A robotic platform for the synthesis of colloidal nanocrystals

Haitao Zhao [1,7] ✉, Wei Chen [1,7], Hao Huang [1,7], Zhehao Sun[1,2,7], Zijian Chen [1], Lingjun Wu [1], Baicheng Zhang [3], Fuming Lai [1], Zhuo Wang[1,4], Mukhtar Lawan Adam [1], Cheng Heng Pang [4], Paul K. Chu[5], Yang Lu [6], Tao Wu[4], Jun Jiang [3] ✉, Zongyou Yin [2] ✉ & Xue-Feng Yu [1] ✉

Morphological control with broad tunability is a primary goal for the synthesis of colloidal nanocrystals with unique physicochemical properties. Here we develop a robotic platform as a substitute for trial-and-error synthesis and labour-intensive characterization to achieve this goal. Gold nanocrystals (with strong visible-light absorption) and double-perovskite nanocrystals (with photoluminescence) are selected as typical proof-of-concept nanocrystals for this platform. An initial choice of key synthesis parameters was acquired through data mining of the literature. Automated synthesis and in situ characterization with further ex situ validation was then carried out and controllable synthesis of nanocrystals with the desired morphology was accomplished. To achieve morphology-oriented inverse design, correlations between the morphologies and structure-directing agents are identified by machine-learning models trained on a continuously expanded experimental database. Thus, the developed robotic platform with a data mining–synthesis–inverse design framework is promising in data-driven robotic synthesis of nanocrystals and beyond.

Colloidal nanocrystals (NCs) have shown great potential in optical, photochemical, electrochemical, optoelectronic and biomedical applications[1–3]. One of the major goals in the synthesis of colloidal NCs is to achieve desired physicochemical properties through morphological control. However, traditional trial-and-error synthesis and labour-intensive characterization procedures restrict the development of morphology-tunable NCs. To reduce the time and effort required, robotic synthesis, in which these manual tasks are conducted by robotic chemists/scientists[4,5], chemical synthesis machines[6,7] or self-driving laboratories[8–12], is being rapidly developed to free up human scientists. Great progress has recently been made in this promising synthetic approach through the combination of robotic techniques[13,14] and artificial intelligence (AI) technologies[15,16].

Several synthetic platforms have been successfully developed for the synthesis of organic materials, for example, small organic molecules[7,17,18], organic cages and catenanes[14], peptides[6], pharmaceutical compounds[13], analgesic lidocaine and other compounds[19]. On these platforms, automated synthesis[4–8,13–17,19,20], robotic characterization[4,8,14,16,20], experimental database generation[4,5,8,15,16,19,21] and AI[4,5,8,15,16,18,19,21–23] have been gradually accomplished. Toward retrosynthesis of targeted materials[15,18,23], the combination of AI and an experimental database generated on a robotic platform is

[1]Materials Interfaces Center, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [2]Research School of Chemistry, Australian National University, Canberra, ACT, Australia. [3]School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, China. [4]Department of Chemical and Environmental Engineering, University of Nottingham Ningbo China, Ningbo, China. [5]Department of Physics, Department of Materials Science and Engineering, and Department of Biomedical Engineering, City University of Hong Kong, Kowloon, China. [6]Department of Mechanical Engineering, City University of Hong Kong, Kowloon, China. [7]These authors contributed equally: Haitao Zhao, Wei Chen, Hao Huang, Zhehao Sun. ✉e-mail: ht.zhao@siat.ac.cn; jiangj1@ustc.edu.cn; zongyou.yin@anu.edu.au; xf.yu@siat.ac.cn
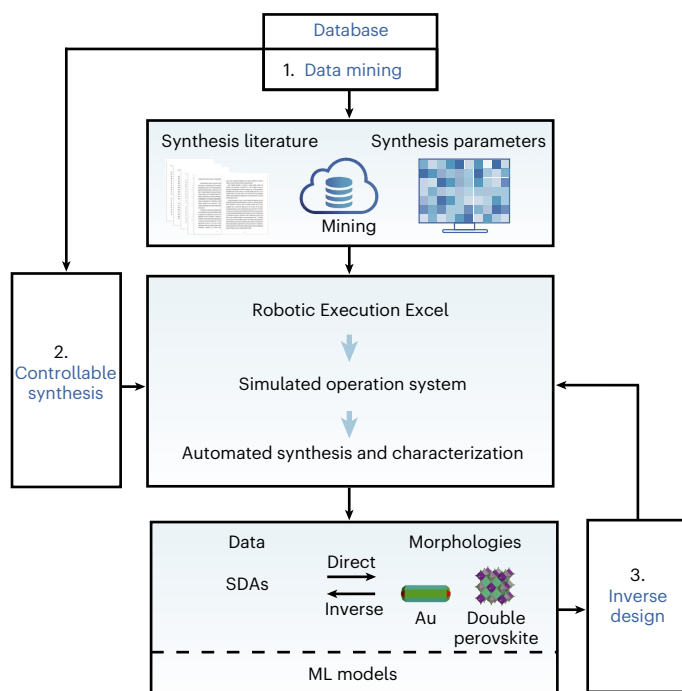
Fig. 1 | Framework of the robotic platform for the synthesis of colloidal NCs. (1) Data mining of the literature. (2) Controllable synthesis of colloidal NCs (consisting of Robotic Execution Excel, a simulated operation system, and automatic synthesis and characterization). (3) Inverse design of targeted NC morphologies (with correlations between SDAs and NC morphologies identified by ML models).

a breakthrough for accelerating the discovery of materials[24]. However, in practice, robot-assisted high-throughput characterization remains less explored, whereas synthesis planning has been integrated with retrosynthesis to streamline robotic synthesis[15]. A software platform that could directly translate the organic chemistry literature into editable code to drive automated synthesis has been reported[19]. Moreover, advanced data-driven models have been recently applied to extract the organic synthesis parameters from patents for synthesis planning[22,23].

The application of robotic synthesis, pioneered through the synthesis of organic materials, is expected to be expanded to the field of inorganic materials[24]. A state-of-the-art robotic chemist has been reported to significantly improve the performance of photocatalysts[4], indicating the great potential for robotic characterization of inorganic materials. The adoption of a robotic platform in the preparation and characterization of lead halide perovskites has been reported[25]. Other lead-containing perovskite (MAPbBr$_3$, FAPbBr$_3$, CsPbI$_3$, CsPbBr$_3$ (ref. [26]), MAPbI$_3$ (refs. [26,27]), FAPbI$_3$ (refs. [26,27]) and MA$_{0.1}$Cs$_{0.05}$FA$_{0.85}$Pb(I$_{0.95}$Br$_{0.05}$)$_3$ (ref. [28])) thin films were further studied by incorporating automated synthesis, robotic characterization and machine learning (ML) techniques. Moreover, the combination of automation and AI has been applied to the discovery of palladium thin films[8]. However, data mining of the literature to drive the robotic search for targeted inorganic materials, especially NCs, has rarely been explored.

In the field of inorganic materials, data-driven initial hypotheses[22,23], robot-assisted synthesis[4,6] and experimental databases[11,18,29] are promising for integration on robotic platforms to progressively acquire knowledge[30], iteratively develop ML models[18] and efficiently reveal data correlations[16]. Nevertheless, limitations still exist that hinder the robotic synthesis of inorganic NCs, for example, the lack of data-driven models for translating synthetic goals into robotic synthesis[4,22,23], experimentally characterized crystal morphologies for realizing controllable synthesis[31,32] and ML model-based correlation

identification for applying an inverse material design approach ('inverse design')[33,34].

