

Large-Language-Model-Based AI Agent for Organic Semiconductor Device Research

Qian Zhang, Yongxu Hu, Jiaxin Yan, Hengyue Zhang, Xinyi Xie, Jie Zhu, Huchao Li, Xinxin Niu, Liqiang Li, Yajing Sun,* and Wenping Hu*

Large language models (LLMs) have attracted widespread attention recently, however, their application in specialized scientific fields still requires deep adaptation. Here, an artificial intelligence (AI) agent for organic field-effect transistors (OFETs) is designed by integrating the generative pre-trained transformer 4 (GPT-4) model with well-trained machine learning (ML) algorithms. It can efficiently extract the experimental parameters of OFETs from scientific literature and reshape them into a structured database, achieving precision and recall rates both exceeding 92%. Combined with well-trained ML models, this AI agent can further provide targeted guidance and suggestions for device design. With prompt engineering and human-in-loop strategies, the agent extracts sufficient information of 709 OFETs from 277 research articles across different publishers and gathers them into a standardized database containing more than 10 000 device parameters. Using this database, a ML model based on Extreme Gradient Boosting is trained for device performance judgment. Combined with the interpretation of the high-precision model, the agent has provided a feasible optimization scheme that has tripled the charge transport properties of 2,6-diphenylidithieno[3,2-*b*:2',3'-*d*]thiophene OFETs. This work is an effective practice of LLMs in the field of organic optoelectronic devices and expands the research paradigm of organic optoelectronic materials and devices.

1. Introduction

In recent years, artificial intelligence (AI) technology, large language models (LLMs) such as OpenAI's generative pre-trained transformer 4 (GPT-4) and Google's Gemini have emerged. These advanced models have been revolutionizing information generation and processing across various fields.^[1,2] They have shown exceptional performance in diverse tasks, including text comprehension, visual recognition, mathematical problem-solving, and coding. Such achievements significantly advance the development of artificial general intelligence. Particularly in specialized fields such as chemistry,^[3–6] material science,^[7–12] biology,^[13–17] and medicine,^[18–22] LLMs have shown immense potential. The close collaboration between AI technologies and scientists now allows for the efficient completion of tasks that were previously labor-intensive and time-consuming, such as literature review, compound screening, and data analysis. This new paradigm of scientific research not only opens new avenues for scientific innovation but also significantly accelerates the pace of scientific discovery, heralding a new era in scientific exploration and technological breakthroughs.^[23,24]

Interest in the field of organic semiconductor devices is surging due to their flexibility and cost-effectiveness, particularly in applications involving flexible electronics.^[25–29] However, optimizing device performance is a challenging task due to the diverse selection of materials and design parameters. Conventional methods often rely on trial and error (Figure S1, Supporting Information), which can be inefficient, costly, and sometimes even dangerous.^[30] Furthermore, these methods fail to provide a comprehensive understanding of complex relationships between materials and device parameters. Actually, these relationships may have already existed in the vast body of published academic literature. The challenge lies in how to thoroughly mine these publications and effectively extract information from them. Specialized natural language processing models, like ChemicalTagger,^[31] Open-Source Chemistry Analysis Routines 4 (OSCAR4),^[32] and ChemDataExtractor,^[33] are designed to automatically extract key information from scientific texts. However, these models

Q. Zhang, Y. Hu, J. Yan, H. Zhang, X. Xie, J. Zhu, H. Li, X. Niu, L. Li, Y. Sun, W. Hu

Key Laboratory of Organic Integrated Circuits
Ministry of Education and Tianjin Key Laboratory of Molecular Optoelectronic Sciences

Department of Chemistry

School of Science

Tianjin University

Tianjin 300072, China

E-mail: syj19@tju.edu.cn; huwp@tju.edu.cn

Q. Zhang, J. Yan, H. Zhang, J. Zhu, X. Niu, Y. Sun

Haihe Lab of ITAI

Tianjin 300051, China

W. Hu

Joint School of National University of Singapore and Tianjin University
Fuzhou, Fujian 350207, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.202405163>

DOI: [10.1002/adma.202405163](https://doi.org/10.1002/adma.202405163)

