ARTICLE   OPEN

Accelerating materials-space exploration for thermal insulators by mapping materials properties via artificial intelligence

Thomas A. R. Purcell [1 ✉], Matthias Scheffler [1,2], Luca M. Ghiringhelli [1,2 ✉] and Christian Carbogno [1 ✉]

Reliable artificial-intelligence models have the potential to accelerate the discovery of materials with optimal properties for various applications, including superconductivity, catalysis, and thermoelectricity. Advancements in this field are often hindered by the scarcity and quality of available data and the significant effort required to acquire new data. For such applications, reliable surrogate models that help guide materials space exploration using easily accessible materials properties are urgently needed. Here, we present a general, data-driven framework that provides quantitative predictions as well as qualitative rules for steering data creation for all datasets via a combination of symbolic regression and sensitivity analysis. We demonstrate the power of the framework by generating an accurate analytic model for the lattice thermal conductivity using only 75 experimentally measured values. By extracting the most influential material properties from this model, we are then able to hierarchically screen 732 materials and find 80 ultra-insulating materials.

INTRODUCTION

Artificial-intelligence (AI) techniques have the potential to significantly accelerate the search for novel, functional materials, especially for applications where different physical mechanisms compete with each other non-linearly, e.g., quantum materials[1], and where the cost of characterizing the materials makes a large-scale search intractable, e.g., thermoelectrics[2]. Due to this inherent complexity, only limited amounts of data are currently available for such applications, which in turn severely limits the applicability and reliability of AI techniques[3]. Using thermal transport as an example, we propose a route to overcome this hurdle by presenting an AI framework that is applicable to scarce datasets and that provides heuristics able to steer further data creation into regions of interest in materials space.

Heat transport, as measured by the temperature-dependent thermal conductivity, κ_L, is a ubiquitous property of materials and plays a vital role for numerous scientific and industrial applications including energy conversion[4], catalysis[5], thermal management[6], and combustion[7]. Finding new crystalline materials with either an exceptionally low or high thermal conductivity is a prerequisite for improving these and other technologies or making them commercially viable at all. Accordingly, finding new thermal insulators and understanding where in materials space to search for such compounds is an important open challenge in this field. From a theory perspective, thermal transport depends on a complex interplay of different mechanisms, especially in thermal insulators, for which strongly anharmonic, higher-order effects can be at play[8]. Despite significant progress in the computational assessment of κ_L in solids[9,10], these ab initio approaches are too costly for a large-scale exploration of material space. For this reason, computational *high-throughput* approaches have so far covered only a small subset of materials[11–13]. Experimentally, an even smaller number of materials have had their thermal conductivities measured, and <150 thermal insulators identified[14,15].

Recently, increased research efforts have been devoted to leveraging AI frameworks to extend our knowledge in this field. In particular, various regression techniques have been proven to successfully interpolate between the existing data and approximate κ_L using only simpler properties[11,14,16,17]; however, using these techniques to extrapolate into new areas of materials space is a known challenge. More importantly, the explainbility of these models is limited by their inherent complexity. Physically motivated, semi-empirical models, e.g., the Slack model[18], perform slightly better in this regard because they encapsulate information about the actuating mechanism. Recent efforts have used AI to extend the capabilities of these models[2,16,19,20] to increase their accuracy in estimating κ_L. However, the applicability of such models is still limited by the physical assumptions entering the original expressions[2,19]. A general model that removes these assumptions and achieves the quantitative accuracy of AI approaches, while retaining the qualitative interpretability of analytical models, is however, still lacking.

In this work, we tackle this challenge by using a symbolic regression technique to quantitatively learn κ_L, using easily calculated materials properties. While symbolic regression methods are typically more expensive to train than other kernel based methods, such as Kernel-Ridge Regression (KRR) and Gaussian Process Regression (GPR), their prediction errors are typically equivalent to other methods and their natural feature reduction and resulting analytical expressions make them a useful method for explainable AI, as further illustrated below[21]. Furthermore, the added cost of training does not affect the evaluation time of the given models, meaning the extra time only has to be spent at the beginning. The inherent uncertainty estimate in methods like GPR, allows for a prediction of where the resulting models are expected to perform worse; however, we also propose a method to get an ensemble uncertainty estimate for symbolic regression that can be applied more generally to these types of models. We further exploit the feature reduction of SISSO and expand upon its

[1]The NOMAD Laboratory at Fritz-Haber-Institut der Max-Planck-Gesellschaft and IRIS-Adlershof of the Humboldt-Universität zu Berlin, Faradayweg 4-6, D-14195 Berlin, Germany. [2]Physics Department and IRIS Adlershof Humboldt Universität zu Berlin, Berlin, Germany. ✉email: purcell@fhi-berlin.mpg.de; ghiringhelli@fhi-berlin.mpg.de; carbogno@fhi-berlin.mpg.de
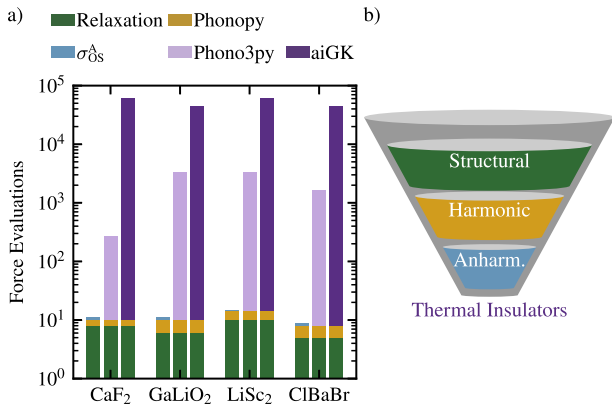
Fig. 1 The motivation for the work is reducing the number of calculations needed to approximate the thermal conductivity of a material. a The number of force evaluations needed to complete each step of a κ_L calculation for four representative materials: (1) Geometry relaxation (green first bar), (2) Harmonic model generation with Phonopy (yellow, second bar), (3) Evaluating κ_L via Phono3py (lavender third bar) or MD (purple fourth bar), and (4) estimating κ_L with the σ^A (light blue). The relaxation step typically acts on the primitive cells (~10 atoms) while all other are done on supercells with ~200 or more atoms. The number of force evaluations for Phono3py assumes all displacements are needed to calculate the third order force constants for version 2.5.1 (**b**) The proposed hierarchical workflow that can screen out materials before the final calculations.

interpretability by using a global sensitivity analysis method to distill out the key material properties that are most important for modelling κ_L and to find the conditions necessary for obtaining an ultra-low thermal conductivity. From here, we use this analysis to learn the conditions needed to screen materials in each step of a hierarchical, high-throughput workflow to discover new thermal insulators. Using this workflow we can then establish qualitative design principles that lend themselves to general application across material space and use them to find 80 materials with an ultra-low κ_L.

RESULTS

Symbolic regression models for thermal conductivity

We use the sure-independence screening and sparsifying operator (SISSO) method as implemented in the SISSO++ code[22]. This method has been used to successfully describe multiple applications including the stability of materials[23], catalysis[24], and glass transition temperatures[25]. To find the best low-dimensional models for a specific target property, in our case the room temperature, lattice thermal conductivity, $\kappa_L(300\,\text{K})$, SISSO first builds an exhaustive set of analytical, non-linear functions, i.e., trillions of candidate descriptors, from a set of mathematical operators and primary features, the set of user-provided properties that will be used to model the target property. Here we are focusing on room temperature data only because that is what is the most abundant in the literature and relevant for potential applications; however, some temperature dependence will be inherently included via the temperature dependence of our anharmonicity factor σ^A. For this application the primary features are both the structural and dynamical properties for seventy-five materials with experimentally measured $\kappa_L(300\,\text{K})$[17,26–43] (see Section IV D and Supplementary Note 1 for more details). By using the experimentally measured values for κ_L we avoid the issues related to the inconsistent reliability of different approaches to calculating κ_L for different material classes[44,45], and hopefully create a universal model for it. For many of the materials of interest here the standard Boltzmann Transport approach will be

unreliable[44,45], but the fully anharmonic ab initio Green-Kubo approach is unnecessarily expensive to use for all materials[45]. Combining theoretical and experimental data in this way allows one to avoid both the cost or unreliability of calculating, κ_L and the challenges of experimentally synthesizing and characterizing candidate materials. As long as all samples are consistent across each feature, AI and ML based models will adapt the computational features to the experimental target.

Figure 1b illustrates the main goal of the work: to learn which primary features are important for modeling κ_L and what thresholds of those indicate where thermal insulators are present. As a result the figure also represents the workflow used to calculate κ_L and generate the primary features for the model. All of the data generated in this workflow will be calculated using ab initio methods, with each step representing an increasing cost of calculation, as shown in Fig. 1a. The total cost of calculating these primary features is several orders of magnitude smaller than explicitly calculating κ_L, either with the Boltzmann Transport Equation or aiGK. While using only compositional and structural features would further reduce the cost of generating them, it comes at the expense of decreasing the reliability and explainability of the models. A goal of this work is to learn the screening conditions needed to remove materials at each step of the workflow in Fig. 1b and only perform the intensive κ_L calculations on the most promising materials. Because of this, we feel that using the features generated from this workflow is the most logical set to use. Importantly, as described in Section IV D we use a consistent and accurate formalism for calculating all features in this workflow, and therefore expect a quantitative agreement between these features and their experimental counterparts. Even if this framework were restricted to explore only high-symmetry materials, the overall cost of the calculations in a supercell would be reduced by a factor of one hundred as shown by the non-green bars in Fig. 1a. In the more general case we would be able to screen closer to 1000 more materials using this procedure over the brute-force workflows of calculating κ_L for all materials. With the learned conditions one could then create a prescreening procedure by learning models for each of the relevant structural or harmonic properties using only compositional inputs, and use those to estimate κ_L[46]; however, that is outside of the scope of this work.

