

Review

Artificial Intelligence for Chalcogenide Materials: Evidence, Validation, and Emerging Closed-Loop Workflows

Marie-Christine Record *  and Pascal Boulet * 

IM2NP, Faculty of Sciences, CNRS, Aix-Marseille University, Avenue Escadrille Normandie-Niemen, 13103 Marseille, France

* Correspondence: m-c.record@univ-amu.fr (M.-C.R.); pascal.boulet@univ-amu.fr (P.B.)

Abstract

Chalcogenides challenge materials artificial intelligence (AI) because their properties depend on defects, disorder, metastability, interfaces, processing history, and operating conditions, as well as composition and structure. This review examines AI applications across transition-metal dichalcogenides, photovoltaic materials, thermoelectrics, phase-change materials, chalcogenide perovskites, and glasses. It combines structured bibliographic analysis with critical assessment of representative studies rather than exhaustive evaluation of all retrieved publications. Its central contribution is a claim-specific, five-level evidence framework: Level 0, method transfer or proposal; Level 1, internal retrospective validation; Level 2, domain-aware or external validation; Level 3, prospective computational or experimental validation; and Level 4, closed-loop operation in which new results guide subsequent decisions. Evidence levels characterize individual claims, not entire papers. Descriptor-based models, graph neural networks, and machine-learned interatomic potentials offer complementary capabilities, but predictive accuracy alone does not establish transferability or practical reliability. Prospective validation is documented in selected prediction, synthesis, and processing applications. Among the chalcogenide-specific studies examined, however, no workflow met the Level 4 criteria for repeated model-updated closed-loop operation. Progress requires traceable datasets, leakage-resistant evaluation, uncertainty assessment under relevant distribution shifts, physically meaningful validation, and stronger connections between modelling, synthesis, processing, and characterization. The framework clarifies existing evidence and priorities for future validation.

Keywords: chalcogenides; artificial intelligence; machine learning; materials informatics; graph neural networks; machine-learned interatomic potentials; out-of-distribution validation; active learning; autonomous laboratories



Academic Editor: Zhimei Sun

Received: 27 August 2026

Revised: 16 October 2026

Accepted: 19 October 2026

Published: 27 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Chalcogenides are a broad group of materials containing sulfur, selenium, or tellurium. They include layered transition-metal dichalcogenides (TMDCs), chalcopyrite and kesterite photovoltaic materials, thermoelectric tellurides and selenides, Ge-Sb-Te phase-change alloys, chalcogenide perovskites, and amorphous glasses. In these materials, composition, structure, dimensionality, bonding, defects, interfaces, and processing conditions can be changed over a wide range. This gives chalcogenides many possible applications, including photovoltaics, thermoelectric conversion, non-volatile memory, photonics, sensing, and optoelectronics [1–12].

However, this diversity also makes chalcogenides difficult for data-driven modelling. A random train-test split can give good results when similar compositions or structures are present in both the training and test sets. But in real applications, the model may need to predict a new chemistry, defect population, amorphous structure, interface, temperature range, or processing condition. Studies using out-of-distribution and temporal tests show that prediction errors usually increase in these situations. They also show that the relative performance of graph models can change outside the training distribution [13,14]. Studies on uncertainty estimation reach a similar conclusion: uncertainty and distance from the training domain should be tested under the type of distribution shift that is important for the final application [15,16]. This is especially important for chalcogenides because disorder, metastability, and processing history often directly affect their properties.

For this reason, the important question is not only whether machine learning can predict a property. We also need to ask what the model has really learned and under which conditions its predictions remain reliable. This depends on the source of the reference data, the model representation, the train-test split, the uncertainty estimate, and the calculations or experiments used for validation.

For example, a model trained on semilocal density-functional-theory (DFT) band gaps may reproduce the DFT values very well, but it can still keep the systematic difference between these values and experiment. A general interatomic potential may work well near equilibrium but fail for defects, liquids, amorphous states, or bond-breaking events that were not included in the training data. In the same way, a synthesis optimization model may find useful conditions in one experimental system without showing that it can work in another reactor or with different precursors. Therefore, similar numerical errors can correspond to very different levels of scientific evidence.

The chalcogenide literature now includes studies with very different levels of maturity. Descriptor-based models have been used for band-gap prediction and composition analysis related to devices [1,17–26]. Graph and deep-learning models have been used for structure-property prediction and transferable representations [27–31]. Machine-learned interatomic potentials (MLIPs) have enabled large-scale simulations of TMDC defects, interfaces, thermal transport, and phase-change processes [4,5,10,32–47]. Active learning and Bayesian optimization have been applied to thermoelectric screening, multinary synthesis, thin-film growth, and processing [48–59]. Generative models and autonomous laboratories also point toward inverse design and closed-loop experiments, although the strongest demonstrations are still mainly for general inorganic materials rather than chalcogenides specifically [60–62].

Several reviews already cover important parts of this field. Yin et al. reviewed machine learning for two-dimensional materials, including materials preparation, atomistic-structure analysis, and property prediction [63]. Wang et al. reviewed machine learning for thermoelectric materials, with a focus on the prediction and optimization of electronic and thermal transport properties [64]. Zhang et al. focused on chalcogenide phase-change materials, including high-throughput analysis, machine-learned interatomic potentials, phase-transition dynamics, and device design [65]. Liu et al. reviewed machine learning for glass science and engineering, including composition-property modelling and combined physical and data-driven approaches [66].

The aim of the present review is to complement these previous reviews. Our main question is not which machine-learning model performs best for one material family. Instead, we ask what level of scientific evidence supports the different claims made across chalcogenide families and tasks. We therefore compare transition-metal dichalcogenides, photovoltaic materials, thermoelectrics, phase-change materials, chalcogenide perovskites, glasses, and related systems using the same evidence framework.

This approach has three main features. First, we combine quantitative bibliographic analysis with a detailed examination of representative primary studies. This allows us to separate the amount of literature in a field from the strength of the evidence provided by individual studies.

Second, we use the same claim-specific validation framework for different applications, including property prediction, atomistic modelling, characterization, synthesis optimization, generative design, and closed-loop experiments. We distinguish between internal validation, transfer to another domain, prospective validation, and repeated autonomous operation, even when similar machine-learning methods are used.

Third, we compare common validation problems across material families. These include data provenance, data leakage, distribution shift, uncertainty calibration, disorder, differences in data fidelity, and the timing of model recommendations relative to confirmatory calculations or experiments. We then use these problems to define practical reporting and validation requirements.

The main contribution of this review is therefore a cross-family analysis of scientific evidence and validation, rather than a list of machine-learning algorithms.

Throughout the review, we ask several questions. What physical or experimental quantity is the model learning, and how reliable are the reference data? What variable is really excluded from the training set? Does the evaluation test simple interpolation, new chemistry or structure, or real prospective use? Is the uncertainty useful under the distribution shift expected in practice? Was the model recommendation made before the calculation or experiment used to confirm it? Finally, for autonomous workflows, was the new experimental result used to update the next decision?

Using the same questions for different chalcogenide families makes it possible to compare strengths and limitations without assuming that prediction metrics from very different tasks are directly comparable.

2. Review Design and Evidence Framework

2.1. Scope and Review Type

This work is a structured critical review combined with quantitative bibliographic analysis. It is not a fully systematic review in which every retrieved record is examined in detail.

We used reproducible rules based on publication metadata to describe publication trends and to identify material families and scientific tasks that were clearly mentioned in the titles.

The scientific discussion is based on representative studies. For these studies, we checked the material system, the role of AI, the data source, the validation method, and the main contribution. Abstracts or full texts were examined when necessary.

We also included important papers on materials AI and studies on other inorganic materials when they were useful for understanding or comparing the chalcogenide literature.

The main analysis covered the period 2015–2026. This period corresponds to the rapid development of modern materials informatics, graph neural networks, machine-learned interatomic potentials (MLIPs), active learning, and generative models.

Some earlier references were also included when they were important for scientific context, for example for benchmark chalcogenide properties, important databases, or foundational models [8,12,27,67–71].

2.2. Search Strategy and Corpus Construction

CAS SciFinder-n (American Chemical Society) was used as the main bibliographic database because it provides broad coverage of chemistry and materials science.

Nine overlapping searches were performed. They covered general AI in chalcogenides, transition-metal dichalcogenides, photovoltaics, phase-change materials, thermoelectrics, property prediction, graph neural networks, synthesis optimization, and XRD/Raman/SEM/TEM or spectroscopic analysis.

No publication-year limit was used during the searches. The period 2015–2026 was used only for the temporal analysis.

The original searches were performed on 22 July 2026. During the peer-review revision, we found that the original export files were incomplete. We therefore repeated all nine searches on 16 September 2026 and exported all non-patent results in several batches.

Before analysis, we checked that the exported records matched the total numbers displayed in SciFinder. The exact search strings and the complete counts for each query are given in Supplementary Table S1.

DOI was used as the main identifier when available. When a DOI was missing, we used a normalized version of the title as a conservative alternative. Title normalization removed differences in capital letters, punctuation, and spacing before comparison.

After removing duplicates across all searches, the dataset contained 6314 unique records from 7222 non-patent search results.

As a reproducibility check, we applied the same deduplication method to the original July export files. This reproduced exactly the previously reported total of 756 unique records. The difference between the original and revised datasets therefore came from the incomplete exports, not from a change in the deduplication method.

Because the nine searches overlap, the number of records from each query cannot be used directly to estimate how common each material family is.

During the revision, we also found another issue with the search design. Query Q6 does not contain an explicit AI or machine-learning term. Records found only by Q6 were therefore excluded from the main AI-explicit bibliometric analysis.

This decision was introduced during the peer-review revision and was not part of the original search design. The complete dataset was still kept for sensitivity analysis and broader context.

We therefore used two datasets in the analysis:

- The complete dataset, containing 6314 unique records;
- The main AI-explicit subset, containing 4066 unique records found by at least one query other than Q6.

A total of 2248 unique records were found only by Q6.

Three additional targeted searches were performed on 9 August 2026. They focused on failed experiments, MatterGen, and out-of-distribution graph neural networks.

These searches were used only to support the methodological discussion. They were not included in the bibliometric counts, temporal analysis, or title-based material-family and task maps.

The results of the nine SciFinder searches are summarized in Table 1.

Table 1. Updated CAS SciFinder search results.

Query	Topic	Total Results	Non-Patent Records	Patents	Unique Records in the Search
Q1	General AI-chalcogenides	2229	2096	133	2093
Q2	TMDCs	1365	1318	47	1316
Q3	Photovoltaic chalcogenides	146	139	7	139

Table 1. Cont.

Query	Topic	Total Results	Non-Patent Records	Patents	Unique Records in the Search
Q4	Phase-change materials	1018	854	164	854
Q5	Thermoelectrics	115	115	0	115
Q6	Property prediction	2345	2317	28	2310
Q7	Graph-based models	40	39	1	39
Q8	Synthesis optimization	103	99	4	98
Q9	XRD/Raman/SEM/TEM/spectroscopy	252	245	7	245
Total retrieval occurrences		7613	7222	391	—

2.3. Screening, Representative-Study Selection, and Title-Coding Quality Control

The literature was analysed in two complementary ways.

First, the 4066-record AI-explicit subset was analysed quantitatively using conservative multi-label rules based on titles. A material family or scientific task was assigned only when it was clearly mentioned in the title. A record could belong to more than one category, and unclear cases were left unassigned.

For this reason, these counts should be considered lower bounds based on explicit title information. They should not be interpreted as exact estimates of how common each family or task is in the full literature.

Second, representative studies were selected for the critical scientific discussion. A study was included when it provided:

- (i) A direct chalcogenide example with a clear AI role and validation method;
- (ii) Evidence supporting the highest validation level assigned to a family/task combination;
- (iii) An important methodological limit or failure mode; or;
- (iv) A broader inorganic example that was useful for interpreting a chalcogenide-specific claim.

Studies in the fourth group were treated as transfer benchmarks and not as direct chalcogenide evidence.

Titles, abstracts, bibliographic information, and publisher information were checked when needed. Full texts were examined when necessary to understand the role of the model, the source of the reference data, the order between prediction and confirmation, the independence of the validation data, or other important details.

The scientific discussion is therefore based on representative studies. It is not a complete full-text analysis of all 6314 records.

To evaluate the reliability of the title-based classification, two independent annotators performed an abstract-level audit. The audit included 120 category decisions: 70 material-family decisions and 50 scientific-task decisions.

For each category, five title-positive records and five negative or borderline records were included. The abstracts were classified as Yes, No, or Unclear.

For material-family coding, the two annotators agreed on 80.0% of the three-class decisions, with Cohen's $\kappa = 0.634$. After removing decisions containing an Unclear label, 58 Yes/No decisions remained. Agreement increased to 94.8%, with $\kappa = 0.881$.

For scientific-task coding, agreement was 72.0%, with $\kappa = 0.463$. After removing decisions containing an Unclear label, 47 Yes/No decisions remained. Agreement was 76.6%, with $\kappa = 0.518$.

The audit showed that material-family classification was generally more reproducible than scientific-task classification.

We therefore refined some task definitions, especially for characterization, synthesis/process, general prediction terms, and atomistic modelling, before producing the final maps.

The audit results were not used to correct or extrapolate the full dataset counts.

The final classification rules and the main difficult cases identified during the audit are reported in Section S3 of the Supplementary Materials, particularly in Tables S3A and S3B.

2.4. Evidence Ladder

Representative claims are classified using a five-level evidence framework. This framework helps distinguish retrospective prediction from prospective validation and repeated closed-loop use.

The evidence level is assigned to a specific claim, not to the paper as a whole. The five evidence levels and their associated validation criteria are summarized in Table 2.

Table 2. Evidence ladder used for the critical synthesis.

Level	Evidence Type	Minimum Evidence Needed	Main Remaining Risk
0	Method transfer or proposal	The method is discussed using a non-chalcogenide example or a proposed workflow.	No direct validation for chalcogenides.
1	Internal retrospective validation	Performance is tested on held-out data or by cross-validation within the same dataset.	Data leakage, prediction of very similar cases, or bias in the reference data.
2	Domain-aware or external validation	The model is tested on a different chemistry, structure, time period, instrument, data source, or another scientifically relevant domain.	Distribution shift may remain, and uncertainty may not be fully calibrated.
3	Prospective computational or experimental validation	The model makes a prediction, ranking, hypothesis, or recommendation before a new calculation, synthesis, process change, or measurement is performed to test it.	The number of tests may be small, and transfer outside the tested domain may remain uncertain.
4	Closed-loop operation	Several cycles of data collection, prediction, decision, and action are repeated, and new results are used to guide the next decision.	Robustness, reproducibility, governance, and human oversight at the system level.

The evidence ladder is not a general score of paper quality. For example, a well-validated Level 1 model may be sufficient for prediction within a limited and well-defined domain. In contrast, a Level 4 system may still include some weaker models or components.

For this reason, we avoid some simple interpretations. A generated candidate should not be considered a discovery until novelty, structure, disorder, stability, and synthesizability have been checked.

One model recommendation that is confirmed prospectively is classified as Level 3, not Level 4. Level 4 requires that the new result is used to guide the next decision.

The presence of experimental data in a paper is also not enough to prove prospective validation. If the model was trained or used only after the experimental results were already known, the evidence is still retrospective.

Similarly, a low prediction error does not prove that a model will work well outside its training domain. The test design must match the conditions in which the model is expected to be used [13,14].

Calibration, applicability-domain analysis, and OOD-aware uncertainty evaluation are related, but they test different aspects of model reliability [15,16,72,73].

2.5. Quantitative Mapping

The main quantitative analysis used the 4066-record AI-explicit subset.

Among these records, 3995 were published between 2015 and 2026. Of these, 3170 records (79.3%) were published between 2023 and 2026, and 2754 records (68.9%) were published between 2024 and 2026. The annual distribution of publications in the AI-explicit subset is shown in Figure 1.

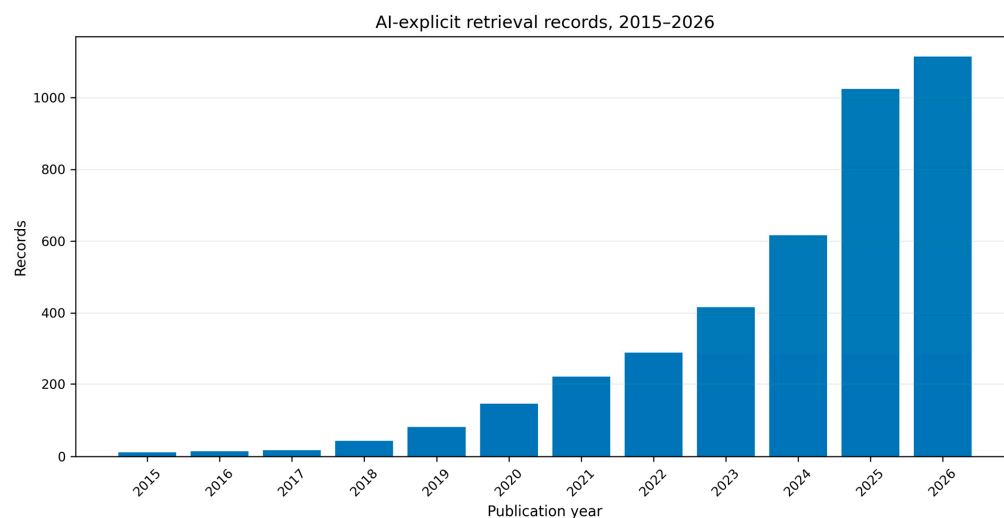


Figure 1. Number of records published each year in the main AI-explicit SciFinder subset between 2015 and 2026 ($n = 3995$). The 2026 value represents only part of the year. Seven records dated 2027 were kept in the complete dataset but were not included in this 2015–2026 analysis.

