= \left(m_a^{C_t,C_v} \oplus m_b^{C_t,C_v}\right)(\omega) \\ &= \frac{\sum\limits_{\forall\omega_k\cap\omega_h=\omega} m_a^{C_t,C_v}(\omega_k) \times m_b^{C_t,C_v}(\omega_h)}{1 - \sum\limits_{\forall\omega_k\cap\omega_h=\emptyset} m_a^{C_t,C_v}(\omega_k) \times m_b^{C_t,C_v}(\omega_h)}, \end{aligned} \tag{1}$$

where ω, ω_k and ω_h are subsets of Ω_{sim}. Note that the Dempster rule is commutative and yields the same result by changing the order of $m_a^{C_t,C_v}$ and $m_b^{C_t,C_v}$.

All the q obtained mass functions corresponding to the q collected pieces of evidence from available sources are then combined using the Dempster rule to assign the final mass m^{C_t,C_v} as follows:

$$m^{C_t,C_v}(\omega) = \left(m_1^{C_t,C_v} \oplus m_2^{C_t,C_v} \oplus \cdots \oplus m_q^{C_t,C_v}\right)(\omega). \tag{2}$$

III. ILLUSTRATIVE EXAMPLES

The following examples provide explanations of how the evidence theory work to learn the similarity and infer the HEA formation for new element combinations, identifying equiatomic alloys.

Example 1: Suppose we have collected four pairs of alloys from experiments. Three of those pairs are alloys that both form HEA phase: $pair_1 = (\{A^1, B^1, C^1, D\}, \{A^1, B^1, C^1, E\})$; $pair_2 = (\{A^2, B^2, C^2, D\}, \{A^2, B^2, C^2, E\})$; and $pair_3 = (\{A^3, B^3, C^3, D\}, \{A^3, B^3, C^3, E\})$. The fourth pair $pair_4 = (\{A^4, B^4, C^4, D\}, \{A^4, B^4, C^4, E\})$ is different from the other three, in which $\{A^4, B^4, C^4, D\}$ forms HEA phase while $\{A^4, B^4, C^4, E\}$ does not form HEA phase. We consider each pair as a source of evidence support that $\{D\}$ is similar to $\{E\}$ in term of substitutability to form the HEA phase. Each evidence is modeled using mass function as follows:

$$\begin{aligned} &m_{pair_1}^{\{C\},\{D\}}(\{similar\}) = 0.1, \\ &m_{pair_1}^{\{C\},\{D\}}(\{dissimilar\}) = 0, \\ &m_{pair_1}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9 \end{aligned}$$

$$m_{pair_2}^{\{C\},\{D\}}(\{similar\}) = 0.1,$$
$$m_{pair_2}^{\{C\},\{D\}}(\{dissimilar\}) = 0,$$
$$m_{pair_2}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9$$
$$m_{pair_3}^{\{C\},\{D\}}(\{similar\}) = 0.1,$$
$$m_{pair_3}^{\{C\},\{D\}}(\{dissimilar\}) = 0,$$
$$m_{pair_3}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9$$
$$m_{pair_4}^{\{C\},\{D\}}(\{similar\}) = 0,$$
$$m_{pair_4}^{\{C\},\{D\}}(\{dissimilar\}) = 0.1,$$
$$m_{pair_4}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9$$

The three pieces of evidence are combined using the Dempster' rule of combination to accumulate the believe that $\{D\}$ is similar to $\{E\}$:

$$m^{\{C\},\{D\}}(\{similar\}) = 0.25,$$
$$m^{\{C\},\{D\}}(\{dissimilar\}) = 0.075,$$
$$m^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.675$$

Next, if we observed (included in the data) that the HEA phase exists for alloy $\{G, H, I, D\}$, the ERS (which focuses on finding some chance for discovering new combination of elements that the HEA phase exist and ignores the belief regarding $\neg HEA$) will consider that there is some believe that the HEA phase also exists for $\{G, H, I, E\}$ (by substituting $\{D\}$ with $\{E\}$). The evidence is modeled using mass function as follows:

$$m_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}^{\{G,H,I,E\}}(\{\neg HEA\}) = 0,$$
$$m_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}^{\{G,H,I,E\}}(\{HEA\}) = m^{C,D}(\{similar\}) = 0.25,$$
$$m_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}^{\{G,H,I,E\}}(\{HEA, \neg HEA\}) = 1 - m^{C,D}(\{similar\}) = 0.75$$

Example 2: In a same manner but for an extrapolative recommendation: if the HEA phases exist for all the alloys in the three following pairs: $pair_1 = (\{A^1, B^1, C\}, \{A^1, B^1, D, E\})$, $pair_2 = (\{A^2, B^2, C\}, \{A^2, B^2, D, E\})$, $pair_3 = (\{A^3, B^3, C\}, \{A^3, B^3, D, E\})$. In the fourth pair $pair_4 = (\{A^4, B^4, C\}, \{A^4, B^4, D, E\})$, $\{A^4, B^4, C\}$ forms HEA phase while $\{A^4, B^4, D, E\}$ does not form HEA phase. The algorithm will accumulate the believe that $\{C\}$ is similar to $\{D, E\}$ as follows:

$$m^{\{C\},\{D,E\}}(\{similar\}) = 0.25,$$
$$m^{\{C\},\{D,E\}}(\{dissimilar\}) = 0.075,$$
$$m^{\{C\},\{D,E\}}(\{similar, dissimilar\}) = 0.675$$

Consequently, if we observed (included in the data) that the HEA phase exists for $\{G, H, I, C\}$, the algorithm (which focuses on finding some chance for discovering new combination of elements that the HEA phase exist and ignores the belief regarding $\neg HEA$) will consider that there is some believe that the HEA phase also exists for $\{G, H, I, D, E\}$ (by substituting $\{C\}$ with $\{D, E\}$).

$$m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{\neg HEA\}) = 0,$$
$$m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{HEA\}) = m^{C,D}(\{similar\}) = 0.25,$$
$$m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{HEA, \neg HEA\}) = 1 - m^{C,D}(\{similar\}) = 0.75$$

IV. ALLOYS MAP RECONSTRUCTION FOR VISUAL ANALYSIS OF KNOWLEDGE TRANSFER

Alloy design involves high costs and risks, requiring decision support systems that can both extrapolate to unexplored compositional spaces and provide interpretable insights. In this study, we present a visual analysis method to illustrate the effect of expert knowledge transfer in materials extrapolation. Specifically, we employ t-distributed stochastic neighbor embedding (t-SNE)[2] to organize alloys based on shared compositions and phase behavior, integrating expert-derived knowledge within a hybrid distance matrix. By combining (1) empirical and knowledge-based dissimilarity scores with (2) Jaccard distances (for compositional overlap), this visualization exposes chemical relationships and mechanisms underlying phase stability. In doing so, it not only aids hypothesis generation and validation but also demonstrates how leveraging expert insights can guide researchers toward promising new alloys.

To identify the subgroup to which unobserved alloys belong, we propose an approach that constructs an embedding space for both observed and unobserved alloys based on a combined distance matrix $\overline{M}$. This matrix is computed as the element-wise product of the distance matrix obtained from our proposed substitutability-based similarity measurement

and the distance matrix derived using the Jaccard index for the binary descriptors of the alloys.

