Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
AI-Driven Chemicals and Materials Design—Article

Agentic Robotic Boxes for Perovskite Solar Cell Fabrication with Recipe Language Model

Zijian Chen [a,c,d,#], Wenjin Yu [b,h,#], Chuang Wu [c,#], Feibei Chen [a,c,#], Zixuan Wang [c,#], Chao Zhou [c], Yimeng You [c], Shaojie Li [c], Qiyuan Zhu [c], Ning Ma [a,c], Yao Sun [a,c], Donghui Li [c], Billy Fanady [c], Shengchou Jiang [c], Zhongliang Yan [e], Shumin Zhou [c], Liang Li [f], Chang-Yu Hsieh [g], Yang Bai [e], Lixin Xiao [h], Chi-yung Chung [a], Ching-chuen Chan [a], Zhanfeng Cui [i], Michael Grätzel [b,*], Haitao Zhao [a,c,*]

[a] *Research Centre for Materials Intelligent Manufacturing, State Key Laboratory of Ultra-Precision Machining Technology, Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University (PolyU), Hong Kong 999077, China*
[b] *Laboratory of Photonics and Interfaces, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland*
[c] *Physical AI Evolution (PAIEvo) Laboratory, AI for Energy Key Laboratory, PolyU–WIT Research Centre for Materials Intelligent Manufacturing, Wenzhou Institute of Technology (WIT), Wenzhou 325000, China*
[d] *Department of Chemical and Environmental Engineering, University of Nottingham Ningbo China, Ningbo 315100, China*
[e] *Faculty of Materials Science and Energy Engineering, Shenzhen University of Advanced Technology, Shenzhen 518107, China*
[f] *State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of New Energy, North China Electric Power University, Beijing 102206, China*
[g] *College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China*
[h] *State Key Laboratory of Artificial Microstructure and Mesoscopic Physics, School of Physics, Peking University, Beijing 100871, China*
[i] *Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK*

ARTICLE INFO

Article history:
Received 25 March 2026
Revised 3 April 2026
Accepted 7 April 2026
Available online xxxx

Keywords:
Robotics
Language agent
Recipe language model
Materials intelligence
Perovskite solar cells
Device fabrication

ABSTRACT

Perovskite solar cells (PSCs) have been undergoing rapid development with the vast combinatorial exploration of recipes; however, the related research suffers from time-consuming trial-and-error synthesis and labor-intensive fabrication. As a promising alternative, interconnected robotic boxes that integrate fabrication and characterization enable high-throughput experimentation and data collection; however, the resulting numerical datasets are often insufficiently analyzed and fail to provide effective feedback for semantic recipe optimization. Here, we conceived and realized an emerging scientific tool of robotic boxes enabled by a domain-specific recipe language model (RLM) and a coordinating language agent for PSCs research. The developed agent features two loops of seven artificial intelligence (AI) layers, in which both numerical and semantic recipes were continuously learned and optimized from the literature and robotic corpora for iterative fine-tuning of the RLM. Guided by the agent, 11 robotic boxes executed the controllable synthesis, fabrication, and characterization of 50 764 PSCs, increasing the power conversion efficiency (PCE) to 27.0% (26.5% certified). Simultaneously, more than 578 million tokens were generated and augmented to improve the ability to recommend a recipe and mechanistic reasoning, achieving an overall score of about 80% based on the dedicated evaluation criteria. Thus, such agentic robotic boxes provide an advanced tool for the next-generation synthesis, fabrication, characterization, and even mechanistic reasoning of PSCs and beyond.

1. Introduction

Over the past decade, perovskite solar cells (PSCs) have undergone rapid development [1], with more than a hundred thousand recipes explored to increase device performance [2]. However, the synthetic formulas are complex and span large chemical spaces for perovskite compositions and numerous additives [3–5]. Moreover, the environment-sensitive nature of perovskite crystallization significantly

* Corresponding authors.
E-mail addresses: michael.graetzel@epfl.ch (M. Grätzel), hai-tao.zhao@polyu.edu.hk (H. Zhao).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.eng.2026.04.002

complicates the fabrication process [6,7], and the correlated physical and chemical mechanisms remain underexplored [8,9]. Hence, the fabrication process must shift from conventional methods to next generation scientific tools that integrate robotic hardware for systematic high-throughput synthesis, fabrication, and characterization [10].

To date, various robotic systems, such as mobile robots [11,12], robotic platforms [13,14], synthesis machines [15–17], portable platforms [18,19], and self-driving laboratories [20], have been introduced for materials synthesis and characterization. These systems can investigate individual aspects of PSCs, including the perovskite composition and crystallization [21–23], antisolvents [24,25], and hole-transport materials [26,27]. However, for the fabrication of whole multi-layered PSC devices, fully enclosed and interconnected robotic boxes remain unexplored. This limitation primarily arises from the lack of techniques that can handle the exponentially increasing datasets and provide effective and timely feedback for fabrication-recipe optimization and mechanistic reasoning at the PSC device level.

Numerous artificial intelligence (AI) models that enable robotic systems to conduct data-driven PSC research, including theoretical calculations [28], systematic statistics [29], multiple linear regression (LR) with expectation maximization [30], gradient boosting decision tree [23], Gaussian process regression [31], and Bayesian optimization [32], have been developed to bridge the gap between extremely fast robotic automated data collection and the limited analytical capability. However, the coupling between these AI models and robotic platforms is lacking in terms of interaction and reasoning [33], hindering the exploration of recipes with multiple variables and comprehensive mechanisms. Recently, the impressive reasoning capability of large language models (LLMs) has been documented in areas such as mathematics [34], programming [35], rare disease diagnosis [36], and general chemical research [37]; however, these advances with domain specific language models have not been applied with robotics to materials research, such as optoelectronic fabrication.

Here, we conceived and realized a novel scientific tool, integrating with 11 robotic boxes driven by a domain-specific recipe language model (RLM) for PSC research, which can handle complex tasks, including recipe recommendation, robotic execution, and mechanistic reasoning. Moreover, a language agent is developed with two loops of seven AI layers to operate the robotic boxes, in which both numerical and semantic recipes are continuously learned and optimized from the literature and robotic corpora for the gradual fine-tuning of the RLM. Formulas and parameters (FPs) from more than 60 000 PSC-related studies are first encoded into machine-readable recipes as initial RLM training sources. The recommendations are then generated and compiled into robotic instructions and iteratively executed across 11 robotic boxes. Following the fabrication of 50 764 PSCs, we achieved a maximum power conversion efficiency (PCE) of 27.0% (certified as 26.5%). During this research, high-throughput robotic characterizations were conducted within these boxes to generate and augment more than 578 million tokens, derived from tokenized FPs paired with their corresponding mechanisms. Through the closed loop of the RLM and robotic boxes, the abilities of recipe recommendation and mechanistic reasoning for PSCs are progressively improved.

2. Results

2.1. Framework of agentic robotic boxes

In our framework, the interdisciplinary workflow starts from encoded FPs as recipe inputs (Fig. 1(a)), proceeds through seven AI layers with the language agent to train a domain-specific RLM (Fig. 1(b)), executes synthesis and fabrication within 11 interconnected robotic boxes (Fig. 1(c)), and produces *in-situ* characterization and device performance assessment as mechanistic outputs (Fig. 1(d)). Together, the encoded recipes, RLM, robotics, and characterized results form a closed recommendation–synthesis–fabrication–characterization–mechanism loop for exploring the large space of the recipes and their underlying mechanisms.