The complete dataset of 6314 records was also analysed as a sensitivity check.

In this complete dataset, 5589 records were published between 2015 and 2026. Among them, 3942 records (70.5%) were published between 2023 and 2026, and 3365 records (60.2%) were published between 2024 and 2026.

Seven records from the new export had the publication year 2027. These records were kept in the complete dataset, but they were not included in the 2015–2026 temporal analysis.

The previous problem of missing publication years in the thermoelectric export was solved by the complete September re-export. Therefore, no material family is excluded from the revised temporal analysis because of missing publication years.

Using simple multi-label title rules, we identified 708 TMDC records, 128 thermoelectric or thermal-transport chalcogenide records, 96 photovoltaic chalcogenide records, 48 phase-change chalcogenide records, 16 chalcogenide-perovskite records, 11 chalcogenide-glass records, and 6 high-entropy-chalcogenide records.

These numbers are minimum estimates based only on information clearly stated in the titles. Some records may belong to more than one category.

The final task counts were 116 records for property prediction/classification, 36 for synthesis/process optimization, 43 for screening/discovery, 103 for atomistic/MLIP studies, and 38 for characterization/interpretation.

The family–task table (Supplementary Table S3C) is descriptive only. A zero means that no title matched both the predefined family and task rules. It does not mean that this topic is absent from the literature. The distribution of records across the main chalcogenide families is shown in Figure 2.

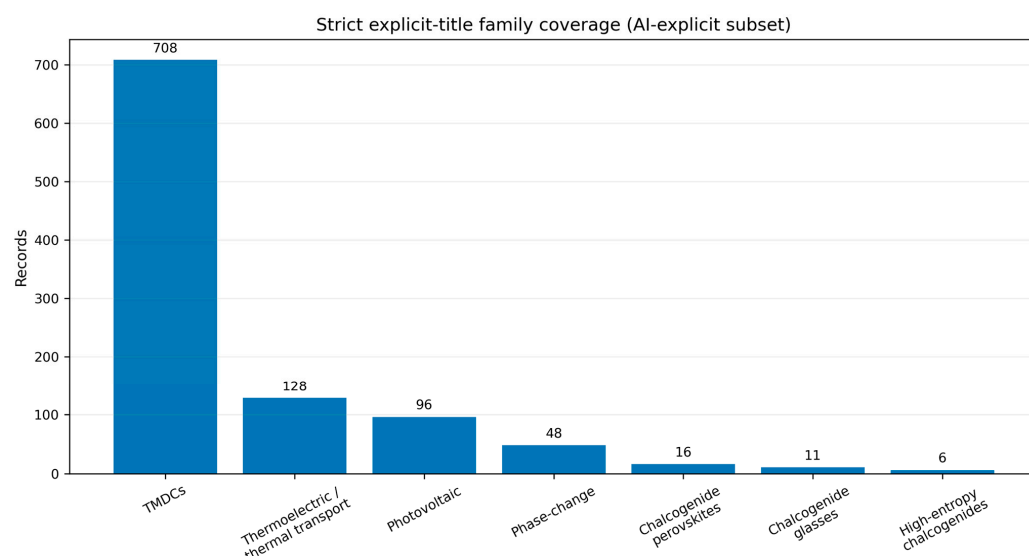


Figure 2. Minimum numbers of records for the main chalcogenide families in the 4066-record AI-explicit subset, based only on clear information in the titles. A record can belong to more than one category. If a material family is not mentioned in the title, it may still be discussed in the full article.

The distribution of records across material families and scientific tasks is shown in Figure 3.

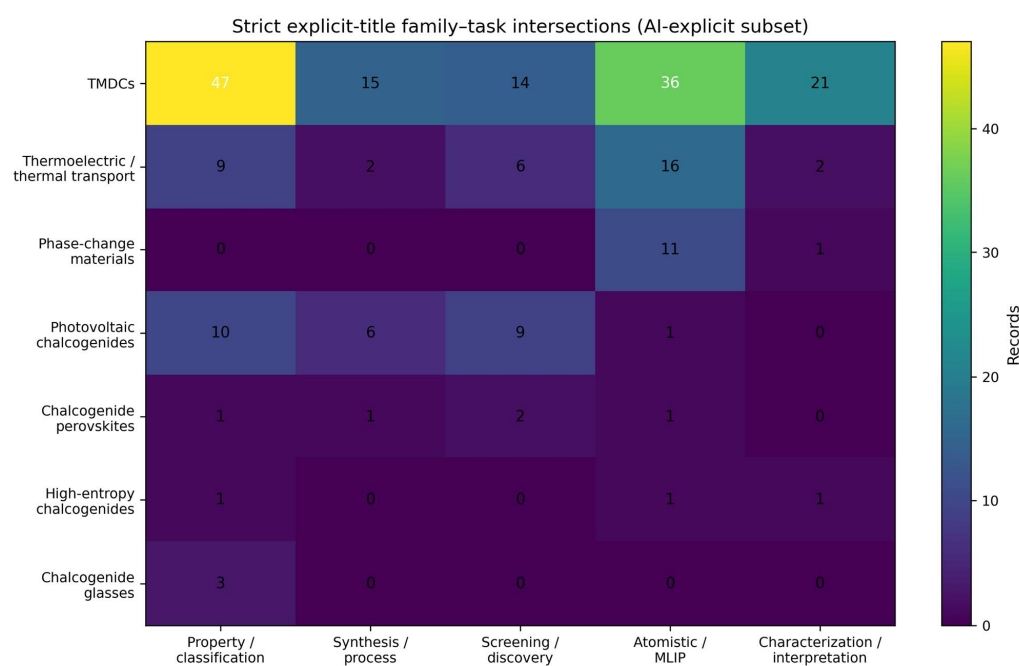


Figure 3. Number of records that match both a major chalcogenide family and a scientific task in the 4066-record AI-explicit subset. The counts are minimum estimates based on the title rules described in Supplementary Section S3. A zero means that no title matched both categories. It does not mean that this topic is absent from the literature.

2.6. Limitations and Reproducibility

The quantitative analysis has several limitations.

First, SciFinder was used as the main bibliographic database. The results may therefore be different from those obtained with Scopus, Web of Science, Dimensions, or other databases. The nine searches also overlap and should not be considered as separate material categories.

Second, Q6 was designed as a broad property-prediction search, but it does not contain an explicit AI or machine-learning term. During the peer-review revision, records found only by Q6 were therefore excluded from the main AI-explicit analysis.

Because this decision was made during the revision and was not part of the original search design, we report both datasets: the main AI-explicit subset of 4066 records and the complete dataset of 6314 records.

Third, the title-based classification is intentionally conservative. If a material family or scientific task is mentioned only in the abstract or full text, it may not be identified. Some material families are also easier to recognize from titles because they have more specific names or chemical formulas.

The abstract-level annotation audit was used to evaluate this limitation, but the audit results were not used to correct the final record counts.

Fourth, the detailed scientific discussion is based on representative studies and is not an exhaustive analysis of all publications. The highest validation levels identified should therefore not be interpreted as representing the frequency or average quality of evidence in the complete literature.

Fifth, 2026 is only a partial publication year. In addition, seven records dated 2027 were not included in the predefined 2015–2026 temporal analysis.

To improve reproducibility, we provide a machine-readable dataset with one row for each of the 6314 unique records in the complete dataset.

The dataset includes the record identifier, DOI, publication year, snapshot date, query labels, membership in the AI-explicit subset, final material-family and task labels, and a SHA-256 hash of the normalized title used when DOI information was not available.

Raw CAS record information is not shared when reuse is restricted by the database licence.

The exact search strings, deduplication rules, classification rules, annotation-audit method, and data used to create the figures are provided in the Supplementary Materials.

3. Why Chalcogenides Stress-Test Materials AI

3.1. A Family Label Hides Different Learning Problems

The term “chalcogenide” includes several very different machine-learning problems. A layered TMDC described by a relaxed crystal structure, an amorphous Ge–Sb–Te network during crystallization, a multinary photovoltaic material with antisite disorder, and a thermoelectric material affected by texture and processing cannot be treated as the same type of composition-property problem.

The type of data, and therefore the most useful representation, changes from one material family to another.

TMDCs have relatively many well-defined crystalline structures. However, their properties can strongly depend on the number of layers, stacking, strain, moiré patterns, vacancies, grain boundaries, substrates, and interfaces. Monolayer MoS₂ is a well-known example: changing from bulk to monolayer changes the electronic structure from an indirect to a direct band gap [3,67].

Recent AI studies also consider layer-dependent heterostructures, defects, atomistic potentials, thermal transport, and spectroscopic or imaging analysis [3,4,32,34–36,38,40–43,46,47,51,56,74–87]. The main validation question is therefore not only whether the model includes the right composition. It is also whether the training data include the important structural and environmental conditions.

Chalcopyrites and kesterites have different difficulties. Their photovoltaic properties depend on multinary composition, cation disorder, secondary phases, interfaces, and defects.

Machine learning has been used for band-gap prediction, high-throughput screening, device-performance analysis, and defect or process optimization [1,17–22,24,25,88–94].

Some studies use simulated device data instead of direct experimental measurements. These studies can be useful for understanding which parameters are important, but they should not be considered equivalent to the discovery of a new absorber material or to experimental validation of a processing method.

Thermoelectric chalcogenides combine electronic and thermal transport. Their performance is commonly described by the dimensionless figure of merit (ZT), which depends on the Seebeck coefficient, electrical conductivity, temperature, and electronic and lattice thermal conductivity.

Composition, carrier concentration, defects, grain boundaries, texture, and processing can affect several of these properties at the same time.

For this reason, a model trained on measurements obtained at different temperatures or under different processing conditions may learn mixed or misleading relationships. The most useful studies therefore connect model predictions with controlled synthesis, characterization, or transport calculations [9,48,50,52–55,95–98].

Phase-change chalcogenides are strongly affected by non-equilibrium processes. Ge–Sb–Te alloys need to be described in crystalline, liquid, supercooled, amorphous, interfacial, and transition states.

Datasets containing only equilibrium crystal properties are not sufficient for this problem.

Machine-learned interatomic potentials (MLIPs) are especially useful here because they allow larger and longer atomistic simulations than standard first-principles molecular dynamics, while still being connected to DFT reference data within the training domain [5,10,39,44,45].

Their reliability depends on the variety of configurations used for training, validation across different phases, and the ability to detect when the model is used outside its training domain.

Chalcogenide perovskites, glasses, and high-entropy systems still have relatively limited datasets. Studies on perovskites have explored composition and band-gap prediction [2,6,21,90,99,100]. More recent work also includes glass-property prediction and high-entropy thermoelectric materials [7,11,39,48,101–103].

These material families are especially challenging for standard machine-learning representations. Amorphous materials do not have one unique periodic crystal structure, and high-entropy materials contain many different local atomic environments instead of one ordered structure.

Good performance on large datasets of crystalline materials should therefore not be assumed to transfer directly to these systems.

Table 3 summarizes these different data situations and the main validation risks.

3.2. Defects, Disorder, and Processing Are Important Variables

Many materials databases mainly contain relaxed, periodic, and near-equilibrium structures because these data are easier to standardize and reuse.

For chalcogenides, however, the variables that are often missing from databases can be very important for the final properties.

For example, sulfur vacancies can change the electronic structure and reactivity of TMDCs. Antisite defects and interfaces can affect non-radiative recombination in kesterites. Grain boundaries and texture can influence thermoelectric transport. In phase-change materials and glasses, the amorphous structure strongly affects the material behavior [4,10,18,39,44,45,76,77,81,87,104].

A model that ignores these effects may predict the average behavior of a composition correctly, but it may fail to explain differences between individual samples.

This has important consequences for model design. The representation used by the model should match the scientific problem.

Composition descriptors may be sufficient for screening within a limited alloy family. However, studies of defects, interfaces, Raman maps, crystallization, or other complex processes need inputs that include structure, images, spectra, or atomic trajectories.

Experimental metadata are also important. Information such as temperature, atmosphere, precursor type, annealing history, instrument settings, and measurement uncertainty should be kept when these factors can affect the measured property.

Table 3. Main data characteristics and validation risks for different chalcogenide families.

Material Family	Important Variables	Common Data Used	Main AI Risk	Representative Refs.
TMDCs	Layer number, stacking, strain, vacancies, interfaces	DFT structures, microscopy and Raman maps, MLIP trajectories	Similar structures may appear in both training and test data; defects and substrates may be poorly represented	[3,4,34,35,40,47,75–79,83,84,87]
Chalcopyrites and kesterites	Multinary composition, antisite defects, secondary phases, interfaces	DFT and device simulations, experimental process and device data	Simulated device results may be confused with direct materials validation	[1,17–22,24,25,88,89,91–94]
Thermoelectric chalcogenides	Temperature, carrier concentration, microstructure, texture, processing	Calculated transport data and experimental measurements from different conditions	Composition, temperature, and processing effects may be mixed together	[9,48,50,52–55,95–98]
Phase-change materials	Liquid, amorphous and crystalline states, kinetics, interfaces	DFT trajectories and MLIP molecular-dynamics simulations	The model may fail during rare events or for phases not well represented in training	[5,10,33,39,44,45,102,105]
Chalcogenide perovskites	Structural distortions, stability, defects, synthesis conditions	Mostly calculated data or small experimental datasets	Limited data and uncertainty about whether predicted materials can be synthesized	[2,6,21,90,99,100]
Glasses	Network structure, composition, thermal history	Small composition-property datasets or generated structures	No unique structural representation and limited external validation	[7,11,39,102,103]
High-entropy chalcogenides	Local chemical order and different atomic configurations	Limited experimental data across very large composition spaces	High risk of prediction outside the training domain	[48,101]

3.3. Large Datasets Do Not Always Mean Reliable Real-World Use

Large computational datasets can make crystal-property prediction look more advanced than it really is. A large dataset does not automatically guarantee independent data, good coverage of the important domains, or direct relevance to experiments.

For example, very similar crystal structures may appear in both the training and test sets. DFT calculations may also contain systematic errors that depend on the method used. In addition, relaxed crystal structures do not show whether a material can actually form under real experimental conditions.

In MatBench, five types of out-of-distribution (OOD) tests showed that model performance decreased when the test data differed from the training data. Models that performed best on standard tests were not always the most reliable under distribution shift [13].

A temporal study using Materials Project data showed a similar decrease in performance when models trained on older data were tested on newer data. Measures based on feature space and model committees helped identify unfamiliar regions. Adding a small amount of new targeted data also improved model performance [14].

For materials discovery, a low mean absolute error from a random train-test split is therefore only one part of the evidence.

The test design should clearly show what is different between the training and test data. Depending on the application, this can include leaving out a complete chemistry, separating structural prototypes, grouping compositions, using a temporal split, or testing on different instruments or laboratories.

It is also useful to report model uncertainty or a measure of distance from the training domain. Failure cases should be shown when possible instead of being hidden in a single average error.

These checks become especially important when a model is used on materials or conditions that are very different from its training data.

4. Data, Representations, and Models

4.1. Database Choice Is an Important Scientific Decision

Large materials databases such as the Materials Project, OQMD, AFLOW, NOMAD, JARVIS, and C2DB provide important computational data for materials-informatics studies [68–70,106–108].

However, these databases are not equivalent. They use different calculation methods, contain different properties, and have different levels of information about data origin, versioning, and intended use.

The important question is therefore not only which database is the largest. We also need to ask whether the calculations, data selection, and chemical coverage are suitable for the scientific problem and for the conditions in which the model will be used.

The Materials Project and OQMD provide broad data on inorganic crystals and thermodynamic properties. AFLOW focuses on automated high-throughput calculations and descriptors. NOMAD stores a wide range of raw computational results and is useful for tracking data origin and comparing workflows. JARVIS provides datasets and benchmarks for several properties, while C2DB focuses on monolayer materials.

However, none of these databases fully describes all important aspects of chalcogenides, such as synthesis history, defects, amorphous structures, spectroscopy, microscopy, and experimental uncertainty.

The main data resources and their limitations are summarized in Table 4.

Table 4. Main data resources for chalcogenide AI. The usefulness of each database depends on the dataset version, data filters, target property, and chemical coverage.

Resource	Main Content	Useful Applications for Chalcogenides	Main Limitation	Ref.
Materials Project	Standardized DFT data for crystal structures and properties	Crystal structures, formation energies, model pretraining, and benchmarks	Mainly equilibrium calculations; band gaps depend on the DFT method; limited information on processing conditions	[68]
OQMD	Large DFT database including reported and hypothetical compounds	Formation-energy screening and exploration of chemical space	Mostly computational data with limited experimental information	[69]
AFLOW	Automated calculations and calculated descriptors	Descriptor-based screening and comparison of computational workflows	Available properties and workflows vary between subsets	[70]
NOMAD	Raw and processed calculations from different simulation codes	Data-provenance studies, comparison of workflows, and model pretraining	Data are heterogeneous and need careful normalization	[106]
JARVIS	Curated computational datasets for several properties	Benchmarking electronic, optical, mechanical, and topological properties	Some datasets are smaller or have a different distribution from Materials Project or OQMD	[107]
C2DB	Standardized calculations for two-dimensional materials	Screening of TMDCs and other monolayer materials	Limited to two-dimensional structures and specific computational workflows	[108]

FAIR principles improve how data can be found, accessed, combined, and reused [71]. However, FAIR data are not automatically ready for machine learning.