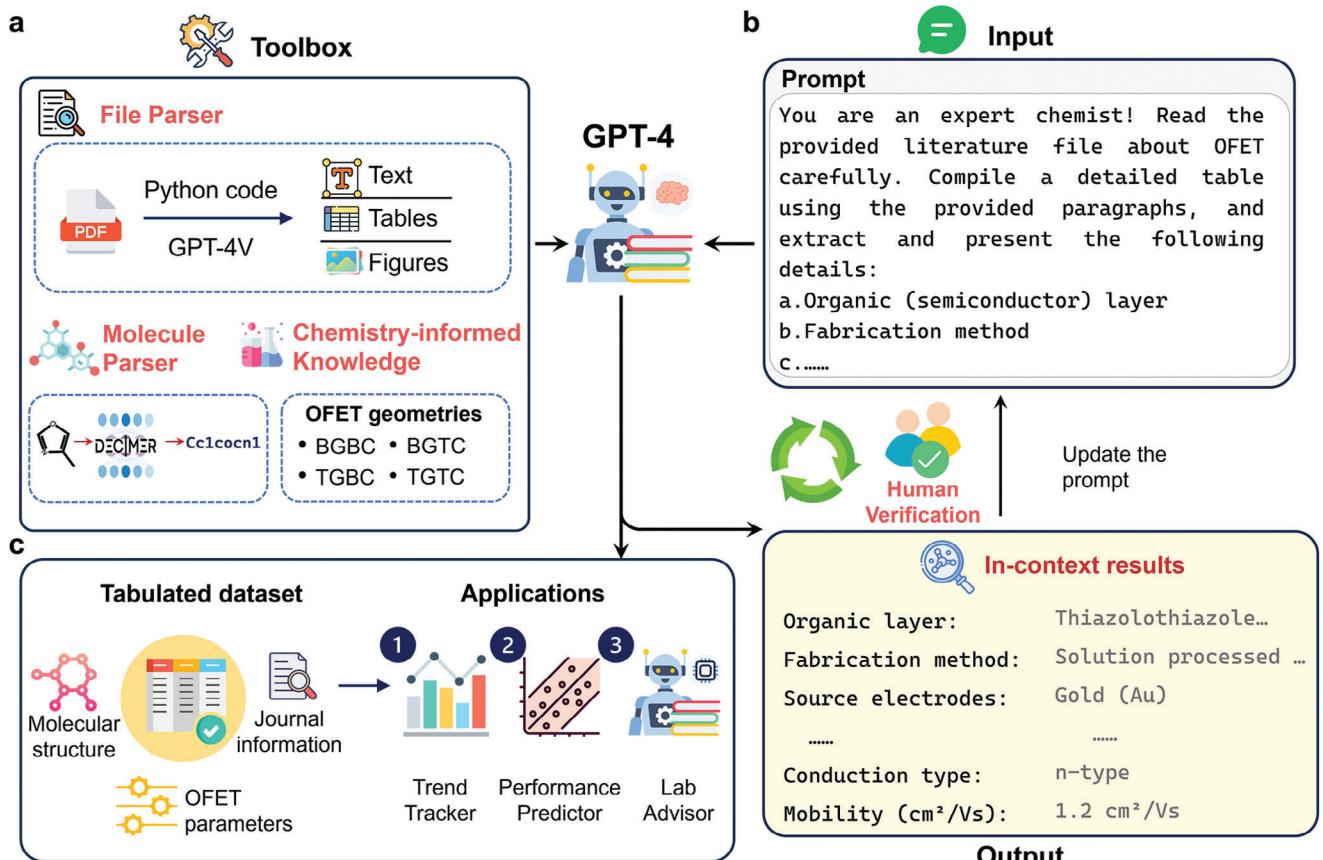


Figure 1. Schematic representation of the proposed LLM-based AI agent. **a)** The toolbox for data preprocessing. **b)** Human-in-the-loop prompt engineering strategy. **c)** Establishment of the tabulated dataset and further applications.

encounter several challenges, including the need for manual input, the integration of domain knowledge with expertise in computer and data science, and the extraction of information from images. By contrast, multimodal LLMs are convenient to use and can process diverse nontextual data, including images, audios, and videos.^[34] By harnessing LLMs for in-depth analysis of scientific literature and employing advanced machine learning (ML) strategies, our aim is to distill decades of research into actionable insights that can directly inform future studies and innovation.

In this study, we develop an AI agent to assist organic field-effect transistor (OFET) research by integrating GPT-4 model with machine learning algorithms. This integration has led to a novel workflow that guides the OFET design. An AI agent, in this context, refers to a system capable of performing different tasks or making decisions based on its programming and the data it processes. Our study focuses on three primary areas: First, we introduce a GPT-4-based text mining methodology adept at extracting OFET design parameters from scientific texts, tables, and images. Second, leveraging the extracted dataset of OFET design parameters, we develop three applications: a Research Trend Tracker, a Performance Predictor, and a Lab Advisor. Notably, the Performance Predictor can not only predict OFET performance but also decipher the critical factors influencing their charge carrier mobility using SHapley Additive exPlanations (SHAP) method.^[35] Finally, experimental val-

idation has confirmed a significant enhancement in the performance of 2,6-diphenyldithieno[3,2-*b*:3',2'-*d*]thiophene (DP-DTT) OFETs, attributed to the insights from interpretability analysis and guidance provided by the Lab Advisor. This study highlights the transformative potential of LLMs in the development of organic semiconductor devices and paves the way for accelerating scientific discovery and technological innovation using AI.

2. Results and Discussion

2.1. Overview of LLM-Based Agent for OFETs

Nowadays, a few studies have used text mining techniques to extract material basic properties or chemical synthesis methods from scientific literatures to build the specific datasets. However, different from these researches often with less target parameters and readable expressions, the fabrication of OFET involves more complicated design parameters. It necessitates gathering information across textual, tabular, and graphical formats (Figure S2, Supporting Information), which significantly increases the challenge of building the parameters database. To address this issue, we leverage the image processing capabilities of GPT-4 V, a state-of-the-art multimodal LLM. As demonstrated in Figure 1a, our file parser dissects academic literatures in portable document format (PDF) into textual content, tabular data, and image data

using prewritten Python code and the GPT-4 V model (Figures S3–S5, Supporting Information). It is important to note that numerous OFET setup diagrams include crucial fabrication parameters. These parameters encompass the geometric configuration of the OFET, the materials of source and drain electrodes, among others. Consequently, this study primarily emphasizes the image data content of depicting OFET structures. Furthermore, to facilitate the subsequent analysis of the molecular characteristics of semiconductor materials, we employed the deep learning for chemical image recognition (DECIMER)^[36] tool to convert molecular images from literature into simplified molecular-input line-entry system (SMILES)^[37] format text. Finally, to mitigate potential hallucination^[38] from the GPT-4 and enhance the accuracy of text mining, we incorporated common chemical knowledge into the prompts, such as the four most common configurations: bottom-gate bottom-contact (BGBC), bottom-gate top-contact (BGTC), top-gate bottom-contact (TGBC), and top-gate top-contact (TGTC)^[39] (see more details in Note S1 in the Supporting Information).

When preparing research articles for analysis, it is essential to consider the variety in OFET design parameters. This includes differences in fabrication methods, material choices, and testing conditions mentioned within the same paper, as well as the distinct writing styles utilized by various researchers. Notably, due to the absence of a standardized format for reporting OFET fabrication, various research groups and journals have adopted a range of reporting templates. Against this backdrop, we identified 14 key OFET design parameters. These encompass semiconductor material parameters (such as names of the semiconductor materials), device fabrication parameters (including fabrication method, thickness of the semiconductor layer, materials of the source and drain electrodes, and materials and thickness of the dielectric layer and gate), and device performance parameters (such as testing environment, on/off current ratio, threshold voltage, and the maximum mobility). These parameters are detailed in Table S1 in the Supporting Information. Previous ChatGPT-based text mining studies often had to segment the text into specified numbers of tokens due to input token limitations,^[6,7,13] then feed them to ChatGPT in sequences. This strategy potentially disrupted the continuity of context processing by ChatGPT and affected the final accuracy. Fortunately, with the release of the latest GPT-4 model by OpenAI, it is now possible to process inputs up to 128 000 tokens. This capacity enables the inclusion of the full text of an article and its Supporting Information, accommodating texts of various lengths and complexities.