In practice, we model the $\log(\kappa_L(300\,\text{K}))$ instead of $\kappa_L(300\,\text{K})$ itself to better handle the wide range of possible thermal conductivities. The parity plot in Fig. 2(a) illustrates the performance of the identified SISSO model when the entire dataset is used (see Section IV A for more details). The resulting expression is characterized by d_1 and d_2

$$\log\left(\kappa^{SISSO}(300\,\text{K})\right) = a_0 + a_1 d_1 + a_2 d_2$$
$$d_1 = \frac{\left(m_{avg} + 200.3\text{Da}\right)^2}{\sqrt{\mu}\left(V_m + 218.9\text{Å}^3\right)^3 \Theta_{D,\infty} \sigma^A}$$
$$d_2 = \sigma^A \frac{V_m \rho}{m_{avg}} + e^{\frac{-\omega_{\Gamma,max}}{27.11\text{THz}}} + e^{\sigma^A} \tag{1}$$

where $a_0 = 6.327$, $a_1 = -8.219 \times 10^4$, and $a_2 = -1.704$ are constants found by least-square regression and all variables are defined in Table 1. We find that this model has a training root-mean squared error (RMSE) of 0.14, with an R^2 of 0.98 for $\log\left(\kappa^{SISSO}(300\,\text{K})\right)$. To better understand how these error terms translate to $\kappa_L(300\,\text{K})$, we also use the average factor difference (AFD)

$$\text{AFD} = 10^x \tag{2a}$$

$$x = \frac{1}{n}\sum_i^n \left| \log(\kappa_L) - \log\left(\kappa_L^{pred}\right) \right|, \tag{2b}$$

where n is the number of training samples. Here, we find an AFD of 1.30 that is on par if not smaller than models previously found
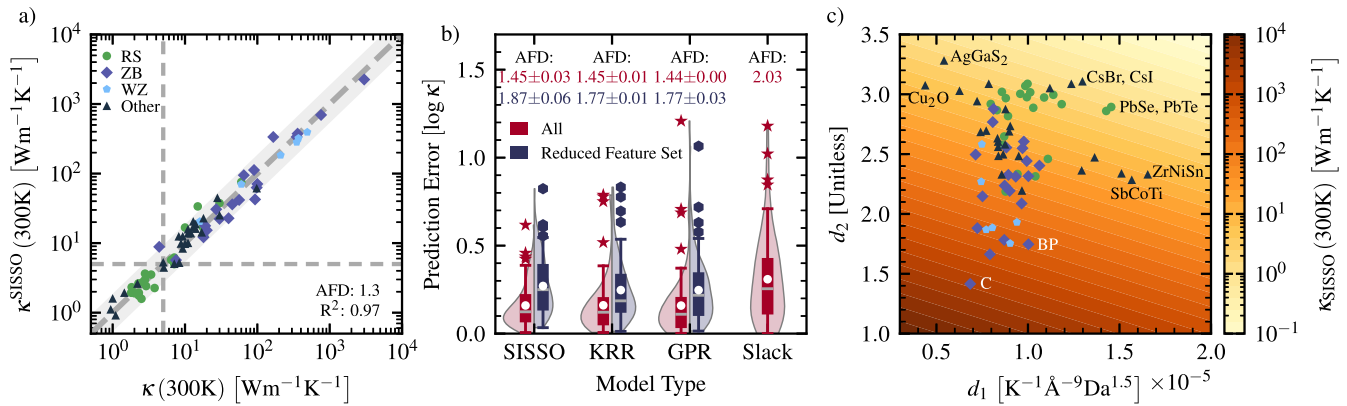
Fig. 2 Error evaluation for the presented models. a Comparison of the predicted $\kappa^{\text{SISSO}}(300\,\text{K})$ against the measured $\kappa_L(300\,\text{K})$ for the model trained against all data. The gray shaded region corresponds to the 95% confidence interval. **b** Violin plots of the mean prediction error of all samples for the SISSO, KRR, and GPR models using all features (red, left) and a reduced set including only σ^A, $\Theta_{D,\infty}$, and V_m (blue, right) and the Slack model. Gray lines are the median, white circles are the mean of the distributions, the boxes represent the quartiles, and the whiskers are the minimum and 95% absolute error. For all calculations the parameterization depth and dimension are determined by cross-validation on each training set. The red stars and blue hexagons are the outliers for the box plots. **c** A map of the two-dimensional SISSO model, where the features on the $x-$ and $y-$ axes correspond to the two features selected by SISSO. The labeled points represent the convex-hull of the scatter plot and related points.

Table 1.	List of the primary features used in this calculation.		
Name	Symbol	Unit	Domain
Anharmonicity Score (aiMD)[8]	σ^A	—	[0.075, 1.0]
Anharmonicity Score (one-shot[8])	σ^A_{OS}	—	[0.075, 1.0]
Maximum Phonon Frequency at the Γ-point	$\omega_{\Gamma,max}$	THz	[0.1, 200]
High-Temperature Limit of the Debye Temperature	$\Theta_{D,\infty}$	K	[10, 1000]
Average Phonon Temperature	Θ_P	K	[10, 10,000]
Heat Capacity	C_V	J mol^{-1} K^{-1}	[10, 5000]
Speed of sound	v_s	m s^{-1}	[500, 10,000]
Density	ρ	Da Å^{-3}	[0.25, 10]
Molar Volume	V_m	Å^3	[2.5, 1000]
Minimum Lattice Parameter	L_{min}	Å	[1, 100]
Maximum Lattice Parameter	L_{max}	Å	[1, 100]
Mean Lattice Parameter	L_{avg}	Å	[1, 100]
Reduced Mass	μ	Da	[0.2, 300]
Minimum Atomic Mass	m_{min}	Da	[1, 300]
Maximum Atomic Mass	m_{max}	Da	[1, 300]
Mean Atomic Mass	m_{avg}	Da	[1, 300]
Number of Atoms	n_{at}	$\mathbb{Z}$	[1, 1000]

by other methods (e.g., 1.36 ± 0.03 for a Gaussian Process Regression model[17] and 1.48 for a semi-empirical Debye-Callaway Model[2]). However, differences in the training sets and cross-validation scheme prevent a fair comparison of these studies for the prediction error. To see a complete representation of the training error for all models refer to Supplementary Note 2.

To get a better estimate of the prediction error, we use a nested cross-validation scheme further defined in Section IV E. As expected, the prediction error is slightly higher than the training error with an RMSE of 0.22 ± 0.02 and an AFD of 1.45 ± 0.03. As shown in Fig. 2(b), these errors are comparable to those of a KRR and GPR model trained on the same data, following the procedures listed in Sections IV B and IV C, respectively. We chose to retrain the models using the same dataset and cross-

validation splits in order to single out the effect of the methodology itself, and not changes in the data set and splits. These results show that the performance of SISSO and more traditional regression methods are similar, but the advantage of the symbolic regression models is that only seven of the primary features are selected. Another advantage of the nested cross-validation scheme is that it creates an ensemble of independent models, which can also be used to approximate the uncertainty of the predictions. These results substantiates that our symbolic regression approach performs as well as interpolative methods and it outperforms the Slack model, which was originally developed for elemental cubic solids[18]. Interestingly, offering the features of the Slack model to SISSO does not improve the results, and even some primary features previously thought to be decisive, e.g., the Grüneisen parameter, γ. are not even selected by SISSO (see Supplementary Note 5).

A key advantage of using symbolic regression techniques over interpolative methods such as KRR and GPR is that the resulting models not only yield reliable quantitative predictions, but also allows for a qualitative inspection of the underlying mechanisms. To get a better understanding of how the thermal conductivity changes across materials space we map the model in Fig. 2c. From this map we can see that the thermal conductivity of a material is mostly controlled by d_2 with d_1 providing only a minor correction. While these observed trends are already helpful, the complex non-linearities in both d_1 and d_2 impedes the generation of qualitative design rules. Furthermore, some primary features such as V_m and σ^A enter both d_1 and d_2, with contrasting trends, e.g., σ^A lowers d_1 but increases d_2. To accelerate the exploration of materials space, one must first be able to disentangle the contradicting contributions of the involved primary features.

Extracting physical understanding by identifying the most physically relevant features via sensitivity analysis

The difficulties in interpreting the "plain" SISSO descriptors described above can be overcome by performing a sensitivity analysis or a feature importance study to identify the most relevant primary features that build d_1 and d_2. For this purpose, we employ both the Sobol indices, i.e., the main effect index S_i and the total effect index S_i^{T}[47], and the Shapley Additive Explanations (SHAP)[48] metric for the model predictions. To calculate the Sobol indices we use an algorithm that includes correlative effects first described by Kucherenko et al.[49], and later implemented in
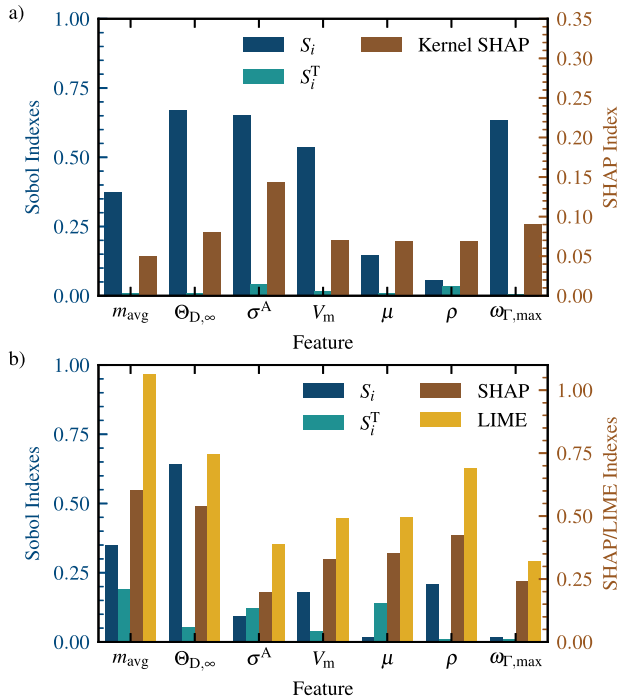
Fig. 3 The feature importance metrics for the models. S_i (first bar, dark blue), S_i^T (second bar, light blue), mean absolute SHAP index (third bar, brown), and LIME index (fourth bar, yellow) for each feature in the model by treating the inputs as (**a**) dependent feature and (**b**) independent features. The Sobol indices are plotted on the left y-axis and the SHAP and LIME indexes are plotted on the right y-axis.