For any pair of alloys A_i and A_j, the combined distance is defined as follows:

$$\overline{M}[i,j] = \left(m^{C_t,C_v}(\{\text{dissimilar}\}) + \frac{m^{C_t,C_v}(\Omega_{sim})}{2} \right) \times (1 - J(A_i, A_j)), \tag{3}$$

where m^{C_t,C_v} represents the substitutability between the two element combinations, $C_t = A_i \setminus (A_i \cap A_j)$ and $C_v = A_j \setminus (A_i \cap A_j)$, obtained through our method, while $J(A_i, A_j)$ denotes the Jaccard similarity coefficient[3] between the element sets of the two alloys.

If one of the alloys in the pair is unobserved in the dataset, our method assigns no similarity information between the alloys, resulting in $m^{C_t,C_v}(\Omega_{sim}) = 1$ and $m^{C_t,C_v}(\{\text{dissimilar}\}) = 0$. In this case, the combined distance simplifies to:

$$\overline{M}[i,j] = \frac{1 - J(A_i, A_j)}{2} \tag{4}$$

This approach allows for the effective grouping of unobserved alloys by leveraging both compositional similarity and substitutability information.

V. ANALYSIS ON DATASET $\mathcal{D}_{0.9Tm}$ FOR EXTRACTION OF HEA FORMATION MECHANISM

Having demonstrated the value of knowledge-enhanced visualization for extrapolation to unexplored compositions (Section 4.3 in the main text), we next apply the t-SNE mapping techniques to uncover the fundamental mechanisms governing HEA formation across the broader compositional landscape on dataset $\mathcal{D}_{0.9Tm}$.

A. Learning about the Substitutability between Elements

Previous experiments demonstrate that integrating multi-source knowledge enables robust extrapolation in predicting HEA formation, even when novel elements are introduced. To further investigate the underlying mechanisms, we analyze the substitutability between elements based on data collected from both DS-source and LLM-source models. Supplementary Figures 3 and 4 presents a hierarchically clustered structure of constituent elements, constructed using substitutability information derived from both sources. This structure is

generated via hierarchical agglomerative clustering (HAC) with the complete linkage criterion, grouping elements based on similar substitutability patterns.

Clustering analysis reveals that all domains consistently agree on the behavior of coinage metals, with Cu, Ag, and Au exhibiting similar substitutability patterns across all sources. Other transition metals also form subclusters, although the exact clustering varies across different domains. A key and recurring pattern is the clear distinction between early transition metals (blue-labeled) and late transition metals (orange-labeled). Furthermore, Al, Si, and As cluster together, suggesting shared substitutability characteristics. While transition metals from groups 4 and 5, including Ta, Nb, Hf, Ti, and Zr, form another distinct cluster, suggesting strong substitutability relationships within this group.

Supplementary Figure 3 presents the hierarchically clustered structure of constituent elements based on LLM-derived substitutability data across four domains: corrosion, mechanical properties, materials science, and metallurgy. Notably, the structure derived from solid-state physics closely resembles that of metallurgy. Further details on solid-state physics are provided in Supplementary Figure 4.

Similar substitutability patterns are observed in datasets $\mathcal{D}_{0.9Tm}$ and $\mathcal{D}_{1350\mathrm{K}}$. The coinage metals form a cohesive group but exhibit a negative synthesized effect on phase stability; most alloys containing Cu, Ag, or Au do not form HEAs under either temperature conditions, as shown in Supplementary Figure 3. Additionally, a group of 14 transition metals, denoted as $\mathcal{E}$, demonstrates consistent substitutability characteristics. The synthesized effects of this group are discussed in Section V B.

Despite general consistency with domain knowledge, certain elements exhibit divergent substitutability behavior. Al and Si exhibit patterns similar to transition metals from groups 4 and 5, along with Pd, under both temperature conditions. Contrarily, As behaves distinctly, separating from both metalloid and transition metal clusters. This deviation impacts predictive performance during extrapolation experiments where alloys containing any of these elements are excluded from the training set, as illustrated in Supplementary Figure 5.

Substitutability data from the magnetization and Curie temperature datasets, $\mathcal{D}_{\mathrm{Mag}}$ and $\mathcal{D}_{T_C}$, reveal two dominant clusters corresponding to early and late transition metals, consistent with other sources. However, Mn, Fe, and Co form a separate subcluster, exhibiting positive effects on magnetic properties such as magnetization and Curie temperature Supplementary Figure 5. This subcluster diverges from LLM-based substitutability data, affecting

the performance of our hybrid framework when analyzing these metals. Supplementary Figure 5 shows that predictive performance for Mn, Fe, and Co is significantly lower than for other elements in the magnetization and Curie temperature datasets.

In summary, the analysis identifies distinct groups of elements exhibiting consistent behavior across stability and key physical properties, such as magnetization and Curie temperature. These findings emphasize that while integrating domain knowledge enhances exploration within the hybrid framework, alignment with underlying physical mechanisms is crucial for robust predictions.

B. Toward Insights into HEA Formation Mechanism

To reveal the underlying formation mechanisms of HEAs, we constructed a map of all alloys in the $\mathcal{D}_{0.9T_m}$ dataset using t-SNE visualization (Supplementary Figure 6a). We then applied Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)[4] with a minimum cluster size of 100 and minimum samples of 5 to identify distinct clusters of HEAs within this map. Supplementary Figure 6b presents a detailed elemental distribution analysis of the four largest clusters detected.

Cluster 1, the largest group, comprises 714 HEAs formed exclusively from a specific subset of 13 transition metals from set $\mathcal{E}$, which contains 14 transition metals that exhibit similar substitutability behavior, as described in Section V A. This cluster accounts for 99% of all possible quaternary alloys derived from these elements, underscoring their strong co-synthesis effect in promoting HEA formation. Notably, HEAs in Cluster 2 feature Fe as the key element, alloyed with Pd, Si, and other elements from $\mathcal{E}$. Cluster 3, the second-largest group with 521 HEAs, is dominated by Ta and Nb combined with other elements from $\mathcal{E}$ except Fe. Meanwhile, Cluster 4 comprises 327 Pd-based alloys, where Pd is alloyed with the transition metals in set $\mathcal{E}$.

To quantify the significance of set $\mathcal{E}$ in HEA formation beyond these cluster patterns, we analyzed how HEA stability correlates with the inclusion of elements from this set. Supplementary Figures 6c and 6d illustrate the progressive destabilization of HEAs as the number of constituent elements from $\mathcal{E}$ decreases. Approximately 98% of all HEAs in the dataset contain at least one element from set $\mathcal{E}$, highlighting the foundational role of these 14 transition metals in stabilizing HEAs. When alloys are composed exclusively of elements

from set $\mathcal{E}$, 99% form stable HEAs; however, this stability rate declines systematically as fewer elements from $\mathcal{E}$ are incorporated (Supplementary Figure 6c). Supplementary Figure 7 provides further granularity, demonstrating that stability rates for alloys combining elements from $\mathcal{E}$ with one of other elements vary substantially from 0% to 84%.

The stabilizing influence of set $\mathcal{E}$ persists even at elevated temperatures, as evidenced by the $\mathcal{D}_{1350\text{K}}$ dataset (Supplementary Figure 6d). In this high-temperature regime, approximately 97% of all HEAs still contain at least one element from set $\mathcal{E}$, though the stability rate of alloys formed exclusively from elements within set $\mathcal{E}$ decreases to 61% compared to 99% at lower temperatures.

These findings establish the identified set of 14 transition metals as a critical foundation for phase stability in HEAs across temperature ranges. Our visualization approach bridges compositional effects with phase stability, providing both fundamental insights and practical guidance for alloy design. For instance, when designing novel quaternary HEAs, prioritizing alloys of elements from Cluster 1 (particularly combinations including Co, Cr, and Ni) would maximize the probability of successful HEA formation. This data-driven, knowledge-guided approach enhances interpretability and offers actionable insights for targeted exploration of the vast compositional space of multi-component alloys.