Specifically, the language agent first encodes the FPs extracted from more than 60 000 PSC-related studies published over the past decade (Fig. 1(a)). Each FP is defined as a discrete unit within the encoded recipes. For instance, F_a^1 represents the cesium (Cs) mole fraction in the perovskite composition, and P_a^1 corresponds to the first-step spin-coating speed during perovskite deposition (Table S1 in Appendix A). The language agent then encoded these machine-readable recipes into structured FP sequences, which were translated into tokens for subsequent fine-tuning of the RLM and execution by the robotic boxes. Compared with traditional data-mining approaches [38,39], our RLM-centric agent provides a straightforward approach to transfer literature-derived recipes in a format that is both RLM-readable and robot-executable, enabling direct interactions between the RLM and robotic boxes.

Building on these tokenized recipes as inputs, we constructed a seven-layer AI architecture operated by a language agent to train the RLM for closed-loop recipe discovery, connecting literature-based and robot-executed experiments (Fig. 1(b)). Given the vast exploration space and the coupled chemical and physical mechanisms involved, the ability of the language model to perform both recipe recommendation and the corresponding mechanistic reasoning is particularly important. The framework operates through two synergistic loops: loop I (literature–RLM), enabling effective RecipeQA corpora extraction from the literature, and loop II (robotics–RLM), allowing real-time accumulation of experimentally grounded recipe–mechanism pairs from robotic execution (Fig. S1 in Appendix A). Although both loops are structured with the same seven layers of learning, generating, RecipeQA, fine-tuning, reasoning, evaluation, and optimization, they differ in their data sources and forms of feedback provided to the RLM.

For the learning layer, the language agent first learns from both the literature and robotic experiments by extracting and encoding FPs with corresponding characterized mechanisms into recipes. For the generating layer, the dedicated language agent produces structured textual outputs for each recipe instance, including literature-derived recipe (L-Recipe) reports and robotic recipe (R-Recipe) reports that comprise FPs, fabrication details, mechanistic descriptions, an optimization summary, and supporting information (Tables S2 and S3 in Appendix A). These recipes are subsequently transcribed and organized into a training corpus as the RecipeQA. Under the coordination of the language agent, domain knowledge from the evolving RecipeQA is incrementally incorporated into a general-purpose foundation model (Qwen3-30B-A3B-Thinking-2507) through parameter-efficient fine-tuning methods such as low-rank adaptation (LoRA) to form a domain-specific RLM [40]. In the reasoning layer, the RLM then generates new recipe recommendations that are physics-informed and mechanistically grounded. These recommended recipes are fed into the evaluation layer, where both numerical and semantic recipes are scored and ranked based on the evaluation criteria. Preference-aligned recipe–mechanism pairs are then incorporated into the optimization layer using methods such as direct preference optimization (DPO; Section S1 in Appendix A), which reinforces preferred recipe–mechanism pairs without requiring explicit reward functions [41]. Through the continuous operation of loop I and loop II, the seven AI layers progressively optimize the RLM, refining the initially weakly aligned foundation model from text-based prior knowledge into a comprehensive mechanism-grounded rea-

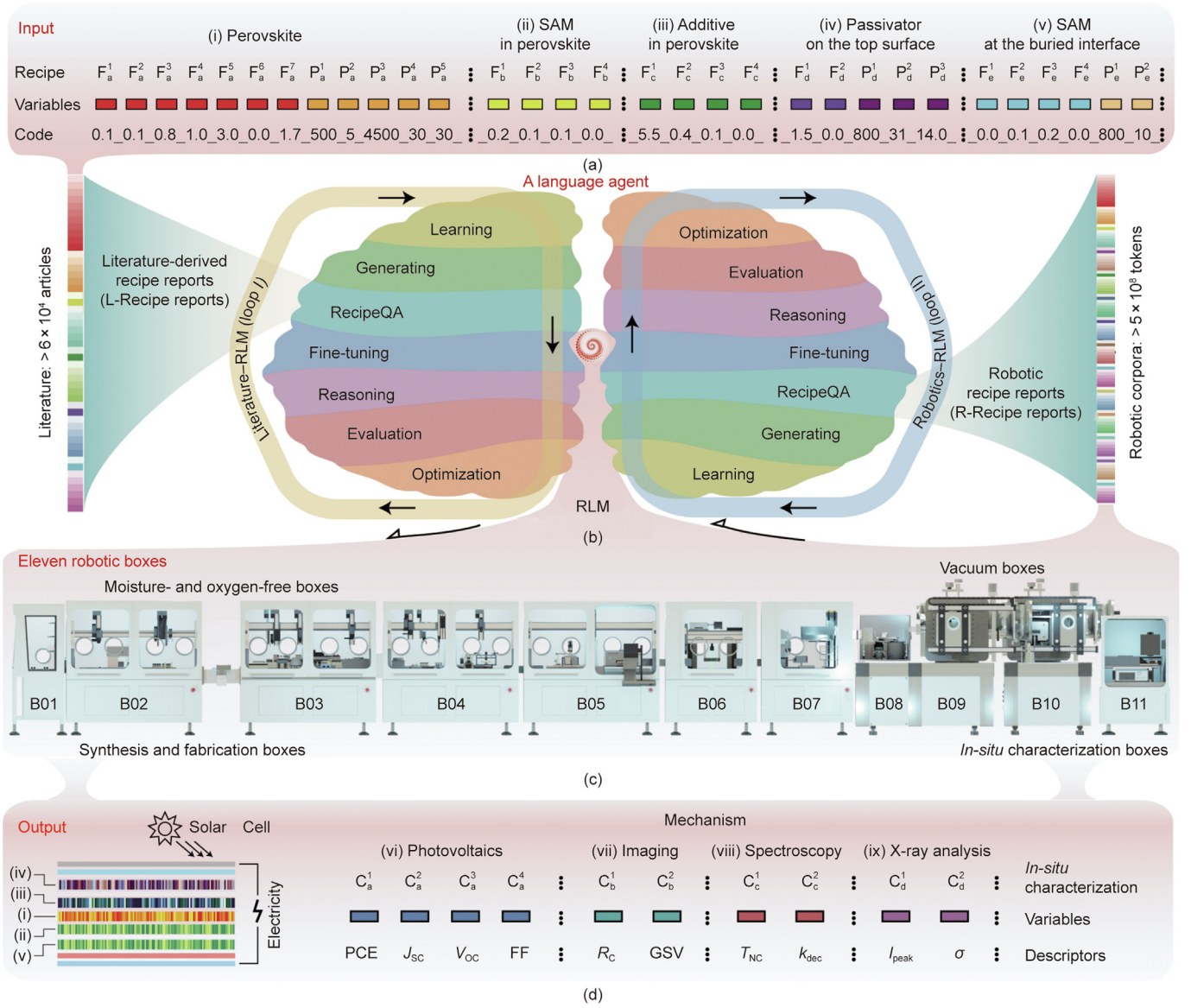

Fig. 1. Framework of 11 robotic boxes with a language agent for the fabrication and characterization of PSCs. (a) Input by encoding FPs as recipes for the synthesis and fabrication of (i) perovskites, (ii) self-assembled molecules (SAMs) and (iii) additives in perovskites, (iv) passivators on the top surface of perovskite layer, and (v) SAMs at the buried interface of perovskite layer. Each recipe is represented as a discrete, structured unit that is both RLM-readable and robot-executable, with FPs explicitly defined. (b) Software architecture of the seven AI layers used to train a domain-specific RLM with the recipes from the literature (loop I) and robotic boxes (loop II). (c) Hardware infrastructure of 11 interconnected robotic boxes, operated by the RLM: B01–B07 for moisture- and oxygen-free solution processed synthesis and fabrication; B08–B10 with a high vacuum for sequential layer deposition; and B11 for *in-situ* characterization. (d) Robotic characterizations of PSCs as outputs from (vi) photovoltaics, (vii) imaging, (viii) spectroscopy, and (ix) X-ray analysis. C_a^1-C_d^2: the PCE result under photovoltaic characterization; J_{SC}: short-circuit current density; V_{OC}: open-circuit voltage; FF: fill factor; R_C: optical coverage ratio; GSV: grayscale value; T_{NC}: time elapsed from annealing to the point at which the photoluminescence (PL) intensity reaches its maximum; k_{dec}: slope of the PL decay from the peak to 30% intensity; I_{peak}: maximum diffracted intensity of the X-ray diffraction (XRD) reflection; σ: residual stress determined from the grazing-incidence X-ray diffraction (GIXRD)-measured lattice strain using the $\sin^2\psi$ method.

soning model. The trained RLM therefore not only interprets and executes experiments but also continuously generates structured recipe–mechanism reports for substantial analysis.