Here we show how these limitations can be addressed by developing a robotic platform through a framework consisting of data-driven automated synthesis, robot-assisted controllable synthesis and morphology-oriented inverse design. We demonstrate this platform by synthesizing inorganic colloidal NCs, with an emphasis on their morphological control. Specifically, gold NCs (with strong visible-light absorption and no photoluminescence) and lead-free double-perovskite NCs (with photoluminescence and weak visible-light absorption) are selected as typical proof-of-concept NCs for research topics that are well known and emerging, respectively. Moreover, the correlations between tunable NC morphologies and key synthesis parameters are identified by ML models trained on a robot-assisted high-throughput experimental database.

Results

Framework of the robotic platform

To overcome the limitations of robotic synthesis and explore the complex tunable morphologies of colloidal NCs, a new robotic platform was specifically developed for high-throughput synthesis and characterization of NCs. To design such a platform, the key steps in a typical research project performed by human scientists, namely, searching the literature, conducting the synthesis and characterization of NCs, and iterating throughout trial-and-error experiments to obtain optimized synthesis parameters, were all considered in the conceptual framework of the robotic platform. This new synthetic framework, as illustrated in Fig. 1, integrates data mining of the initial key synthesis parameters from the literature (Fig. 1(1)), controllable synthesis of colloidal NCs (Fig. 1(2)) and inverse design of targeted NC morphologies (Fig. 1(3)).

In the framework, data mining of related literature is first conducted to provide initial choices of key synthesis parameters of NCs, such as the types or concentrations of surfactants. To illustrate the operation of the whole platform, two typical demonstrations, gold NCs and double-perovskite NCs, are selected for exploring research topics with abundant and relatively little published literature, respectively.

Based on the synthesis parameters obtained from data mining, we conducted high-throughput synthesis and characterization of NCs to investigate morphology-controlled synthesis (controllable synthesis), which has been a research hotspot in materials chemistry[31]. The key synthesis parameters that control the crystal morphology are identified as structure-directing agents (SDAs). The processes integrate Robotic Execution Excel files, a simulated operation system, and automated synthesis and characterization modules. Specifically, the Robotic Execution Excel files are designed for execution of the automated platform, which works as a user-friendly interactive interface between the platform and researchers, in which no programming skills are required. The experimental design is accomplished by writing the Excel file with information about the SDAs. The designed procedures written in the Excel file are pre-examined before the experiment, and monitored in real time during the robotic synthesis process by the simulated operation system (Supplementary Videos 1 and 2).

The platform was built with the desired synthesis and characterization functionalities, as shown in Fig. 2. Photographs and a schematic representation of the platform are shown in Fig. 2a,b, respectively. Compared to the currently available platforms[4–8,13–17,19,20], our platform, which is equipped with rapid optical characterization modules—such as a spectrometer (for ultraviolet–visible–near-infrared absorption measurement), a colour-ultrasensitive mobile camera (to obtain photographs and red–green–blue (RGB) values) and light sources (white light and ultraviolet light)—and integrated with automated synthesis modules and two collaborative robots, is specifically developed for automated synthesis and characterization of colloidal NCs. The typical properties of gold NCs (with strong visible-light absorption) and double-perovskite NCs (with photoluminescence) are robotically in
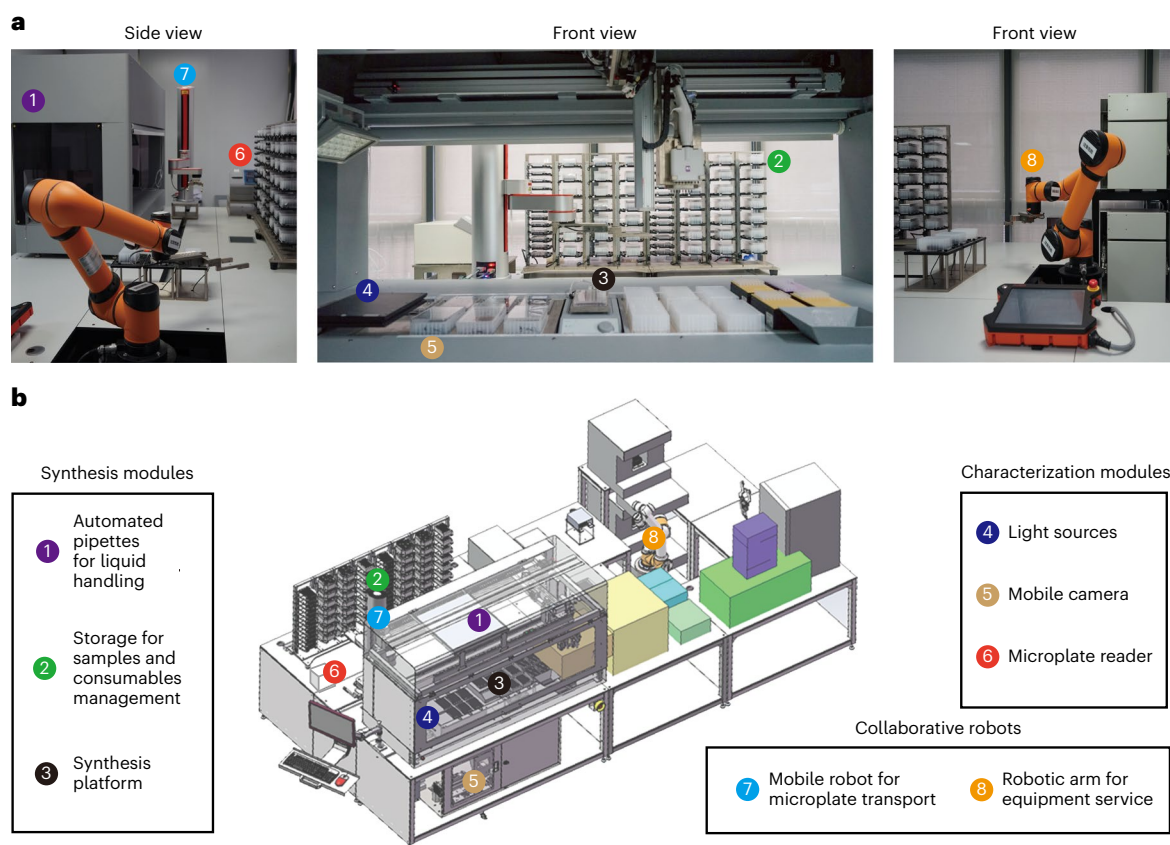
Fig. 2 | Robotic platform for NC synthesis and characterization. a, Side view and front view photographs. **b**, Schematic illustration.

situ characterized by the spectrometer and the mobile camera on the platform. Absorption spectra and photographs (taken under white light or ultraviolet irradiation and the corresponding digitalized RGB values) are automatically acquired for further analysis. Supplementary Videos 1 and 2 show the operation of the platform.

Through automated synthesis and characterization, SDAs and the corresponding characterization results are acquired. ML models are then trained to identify the correlations between SDAs and NC morphologies. Finally, inverse design, in which the desired morphologies are used to predict the SDAs used in robotic synthesis (morphologies → SDAs)[33], is implemented to accelerate the synthesis of targeted NCs. Moreover, the database is continuously expanded through data mining, automated characterization and ML predictions, which provides constructive guidance for achieving morphology-oriented inverse design.