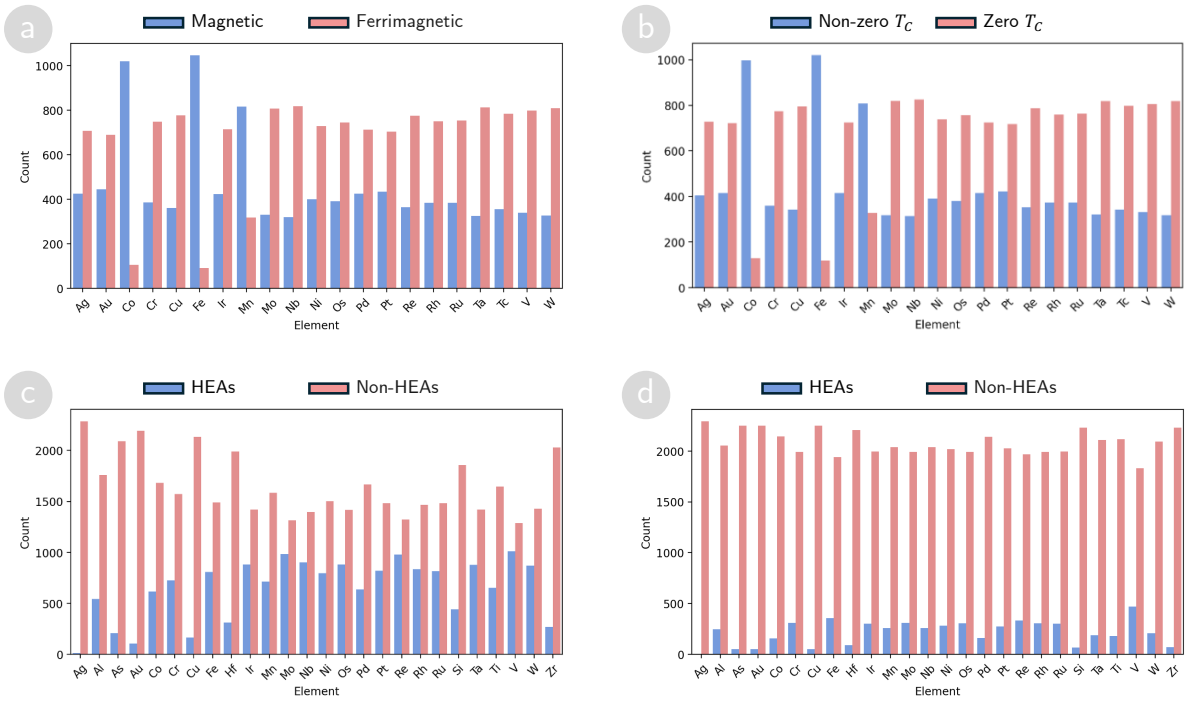
VI. COMPUTATIONAL DATASETS AS ADDITIONAL SOURCES OF EVIDENCE

To provide additional sources of evidence, we use three additional computational datasets consisting of binary, ternary, and quaternary alloys comprising multiple equiatomically combined elements. Supplementary Table I show summary of the computational datasets.

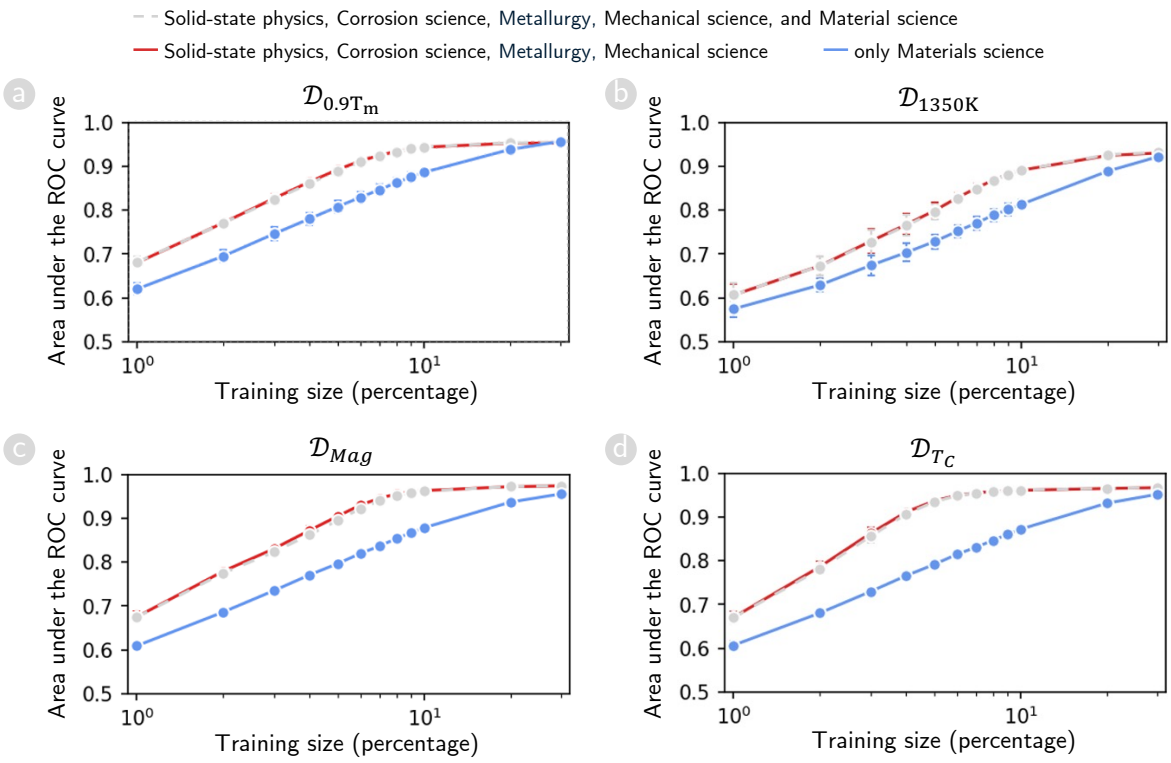
- $\mathcal{D}_{\text{CALPHAD}}$: The order-disorder transition temperatures ($T_C^\star$) and melting temperatures ($T_m^\star$) of the alloys are both predicted using calculated-phase-diagram (CALPHAD) calculations[5–7] based on the temperatures for some binary alloys (three possible for each ternary alloy) found in the Thermo-Calc software SSOL5 database[8]. The alloy in the dataset is considered as an HEA if its melting temperature $T_C^\star$ lower than its order-disorder transition temperature $T_m^\star$. Note that the $\mathcal{D}_{\text{CALPHAD}}$ data set only contains the alloys satisfying $T_C^\star < T_m^\star$.

- $\mathcal{D}_{\mathrm{AFLOW}}$: The order-disorder transition temperatures (T_C^{AFLOW}) of the alloys contained in these data sets are estimated using the automatic flow (AFLOW) convex-hull database[9]. The melting temperatures T_m^{exp} and $T_m^{\star}$ are applied to the binary and ternary alloys, respectively. The alloy is considered as an HEA if $T_C^{\mathrm{AFLOW}} < T_m^{\mathrm{exp}}$ for binary alloys and $T_C^{\mathrm{AFLOW}} < T_m^{\star}$ for ternary, quaternary, and quinary alloys).