The integration of the RLM, language agent, and robotic boxes (Fig. 1(c)) thus forms an integrated AI–robotics framework, in which the AI layers provide a recipe recommendation, mechanistic reasoning, and decision-making, while the robotic modules execute the corresponding synthesis, fabrication, and characterization (Fig. S2 in Appendix A). Closed-loop optimization is implemented at the software level through continuous model training, while the robotic modules are integrated to support subsequent synthesis, fabrication, and characterization. Within this intelligent framework, the crystallization dynamics, surface morphology, crystallography, and optical spectroscopy of the perovskite films, as well as the corresponding device performance, are systematically characterized (Fig. 1(d), Fig. S3 and Table S4 in Appendix A). The coupled robotic characterizations and continuously augmented mechanistic corpora eliminate manual sample transportation, sequential experiments in gloveboxes, fragmented data collection, and subjective discussions about mechanisms, thereby substantially accelerating the process of recipe generation and revealing the underlying mechanism.

2.2. Robotic boxes for controllable fabrication

We upgraded our previous robotic platform, which focused on automated synthesis, to include robotic boxes with full PSC device fabrication functionalities to precisely control the high-

dimensional recipe variables of the FPs required to reveal the underlying mechanism (Fig. 2(a)) [13]. A digital twin is introduced as a real-time software–hardware interface that links the RLM and language agent with the robotic boxes by translating model-generated recipes into executable commands while synchronizing experimental states and feedback (Fig. 2(b), Fig. S4 in Appendix A). The key features of this robotic hardware are a dedicated synthesis–fabrication–characterization streamline, a unique fully enclosed multi-box architecture operating under an inert atmosphere, a modular design for interchangeable experimental tools, interconnected robotic transportation of consumables, and multi-purpose end-effectors (camera, sensors, and actuators). Specifically, 11 robotic boxes are equipped with 101 functional modules, more than 1500 components and 4300 commandable parameters (Tables S5–S7 in Appendix A). The main functionalities are high-throughput synthesis (chemical storage, solid sampling, and liquid handling), robotic fabrication (spin-coating, antisolvent dropping, thermal annealing, laser processing, device transport, vacuum exchange, and thin-film deposition), and multidisciplinary characterization. In this manner, the widely used glovebox-based manual experimental steps are reconstructed into coupled robotic executions. In practice, coordinating fragmented FPs across complex PSC protocols remains nontrivial.

Controllable fabrication is demonstrated through the results from more than 50 000 robotic experiments (Fig. 2(c)), in which the photovoltaic performance was validated against the corresponding RLM-recommended recipe. The four color-coded stages reflect the progressive exploration of the FP space. In stage I, without interface or additive engineering, the PCE spans a broad range (0 to 17.4%). The incorporation of the SAM and additives in stage II narrows the distribution and increases the PCE to approximately 23%. In stage III, interfacial post-treatment passivation produced a remarkable increase, resulting in a PCE of 25.6% (Fig. S5(a) in Appendix A). The final configuration in stage IV, which combines SAM-based hole transport layers with targeted additive and passivation strategies, delivered a PCE of 27.0% (certified as 26.5%, Figs. S5(b) and S6 in Appendix A), which was comparable to the record efficiencies achieved by proficient and professional researchers [42], with more than 20 different recipe variables.

Controllable fabrication is further visualized through an importance–performance analysis to show how systematic FP variations influence device performance, including the PCE, fill factor (FF), open-circuit voltage (V_{OC}), and short-circuit current density (J_{SC}) (Fig. 2(d), Table S8 and Section S2 in Appendix A). In the first stage, the analysis revealed an unfocused random exploration of the recipe space, with perovskite formula variables such as the iodide ion (F_a^5) repeatedly adjusted, consistent with the wide performance distribution and limited controllability observed in early RLM recommendations. From stage II to stage III, the importance map contracted, demonstrating that the robotic system shifted from random exploration to targeted adjustments of key variables, including (4-(3,6-dimethyl-9*H*-carbazol-9-yl)butyl)phosphonic acid (Me-4PACz) SAMs (F_b^9), methylammonium chloride (MACl) additives (F_c^5), and later passivation-related variables such as phenethylammonium iodide (PEAI; F_d^{11}). The concentration of importance into a smaller set of variables, together with the reduced performance variance in Fig. 2(c), indicates the transition of the controllable fabrication process from stochastic exploration to deterministic optimization. In the final stage, where the FP variables were effectively stabilized, optimization focused on decoupled SAM engineering. The stable and high values for the few remaining active variables confirmed their deterministic influence and the precision of robotic control, with the optimized device performance consistently residing in the high-performance regime, confirming that fabrication becomes predictable and controllable. This emergence of predictable control stems from reduced recipe dimensionality combined with targeted, high-impact FP adjustments enabled by the cooperative interaction between the RLM and robotic boxes.

2.3. Robotic characterization augmented mechanism

We integrated physical and chemical characterization, including crystallization dynamics, surface morphology, crystallography, and optical spectroscopy, into the robotic fabrication workflow to facilitate recipe optimization and achieve mechanistic reasoning. Mechanistic reasoning is enabled by organizing characterization descriptors extracted from high-throughput robotic characterizations into a structured RecipeQA (Fig. 3). Specifically, all numerical values reported in the recipe reports, including formulas, parameters, and characterization results, were directly derived from collected measurements obtained from the robotic experiments. Through continuous coordination with ever-changing FPs within agentic robotic boxes, these descriptors are standardized, integrated, and streamed as unified datasets for subsequent mechanistic reasoning. We first integrated visual imaging to quantify the final perovskite films in terms of coverage and grayscale values (stage I, Section S3 in Appendix A). As precursor formulas evolved from imbalanced stoichiometry and suboptimal concentrations toward optimized compositions, the film coverage increased while the grayscale intensity decreased, indicating that the wet films developed a more stable viscosity and more homogeneous nucleation behavior [43] (Fig. 3(a), Figs. S7 and S8 in Appendix A). Combined with spin-coating parameters, these descriptors generate mechanistic reports (R-Recipe reports) linking FPs to film formation through the regulation of solvent removal and supersaturation dynamics.