This framework (Fig. 1) of data mining–synthesis–inverse design will be discussed in detail to demonstrate the capabilities of this robotic platform (literature search, NC synthesis and characterization, and correlation identification) in the following sections.

Data-driven automated synthesis

To plan and perform material synthesis, key synthesis parameters must be known, which are often determined on the basis of literature reports, trial-and-error experiments or the researcher's previous experience. Recently, synthesis parameters, such as solvents and ligands, were predicted by extracting experimental text from patents, using state-of-the-art data-driven models[22,23]. Although manual work cannot be easily fully replaced by data mining, the efficiency can be drastically boosted via automated literature searches compared with traditional trial-and-error experiments, and the dependency on the researcher's expertise can be greatly reduced. However, in practice, data mining of synthesis parameters for automated synthesis, especially for NCs, is an emerging research area.

In the current study, data mining of the literature was applied to drive the platform with the aid of an automated literature recommendation system[35]. Key parameters required for the synthesis of well-known gold NCs and less-known double-perovskite NCs were extracted from the literature in two typical ways, either directly from the specific literature of gold NCs or indirectly from the related literature of other perovskites (Supplementary Fig. 1).

To demonstrate that the platform can address the difficulty of studying morphology-controlling synthesis variables among a large number of parameters in abundant published literature, gold NCs were selected as a typical example (Supplementary Fig. 1). For the known gold NC synthesis protocol, the target of data mining is to determine the most frequently used concentrations of the surfactant (hexadecyltrimethylammonium bromide (CTAB)) and other agents (for example, $AgNO_3$, HCl, L-ascorbic acid (AA), $HAuCl_4$, $NaBH_4$ and gold seeds). Figure 3a and Supplementary Fig. 2 (in detail) show the frequency distribution of the concentrations of key synthesis parameters reported in 1,300 studies regarding this specific synthesis protocol. Among them, the papers with the most frequently used parameters are indicated with blue dots in Fig. 3a. Moreover, L2, corresponding to the highest-frequency parameter, is defined as the middle level, and a linear transformation (Supplementary Table 1) is applied to keep the L2 parameters at the same vertical position, as shown in Fig. 3b. By further fitting the curves obtained by Gaussian expansion, the boundaries of the shaded rectangle are determined as the low (L1) and high (L3) levels. Hence, L1, L2 and L3 are selected as key synthesis parameters for robotic execution of experiments.

Based on the data-mining results, the automated synthesis followed an orthogonal design, as shown in Supplementary Table 2. Ultraviolet-visible-near-infrared absorption spectroscopy (Supplementary Fig. 3) and multivariate analysis of variance (Supplementary Table 3) were performed. Twenty-four initial levels of experimental conditions
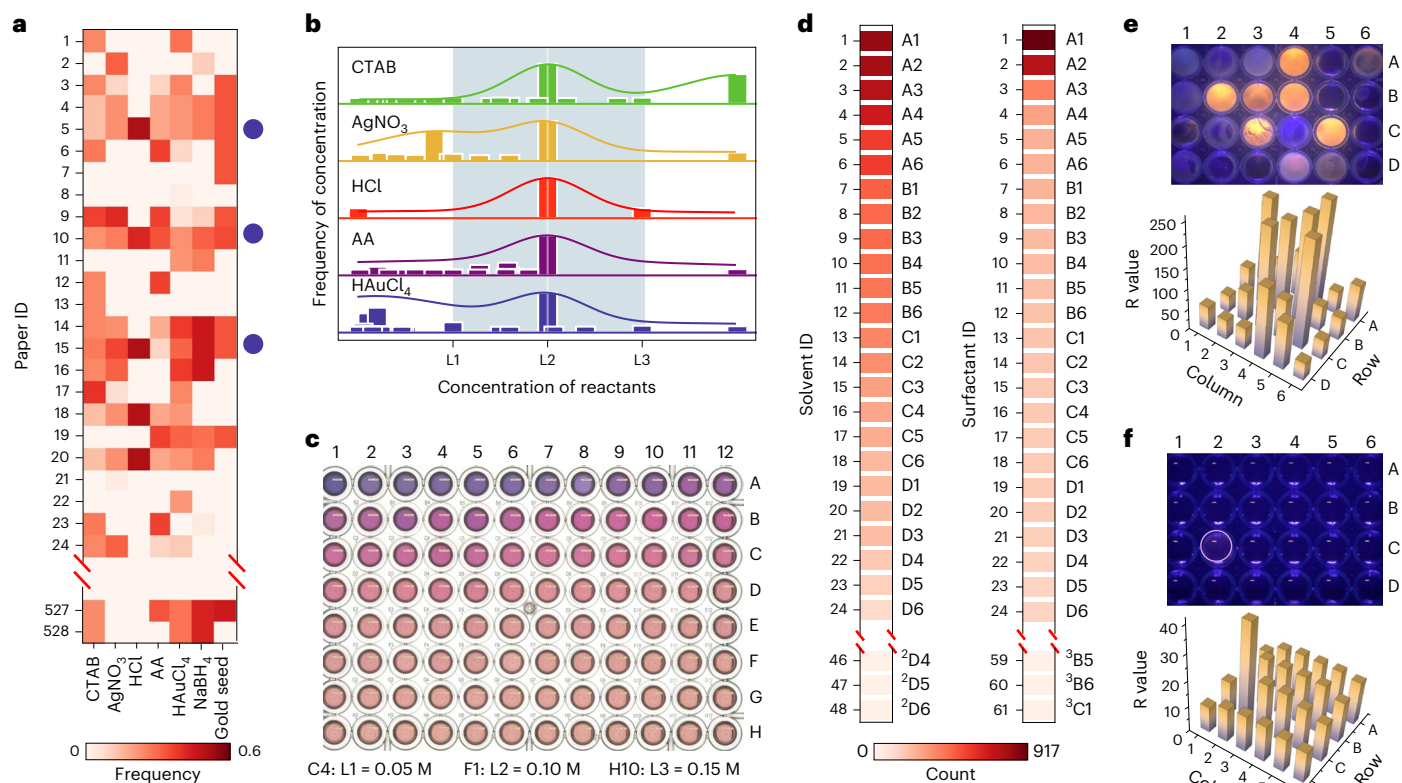
Fig. 3 | Data-driven automated synthesis. a, Data mining of key synthesis parameters for gold NCs (the key synthesis parameters extracted are the concentrations of CTAB, AgNO$_3$, HCl, AA, HAuCl$_4$, NaBH$_4$ and gold seed). Papers with IDs 1–24 and 527–528 are identified and the others are displayed in Supplementary Fig. 2. Blue dots indicate papers with the most frequently used parameters. **b**, Frequency distribution of the concentration with identified low (L1), middle (L2) and high (L3) levels. The most frequently used concentrations are identified as L2, and linear transformation is applied to make all L2 values fall at the same vertical position. L1 and L3 refer to the boundary values of the shaded rectangle covering the specific area of the fitting curves obtained by Gaussian expansion. **c**, Photographs of samples prepared by automated synthesis (96 samples with different concentrations of CTAB; 1–12 and A–H are the microplate labels). **d**, Data mining of key synthesis parameters for double-perovskite NCs (initial candidates of 48 solvents and 61 surfactants, the others are listed in Supplementary Tables 5 and 6; ^{2}D4 is the well in column 4 and row D of the second plate shown in the tables). **e**, Photograph showing 24 samples with different solvents prepared by automated synthesis under ultraviolet irradiation (top) and corresponding R (red) values of the photoluminescense (bottom). **f**, Samples of supernatants extracted with 24 surfactants (top) and corresponding R values of the photoluminescence (bottom).