Prompt engineering (PE) is a precise strategy aimed at designing prompts that guide LLM to accurately generate relevant information. Within the specific domain, implementing PE significantly enhances performance of LLMs.^[40] In this work, we adopted a human-in-the-loop approach to update prompts (Figure 1b). This strategy involves direct user involvement in the optimization process of prompts, allowing for their refinement and adjustment through continuous experimentation and feedback (Figures S6–S9, Supporting Information). Initially, we created a series of starting prompts based on literature reviews and preliminary experiments. Subsequently, by evaluating the responses generated by ChatGPT to these prompts, we identified those that led to the highest quality responses. Through multiple iterations, we continuously collected the latest results from

ChatGPT and manually refined the prompts to adapt to the complexities of extracting OFET parameters. This process not only deepens our understanding of how ChatGPT processes chemical literature information but also significantly increases the accuracy of parameter extraction.

By integrating toolboxes with prompts, 277 academic articles and the corresponding Supporting Information were analyzed using the gpt-4-1106-vision-preview application programming interface (API), encompassing publications from Springer, WILEY, and other publishing groups (Figure 2a). This process enabled us to successfully extract 709 unique OFETs, encompassing a wide range of critical parameters such as organic layer materials, fabrication methods, and device configurations. Additionally, we integrated information such as the digital object identifier (DOI) and publication year of each document into the extracted data. Ultimately, this allowed us to construct a comprehensive table database containing information on 709 OFETs, laying the groundwork for subsequent analysis (Figure 1c). To extract useful insights for OFETs from the database, we have further developed three applications. Using the Trend Tracker, we explore the evolution of OFET technology over the past two decades. We established a Performance Predictor, based on a well-trained ML model, to assess OFET performance. Moreover, the Lab Advisor based on GPT-4 model provides targeted experimental suggestions aimed at reducing the trial-and-error costs in experimental processes.

2.2. Performance Evaluation of Text Mining

To assess the accuracy of our method in text mining OFET literature, we conducted a comprehensive evaluation of the entire database. Specifically, we manually verified the authenticity of 14 key parameters across 709 OFET configurations identified in the collected literature, serving as the basis for evaluating the performance of our text mining approach. This involved detailed scoring of nearly 10 000 parameters. Each parameter was assigned one of three labels: true positive (TP, correctly identified parameters), false positive (FP, incorrectly extracted OFET parameters or irrelevant information), or false negative (FN, missed OFET parameters). It is important to note that not all papers reported every one of these 14 parameters. For instance, some papers did not specify the testing environment, in which case we defaulted to room temperature for GPT. Likewise, if a paper did not specify the thickness of the organic or dielectric layers, we instructed GPT to return N/A, preventing the illusions. Subsequently, we calculated the precision, recall, and F1 score for each parameter, demonstrating that identification across all parameters performed exceptionally well, with all scores exceeding 0.92 (Figure 2b). Particularly, scores were relatively lower in identifying threshold voltage and on/off current ratio, two OFET performance parameters. This is because, unlike mobility, which is generally described in the text, these data often reside in tables within the literature. We use gpt-4-1106-vision-preview API to extract table information from the literature (Figure S5, Supporting Information). Unfortunately, the accuracy in extracting precise text from images, relying on the optical character recognition functionality of multimodal GPT-4 V, is lower compared to direct text extraction (Figure 2c).

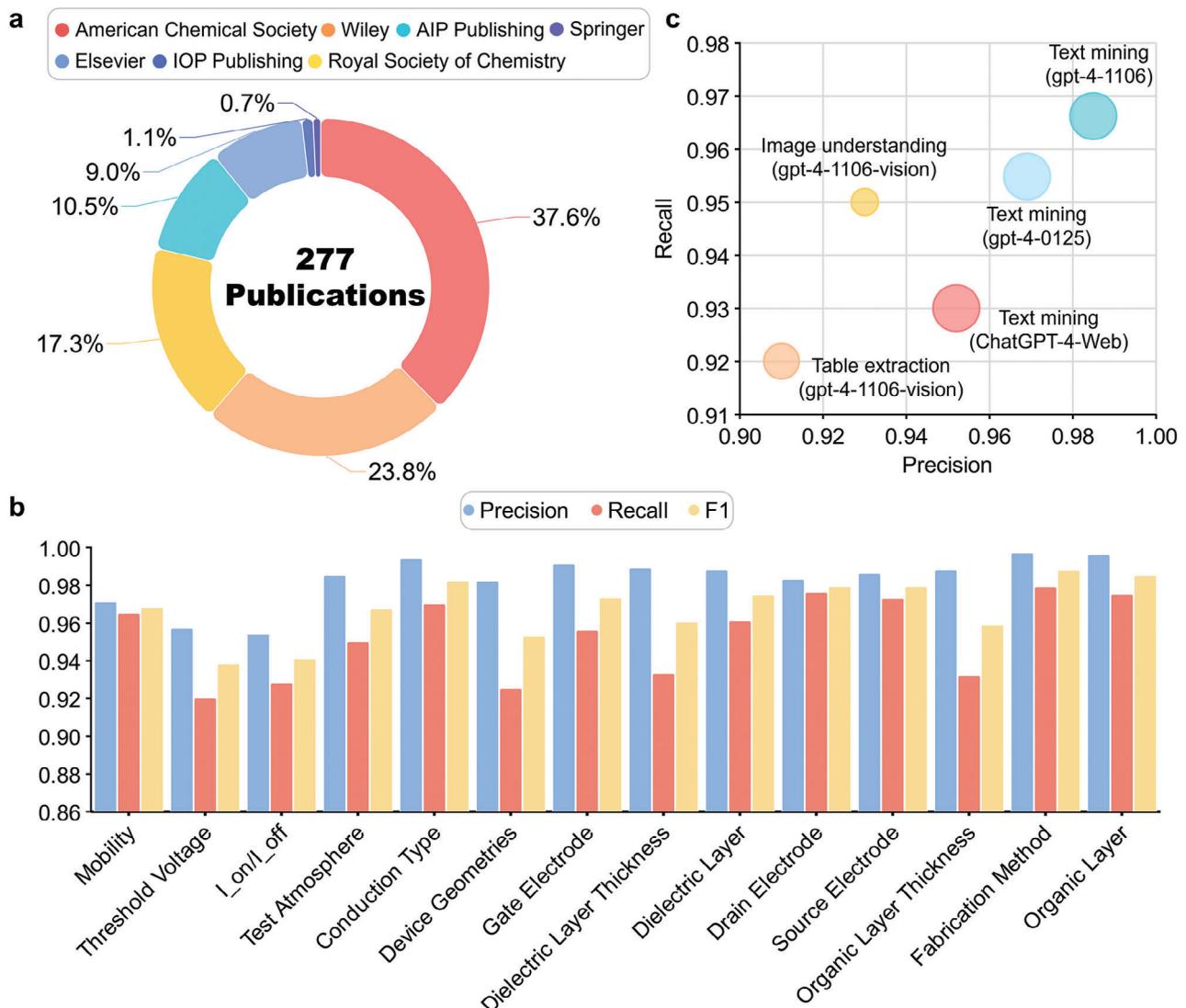


Figure 2. Data description and performance analysis of our agent. a) Distribution of the publishers in the academic literature used in our research. b) Average precision, recall, and F1 scores for each parameter of our agent. c) Comparison of the performance of different tasks and models using 30 papers randomly selected from 277 publications. The names of different models are shown in parentheses. The horizontal and vertical axes represent the precision and recall of each model, respectively. The node size is based on the dataset size rank.