For reliable scientific use, other information is also important, including units, clear target definitions, uncertainty, negative results, data origin, and stable dataset versions.

Failed or poor synthesis conditions can also be very useful because they help define the limits of the process window. The prospective use of failed experiments shown by Raccuglia et al. in a non-chalcogenide synthesis study [109] is therefore also relevant to chalcogenide growth, deposition, and annealing studies.

4.2. Descriptor-Based Models Remain Useful Baselines

Random forests, support-vector machines, gradient boosting, and simple neural networks can work well for small or medium-sized tabular datasets when the descriptors correctly represent the important chemical or process variables.

These methods have several advantages. They usually need less data, train quickly, and make it easier to study which input features influence the prediction. They have been used for kesterite band-gap prediction, CIGS and CZTSSe performance, thermoelectric screening, glass properties, and synthesis or process optimization [1,9,11,17,19,22–24,26,52,53,55–58,79,94,95,98,103,110].

These models are not only old reference methods. For small datasets, a well-regularized tree-based or kernel model can sometimes perform better than a deep neural network. They can also make problems in the data easier to detect.

Simple linear or physically motivated models can also be useful when they match the physics of the problem.

Model selection should be included inside the validation procedure. Otherwise, repeatedly checking the test set during model development can make the final performance appear better than it really is.

Feature-importance methods also need to be interpreted carefully. SHAP values and similar methods show how the trained model uses its input variables. However, they do not prove that one variable causes a change in another.

When input variables are strongly correlated, feature importance can also become unstable. This is important for thermoelectric and photovoltaic datasets because composition and processing variables often change together.

A physical interpretation is therefore stronger when it is also supported by an experiment, a prospective test, or an independent physical mechanism [81,95].

The vanadium-selenite study gives a useful example from another material system. The model was trained using both successful and failed hydrothermal synthesis experiments. It was then tested prospectively with new commercial organic building blocks that had not been used before.

The conditions predicted to produce new inorganic products were successful in 89% of the reported tests [109].

This study does not demonstrate general autonomous discovery for chalcogenides. However, it shows an important principle: relatively simple reaction descriptors can be scientifically useful when failed experiments are kept in the dataset and model recommendations are tested prospectively.

4.3. Graph Neural Networks: Better Representations, but Limited Transfer

Graph neural networks (GNNs) represent atoms as nodes and local interactions as edges. This allows the model to learn a representation of the crystal structure instead of using only predefined descriptors.

The Crystal Graph Convolutional Neural Network (CGCNN) introduced direct prediction of material properties from crystal structures [27]. MEGNet added global state variables and transfer between different properties [28]. ALIGNN included bond-angle

information using line graphs [29]. M3GNet and CHGNet extended graph models toward energy and force prediction and broad pretrained interatomic potentials [30,31].

Graph models have clearly improved the representation of materials. However, the use of a GNN does not automatically mean that the model can transfer to new conditions.

For example, a GNN trained on relaxed periodic crystals may not work well for an amorphous phase or an interface, even if the same chemical elements are present in the training data. Adding angular or charge-related information also does not automatically guarantee transfer to a new oxidation state, defect, or reaction pathway.

Studies using five types of out-of-distribution (OOD) tests showed that the ranking of GNN models can change when the test data differ from the training data. In some cases, older GNN models were more robust than models that performed best on standard tests [13].

A temporal study comparing Materials Project data from MP18 and MP21 also showed lower performance on newer materials. Feature-space and query-by-committee methods helped identify unfamiliar regions and select useful new data [14].

In practice, model robustness depends not only on the graph architecture, but also on how the dataset is built, how the training and test data are separated, and what objective is used during training.

For chalcogenides, graph models are especially useful when reliable crystal structures are available and when the target property depends on local atomic coordination.

For simple composition-based screening, however, a graph model may provide little additional benefit.

In contrast, problems such as defect migration, amorphous structures, phase transitions, and synthesis history contain information that is not present in a single relaxed crystal structure.

Fine-tuning a broad pretrained model can reduce the amount of task-specific data needed, but the final model still needs to be tested to define the domain in which it can be used reliably [31,47].

4.4. Machine-Learned Interatomic Potentials

Machine-learned interatomic potentials (MLIPs) learn energies and forces from quantum-mechanical calculations. They can then be used to speed up molecular dynamics, structural relaxation, and sampling of atomic configurations.

MLIPs are especially useful for chalcogenides because many important processes involve large systems, long simulations, structural disorder, defects, interfaces, or phase transformations. Examples include WSe₂ thermal transport, defects and interfaces in MoS₂ and MoS₂/WS₂, amorphous MoS₂, metal–chalcogenide interdiffusion, GeTe/Sb₂Te₃ transport, and Ge–Sb–Te crystallization [4,5,10,32–47].

A random test set taken from the same trajectory or configuration pool is only a limited validation. The type of validation should depend on how the potential will be used.

Useful tests can include energy and force errors for different phases or thermodynamic states, equations of state, phonons, diffusion, defect energies, structural distributions, phase stability, transport properties, crystallization kinetics, and detection of configurations outside the training domain.

The quantum-mechanical reference method, the way the configurations were generated, and the relationship between the training and validation structures should therefore be reported together with the numerical errors.

The type of potential is also an important modelling choice. The interaction range, cutoff, descriptors, composition of the training database, and fitting method can strongly affect the results.

A study comparing different potential forms for the graphene-to-diamane transformation showed that models can reproduce important equilibrium properties but still predict very different transformation behavior if the interaction form or cutoff does not describe the relevant pathway correctly [111].

More generally, MLIPs should be validated using physical properties in addition to numerical error values [112]. Studies on defects and fracture also show that the predicted behavior can depend on the training database, ML method, and cutoff radius, even when standard error values look acceptable [113].

For studies of transformations, validation should therefore include at least one physical quantity directly related to the process of interest.

Depending on the application, this can include phase ordering, transition temperature, diffusivity, nucleation, crystal-growth kinetics, defect migration, interfacial resistance, or another property related to the transformation.

When possible, comparing different potential families, architectures, or cutoff values gives stronger evidence than agreement with equilibrium properties alone.

The methodological studies above are not direct chalcogenide examples. They are used here to define useful validation requirements for chalcogenide MLIPs.

Ge–Sb–Te provides a useful example. The early GST225 Gaussian approximation potential was tested using quantum-mechanical configurations that were not used for training and using experimentally measurable structural properties [5].

A later GST225 neural-network potential was validated against DFT calculations and physical properties relevant to phase-change processes [10].

A later study on viscosity and dynamical heterogeneity used the same potential and extended its physical validation, rather than training a completely new model [44].

More recent work on Ge-rich Ge–Sb–Te tested compositions that were not included in the training database before using the potential for large-scale phase-separation and crystallization simulations [45].

These studies are convincing not simply because they use large numbers of atoms or long simulation times. Their strength comes from validation against quantum-mechanical calculations or experimental properties that are relevant to the physical problem.

However, the conclusions still depend on the reference electronic-structure method, the configurations included in the training data, and the form of the potential.

Transfer to new compositions, defects, interfaces, or transformation pathways should therefore be tested directly and should not be assumed only because the average energy or force errors are low.

4.5. Active Learning, Bayesian Optimization, Out-of-Distribution Evaluation, and Uncertainty

Active learning and Bayesian optimization are related methods, but they have different goals.

Active learning selects new data that can improve the model efficiently. Bayesian optimization selects new tests or experiments to optimize a target while considering uncertainty.

In computational studies, the method may select a structure with high uncertainty for a more accurate calculation. In experimental studies, it may select the next composition or processing condition by balancing exploration of new conditions and use of already promising conditions.

Chalcogenide applications include power-factor prediction, surface-supported clusters, multinary Chevrel phases, thermoelectric processing and printing, MoS₂ growth, and control of SnSe or SnS morphology [48–59].

The strongest evidence comes from studies in which the model recommends a new action before the result is known, and the new calculation or experiment is then performed and reported.

Retrospective replay is still useful for developing methods, but it provides weaker evidence. This is especially true when the acquisition method is tested using a large existing dataset that would not have been available during a real experiment.

Uncertainty should be tested under the same type of distribution shift expected during real use.

A random train-test split is useful as a basic interpolation test, but it is usually not enough to show that a model can transfer to new materials.

Structure-based out-of-distribution (OOD) tests show that model performance can decrease strongly when training and test materials are separated by structural or compositional similarity instead of randomly [13]. Temporal tests also show lower performance on materials added later and show that model rankings obtained from random splits can change under distribution shift [14].

MatUQ recently introduced SOAP-LOCO, or Smooth Overlap of Atomic Positions Leave-One-Cluster-Out [72]. In this method, structures with similar local environments are grouped together, and complete groups are left out of the training set.

Tests on several materials datasets, GNN models, and uncertainty methods show that predictive accuracy and uncertainty quality are different properties. They also show that the best uncertainty method depends on the target, the model architecture, the amount of data, and the type of OOD shift.

Methods such as deep evidential regression, Monte Carlo dropout, or combinations of these methods should therefore be considered as possible uncertainty approaches, not as methods that are always well calibrated. Other molecular-learning studies also show that evidential and ensemble uncertainties may need calibration before they are used for decisions [73].

For chalcogenide property prediction and computational screening, a random split should therefore be used mainly as an interpolation reference. At least one OOD split related to the intended application should also be included.

If the goal is transfer to new compositions or structure families, a chemistry-aware or prototype-aware split is useful.

For structure-based GNNs, SOAP-LOCO or another local-environment clustering method is useful when the goal is transfer to unfamiliar atomic environments.

A temporal split is also useful when the model will be applied to materials added after the model was developed.

There is no single split that is always “OOD”. The variable that is left out should match the type of novelty claimed in the study.

Experimental applications need different grouping variables.

Spectroscopy, microscopy, and diffraction models should be tested on different specimens, synthesis batches, instruments, and, when important for two-dimensional materials, different substrates. Randomly separating pixels or spectra is usually not enough.

Process and synthesis models should be tested on different batches, reactors, substrates, experimental campaigns, or time periods. Prospective testing should be used when possible.

For MLIPs, important OOD tests can include new compositions, phases, defects, interfaces, thermodynamic states, or transformation pathways. A random split of snapshots from the same trajectories gives only limited evidence of transfer.

The uncertainty metric should also match the scientific decision.

At least three questions should be considered separately.

First, calibration asks whether the predicted uncertainty agrees with what happens in practice. For example, prediction intervals should contain the true value as often as expected. Coverage should be reported together with interval width or sharpness. Conformal or other distribution-free recalibration methods can be useful when the error distribution is difficult to model [15].

Second, error ranking asks whether predictions with higher uncertainty also have larger errors. This can be tested using the correlation between uncertainty and absolute error, or by comparing errors for groups with different uncertainty levels. However, good error ranking does not prove that the uncertainty is well calibrated.

Third, decision utility asks whether uncertainty helps make better decisions. For example, rejecting the most uncertain predictions may reduce the error of the remaining predictions, or an acquisition strategy may reach a target using fewer calculations or experiments.

Graph-kernel and Gaussian-process methods provide another example of how calibrated uncertainty can be combined with material-level and atomic-level OOD detection [16].

Recent OOD and uncertainty benchmarks show why all these questions should not be reduced to one score. MatUQ combines evidential and dropout-based uncertainty measures and evaluates them under specific OOD shifts [72]. However, its results do not show that one uncertainty method is always the best.

Uncertainty methods should therefore be tested for the type of distribution shift expected in the intended application.

In this review, we consider an uncertainty analysis credible when it reports predictive error on a test split that matches real use, checks or recalibrates prediction intervals, tests whether uncertainty follows the real prediction error, and clearly explains what happens when uncertainty becomes large.

The next action may be to collect more data, perform a more accurate calculation, reject the prediction, or ask for human review.

Uncertainty is most useful when it changes a scientific decision, rather than being reported only as an additional model output.

4.6. Generative and Foundation-Model Approaches

Generative models change the usual design process because they can propose new compositions or structures for a given target.

MatterGen uses diffusion modelling to generate inorganic material candidates. In the original study, one generated structure was synthesized, and its measured property was within 20% of the target value [61].

A later critique argued that the highlighted Ta–Cr–O example was not clearly new and could correspond to a known disordered compound already represented in the training data [62].

This discussion is useful because it shows that several questions must be considered separately: Is the material really new? Is it similar to something in the training data? Is structural disorder correctly described? Is the material stable? And can it actually be synthesized?

Autonomous laboratories raise a similar problem.

The original A-Lab study showed repeated cycles of planning, synthesis, X-ray diffraction analysis, and selection of the next experiment in a general inorganic-material setting [60]. We therefore classify this operational loop as Level 4 evidence for repeated closed-loop operation.

However, the 2026 Author Correction clarified that the word “new” in the original paper meant new relative to the prediction platform and did not necessarily mean new to science [60].

Leeman et al. also showed that structural disorder and similarity to known disordered compounds can make novelty claims difficult to interpret [114].

For this reason, we evaluate operational autonomy and material novelty separately.

For the chalcogenide studies reviewed here, generative AI is better considered as a tool for proposing candidates than as proof of autonomous materials discovery.

A proposed material still needs to be checked for similarity to the training or reference data, structural disorder, thermodynamic and kinetic stability, uncertainty, and experimental feasibility.

Foundation models raise a similar issue. Large-scale pretraining can be useful, but transfer to new chemical or structural chalcogenide problems must be demonstrated. It should not be assumed only because the model is large.

Table 5 compares these different model types, their possible uses, and the minimum validation needed.

Table 5. Main model types, suitable uses, and minimum validation.

Model Type	Suitable Use	Minimum Validation Needed	Main Risk for Chalcogenide	Representative Refs.
Linear/tree/kernel models	Small tabular datasets for composition, processing, or spectra	Compare with a simple physical or statistical baseline; use nested validation; use chemistry- or batch-separated tests when transfer is claimed	Mixed effects between variables, correlated descriptors, and incomplete experimental metadata	[1,9,11,17,53,55,79,95,103,110]
CNN/deep spectral or image models	Raman, microscopy, diffraction, and combined maps	Test on different specimens, batches, instruments, and, when important, substrates; check preprocessing and data augmentation	Data leakage between pixels or maps; changes between instruments or substrates	[75–77,79,83–87,105,115–118]
Crystal GNNs	Structure-based equilibrium-property prediction and screening	Compare with a descriptor baseline; use chemistry-, prototype-, or SOAP-LOCO-based OOD tests; keep random split only as an interpolation test; check uncertainty under the expected distribution shift	Bias toward relaxed structures, leakage between similar structures, and poor transfer to new chemistries or prototypes	[13–16,27–29,72]
Pretrained GNN/universal MLIP	Energies, forces, structural relaxation, and molecular dynamics	Test or fine-tune for the target domain; test different phases, defects, and interfaces; evaluate applicability domain or uncertainty	Direct use on configurations that are not represented in the training data	[15,16,30,31,47]
Specialized MLIP	Dynamics and transformations within a defined chemical system	Test energies and forces on independent data from relevant phases; compare potential type, architecture, or cutoff when possible; validate at least one physical property related to the transformation	Failure for rare events, wrong phase order or kinetics, dependence on interaction range, and limited transfer between compositions	[4,5,10,32–46,111–113]
Active learning/Bayesian optimization	Expensive calculations or limited experimental/computational search spaces	Make the recommendation before the result is known; compare with random or expert selection; report sample efficiency or regret; check that uncertainty is suitable for selecting the next test	Biased selection, poorly calibrated uncertainty, or retrospective analysis presented as prospective optimization	[48–59,119]
Generative models	Candidate generation and inverse design under defined constraints	Check similarity with the training set and real novelty; include disorder in structural comparison; test stability, kinetics, and feasibility; independently test a prospectively selected candidate	Bias toward ordered crystals, rediscovery of known or disordered materials, and difficulty of experimental synthesis	[60–62,114]

5. Evidence by Scientific Task

Organizing the discussion by scientific task is more useful than simply listing machine-learning methods.

The same random forest, graph neural network, or machine-learned potential can support very different conclusions. This depends on the property being predicted, the information available when the prediction is made, the way the training and test data are separated, the reference method used for comparison, and the validation procedure.

For each scientific task, we therefore examine three main questions: what is predicted, what is considered new, and how the prediction is tested afterwards.

5.1. Electronic and Optical Property Prediction

Band-gap prediction is one of the most common uses of AI in chalcogenide research. Kesterites, transition-metal dichalcogenides, chalcogenide perovskites, Janus structures, and multinary absorbers have been studied using descriptor models, deep atomic representations, graph models, and combined DFT-ML approaches [1–3,6,21,23,74,89,90,99,120–123].

These methods can reduce the number of expensive calculations and help identify trends between related materials. However, the quality of the prediction depends on the quality and definition of the training data.

For example, a model trained on band gaps calculated with semilocal DFT mainly learns to reproduce that computational method. Good agreement with a test set does not automatically mean good agreement with experimental optical band gaps measured on real samples.

The strongest computational studies usually define a limited chemical space, compare the model with simple baselines, and then check the most promising candidates using more accurate calculations.

Experimental measurements provide a higher level of evidence. Without experimental confirmation, the word “discovery” usually means computational selection of promising candidates rather than proof of material performance.