We extended the mechanistic scope from final-state visual observation to crystallization quality by integrating crystallographic and optical spectroscopic characterization (stage II). X-ray diffraction (XRD) pattern revealed a distinct residual PbI_2 peak (12.6°) at the beginning of stage II, which weakened and broadened upon the introduction of crystallization modulators (Fig. 3(b), Fig. S9 and Section S4 in Appendix A). These results connect additive chemistry to crystallization control: Aromatic ammonium salts suppress residual PbI_2 by reacting with it and forming low-dimensional perovskites [44], whereas chloride modulators induce ordered crystallization by forming intermediate complexes [45]. Furthermore, *in-situ* photoluminescence (PL) spectroscopy revealed the nucleation dynamics. Both the explosive nucleation time and the PL-decay slope decreased from the SAM-free to the single-SAM and co-SAM configurations, indicating progressively accelerated and synchronized heterogeneous nucleation [46] (Fig. 3(c), Figs. S10 and S11 and Section S5 in Appendix A). Although device performance is more directly governed by the hole selectivity of the interfacial SAM [47], variations in these descriptors indicate that perovskite crystallization can be significantly modulated by the properties of the preferentially precipitated SAM layer [48]. Passivators for the perovskite top surface were subsequently introduced in stage III to further reduce interfacial energy loss. High-throughput XRD and grazing-incidence X-ray diffraction (GIXRD) were used to determine the surface phase configuration and lattice stress, revealing that the formation of surface-confined two-dimensional (2D) perovskite phases (layer number $n = 1$ and $n = 2$) was accompanied by a transition of residual stress from tensile to mildly compressive states (Fig. 3(d), Figs. S12 and S13 in Appendix A). These measurements directly link surface chemistry to strain relaxation: Aromatic ammonium passivators induce the formation of surface 2D layers that terminate lattice growth and relax strain, whereas diammonium salts

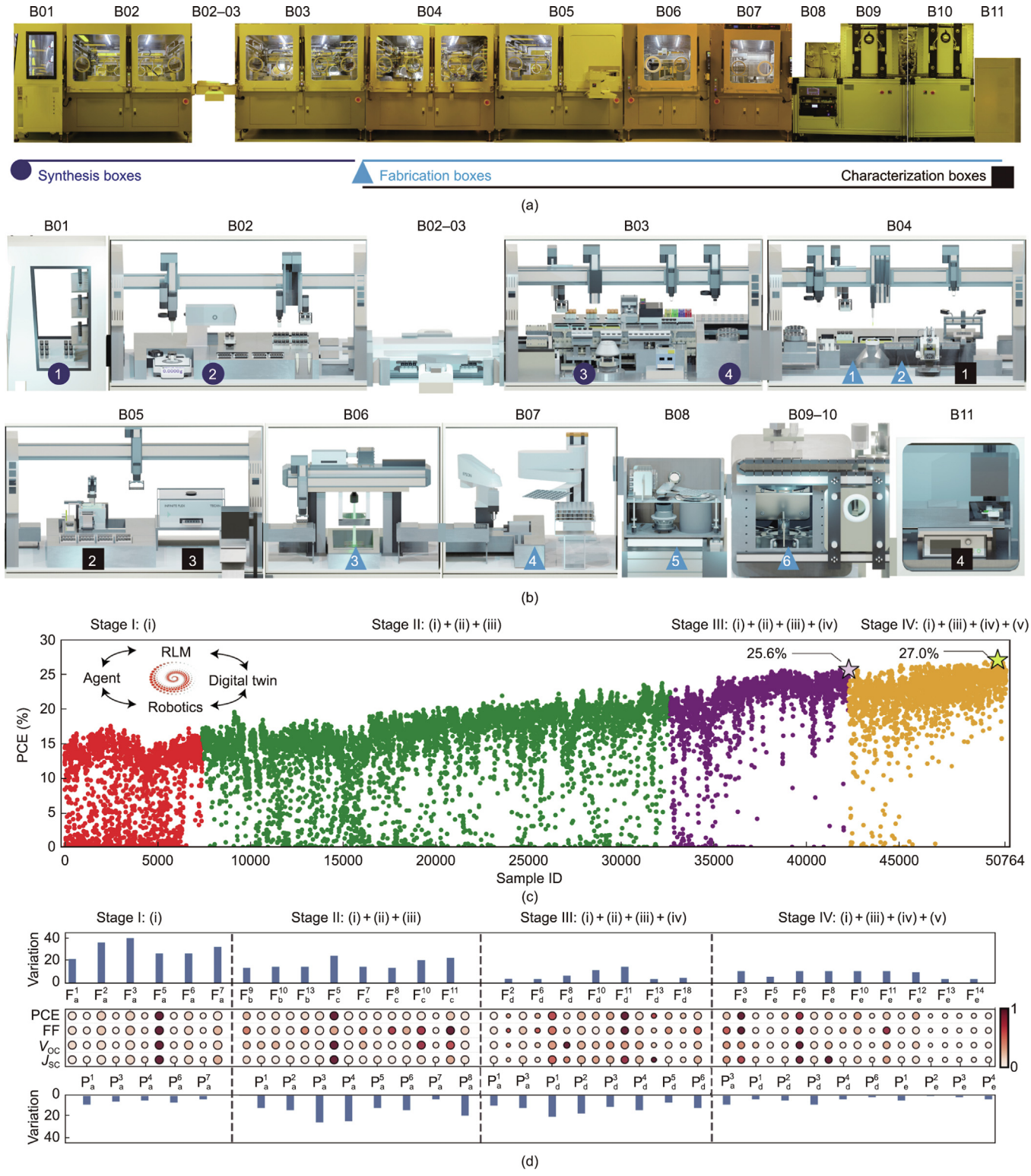

Fig. 2. Eleven robotic boxes enabled the controllable fabrication of PSCs. (a) Photographs of the robotic boxes. (b) Schematic representation of the digital twin system. (c) PCEs from the robotic controllable fabrication. ID: identity. (d) Analysis of the importance and performance of validated robotic FPs (the histogram and size of the circle represent the validated FPs; the color of the circle represents the relative impact of each FP on photovoltaic performance, with darker and larger circles indicating higher-impact variables).

provide additional stabilization through dual-site coordination and stress buffering (Table S9 in Appendix A).

Apart from these dedicated characterization of perovskite films, we also integrated detailed intrinsic descriptors (i.e., V_{OC}, J_{SC}, and FF) in the current density–voltage (J–V) test of the devices to further improve the reasoning chain in R-Recipe reports (Figs. 3(e)–(g)). The optimization of the FPs of the perovskite in stage I primarily increases J_{SC} by approximately 30% from 17.0 to 22.9 mA·cm^{-2} (Fig. 3(f)). This substantial improvement arises from the pronounced increase in overall film quality, as reflected by the sys-