(Supplementary Table 4) were chosen for a further single-factor (adjusting one factor) study based on the results of the orthogonal experiments. To explore the potential impact of all single factors (CTAB, AgNO$_3$, HCl, gold seeds, AA and HAuCl$_4$), 24 levels and 96 extended levels of high-throughput experiments were designed, as shown in Supplementary Table 4, and were carried out to construct a database. A photograph of 96 samples obtained from automated synthesis with the 96 extended levels of CTAB is displayed in Fig. 3c as an example, showing the gradually changing colour of the synthesized gold NCs. The corresponding optical absorption spectra are provided in Supplementary Fig. 4, showing the strong absorption of gold NCs in the visible region. These results suggest that adjusting the concentration of CTAB as a key synthesis parameter can lead to colour change and a peak shift of the longitudinal surface plasmon resonance (LSPR), indicating that these parameters can have a potential effect on morphology manipulation. Hence, a further study of morphology-controlled synthesis will be of great significance.

For the research topic of lead-free double-perovskite NCs, about which there is less published literature (although lead-containing perovskite materials have been widely studied), the target of data mining is to identify potential surfactants and solvents in the related literature for the synthesis of Cs$_2$AgIn$_{1-x}$Bi$_x$Cl$_6$ NCs as another typical demonstration (Supplementary Fig. 1). The bismuth ions are doped into the cubic unit cell of the Cs$_2$AgInCl$_6$ crystal to enhance the crystal quality and photoluminescence efficiency. This lead-free material was selected for its low toxicity and high photoluminescence efficiency. A probe of the potential influence of all solvents (Supplementary Table 5) and surfactants (Supplementary Table 6) with the initial choices obtained from data mining (Fig. 3d) was conducted on the platform.

First, in situ photoluminescence characterization (under irradiation by ultraviolet light with emission at 365 nm) of 48 solvents was conducted in an attempt to screen the double-perovskite samples with the highest photoluminescence efficiency. The images were obtained using a colour-ultrasensitive mobile camera, and the colours were analysed both qualitatively by visual inspection and quantitatively by digital analysis of the RGB values, as shown in Fig. 3e and Supplementary Fig. 5. The R values of the samples with ethanol (A4), ethyl acetate (B2), isopropanol (IPA, B4), diethyl ether (C3), acetic acid (C5) and 1,4-dioxane (^{2}B2) as the solvent exposed to ultraviolet irradiation were higher than 240, and their emissions were much brighter, as shown in Fig. 3e, indicating higher photoluminescence efficiency.

Second, with these six solvents selected, the potential role of 61 surfactants (obtained from data mining, as additives in 366 experiments) in the morphology tuning of NCs was investigated. The supernatant was extracted for photoluminescence analysis under ultraviolet irradiation since crystals of small sizes can be better dispersed in solution (which results in a colour difference in the solution), while larger crystals tend to quickly precipitate. For most samples with different
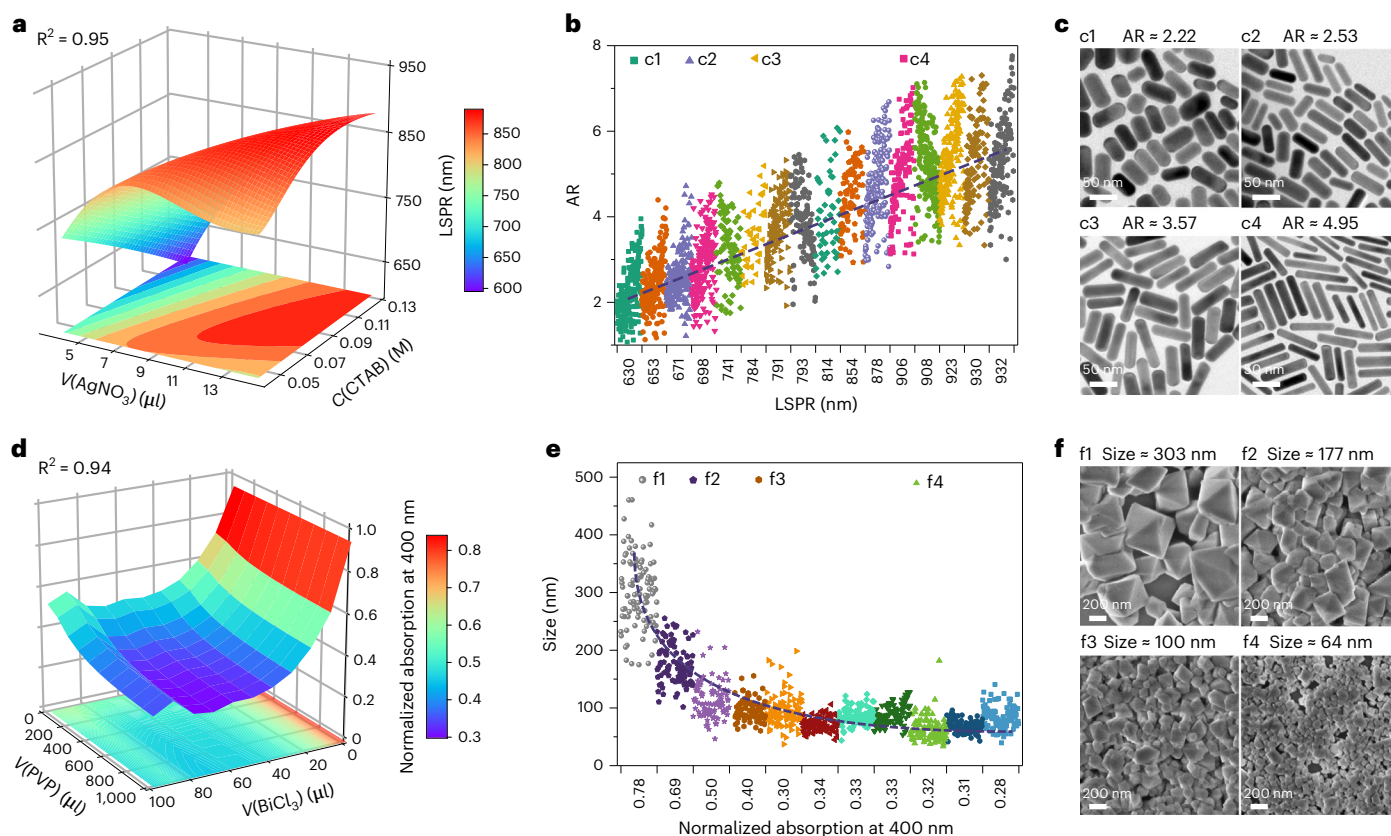
Fig. 4 | Controllable synthesis involving robotic in situ characterization, ML models and ex situ morphology validation. a, Double-factor ML model of robotically in situ characterized LSPR versus SDA content for gold NCs (volume (*V*) and concentration (*C*)). **b**, Identified linear relationship between the LSPR and (AR (the results for c1–c4 are measured from **c**). **c**, TEM images for AR validation (more TEM images are presented in Supplementary Fig. 8). **d**, Double-factor ML model of robotically in situ characterized normalized absorption at 400 nm versus SDA content for double-perovskite NCs. **e**, Identified relationship between the normalized absorption at 400 nm and morphological size (the results for f1–f4 are measured from **f**). **f**, SEM images for size validation (more SEM images are presented in Supplementary Fig. 16).

additives, the photoluminescence colours show little difference due to the precipitation of larger crystals at the bottom. The dispersibility of the extracted supernatant of all the samples was characterized, as shown in Fig. 3f (the rest of the data are shown in Supplementary Fig. 6). The results illustrate that the sample with polyvinyl pyrrolidone (PVP, circled well in Fig. 3f) as the surfactant exhibits the best dispersibility and strongest photoluminescence (R = 39), indicating the possible formation of smaller crystals, which is worth further investigating for controllable synthesis of crystals with a wide range of sizes.