UQLAB[50,51]. The main advantage of this approach is its ability to include correlative effects between the inputs, which if ignored can largely bias or even falsify the sensitivity analysis results[52]. Qualitatively, S_i quantifies how much the variance of $\log(\kappa_L(300\,\mathrm{K}))$ correlates with the variance of a primary feature, $\hat{x}_i$, and S_i^T quantifies how much the variance of $\log(\kappa_L(300\,\mathrm{K}))$ correlates with $\hat{x}_i$ including all interactions between $\hat{x}_i$ and the other primary features. For example, Sobol indices of 0.0 indicate that $\log(\kappa_L(300\,\mathrm{K}))$ is fully independent of $\hat{x}_i$, whereas a value of 1.0 indicates that $\log(\kappa_L(300\,\mathrm{K}))$ can be completely represented by changes in $\hat{x}_i$[51]. Moreover, $S_i^T < S_i$ implies that correlative effects are significant, with an $S_i^T = 0$ indicating that a primary feature is perfectly correlated to the other inputs[51].

The SHAP values constitute a local measure of how each feature influences a given prediction in the data set. This metric is based on the Shapley values used in game theory for assigning payouts to players in a game based on their contribution toward the total reward[48]. In the context of machine learning models each input to the model represents the players and the difference between individual predictions from the global mean prediction of a dataset represents the payouts[53]. The SHAP values then perfectly distribute the difference from the mean prediction to each feature for each sample, with negative values indicating that the feature is responsible for reducing the prediction from the mean and a positive value is responsible for increasing it[53]. A similar metric is the Local Interpretable Model-agnostic Explanations (LIME) values[54]. LIME first defines a local neighborhood for each data point, and then uses a similar algorithm to SHAP to compare each prediction against their corresponding local area. Because of the computational complexity of calculating SHAP values makes their exact calculation intractable with a large number of features, these values can be approximated by the Kernel SHAP method[48]. Originally the Kernel SHAP method assumed feature

independence[48], but was recently advanced to include feature dependence via sampling over a multivariate distribution represented by a set of marginal distributions and a Gaussian Copula[53]. However, there are some cases for small data sets with highly correlated features where the SHAP values are qualitatively different from the true Shapley values[55].

Figure 3 compares the different sensitivity metrics including and excluding feature dependence. To get the global values of the SHAP and LIME indexes we take the mean absolute value for each feature across all 75 materials, but other metrics have been proposed in the literature and it is not clear which one is best[56–58]. However the local information contained in metrics such as SHAP and LIME is an advantage they have over global metrics such as the Sobol indexes as it allows for the identification of regions in the material space that do not follow the global trends. Comparing the plots in Fig. 3a and b illustrates the importance of not treating the input primary features as independent, as all four sensitivity analysis metrics are qualitatively wrong under that assumption. This is likely a result of sampling over physically unreachable parts of the feature space, e.g., areas with a high density, low mass, and high molar volume, and suggests that caution should be used when applying these techniques to highly correlated datasets. The impact of this is demonstrated in Supplementary Fig. 3, where we explicitly simplify the model to remove some of the dependencies. All three indexes that include correlative effects show that σ^A, V_m, $\Theta_{D,\infty}$, and $\omega_{\Gamma,max}$ predominately control the variance of $\kappa^{SISSO}(300\,\mathrm{K})$. The main difference between S_i and the kernel SHAP metrics is the relative importance of $\Theta_{D,\infty}$ and $\omega_{\Gamma,max}$ when compared against V_m and σ^A. The difference between these results could be from the the Sobol indexes globally sampling the region of $\Theta_{D,\infty} > 1300\,\mathrm{K}$ instead of relying on the two materials in that regime or S_i over-estimating its importance because the higher correlation between $\Theta_{D,\infty}$ and the other inputs. In fact, the low values of S_i^T also imply that there are significant correlative effects in place between these inputs, and no single feature can be singled out as primarily responsible for changes in $\kappa^{SISSO}(300\,\mathrm{K})$. For instance, the similarity between the importance of $\omega_{\Gamma,max}$ and $\Theta_{D,\infty}$ is because they are strongly correlated to each other, and only one of them needs to be considered (see the Supplementary Fig. 2). The importance of these features is further substantiated in Fig. 2b, where we compare the performance of the models calculated using the full dataset and one that only includes σ^A, V_m, and $\Theta_{D,\infty}$. For all tested models, we see only a slight deterioration in performance with a predictive AFD of 1.87, 1.77, and 1.77 for the SISSO, KRR, and GPR models, respectively, compared to 1.45 for the models trained with all features. This result highlights that the trends and the underlying mechanisms describing the dependence of $\kappa_L(300\,\mathrm{K})$ in materials space are fully captured by those features alone.

Even more importantly, our model captures the interplay between these features across materials, as demonstrated in the maps in Fig. 4. These maps showcase the strong correlation between $\kappa^{SISSO}(300\,\mathrm{K})$ and σ^A, V_m, and $\Theta_{D,\infty}$, and that materials with high anharmonicity, low-energy vibrational modes, and a large molar volume will be good thermal insulators. Figure 4 shows the expected value of $\kappa^{SISSO}(300\,\mathrm{K})$, $E_{\hat{\mathcal{X}}}(\kappa^{SISSO}(300\,\mathrm{K})|\hat{\mathcal{X}})$, for different sets of input features, $\hat{\mathcal{X}}$, shown on the axes of each plot. We then overlay the maps with the actual values of each input for all materials in the training set to evaluate the trends across different groups of materials. Figure 4c confirms that σ^A is already a good indicator for finding thermal insulators, with most of the materials having $\kappa_L(300\,\mathrm{K})$ within one standard deviation of the expected value. For the more harmonic materials with $\sigma^A < 0.2$, the vanishing degree of anharmonicity is, alone, not always sufficient for quantitative predictions. In this limit, a combination of σ^A and V_m can produce correct predictions for the otherwise underestimated white triangles with a $\sigma^A < 0.2$, as seen in Fig. 4a. In order to fully describe the low thermal conductivity of the
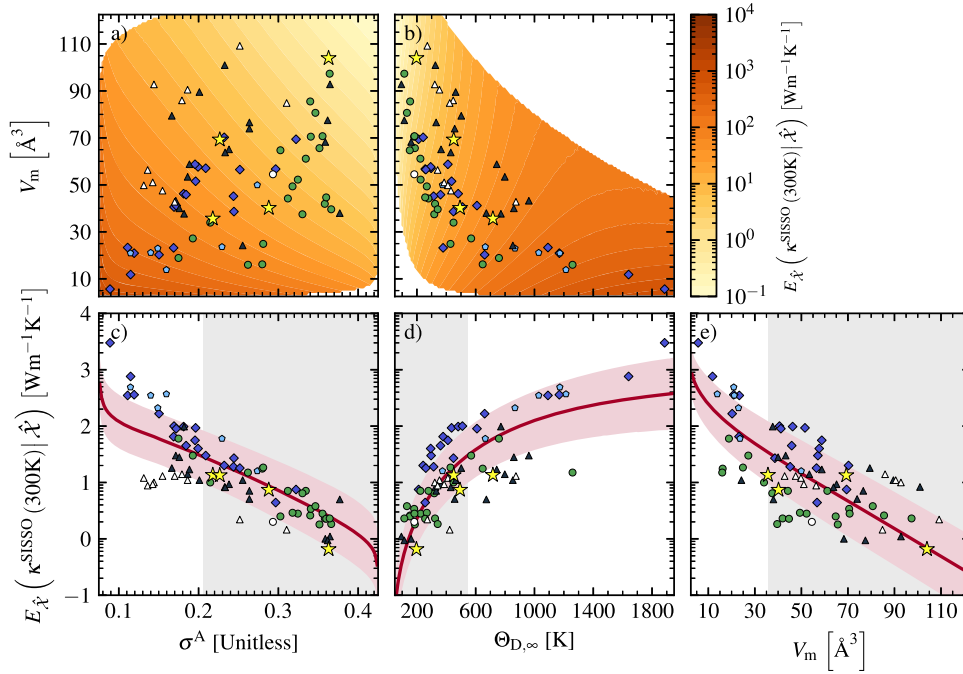
Fig. 4 The expected value of $\kappa^{SISSO}(300\,K)$ relative to select primary features. The expected value of $\kappa^{SISSO}(300\,K)$, $E_{\hat{\mathcal{X}}}\left(\kappa^{SISSO}(300\,K)\big|\hat{\mathcal{X}}\right)$, where $\hat{\mathcal{X}}$ is (**a**) $\{\sigma^A, V_m\}$, (**b**) $\{\Theta_{D,\infty}, V_m\}$, (**c**) $\{\sigma^A\}$, (**d**) $\{\Theta_{D,\infty}\}$, and (**e**) $\{V_m\}$. $E_{\hat{\mathcal{X}}}\left(\kappa^{SISSO}(300\,K)\big|\hat{\mathcal{X}}\right)$ is calculated by sampling over the multivariate distributions used for the sensitivity analysis, and binning the input data until there are at least 10,000 samples in each bin. The red line in (**c**–**e**) corresponds to $E_{\hat{\mathcal{X}}}\left(\kappa^{SISSO}(300\,K)\big|\hat{\mathcal{X}}\right)$ and the pink shaded region is one standard deviation on either side of the line. The gray shaded regions represent where a thermal conductivity of $10\,Wm^{-1}\,K^{-1}$ or lower is within one standard deviation of the expected value. On all maps all materials in the training set are displayed. The green circles correspond to rock-salts, the blue diamonds are zincblende, the light blue pentagons are wurtzites, and black triangles are all other materials. All points with a $\kappa_L(300\,K)$ less than one standard deviation below the expected value based on σ^A are highlighted in white. The points in (**c**–**e**) correspond to the actual values of $\kappa_L(300\,K)$ for each material. Additionally we include four materials outside of the training set (yellow stars) whose thermal conductivities we calculate using ab initio molecular dynamics.

remaining highlighted materials both $\Theta_{D,\infty}$ and V_m are needed as can be seen in Fig. 4a, b, d and e. Generally, this reflects that the *three* properties σ^A, $\Theta_{D,\infty}$, and V_m are the target properties to optimize to obtain ultra-low thermal conductivities.