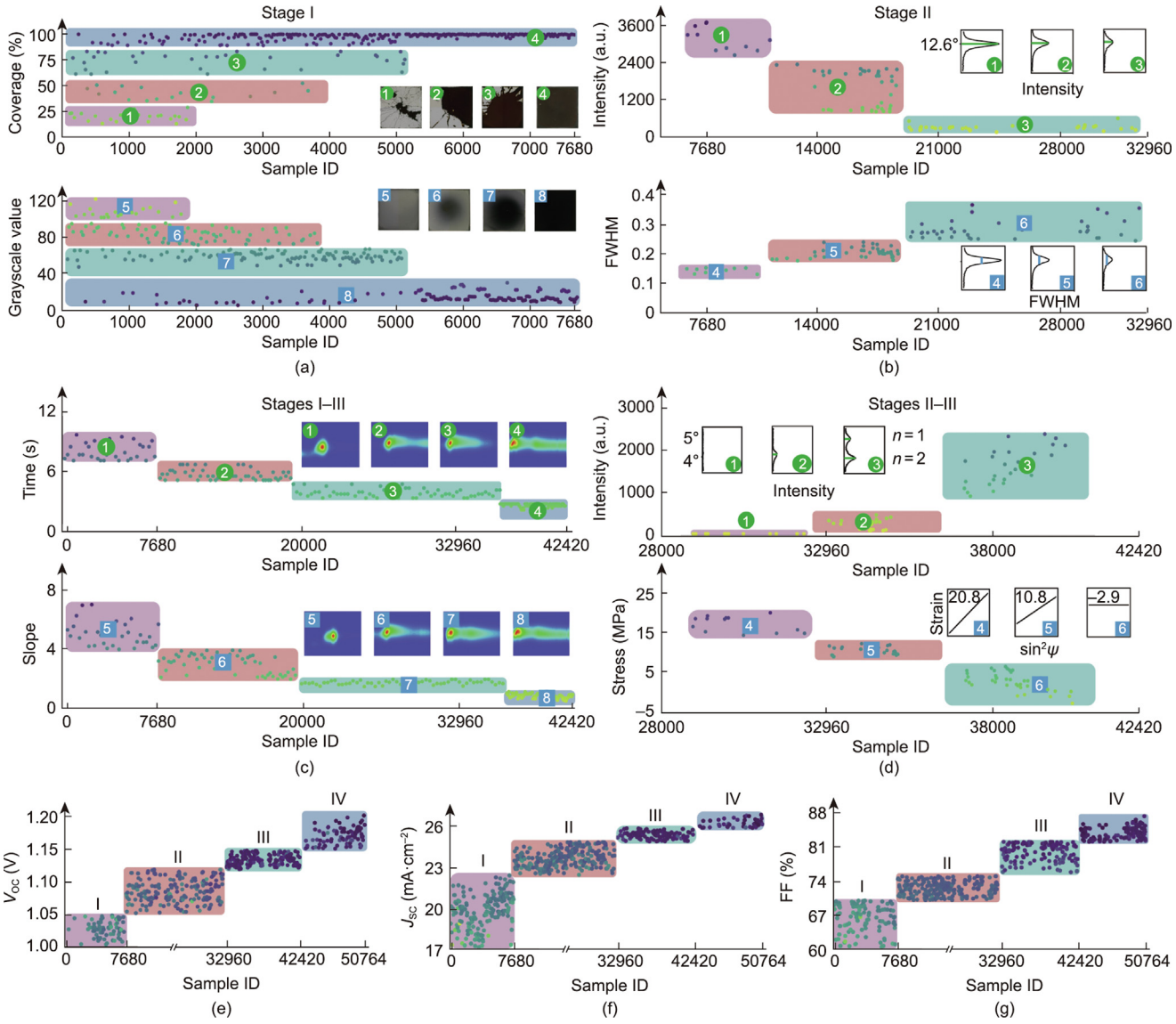

Fig. 3. Robotic characterization of PSCs for mechanistic studies and corpora augmentation. High-throughput robotic characterization of (a) the coverage and grayscale values of perovskite thin film, (b) XRD peak intensity and full width at half maximum (FWHM) of PbI_2, (c) explosion nucleation time and PL-decay slope during film formation, (d) XRD peak intensity of two-dimensional (2D) perovskite and residual stress of thin films, (e) V_{OC}, (f) J_{SC}, and (g) FF of PSCs. Background colors denote different recipe variables in panels (a–d) and optimization progress in the panels.

tematic increase in film coverage and decrease in the grayscale value. The introduction of SAM and additives in stage II leads to a pronounced increase in the V_{OC} from 1.05 V to approximately 1.13 V (Fig. 3(e)). The reduced impurity-phase segregation and improved phase uniformity mitigate the formation of recombination-active PbI_2-rich regions at grain boundaries and thereby reduce bulk nonradiative recombination losses. In addition, accelerated and synchronized heterogeneous nucleation further improves the film quality and thus reduces the number of grain boundary-associated trap states. The introduction of surface passivators in stage III resulted in a clear increase in the FF from 75% to approximately 82% (Fig. 3(g)). The formation of a 2D/three-dimensional (3D) perovskite structure suppresses surface defect states and establishes an energetic barrier, thereby reducing surface recombination and mitigating interfacial carrier accumulation [49]. In addition, the alleviation of tensile stress reduces strain-induced defect formation and suppresses stress-driven nonradiative recombination pathways, which distort the *J*–*V* characteristics and directly contribute to the substantial increase in the FF [50]. In stage IV, the decoupled SAM interlayer is implemented at the buried interface, increasing V_{OC} from 1.15 V to approximately 1.20 V (Fig. 3(e)). Defects at the buried interface are passivated, and electrons are eliminated at the hole-selective contact, thereby suppressing interfacial nonradiative recombination and enabling V_{OC} to approach its upper limit. These mechanistic descriptions were generated by the RLM in the reasoning layer, where the model performs RLM-guided reasoning using distilled chemical and physical knowledge from the generating and RecipeQA layers. As a result, high-throughput robotic characterizations transform controlled fabrication outcomes into mechanism-resolved evidence that is systematically structured into an expanded mechanistic corpus, enabling the RLM to upgrade base on experimentally validated relationships between recipe variables and device performance.

2.4. Robotic recipes for language model training

By inputting substantial domain expertise, both numerical and semantic recipes are essential for accelerating the discovery of PSCs. Through continuous closed-loop interactions between the robotic boxes and the RLM-coordinated AI layers, the recipes are systematically evaluated and optimized, enabling progressive training of the domain-specific RLM to power the robotic boxes. Within this model, the recipes expand from numerical type to semantic type with a mechanism through robotic experimentation, and the capability of the RLM advances from factual recall to mechanism-grounded reasoning driven by progressively enriched RecipeQA. Together, these processes transform recipe accumulation into mechanism-informed domain knowledge, resulting in a self-improving loop that unites FPs and the mechanistic understanding of PSCs.

We divide the language model training into four stages based on the progressively integrated characterizations and agentic robotic system modifications (Fig. 4(a)). In stage I (sample identities (IDs) 1–7680), the system operated without reasoning and optimization layers, focusing primarily on reproducing literature-reported experiments and exploring initial recipes to improve the PCE of the device. Although partial characterization data, such as film coverage and grayscale values, were incorporated, they served only as supplementary descriptors within the evaluation layer. Consequently, the recipe recommendations at this stage relied on direct extrapolation from historical experimental data, without an explicit mechanism or goal-oriented optimization, as optimization was guided by observed performance changes rather than mechanistic interpretations. Under these conditions, extreme gradient boosting (XGB), support vector regression (SVR), and LR were used for PCE prediction (Fig. 4(b), Section S6 in Appendix A), rather than a single model, because they represent different modeling assumptions, from linear to increasingly nonlinear relationships, thereby reducing model-specific bias and improving robustness. In addition, the FPs were restricted to perovskite-related number of variable (formula: 7; parameter: 8), which together constrained the achievable PCE to less than 18% (Table S10 in Appendix A). Training at this stage was dominated by the RecipeQA without mechanism (Fig. S14 in Appendix A), establishing only direct mappings between FPs and device performance. As a result, the RecipeQA corpus expanded to 16 million tokens (Fig. 4(c)); however, the increase in scale did not translate into a qualitative improvement in recipe recommendations and mechanistic reasoning. Despite modest improvements in perovskite film quality, the RLM lacked a mechanistic understanding or interpretative capability regarding perovskite crystallization, underscoring the necessity of introducing explicit reasoning and optimization layers and mechanism-aware supervision in subsequent stages.