Robot-assisted controllable synthesis
When using this robotic platform, the key to controllable synthesis is to establish correlations between SDAs for automated synthesis and the corresponding crystal morphologies on the nanoscale. To achieve this goal, in situ robotic synthesis and characterization, ML prediction and ex situ transmission electron microscopy (TEM) or scanning electron microscopy (SEM) characterization (for morphology validation) were combined (Fig. 4). The correlations between the SDAs and the morphologies were determined by constructing ML models based on a dataset generated from rapid in situ characterization and validated by a small dataset generated from ex situ characterization. Based on the initial results from data-driven robotic synthesis, controllable synthesis of the two typical examples (gold NCs and double-perovskite NCs) was further explored.

On the basis of single-factor experiments, the ML models identified correlations between the robotically in situ characterized LSPR and SDA content for gold NCs, which are presented in Supplementary

Fig. 7 (the corresponding coefficients are listed in Supplementary Tables 7 and 8). CTAB, $AgNO_3$ and HCl showed greater effects on the LSPR than the gold seeds, AA and $HAuCl_4$. Therefore, CTAB, $AgNO_3$ and HCl were identified as SDAs, which could be used as key synthesis parameters to control the morphology during robotic synthesis. For example, the LSPR peak shift in Supplementary Fig. 4 could be adjusted by the factor CTAB using this platform. As a result, the relationships between the SDAs and the characterized LSPR were verified to be the key to achieve controllable synthesis.

To further explore the manipulation mechanism, double-factor (that is, investigation of two identified SDAs) experiments were carried out (Supplementary Table 9). The LSPR was further extracted and then used to train the double-factor ML models. CTAB and $AgNO_3$ exhibited synergistic effects on the tunability of the LSPR (Fig. 4a), which is consistent with the observation that CTAB and Ag^+ form a face-specific capping agent for tuning the morphology[36,37].

Furthermore, the morphological aspect ratio (AR) (Fig. 4b) was measured and calculated based on TEM results (Fig. 4c and Supplementary Fig. 8). A linear relationship between the LSPR and morphological AR was observed, as shown in Fig. 4b. These results further confirm that the SDA-based parameter can be used as an input, while the AR can be used as the output, providing data useful for controllable synthesis of NCs. The double-factor ML models of AR versus SDA contents are presented in Supplementary Fig. 9 (the corresponding coefficients are listed in Supplementary Tables 10 and 11). Interestingly, CTAB and HCl in the double-factor experiment (Supplementary Fig. 9f) exhibited similar behaviour in morphology manipulation (in a
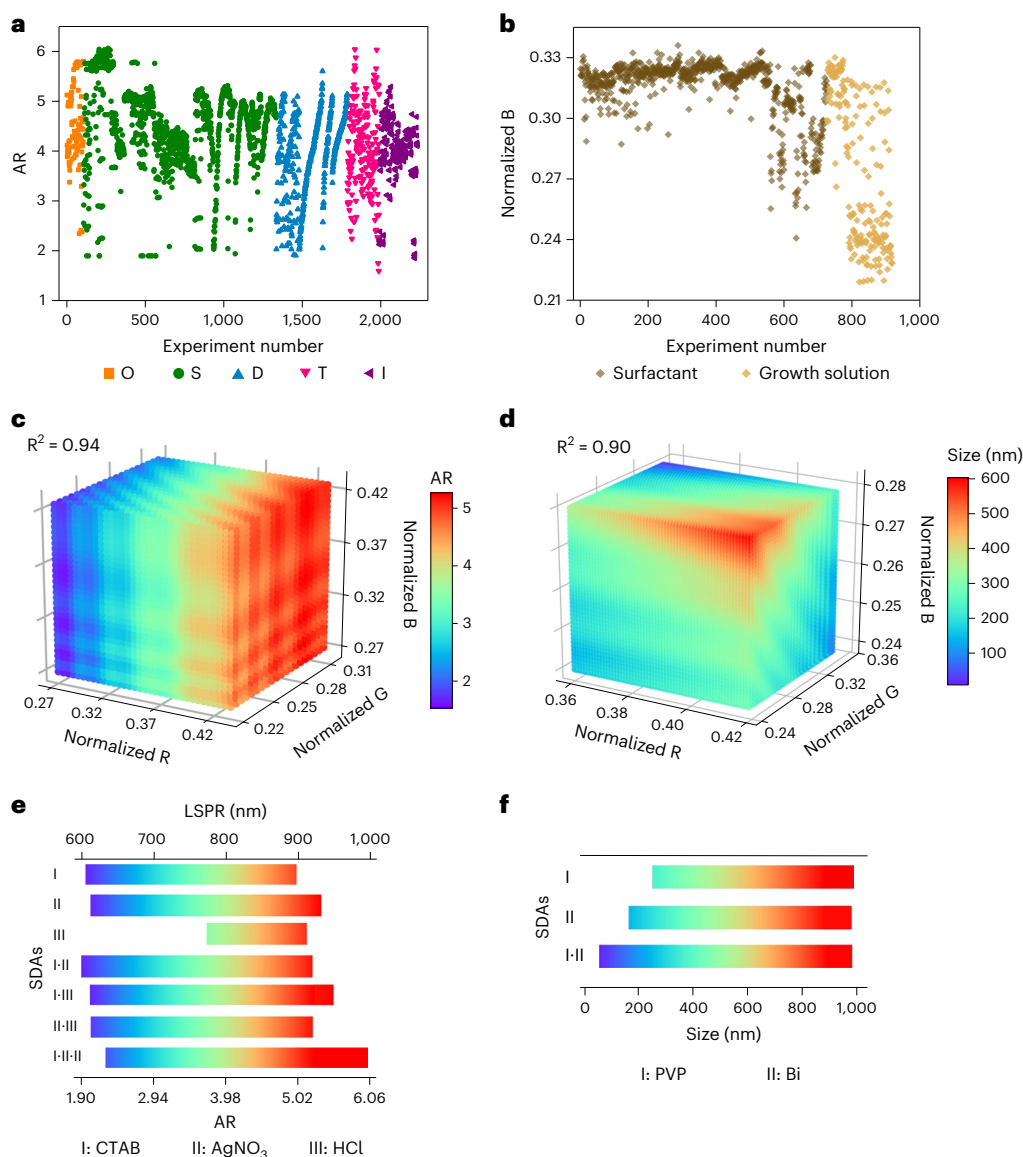
Fig. 5 | Experimental database and ML models facilitate inverse design. **a**, Graphical illustration of the database for synthesizing gold NCs: O, S, D, T and I represent the orthogonal, single-factor, double-factor, triple-factor and inverse design experiments, respectively. **b**, Graphical illustration of the database for synthesizing double-perovskite NCs. **c**, ML normalized RGB–AR model for gold NCs. R, G, B represent red, green and blue, respectively. **d**, ML normalized RGB–size model for double-perovskite NCs. R, G, B represent red, green and blue, respectively. **e**, The ML-predicted correlations between SDAs (as inputs) and the AR or LSPR (as output) were identified for inverse design of targeted gold NCs. **f**, The ML-predicted correlations between SDAs (as inputs) and size (as output) were identified for inverse design of the double-perovskite NCs from microcrystals.