These results can also be rationalized within our current understanding of thermal transport and showcase which physical mechanisms determine κ_L in material space. Qualitatively, it is well known that good thermal conductors typically exhibit a high degree of symmetry with a smaller number of atoms, e.g., diamond and silicon, whereas thermal insulators, e.g., glass-like materials, are often characterized by an absence of crystal symmetries and larger primitive cells. In our case, this trend is quantitatively captured via V_m, which reflects that larger unit cells have smaller thermal conductivities. Furthermore, it is well known that phonon group velocities determine how fast energy is transported through the crystal in the harmonic picture[59], and that it is limited by scattering events arising due to anharmonicity. In our model, these processes are captured by $\Theta_{D,\infty}$, which describes the degree of dispersion in the phonon band structure, and the anharmonicity measure, σ^A respectively. In this context, it is important to note that, in spite of the fact that these qualitative mechanisms were long known, there had hitherto been no agreement on which material property would quantitatively capture these mechanisms best across material space. For instance, both the γ, the lattice thermal expansion coefficient, and now σ^A, have been used to describe the anharmonicity of a material. However, when both γ and σ^A are included as primary features, only σ^A is chosen (see Supplementary Note 5 for more details). This result indicates that the σ^A measure is the more sensitive choice for modeling the strength of anharmonic effects. While γ also depends on anharmonic effects, they are also influenced by the bulk modulus, the density, and the specific heat of a material.

Validating the predictions with ab initio Green-Kubo calculations

To confirm that the discovered models produce physically meaningful predictions, we validate the estimated thermal conductivity of four materials using the ab initio Green-Kubo method (aiGK)[10,45]. This approach has recently been demonstrated to be highly accurate when compared to experiments[45], using similar DFT settings for what is done in this work. In particular aiGK is highly accurate in the low thermal conductivity regime that we are studying here. For details of how we calculate κ_L see the methodology in Section IV J. For this purpose, we chose $BrBaCl$, $LiScS_2$, CaF_2, and $GaLiO_2$, since these materials represent a broad region of the relevant feature space that also test the boundary regions of the heuristics found by the sensitivity analysis and mapping, as demonstrated by the yellow stars in Fig. 4. Figure 5 shows the convergence of the thermal conductivity of the selected materials, as calculated from three aiMD trajectories. All of the calculated thermal conductivities fall within the 95% confidence interval of the model, with the predictions for both CaF_2 and $ClBaBr$ being especially accurate. The better performance of the model for these materials is expected, as they are more similar to the training data than the hexagonal Caswellsilverite like materials. In addition, quantum nuclear effects play a more important role in $LiScS_2$ and $GaLiO_2$ than CaF_2 and $ClBaBr$, which can also explain why those predictions are worse than CaF_2 and $ClBaBr$. Overall these results demonstrate the predictive power of the discussed model.
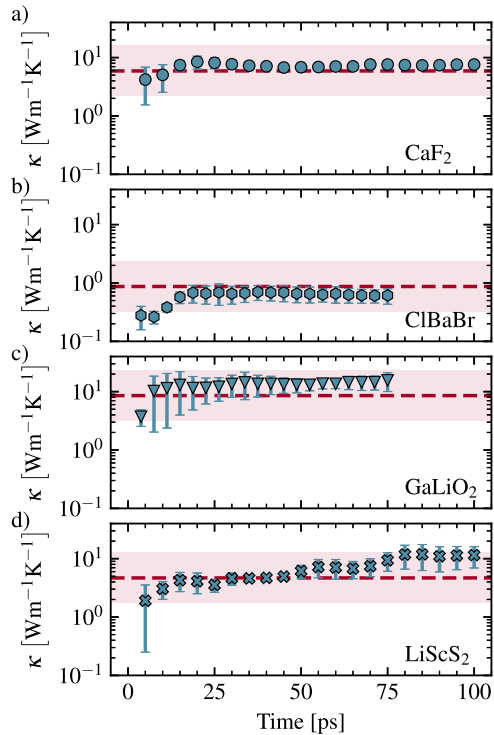
Fig. 5 Validation of the predictions of the model. The convergence of the calculated thermal conductivity of (**a**) CaF_2, (**b**) ClBaBr, (**c**) $GaLiO_2$ (**d**) $LiScS_2$. All aiGK calculations were done using the average of three 75 ps (ClBaBr and $GaLiO_2$) or 100 ps (CaF_2 and $LiScS_2$) molecular dynamics trajectories. The dashed lines are the values of the thermal conductivities predicted by Eq. (1) and the shaded region is the 95% confidence interval of the prediction based on the RMSE obtained in Fig. 2b.

Discovering improved thermal insulators

Using the information gained from the sensitivity analysis and statistical maps of the model, we are now able to design a hierarchical and efficient high-throughput screening protocol split into three stages: structure optimization, harmonic model generation, and anharmonicity quantification. We demonstrate this procedure by identifying possible thermal insulators within a set of 732 materials, within those compounds available in the materials project[60] that feature the same crystallographic prototypes[61,62] as the ones used for training. Once the geometry is optimized we remove all materials with $V_m < 35.5$ Å(60 materials) and all (almost) metallic materials (GGA bandgap < 0.2 eV), and are left with 302 candidate compounds. We then generate the converged harmonic model for the remaining materials and screen out all materials with $\Theta_{D,\infty} > 547$ K or have an unreliable harmonic model, e.g., materials with imaginary harmonic modes, leaving 148 candidates. Finally we evaluate the anharmonicity, σ^A, for the remaining materials (see Section IV D) and exclude all materials with $\sigma^A < 0.206$, and obtain 110 candidate thermal insulators. To avoid unnecessary calculations, we first estimate σ^A via σ^A_{OS} and then refine it via aiMD when $\sigma^A_{OS} > 0.4$[8]. For these candidate materials, we evaluate $\kappa^{SISSO}(300\,K)$ using Eq. (1). Of the 110 materials that passed all checks, 96 are predicted to have have a $\kappa^{SISSO}(300\,K)$ below 10 Wm^{-1}K^{-1}, illustrating the success of this method.

Finally, let us emphasize that the proposed strategy is not limited to the discovery of thermal insulators, but can be equally used to find, e.g., good thermal conductors. This is demonstrated in Fig. 6, in which we predict the thermal conductivity of all non-metallic and stable materials using the SISSO and KRR models. Generally, both the SISSO and KRR models agree with each other
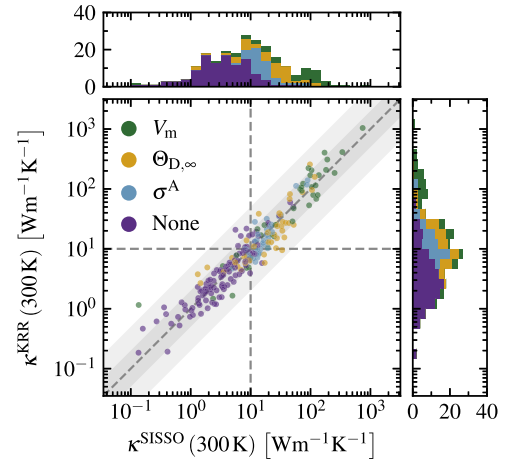

Fig. 6 A scatter plot of the prediction of both the SISSO and KRR generated models for an additional 227 materials from the same classes as the training set. σ^A is estimated via σ^A_{OS} for all materials with a $\sigma^A_{OS} \leq 0.4$ in this screening. The dataset is split up into four subsets based on if the V_m test failed (top, green), $\Theta_{D,\infty}$ test failed (second from top, yellow), σ^A test failed (third from top, blue), or none of the tests failed (bottom, purple). The outlets correspond to the histogram of all predictions using the same break down. The darker shaded region represents where both predictions are within a factor of 2 of each other and the lighter shaded region where both predictions are within a factor of 5 of each other.

with only 28 of the 227 materials having a disagreement larger than a factor of two and one ($LiHF_2$) with a disagreement larger than a factor of 5, further illustrating the reliability of these predictions. We expect that the large deviation for $LiHF_2$ is a result of the large σ^A value for that material (0.54), which is significantly larger than the maximum in the training data. We can see from the outset histograms of both the SISSO and KRR models that the hierarchical procedure successfully finds the good thermal insulators, with only 26 of the 122 materials with a $\kappa_L(300\,K) \leq 10$ Wm^{-1}K^{-1} and 10 of the 80 materials with a $\kappa_L(300\,K) \leq 5$ Wm^{-1}K^{-1} not passing all tests. Of these eight only the thermal insulating behavior of $CuLiF_2$ and Sr_2HN can not be described by the values of the other two tests that passed. Conversely, materials that do not pass the test show high conductivities. When one of the tests fail the average estimated value of $\log(\kappa_L(300\,K))$ increases to 1.38 ± 0.490 (24.0 Wm^{-1}K^{-1}), with a range of 0.95 Wm^{-1}K^{-1} to 741.3 Wm^{-1}K^{-1}. In particular, screening the materials by their molar volumes alone is a good marker for finding strong thermal conductors as all of the 15 materials with $\kappa_L(300\,K) \geq 100$ Wm^{-1}K^{-1} have a $V_m \leq 45$ Å^3.

DISCUSSION

We have developed an AI framework to facilitate and accelerate material space exploration, and demonstrate its capabilities for the urgent problem of finding thermal insulators. By combining symbolic regression and sensitivity analysis, we are able to obtain accurate predictions for a given property using relatively easy to calculate materials properties, while retaining strong physical interpretability. Most importantly, this analysis enables us to create hierarchical, high-throughput frameworks, which we used to screen over a set of more than 700 materials and find a group of ~100 possible thermal insulators. Notably, almost all of the good thermal conductors in the set of candidate materials are discarded within the first iteration of the screening, in which we only discriminate by molar volume, i.e., with an absolutely negligible computational cost compared to full calculations of κ_L. Accordingly, we expect this approach to be extremely useful in a wide range of materials problems beyond thermal transport,

especially whenever (i) few reliable data are available, (ii) additional data are hard to produce, and/or (iii) multiple physical mechanisms compete non-trivially, limiting the reliability of simplified models.