Structure prediction for ternary chalcogenides [123] also has an additional challenge. A predicted structure is useful only if its stability or possibility of formation is also considered.

Photovoltaic-device studies need another important distinction.

Models trained on experimental CIGS or CZTSSe data can help identify relationships between composition, processing, interfaces, and device performance [17–22,24,25,88,94].

In contrast, models trained on SCAPS-1D simulation results mainly learn to reproduce the simulator [90–93,121]. These surrogate models can make sensitivity studies faster, but they do not prove that the simulator assumptions are correct or that the predicted optimum can be achieved experimentally.

Training and test data generated by the same deterministic simulator can also be very similar. For this reason, the test set should be separated by parameter range or device architecture when possible. The predicted optimum should then be checked using an independent simulation and, when possible, experiments.

AI has also been used to predict other chalcogenide properties, including quantum yield, refractive index, glass-transition temperature, density, and nonlinear-optical performance [7,11,82,102,103,124].

For these properties, sample preparation and measurement conditions can be as important as composition.

A model may perform well on data from one laboratory but fail for samples prepared using another method. Tests on new compositions or independent datasets, such as those used in some glass-property studies [103], therefore provide stronger evidence of transfer than a low error from a random train-test split alone.

5.2. Thermoelectric Discovery and Process Optimization

Thermoelectric performance depends on several properties at the same time. The dimensionless figure of merit is

$$ZT = S^2 \sigma T / (\kappa_e + \kappa_l)$$

where S is the Seebeck coefficient, σ is the electrical conductivity, and κ_e and κ_l are the electronic and lattice parts of the thermal conductivity.

A high power factor, $S^2 \sigma$, or a low lattice thermal conductivity can be useful, but these two quantities alone do not guarantee a high ZT . Carrier concentration, bipolar transport, anisotropy, microstructure, contact resistance, and mechanical properties can also change the final ranking of materials.

Computational and data-driven studies have predicted power factor, thermal conductivity, and DFT-based transport properties for several chalcogenide families [9,12,42,43,53,95–98,124].

These results are easier to interpret when the target property is clearly defined and when electronic and thermal properties are later combined in a physically consistent way.

Public datasets can be useful for screening, but important differences in temperature, crystal direction, sample density, and measurement method should be kept in the data rather than mixed together as if all measurements were equivalent [95].

Microstructure is a good example of why a low predicted lattice thermal conductivity cannot be considered independently of the real sample.

Grain boundaries can strongly affect phonon transport. However, the size of this effect and its dependence on boundary angle can change with grain-boundary structure, dislocation content, and material type.

For monolayer WSe₂, nonequilibrium molecular-dynamics simulations showed that thermal conductivity decreased from $23.7 \pm 0.96 \text{ W m}^{-1} \text{ K}^{-1}$ in the pristine material to about $10.1 \text{ W m}^{-1} \text{ K}^{-1}$ at a misorientation angle of 21.79° . This corresponds to a decrease of more than 55% [125].

In the same study, the calculated Kapitza resistance increased from about 0.31 to $5.08 \text{ m}^2 \text{ K GW}^{-1}$. These results show that grain-boundary density and geometry can strongly change lattice thermal transport.

However, the effect of the misorientation angle is not always monotonic.

For 517 grain boundaries in monolayer MoS₂, nonequilibrium molecular-dynamics calculations showed a U-shaped variation of thermal conductance for misorientation angles between about 5° and 52° [126].

The misorientation angle can therefore be a useful description of microstructure, but it is not a universal measure of thermal resistance.

Experimental evidence is also available for thermoelectric SnTe.

Frequency-domain thermorefectance combined with EBSD showed local decreases of about $0.5\text{--}1 \text{ W m}^{-1} \text{ K}^{-1}$ in thermal conductivity near individual grain boundaries, compared with a bulk value of about $8 \text{ W m}^{-1} \text{ K}^{-1}$ [127].

The measured grain-boundary thermal resistance increased from about $51 \pm 2 \text{ m}^2 \text{ K GW}^{-1}$ for misorientation angles below 10° to $85 \pm 20 \text{ m}^2 \text{ K GW}^{-1}$ for angles above 48° .

A model using the measured grain-boundary resistance also reproduced the effect of grain size on the thermal conductivity of polycrystalline SnTe.

These studies show a direct link between microstructure and the thermal-conductivity term in ZT .

Reducing lattice thermal conductivity through grain-boundary scattering can be useful. However, grain boundaries can also affect electrical transport. For this reason, a decrease in total thermal conductivity does not automatically mean an increase in ZT .

Grain size, grain-boundary type, texture, and processing history should therefore be included as physical descriptors or grouping variables when machine-learning models are used to predict thermoelectric transport in real samples.

Thermoelectric research also includes several prospective examples.

High-throughput synthesis was combined with ML-assisted image analysis for Cu–Sn–S compounds [50]. Because the synthesis and original imaging were performed before the ML analysis, this study is classified as Level 3 only for analysis-guided phase identification and the later follow-up experiments, not for prospective selection of the synthesis conditions.

Active learning has also been used to select power-factor calculations [53] and high-entropy thermoelectric experiments [48].

For Ag₂Se-based devices and composites, models were used to guide flash-sintering conditions, material combinations, and three-dimensional printing. The suggested conditions were then tested by fabrication and characterization [52,54,55].

These studies reach Level 3 when the model recommendation is made before the confirming experiment and the experimental result is reported.

Level 4 requires repeated cycles in which each new result is used to update the model or guide the next decision.

5.3. Phase-Change Materials and Chalcogenide Glasses

Phase-change materials are a good example of how AI can extend a well-defined physical calculation.

The important atomic configurations include liquid, supercooled, amorphous, and crystalline states. Nucleation and crystal growth also require large systems and long simulation times, which are difficult to study with standard first-principles molecular dynamics.

Specialized MLIPs for Ge₂Sb₂Te₅ and related Ge–Sb–Te alloys have made it possible to study crystallization, viscosity, dynamical heterogeneity, thermal transport, and phase separation at much larger scales [5,10,33,44,45].

The main advantage of these MLIPs is not only faster prediction. They also make simulations possible that would otherwise be too expensive.

However, large simulation size and long simulation time do not prove that the model is accurate. The strongest evidence comes from comparisons with quantum-mechanical calculations and physical properties for the relevant states.

For GST225, the early Gaussian approximation potential was tested using quantum-mechanical configurations that were not used for training. It was also compared with experimentally accessible structural properties [5].

A later neural-network potential was trained on a much larger DFT database. It was tested for liquid, amorphous, cubic, and hexagonal phases. The validation included structural distributions, thermodynamic properties, diffusion, melting, and crystallization kinetics [10].

For example, the predicted melting temperature was about 920–930 K, compared with about 900 K experimentally. The predicted glass-transition temperature was about 500 K, compared with a recent experimental value of about 473 K. The predicted homogeneous crystal-growth velocity was about 4.0 m s^{−1} at 600 K, compared with an experimental estimate of about 2.5 m s^{−1}.

A later study on viscosity extended the physical validation of the same potential. It did not provide a completely independent new fit [44].

More recent work on Ge-rich Ge–Sb–Te gives a stronger test of transfer between compositions. Several compositions that were not included in the training database were

first compared with DFT before the potential was used for large-scale phase-separation and crystallization simulations [45].

This supports Level 2 domain-aware validation within the Ge–Sb–Te system, but it does not prove transfer to completely different chemical systems.

These studies also show that the type of potential matters. A model can have low average energy or force errors and still predict the wrong phase order, diffusion behavior, nucleation, or transformation pathway.

The potential architecture, interaction range, reference database, and training method should therefore be considered important modelling choices that need physical validation [111–113].

When possible, comparisons between different potential types, architectures, or cut-off values should be combined with at least one physical property directly related to the transformation.

Other studies focus on different parts of the phase-change problem. These include glass classification, prediction of refractive index, density, or glass-transition temperature, and deep-learning methods for in-line GeSbTe metrology [7,11,102,103,105].

Composition–property models are mainly useful for selecting or improving formulations, while spectral metrology is mainly used for process monitoring.

These applications should not be confused with inverse design.

For example, Bayesian optimization of a GST optical emitter [128] optimizes the geometry of a device that already contains a known phase-change material. It does not discover a new phase-change composition.

5.4. Defects, Interfaces, Disorder, and Transport

Defects and interfaces strongly affect the properties and degradation of chalcogenides.

MLIPs have been used to study sulfur vacancies in MoS₂, defective or amorphous TMDCs, interdiffusion at metal–chalcogenide contacts, lateral heterostructures, CdTe-related defects, and Cu clustering [4,37,38,40,46,47,104].

Image-based models have also been used to detect and describe atomic-scale defects in MoS₂ and WSe₂ [76,77]. Explainable and causal approaches have linked defects and processing conditions to mechanical properties, non-radiative recombination, and nanoparticle properties [18,81,87,129].

These studies include information that is often missing from equilibrium materials databases.

For example, a defect model may need information about charge state, chemical potentials, local atomic relaxation, defect concentration, and the simulation method used. An interface may require information about atomic alignment, strain, surface termination, intermixing, and environment.

If these variables are missing, a model may still find useful correlations, but these correlations may not represent the real physical mechanism that controls the property.

Transport properties have similar difficulties because they depend on several length and time scales.

Thermal conductivity can depend on temperature, system size, boundary scattering, isotope content, anharmonic effects, and microstructure.

Chalcogenide MLIPs have been used to study WSe₂ nanostructures, MoS₂(1–x)Se_{2x} alloys, CaZrS₃, GeTe/Sb₂Te₃ superlattices, TMDC ripples, and higher-order phonon scattering [32,33,36,41–43,97,100].

These results should be compared under similar physical conditions and, when possible, with both first-principles calculations and experimental measurements.

Grain boundaries are a useful test for atomistic potentials used in transport studies.

In monolayer WSe₂, simulations showed that grain boundaries can reduce thermal conductivity by more than a factor of two compared with the perfect material [125].

In MoS₂, however, grain-boundary conductance changes non-monotonically with the misorientation angle [126]. This shows that a simple rule such as “larger angle means larger resistance” is not always valid.

In bulk SnTe, spatially resolved experiments directly linked the properties of individual grain boundaries to measurable thermal resistance [127].

For this reason, a potential used to study polycrystalline transport should not be validated only with perfect-crystal lattice parameters, equilibrium phonons, or average energy and force errors.

Useful validation can also include grain-boundary thermal resistance, thermal conductivity as a function of grain size or boundary type, defect-sensitive phonon transport, and temperature dependence.

The same care is needed for properties that depend strongly on the environment.

For example, machine learning has been used to predict the tribological performance of WS₂ coatings under different operating conditions [130].

To test whether such a model can transfer to new conditions, variables such as load, humidity, counterface material, film preparation, and wear history should be treated as important parts of the scientific domain, not simply as equivalent rows in a dataset.

5.5. Characterization and Metrology

Raman spectroscopy is one of the most developed characterization applications in the chalcogenide studies reviewed here.

Supervised and unsupervised models have been used to classify MoS₂ spectra, connect Raman and photoluminescence maps, estimate twist angle, study anharmonicity, identify phase-change states, and support temperature sensing [75,79,83,84,86,102,105].

This is useful because spectroscopic measurements are fast, contain a lot of information, and can be included relatively easily in process monitoring.

The main risk is data leakage between measurements from the same sample or fabrication conditions.

For example, spectra from nearby pixels, repeated scans, or the same flake should not be randomly divided between training and test sets if the final goal is to predict a new sample or use a different instrument.

For a pixel-level model, complete maps should be kept separate for testing. For use on new samples, independent specimens should be tested. For manufacturing applications, different batches, measurement days, and preferably different instruments should also be used for validation.

For two-dimensional chalcogenides, the substrate is another important variable and should not be treated only as descriptive information.

Studies of MoS₂ growth show that the substrate can change nucleation and preferred orientation. Sapphire and SiO₂ therefore do not represent the same growth conditions [131].

Different crystallographic planes of sapphire can also give different orientations and morphologies [132]. In addition, the surface termination of C-plane sapphire can influence orientation selectivity [133].

For this reason, a Raman, photoluminescence, microscopy, or defect-classification model trained on material grown on one substrate should not automatically be expected to work on another substrate.

The difference is not only due to the instrument. The substrate can change nucleation, crystal orientation, domain size, grain-boundary density, strain, and interface interactions.

For thin-film and 2D characterization, important grouping variables can therefore include specimen, synthesis batch, substrate identity, substrate crystallinity and orientation, surface preparation or termination, growth reactor, and measurement instrument when these variables can affect the target.

Data-processing steps are also part of the model. Preprocessing, peak alignment, baseline correction, data augmentation, and rejection rules should therefore be reported.

Microscopy and diffraction also offer useful applications, although there are fewer chalcogenide-specific studies.

Deep-learning methods have been used to identify defects in STM and electron-microscopy images and to measure nanoscale changes [76,77,87].

General methods developed for small X-ray diffraction datasets and automated nanoparticle TEM analysis provide useful examples [117,118]. However, they should be considered as transfer methods until they are validated on the intended chalcogenide samples.

The same is true for photoluminescence classification across different semiconductor families [115] and dynamic fibre-optics modelling [116].

For characterization tasks, the evidence level mainly depends on how the model is tested and used.

Retrospective classification within a similar data distribution corresponds to Level 1.

Testing on different specimens, substrates, instruments, or other clearly defined domains provides Level 2 evidence.

A model that is used prospectively to select a new synthesis or measurement reaches Level 3.

Level 4 requires repeated cycles in which data acquisition, model prediction, decision, and experimental action are directly connected.

5.6. Synthesis Optimization and Closed-Loop Workflows

Synthesis is the stage where model predictions meet the full complexity of an experimental system.

Machine-learning methods have been used for Chevrel-phase composition selection, Cu-Sn-S analysis, CZTSSe fabrication, Ag₂Se processing and printing, MoS₂ layer control, WSe₂ film-growth characterization, and control of SnSe or SnS morphology [19,49,50,52,54–58,80,85].

These studies show that AI can already be useful in limited experimental spaces where the number of variables is controlled and characterization is relatively fast.

Recent TMDC growth studies provide stronger prospective evidence.

A multiscale WSe₂ MOCVD study combined computational fluid dynamics, phase-field modelling, and reactive molecular dynamics. It identified effects of precursor gradients, predicted that substrate rotation could improve growth, and then tested this prediction experimentally [134].

A Bayesian-optimization study for WS₂ CVD reported an 86.6% increase in photoluminescence intensity over 13 rounds and compared the results with random search [119].

Both studies are classified as Level 3 because the model-guided decisions were made before the confirming measurements.

Real-time spectroscopic reflectance during TMDC MOCVD is another useful tool. It can detect growth, coalescence, and thickness-related changes during the experiment [135].

However, this is not Level 4 in our framework because repeated AI-driven cycles of decision and experimental action were not demonstrated.

General synthesis prediction is more difficult because many experimental variables are not standardized between laboratories. These include precursor state, reactor geometry, atmosphere, substrate, heating history, operator decisions, and failed experiments.

For this reason, we use the term “closed loop” only for a repeated operational cycle.

The result of one experiment must be used to update the model or decision process. The next action must then be selected before its result is known. The workflow should also include a stopping rule or a rule for human intervention.

A single model-selected experiment is useful Level 3 evidence, but it is not Level 4.

A-Lab provides a broader inorganic example of repeated closed-loop laboratory operation [60].

Its evidence should be separated into two different claims. The repeated planning, execution, analysis, and selection cycle supports Level 4 for closed-loop operation.

However, the 2026 Author Correction explains that the word “new” in the original paper did not necessarily mean new to science [60].

The independent analysis by Leeman et al. also highlighted problems related to novelty and structural disorder [114].

For this reason, we do not consider A-Lab as Level 4 evidence for material novelty.

The failed-experiment study provides another useful example from a broader inorganic system. A reaction model was trained using both successful and failed hydrothermal experiments and was then tested prospectively with new building blocks. The study reported an 89% success rate [109].

However, neither A-Lab nor this failed-experiment study demonstrates autonomous discovery across chalcogenide families.

A more realistic short-term goal for chalcogenides is to connect AI directly to instruments for optimization within a clearly defined material family, substrate, reactor, and process range.

These limited workflows can reduce the number of experiments needed, improve reproducibility, and keep better records of unsuccessful conditions without claiming general laboratory autonomy. Examples of studies classified according to the different evidence levels are summarized in Table 6.

Table 6. Examples of evidence levels in AI studies on chalcogenides. Levels apply to specific claims. For prospective evidence, the model prediction or recommendation must be made before the confirming calculation or experiment.

Material/Claim	Data and Role of AI	Main Validation	Level	Ref.
Kesterite band gaps	DFT data with descriptor-based regression	Tested on held-out calculations within the same chemical family	1	[1]
Chalcogenide double perovskites	ML used to select promising absorber candidates	ML-selected candidates were then tested with more accurate stability and optical calculations	3	[6]
CdSe/CdS nanoparticles	Causal ML applied to synthesis–property data	A new experimental result followed a model-guided hypothesis	3	[129]
Multinary Chevrel phases	ML used to select compositions and predict stability	Model-selected compositions were synthesized and characterized prospectively	3	[49]
Cu–Sn–S phase discovery/characterization	Active-learning segmentation and unsupervised analysis of existing high-throughput images	ML identified candidate phases that were later studied in more detail	3	[50]
MoS ₂ Raman analysis	Classification and regression of Raman spectra	Tested retrospectively on measured spectra	1	[75,79]
TMDC atomic defects	Deep learning on STM and microscopy data	Tested on held-out labels; transfer to new samples or instruments was generally not demonstrated	1	[76,77]
CZTSSe photovoltaics	Experimental process and device data analysed with ML	Model-predicted conditions were later fabricated and tested	3	[18,19]

Table 6. Cont.