Additionally, we compared the efficiency of our method with the time typically required by a postdoctoral researcher specialized in the OFET field to manually extract parameters from the literature. The results indicate that our method can decrease the time required for this process by a factor of six (Figure S10 Supporting Information). This increase in efficiency can free researchers from the tedium of labor-intensive tasks and allow them to focus on more specialized and innovative work.

2.3. Developing Applications Using Established Dataset

2.3.1. Research Trend Tracking in the OFET Field

Upon establishing the final tabular dataset, we first conducted a data visualization analysis of the publication years, organic

semiconductor materials, and their corresponding device mobilities. This analysis enables us to track the development trends in the field more clearly and comprehensively compared to traditional literature reviews. As illustrated in Figure 3a, the past two decades have witnessed significant changes in the volume of literature related to organic semiconductor materials. Notably, research related to OFET semiconductor materials peaks around 2007. Before 2010, studies predominantly focus on acenes, chalcogen-containing compounds, and nitrogen-containing heterocycles. In recent years, however, the focus has largely shifted toward chalcogen-containing heterocyclic semiconductors, as indicated by the line plot. Given that our literature selection was primarily targeted at the OFET semiconductor material, the focus of most publications post-2010 has shifted toward device fabrication methods. For instance, new fabrication processes such as photolithography and vacuum deposition methods have

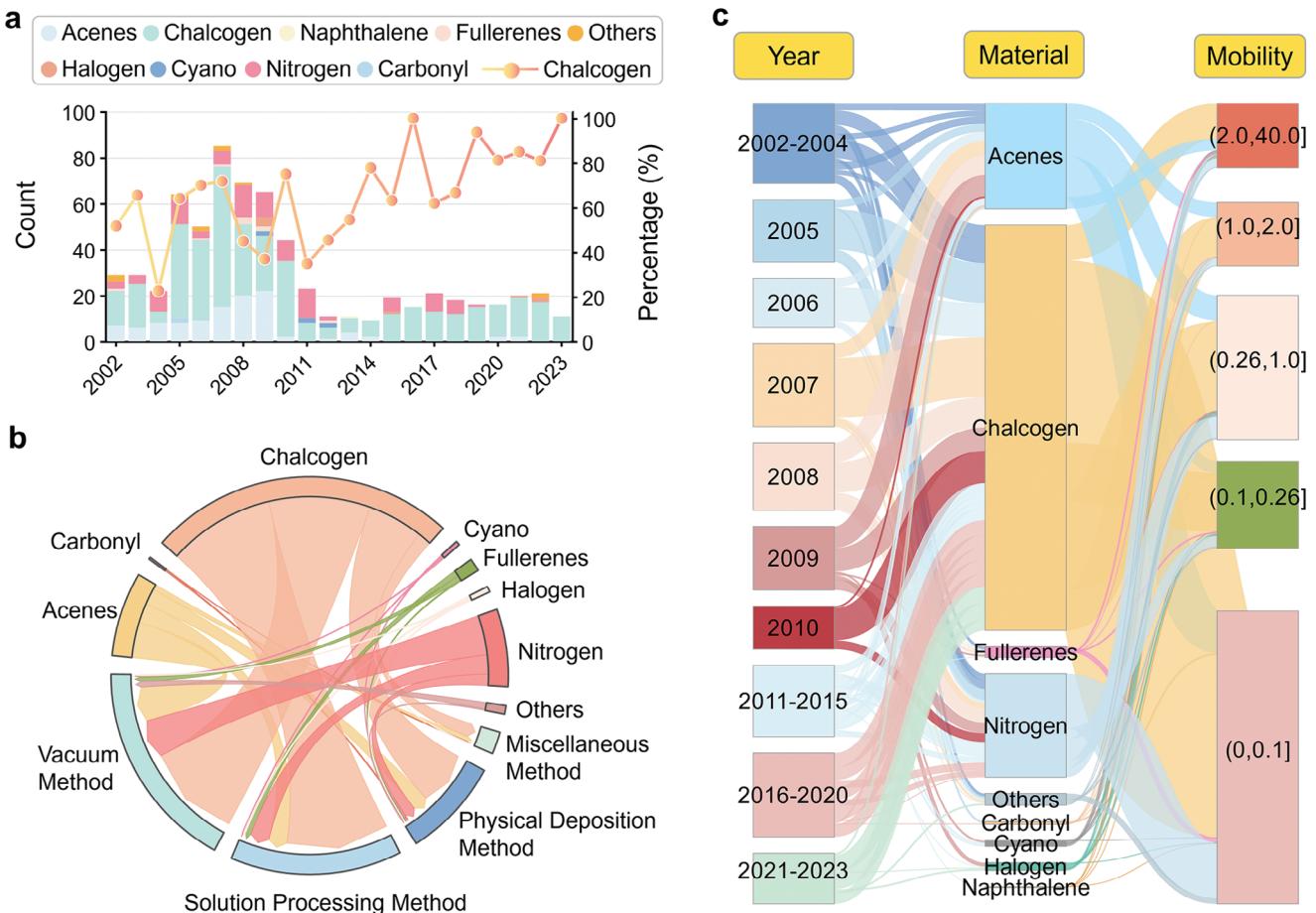


Figure 3. Trends and explorations in organic semiconductor materials. a) Stacked histogram showing the distribution of various organic semiconductor materials published in 277 articles over the past two decades. The categories of different semiconductor materials are taken from reference.^[39] The line graph represents the percentage of chalcogen semiconductor materials for different years. b) Chord plot associating organic semiconductor materials with the fabrication methods in the scientific literature. c) Sankey diagram illustrating the relationship between organic semiconductor materials and carrier mobilities of the corresponding OFETs. The carrier mobilities are divided into five intervals.

gradually emerged, which help reduce defects in the semiconductor layer and enhance the crystalline quality of materials.