Although the proposed approach is already reliable for small dataset sizes, it obviously becomes more so when applied to larger ones. Here, the identified heuristics can substantially help steer data creation toward more interesting parts of material space. Along these lines, it is possible to iteratively refine both the SISSO model and the rules from the sensitivity analysis during material space exploration while the dataset grows. Furthermore, one can also apply the proposed procedure to the most influential primary features in a recursive fashion, learning new expressions for the computationally expensive features, e.g., σ^A, using simpler properties. In turn, this will further accelerate material discovery, but also allow for gaining further physical insights. Most importantly, this method is not limited to just the thermal conductivity of a material, and can be applied to any target property. Further extending this framework to include information about where the underlying electronic structure calculations are expected to fail, also provides a means of accelerating materials discovery more generally[63].

METHODS
SISSO

We use SISSO to discover analytical expressions for $\kappa_L(300\,\text{K})$[64]. SISSO finds low-dimensional, analytic expressions for a target property, P, by first generating an exhaustive set of candidate features, $\hat{\Phi}$, for a given set of primary features, $\hat{\Phi}_0$, and operators $\hat{\mathcal{H}}_m$, and then performing an ℓ_0-regularization over a subset of those features to find the n-dimensional subset of features, whose linear combination results in the most descriptive model. $\hat{\Phi}$ is recursively built in rungs, $\hat{\mathcal{F}}_r$, from $\hat{\Phi}_0$ and $\hat{\mathcal{H}}_m$, by applying all elements, $\hat{h}^m$, of $\hat{\mathcal{H}}^m$ on all elements $\hat{f}_i$ and $\hat{f}_j$ of $\hat{\mathcal{F}}_{r-1}$

$$\hat{\mathcal{F}}_r \equiv \hat{h}^m\left[\hat{f}_i, \hat{f}_j\right], \forall\ \hat{h}^m \in \hat{\mathcal{H}}^m\ \text{and}\ \forall\ \hat{f}_i, \hat{f}_j \in \hat{\mathcal{F}}_{r-1}.$$

$\hat{\Phi}_r$ is then the union of $\hat{\Phi}_{r-1}$ and $\hat{\mathcal{F}}_r$. Once $\hat{\Phi}$ is generated, the n_{SIS} features most correlated to P are stored in $\hat{\mathcal{S}}_1$, and the best one-dimensional models are trivially extracted from the top elements of $\hat{\mathcal{S}}_1$. Then the n_{SIS} features most correlated to any of the residuals, $\boldsymbol{\Delta}_1^i$, of the n_{res} best one-dimensional descriptors are stored in $\hat{\mathcal{S}}_2$. We define this projection as

$$s = \max(s_0, s_1, ..., s_i, ..., s_{nres}) \tag{3}$$

$$s_i = R^2\left(\hat{\phi}, \boldsymbol{\Delta}_1^i\right), \tag{4}$$

where $\hat{\phi} \in \hat{\Phi}$, and R is the Pearson correlation function. We call this approach the multiple residual approach, which was first introduced by the authors[65] and later fully described in Ref. [66]. From here, the best two-dimensional models are found by performing an ℓ_0-regularized optimization over $\hat{\mathcal{S}}_1 \cup \hat{\mathcal{S}}_2$[67]. This process is iteratively repeated until the best n-dimensional descriptor is found[64].

For this application $\hat{\mathcal{H}}_m$ contains: $A+B$, $A-B$, $A*B$, $\frac{A}{B}$, $|A-B|$, $|A|$, $(A)^{-1}$, $(A)^2$, $(A)^3$, $\sqrt{A}$, $\sqrt[3]{A}$, $\exp(A)$, $\exp(-1.0*A)$, and $\ln(A)$. In addition to ensure the units of the primary features do not affect the final results, we additionally include the following operators: $(A+\beta)^{-1}$, $(A+\beta)^2$, $(A+\beta)^3$, $\sqrt{\alpha A+\beta}$, $\sqrt[3]{A+\beta}$, $\exp(\alpha A)$, $\exp(-1.0*\alpha A)$, and $\ln(\alpha A+\beta)$, where α and β are scaling and bias constants used to adjust the input data on the fly. We find the optimal α and β terms using non-linear optimization for each of these operators[22,66,68]. To ensure that the parameterization does not result in mathematically invalid equations for data points outside of the training set, the range of each candidate feature is derived from the range of the primary features, and the upper and lower bounds for the features are set appropriately. When

generating new expressions these ranges are then used as a domain for the operations, and any expression that would lead to invalid results are excluded[66]. The range of the primary features are set to be physically relevant for the systems we are studying and are listed in Table 1. Hereafter, we call the use of these operators parametric SISSO. For more information please refer to[66].

All hyperparameters were set following the cross-validation procedures described in Section IV E.

Kernel-Ridge regression

To generate the kernel-ridge regression models we used the utilities provided by scikit-learn[69], using a radial basis function kernel with optimized regularization term and kernel length scale. The hyperparameters were selected using with a 141 by 141 point logarithmic grid search with possible parameters ranging from 10^{-7} to 10^0. Before performing the analysis each input feature, $\mathbf{x}_i$ is standardized

$$\mathbf{x}_i^{Stand} = \frac{\mathbf{x}_i - \mu_i}{\sigma_i} \tag{5}$$

where $\mathbf{x}_i^{stand}$ is the standardized input feature, μ_i is the mean of the input feature for the training data, and σ_i is the standard deviation of the input feature for the training data.

Gaussian process regression

To generate the Gaussian Process Regression Models we used the utilities provided by scikit-learn[69], using a radial basis function kernel with an optimized regularization term and kernel length scale. The hyperparameters were selected using with a 141 by 141 point logarithmic grid search with possible parameters ranging from 10^{-7} to 10^0. Before performing the analysis each input feature, $\mathbf{x}_i$ is standardized

$$\mathbf{x}_i^{Stand} = \frac{\mathbf{x}_i - \mu_i}{\sigma_i} \tag{6}$$

where $\mathbf{x}_i^{stand}$ is the standardized input feature, μ_i is the mean of the input feature for the training data, and σ_i is the standard deviation of the input feature for the training data. All uncertainty values were taken from the results of the GPR predictions, and in the case of the nested cross-validation the uncertainty was propagated using

$$\kappa_{GPR}^{pred} = \frac{1}{3}\sum_{i=1}^{3}\kappa_{GPR,i}^{pred} \tag{7}$$

$$\sigma_{GPR}^{pred} = \frac{1}{3}\sqrt{\sum_{i=1}^{3}\left(\sigma_{GPR,i}^{pred}\right)^2}, \tag{8}$$

where κ_{GPR}^{pred} and $\kappa_{GPR,i}^{pred}$ are the respective prediction and uncertainty of the ith GPR model for a given data point and κ_{GPR}^{pred} and κ_{GPR}^{pred} are the respective mean prediction and uncertainty for a prediction.

Creating the dataset

In this study we focus on only room-temperature data for κ_L, since values for other temperatures are even scarcer. However, we note that an explicit temperature dependence can be straightforwardly included using multi-task SISSO[70], and it is at least partially included via, the anharmonicity factor, σ^A[8] (see below for more details). For $\kappa_L(300\,\text{K})$, we have compiled a list of seventy-five materials from the literature (see Supplementary Table 1 for complete list with references), whose thermal conductivity has been experimentally measured. This list was curated from an initial set of over 100 materials, from which we removed all samples that are either thermodynamically unstable or are electrical

conductors. This list of materials covers a diverse set of fourteen different binary and ternary crystal structure prototypes[61,62,71].

With respect to the primary features, $\hat{\Phi}_0$, compound specific properties are provided for each material. All primary features can be roughly categorized in two classes: Structural parameters that describe the equilibrium structure and dynamical parameters that characterize the nuclear motion. For the latter case, both harmonic and anharmonic properties have been taken into account. As shown in Supplementary Note 5, additional features, such as the parameters entering the Slack model, i.e., γ, Θ_a, and V_a, can be included. However, these features do not benefit the model and when included only V_a, and not γ or Θ_a are selected. For a complete list of all primary features, and their definitions refer to Table 1.

The structural parameters relate to either the mass of the atoms (μ, m_{min}, m_{max}, m_{avg}), the lattice parameters of the primitive cell (V_m, L_{min}, L_{max}, L_{avg}), the density of the materials (ρ), or the number of atoms in the primitive cell (n_{at}). For all systems a generalization of the reduced mass, μ, is used so it can be extended to non-binary systems,

$$\mu = \left(\sum_i^{n_{emp}} \frac{1}{m_i} \right)^{-1}, \tag{9}$$

where n_{emp} is the number of atoms in the empirical formula and m_i is the mass of atom, i. Similarly, the molar volume, V_m, is calculated by

$$V_m = \frac{V_{prim}}{Z}, \tag{10}$$

where V_{prim} is the volume of the primitive cell and $Z = \frac{n_{at}}{n_{emp}}$. Finally, ρ is calculated by dividing the total mass of the empirical cell by V_m

$$\rho = \sum_i^{n_{emp}} \frac{m_i}{V_m}. \tag{11}$$

All of the harmonic properties used in these models are calculated from a converged harmonic model generated using phonopy[72]. For each material, the phonon density of states of successively larger supercells are compared using a Tanimoto similarity measure

$$S = \frac{g_{p,L}(\omega) \cdot g_{p,S}(\omega)}{\|g_{p,L}(\omega)\|^2 + \|g_{p,S}(\omega)\|^2 - g_{p,L}(\omega) \cdot g_{p,S}(\omega)}, \tag{12}$$

where S is the similarity score, $g_{p,L}(\omega)$ is the phonon density of states of the larger supercell, $g_{p,S}(\omega)$ is the phonon density of states of the smaller supercell, $A(\omega) \cdot B(\omega) = \int_0^\infty A(\omega)B(\omega)d\omega$, and $\|A(\omega)\|^2 = \int_0^\infty A^2(\omega)d\omega$. If $S > 0.80$, then the harmonic model is considered converged. From here C_V is calculated from phonopy as a weighted sum over the mode dependent heat capacities. Both approximations to the Debye temperature are calculated from the moments of the phonon density of states