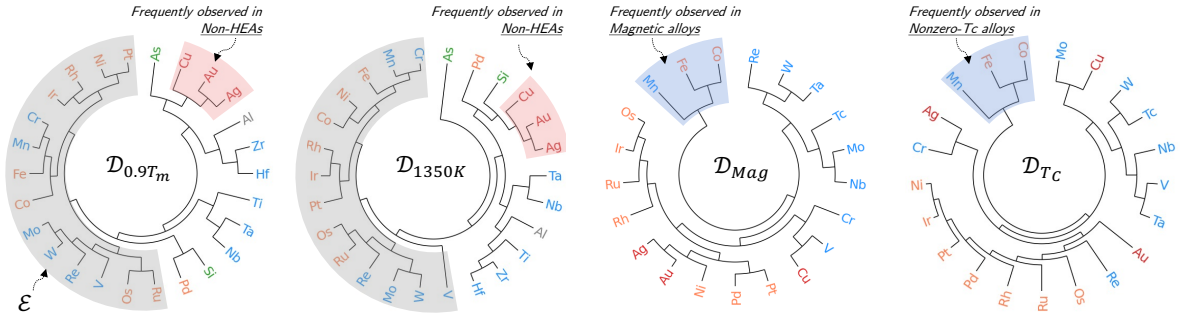
- $\mathcal{D}_{\mathrm{LTVC}}$: These data sets contain the same alloys as those contained in data sets $\mathcal{D}_{\mathrm{AFLOW}}$, $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quaternary}}$, and $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quinary}}$, respectively. However, the properties of the alloys contained in these data sets are predicted using the method of Lederer, Toher, Vecchio, and Curtarolo (LTVC)[10]. *Ab-initio* calculations are used to estimate the order-disorder transition temperatures (T_C^{LTVC}) of the alloys contained in these data sets. In addition, the T_m values are the same as those of the alloys contained in the AFLOW data sets. Any alloy in these data sets is predicted as an HEA if $T_C^{\mathrm{LTVC}} < T_m^{\mathrm{exp}}$ for binary alloys and $T_C^{\mathrm{LTVC}} < T_m^{\star}$ for ternary, quaternary, and quinary alloys.



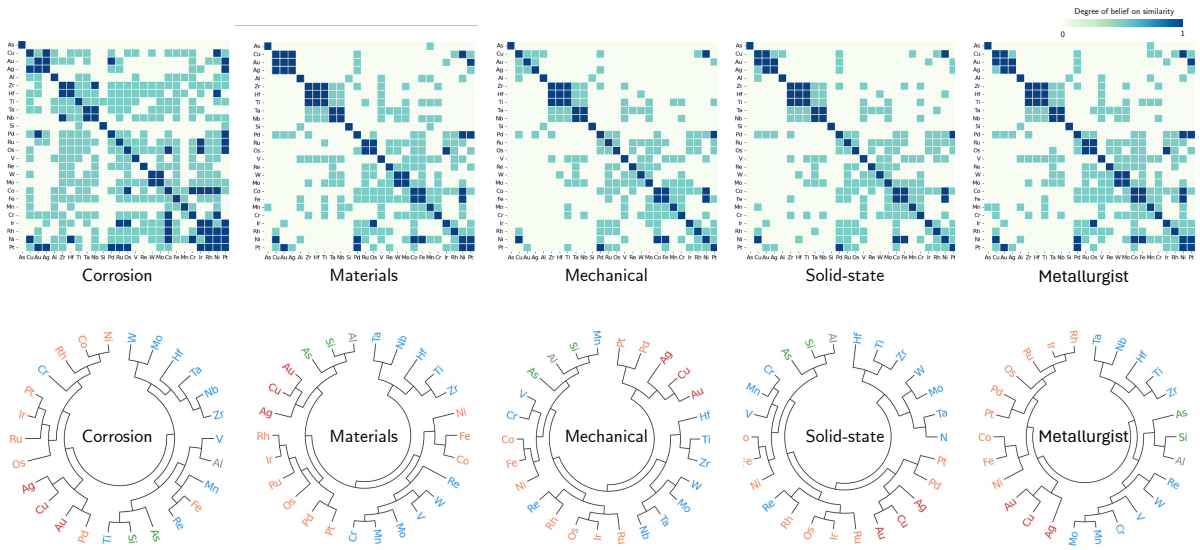
Supplementary Figure 1. Proportions of constituent elements in the four alloys datasets $\mathcal{D}_{\text{Mag}}$ (a), $\mathcal{D}_{T_C}$(b), $\mathcal{D}_{0.9T_m}$ (c), and $\mathcal{D}_{1350\text{K}}$ (d).



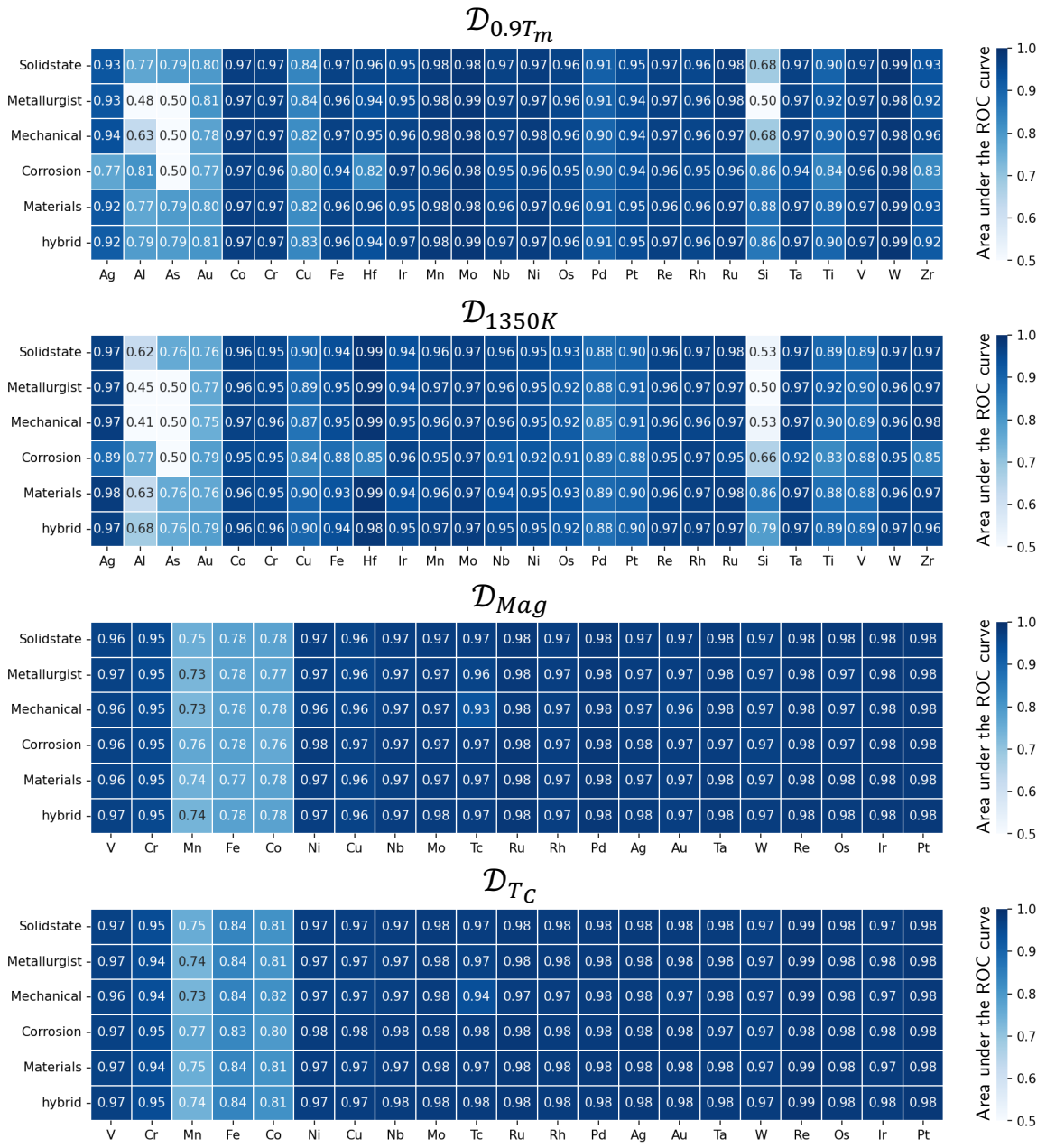
Supplementary Figure 2. **Performance comparison of explicit versus implicit domain integration.** Area under ROC curves for predicting HEA stability ($\mathcal{D}_{0.9T_m}$, $\mathcal{D}_{1350K}$) and magnetic properties ($\mathcal{D}_{Mag}$, $\mathcal{D}_{T_C}$) using three domain integration strategies: (i) systematic combination of four specialized domains (solid-state physics, corrosion science, metallurgy, materials mechanics) shown in red, (ii) only using materials science, which serves as an integrative field that synthesizes perspectives from four specialized domains, shown in blue, and (iii) all five domains including materials science shown in gray.