Each category of semiconductor materials features a variety of fabrication methods, and potential correlations between these are illustrated in Figure 3b. To reveal the development trends between the types of organic semiconductor materials and their corresponding OFET mobilities, we constructed a Sankey diagram (Figure 3c). An analysis of the mobility ranges associated with different organic semiconductor materials demonstrated complex and diverse impacts on mobility within the OFET domain. Nearly all types of organic semiconductors spanned across all five mobility intervals, with the highest proportion of materials falling within the (0,1) mobility range. In the higher mobility intervals, chalcogen-containing heterocyclic semiconductors, including high-performance materials such as dinaphtho[2,3-*b*:2',3'-*f*]thiophene (DNTT)^[41] and 2,6-diphenyl anthracene (2,6-DPA)^[42] predominate. These materials typically exhibit higher levels of conjugation and intermolecular interactions, which lead to reduced reorganization energies and increased transfer integrals. Additionally, we explore the relationships between different years, fabrication methods, and mobil-

ities (Figure S11, Supporting Information). We find that these factors also span across all five mobility intervals. Although the choice of organic semiconductor material is often considered to have a decisive impact on device performance, the occurrence of the same material type exhibiting different mobilities highlights the importance of considering other device factors. These include fabrication methods, electrode material choices, and testing environments, which are crucial and should not be overlooked.

2.3.2. Performance Prediction and Feature Importance Analysis

To facilitate the rapid prediction of device performance, we integrated the molecular characteristics of semiconductor materials with device fabrication parameters to train a ML model. Because organic semiconductors with a maximum mobility over $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ are potentially beneficial for several applications,^[43,44] we divided OFETs performance into two groups: high mobility (reported maximum mobility over $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, comprising 149 samples) and low mobility (reported maximum mobility below $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, comprising 560 samples). Given the significant

Table 1. F1 score of five common machine learning models using various combinations of input parameters.

Input	XGBoost	RF	MLP	SVM	LR
Fabrication parameters	0.71	0.68	0.73	0.69	0.65
MACCS + Fabrication	0.75	0.79	0.8	0.78	0.75
Morgan + Fabrication	0.86	0.78	0.71	0.65	0.73
MACCS + Morgan + Fabrication	0.88	0.76	0.65	0.67	0.75

class imbalance in the dataset, we employed an undersampling strategy to randomly select 149 samples from the low mobility group, thus creating a balanced dataset (149 samples per category). Based on this, we used 80% of the data for training and 20% for testing, developing a binary classification model to assess OFET performance.

The input data for the ML model consisted of two parts: the molecular fingerprints of semiconductor materials, and the OFET device fabrication parameters processed through one-hot encoding (see more details in Note S2 and Table S2 in the Supporting Information). We selected two commonly used types of molecular fingerprints: molecular access system (MACCS)^[45] and Morgan.^[46] The MACCS fingerprint relies on a predefined set of 166 structural features (keys). Each feature represents a specific structural pattern or atomic combination potentially present in the molecule, such as certain functional groups or ring structures. By contrast, the Morgan fingerprint is based on the connectivity of molecules, generated by considering the information of atoms and their neighborhoods to capture finer structural details within molecules. Table 1 displays the input features formed by combining different fingerprints with the fabrication parameters of OFET devices. It also compares device mobility performance predictions using five common machine learning algorithms: Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Logistic Regression (LR). The results indicate that the XGBoost model, trained with both MACCS and Morgan fingerprints along with OFET device parameters as input features, achieved the best F1 score of 0.88. Moreover, the XGBoost model outperformed in all metrics, achieving an area under the receiver operating characteristic (AUROC) of 0.92, a precision of 0.93, and a recall of 0.83 (Figure S12, Supporting Information).

Interpreting the explicit relationship between input features and carrier mobility in black-box ML model presents significant challenges. To address this, we employed the SHAP^[35] method for the interpretability analysis of XGBoost model. This approach, grounded in game theory, quantifies the impact of each feature on the model's output through a generalized measure. SHAP analysis not only ranks features based on their importance but also delineates their positive and negative impacts on the model's output, offering insights into both global (feature importance) and local (individual contributions) levels. The summarization of global SHAP analysis results among the whole dataset highlighted the Morgan fingerprint characteristics (specifically at indices 1357 and 33) as the most influential features, followed by the device fabrication parameters (the dielectric layer and gate

electrode materials), as evidenced in Figure S13 (Supporting Information).

To further elucidate how selected features influence the performance of OFETs, we conducted separate importance analyses for molecular characteristics and OFET fabrication parameters. Here, we analyzed two representative OFET configurations: a high-mobility setup^[47] using dibenzo[*d,d'*]thieno[3,2-*b*;4,5-*b'*]dithiophene (DBTDT) as the organic semiconductor (Figure 4a) and a low-mobility setup^[48] using DP-DTT as the semiconductor material (Figure 4b). These configurations utilize semiconductor molecules with the same backbone structure. We first conducted a global feature importance analysis for the two OFET configurations (see Figures S14 and S15 in the Supporting Information). Since molecular fingerprints are represented only by site numbers, which do not intuitively reflect the molecular structure, we addressed this limitation by analyzing the SHAP values for each site in the Morgan fingerprints. We then mapped these values back to the average SHAP values for each atom within the molecules, thereby assessing the importance of each atom in the molecular structure. Our analysis reveals that in the DBTDT molecule, the linkage points between the five-membered rings and phenyl rings can enhance carrier transport (Figure 4a). This enhancement is likely due to strong π - π interactions and effective orbital overlap at these points, which are crucial for the delocalization of π electrons. Such delocalization directly facilitates the charge transport necessary for high mobility in organic semiconductors. Conversely, in the DP-DTT molecule, the sulfur atoms within the three five-membered rings have a more significant negative impact on carrier transport (Figure 4b). This could be due to the increased molecular size leading to weakened conjugation and enhanced steric hindrance. Additionally, the increased distance between the two phenyl rings, compared to the backbone structure, results in the carbon atoms on the phenyl rings having a more pronounced negative impact on mobility.