$$\langle \varepsilon^n \rangle = \frac{\int d\varepsilon \, g_p(\varepsilon) \, \varepsilon^n}{\int d\varepsilon g_p(\varepsilon)} \tag{13}$$

$$\Theta_P = \frac{1}{k_B} \langle \varepsilon \rangle \tag{14}$$

$$\Theta_{D,\infty} = \frac{1}{k_B} \sqrt{\frac{5}{3} \langle \varepsilon^2 \rangle}, \tag{15}$$

where $g_p(\varepsilon)$ is the phonon density of states at energy ε[73]. Finally v_s is approximated from the Debye frequency, ω_D, by[20]

$$v_s = \left(\frac{V_a}{6\pi^2} \right)^{1/3} \omega_D, \tag{16}$$

where ω_D is approximated as

$$\omega_D = \sqrt[3]{\frac{9n_{at}}{a}} \tag{17}$$

and a is found by fitting $g_p(\omega)$ in the range $\left[0, \frac{\omega_{\Gamma,max}}{8} \right]$ to

$$g_{p,D}(\omega) = a\omega^2. \tag{18}$$

To measure the anharmonicity of the materials we use σ^A as defined in[8]

$$\sigma^A(T) = \sqrt{\frac{\sum_{I,a} \left\langle \left(F_{I,a} - F_{I,a}^{ha} \right)^2 \right\rangle_{(T)}}{\sum_{I,a} \left\langle F_{I,a}^2 \right\rangle_{(T)}}}, \tag{19}$$

in which $\langle \cdot \rangle_{(T)}$ denotes the thermodynamic average at a temperature T, $F_{I,a}$ is the a component of the force calculated from density functional theory (DFT) acting on atom I, and $F_{I,a}^{ha}$ is the same force approximated by the harmonic model[8]. First we calculate σ_{OS}^A, which uses an approximation to the thermodynamic ensemble average using the one-shot method proposed by Zacharias and Giustino[74]. In the one-shot approach the atomic positions are offset from their equilibrium positions by a vector ΔR,

$$\Delta R_I^a = \frac{1}{\sqrt{M_I}} \sum_s \zeta_s \langle A_s \rangle e_{sI}^a, \tag{20}$$

where I is the atom number, a is the component, $\mathbf{e}_s$ are the harmonic eigenvectors, $\langle A_s \rangle = \sqrt{2k_BT}/\omega_s$ is the mean mode amplitude in the classical limit[75], and $\zeta_s = (-1)^{s-1}$[74]. These displacements correspond to the turning-points of the oscillation estimated from the harmonic force constants, and is a good approximation to σ^A in the harmonic limit. Because of this, if $\sigma_{OS}^A < 0.2$ we accept that value as the true σ^A. Otherwise we calculate σ^A using aiMD in the canonical ensemble at 300 K for 10 ps, using the Langevin thermostat. When performing the high-throughput screening the threshold for when to use aiMD is increased to 0.4 because that is the point where σ_{OS}^A becomes qualitatively unreliable[8].

All electronic structure calculations are done using FHI-aims[76]. All geometries are optimized with symmetry-preserving, para-metric constraints until all forces are converged to a numerical precision better than 10^{-3} eV/Å[77]. The constraints are generated using the AFlow XtalFinder Tool[71]. All calculations use the PBEsol functional to calculate the exchange-correlation energy and an SCF convergence criteria of 10^{-6} eV/Å and 5×10^{-4} eV/Å for the density and forces, respectively. Relativistic effects are included in terms of the scalar atomic ZORA approach and all other settings are taken to be the default values in FHI-aims. For all calculations we use the *light* basis sets and numerical settings in FHI-aims. These settings were shown to ensure a convergence in lattice constants of ± 0.1 Å and a relative accuracy in phonon frequencies of 3%[8].

All primary features are calculated using the workflows defined in FHI-vibes[78].

Error evaluation

To estimate the prediction error for all models we perform a nested cross-validation, where the data are initially separated into different training and test sets using a ten-fold split. Two hyperparameters (maximum dimension and parameterization depth) are then optimized using a fivefold cross-validation on each of the training sets, and the overall performance of the model is evaluated on the corresponding test set. The size of the SIS subspace, number of residuals, and rung were all set to 2000, 10, and 3, respectively, because they did not have a large impact

on the final results. We then repeat the procedure three times and average over each iteration to get a reliable estimate of the prediction error for each sample[79].

Calculating the inputs to the Slack model

The individual components for the Slack model were the same as the ones used for the main models, with the exception of γ, V_a and Θ_a. For Θ_a, we first calculate the Debye temperature, Θ_D

$$\Theta_D = \frac{\hbar \omega_D}{k_B} \tag{21}$$

where ω_D is the same Debye frequency used for calculating v_s (see Section IV D), k_B is the Boltzmann constant, and $\hbar$ is Planck's constant. From here we calculate Θ_a using

$$\Theta_a = \frac{\Theta_D}{\sqrt[3]{n_{at}}}. \tag{22}$$

We use the phonopy definition of Θ_D instead of $\Theta_{D,\infty}$ because it is better aligned to the original definition of Θ_a. However, it is not used in the SISSO training because the initial fitting procedure to find ω_D does not produce a unique value for Θ_D and it is already partially included via v_s. To calculate the thermodynamic Grüneisen parameter we use the utilities provided by phonopy[72]. The atomic volume was calculated by taking the volume of the primitive cell and dividing it by the total number of atoms.

Calculating the Sobol indexes

Formally, the Sobol indices are defined as

$$S_i = \frac{\mathrm{Var}_{\hat{x}_i}\left(E_{\widetilde{\mathcal{X}_i}}(\log(\kappa_L(300\,\mathrm{K}))|\hat{x}_i)\right)}{\mathrm{Var}(\log(\kappa_L(300\,\mathrm{K})))} \tag{23}$$

$$S_i^{\mathrm{T}} = 1 - \frac{\mathrm{Var}_{\widetilde{\mathcal{X}_i}}\left(E_{\hat{x}_i}\left(\log(\kappa_L(300\,\mathrm{K}))|\widetilde{\mathcal{X}_i}\right)\right)}{\mathrm{Var}(\log(\kappa_L(300\,\mathrm{K})))} \tag{24}$$

where $\hat{x}_i \in \hat{\mathcal{X}}$ is one of the inputs to the model, $\mathrm{Var}_a(B)$ is the variance of B with respect to a, $E_a(B)$ is the mean of B after sampling over a, and $\widetilde{\mathcal{X}_i}$ is the set of all variables excluding $\hat{x}_i$.

Normally, it is assumed that all elements of $\hat{\mathcal{X}}$ are independent of each other, and this assumption is preserved when calculating S_i and S_i^T in Fig. 3b. As a result of this, the variance of $\log(\kappa^{\mathrm{SISSO}}(300\,\mathrm{K}))$ and the required expectation values would be calculated from sampling over an n_v-dimensional hypercube covering the full input range, ignoring the correlation between the input variables. However, in order to properly model the correlative effects between elements of $\hat{\mathcal{X}}$, Kucherenko et al. modify this sampling approach[49,51]. The first step of the updated algorithm is to fit the input data to a set of marginal univariate distributions coupled together via a copula[49,51]. The algorithm then samples over an n_v-dimensional unit-hypercube and transforms these samples into the correct variable space using a transform defined by the fitted distributions and copulas (see Supplementary Note 3 for more details). It was later demonstrated that when using the approach proposed by Kucherenko and coworkers to calculate the Sobol indices, S_i includes effects from the dependence of $\hat{x}_i$ on those in $\widetilde{\mathcal{X}_i}$, while S_i^T is independent of these effects[80]. We use this updated algorithm to calculate S_i and S_i^T in Fig. 3a. In both cases we use the implementation in UQLab[50] to calculate S_i and S_i^T.

Calculating the SHAP indexes

The SHAP values are calculated by treating the features as independent variables using the original method proposed by Lundberg and Lee[48], as implemented in the python package SHAP, and as dependent variables using shapr by Aas, et al.[53]. The SHAP values are an extension of the Shapley values from cooperative game theory, that distributes the contribution, $v(\mathcal{S})$, of each player or subset of players, $\mathcal{S} \subseteq \mathcal{M} = \{1, \cdot, M\}$, where $\mathcal{M}$ is the set of all players[48,53]. The Shapley value, $\phi_j(v) = \phi_j$, can then be calculated by taking a weighted mean over the contribution function differences for all $\mathcal{S}$ not containing the player, j,

$$\phi_j = \sum_{\mathcal{S} \subseteq \mathcal{M} \setminus \{j\}} \frac{|\mathcal{S}|!(M - |\mathcal{S}| - 1)!}{M!} (v(\mathcal{S} \cup \{j\}) - v(\mathcal{S})), j = 1, \cdots, M, \tag{25}$$

where $|\mathcal{S}|$ is the number of members in $\mathcal{S}$[53]. For a machine learning problem with a training set $\{y^i, x^i\}_{i=1,\cdots,n_{\mathrm{train}}}$, where y^i is the property value and x^i are the target property value and input feature values for the ith data point in the training set with n_{train} data points[48,53], we can explain the prediction of the model, $f(x^*)$ for a particular point, x^*, with

$$f(x^*) = \phi_0 + \sum_{j=1}^{M} \phi_j^*, \tag{26}$$

where ϕ_0 is the mean prediction and ϕ_j^* is the Shapley value for the jth feature for a prediction $x = x^*$. Essentially the Shapley value for the model describes the difference between a prediction, $y^* = f(x^*)$, and the mean of all predictions[48,53]. The contribution function is then defined as

$$v(\mathcal{S}) = E\left[f(x)|x_{\mathcal{S}} = x_{\mathcal{S}}^*\right], \tag{27}$$

which is the expectation value of the model conditional on $x_{\mathcal{S}} = x_{\mathcal{S}}^*$[48,53]. The expectation value can be calculated as