Stage II (sample IDs 7681–32960) incorporated the full seven AI layers, enabling the extension of mechanistic knowledge beyond the literature through elaborate crystallographic and optical spectroscopic characterizations. At this stage (Fig. 4(a)), the FPs were expanded to include SAMs and additives in the perovskite precursor (formula: 38; parameter: 8), aiming to increase phase purity and regulate perovskite crystallization. The RecipeQA was upgraded from numerical descriptions to coherent, method-style process narratives. The reasoning layer organized discrete FPs into logically consistent fabrication workflows, while lightweight mechanistic knowledge, such as defect passivation, crystallization modulation, and interfacial recombination suppression, was introduced to contextualize parameter modifications. These refinements resulted in a best PCE of approximately 23%, whereas cumulative RecipeQA augmentation expanded the training corpus to approximately 104 million tokens (Fig. 4(c)). The improved recipe is contributed by the combined introduction of reasoning and optimization layers, expansion of the tokenized FP, and augmentation of mechanism, thereby increasing the training tokens.

The subsequent stages (stage III: sample IDs 32961–42420; stage IV: sample IDs 42421–50764) further broadened the experimental space (formula: 70; parameter: 18) from bare perovskite crystallization to interfacial regulation through post-treatment strategies and individual SAM hole-selective layers, thereby increasing device performance to a level comparable with that in state-of-the-art reports. Training in these two stages focused on establishing relational chains that link FPs, multimodal characterizations, and device performance. RecipeQA incorporated structured mechanistic reasoning, guiding the RLM to generate coherent links from the recipe recommendation to device-level outcomes (Fig. S15 in Appendix A). Validated by consistent high-throughput robotic experimentation feedback to these relationships, a stabilized coefficient of determination (R^2) of 0.80 was achieved (Fig. 4(b)). Ultimately, the total token count reached 578 million (Fig. 4(c)), indicating the transition toward a device-scale and experimentally grounded optimization framework.

We established a dedicated set of criteria to assess the capability of the RLM from two complementary aspects: recipe recommendation and mechanistic reasoning (Table S11 in Appendix A). The quality of the recommended recipe was evaluated based on the integrity of the recipe, formula rationality, parameter rationality, and experimental validation (Table S12 in Appendix A). The mechanistic reasoning capability was assessed in terms of domain knowledge, mechanism integrity, interpretation, comprehensiveness, and coherence (Table S13 in Appendix A). Initial assessments of RLM-1.1 revealed a lack of domain-specific knowledge in the training corpus (Table S14 in Appendix A), with unsatisfactory reasoning scores of 9.0% and 5.4% (Fig. 4(d)). Following the integration of the reasoning and optimization layers in stage II, the mechanistic reasoning capability of RLM-1.2 markedly improved (reasoning score: 14.9%) due to the structured enrichment of the training corpus. The corresponding recipe recommendation performance also increased to 16.7%, benefiting from the expanded FP space and enhanced recipe integrity. However, the domain knowledge focused solely on perovskite crystallization, which was insufficient to further increase the device performance to state-of-the-art levels, and interfacial energetic losses remained inadequately analyzed (Table S15 in Appendix A). The incorporation of interfacial characterization and a statistical analysis of device-level descriptors were performed to address this limitation and introduced explicit mechanistic information, such as interfacial energetic alignment, carrier extraction efficiency, and nonradiative recombination pathways, thereby allowing RLM-1.3 to reach the expert level at the device scale. Consequently, recipe recommendation (35.7%) and mechanistic reasoning (45.2%) together yielded an overall score of about 80% under the evaluation criteria (Tables S12 and S13), as demonstrated by a representative sample (Table S16 in Appendix A). Based on this mechanism-resolved evidence, the RLM was progressively refined not only through experimentally validated relational chains but also through statistically derived preferences learned from large-scale datasets. The advanced RLM generated increasingly appropriate recipe recommendations (Fig. S16 in Appendix A), enabled by its enhanced recognition and interpretation of both knowledge from the literature and experimental outcomes. This closed-loop process resulted in progressively more stable, coherent, and mechanism-grounded recipe optimization. Together, these bidirectional interactions enhanced both the mechanistic reasoning capability of the RLM and the controllable fabrication performance of the robotic modules.

3. Conclusions

In this study, we conceived and realized robotic boxes for fabricating PSCs with a language agent as a unified framework for device-scale discovery. This framework addresses a central limita-

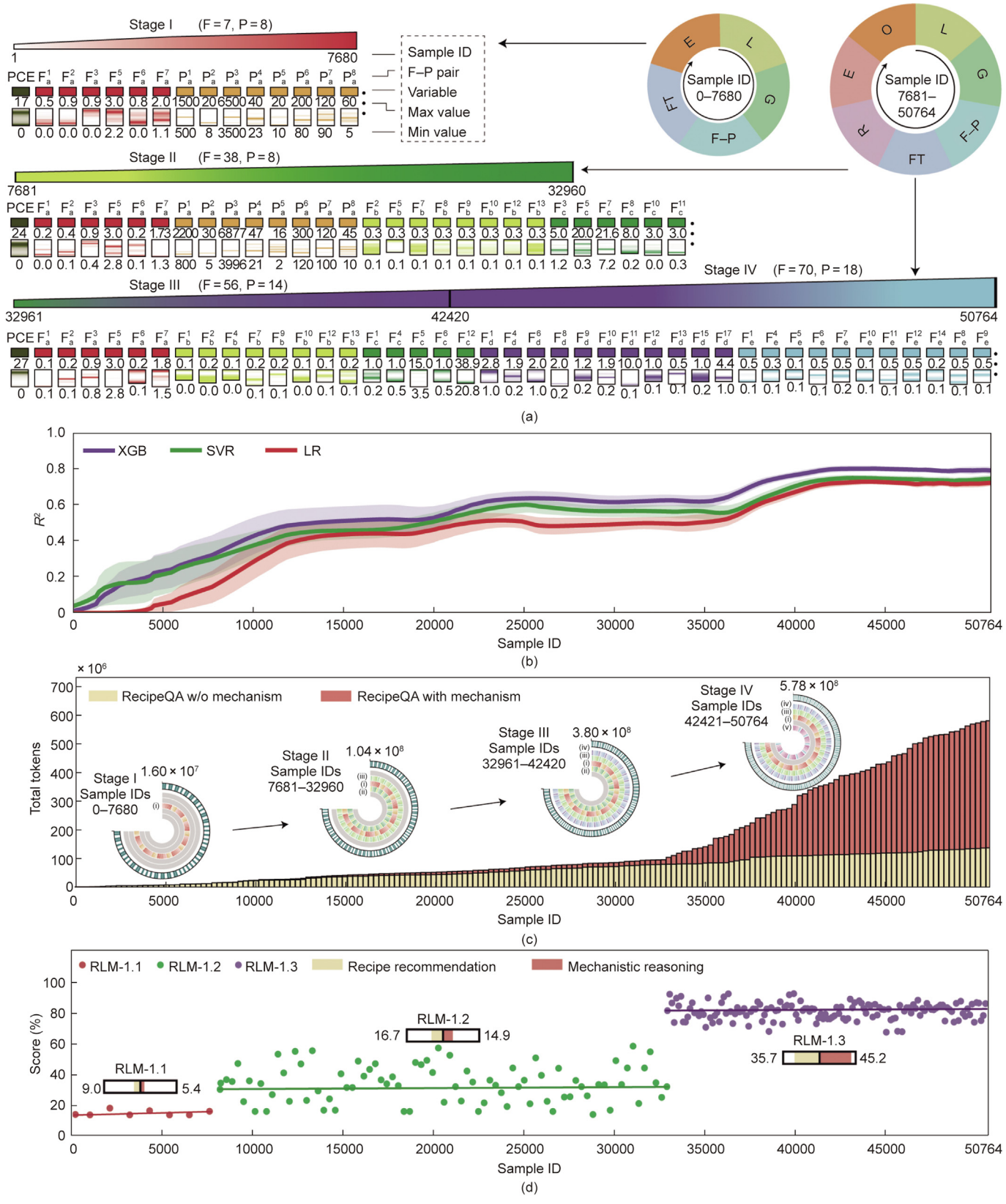

Fig. 4. Recipes (numerical and semantic) facilitated language model training. (a) Numerical FPs from the robotic boxes. Each colored block represents a component in the encoded recipe, with the color denoting the functional module (perovskite, SAM, additive, or passivator) and the shading intensity reflecting the sampling probability across its numerical range. F: formula; P: parameter; L: learning; G: generating; FT: fine-tuning; R: reasoning; E: evaluation; O: optimization. (b) Numerical evaluation of the predicted recipes by three types of algorithms: extreme gradient boosting (XGB), support vector regression (SVR), and LR. (c) Semantic recipes generated from robotic experimental loops I and II. w/o: without. (d) Semantic evaluation of the trained language model with the capabilities of recipe recommendation and mechanistic reasoning.