In addition to the molecular features, we also analyzed the device fabrication characteristics of the two OFETs. As shown in Figure 4c,d, a feature is assigned a value of 1 if a specific technique or material is utilized, and 0 if not. This analysis reveals key factors affecting the carrier mobility of OFETs, particularly highlighting the influence of fabrication methods and dielectric layer material choices. Specifically, the selection of physical deposition techniques is identified as critical for achieving high mobility (Figure 4c), whereas the use of vacuum technology and pure silica dielectric layers are associated with low mobility (Figure 4d). These findings indicate that the choice of fabrication methods and gate electrode materials should be a priority when optimizing OFET parameters.

2.3.3. Lab Advisor for OFET Fabrication

LLMs are typically vast, general-purpose models trained on extensive text corpora and can easily produce misleading and untargeted answers when handling highly specialized, domain-specific queries (Figures S16–S18, Supporting Information). By integrating a domain knowledge base and other useful tools, LLMs can be improved to respond accurately to specific questions, thereby enabling them to provide more pertinent advice. Leveraging a previously established database

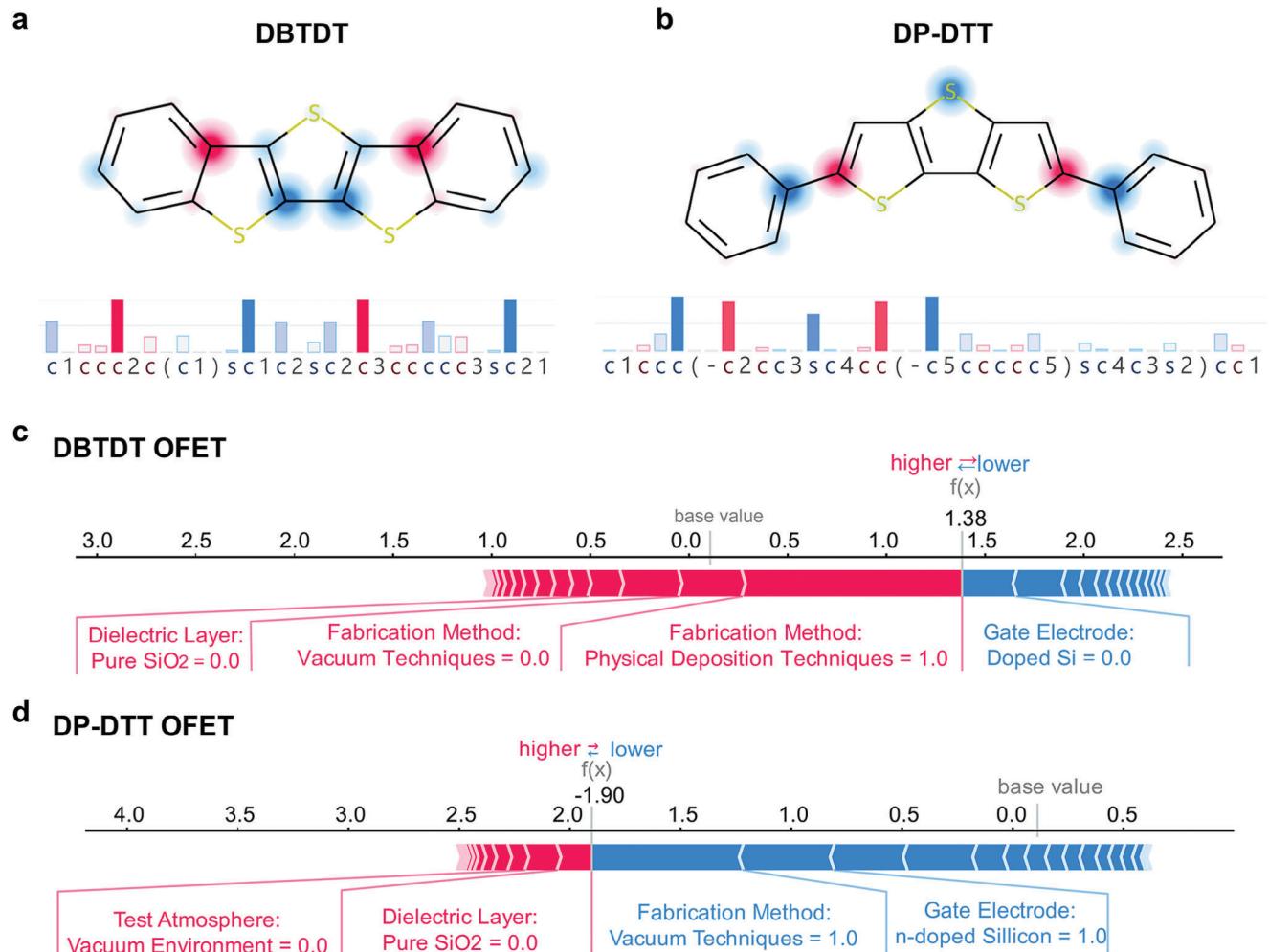


Figure 4. Individual SHAP analysis of predictions from the XGBoost model. a) Molecular characteristic analysis of DBTDT. b) Molecular characteristic analysis of DP-DTT. Here, red represents a positive contribution, blue represents a negative contribution, and the depth of the color indicates the magnitude of the absolute value of the contribution. c) Device parameter analysis for the high mobility setup with DBTDT. d) Device parameter analysis for the low mobility setup with DP-DTT. Features contributing positively or negatively to the predicted high mobility probability are indicated with red and blue arrows, respectively. These arrows illustrate how different features influence the prediction, moving it away from or closer to the baseline, with the arrow length reflecting the feature's contribution magnitude.

of precise OFET fabrication parameters extracted from published literature, we developed a GPT-4-based Lab Advisor using the Assistants API of OpenAI (Figure 5a). Lab Advisor can provide more targeted responses compared with native GPT-4 model (Figure 5b), promising to be a powerful AI assistant in the practical execution of OFET fabrication experiments.

To investigate the properties of the DP-DTT molecule from a theoretical perspective, we conducted first-principle calculations. Figure 6a illustrates the distribution of frontier molecular orbitals and the highest occupied molecular orbital (HOMO)/lowest unoccupied molecular orbital (LUMO) energy levels of the target molecule, indicating a HOMO–LUMO gap of 2.20 eV. Figure 6b shows the transfer integrals of DP-DTT crystals in different transport channels. These results suggest that the intrinsic mobility of the molecule could theoretically reach $1.22 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, whereas the maximum mobility reported in the literature for devices us-

ing DP-DTT as the semiconductor layer is only $0.42 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.^[48] Therefore, these findings imply that the DP-DTT OFETs currently reported in the literature do not fully exploit the intrinsic properties of the material, indicating significant potential for optimization.

