Supplementary Information (SI) for Journal of Materials Chemistry A.

Supporting Information:

Synergistic Prediction of Phase and Hardness in High-Entropy Alloys via the Integration of Machine Learning and Active Learning Strategies

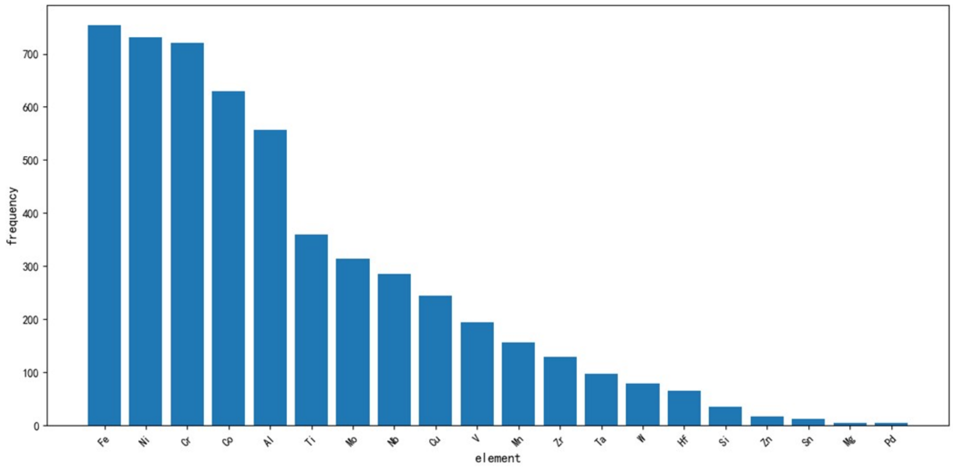

Fig.S1 Frequency of element occurrence in the dataset

Fig. S1 shows the frequency of occurrence of different elements in the dataset. Elements such as Fe, Ni, Cr, and Co appear more frequently, as they are commonly present in typical high-entropy alloy systems. In contrast, elements such as Sn, Mg, and Pd show relatively low occurrence frequencies, reflecting the uneven distribution of alloy compositions in the existing experimental data. Therefore, materials descriptors were constructed to improve the generalization capability of the model.

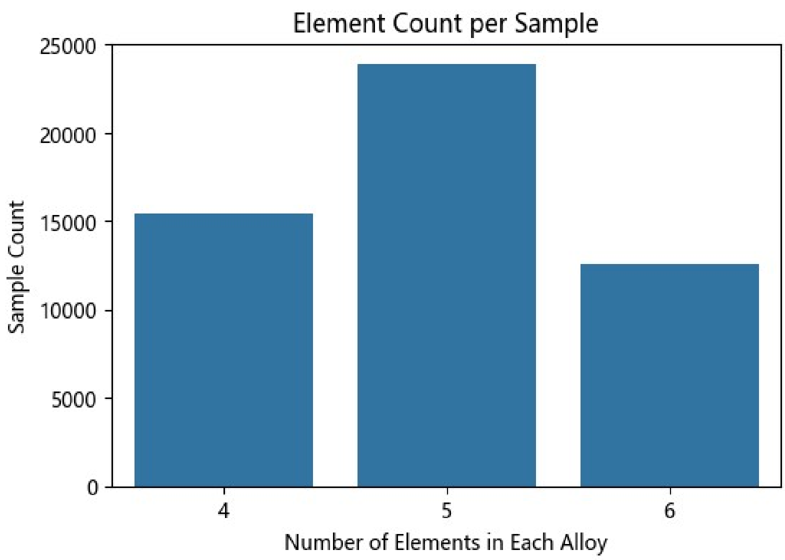

Fig.S2 Statistics on the number of elements in generated samples

Fig. S2 presents the statistics on the number of constituent elements in the generated samples. Each generated sample contains 4–6 elements, providing an overview of the composition of candidate alloys used for active learning. The results show that the generated samples cover alloy systems with different numbers of principal elements, thereby supporting the exploration of a diverse high-entropy alloy compositional space.

Table.S1 Final selected features used for model training

Feature name	
r_{van}	ΔR_{Ω}
δTE	$\delta \Delta H_m$
R_{Ω}	FIE
δG	Xe
Λ	AM
AW	AM-HT
VEC	LC
SHC	FG

Table S1 lists the final feature subset used for model training. Through variance threshold filtering, Pearson correlation analysis, sequential forward feature selection, and genetic algorithm-based feature selection, redundant and weakly relevant features in the initial feature set were progressively removed. Finally, 16 key features that make important contributions to phase and hardness classification were retained, including 12 materials descriptors and 4 one-hot encoded processing features. This feature subset reduces the input dimensionality and model complexity while preserving effective information closely related to phase formation and property variations in high-entropy alloys, thereby providing the basis for subsequent machine learning model training and prediction.