Material/Claim	Data and Role of AI	Main Validation	Level	Ref.
Ag ₂ Se devices/composites	ML used to guide sintering, formulation, or printing	Prospective fabrication and device testing	3	[52,54,55]
GST225 phase transformation	Specialized MLIP trained on DFT data from several phases	Tested against held-out/reference DFT and structural, thermodynamic, and kinetic data from DFT or experiment	2	[5,10,44]
Ge-rich Ge–Sb–Te	MLIP tested for transfer between compositions	Compositions not included in training were tested before large-scale phase-separation and crystallization simulations	2	[45]
Chalcogenide-glass properties	Composition–property models	Tested on held-out or new experimental compositions	2	[11,103]
MoS ₂ growth	Regression and Bayesian optimization of growth conditions	Selected growth conditions were tested prospectively	3	[56,80]
WSe ₂ wafer-scale MOCVD	Process model combining CFD, phase-field modelling, and reactive MD	The predicted benefit of substrate rotation was tested experimentally	3	[134]
WS ₂ CVD optimization	Bayesian optimization using photoluminescence intensity	Prospective 13-round campaign with 86.6% PL improvement and comparison with random search	3	[119]
TMDC in situ MOCVD monitoring	Real-time spectroscopic reflectance	Growth, coalescence, and thickness-related signals were demonstrated, but autonomous feedback was not	2	[135]
SnSe/SnS nanoplates	Data-driven model relating process conditions to morphology	Controlled growth experiments	3	[57,58]
Universal potential on WS ₂ /MoS ₂	Pretrained CHGNet model	Compared with DFT and EXAFS for the target materials	2	[47]
Failed-experiment synthesis *	Reaction model trained on successful and failed experiments	Prospectively tested with previously unused building blocks	3 *	[109]
MatterGen candidate validation *	Generative model used to propose candidates	Computational filtering and selected synthesis were performed, but the novelty claim was later questioned	3 * for prospective validation, not for novelty	[61,62]
A-Lab autonomous operation *	Integrated planning, synthesis, XRD analysis, and selection of the next experiment	Repeated laboratory cycles were demonstrated	4 *	[60]
A-Lab material novelty *	Candidate novelty relative to the prediction platform	The correction clarified that “new” did not necessarily mean new to science	1 *	[60,114]

Note: * indicates broader-inorganic examples used as methodological benchmarks.

6. Cross-Family Comparative Assessment

6.1. What Is Mature, and What Is Still Developing?

It is difficult to compare algorithms directly across the chalcogenide literature because the reported errors concern different properties, units, chemical spaces, data splits, and data-generation methods.

For this reason, we compare mainly the validation pathway. We ask what information was available during model training, what scientific variable was really left out, and what happened after the model made a prediction or recommendation.

Table 7 shows the highest representative validation level found among the studies examined for each material-family and task combination.

Table 7. Highest evidence level found among the studies examined in this review. R = Level 1; X = Level 2; P = Level 3; L = Level 4; — = no sufficient direct evidence found in the reviewed studies.

Material Family	Property Prediction/ Screening	ML Atomistic Modelling	Characterization	Synthesis/Process	Generative/ Closed Loop
TMDs and 2D heterostructures	X [3,78,82]	X [4,32,34–38,40–43,46,47]	R [75–77,79,83–85,87]	P [56,80,119,134]	—
Chalcopyrites, kesterites, and related photovoltaic materials	P [24,89]	X [104]	R [25]	P [18,19]	—
Thermoelectric chalcogenides	P [53,55,96]	X [97]	P [50]	P [52,54,55]	—
Phase-change materials and chalcogenide glasses	X [11,103]	X [5,10,39,44,45]	R [102,105]	—	—
Chalcogenide perovskites and other multinary chalcogenides	P [6,99,120]	X [100]	—	P [49]	—
Broader inorganic examples *	X [13–16,72,73]	X [30,31,111–113]	X [117,118]	L (operation) [60]	P/L [60–62,114]

Notes: A-Lab is classified as Level 4 only for repeated autonomous laboratory operation. Its material-novelty claim is considered separately as Level 1. * These studies are included as methodological benchmarks and do not provide direct evidence for chalcogenides.

Several areas have reached Level 3 prospective evidence. These include ML-selected photovoltaic and multinary candidates that were later tested with more accurate calculations [6,24,89,99,120], experimentally tested photovoltaic fabrication conditions [18,19], thermoelectric synthesis or characterization workflows [50,52,54,55], and model-guided TMDC growth [56,80,119,134].

For TMDC processes, the evidence is strengthened by experimentally tested multiscale MOCVD modelling and prospective Bayesian optimization. Real-time optical monitoring during growth is also an important step toward future instrument-connected feedback [119,134,135].

Phase-change materials provide strong Level 2 evidence for ML-based atomistic modelling. This includes structural, thermodynamic, and kinetic validation and, in more recent studies, explicit tests on compositions that were not included in the training database [5,10,44,45].

These evidence levels apply to specific claims.

For example, the Cu–Sn–S study reaches Level 3 for ML-guided phase identification followed by additional characterization. However, it does not reach Level 3 for prospective synthesis selection because the synthesis experiments were performed before the ML analysis [50].

A-Lab provides Level 4 evidence for repeated closed-loop laboratory operation in a broader inorganic system. However, the material-novelty claim is considered separately at Level 1 after the Author Correction [60,114,115].

Among the chalcogenide-specific studies examined in this review, we did not find a Level 4 workflow with repeated model-updated cycles of data acquisition, decision, and experimental action.

This conclusion applies only to the studies examined using our criteria. It does not mean that no such experiment exists.

Broader inorganic studies show that closed-loop operation is possible, but they do not directly prove transfer to chalcogenide systems with defects, disorder, amorphous structures, substrate effects, or strong process dependence.

The five evidence levels used in this review are illustrated in Figure 4.

Evidence from broader inorganic materials cannot be assumed to apply directly to chalcogenides.

A dash means that we did not find enough direct evidence among the studies examined in this review. It does not mean that no such study exists in the complete literature.

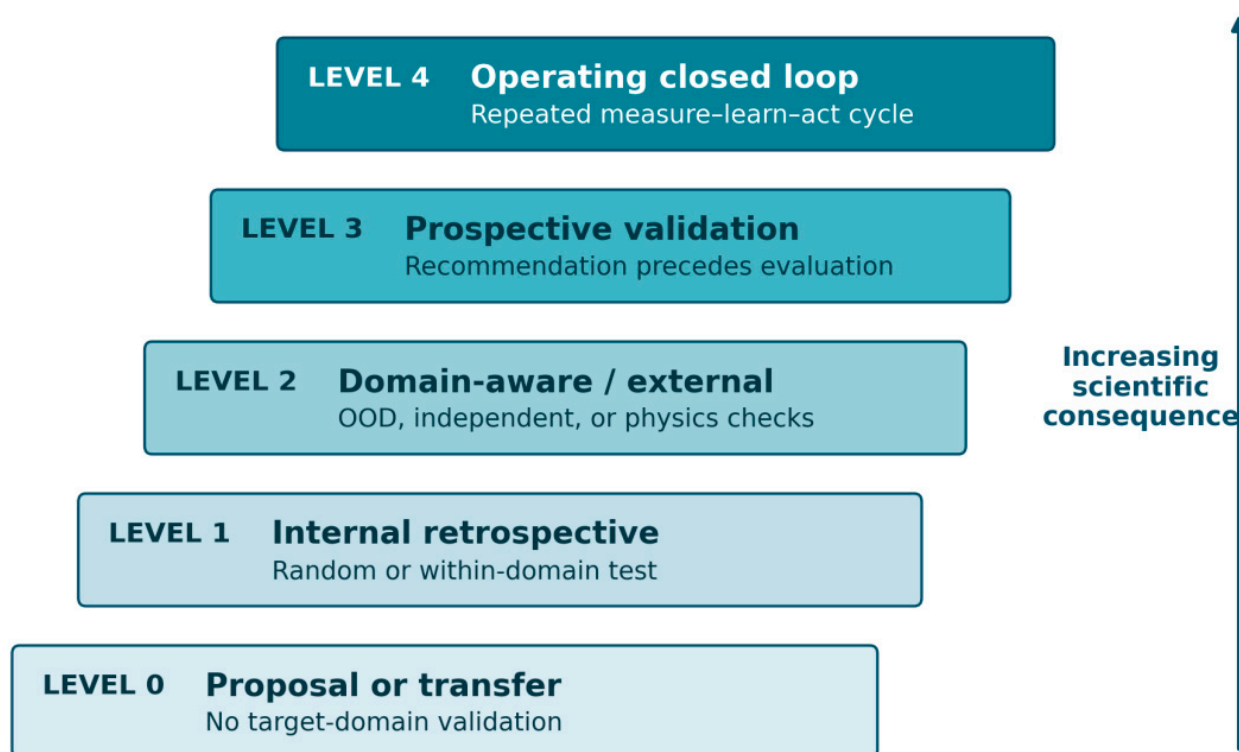


Figure 4. Evidence levels used in this review. Evidence increases from a methodological proposal or transfer example (Level 0) to repeated closed-loop operation (Level 4). Moving from one level to the next is not automatic. Each level needs a different type of validation, and a high evidence level shown for one material system cannot be assumed to apply directly to another.

6.2. Recurrent Failure Modes

Information leakage. Very similar structures, compositions, pixels, spectra, or process records can appear in both the training and test sets. With a random split, the model may only be interpolating between very similar examples instead of showing real transfer.

The data should therefore be grouped according to what is really new in the scientific problem, for example structural prototype, composition family, specimen, substrate, batch, laboratory, instrument, reactor, or time. Duplicate records should also be removed before splitting the data [13,14].

Label and fidelity mismatch. A model can reproduce a computational method very well and still differ systematically from experiment.

For computational data, the DFT functional, pseudopotential, spin treatment, convergence settings, and structural state are part of the label definition. Experimental data also depend on temperature, uncertainty, specimen condition, and measurement method.

When data from different fidelity levels are combined, the relationship between these data sources should remain clear.

Narrow applicability domains. Models may fail when they are applied to new oxidation states, defects, phases, compositions, substrates, or processing conditions.

A good average score can therefore hide large errors in some regions of the data space.

Distance from the training domain, calibrated prediction intervals, error analysis for different groups, and clear rules for rejecting uncertain predictions can help identify this problem [15,16,72,73].

Universal potentials have the same limitation. They should be tested carefully on the target system before zero-shot molecular dynamics is considered reliable [30,31,47].

Uncontrolled confounding. In experimental datasets, composition, processing, thickness, morphology, substrate, and measurement conditions often change at the same time.

Feature-importance methods can show which variables are used by the model, but they do not prove that changing one variable will directly cause a change in the target property.

Causal conclusions therefore need an appropriate study design, such as a causal model, a suitable natural experiment, or a prospective experiment [81,129].

Selective reporting. Successful candidates and high-performing models are more likely to be reported than failed experiments, negative results, or unstable model runs.

The vanadium-selenite study showed the value of keeping failed synthesis experiments in the dataset before prospective testing [109]. This study is a broader inorganic example rather than direct chalcogenide evidence.

Keeping failed experiments, acquisition histories, and stopping decisions gives a more realistic view of the limits of the experimental space.

Weak baselines and incomplete tests. Comparing several complex models is not enough if a simple model can answer the same question.

Simple descriptor-based or physically motivated models, learning curves, and tests where some descriptors or data types are removed can provide useful information.

For MLIPs, the potential family, model architecture, training database, and cutoff may also need to be tested separately [111–113].

For active learning, random selection and a simple expert rule are useful baselines.

Mismatch between metric and decision. Mean absolute error gives the same importance to all predictions, but real discovery or process decisions may depend more on ranking, thresholds, rare candidates, or safe use outside the training domain.

The evaluation should therefore also include a measure related to the final decision, such as recall of useful candidates, calibration near a selection threshold, regret, sample efficiency, or the fraction of experiments that meet a predefined success criterion.

Novelty and structural disorder. A material that is new to a model or database is not necessarily new to science. Also, an ordered predicted structure may not fully describe the real experimental material.

The A-Lab critique and the later Author Correction illustrate these two problems [60,114].

Automatic structure identification can be difficult when the material is compositionally disordered or only partly ordered.

This is especially important for chalcogenides because antisite disorder, cation or configurational disorder, high-entropy compositions, amorphous structures, and metastable local structures can directly affect their properties.

Generative and autonomous workflows should therefore consider candidate novelty, structural identification, disorder, synthesizability, and prospective confirmation as separate questions instead of combining them into one “discovery” claim.

The MatterGen discussion gives another example of why similarity to the training data and structural disorder should be checked carefully [61,62].

6.3. Minimum Reporting Standard

To make studies easier to compare and reuse, the following information should be clearly reported:

- (i) The data source, dataset version, and inclusion rules;
- (ii) What each data point represents and which variables are used to group the data;
- (iii) Preprocessing steps fitted only on the training data;
- (iv) A train–test split that matches the type of novelty or real use being claimed;
- (v) Both simple baseline models and more advanced models;
- (vi) Uncertainty information or a clearly defined applicability domain;
- (vii) Prediction errors for different chemistries, structures, or operating conditions;

- (viii) Whether model recommendations were made prospectively and when they were made;
- (ix) Code, model settings, and machine-readable results when licensing allows this.

For experimental studies, complete information about synthesis and characterization should also be reported. Failed experiments and changes from the planned experimental sequence are important as well.

For thin-film and two-dimensional chalcogenides, additional information should include the substrate type and composition, whether the substrate is crystalline or amorphous, its crystallographic plane or orientation, surface termination or preparation, and any pretreatment when relevant [131–133].

For process and synthesis models, important information can include reactor geometry and configuration, precursor identity and delivery method, precursor flux or concentration, carrier-gas conditions, pressure, temperature history, substrate position and rotation, synthesis batch, laboratory, and instrument identifiers [134,136–138].

These variables should not be treated only as descriptive information. When they can affect the result, they should also be considered when separating training and test data.

If a model is claimed to transfer to new substrates, reactors, instruments, laboratories, or material chemistries, its performance should be reported separately for each new domain.

Any recalibration or fine-tuning after transfer should also be reported, together with the amount of additional data needed.

7. Research Agenda: From Better Benchmarks to Bounded Autonomy

The next research needs can be divided into short-, medium-, and longer-term goals. These time periods should be considered as milestones, not as exact predictions.

Progress will also be different across computational modelling, characterization, and experimental synthesis.

7.1. Near Term (1–2 Years): Make Validation Match the Claim

In the near term, an important priority is to make existing datasets easier to check and reproduce, rather than only developing new model architectures.

A useful benchmark should clearly define the target property, the origin of the data, what each data point represents, how duplicates are treated, and what is considered new.

For crystal-property prediction, random train-test splits should be combined with tests based on structural prototype, chemistry, or time.

For spectroscopy and microscopy, training and test data should come from different specimens, batches, and instruments.

For synthesis datasets, complete process histories and failed experiments should also be included.

These changes are relatively simple, but they can strongly affect the measured model performance [13–16,109].

Chalcogenide-specific benchmark datasets would be especially useful for problems that are not well represented in general inorganic databases. These include charged and neutral defects, different crystal phases, layered structures, amorphous configurations, interfaces, and measurements that strongly depend on processing conditions.

The same fixed train-test splits could then be used to compare descriptor-based models, current graph models, and pretrained potentials.

Model documentation should also clearly describe the intended use, domains where the model should not be used, known failure modes, uncertainty behavior, and the versions of important software or data dependencies.

7.2. Medium Term (3–5 Years): Connect Data Types and Laboratories

In the medium term, an important challenge is to connect process data and data from different fidelity levels while keeping their origin clear.

Computed structures and energies can be combined with experimental composition, diffraction, spectra, microscopy, processing history, and device results. However, the laboratory, substrate, reactor, instrument, fidelity level, temperature, and uncertainty should remain clearly identified.

Without this information, a model may appear to transfer well while actually learning patterns that are specific to one laboratory, reactor, or instrument.

For MLIPs, the goal should be to move beyond potentials developed only for one specific study. More general workflows should monitor uncertainty for defects, interfaces, liquid–glass–crystal transitions, and reactive processes.

Broad pretrained potentials can be used as a starting point. Active learning and targeted quantum calculations can then be used to improve the model in the configuration space that is actually explored [15,30,31,47].

The choice of potential family or cutoff should also be tested when predictions of transformations or defects strongly depend on the interaction model [111–113].

Cross-domain validation is especially important for TMDC synthesis.

General CVD modelling shows that growth conditions do not transfer easily between different reactor types and operating conditions. Morphology can also depend on reactor-specific transport and precursor concentrations [136].

Studies on vertical CVD also show that changes in reactor design and precursor delivery can strongly affect uniformity and reproducibility [137].

A recent MOCVD study provides a useful chalcogenide example under matched conditions.

MoS₂, WS₂, and WSe₂ monolayers were grown on 2-inch c-plane sapphire in the same cold-wall reactor under similar thermofluidic conditions. Even under these controlled conditions, the materials showed different nucleation densities, domain sizes, and lateral growth rates [138].