tion in PSC research, that the disconnect between high-throughput robotic experimentation and effective feedback for recipe optimization and mechanistic reasoning. By encoding FPs from over 60 000 PSC-related studies into machine-readable recipes, it enables closed-loop recommendation, synthesis, fabrication, characterization, and mechanism analysis at the device scale.

Three advances introduced in this study enable this framework. First, the robotic boxes achieved controllable fabrication of full PSC devices, executing 50 764 experiments and delivering a champion PCE of 27.0% (26.5% certified), which demonstrates the feasibility of device-scale fabrication under enclosed and coordinated robotic operation. Second, robotic characterization of PSCs augments the mechanism by transforming high-throughput experimental outputs into more than 500 million tokens of recipe–mechanism pairs, extending mechanistic corpora beyond the literature and shifting analysis from numerical descriptors to semantic mechanistic reasoning. Third, by learning from both literature-derived and robot-generated corpora through two loops of seven AI layers, the RLM evolved from basic aligned literature-based reasoning into mechanism-grounded reasoning.

In addition to PSC optimization, this approach changes how complex materials can be explored. Methodologically, it shifts device fabrication from trial processes toward language model reasoning-guided experimentation. At the software level, it introduces transferable recipes linking numerical and semantic variables with mechanisms. At the hardware level, it demonstrates fully enclosed, interconnected robotic modules capable of device-scale optimization with continuous mechanistic validation.

More broadly, this work highlights a paradigm shift from manual discovery, providing a scalable architectural foundation of materials intelligence. In the longer term, such AI and robotics systems could be deployed in extreme environments to support on-site materials intelligent manufacturing.

CRediT authorship contribution statement

Zijian Chen: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Data curation. **Wenjin Yu:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Chuang Wu:** Validation, Software, Methodology, Investigation. **Feibei Chen:** Writing – review & editing, Methodology, Investigation. **Zixuan Wang:** Writing – original draft, Visualization, Project administration, Investigation, Data curation. **Chao Zhou:** Visualization, Software. **Yimeng You:** Visualization, Software. **Shaojie Li:** Data curation. **Qiyuan Zhu:** Visualization, Software. **Ning Ma:** Visualization, Software. **Yao Sun:** Software. **Donghui Li:** Visualization. **Billy Fanady:** Writing – review & editing. **Shengchou Jiang:** Software. **Zhongliang Yan:** Data curation. **Shumin Zhou:** Supervision. **Liang Li:** Supervision, Methodology. **Chang-Yu Hsieh:** Writing – review & editing, Software. **Yang Bai:** Writing – review & editing. **Lixin Xiao:** Writing – review & editing, Supervision. **Chi-yung Chung:** Writing – review & editing, Conceptualization. **Ching-chuen Chan:** Writing – review & editing, Conceptualization. **Zhanfeng Cui:** Writing – review & editing, Conceptualization. **Michael Grätzel:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Haitao Zhao:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by XtaiPi Future Materials Pilot Platform, Future Materials AI Accelerator, the Robotic AI-Scientist Platform of Chinese Academy of Sciences, the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government, Wenzhou Key Laboratory of AI Energy and the Wenzhou Science and Technology Plan Project (G20240040 and ZG2024053), the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie Actions grant agreement (101281154).

We thank Meng Zhang, Shouying Mu, Yixiao Lei, Xidong Chen, Richeng Jin, Xianju Lu, Peng Liu, Zhengda He, Jinfei Ling, Zhaohui Bao, and Ziye Chen for their contributions to this work.

Data availability

The codes and pipelines used for this study are provided at GitHub (https://github.com/MI-Materials-Intelligence/Recipe-Language-Model).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2026.04.002.

References

[1] National Laboratory of the Rockies. Best research-cell efficiency chart. Golden: National Laboratory of the Rockies; 2026.
[2] Jacobsson TJ, Hultqvist A, García-Fernández A, Anand A, Al-Ashouri A, Hagfeldt A, et al. An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles. Nat Energy 2022;7(1):107–15.
[3] Zhang H, Pfeifer L, Zakeeruddin SM, Chu J, Grätzel M. Tailoring passivators for highly efficient and stable perovskite solar cells. Nat Rev Chem 2023;7(9):632–52.
[4] Tian W, Wang R, Yang D, Xue J. Organic A-cations in metal halide perovskite photovoltaics. Nat Rev Chem 2026;10(1):50–71.
[5] Maqsood A, Näsström H, Chen C, Qiutong L, Luo J, Chakraborty R, et al. Towards an interoperable perovskite description or how to keep track of 300 perovskite ions. Nat Commun 2025;16:8725.
[6] Zou Y, Yu W, Qu B, Chen Z, Wei M, Xiao L. Ambient fabrication of perovskites for photovoltaics. Nat Rev Mater 2025;10(6):400–2.
[7] Yang C, Hu W, Liu J, Han C, Gao Q, Mei A, et al. Achievements, challenges, and future prospects for industrialization of perovskite solar cells. Light Sci Appl 2024;13(1):227.
[8] Bai Y, Huang Z, Zhang X, Lu J, Niu X, He Z, et al. Initializing film homogeneity to retard phase segregation for stable perovskite solar cells. Science 2022;378(6621):747–54.
[9] Li S, Jiang Y, Xu J, Wang D, Ding Z, Zhu T, et al. High-efficiency and thermally stable $FACsPbI_3$ perovskite photovoltaics. Nature 2024;635(8037):82–8.
[10] Zhang X, Chen Z, Chen F, Fanady B, Wang B, Ni Z, et al. Material intelligence by the convergence of artificial intelligence and robotic platforms. Nexus 2025;2(3):100083.
[11] Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, et al. A mobile robotic chemist. Nature 2020;583(7815):237–41.
[12] Dai T, Vijayakrishnan S, Szczypiński FT, Ayme JF, Simaei E, Fellowes T, et al. Autonomous mobile robots for exploratory synthetic chemistry. Nature 2024;635(8040):890–7.
[13] Zhao H, Chen W, Huang H, Sun Z, Chen Z, Wu L, et al. A robotic platform for the synthesis of colloidal nanocrystals. Nat Synth 2023;2(6):505–14.
[14] Zhang Z, Ren Z, Hsu CW, Chen W, Hong ZW, Lee CF, et al. A multimodal robotic platform for multi-element electrocatalyst discovery. Nature 2025;647(8089):390–6.
[15] Angelone D, Hammer AJS, Rohrbach S, Krambeck S, Granda JM, Wolf J, et al. Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. Nat Chem 2021;13(1):63–9.
[16] Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, et al. Organic synthesis in a modular robotic system driven by a chemical programming language. Science 2019;363(6423):eaav2211.
[17] Angello NH, Rathore V, Beker W, Wołos A, Jira ER, Roszak R, et al. Closed-loop optimization of general reaction conditions for heteroaryl Suzuki–Miyaura coupling. Science 2022;378(6618):399–405.
[18] Manzano JS, Hou W, Zalesskiy SS, Frei P, Wang H, Kitson PJ, et al. An autonomous portable platform for universal chemical synthesis. Nat Chem 2022;14(11):1311–8.