collaborative manner) to that of CTAB and AgNO$_3$ in their double-factor experiment (Supplementary Fig. 9c). In contrast, for the double-factor experiment investigating AgNO$_3$ and HCl (parameters in Supplementary Table 11), AgNO$_3$ dominated the morphology manipulation (Supplementary Fig. 9i).

The design of triple-factor experiments (parameters in Supplementary Table 12) and the developed ML models are shown in Supplementary Tables 13 and 14. By varying three SDAs in the triple-factor experiments, a more complex AR response profile was observed, as shown in Supplementary Fig. 10. Therefore, the correlation between the SDAs and morphological AR in the controllable synthesis of gold NCs was confirmed.

For controllable synthesis of double-perovskite crystals, single-factor and double-factor experiments were performed to tune the crystal size from microcrystals to NCs. PVP as a surfactant shows great promise in reducing the size of the double-perovskite crystals,

as suggested by the results of the surfactant screening experiments (Fig. 3f). In addition, considering that the main ions might also contribute to the formation of crystals, single-factor experiments (parameters in Supplementary Table 15) were conducted to verify the effects of the CsCl, InCl$_3$ and BiCl$_3$ additives together with PVP as the surfactant based on the crystal structure (Supplementary Fig. 11) and photoluminescence (Supplementary Fig. 12) results. In the presence of PVP (0–1,000 μl) and BiCl$_3$ (0–100 μl), the pure phase of Cs$_2$AgIn$_{1-x}$Bi$_x$Cl$_6$ (0 ≤ x ≤ 1) (Supplementary Fig. 11g) was obtained, as confirmed by X-ray diffraction patterns (Supplementary Fig. 11d–f). Moreover, the supernatants of the samples with varied PVP and BiCl$_3$ contents exhibited photoluminescence changes (Supplementary Fig. 12). These experiments reveal the vital roles of PVP and BiCl$_3$ in the growth of Cs$_2$AgIn$_{1-x}$Bi$_x$Cl$_6$ double perovskites as SDAs.

Therefore, double-factor experiments (PVP and BiCl$_3$ contents) were designed to further investigate the factor effects in tuning the

crystal size, as shown in Supplementary Table 16. Ultraviolet–visible–near-infrared absorption spectra and colour analysis of the samples show similar trends, which can be elucidated by Mie scattering theory[38]. In situ ultraviolet–visible–near-infrared absorption spectroscopy characterization was conducted (Supplementary Fig. 13). The spectra were normalized according to the absorption peak, and the absorption intensity at 400 nm, as shown in Supplementary Fig. 14, was extracted as an indicator of the crystal size[39]. The double-factor ML model of the characterized absorption versus SDA content is presented in Fig. 4d, showing that the intensity at 400 nm decreased with increasing PVP and $BiCl_3$ contents (validated in Supplementary Fig. 15). The correlation between the normalized absorption at 400 nm and the crystal size was further evaluated, as shown in Fig. 4e. The size was validated by SEM characterization (down to 64 nm) (Fig. 4f and Supplementary Fig. 16). These results indicate that the size of the $Cs_2AgIn_{1-x}Bi_xCl_6$ double-perovskite crystal can be manipulated by changing both the PVP and $BiCl_3$ contents (Supplementary Fig. 17), which is consistent with the SEM validation and theoretical calculations (Supplementary Fig. 18). Hence, the correlations between the SDAs (PVP and $BiCl_3$ contents) as inputs and crystal sizes as the output for controllable synthesis of double-perovskite NCs were identified.

Morphology-oriented inverse design

Utilizing the data acquired from controllable synthesis, the robotic platform was further developed with the aim of inversely designing targeted NC morphologies based on the correlations between SDAs and the morphologies identified by ML models. This platform continues to be improved by receiving more robotically characterized data to realize the ultimate goal of morphology-oriented inverse design of NCs. With the aid of the platform, over 2,300 gold NC samples and 1,000 double-perovskite NC samples were synthesized and in situ characterized; graphical diagrams of the databases are shown in Fig. 5a,b, respectively. The experimental database is considered to be crucial in supporting the inverse design process.

At the same time, the results obtained from in situ colour characterization using the colour-ultrasensitive mobile camera contributed to the generation of another potential dataset for rapid inverse design. Colour information (RGB values) was extracted from photographs of gold NCs (Supplementary Fig. 19) and double-perovskite NCs (Supplementary Fig. 20). The corresponding RGB values are presented in Supplementary Tables 17 and 18, and the ML models of gold NCs and double-perovskite NCs were trained based on the normalized RGB values, as presented in Fig. 5c,d, respectively. ML models with R^2 values of 0.94 and 0.90 were obtained for gold and double-perovskite NCs (the formulas and coefficients of the models are provided in Supplementary Tables 19 and 20), respectively. The results indicate that colour analysis, which is typically achieved by visual inspection by a professional scientist, can serve as another indirect indicator for rapid morphological identification. In this way, the platform can be used for digitalization of colour in a similar way as a scientist but without bias, thus contributing to inverse design of NCs with colour features.

Based on the experimental database and ML models, the morphologies (AR or size) as 'input' and identified SDAs as 'output' are presented for gold NCs and double-perovskite NCs in Fig. 5e,f, respectively. In Fig. 5e,f, the effective ranges of the typical morphological AR (for gold NCs, from 1.90 to 6.00) and size (for double-perovskite NCs, down to 64 nm) are graphically displayed. Morphological control with broad tunability within these ranges is revealed for both gold NCs and double-perovskite NCs. Finally, the achieved inverse design of a targeted NC morphology with the corresponding SDAs is also illustrated in Fig. 5e,f.

The effective inverse design of gold NCs (Supplementary Fig. 21) and double-perovskite NCs (Supplementary Fig. 22) shows the promise of robotic synthesis of targeted NCs. In particular, we demonstrate the successful synthesis of targeted gold NCs (AR = 4.06 ± 0.41) and of nanosized (78 nm) and microsized (749 nm) double perovskites

using the inverse design strategy, as given in Supplementary Tables 21 and 22. Therefore, this study reveals that inverse design can be achieved by making use of databases (SDA-based parameters from robotic synthesis, robotic in situ characterization results and ex situ validation results) and the corresponding ML models.

Discussion

To address the limitations of automated synthesis and characterization of colloidal NCs with morphological control on the nanoscale, a robotic platform framework is established by integrating data mining, controllable synthesis and inverse design of targeted NC morphologies.