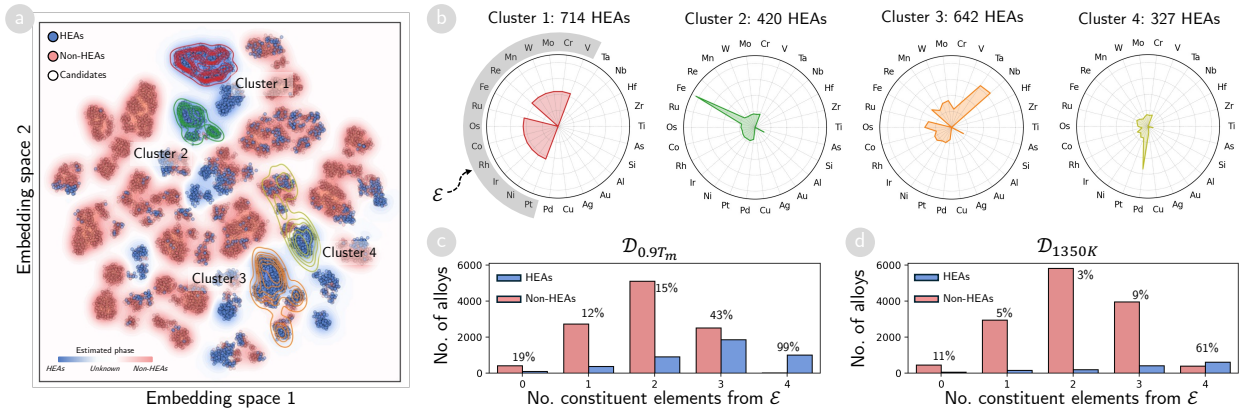
Supplementary Figure 3. **Substitutability tree of elements.** Circular tree visualizes the hierarchical clustering of all constituent elements, formed using HAC with the "complete" linkage criterion. Substitutability information is derived from four quaternary alloy datasets. Blue labels represent early transition metals, orange labels indicate represent transition metals, and red represent coinage metals, including Cu, Ag, and Au.



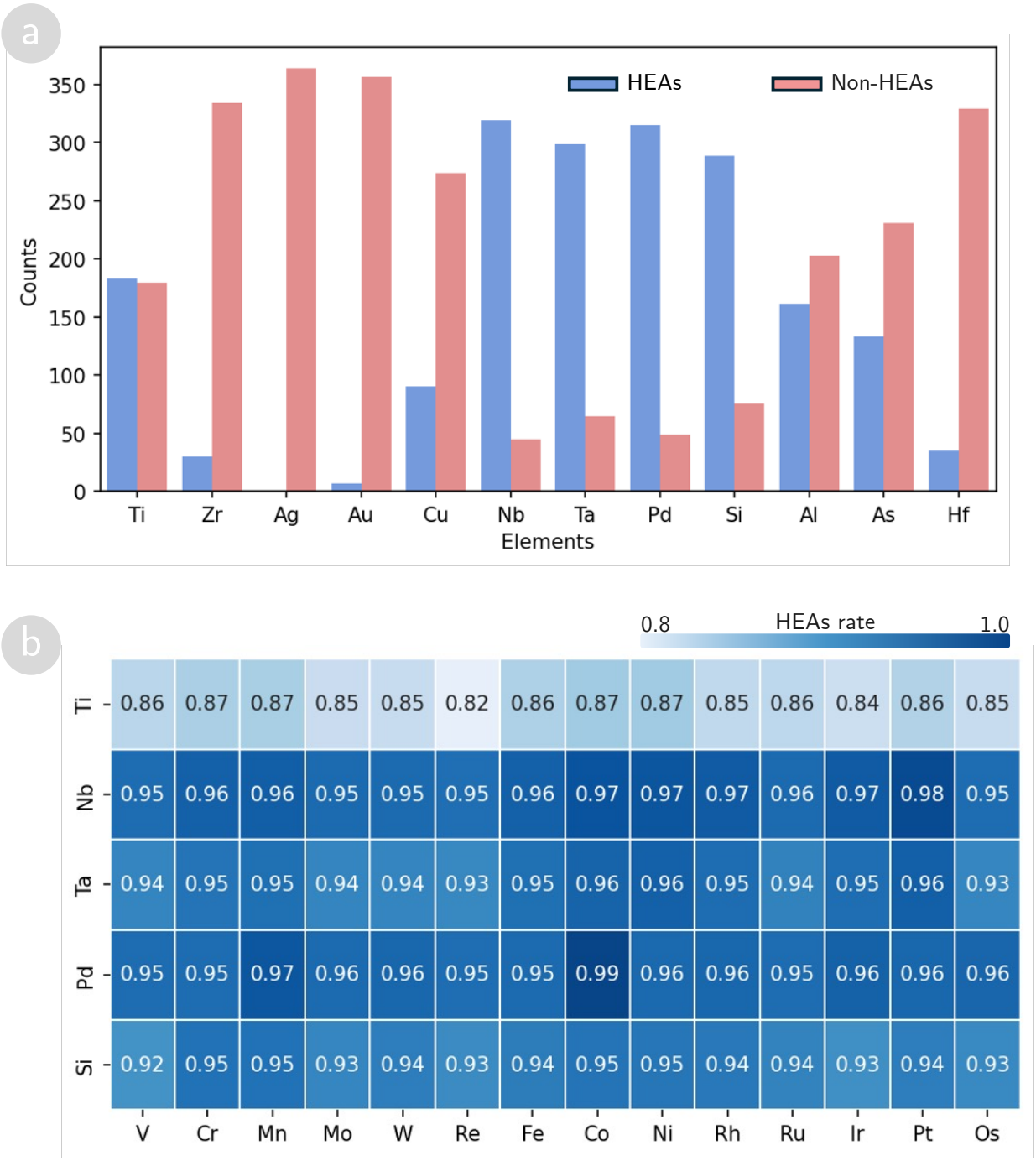
Supplementary Figure 4. (Top) Heatmap illustrating similarity matrices derived from domain knowledge distilled by GPT-4o. The heatmap visualizes the pairwise similarity between elements, based on substitutability evidence extracted from domain-specific knowledge using GPT-4o. The color intensity indicates the degree of similarity, with higher values representing stronger substitutability relationships. (Bottom) Circular hierarchical clustering (HAC) of elements based on substitutability between elements.The circular dendrogram displays the hierarchical clustering of all constituent elements, constructed using hierarchical agglomerative clustering (HAC) with the "complete" linkage criterion. The substitutability information is derived from LLM-based knowledge. Blue labels represent early transition metals, orange labels indicate late transition metals, and red labels denote coinage metals, including copper (Cu), silver (Ag), and gold (Au).