[19] Rohrbach S, Šiaučiulis M, Chisholm G, Pirvan PA, Saleeb M, Mehr SHM, et al. Digitization and validation of a chemical synthesis literature database in the ChemPU. Science 2022;377(6602):172–80.
[20] Wu T, Kheiri S, Hickman RJ, Tao H, Wu TC, Yang ZB, et al. Self-driving lab for the photochemical synthesis of plasmonic nanoparticles with targeted structural and optical properties. Nat Commun 2025;16:1473.
[21] Zhang J, Wu J, Zhao Y, Zou Y, Barabash A, Wu Z, et al. Revealing the crystallization and thermal-induced phase evolution in aromatic-based quasi-2D perovskites using a robot-based platform. ACS Energy Lett 2023;8(8):3595–603.
[22] Higgins K, Valleti SM, Ziatdinov M, Kalinin SV, Ahmadi M. Chemical robotics enabled exploration of stability in multicomponent lead halide perovskites via machine learning. ACS Energy Lett 2020;5(11):3426–36.
[23] Zhao Y, Zhang J, Xu Z, Sun S, Langner S, Hartono NTP, et al. Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning. Nat Commun 2021;12:2191.
[24] Higgins K, Ziatdinov M, Kalinin SV, Ahmadi M. High-throughput study of antisolvents on the stability of multicomponent metal halide perovskites through robotics-based synthesis and machine learning approaches. J Am Chem Soc 2021;143(47):19945–55.
[25] Gu E, Tang X, Langner S, Duchstein P, Zhao Y, Levchuk I, et al. Robot-based high-throughput screening of antisolvents for lead halide perovskites. Joule 2020;4(8):1806–22.
[26] Wu J, Torresi L, Hu M, Reiser P, Zhang J, Rocha-Ortiz JS, et al. Inverse design workflow discovers hole-transport materials tailored for perovskite solar cells. Science 2024;386(6727):1256–64.
[27] Wu J, Zhang J, Hu M, Reiser P, Torresi L, Friederich P, et al. Integrated system built for small-molecule semiconductors via high-throughput approaches. J Am Chem Soc 2023;145(30):16517–25.
[28] Xu J, Chen H, Grater L, Liu C, Yang Y, Teale S, et al. Anion optimization for bifunctional surface passivation in perovskite solar cells. Nat Mater 2023;22(12):1507–14.
[29] Deng C, Tang L, Luo P, Li H, Yang L, Liu Z, et al. Unveiling the statistical behaviors of metal–halide perovskites from films to devices through a high-throughput experimental platform. InfoMat 2026;8(1):e70039.
[30] Meftahi N, Surmiak MA, Fürer SO, Rietwyk KJ, Lu J, Raga SR, et al. Machine learning enhanced high-throughput fabrication and optimization of quasi-2D Ruddlesden–Popper perovskite solar cells. Adv Energy Mater 2023;13(38):2203859.
[31] Wang Y, Perea-Puente S, Le Corre VM, Wu Z, Sytnyk M, These A, et al. Hybrid learning enables reproducible >24% efficiency in autonomously fabricated perovskites solar cells. Adv Energy Mater 2026;16(4):e04340.
[32] Zhang J, Liu B, Liu Z, Wu J, Arnold S, Shi H, et al. Optimizing perovskite thin-film parameter spaces with machine learning-guided robotic platform for high-performance perovskite solar cells. Adv Energy Mater 2023;13(48):2302594.
[33] Tom G, Schmid SP, Baird SG, Cao Y, Darvish K, Hao H, et al. Self-driving laboratories for chemistry and materials science. Chem Rev 2024;124(16):9633–732.
[34] Romera-Paredes B, Barekatain M, Novikov A, Balog M, Kumar MP, Dupont E, et al. Mathematical discoveries from program search with large language models. Nature 2024;625(7995):468–75.
[35] Yuksekgonul M, Bianchi F, Boen J, Liu S, Lu P, Huang Z, et al. Optimizing generative AI by backpropagating language model feedback. Nature 2025;639(8055):609–16.
[36] Zhao W, Wu C, Fan Y, Qiu P, Zhang X, Sun Y, et al. An agentic system for rare disease diagnosis with traceable reasoning. Nature 2026;651(8106):775–84.
[37] Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. Nature 2023;624(7992):570–8.
[38] Vaucher AC, Zipoli F, Geluykens J, Nair VH, Schwaller P, Laino T. Automated extraction of chemical synthesis actions from experimental procedures. Nat Commun 2020;11:3601.
[39] Vaucher AC, Schwaller P, Geluykens J, Nair VH, Iuliano A, Laino T. Inferring experimental procedures from text-based representations of chemical reactions. Nat Commun 2021;12:2573.
[40] Hu E, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. LoRA: low-rank adaptation of large language models. In: Proceedings of the 10th international conference on learning representations, 2022 Apr 25–29, online conference; 2022.
[41] Rafailov R, Sharma A, Mitchell E, Ermon S, Manning CD, Finn C. Direct preference optimization: your language model is secretly a reward model. In: Proceedings of the 37th conference on neural information processing systems, 2023 Dec 10–16, New Orleans, LA, USA. Red Hook: Curran Associates, Inc.; 2023. p. 53728–41.
[42] Jiang W, Qu G, Huang X, Chen X, Chi L, Wang T, et al. Toughened self-assembled monolayers for durable perovskite solar cells. Nature 2025;646(8083):95–101.
[43] Zhang C, Park NG. Materials and methods for cost-effective fabrication of perovskite photovoltaic devices. Commun Mater 2024;5(1):194.
[44] Kim DH, Muzzillo CP, Tong J, Palmstrom AF, Larson BW, Choi C, et al. Bimolecular additives improve wide-band-gap perovskites for efficient tandem solar cells with CIGS. Joule 2019;3(7):1734–45.
[45] Park J, Kim J, Yun HS, Paik MJ, Noh E, Mun HJ, et al. Controlled growth of perovskite layers with volatile alkylammonium chlorides. Nature 2023;616(7958):724–30.
[46] Du S, Huang H, Lan Z, Cui P, Li L, Wang M, et al. Inhibiting perovskite decomposition by a creeper-inspired strategy enables efficient and stable perovskite solar cells. Nat Commun 2024;15:5223.
[47] Li Z, Sun X, Zheng X, Li B, Gao D, Zhang S, et al. Stabilized hole-selective layer for high-performance inverted p–i–n perovskite solar cells. Science 2023;382(6668):284–9.
[48] Tan Q, Li Z, Luo G, Zhang X, Che B, Chen G, et al. Inverted perovskite solar cells using dimethylacridine-based dopants. Nature 2023;620(7974):545–51.
[49] Jiang Q, Zhao Y, Zhang X, Yang X, Chen Y, Chu Z, et al. Surface passivation of perovskite film for efficient solar cells. Nat Photonics 2019;13(7):460–6.
[50] Zhu C, Niu X, Fu Y, Li N, Hu C, Chen Y, et al. Strain engineering in perovskite solar cells and its impacts on carrier dynamics. Nat Commun 2019;10:815.