For automated synthesis, data mining of the literature was conducted to provide initial choices of key synthesis parameters, for example, the concentrations of the known surfactants for gold NCs and the types of unknown surfactants for lead-free double-perovskite NCs. High-throughput automated synthesis, such as single-factor and double-factor experiments, was then systematically conducted. In these processes, accessible large (in situ characterized ultraviolet–visible–near-infrared absorption spectra and RGB results) and small (ex situ TEM and SEM validations) datasets were generated to continuously expand the experimental databases. Through a sequence of iterations, the corresponding experimental database was constructed for training the ML models, which enabled controllable synthesis of morphology-tunable NCs. These developed ML models could be used to identify the complex correlations between the SDAs and crystal morphologies in the controllable synthesis of gold NCs and double-perovskite NCs. The experimental databases and ML models are critical for supporting the inverse design process. Furthermore, inverse design of targeted morphologies with the ML-predicted SDA-based synthesis parameters (morphologies → SDAs) was successfully demonstrated for both gold NCs (with strong visible-light absorption) and double-perovskite NCs (with intensive photoluminescence).

Comprising data-driven robotic synthesis, robot-assisted controllable synthesis and morphology-oriented inverse design, this new synthetic framework was developed for synthesis of inorganic NCs. Training individuals to be highly qualified scientists with the expertise for conducting crystal synthesis and characterization on the nanoscale entails considerable cost. The outcomes of NC morphology manipulation can be diverse, as they depend heavily on the scientists' experiences. Moreover, most material synthesis and discovery involve trial-and-error synthesis and labour-intensive characterization[19,40]. The prototype of this robotic platform specifically demonstrated for synthesis and characterization of NCs is a start toward reducing the manual tasks. In the current work, initial selection of key synthesis parameters from a literature search, high-throughput synthesis and in situ characterization, synthesis parameter–crystal morphology correlation identification, and inverse design of morphology-tunable NCs were achieved, which are comparable to those performed by an experienced scientist in these fields. This synthetic approach is believed to be promising for digital synthesis of NCs from data to a crystal with the desired morphology using the robotic platform.

Methods
Data mining

The initial choices of key synthesis parameters were obtained by data mining the existing literature through our automated literature recommendation system to plan the automated synthesis. The whole process consists of several steps, as shown in Supplementary Fig. 1: (1) downloading literature from publishers by keywords; (2) using rules to locate target paragraphs; (3) splitting words and sentences in paragraphs by ChemicalTagger[41]; (4) applying a four-step cascading tagger (chemical named entity recognition by OSCAR4[35], additional recognition of chemical entities based on a domain dictionary, identification of the units of compound properties based on regular expressions and tagging of English parts of speech (POS) by OpenNLP

(https://opennlp.apache.org); (5) applying phrase parsing;[41] and (6) statistically analysing the distribution of the extracted key synthesis parameters. In this work, the workflow of the automated literature recommendation system was demonstrated with two typical topics: gold NCs and double-perovskite NCs (details are provided in the Supplementary Methods).

Robotic platform

The robotic platform was designed and assembled in our laboratory with a series of modules capable of performing robot-assisted high-throughput synthesis and in situ characterizations. Automated pipettes, sample and consumables storage, a synthesis platform, light sources, colour-ultrasensitive mobile cameras, and a microplate reader are integrated with a mobile robot and a robotic arm, as shown in Fig. 2. Based on the synthesis parameters obtained from data mining, the execution of automated experiments is programmed in Robotic Execution Excel files and then checked and monitored by a simulated operation system to conduct synthesis and characterization of NCs. Two typical operation videos for gold NCs (under white light for colour capture) and double-perovskite NCs (under ultraviolet light for fluorescence capture) are provided in Supplementary Videos 1 and 2.

Preparation of chemical solutions

Standard solutions, consisting of CTAB (0.2 M), $HAuCl_4$ (0.02 M), $AgNO_3$ (0.01 M), HCl (3 M) and AA (0.2 M) for the synthesis of gold NCs, and CsCl (0.2 M in 36–38% HCl), AgCl (5 mM in 36–38% HCl), $InCl_3$ (0.1 M in 36–38% HCl) and $BiCl_3$ (0.1 M in 36–38% HCl) for the synthesis of double-perovskite NCs, were first manually prepared in volumetric flasks (100 ml). Then, the standard solutions were transferred to the plate on the platform and automatically diluted in proportion by execution of the designed Robotic Execution Excel files. As a result, the desired concentrations of precursors could be obtained for the synthesis of gold NCs and double-perovskite NCs. Details are provided in the Supplementary Methods.

Automated synthesis of gold NCs

For a typically automated synthesis of gold NCs, certain amounts of CTAB (0.1 M, 1 ml), $HAuCl_4$ (0.01 M, 50 μl), $AgNO_3$ (0.01 M, 10 μl), HCl (1 M, 20 μl) and AA (0.1 M, 8 μl) were pipetted into a well of a 96-well microplate by automatic pipettes. A 4 min mixing process was set for each addition of chemicals (except for the gold seeds). After that, 2.4 μl of preprepared seed solution was injected into the growth solution and then gently stirred for 10 s. The microplate was transferred to a furnace by the robotic arm and kept undisturbed at 28 °C for 12 h. After the growth of gold NCs for 12 h, the microplate was taken out of the furnace by the robotic arm. Next, 200 μl of the mixed solution was aspirated into a well of a 96-well transparent microplate for ultraviolet–visible–near-infrared absorption measurement and colour characterization. Details of the synthesis parameters and experimental design for the gold NCs are given in the Supplementary Methods. A video showing the automated synthesis of the gold NCs is also provided in the Supplementary Video 1.

Automated synthesis of double-perovskite NCs

For a typical automated synthesis of $Cs_2AgIn_{1-x}Bi_xCl_6$ double perovskite, 1 ml of PVP solution (20 mg ml^{-1} in 36–38% HCl), 500 μl of AgCl (5 mM in 36–38% HCl), 500 μl of CsCl (0.2 M in 36–38% HCl), 90 μl of $InCl_3$ (0.1 M in 36–38% HCl) and 10 μl of $BiCl_3$ (0.1 M in 36–38% HCl) were added into one well of a 24-deep-well plate. For experiments of varied V(PVP), a complementary HCl (36–38%) solution was added to ensure that the total volume of the PVP solution and HCl was 1,000 μl. The solution was then mixed for 30 s. Next, 250 μl of the solution in each well was transferred into the corresponding well in another 24-well transparent plate, followed by the addition of 1,250 μl of isopropanol as a solvent, leading to the immediate formation of NCs/microcrystals. The mixed solution was then taken for further characterization. Details of the synthesis parameters and experimental design, and a video showing the automated synthesis of the double perovskite, are provided in Supplementary Methods and Supplementary Video 2.

ML models

ML models were employed for identification of the correlation between the synthesis parameters and the corresponding morphologies from an experimental database, which were characterized by an ultraviolet-visible-near-infrared absorption spectrometer and a colour-ultrasensitive mobile camera installed on the robotic platform. The sure independence screening and sparsifying operator (SISSO) approach[42,43], a supervised ML algorithm that is a compressed sensing-based approach, was used to determine the critical correlation. For the construction of the experimental feature spaces during ML training, the set of implemented operators was:

$$\text{opset} \equiv \{(+)(-)(*)(/)(\^{-1})(\^2)(\^3)(\exp)(\log)(\sqrt{})(\text{cbrt})(\sin)(\cos)\}$$

The detailed ML models with the corresponding coefficients are given in the Supplementary Information (Supplementary Tables 7, 8, 10, 11, 13, 14, 19 and 20) for gold NCs and double-perovskite NCs.