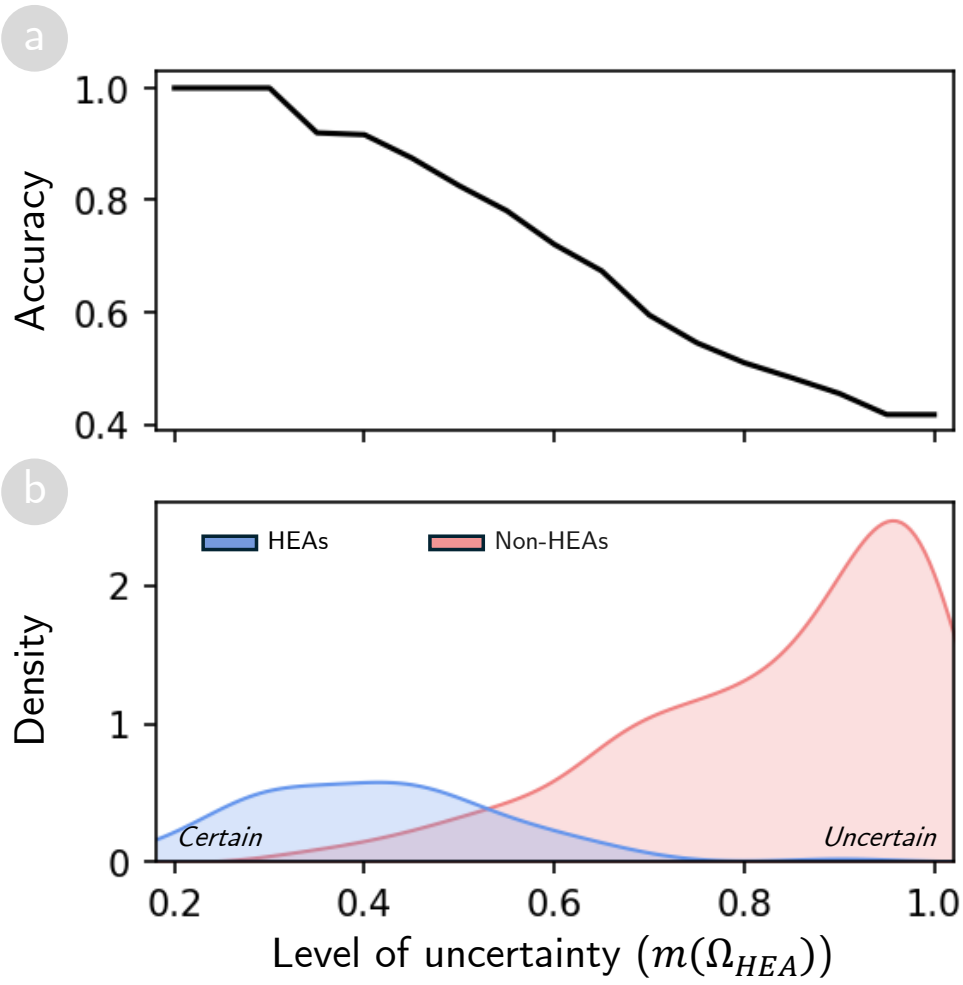
Supplementary Figure 5. The area under the ROC curve (AUC) for extrapolation experiments on constituent elements across the four alloy datasets: $\mathcal{D}_{\text{Mag}}$, $\mathcal{D}_{T_C}$, $\mathcal{D}_{0.9T_m}$, and $\mathcal{D}_{1350\text{K}}$.



Supplementary Figure 6. **Towards understanding high-entropy alloys (HEA) formation mechanism**. (a) Alloy map generated from element substitutability patterns extracted using the DS-source model on dataset $\mathcal{D}_{0.9T_m}$. Marker colors indicate phase formation: blue for HEA-forming alloys, red for non-HEA-forming alloys, and white for Os-based candidate alloys. Background coloration represents the predicted HEA formation probability according to the DS-source model. Kernel density estimations highlight the four largest HEA clusters identified by HDBSCAN. (b) Radar charts displaying elemental distribution patterns across the four quaternary HEA clusters. (c–d) Bar plots showing the relationship between HEA stability and the number of constituent elements from $\mathcal{E}$ in datasets $\mathcal{D}_{0.9T_m}$ (c) and $\mathcal{D}_{1350\text{K}}$ (d), with percentages indicating HEA formation rates for each configuration.



Supplementary Figure 7. Analysis of synthesized effects between transition metals in $\mathcal{E}$ and other elements. (a) Summary of the number of single-phase and multi-phase alloys formed by combining transition metals in $\mathcal{E}$ with other elements. (b) Heatmap showing the stability rate (proportion of single-phase alloys) for alloys formed exclusively from modified sets of $\mathcal{E}$, where one element is replaced by silicon (Si), palladium (Pd), tantalum (Ta), niobium (Nb), or titanium (Ti).



Supplementary Figure 8. **Empirical validation of uncertainty quantification mechanism preventing overfitting in data-sparse boride predictions.** (a) Prediction accuracy as a function of uncertainty threshold $m(\Omega_{HEA})$. As the threshold increases (accepting candidates with higher uncertainty for evaluation), prediction accuracy systematically degrades, confirming that high uncertainty values successfully flag regions where evidence is insufficient. Accuracy measured against DEED ab-initio descriptor as independent validation benchmark. (b) Distribution of DEED-predicted stable (blue) versus unstable (red) candidates across uncertainty levels. Stable candidates predominantly exhibit low uncertainty (high confidence), while unstable candidates show high uncertainty with most assigned near-total ignorance $m(\Omega_{HEA}) \approx 1$. Despite training data bias toward stable compositions (15 of 19 training samples are stable), the framework correctly identifies unstable candidates as highly uncertain rather than confidently mispredicting them as stable, demonstrating effective overfitting prevention through explicit uncertainty quantification.

Supplementary Table I. Summary of computational data sets used as additional sources of evidence about elemental substituability. No. alloys: number of alloys included in each data set; No. positive label: number of the alloys confirmed or estimated to form single phase solution in each data set; The percentage values in parentheses represent the proportion of positive labels within each dataset.

Datasets	**No. alloys**	**Positive label**	**No. positive label**
$\mathcal{D}_{\mathrm{CALPHAD}}$[11,12]	243 ternary alloys	HEA	243 (100%)
$\mathcal{D}_{\mathrm{AFLOW}}$[7]	117 binary alloys	HEA	60 (51%)
	441 ternary alloys	HEA	234 (53%)
	1,110 quaternary alloys	HEA	754 (68%)
$\mathcal{D}_{\mathrm{LTVC}}$[10]	117 binary alloys	HEA	58 (49%)
	441 ternary alloys	HEA	148 (33%)
	1,110 quaternary alloys	HEA	480 (43%)

Supplementary Table II. Criteria for the formation of single-phase HEAs defined by various empirical rules (ERs), free-energy models (FEMs), and prediction model based on valence electron concentration (VEC).