$$E\left[f(x)|x_{\mathcal{S}} = x_{\mathcal{S}}^*\right] = E\left[f\left(x_{\widetilde{\mathcal{S}}}, x_{\mathcal{S}}\right)\Big|x_{\mathcal{S}} = x_{\mathcal{S}}^*\right]$$
$$= \int f\left(x_{\widetilde{\mathcal{S}}}, x_{\mathcal{S}}\right) p\left(x_{\widetilde{\mathcal{S}}}\Big|x_{\mathcal{S}} = x_{\mathcal{S}*}\right) dx_{\widetilde{\mathcal{S}}}, \tag{28}$$

where $x_{\widetilde{\mathcal{S}}}$ is the subset of all features not included in $\mathcal{S}$ and $p\left(x_{\widetilde{\mathcal{S}}}\Big|x_{\mathcal{S}} = x_{\mathcal{S}*}\right)$ is the conditional probability distribution of $x_{\widetilde{\mathcal{S}}}$ given $x_{\mathcal{S}} = x_{\mathcal{S}*}$[48,53]. In the case where the features are treated independently, $p\left(x_{\widetilde{\mathcal{S}}}\Big|x_{\mathcal{S}} = x_{\mathcal{S}*}\right)$ is replaced by $p\left(x_{\widetilde{\mathcal{S}}}\right)$ and $v(\mathcal{S})$ can be approximated by Monte Carlo integration

$$v(\mathcal{S}) = \frac{1}{K} \sum_{k=1}^{K} f\left(x_{\widetilde{\mathcal{S}}}^k, x_{\mathcal{S}}^*\right), \tag{29}$$

where $x_{\widetilde{\mathcal{S}}}^k$ are samples from the training data, and K is the number of samples taken[48,53]. To include feature dependence the marginal distributions of the training data are converted into a Gaussian copula and that is used to generate samples for the Monte Carlo integration[53].

Because the number of subsets that need to be explored grows as 2^M for the number of features, calculating the exact Shapley values for a large number of inputs becomes intractable. To remove this constraint the problem can be approximated as the optimal solution of a weighted least squares problem, which can be described as Kernel SHAP, which is described in refs. [48,53].

Calculating the LIME indexes

For the LIME values we use the LIME package in python[54]. The values were calculated using the standard tabular explainer using all features in the model and the mean absolute value of each prediction for each feature was used to asses the global feature importance. The methodology assumes the features are independent and for algorithmic details see ref. [54]

Calculating the thermal conductivity

To calculate κ_L, we use the ab initio Green-Kubo (aiGK) method[10,81]. The aiGK method calculates the $\alpha\beta$ component of the thermal conductivity tensor, $\kappa^{\alpha\beta}$, of a material for a given

volume V, pressure p, and temperature T with

$$\kappa^{\alpha\beta}(T,p) = \frac{V}{k_B T^2} \lim_{\tau \to \infty} \int_0^\tau \langle G[\mathbf{J}]^{\alpha\beta}(\tau') \rangle_{(T,p)} d\tau' \tag{30}$$

where k_B is Boltzmann's constant, $\langle \cdot \rangle_{(T,p)}$ denotes an ensemble average, $\mathbf{J}(t)$ is the heat flux, and $G[\mathbf{J}]$ is the time-(auto)correlation functions

$$G[\mathbf{J}]^{\alpha\beta} = \lim_{t_0 \to \infty} \frac{1}{t_0} \int_0^{t_0-\tau} J^\alpha(t) J^\beta(t+\tau) dt. \tag{31}$$

The heat flux of each material is calculated from aiMD trajectories using the following definition

$$\mathbf{J}(t) = \sum_I \boldsymbol{\sigma}_I \dot{\mathbf{R}}_I, \tag{32}$$

where $\mathbf{R}_I$ is the position of the ith-atom and $\boldsymbol{\sigma}_I$ is the contribution of the ith atom to the stress tensor, $\boldsymbol{\sigma} = \sum_I \boldsymbol{\sigma}_I$[10]. From here κ_L is calculated as

$$\kappa_L = \frac{1}{3}\text{Tr}[\boldsymbol{\kappa}] \tag{33}$$

All calculations were done using both FHI-vibes[78] and FHI-aims with the same settings as the previous calculations[8] (see Section IV D for more details). The molecular dynamics calculations were done using a 5 fs time step in the NVE ensemble, with the initial structures taken from a 10 ps NVT trajectory. Three MD calculations were done for each material and the κ_L was taken to be the average of all three runs.

DATA AVAILABILITY
All raw electronic structure data can be found on the NOMAD archive (https://doi.org/10.17172/NOMAD/2022.04.27-1)[82]. All processed data and figure creation scripts can be found on figshare (https://doi.org/10.6084/m9.figshare.22068749.v4)[83]. A reproduction notebook can be found on the NOMAD AI Toolkit (https://nomad-lab.eu/aitutorials/kappa-sisso).

CODE AVAILABILITY
SISSO++[22] and FHI-vibes[78] were used to generate all data and analysis in the paper and are freely available online in the cited publications. All electronic structure calculations were done using FHI-aims[76], which is freely available for use for academic use (with a voluntary donation) (https://fhi-aims.org/get-the-code-menu/get-the-code). The Sobol indexes are calculated with UQLab[50,51] (https://www.uqlab.com/download) and the kernel SHAP values were found with shapr[53] (https://github.com/NorskRegnesentral/shapr) which are open source. The python SHAP library[48] was also used for the independent SHAP values, and is open source (https://github.com/slundberg/shap).

REFERENCES
1. Stanev, V., Choudhary, K., Kusne, A. G., Paglione, J. & Takeuchi, I. Artificial intelligence for search and discovery of quantum materials. *Commun. Mater.* **2**, 105 (2021).
2. Miller, S. A. et al. Capturing Anharmonicity in a Lattice Thermal Conductivity Model for High-Throughput Predictions. *Chem. Mater.* **29**, 2494 (2017).
3. Gomes, C. P., Selman, B. & Gregoire, J. M. Artificial intelligence for materials discovery. *MRS Bull.* **44**, 538 (2019).
4. Zhang, Q., Uchaker, E., Candelaria, S. L. & Cao, G. Nanomaterials for energy conversion and storage. *Chem. Soc. Rev.* **42**, 3127 (2013).
5. Christian Enger, B., Lødeng, R. & Holmen, A. A review of catalytic partial oxidation of methane to synthesis gas with emphasis on reaction mechanisms over transition metal catalysts. *Appl. Catal. A Gen.* **346**, 1 (2008).
6. Wu, W. et al. Preparation and thermal conductivity enhancement of composite phase change materials for electronic thermal management. *Energy Convers. Manag.* **101**, 278 (2015).
7. Pollock, T. M. Alloy design for aircraft engines. *Nat. Mater.* **15**, 809 (2016).
8. Knoop, F., Purcell, T. A. R., Scheffler, M. & Carbogno, C. Anharmonicity measure for materials. *Phys. Rev. Mater.* **4**, 083809 (2020).
9. Broido, D. A., Malorny, M., Birner, G., Mingo, N. & Stewart, D. A. Intrinsic lattice thermal conductivity of semiconductors from first principles. *Appl. Phys. Lett.* **91**, 231922 (2007).
10. Carbogno, C., Ramprasad, R. & Scheffler, M. Ab Initio Green-Kubo Approach for the Thermal Conductivity of Solids. *Phys. Rev. Lett.* **118**, 175901 (2017).
11. Carrete, J., Li, W., Mingo, N., Wang, S. & Curtarolo, S. Finding Unprecedentedly Low-Thermal-Conductivity Half-Heusler Semiconductors via High-Throughput Materials Modeling. *Phys. Rev. X* **4**, 11019 (2014).
12. Seko, A. et al. Prediction of Low-Thermal-Conductivity Compounds with First-Principles Anharmonic Lattice-Dynamics Calculations and Bayesian Optimization. *Phys. Rev. Lett.* **115**, 205901 (2015).
13. Xia, Y. Revisiting lattice thermal transport in PbTe: The crucial role of quartic anharmonicity. *Appl. Phys. Lett.* **113**, 073901 (2018).
14. Zhu, T. et al. Charting lattice thermal conductivity for inorganic crystals and discovering rare earth chalcogenides for thermoelectrics. *Energy Environ. Sci.* **14**, 3559 (2021).
15. Springer Materials. http://materials.springer.com.
16. Zhang, Y. & Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Comput. Mater.* **4**, 25 (2018).
17. Chen, L., Tran, H., Batra, R., Kim, C. & Ramprasad, R. Machine learning models for the lattice thermal conductivity prediction of inorganic materials. *Comput. Mater. Sci.* **170**, 109155 (2019).
18. Slack, G.A. The Thermal Conductivity of Nonmetallic Crystals. In *Solid State Phys. - Adv. Res. Appl.*, vol. 34, 1–71 (Academic Press, 1979).
19. Yan, J. et al. Material descriptors for predicting thermoelectric performance. *Energy Environ. Sci.* **8**, 983 (2015).
20. Toberer, E. S., Zevalkink, A. & Snyder, G. J. Phonon engineering through crystal chemistry. *J. Mater. Chem.* **21**, 15843 (2011).
21. Wang, Y., Wagner, N. & Rondinelli, J. M. Symbolic regression in materials science. *MRS Commun.* **9**, 793–805 (2019).
22. Purcell, T. A. R., Scheffler, M., Carbogno, C. & Ghiringhelli, L. M. SISSO++: A C++ Implementation of the Sure-Independence Screening and Sparsifying Operator Approach. *J. Open Source Softw.* **7**, 3960 (2022).
23. Schleder, G. R., Acosta, C. M. & Fazzio, A. Exploring Two-Dimensional Materials Thermodynamic Stability via Machine Learning. *ACS Appl. Mater. Interfac.* **12**, 20149 (2020).
24. Han, Z.-K. et al. Single-atom alloy catalysts designed by first-principles calculations and artificial intelligence. *Nat. Commun.* **12**, 1833 (2021).
25. Pilania, G., Iverson, C. N., Lookman, T. & Marrone, B. L. Machine-Learning-Based Predictive Modeling of Glass Transition Temperatures: A Case of Polyhydroxyalkanoate Homopolymers and Copolymers. *J. Chem. Inf. Model.* **59**, 5013 (2019).
26. Morelli, D. T. & Slack, G. A. High Lattice Thermal Conductivity Solids. In *High Therm. Conduct. Mater.*, 37–68 (Springer, New York, NY, New York, 2006).
27. Slack, G. A. Thermal Conductivity of MgO, Al_2O_3, $MgAl_2O_4$, and Fe_3O_4 Crystals from 3° to 300°K. *Phys. Rev.* **126**, 427–441 (1962).
28. Martin, J. Thermal conductivity of Mg_2Si, Mg_2Ge and Mg_2Sn. *J. Phys. Chem. Solids* **33**, 1139–1148 (1972).
29. Takahashi, T. & Kikuchi, T. Porosity dependence on thermal diffusivity and thermal conductivity of lithium oxide Li_2O from 200 to 900°C. *J. Nucl. Mater.* **91**, 93–102 (1980).
30. Turkes, P., Pluntke, C. & Helbig, R. Thermal conductivity of SnO_2 single crystals. *J. Phys. C. Solid State Phys.* **13**, 4941–4951 (1980).
31. Gerlich, D. & Andersson, P. Temperature and pressure effects on the thermal conductivity and heat capacity of CsCl, CsBr and CsI. *J. Phys. C. Solid State Phys.* **15**, 5211 (1982).
32. Williams, R. K., Graves, R. S. & McElroy, D. L. Thermal Conductivity of Cr_2O_3 in the Vicinity of the Neel Transition. *J. Am. Ceram. Soc.* **67**, C–151 (2006).
33. Valeri-Gil, M. & Rincón, C. Thermal conductivity of ternary chalcopyrite compounds. *Mater. Lett.* **17**, 59 (1993).
34. Morelli, D. T. et al. Low-temperature transport properties of p -type $CoSb_3$. *Phys. Rev. B* **51**, 9622–9628 (1995).
35. Hohl, H. et al. Efficient dopants for ZrNiSn-based thermoelectric materials. *J. Phys. Condens. Matter* **11**, 1697–1709 (1999).
36. Young, D. P., Khalifah, P., Cava, R. J. & Ramirez, A. P. Thermoelectric properties of pure and doped FeMSb (M=V,Nb). *J. Appl. Phys.* **87**, 317–321 (2000).
37. Li, J.-G., Ikegami, T. & Mori, T. Fabrication of transparent Sc_2O_3 ceramics with powders thermally pyrolyzed from sulfate. *J. Mater. Res.* **18**, 1816–1822 (2003).
38. Kawaharada, Y., Kurosaki, K., Muta, H., Uno, M. & Yamanaka, S. High temperature thermoelectric properties of CoTiSb half-Heusler compounds. *J. Alloy. Compd.* **384**, 308–311 (2004).