Substrate rotation also changed mass transport, deposition rate, and thickness uniformity.

This means that process transfer can already be difficult between chemically related TMDCs in the same reactor, even before considering differences between reactors or laboratories.

A useful next step would therefore be to build matched-condition datasets covering several chemistries, reactors, and laboratories.

These datasets should include complete information about precursors, substrates, reactors, thermofluidic conditions, and characterization.

A process model trained in one domain should then be tested, without refitting, on a predefined new chemistry, substrate, reactor, or laboratory.

If recalibration is needed, the amount of new data required and the recovered model performance should both be reported.

These tests would help separate three different questions: whether a model works within one reactor, whether it transfers to related chemistries or substrates under similar conditions, and whether it transfers to different reactors or laboratories.

The last case is the strongest practical test of model generalization for synthesis.

Experimental optimization should also compare model-guided choices with expert or random selection.

Every recommendation, experimental result, intervention, override, and stopping decision should be recorded.

Instrument-connected measurements, such as in situ TMDC reflectance [135], could provide a useful step from prospective Level 3 optimization toward bounded Level 4 feedback, but only if the measurement is actually used to choose the next action.

7.3. Longer Term (>5 Years): Interoperable, Human-Governed Closed Loops

In the longer term, one workflow could combine literature and database searches, candidate generation, multi-fidelity simulation, robotic synthesis, rapid characterization, and adaptive experimental design.

However, such a system would still need a clearly defined domain. The material family, allowed actions, safety rules, objective, uncertainty policy, and conditions for human intervention should be decided in advance.

If the system moves outside this domain, it may be better to stop, ask for human input, or request higher-fidelity data instead of continuing automatically.

Generative methods will become more useful if novelty, disorder, metastability, synthesizability, and processing conditions are included directly in the validation process.

For example, an ordered crystal that is very similar to the training data and cannot be synthesized has limited value for discovery.

In contrast, even a relatively simple model can provide strong evidence if it proposes a reproducible composition or process outside the original data domain and this result is confirmed by independent testing.

In the future, connected laboratories could also test whether models transfer between different reactors and institutions.

Before this becomes practical, several issues need to be solved, including common data formats, calibration standards, licensing, cybersecurity, and data origin.

Table 8 summarizes these priorities as practical goals, measurable success criteria, and approximate time horizons.

Table 8. Main priorities for reliable AI research on chalcogenide materials.

Priority	Main Deliverable	How Success Can Be Measured	Time Horizon
Benchmarks that avoid data leakage	Versioned datasets with fixed random and application-relevant splits, for example by chemistry, structural prototype, SOAP-LOCO, time, specimen, instrument, or substrate	Performance and uncertainty reported for every split; duplicate-removal and grouping rules clearly described	1–2 y
Process, substrate, and defect metadata	Machine-readable data formats for defects, interfaces, substrate state, synthesis, reactor, and characterization conditions	Important variables, units, uncertainty, data origin, and missing data clearly reported	1–2 y
Applicability-domain control	Calibrated uncertainty intervals, measures of distance from the training domain, and rules for stopping or requesting higher-quality data	Coverage and risk tested under the same type of distribution shift expected in real use	1–3 y
Multi-fidelity learning	Connected DFT, higher-level calculations, experimental data, and process data	Errors reported separately for each fidelity level, with transfer improvements shown clearly and systematic differences not hidden	2–4 y
Transferable MLIP workflows	Potentials with uncertainty monitoring for defects, interfaces, and transformations; sensitivity to potential family and cutoff tested when possible	Stable simulations and good agreement for observables related to transformations, defects, transport, or kinetics	2–5 y
Prospective synthesis studies	Model-guided experimental campaigns planned and recorded before the results are known	Sample efficiency or regret compared with random or expert selection; failed experiments also kept	2–5 y

Table 8. *Cont.*

Priority	Main Deliverable	How Success Can Be Measured	Time Horizon
Validation across domains and laboratories	Matched-condition datasets covering several chemistries, substrates, reactors, and laboratories	Transfer tested on a predefined new domain; performance before and after recalibration reported, together with the amount of new data needed	3–5 y
Constrained generative design	Candidates checked for similarity to training data, disorder, validity, stability, kinetics, and practical feasibility	Selected candidates pass independent calculations and synthesis/characterization tests; novelty is checked separately	3–6 y
Bounded closed-loop systems	Instrument-connected cycles of data acquisition, analysis, decision, and action within a clearly defined material or process domain	Several complete cycles, with recorded recommendations, interventions, overrides, stopping rules, baseline comparison, and reproducible results	3–7 y

This also changes how progress should be described.

A study does not need full autonomy to be important.

Reliable prediction within a clearly defined domain, a transferable potential, an externally validated spectroscopy model, or one synthesis recommendation confirmed prospectively can all be important results, as long as the limits of the evidence are clearly stated.

8. Conclusions

Chalcogenides show why AI in materials science cannot be evaluated only with standard equilibrium-crystal benchmarks.

These materials cover a wide range of compositions, but many important properties also depend on factors that are not always well represented in common databases. These include defects, interfaces, disorder, metastability, grain boundaries, substrate state, processing history, temperature, reactor conditions, and the specific structure of each sample.

For this reason, high benchmark accuracy is useful, but it is not enough to show that a model is ready for scientific or practical use.

Several main conclusions can be drawn from the studies examined in this review.

Conventional descriptor-based models are still useful for small and well-defined composition or process datasets. They also remain important baseline models.

Graph neural networks are powerful for structure-based predictions when the training and application domains are similar. However, good performance with a random split does not prove that the model will transfer to new chemistries, structural prototypes, atomic environments, specimens, substrates, or process conditions.

Machine-learned interatomic potentials provide especially useful examples for chalcogenides. They allow simulations of phase transformations, defects, interfaces, and transport over larger length and time scales.

However, their reliability should not be judged only from energy and force errors. They should also be tested against physical properties related to transformations, defects, transport, or kinetics. When possible, the effects of model architecture, interaction range, and training database should also be tested.

Prospective evidence is becoming stronger in some well-defined computational and experimental domains.

Examples at Level 3 include ML-selected candidates followed by higher-fidelity calculations, model-guided photovoltaic fabrication, thermoelectric processing and characterization, multiscale TMDC process modelling, and Bayesian optimization of growth.

Real-time measurements also provide a promising route toward instrument-connected feedback.

However, these examples should be separated from Level 4 operation. Level 4 requires several repeated cycles in which new experimental observations are used to update the model or the next decision.

Generative candidate design, scientific novelty, prospective validation, and autonomous operation are also different claims and should be evaluated separately.

The A-Lab Author Correction clearly illustrates this point. Repeated autonomous laboratory operation still provides Level 4 evidence for operation in a broader inorganic system. However, being new to a prediction platform does not automatically mean that a material is new to science.

Structural disorder, possible alternative phase assignments, and similarity to the training or reference data should therefore be checked separately before a generated or autonomously synthesized material is described as a new discovery.

Among the chalcogenide-specific studies examined in this review, Level 4 closed-loop operation has not yet been demonstrated. Although broader inorganic studies show the potential of autonomous experimentation, its application to chalcogenide systems still requires direct validation. This gap in evidence should not be interpreted as a limitation of autonomous approaches themselves, but rather as a need for further experimental demonstration.

The most immediate opportunities are therefore more focused and easier to test. These include OOD benchmarks that match the real application, calibrated uncertainty with clear rules for stopping or requesting more information, MLIP validation based on transformation-related properties, and models that keep information about substrates, grain boundaries, reactors, precursors, and processing conditions.

Prospective studies should also compare model-guided decisions with clear random or expert baselines. Transfer should be tested under matched conditions across different chemistries, substrates, reactors, instruments, and laboratories.

Failed experiments and unsuccessful model decisions should also be kept because they help define the real limits of what can be achieved.

With these elements in place, AI can become more reliable not only for predicting material properties, but also for choosing which calculation, measurement, or experiment should be performed next.

Just as importantly, it should be able to recognize when the available evidence is not sufficient to support a reliable decision.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/aimater1030007/s1>, The following supporting information accompanies this article: Table S1: Primary CAS SciFinder searches; Table S2: Complementary targeted searches; Table S3A: Operational title-coding definitions—Material-family coding; Table S3B: Operational title-coding definitions—Scientific-task coding; Table S3C: Family-task intersections; Table S4: Validation evidence for representative Ge–Sb–Te machine-learned interatomic potentials; Table S5: Recommended OOD splits and uncertainty diagnostics; Table S6: Comparison with selected previous reviews. Machine-readable derived bibliometric data are additionally provided in XLSX format.

Author Contributions: Conceptualization, M.-C.R. and P.B.; methodology, M.-C.R. and P.B.; investigation, M.-C.R. and P.B.; formal analysis, M.-C.R.; data curation, M.-C.R.; validation, P.B.; writing—original draft preparation, M.-C.R.; writing—review and editing, M.-C.R. and P.B.; supervision, M.-C.R.; project administration, M.-C.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The primary bibliographic records were obtained from CAS SciFinder and remain subject to the applicable CAS database licence. The complete search strategies, retrieval counts, deduplication rules, title-coding definitions, annotation-audit methodology, and figure-source variables are reported in the Supplementary Materials. A machine-readable derived dataset containing 6314 globally deduplicated records is provided with the article in XLSX format. The released fields include record identifier, DOI, publication year, snapshot date, SciFinder query labels, membership in the 4066-record AI-explicit subset, final material-family and scientific-task flags, and a SHA-256 digest of the normalized title used for fallback matching. Raw CAS titles, abstracts, and other licensed database content are not redistributed.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.6 Sol; accessed on August 2026) for language editing, abstract condensation, and formatting assistance. The authors reviewed and edited all output and take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
CIGS	Cu(In,Ga)Se ₂ ; copper indium gallium selenide
CNN	Convolutional neural network
CZTSSe	Cu ₂ ZnSn(S,Se) ₄ ; copper zinc tin sulfoselenide
DFT	Density-functional theory
DOI	Digital Object Identifier
EXAFS	Extended X-ray absorption fine structure
FAIR	Findable, Accessible, Interoperable, Reusable
GNN	Graph neural network
ML	Machine learning
MLIP	Machine-learned interatomic potential
MP18-to-MP21	Materials Project-2018.06.01-to-Materials Project-2021.11.10 versions
OOD	Out-of-distribution
PV	Photovoltaic
SEM	Scanning electron microscopy
SHAP	SHapley Additive exPlanations
STM	Scanning tunnelling microscopy
TEM	Transmission electron microscopy
TMDC	Transition-metal dichalcogenide
XRD	X-ray diffraction

References

1. Weston, L.; Stampfl, C. Machine learning the band gap properties of kesterite I2-II-IV-V4 quaternary compounds for photovoltaics applications. *Phys. Rev. Mater.* **2018**, *2*, 085407. [[CrossRef](#)]
2. Mechraoui, B.K.; Benghia, A.; Hebboul, Z.; Lefkaier, I.K.; Halit, M.; Arar, R. From data to discovery: Exploring novel photovoltaic perovskite chalcogenide crystals MgMS₃ (M: Zr, Ti, Hf) via machine learning and density functional theory calculations. *Mater. Today Commun.* **2025**, *42*, 111517. [[CrossRef](#)]
3. Wang, T.; Tan, X.; Wei, Y.; Jin, H. Unveiling the layer-dependent electronic properties in transition-metal dichalcogenide heterostructures assisted by machine learning. *Nanoscale* **2022**, *14*, 2511–2520. [[CrossRef](#)] [[PubMed](#)]
4. Hložný, A.; Brndiar, J.; Casula, M.; Stich, I. Structure and dynamics of sulfur vacancies in monolayer MoS₂ studied by DFT-based machine learning potentials. *J. Chem. Phys.* **2025**, *163*, 214118. [[CrossRef](#)] [[PubMed](#)]

5. Mocanu, F.C.; Konstantinou, K.; Lee, T.H.; Bernstein, N.; Deringer, V.L.; Csányi, G.; Elliott, S.R. Modeling the Phase-Change Memory Material, $\text{Ge}_2\text{Sb}_2\text{Te}_5$, with a Machine-Learned Interatomic Potential. *J. Phys. Chem. B* **2018**, *122*, 8998–9006. [[CrossRef](#)] [[PubMed](#)]
6. Agiorgousis, M.L.; Sun, Y.-Y.; Choe, D.-H.; West, D.; Zhang, S. Machine Learning Augmented Discovery of Chalcogenide Double Perovskites for Photovoltaics. *Adv. Theory Simul.* **2019**, *2*, 1800173. [[CrossRef](#)]
7. Li, H.; Liu, P.; Liu, S.; Yang, B. Deep learning-assisted attribute prediction of chalcogenide glasses based on graph classification. *Sci. Rep.* **2025**, *15*, 19360. [[CrossRef](#)] [[PubMed](#)]
8. Poudel, B.; Hao, Q.; Ma, Y.; Lan, Y.; Minnich, A.; Yu, B.; Yan, X.; Wang, D.; Muto, A.; Vashaee, D.; et al. High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys. *Science* **2008**, *320*, 634–638. [[CrossRef](#)] [[PubMed](#)]
9. Gan, Y.; Wang, G.; Zhou, J.; Sun, Z. Prediction of thermoelectric performance for layered IV-V-VI semiconductors by high-throughput ab initio calculations and machine learning. *npj Comput. Mater.* **2021**, *7*, 176. [[CrossRef](#)]
10. Abou El Kheir, O.; Bonati, L.; Parrinello, M.; Bernasconi, M. Unraveling the crystallization kinetics of the $\text{Ge}_2\text{Sb}_2\text{Te}_5$ phase change compound with a machine-learned interatomic potential. *npj Comput. Mater.* **2024**, *10*, 33. [[CrossRef](#)]
11. Belciu, M.-I.; Velea, A. Ensemble Machine Learning for the Prediction and Understanding of the Refractive Index in Chalcogenide Glasses. *Molecules* **2025**, *30*, 1745. [[CrossRef](#)] [[PubMed](#)]
12. Zhao, L.-D.; Lo, S.-H.; Zhang, Y.; Sun, H.; Tan, G.; Uher, C.; Wolverton, C.; Dravid, V.P.; Kanatzidis, M.G. Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals. *Nature* **2014**, *508*, 373–377. [[CrossRef](#)] [[PubMed](#)]
13. Omeel, S.S.; Fu, N.; Dong, R.; Hu, M.; Hu, J. Structure-based out-of-distribution (OOD) materials property prediction: A benchmark study. *npj Comput. Mater.* **2024**, *10*, 144. [[CrossRef](#)]
14. Li, K.; DeCost, B.; Choudhary, K.; Greenwood, M.; Hattrick-Simpers, J. A critical examination of robustness and generalizability of machine learning prediction of materials properties. *npj Comput. Mater.* **2023**, *9*, 55. [[CrossRef](#)]
15. Musielewicz, J.; Lan, J.; Uyttendaele, M.; Kitchin, J.R. Improved Uncertainty Estimation of Graph Neural Network Potentials Using Engineered Latent Space Distances. *J. Phys. Chem. C* **2024**, *128*, 20799–20810. [[CrossRef](#)]
16. Mamun, O.; Yang, C.; Yue, S. Deep graph kernel learning for material & atomic level uncertainty quantification in adsorption energy prediction. *Digit. Discov.* **2026**, *5*, 1568–1578. [[CrossRef](#)]
17. Zhu, C.; Liu, W.; Li, Y.; Huo, X.; Li, H.; Guo, K.; Qiao, B.; Zhao, S.; Xu, Z.; Zhao, H.; et al. Key factors governing the device performance of CIGS solar cells: Insights from machine learning. *Sol. Energy* **2021**, *228*, 45–52. [[CrossRef](#)]
18. Karade, V.C.; Gour, K.S.; He, M.; Jang, J.; Enkhbat, T.; Lee, M.; Kim, J.; Sutar, S.S.; Dongale, T.D.; Ellingson, R.J.; et al. Machine Learning Drives a Path to Defect Engineering for Suppressing Nonradiative Recombination Losses in $\text{Cu}_2\text{ZnSn}(\text{S,Se})_4$ Solar Cells. *ACS Appl. Mater. Interfaces* **2025**, *17*, 35382–35395. [[CrossRef](#)] [[PubMed](#)]
19. Karade, V.C.; Sutar, S.S.; Shin, S.W.; Suryawanshi, M.P.; Jang, J.S.; Gour, K.S.; Kamat, R.K.; Yun, J.H.; Dongale, T.D.; Kim, J.H. Machine Learning Assisted Analysis, Prediction, and Fabrication of High-Efficiency CZTSSe Thin Film Solar Cells. *Adv. Funct. Mater.* **2023**, *33*, 2303459. [[CrossRef](#)]
20. Maoucha, A.; Berghout, T.; Djeflal, F.; Ferhati, H. Machine learning-assisted investigation of CIGS thin-film solar cell degradation using deep learning analysis. *J. Phys. Chem. Solids* **2025**, *199*, 112526. [[CrossRef](#)]
21. Sharma, S.; Ward, Z.D.; Bhimani, K.; Sharma, M.; Quinton, J.; Rhone, T.D.; Shi, S.-F.; Terrones, H.; Koratkar, N. Machine Learning-Aided Band Gap Engineering of BaZrS_3 Chalcogenide Perovskite. *ACS Appl. Mater. Interfaces* **2023**, *15*, 18962–18972. [[CrossRef](#)] [[PubMed](#)]
22. Maoucha, A.; Berghout, T.; Djeflal, F.; Ferhati, H. Machine learning-guided analysis of CIGS solar cell efficiency: Deep learning classification and feature importance evaluation. *Sol. Energy* **2025**, *287*, 113251. [[CrossRef](#)]
23. Cao, J.; Sun, T.; Feng, J.; Yuan, J. Deep learning representation learning model and its applications in band gap prediction of quaternary chalcogenides. *Comput. Mater. Sci.* **2025**, *260*, 114250. [[CrossRef](#)]
24. Zhang, Z.; Xiong, Q.; Liu, S. Enhanced design of kesterite solar cells through high-throughput screening and machine learning approaches. *J. Phys. Chem. Lett.* **2024**, *15*, 9795–9802. [[CrossRef](#)] [[PubMed](#)]
25. Gari-Galindez, J.; Atlan, F.; Andrade-Arvizu, J.; Fonoll-Rubio, R.; Payno, D.; Grau-Luque, E.; Pérez-Rodríguez, A.; Becerril-Romero, I.; Guc, M.; Izquierdo-Roca, V.; et al. Revealing the Impact of CZTSe/CdS Interface Fluctuations on PV Device Performance through Big Data Analysis Assisted by Machine Learning Methods. *Small Methods* **2025**, *9*, 2400661. [[CrossRef](#)] [[PubMed](#)]
26. Maamria, S.; Chiba, I.K.; Bouazra, A.; Said, M. Artificial intelligence-assisted modeling of geometrical effects on the optoelectronic properties of concentric and non-concentric CdSe/CdS core-shell quantum dots. *Phys. Lett. A* **2026**, *574*, 131354. [[CrossRef](#)]
27. Xie, T.; Grossman, J.C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. [[CrossRef](#)] [[PubMed](#)]
28. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S.P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem. Mater.* **2019**, *31*, 3564–3572. [[CrossRef](#)]