Models	**Criteria**	**References**
ER1	$-11.6 < \Delta H_{\mathrm{mix}} < 3.2$ kJ/mol, $\delta < 0.066$	[13]
ER2	$-11.6 < \Delta H_{\mathrm{mix}} < 3.2$ kJ/mol, $\gamma < 1.175$[a]	[14]
ER3	$\Delta S_{\mathrm{mix}}/\delta^2 > 9.6$ kJ/mol/K	[15]
FEM1	$-T_{\mathrm{ann}}\Delta S_{\mathrm{mix}} < \min(\Delta H^{\mathrm{IM}}_{\mathrm{f,ij}}) < 3.57$ meV/atom	[16]
FEM2	$1 + 0.4 \times T_{\mathrm{ann}}\Delta S_{\mathrm{mix}}/\lvert\Delta H_{\mathrm{mix}}\rvert < \Delta H^{\mathrm{IM}}_{\mathrm{f}}/\Delta H_{\mathrm{mix}}$	[5]
VEC	Single-phase if VEC ≤ 6.87 or VEC ≥ 8, mixed phase otherwise	[17]

[a] $\gamma = \frac{\omega_s}{\omega_l}$, where $\omega_{s,l} = 1 - \sqrt{\frac{(r_{s,l}+\bar{r})^2 - \bar{r}^2}{(r_{s,l}+\bar{r})^2}}$. $r_{s,l}$ refers to the radius of the smallest and the largest atoms, respectively.

Supplementary Table III: **Detailed breakdown of 45 element pairs where LLM predictions differ from Hume–Rothery rules.** For all listed pairs, LLMs predict mutual substitutability, whereas Hume–Rothery rules classify them as non-substitutable. Columns report: (1) element pair, (2) size mismatch (Δr), (3) electronegativity difference ($\Delta\chi$), (4) valency compatibility (V), (5) crystal-structure compatibility (S), (6) binary phase outcome from the phase-diagram database[18], and (7) representative quaternary or quinary alloy systems[19–21] that are experimentally confirmed as single-phase solid solutions, in which the element pair provides direct experimental evidence for substitutability of the corresponding elements.

Pair	Δr	$\Delta\chi$	V	S	**Binary system**	**Representative alloy systems**
(Au-Pd)	5.11	0.20	dissimilar	dissimilar	Single phase	Au-Cu-Ni-Pt, Cu-Ni-Pd-Pt
(Au-Pt)	3.60	0.20	dissimilar	dissimilar	Single phase	Au-Cu-Ni-Pd, Cu-Ni-Pd-Pt
(Co-Cr)	2.40	0.22	similar	similar	Single phase	Al-Co-Fe-Ni, Al-Cr-Fe-Ni
(Cr-Mn)	0.79	0.11	similar	similar	Single phase	Co-Cr-Fe-Ni, Co-Fe-Mn-Ni
(Cr-Ni)	3.23	0.25	similar	similar	Single phase	Co-Cr-Cu-Fe, Co-Cu-Fe-Ni
(Cu-Rh)	4.69	0.38	dissimilar	dissimilar	Single phase	Cu-Ni-Pd-Pt, Ni-Pd-Pt-Rh
(Hf-Nb)	8.90	0.30	dissimilar	similar	Single phase	Hf-Ta-Ti-Zr, Nb-Ta-Ti-Zr
(Hf-Ta)	8.90	0.20	dissimilar	similar	Single phase	Hf-Nb-Ti-Zr, Nb-Ta-Ti-Zr
(Nb-Ti)	0.68	0.06	dissimilar	similar	Single phase	Hf-Nb-Ta-Zr, Hf-Ta-Ti-Zr
(Ti-V)	9.70	0.09	similar	similar	Single phase	Mo-Nb-Ta-Ti, Mo-Nb-Ta-V
(Co-Ni)	0.81	0.03	similar	similar	Single phase	–
(Co-Pd)	9.60	0.32	similar	similar	Single phase	–
(Cr-Ti)	14.84	0.12	similar	similar	Single phase	–
(Fe-Ni)	1.61	0.08	similar	similar	Single phase	–
(Cr-Nb)	14.06	0.06	dissimilar	dissimilar	Multi-phase	Al-Cr-Ti-V, Al-Nb-Ti-V
(Fe-Mn)	0.79	0.28	similar	similar	Multi-phase	Co-Cr-Fe-Ni, Co-Cr-Mn-Ni
(Mo-Re)	1.46	0.26	similar	similar	Multi-phase	Mo-Nb-Ta-Ti-V, Nb-Re-Ta-Ti-V
(Mo-Ta)	5.04	0.66	dissimilar	dissimilar	Multi-phase	Al-Mo-Nb-Ti, Al-Nb-Ta-Ti
(Pd-Ru)	2.24	0.00	similar	similar	Multi-phase	Co-Cu-Fe-Ni-Pd, Co-Cu-Fe-Ni-Ru
(Pt-Ru)	3.73	0.00	similar	similar	Multi-phase	Co-Cu-Fe-Ni-Pt, Co-Cu-Fe-Ni-Ru

Continued on next page

Supplementary Table III – continued from previous page

Pair	Δr	$\Delta\chi$	V	S	Binary system	Representative alloy systems
(Re-W)	1.46	0.20	similar	similar	Multi-phase	Nb-Re-Ta-Ti-V, Nb-Ta-Ti-V-W
(Ta-Ti)	0.68	0.04	dissimilar	similar	Multi-phase	Hf-Nb-Ta-Zr, Hf-Nb-Ti-Zr
(Ag-Pd)	5.11	0.27	dissimilar	dissimilar	Multi-phase	–
(Ag-Pt)	3.60	0.27	dissimilar	dissimilar	Multi-phase	–
(As-Tc)	8.82	0.08	dissimilar	similar	Multi-phase	–
(As-Zr)	8.11	0.85	dissimilar	similar	Multi-phase	–
(Co-Fe)	0.80	0.05	similar	similar	Multi-phase	–
(Co-Os)	8.00	0.32	dissimilar	dissimilar	Multi-phase	–
(Co-Re)	9.60	0.02	dissimilar	dissimilar	Multi-phase	–
(Co-Rh)	7.20	0.40	similar	similar	Multi-phase	–
(Co-Tc)	8.80	0.22	dissimilar	dissimilar	Multi-phase	–
(Fe-Tc)	7.94	0.27	dissimilar	similar	Multi-phase	–
(Ir-Os)	0.74	0.00	similar	similar	Multi-phase	–
(Ir-Re)	0.74	0.30	similar	similar	Multi-phase	–
(Ir-Ru)	1.49	0.00	similar	similar	Multi-phase	–
(Mo-Nb)	5.04	0.56	dissimilar	dissimilar	Multi-phase	–
(Mo-Tc)	2.21	0.06	similar	similar	Multi-phase	–
(Nb-W)	5.04	0.10	dissimilar	dissimilar	Multi-phase	–
(Ni-Rh)	8.06	0.37	dissimilar	dissimilar	Multi-phase	–
(Ni-Ru)	8.06	0.29	dissimilar	similar	Multi-phase	–
(Pd-Si)	0.73	0.30	similar	similar	Multi-phase	–
(Pt-Rh)	3.73	0.08	dissimilar	dissimilar	Multi-phase	–
(Re-Ta)	6.57	0.40	dissimilar	similar	Multi-phase	–
(Rh-Ru)	0.00	0.08	similar	similar	Multi-phase	–
(Ta-W)	5.04	0.20	dissimilar	dissimilar	Multi-phase	–

1 Hume–Rothery criteria used in this work: Δr is the atomic size mismatch (in %) between the two elements;
2 $\Delta\chi$ is the difference in Pauling electronegativity; V denotes valency compatibility (*similar* if the nominal
3 valences are equal, otherwise *dissimilar*); S denotes crystal-structure compatibility (*similar* if the elements

share the same prototype structure at ambient conditions, otherwise *dissimilar*). Hume–Rothery rules classify a pair as non-substitutable when at least one of the following is violated: $\Delta r > 15\%$, $\Delta\chi > 0.55$, V is dissimilar, or S is dissimilar. "Binary system" indicates whether the corresponding binary phase diagram is reported as a single-phase or multi-phase field, and "Representative alloy systems" lists quaternary/quinary solid solutions that are experimentally confirmed to be single-phase and contain the given element pair.