39. Víllora, E. G., Shimamura, K., Yoshikawa, Y., Ujiie, T. & Aoki, K. Electrical conductivity and carrier concentration control in β-Ga_2O_3 by Si doping. *Appl. Phys. Lett.* **92**, 202120 (2008).

40. Toher, C. et al. High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasiharmonic Debye model. *Phys. Rev. B* **90**, 174107 (2014).

41. Lu, Y. et al. Fabrication of thermoelectric $CuAlO_2$ and performance enhancement by high density. *J. Alloy. Compd.* **650**, 558 (2015).

42. Huang, W. et al. Investigation of thermodynamics properties of chalcopyrite compound $CdGeAs_2$. *J. Cryst. Growth* **443**, 8 (2016).

43. Pantian, S., Sakdanuphab, R. & Sakulkalavek, A. Enhancing the electrical conductivity and thermoelectric figure of merit of the p-type delafossite $CuAlO_2$ by Ag_2O addition. *Curr. Appl. Phys.* **17**, 1264 (2014).

44. Xia, Y. et al. High-Throughput Study of Lattice Thermal Conductivity in Binary Rocksalt and Zinc Blende Compounds including Higher-Order Anharmonicity. *Phys. Rev. X* **10**, 041029 (2020).

45. Knoop, F., Purcell, T. A. R., Scheffler, M. & Carbogno, C. Anharmonicity in Thermal Insulators - An Analysis from First Principles. *Phys. Rev. Lett.* **130**, 236301 (2023).

46. Foppa, L., Purcell, T. A., Levchenko, S. V., Scheffler, M. & Ghringhelli, L. M. Hierarchical Symbolic Regression for Identifying Key Physical Parameters Correlated with Bulk Properties of Perovskites. *Phys. Rev. Lett.* **129**, 55301 (2022).

47. Sobol', I. M. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414 (1993).

48. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In (eds Guyon, I. et al.) *Advances in Neural Information Processing Systems 30*, 4765–4774 (Curran Associates, Inc., 2017).

49. Kucherenko, S., Tarantola, S. & Annoni, P. Estimation of global sensitivity indices for models with dependent variables. *Comput. Phys. Commun.* **183**, 937 (2012).

50. Marelli, S. & Sudret, B.*UQLab: A Framework for Uncertainty Quantification in Matlab*, 2554–2563 (American Society of Civil Engineers, Reston, VA, 2014).

51. Wiederkehr, P.*Global Sensitivity Analysis with Dependent Inputs*. Ph.D. thesis, (ETH Zurich, 2018).

52. Razavi, S. et al. The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environ. Model. Softw.* **137**, 104954 (2021).

53. Aas, K., Jullum, M. & Løland, A. Explaining individual predictions when features are dependent: More accurate approximations to shapley values. *Artif. Intell.* **298**, 103502 (2021).

54. Ribeiro, M. T., Singh, S. & Guestrin, C. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17, 2016*, 1135–1144 (2016).

55. Roder, J., Maguire, L., Georgantas, R. & Roder, H. Explaining multivariate molecular diagnostic tests via shapley values. *BMC Med. Inform. Decis. Mak.* **21**, 1–18 (2021).

56. Lee, Y. G., Oh, J. Y., Kim, D. & Kim, G. Shap value-based feature importance analysis for short-term load forecasting. *J. Electr. Eng. Technol.* **18**, 579–588 (2022).

57. Ittner, J., Bolikowski, L., Hemker, K. & Kennedy, R. Feature synergy, redundancy, and independence in global model explanations using shap vector decomposition. Preprint at https://arxiv.org/abs/2107.12436v1 (2021).

58. Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine learning models using improved shapley additive explanation. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, 546 (Association for Computing Machinery, New York, NY, USA, 2019).

59. Peierls, R. E. *Quantum theory of solids* (Oxford University Press, 1955).

60. Jain, A. et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

61. Mehl, M. J. et al. The AFLOW Library of Crystallographic Prototypes: Part 1. *Comput. Mater. Sci.* **136**, S1 (2017).

62. Hicks, D. et al. The AFLOW Library of Crystallographic Prototypes: Part 2. *Comput. Mater. Sci.* **161**, S1 (2019).

63. Duan, C., Liu, F., Nandy, A. & Kulik, H. J. Putting Density Functional Theory to the Test in Machine-Learning-Accelerated Materials Discovery. *J. Phys. Chem. Lett.* **12**, 4628 (2021).

64. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2**, 83802 (2018).

65. Foppa, L. et al. Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence. *MRS Bull.* **46**, 1016 (2021).

66. Purcell, T.A., Scheffler, M. & Ghiringhelli, L.M. Recent advances in the sisso method and their implementation in the sisso++ code.Preprint at https://arxiv.org/abs/2305.01242 (2023).

67. Ghiringhelli, L. M. et al. Learning physical descriptors for materials science by compressed sensing. *N. J. Phys.* **19**, 023017 (2017).

68. Johnson, S. G. The NLopt nonlinear-optimization package. http://github.com/stevengj/nlopt.

69. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825 (2011).

70. Ouyang, R., Ahmetcik, E., Carbogno, C., Scheffler, M. & Ghiringhelli, L. M. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *J. Phys. Mater.* **2**, 24002 (2019).

71. Hicks, D. et al. AFLOW-XtalFinder: a reliable choice to identify crystalline prototypes. *npj Comput. Mater.* **7**, 30 (2021).

72. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1 (2015).

73. Pässler, R. Basic moments of phonon density of states spectra and characteristic phonon temperatures of group IV, III-V, and II-VI materials. *J. Appl. Phys.* **101**, 093513 (2007).

74. Zacharias, M. & Giustino, F. One-shot calculation of temperature-dependent optical spectra and phonon-induced band-gap renormalization. *Phys. Rev. B* **94**, 75125 (2016).

75. Dove, M. *Introduction to lattice dynamics* (Cambridge University Press, 1993).

76. Blum, V. et al. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **180**, 2175 (2009).

77. Lenz, M.-O. et al. Parametrically constrained geometry relaxations for high-throughput materials science. *npj Comput. Mater.* **5**, 123 (2019).

78. Knoop, F., Purcell, T. A. R., Scheffler, M. & Carbogno, C. FHI-vibes: Ab Initio Vibrational Simulations. *J. Open Source Softw.* **5**, 2671 (2020).

79. Krstajic, D., Buturovic, L. J., Leahy, D. E. & Thomas, S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminform.* **6**, 10 (2014).

80. Mara, T. A., Tarantola, S. & Annoni, P. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environ. Model. Softw.* **72**, 173 (2015).

81. Ravichandran, N. K. & Broido, D. Unified first-principles theory of thermal properties of insulators. *Phys. Rev. B* **98**, 085205 (2018).

82. Purcell, T. A., Scheffler, M., Ghiringhelli, L. M. & Carbogno, C. Thermal Conductivity Screening Data https://doi.org/10.17172/NOMAD/2022.04.27-1 (2022).

83. Purcell, T. A., Scheffler, M., Ghiringhelli, L. M. & Carbogno, C. Accelerating Materials-Space Exploration for Thermal Insulators by Mapping Materials Properties via Artificial Intelligence: Figures https://doi.org/10.6084/m9.figshare.22068749.v4 (2023).

AUTHOR CONTRIBUTIONS

T.A.R.P. implemented all methods and performed all calculations. T.A.R.P. and C.C. ideated the workflow. M.S., L.M.G., and C.C. supervised the project. All authors analyzed the data and wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-023-01063-y.

Correspondence and requests for materials should be addressed to Thomas A. R. Purcell, Luca M. Ghiringhelli or Christian Carbogno.