29. Choudhary, K.; DeCost, B. Atomistic Line Graph Neural Network for improved materials property predictions. *npj Comput. Mater.* **2021**, *7*, 185. [\[CrossRef\]](#)
30. Chen, C.; Ong, S.P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718–728. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C.J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **2023**, *5*, 1031–1041. [\[CrossRef\]](#)
32. Chan, H.; Sasikumar, K.; Srinivasan, S.; Cherukara, M.; Narayanan, B.; Sankaranarayanan, S.K.R.S. Machine learning a bond order potential model to study thermal transport in WSe₂ nanostructures. *Nanoscale* **2019**, *11*, 10381–10392. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Wang, B.; Li, K.; Zhang, W.; Sun, Y.; Zhou, J.; Sun, Z. Thermal Transport of GeTe/Sb₂Te₃ Superlattice by Large-Scale Molecular Dynamics with Machine-Learned Potential. *J. Phys. Chem. C* **2025**, *129*, 6386–6396. [\[CrossRef\]](#)
34. Han, L.; Qin, G.; Jia, B.; Chen, Y.; Ma, X.; Lu, P.; Guan, P. A potential function of MoS₂ based on machine learning. *Comput. Mater. Sci.* **2023**, *228*, 112312. [\[CrossRef\]](#)
35. Shaidu, Y.; Naik, M.H.; Louie, S.G.; Neaton, J.B. Transferable dispersion-aware machine learning interatomic potentials for multilayer transition metal dichalcogenide heterostructures. *npj Comput. Mater.* **2025**, *11*, 273. [\[CrossRef\]](#)
36. Gu, X.; Zhao, C. Thermal conductivity of single-layer MoS₂(1-x)Se₂x alloys from molecular dynamics simulations with a machine-learning-based interatomic potential. *Comput. Mater. Sci.* **2019**, *165*, 74–81. [\[CrossRef\]](#)
37. Achar, S.K.; Schneider, J.; Stewart, D.A. Using Machine Learning Potentials to Explore Interdiffusion at Metal–Chalcogenide Interfaces. *ACS Appl. Mater. Interfaces* **2022**, *14*, 56963–56974. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Kety, K.; Namsrai, T.; Nawaz, H.; Rostami, S.; Seriani, N. Amorphous MoS₂ from a machine learning inter-atomic potential. *J. Chem. Phys.* **2024**, *160*, 204709. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Moon, M.; Hwang, S.; Kim, J.; Park, Y.; Hong, C.; Han, S. Unveiling Defect Motifs in Amorphous GeSe Using Machine Learning Interatomic Potentials. *ACS Appl. Mater. Interfaces* **2025**, *17*, 57260–57272. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Liu, X.; Wang, B.; Jia, K.; Wang, Q.; Wang, D.; Xiong, Y. First-principles-based machine learning interatomic potential for molecular dynamics simulations of 2D lateral MoS₂/WS₂ heterostructures. *J. Appl. Phys.* **2024**, *135*, 205107. [\[CrossRef\]](#)
41. Zhang, G.; Dong, S.; Yang, C.; Han, D.; Xin, G.; Wang, X. Revisiting four-phonon scattering in WS₂ monolayer with machine learning potential. *Appl. Phys. Lett.* **2023**, *123*, 052205. [\[CrossRef\]](#)
42. de Souza, G.B.G.; Hancock, S.B.; Landau, D.P.; Abate, Y.; de Paiva, R.; Nascimento, V.B. Revealing the role of ripples in phonon modes for MoS₂ and MoSe₂: Insights from molecular dynamics and machine learning. *Phys. Chem. Chem. Phys.* **2025**, *27*, 16768–16782. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Kocabas, T.; Keceli, M.; Gurel, T.; Milosevic, M.V.; Sevik, C. Thermal conductivity limits of MoS₂ and MoSe₂: Revisiting high-order anharmonic lattice dynamics with machine learning potentials. *Appl. Phys. Rev.* **2025**, *12*, 041424. [\[CrossRef\]](#)
44. Marcorini, S.; Pomodoro, R.; El Kheir, O.A.; Bernasconi, M. Viscosity, breakdown of Stokes–Einstein relation, and dynamical heterogeneity in supercooled liquid Ge₂Sb₂Te₅ from simulations with a neural network potential. *J. Chem. Phys.* **2025**, *163*, 154501. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Abou El Kheir, O.; Baratella, D.; Bernasconi, M. On phase separation and crystallization of Ge-rich GeSbTe alloys from atomistic simulations with a machine learning interatomic potential. *J. Mater. Chem. C* **2026**, *advance article*. [\[CrossRef\]](#)
46. Zhou, J.; Fu, Y.; Liu, L.; Liu, C. Constant-Potential Machine Learning Molecular Dynamics Simulations Reveal Potential-Regulated Cu Cluster Formation on MoS₂. *J. Phys. Chem. C* **2025**, *129*, 6414–6422. [\[CrossRef\]](#)
47. Zguns, P.; Pudza, I.; Kuzmin, A. Benchmarking CHGNet Universal Machine Learning Interatomic Potential against DFT and EXAFS: The Case of Layered WS₂ and MoS₂. *J. Chem. Theory Comput.* **2025**, *21*, 8142–8150. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Jang, H.; Lee, W.; Kim, H.; Cha, S.; Shin, H.; Lee, W.B.; Oh, M.; Jung, Y.S.; Kim, Y. Active Learning-Guided Accelerated Discovery of Ultra-Efficient High-Entropy Thermoelectrics. *Adv. Mater.* **2025**, *38*, e15054. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Singstock, N.R.; Ortiz-Rodriguez, J.C.; Perryman, J.T.; Sutton, C.; Velazquez, J.M.; Musgrave, C.B. Machine Learning Guided Synthesis of Multinary Chevrel Phase Chalcogenides. *J. Am. Chem. Soc.* **2021**, *143*, 9113–9122. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Sheng, Y.; Deng, T.; Qiu, P.; Shi, X.; Xi, J.; Han, Y.; Yang, J. Accelerating the Discovery of Cu–Sn–S Thermoelectric Compounds via High-Throughput Synthesis, Characterization, and Machine Learning-Assisted Image Analysis. *Chem. Mater.* **2021**, *33*, 6918–6924. [\[CrossRef\]](#)
51. Zhao, L.; Ren, Y.; Shi, X.; Liu, H.; Yu, Z.; Gao, J.; Zhao, J. Unveiling the unexpected sinking and embedding dynamics of surface supported Mo/S clusters on 2D MoS₂ with active machine learning. *Smart Mol.* **2024**, *3*, e20240018. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Saeidi-Javash, M.; Wang, K.; Zeng, M.; Luo, T.; Dowling, A.W.; Zhang, Y. Machine learning-assisted ultrafast flash sintering of high-performance and flexible silver–selenide thermoelectric devices. *Energy Environ. Sci.* **2022**, *15*, 5093–5104. [\[CrossRef\]](#)
53. Sheng, Y.; Wu, Y.; Yang, J.; Lu, W.; Villars, P.; Zhang, W. Active learning for the power factor prediction in diamond-like thermoelectric materials. *npj Comput. Mater.* **2020**, *6*, 171. [\[CrossRef\]](#)

54. Song, K.; Xu, G.; Tanvir, A.N.M.; Wang, K.; Bappy, M.O.; Yang, H.; Shang, W.; Zhou, L.; Dowling, A.W.; Luo, T.; et al. Machine learning-assisted 3D printing of thermoelectric materials of ultrahigh performances at room temperature. *J. Mater. Chem. A* **2024**, *12*, 21243–21251. [\[CrossRef\]](#)
55. Shang, W.; Zeng, M.; Tanvir, A.N.M.; Wang, K.; Saeidi-Javash, M.; Dowling, A.; Luo, T.; Zhang, Y. Hybrid Data-Driven Discovery of High-Performance Silver Selenide-Based Thermoelectric Composites. *Adv. Mater.* **2023**, *35*, e2212230. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Wang, Y.; Li, G.; Natarajan, A.H.; Mukerjee, S.; Jin, X.; Kar, S. Accelerating Optimal Synthesis of Atomically Thin MoS₂: A Constrained Bayesian Optimization Guided Brachistochrone Approach. *Adv. Mater. Technol.* **2024**, *10*, 2401465. [\[CrossRef\]](#)
57. Luo, H.; Pan, W.; Liu, J.; Wang, H.; Zhang, S.; Ren, Y.; Yuan, C.; Lei, W. Controlled growth of high-quality SnSe nanoplates assisted by machine learning. *J. Mater. Chem. A* **2024**, *13*, 257–266. [\[CrossRef\]](#)
58. Luo, H.; Liu, J.; Wang, H.; Zhang, S.; Pan, W.; Ren, Y.; Yuan, C.; Lei, W. Data-driven strategy for size-controllable growth of tin(II) sulfide nanoplates via machine learning. *J. Alloys Compd.* **2025**, *1047*, 185023. [\[CrossRef\]](#)
59. Abuawwad, N.; Zhang, Y.; Lounis, S.; Zhang, H. Kalman filter enhanced active learning sampling for inelastic neutron scattering: The case of CrSBr. *Phys. Rev. B* **2025**, *111*, 054404. [\[CrossRef\]](#)
60. Szymanski, N.J.; Rendy, B.; Fei, Y.; Kumar, R.E.; He, T.; Milsted, D.; McDermott, M.J.; Gallant, M.; Cubuk, E.D.; Merchant, A.; et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature* **2023**, *624*, 86–91. Correction in *Nature* **2026**, *650*, E1. <https://doi.org/10.1038/s41586-025-09992-y>. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Wang, Z.; Shysheya, A.; Crabbé, J.; Ueda, S.; et al. A generative model for inorganic materials design. *Nature* **2025**, *639*, 624–632. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Juelsholt, M. Continued challenges in high-throughput materials predictions: MatterGen predicts compounds from the training dataset. *Mater. Horiz.* **2026**, *13*, 5672–5679. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Yin, H.; Sun, Z.; Wang, Z.; Tang, D.; Pang, C.H.; Yu, X.; Barnard, A.S.; Zhao, H.; Yin, Z. The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning. *Cell Rep. Phys. Sci.* **2021**, *2*, 100482. [\[CrossRef\]](#)
64. Wang, X.; Sheng, Y.; Ning, J.; Xi, J.; Xi, L.; Qiu, D.; Yang, J.; Ke, X. A Critical Review of Machine Learning Techniques on Thermoelectric Materials. *J. Phys. Chem. Lett.* **2023**, *14*, 1808–1822. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Zhang, X.; Li, K.; Zhou, J.; Sun, Z. Chalcogenide phase-change materials: Unveiling new horizons with big data and machine learning. *J. Mater. Chem. C* **2025**, *13*, 14168–14186. [\[CrossRef\]](#)
66. Liu, H.; Fu, Z.; Yang, K.; Xu, X.; Bauchy, M. Machine learning for glass science and engineering: A review. *J. Non-Cryst. Solids X* **2019**, *4*, 100036. [\[CrossRef\]](#)
67. Mak, K.F.; Lee, C.; Hone, J.; Shan, J.; Heinz, T.F. Atomically thin MoS₂: A new direct-gap semiconductor. *Phys. Rev. Lett.* **2010**, *105*, 136805. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002. [\[CrossRef\]](#)
69. Saal, J.E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509. [\[CrossRef\]](#)
70. Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R.H.; Nelson, L.J.; Hart, G.L.; Sanvito, S.; Buongiorno-Nardelli, M.; et al. AFLOW- LIB.ORG: A Distributed Materials Properties Repository from High-Throughput ab initio Calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235. [\[CrossRef\]](#)
71. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Tan, L.; Wang, X.; Zou, Y.; Chen, P.; Zou, Q. MatUQ: A benchmark for uncertainty-aware out-of-distribution materials property prediction with graph neural networks. *npj Comput. Mater.* **2026**. [\[CrossRef\]](#)
73. Garain, B.C.; Pinheiro, M., Jr.; de Oliveira Bispo, M.; Barbatti, M. Uncertainty Calibration in Molecular Machine Learning: Comparing Evidential and Ensemble Approaches. *Chem. Eur. J.* **2026**, *32*, e03299. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Sun, T.; Yuan, J.-M. Prediction of band gap of transition metal sulfide with Janus structure by deep learning atomic feature representation method. *Acta Phys. Sin.* **2023**, *72*, 028901. [\[CrossRef\]](#)
75. Mao, Y.; Dong, N.; Wang, L.; Chen, X.; Wang, H.; Wang, Z.; Kislyakov, I.M.; Wang, J. Machine learning analysis of Raman spectra of molybdenum disulfide. *Nanomaterials* **2020**, *10*, 2223. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Wu, S.; Chen, G.; Shen, S.; Yan, J. Point Defect Detection and Classification in MoS₂ Scanning Tunneling Microscopy Images: A Deep Learning Approach. *Molecules* **2025**, *30*, 2644. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Smalley, D.; Lough, S.D.; Holtzman, L.N.; Holbrook, M.; Hone, J.C.; Barmak, K.; Ishigami, M. Determining the density and spatial descriptors of atomic scale defects of 2H-WSe₂ with ensemble deep learning. *APL Mach. Learn.* **2024**, *2*, 036104. [\[CrossRef\]](#)
78. Yang, S.; Chen, J.; Liu, C.-F.; Chen, M. Evolution of flat bands in MoSe₂/WSe₂ moiré lattices: A study combining machine learning and band unfolding methods. *Phys. Rev. B* **2024**, *110*, 235410. [\[CrossRef\]](#)