Supplementary Table IV. Classification report of our proposed framework for extrapolation experiment of Os-based alloys on the dataset $\mathcal{D}_{0.9T_m}$.

	precision	**recall**	**f1-score**	**support**
Single-phase	0.90	0.78	0.84	882 alloys
Multi-phase	0.88	0.95	0.91	1,418 alloys
Accuracy			0.88	2,300 alloys
Macro average	0.89	0.87	0.87	2,300 alloys
Micro average	0.89	0.88	0.88	2,300 alloys

Supplementary Table V. Parameter grid for logistic regression hyperparameter tuning.

Hyperparameter	**Possible Values**
Penalty	L1, L2
Solver	liblinear (L1 and L2), saga (L1 and L2), newton-cg (L2), lbfgs (L2), sag (L2)
Dual Setting	True (liblinear), False (all solver)
Regularization Parameter C	0.01, 0.1, 1, 10, 100

ADDITIONAL SUPPLEMENTARY FILES

1. File Name: SupplementaryFile1_Prompts.txt

   Description: Detailed prompts employed for state-of-the-art LLM, including GPT-4o, GPT-4.5, Claude Opus4, and Grok3 to assess element substitutability based on expert knowledge within five critical scientific domains (*corrosion science*, *materials mechanics*, *metallurgy*, *solid-state physics*, and *materials science*) that play essential roles in understanding and optimizing high-entropy alloys (HEAs).

2. File Name: SupplementaryFile2_ChatGPT4o.txt

   Description: Comprehensive GPT-4o responses evaluating elemental substitutability patterns through systematic extraction of expert knowledge from five key scientific domains relevant to high-entropy alloy design: *corrosion science*, *materials mechanics*, *metallurgy*, *solid-state physics*, and *materials science*.

3. File Name: SupplementaryFile3_ChatGPT45.txt

   Description: Comprehensive GPT-4.5 responses evaluating elemental substitutability patterns through systematic extraction of expert knowledge from five key scientific domains relevant to high-entropy alloy design: *corrosion science*, *materials mechanics*, *metallurgy*, *solid-state physics*, and *materials science*.

4. File Name: SupplementaryFile4_ClaudeOpus4.txt

   Description: Comprehensive Claude Opus 4 responses evaluating elemental substitutability patterns through systematic extraction of expert knowledge from five key scientific domains relevant to high-entropy alloy design: *corrosion science*, *materials mechanics*, *metallurgy*, *solid-state physics*, and *materials science*.

5. File Name: SupplementaryFile5_Grok3.txt

   Description: Comprehensive Grok3 responses evaluating elemental substitutability patterns through systematic extraction of expert knowledge from five key scientific domains relevant to high-entropy alloy design: *corrosion science*, *materials mechanics*, *metallurgy*, *solid-state physics*, and *materials science*.

REFERENCES

[1]A. P. Dempster, "A generalization of bayesian inference," Journal of the Royal Statistical Society: Series B (Methodological) **30**, 205–232 (1968), https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1968.tb00722.x.

[2]L. van der Maaten and G. Hinton, "Viualizing data using t-sne," Journal of Machine Learning Research **9**, 2579–2605 (2008).

[3]M. LEVANDOWSKY and D. WINTER, "Distance between sets," Nature **234**, 34–35 (1971).

[4]R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, edited by J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013) pp. 160–172.

[5]O. Senkov and D. Miracle, "A new thermodynamic parameter to predict formation of solid solution or intermetallic phases in high entropy alloys," Journal of Alloys and Compounds **658** (2015), 10.1016/j.jallcom.2015.10.279.

[6]D. Alman, "Searching for next single-phase high-entropy alloy compositions," Entropy **15**, 4504–4519 (2013).

[7]F. Zhang, C. Zhang, S. Chen, J. Zhu, W. Cao, and U. Kattner, "An understanding of high entropy alloys from phase diagram calculations," Calphad **45**, 1 – 10 (2014).

[8]J.-O. Andersson, T. Helander, L. Höglund, P. Shi, and B. Sundman, "Thermo-calc & dictra, computational tools for materials science," Calphad **26**, 273 – 312 (2002).

[9]C. Nyshadham, C. Oses, J. E. Hansen, I. Takeuchi, S. Curtarolo, and G. L. Hart, "A computational high-throughput search for new ternary superalloys," Acta Materialia **122**, 438 – 447 (2017).

[10]Y. Lederer, C. Toher, K. S. Vecchio, and S. Curtarolo, "The search for high entropy alloys: A high-throughput ab-initio approach," Acta Materialia **159**, 364 – 383 (2018).

[11]O. N. Senkov, J. D. Miller, D. B. Miracle, and C. Woodward, "Accelerated exploration of multi-principal element alloys with solid solution phases," Nature Communications **6**, 6529 (2015).

[12]D. Alman, "Searching for next single-phase high-entropy alloy compositions," Entropy **15**, 4504–4519 (2013).

[13]S. Guo, Q. Hu, C. Ng, and C. Liu, "More than entropy in high-entropy alloys: Forming solid solutions or amorphous phase," Intermetallics **41**, 96–103 (2013).

[14]W. Zhijun, Y. Huang, Y. Yang, J. Wang, and C. Liu, "Atomic-size effect and solid solubility of multi-

component alloys," Scripta Materialia **94** (2015), 10.1016/j.scriptamat.2014.09.010.

[15]A. K. Singh, N. Kumar, A. Dwivedi, and A. Subramaniam, "A geometrical parameter for the formation of disordered solid solutions in multi-component alloys," Intermetallics **53**, 112–119 (2014).

[16]M. C. Troparevsky, J. R. Morris, P. R. C. Kent, A. R. Lupini, and G. M. Stocks, "Criteria for predicting the formation of single-phase high-entropy alloys," Phys. Rev. X **5**, 011041 (2015).

[17]S. Guo, C. Ng, J. Lu, and C. T. Liu, "Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys," Journal of Applied Physics **109**, 103505 (2011), https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/1.3587228/15077138/103505_1_online.pdf.

[18]H. O. M. S. E. Mueller, "Binary alloy phase diagrams," in *Alloy Phase Diagrams* (ASM International, 2016).

[19]W. Chen, A. Hilhorst, G. Bokas, S. Gorsse, P. J. Jacques, and G. Hautier, "A map of single-phase high-entropy alloys," Nature Communications **14**, 2856 (2023).

[20]C. K. H. Borg, C. Frey, J. Moh, T. M. Pollock, S. Gorsse, D. B. Miracle, O. N. Senkov, B. Meredig, and J. E. Saal, "Expanded dataset of mechanical properties and observed phases of multi-principal element alloys," Scientific Data **7**, 430 (2020).

[21]G. Khanna R, M. K. Singh, D. K. Rai, and S. Samal, "Light weight single-phase al-cr-ti-v multiprincipal element alloy as fast and efficient electrocatalyst," Materials Letters **365**, 136404 (2024).