Based on our prior SHAP analysis results (Figure 4d), we focused on optimizing the fabrication method and gate electrode. Following data-driven recommendations from Lab Advisor (Figure 5b), we adopted the physical vapor transportation method for device fabrication and chose p-doped Si as the gate electrode. We then manufactured DP-DTT OFETs with a bottom-gate/top-contact configuration (Figure 6c) and assessed their electrical performance. The transfer and output curves (Figure 6d,e) indicate that these improvements tripled the charge transport properties of DP-DTT OFETs, from 0.42 to $1.10 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (Table 2 and Figure S19 (Supporting Information)). These results not only validate the accuracy of our conclusions

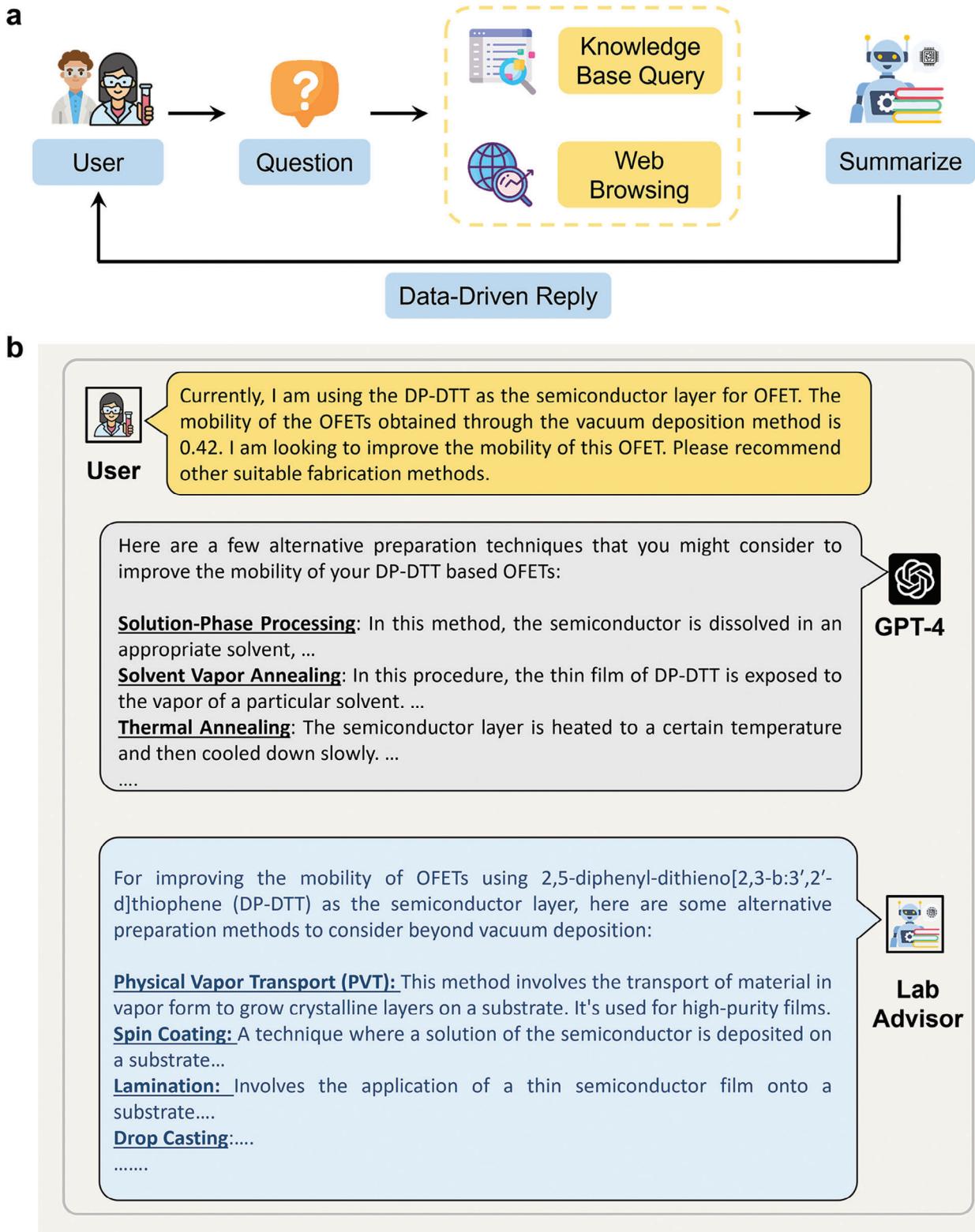


Figure 5. Workflow and examples of Lab Advisor. a) Data-driven Lab Advisor workflow that integrates database queries and web browsing to obtain external information. b) A conversation with the data-driven Lab Advisor and native GPT-4 model.