79. Lu, A.; Martins, L.G.P.; Shen, P.; Chen, Z.; Park, J.; Xue, M.; Han, J.; Mao, N.; Chiu, M.; Palacios, T.; et al. Unraveling the Correlation between Raman and Photoluminescence in Monolayer MoS₂ through Machine-Learning Models. *Adv. Mater.* **2022**, *34*, e2202911. [[CrossRef](#)] [[PubMed](#)]
80. Lu, M.; Ji, H.; Chen, Y.; Gao, F.; Liu, B.; Long, P.; Deng, C.; Wang, Y.; Tao, J. Machine learning assisted layercontrolled synthesis of MoS₂. *J. Mater. Chem. C* **2024**, *12*, 8893–8900. [[CrossRef](#)]
81. Anuragi, A.; Das, A.; Baski, A.; Maithani, V.; Mukherjee, S. Machine learning predicted inelasticity in defective two-dimensional transition metal dichalcogenides using SHAP analysis. *Phys. Chem. Chem. Phys.* **2024**, *26*, 15316–15331. [[CrossRef](#)] [[PubMed](#)]
82. Khor, J.W.P.; Tran, T.T.; Sharbirin, A.S.; Yap, S.X.B.; Choo, H.; Kim, J. Prediction of Quantum Yields of Monolayer WS₂ by Machine Learning. *Adv. Opt. Mater.* **2024**, *12*, 2302195. [[CrossRef](#)]
83. Tang, J.; Wang, Z.; Tan, Q.; Cao, J.; Ling, X. Determining the twist angle of stacked MoS₂ layers using machine learning-assisted low-frequency interlayer Raman fingerprints. *J. Raman Spectrosc.* **2023**, *54*, 1021–1029. [[CrossRef](#)]
84. Okeke, C.; Juma, I.; Cobarrubia, A.; Schottle, N.; Maddah, H.; Mortazavi, M.; Behura, S.K. Probing anharmonic phonons in WS₂ van der Waals crystal by Raman spectroscopy and machine learning. *iScience* **2023**, *26*, 107174. [[CrossRef](#)] [[PubMed](#)]
85. Moses, I.A.; Wu, C.; Reinhart, W.F. Crystal growth characterization of WSe₂ thin film using machine learning. *Mater. Today Adv.* **2024**, *22*, 100483. [[CrossRef](#)]
86. Roy, P.; Chaudhuri, P.R. Raman-induced wavelength shift in chalcogenide microstructure fiber: Temperature sensing and machine learning analysis. *Appl. Phys. B* **2024**, *130*, 221. [[CrossRef](#)]
87. Suh, H.C.; Yoo, J.; Yeo, K.; Kim, D.H.; Won, Y.S.; Kim, T.; Cho, Y.; Kim, K.K.; Lee, S.M.; Yang, H.; et al. Probing nanoscale structural perturbation in a WS₂ monolayer via explainable artificial intelligence. *Appl. Phys. Rev.* **2025**, *12*, 021406. [[CrossRef](#)]
88. Giliberti, G.; Farias-Basulto, G.; Jäger, K.; Mehlhop, T.; Kaufmann, C.A.; Di Carlo, A. Assessment and Optimization of 2T Perovskite/CIGS Tandems via Data-Driven and Optoelectronic Modelling. *Nano-Micro Lett.* **2026**, *18*, 312. [[CrossRef](#)] [[PubMed](#)]
89. Singh, N.; Ubaid, M.; Nayak, P.K.; He, J.; Ghosh, D.; Wolverson, C.; Pal, K. Data-driven discovery of high-performance quaternary chalcogenide photovoltaics. *Phys. Rev. Appl.* **2026**, *25*, 034073. [[CrossRef](#)]
90. Ahamed, B.B.; Pal, D.; Ghosh, P.; De, A. An innovative machine learning based approach for predicting the efficiency of a chalcogenide perovskite semiconductor photovoltaic cell. *Opt. Quantum Electron.* **2025**, *57*, 547. [[CrossRef](#)]
91. Ghosh, A.; Hasan, R.M.; Moumita, M.; Apu, K.U.; Begum, S.; Rahman, A.; Bappy, A.; Ahmed, S.; Kumar, Y.A.; Hassan, A.A.; et al. Discovering the relationship between sulfide and selenide-based HTLs and cubic Ba₃SbI₃ solar cells with SCAPS-1D and machine learning modelling. *Inorg. Chem. Commun.* **2025**, *173*, 113782. [[CrossRef](#)]
92. Joya, J.A.; Hossain, A.M.; Chakma, K.; Hussain, S. Numerical study of chalcopyrite n-CuAlSe₂/p-CuGaSe₂ solar cell with an additional CuTiTe₂ back absorber layer using a SCAPS-1D simulator and machine learning approach. *Semicond. Sci. Technol.* **2026**, *41*, 025006. [[CrossRef](#)]
93. Merad, F.; Kanoun, A.-A.; Goumri-Said, S.; Kanoun, M.B. Performance optimization of CIGS thin-film solar cells via charge transport layer engineering: SCAPS-1D simulations coupled with machine learning. *Results Surf. Interfaces* **2026**, *23*, 100813. [[CrossRef](#)]
94. Cui, J.; Sun, X.; Li, C.; Zhang, Y.; Zhou, C.; E, D. Whole-process performance prediction and optimization of CIGS solar cells via interpretable machine learning. *J. Energy Chem.* **2026**, *118*, 324–338. [[CrossRef](#)]
95. Barua, N.K.; Hall, E.; Cheng, Y.; Oliynyk, A.O.; Kleinke, H. Interpretable Machine Learning Model on Thermal Conductivity Using Publicly Available Datasets and Our Internal Lab Dataset. *Chem. Mater.* **2024**, *36*, 7089–7100. [[CrossRef](#)]
96. Fan, T.; Oganov, A.R. Combining machine-learning models with first-principles high-throughput calculations to accelerate the search for promising thermoelectric materials. *J. Mater. Chem. C* **2025**, *13*, 1439–1448. [[CrossRef](#)]
97. Minhas, H.; Sharma, R.K.; Pathak, B. Exploring Combined Stereochemically Active Lone-Pair and Rattling Effects in Thermoelectrics with Machine Learning Potentials. *J. Phys. Chem. Lett.* **2025**, *16*, 5051–5061. [[CrossRef](#)] [[PubMed](#)]
98. Hariharan, M.; Eithiraj, R. Machine learning model used to predict DFT-computed thermoelectric parameters of chalcogenide halides CuHgSCl and CuHgSI. *Mater. Today Commun.* **2025**, *43*, 111705. [[CrossRef](#)]
99. Hu, J.; Wang, C.; Li, Q.; Sa, R.; Gao, P. Accelerated design of photovoltaic Ruddlesden–Popper perovskite Ca₆Sn₄S_{14–x}O_x using machine learning. *APL Mater.* **2020**, *8*, 111109. [[CrossRef](#)]
100. Wang, Y.; Tang, J.; Li, G.; Zheng, J.; Song, X.; Wang, Q.; Cui, Z.; Cheng, L.; Guo, R. Anisotropic thermal transport in chalcogenide perovskite CaZrS₃ from machine learning interatomic potential. *Eng. Sci.* **2023**, *25*, 952. [[CrossRef](#)]
101. Li, S.; Yao, K.; Peng, H.; Xie, Y.; Fu, X. Unsupervised machine learning guides the design of MnFe–CoNiMg high-entropy alloys for optimizing MoS₂ interface properties. *J. Mater. Res. Technol.* **2026**, *42*, 4154–4164. [[CrossRef](#)]
102. Xu, Q.; Xu, M.; Tang, S.; Yuan, S.; Xu, M.; Zhang, W.; Li, X.; Wang, Z.; Miao, X.; Wang, C.; et al. Machine learning for discrimination of phase-change chalcogenide glasses. *InfoMat* **2025**, *7*, e70006. [[CrossRef](#)]
103. Klinkov, V.A.; Pavlov, E.A.; Semencha, A.V. Machine learning models for predicting density and glass transition temperature of chalcogenide glasses: Comparison and validation on novel compositions. *Materialia* **2026**, *48*, 102771. [[CrossRef](#)]

104. Rahman, H.M.; Biswas, M.; Mannodi-Kanakkithodi, A. DeFecT-FF: A machine learning force field framework for high throughput defect modeling in CdTe-based solar cells. *Phys. Chem. Chem. Phys.* **2026**, *28*, 10718–10730. [[CrossRef](#)] [[PubMed](#)]
105. de Vecchy, J.; Tortai, J.-H.; Besacier, M.; Le Cunff, D.; Pelissier, B. Hybrid metrology investigation combining Raman and ellipsometry spectroscopy applied to in line GeSbTe crystallization measurements and deep learning approaches for accurate prediction. *Microelectron. Eng.* **2025**, *300*, 112345. [[CrossRef](#)]
106. Draxl, C.; Scheffler, M. The NOMAD Laboratory: From Data Sharing to Artificial Intelligence. *J. Phys. Mater.* **2019**, *2*, 036001. [[CrossRef](#)]
107. Choudhary, K.; Garrity, K.F.; Reid, A.C.E.; DeCost, B.; Biacchi, A.J.; Walker, A.R.H.; Trautt, Z.; Hatrick-Simpers, J.; Kusne, A.G.; Centrone, A.; et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Comput. Mater.* **2020**, *6*, 173. [[CrossRef](#)]
108. Hastrup, S.; Strange, M.; Pandey, M.; Deilmann, T.; Schmidt, P.S.; Hinsche, N.F.; Gjerding, M.N.; Torelli, D.; Larsen, P.M.; Riis-Jensen, A.C.; et al. The Computational 2D Materials Database: High-throughput modeling and discovery of atomically thin crystals. *2D Mater.* **2018**, *5*, 042002. [[CrossRef](#)]
109. Raccuglia, P.; Elbert, K.C.; Adler, P.D.F.; Falk, C.; Wenny, M.B.; Mollo, A.; Zeller, M.; Friedler, S.A.; Schrier, J.; Norquist, A.J. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73–76. [[CrossRef](#)] [[PubMed](#)]
110. Ahmed, M.A.O.; Alotaibi, H.; Gami, F.; Fangary, M.M. Bridging laboratory findings and artificial intelligence for the design of TlInTe₂ crystals. *Sci. Rep.* **2026**, *16*, 15858. [[CrossRef](#)] [[PubMed](#)]
111. Sakib, N.; Alam, M.R.; Paul, S.; Neshani, S.; Momeni, K. From graphene to diamane: How interatomic potentials shape the transition. *Comput. Mater. Sci.* **2025**, *254*, 113927. [[CrossRef](#)]
112. Morrow, J.D.; Gardner, J.L.A.; Deringer, V.L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **2023**, *158*, 121501. [[CrossRef](#)] [[PubMed](#)]
113. Zhang, L.; Csányi, G.; van der Giessen, E.; Maresca, F. Efficiency, accuracy, and transferability of machine learning potentials: Application to dislocations and cracks in iron. *Acta Mater.* **2024**, *270*, 119788. [[CrossRef](#)]
114. Leeman, J.; Liu, Y.; Stiles, J.; Lee, S.B.; Bhatt, P.; Schoop, L.M.; Palgrave, R.G. Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis. *PRX Energy* **2024**, *3*, 011002. [[CrossRef](#)]
115. Yu, Y.; McCluskey, M.D. Classification of Semiconductors Using Photoluminescence Spectroscopy and Machine Learning. *Appl. Spectrosc.* **2022**, *76*, 228–234. [[CrossRef](#)] [[PubMed](#)]
116. Fu, J.; Liu, Y.; Wang, Q.; Pan, M.; Cheng, T. A Deep Learning Approach for Dynamic Modeling of Stimulated Raman Scattering in Chalcogenide Microstructured Optical Fibers. *IEEE Trans. Neural Netw. Learn. Syst.* **2026**. online ahead of print. [[CrossRef](#)] [[PubMed](#)]
117. Oviedo, F.; Ren, Z.; Sun, S.; Settens, C.; Liu, Z.; Hartono, N.T.P.; Ramasamy, S.; DeCost, B.L.; Tian, S.I.P.; Romano, G.; et al. Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. *npj Comput. Mater.* **2019**, *5*, 60. [[CrossRef](#)]
118. Wang, X.; Li, J.; Ha, H.D.; Dahl, J.C.; Ondry, J.C.; Moreno-Hernandez, I.; Head-Gordon, T.; Alivisatos, A.P. Auto-Detect-mNP: An Unsupervised Machine Learning Algorithm for Automated Analysis of Transmission Electron Microscope Images of Metal Nanoparticles. *JACS Au* **2021**, *1*, 316–327. [[CrossRef](#)] [[PubMed](#)]
119. Zhang, F.; Tamura, R.; Zeng, F.; Kozawa, D.; Kitaura, R. Bayesian Optimization for Controlled Chemical Vapor Deposition Growth of WS₂. *ACS Appl. Mater. Interfaces* **2024**, *16*, 59109–59115. [[CrossRef](#)] [[PubMed](#)]
120. Singh, N.; Samanta, K.; Maharana, S.K.; Pal, K.; Tretiak, S.; Talapatra, A.; Ghosh, D. High-throughput and data-driven search for stable optoelectronic AMSe₃ materials. *J. Mater. Chem. A* **2025**, *13*, 9192–9210. [[CrossRef](#)]
121. Tarekuzzaman, M.; Ishraq, H.M.; Sarker, A.S.; Hasan, Z.M.; Hasan, S.S. Next-generation solar absorber candidate rubidium palladium telluride: A study based on DFT, SCAPS-1D, and machine learning approaches. *J. Mater. Sci.* **2025**, *61*, 960–1002. [[CrossRef](#)]
122. Hassan, A.U.; Aljaafreh, M.J. A machine learning-quantum chemical quest to predict the photovoltaic properties of 2D molybdenum sulfide super structures. *Opt. Mater.* **2025**, *164*, 117056. [[CrossRef](#)]
123. Li, T.; Park, H.; Walsh, A. Data-Driven Crystal Structure Prediction for Ternary Metal Chalcogenides. *Chem. Mater.* **2026**, *38*, 161–170. [[CrossRef](#)] [[PubMed](#)]
124. Wu, Q.; Kang, L.; Lin, Z. A Machine Learning Study on High Thermal Conductivity Assisted to Discover Chalcogenides with Balanced Infrared Nonlinear Optical Performance. *Adv. Mater.* **2023**, *36*, 2309675. [[CrossRef](#)] [[PubMed](#)]
125. Sakib, N.; Alam, M.R.; Neshani, S.; Momeni, K. Thermal transport properties of WSe₂ monolayer with tilt grain boundaries. *FlatChem* **2025**, *54*, 100964. [[CrossRef](#)]
126. Xu, K.; Liang, T.; Zhang, Z.; Cao, X.; Han, M.; Wei, N.; Wu, J. Grain boundary and misorientation angle-dependent thermal transport in single-layer MoS₂. *Nanoscale* **2022**, *14*, 1241–1249. [[CrossRef](#)] [[PubMed](#)]
127. Isotta, E.; Jiang, S.; Möller, G.; Zevalkin, A.; Snyder, G.J.; Balogun, O. Microscale Imaging of Thermal Conductivity Suppression at Grain Boundaries. *Adv. Mater.* **2023**, *35*, 2302777. [[CrossRef](#)] [[PubMed](#)]

128. Song, K.; Ai, Q.; Guo, Y.; Xie, M.; Shuai, Y. Nonvolatile Tunable Wavelength-Selective Emitter with Phase-Changing Material $\text{Ge}_2\text{Sb}_2\text{Te}_5$ Designed by Bayesian Optimization Method. *Phys. Status Solidi (RRL)—Rapid Res. Lett.* **2023**, *17*, 2300099. [[CrossRef](#)]
129. Liu, R.; Hao, J.; Li, J.; Wang, S.; Liu, H.; Zhou, Z.; Delville, M.-H.; Cheng, J.; Wang, K.; Zhu, X. Causal Inference Machine Learning Leads Original Experimental Discovery in CdSe/CdS Core/Shell Nanoparticles. *J. Phys. Chem. Lett.* **2020**, *11*, 7232–7238. [[CrossRef](#)] [[PubMed](#)]
130. Wang, G.; Song, J.; Zhao, G.; Ding, Q.; Yin, T.; Wang, H. Tribological performance prediction of WS_2 coating under different conditions by machine learning. *Wear* **2023**, 532–533, 205092. [[CrossRef](#)]
131. Paul, S.; Torsi, R.; Robinson, J.A.; Momeni, K. Effect of the Substrate on MoS_2 Monolayer Morphology: An Integrated Computational and Experimental Study. *ACS Appl. Mater. Interfaces* **2022**, *14*, 18835–18844. [[CrossRef](#)] [[PubMed](#)]
132. Peters, L.; Ó Coileáin, C.; Dluzynski, P.; Siris, R.; Duesberg, G.S.; McEvoy, N. Directing the Morphology of Chemical Vapor Deposition-Grown MoS_2 on Sapphire by Crystal Plane Selection. *Phys. Status Solidi (A)* **2020**, *217*, 2000073. [[CrossRef](#)]
133. Park, Y.; Ahn, C.; Ahn, J.-G.; Kim, J.H.; Jung, J.; Oh, J.; Ryu, S.; Kim, S.; Kim, S.C.; Kim, T.; et al. Critical Role of Surface Termination of Sapphire Substrates in Crystallographic Epitaxial Growth of MoS_2 Using Inorganic Molecular Precursors. *ACS Nano* **2023**, *17*, 1196–1205. [[CrossRef](#)] [[PubMed](#)]
134. Momeni, K.; Ji, Y.; Nayir, N.; Sakib, N.; Zhu, H.; Paul, S.; Choudhury, T.H.; Neshani, S.; van Duin, A.C.T.; Redwing, J.M.; et al. A computational framework for guiding the MOCVD-growth of wafer-scale 2D materials. *npj Comput. Mater.* **2022**, *8*, 240. [[CrossRef](#)]
135. Tang, S.; Wang, Y.; Fiadziushkin, H.; Ding, Y.; Ghiami, A.; Heuken, M.; Vescan, A.; Kalisch, H. Real-Time Monitoring of 2D TMDC MOCVD: An In Situ Spectroscopic Reflectance Approach. *Adv. Mater. Interfaces* **2025**, *12*, e00349. [[CrossRef](#)]
136. Rajan, A.G.; Warner, J.H.; Blankshtein, D.; Strano, M.S. Generalized Mechanistic Model for the Chemical Vapor Deposition of 2D Transition Metal Dichalcogenide Monolayers. *ACS Nano* **2016**, *10*, 4330–4344. [[CrossRef](#)] [[PubMed](#)]
137. Tang, L.; Li, T.; Luo, Y.; Feng, S.; Cai, Z.; Zhang, H.; Liu, B.; Cheng, H.-M. Vertical Chemical Vapor Deposition Growth of Highly Uniform 2D Transition Metal Dichalcogenides. *ACS Nano* **2020**, *14*, 4646–4653. [[CrossRef](#)] [[PubMed](#)]
138. Choudhury, T.; Trainor, N.; Chen, C.; Chubarov, M.; Bachu, S.; Momeni, K.; Lundh, J.S.; Hickey, D.R.; Zhang, T.; Sebastian, A.; et al. Comparison of the MOCVD growth and properties of wafer-scale transition metal dichalcogenide epitaxial monolayers. *2D Mater.* **2025**, *12*, 045009. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.