Data availability

The data that support the findings of this study are available in the Supplementary Information (Supplementary Methods, Supplementary Figs. 1–22 and Supplementary Tables 1–22), Supplementary Videos 1 and 2 (synthesis of colloidal gold nanocrystals and double-perovskite nanocrystals on the robotic platform). Source data are provided with this paper.

Code availability

The computer code, algorithm and related data to generate the results that are reported in this paper and are central to its main claims are available in the Zenodo repository with the digital object identifier: https://doi.org/10.5281/zenodo.7353405. The algorithms for chemical name entity recognition[35], expressions and grammatical structures[41], and SISSO[42,43] are adaptable as described in detail in Refs. [35,41–43].

References

1. Li, P. et al. Evaporative self-assembly of gold nanorods into macroscopic 3D plasmonic superlattice arrays. *Adv. Mater.* **28**, 2511–2517 (2016).

2. Talapin, D. V., Lee, J.-S., Kovalenko, M. V. & Shevchenko, E. V. Prospects of colloidal nanocrystals for electronic and optoelectronic applications. *Chem. Rev.* **110**, 389–458 (2010).

3. Parak, W. J. et al. Biological applications of colloidal nanocrystals. *Nanotechnology* **14**, R15–R27 (2003).

4. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).

5. King, R. D. et al. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252 (2004).

6. Angelone, D. et al. Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. *Nat. Chem.* **13**, 63–69 (2021).

7. Service, R. F. The synthesis machine. *Science* **347**, 1190–1193 (2015).

8. MacLeod, B. P. et al. A self-driving laboratory advances the Pareto front for material properties. *Nat. Commun.* **13**, 995 (2022).

9. MacLeod, B. P. et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **6**, eaaz8867 (2020).

10. Häse, F., Roch, L. M. & Aspuru-Guzik, A. Next-generation experimentation with self-driving laboratories. *Trends Chem.* **1**, 282–291 (2019).

11. Häse, F., Roch, L. M. & Aspuru-Guzik, A. Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chem. Sci.* **9**, 7642–7655 (2018).

12. Epps, R. W., Volk, A. A., Ibrahim, M. Y. S. & Abolhasani, M. Universal self-driving laboratory for accelerated discovery of materials and molecules. *Chem* **7**, 2541–2545 (2021).

13. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).

14. Greenaway, R. L. et al. High-throughput discovery of organic cages and catenanes using computational screening fused with robotic synthesis. *Nat. Commun.* **9**, 2849 (2018).

15. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).

16. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).

17. Li, J. et al. Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).

18. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).

19. Mehr, S. H. M., Craven, M., Leonov, A. I., Keenan, G. & Cronin, L. A universal system for digitization and automatic execution of the chemical synthesis literature. *Science* **370**, 101–108 (2020).

20. Bédard, A.-C. et al. Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).

21. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241 (2016).

22. Vaucher, A. C. et al. Automated extraction of chemical synthesis actions from experimental procedures. *Nat. Commun.* **11**, 3601 (2020).

23. Vaucher, A. C. et al. Inferring experimental procedures from text-based representations of chemical reactions. *Nat. Commun.* **12**, 2573 (2021).

24. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).

25. Gu, E. et al. Robot-based high-throughput screening of antisolvents for lead halide perovskites. *Joule* **4**, 1806–1822 (2020).

26. Higgins, K., Ziatdinov, M., Kalinin, S. V. & 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.* **143**, 19945–19955 (2021).

27. Zhao, Y. et al. Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning. *Nat. Commun.* **12**, 2191 (2021).

28. Zhao, Y. et al. A bilayer conducting polymer structure for planar perovskite solar cells with over 1,400 hours operational stability at elevated temperatures. *Nat. Energy* **7**, 144–152 (2022).

29. Skilton, R. A. et al. Remote-controlled experiments with cloud chemistry. *Nat. Chem.* **7**, 1–5 (2015).

30. Hammer, A. J. S., Leonov, A. I., Bell, N. L. & Cronin, L. Chemputation and the standardization of chemical informatics. *JACS Au* **1**, 1572–1587 (2021).

31. Li, S.-H., Qi, M.-Y., Tang, Z.-R. & Xu, Y.-J. Nanostructured metal phosphides: from controllable synthesis to sustainable catalysis. *Chem. Soc. Rev.* **50**, 7539–7586 (2021).

32. Lai, F. et al. Machine learning-aided crystal facet rational design with ionic liquid controllable synthesis. *Small* **17**, 2100024 (2021).

33. Zunger, A. Inverse design in search of materials with target functionalities. *Nat. Rev. Chem.* **2**, 0121 (2018).

34. Wang, Z. et al. Data-driven materials innovation and applications. *Adv. Mater.* **34**, 2104113 (2022).

35. Jessop, D. M., Adams, S. E., Willighagen, E. L., Hawizy, L. & Murray-Rust, P. OSCAR4: a flexible architecture for chemical text-mining. *J. Cheminform.* **3**, 41 (2011).

36. Jackson, S. R., McBride, J. R., Rosenthal, S. J. & Wright, D. W. Where's the silver? Imaging trace silver coverage on the surface of gold nanorods. *J. Am. Chem. Soc.* **136**, 5261–5263 (2014).

37. Kim, F., Song, J. H. & Yang, P. Photochemical synthesis of gold nanorods. *J. Am. Chem. Soc.* **124**, 14316–14317 (2002).

38. Bohren C F, Huffman D. R. *Absorption and Scattering of Light by Small Particles* (Wiley, 2008).

39. Saha, S. K. et al. Scalable submicrometer additive manufacturing. *Science* **366**, 105–109 (2019).

40. Cheetham, A. K., Seshadri, R. & Wudl, F. Chemical synthesis and materials discovery. *Nat. Synth.* **1**, 514–520 (2022).

41. Hawizy, L., Jessop, D. M., Adams, N. & Murray-Rust, P. ChemicalTagger: a tool for semantic text-mining in chemistry. *J. Cheminform.* **3**, 17 (2011).

42. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2**, 083802 (2018).

43. Ouyang, R., Ahmetcik, E., Carbogno, C., Scheffler, M. & Ghiringhelli, L. M. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *J. Phys. Mater.* **2**, 024002 (2019).

Acknowledgements

Author contributions

H.Z. conceived the idea and designed the robotic platform and experiments. H.Z., J.J., Z.Y. and X.-F.Y. supervised the project. W.C. and H.H. planned and performed the experiments, and collected and analysed the data. Z.S. and F.L. contributed to the structural analysis and theoretical calculations. Z.C. and L.W. helped with the synthesis and characterization of the perovskite nanocrystals. B.Z. contributed to the data mining. Z.W. and M.L.A. performed ML training. C.H.P., P.K.C., Y.L., T.W., J.J., Z.Y. and X.-F.Y. edited the paper, and all authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44160-023-00250-5.

Correspondence and requests for materials should be addressed to Haitao Zhao, Jun Jiang, Zongyou Yin or Xue-Feng Yu.

Peer review information *Nature Synthesis* thanks Teodoro Laino, Yuebing Zheng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary handling editor is Peter Seavill, in collaboration with the *Nature Synthesis* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.