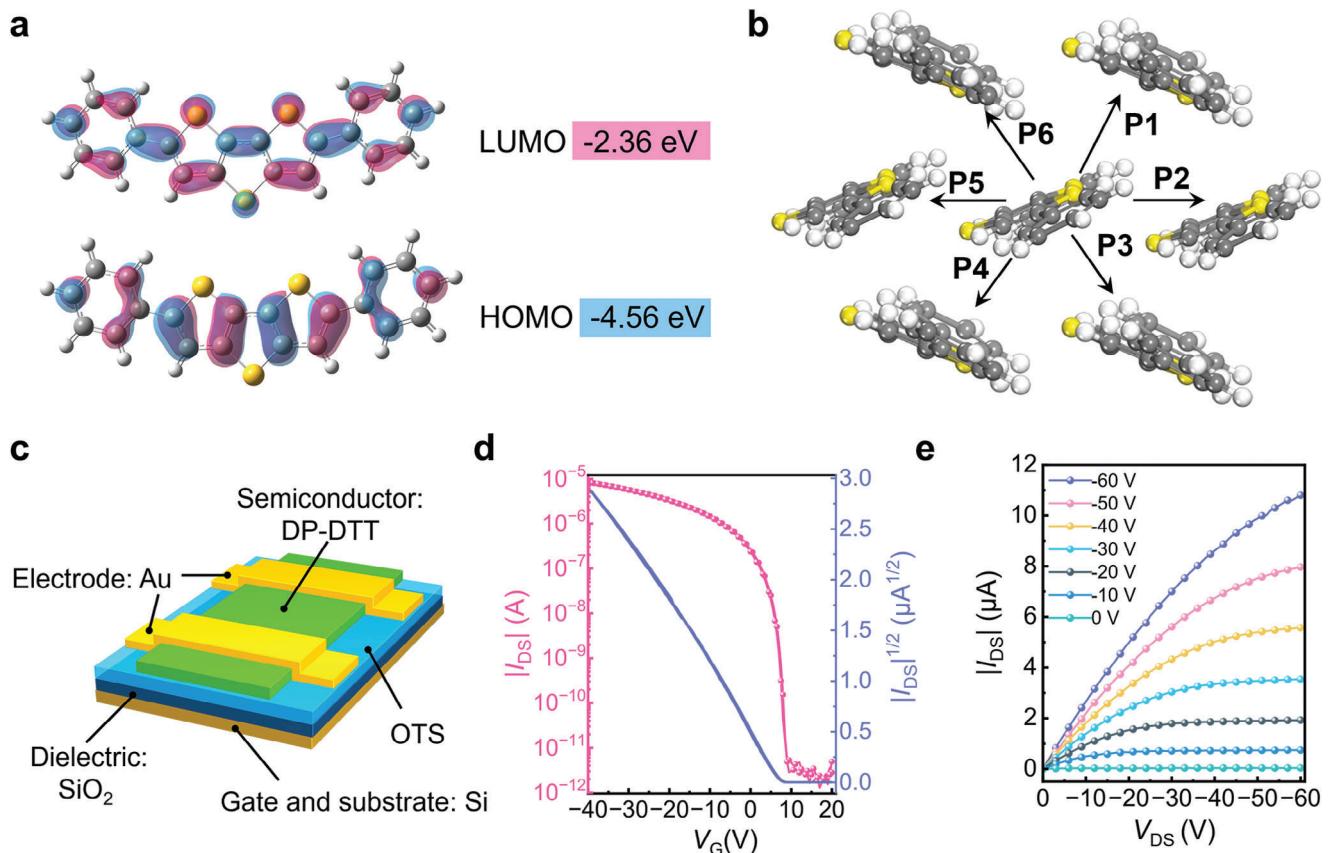


Figure 6. First-principle calculations of DP-DTT and OFET performance after parameter optimization. a) Energy level diagram of the DP-DTT molecule. b) The packing mode of DP-DTT crystals in the xy-plane, with transfer integrals for P1 and P4 at 32.34 meV, P2 and P5 at 3.23 meV, and P3 and P6 at 32.34 meV. c) Schematic illustration of the OFET device incorporating DP-DTT as the semiconductor layer. d) Transfer and e) output characteristic curves of the OFETs. The gate voltage sweep range is from 0 to 60 V, with output curves spaced at 10 V intervals.

derived from machine learning SHAP analysis but also the effectiveness of the Lab Advisor's recommendations. This evidence underscores the considerable potential of our proposed LLM-driven paradigm for advancing OFET device research and its practical applications.

3. Conclusions and Future Outlook

In this study, we have developed an intelligent agent for OFET research by integrating the GPT-4 model with ML technology,

marking a significant milestone in the field of organic semiconductor device research. We ingeniously designed a human-in-the-loop prompt engineering strategy. Utilizing this strategy, along with a pre-coded toolbox, the agent could accurately extract key design parameters from 277 OFET-related academic publications, leading to the successful construction of a database containing 709 unique OFET configurations. Utilizing this database, we further developed three key applications. First, through data visualization techniques, we can more comprehensively track the evolution of semiconductor materials and fabrication methods within the OFET field. Second, we

Table 2. Performance comparison of different DP-DTT OFETs (μ_{\max} : maximum carrier mobility, I_{on}/I_{off} : current on/off ratio, V_T : threshold voltage).

Refs.	μ_{\max} [cm ² V ⁻¹ s ⁻¹]	I_{on}/I_{off}	V_T [V]	Method	Gate electrode	Dielectric
[48]	0.42	5.0×10^6	-23.4	Vacuum sublimation	n-doped Si	Substrate temperatures: 70 °C Surface treatment: OTS
[48]	0.22	10^4	-28.4	Vacuum sublimation	n-doped Si	Substrate temperatures: 25 °C Surface treatment: OTS
[49]	0.36	10^4	-6.3	Physical vapor transport	p-doped Si	Parylene-N as dielectric
[49]	0.20	1.4×10^5	-20	Vacuum sublimation	p-doped Si	Surface treatment: C8-OTS
This work	1.10	4.0×10^6	7.5	Physical vapor transport	p-doped Si	Surface treatment: C18-OTS

trained a machine learning classifier using this database to predict OFET performance and identified key experimental factors affecting device performance through SHAP analysis. Third, we built a GPT-4-based lab assistant using the domain knowledge behind the OFET datasets to provide advice on experimental design. Experimental results demonstrated that with the assistance of the agent, the maximum carrier mobility of DP-DTT OFETs was tripled. This work is an effective practice of LLM in the field of organic optoelectronic devices and expands the research paradigm of organic optoelectronic materials and devices.

Data-driven models will be enhanced by incorporating more high-quality training data. In the future, we plan to expand our dataset with additional publications, enabling our agent to generate more accurate and comprehensive conclusions. We invite researchers to contribute to enriching the database and broadening the scope of application scenarios. This effort will support the advancement of organic semiconductor device technology. Furthermore, we will continue to explore to extract additional material parameters, such as morphology and crystallinity, as well as the impact of different mobility measurement methods, in order to continually enhance the performance of our AI agent. We hope this framework will encourage similar research in other fields, fostering interdisciplinary scientific and technological innovation. Moving forward, we aim to further develop this framework by including more research areas, such as organic light emitting diode technology, and exploring a broader array of applications.

4. Experimental Section

Data Collection and Processing: To eliminate the impact of different measurement methods on mobility and provide useful suggestions for experimental research in device fabrication, literature was used where mobility was calculated using the field-effect method, as it was the most commonly used and device-related method, to construct the OFET database. Papers related to OFET fabrication were retrieved from the Web of Science and these documents were downloaded (for a detailed description of the selection process, please see Notes S3 in the Supporting Information). Given the importance of timeliness in research, the dataset primarily focused on papers published within the last 20 years. Typically, the details of OFET fabrication were included in the main text of the papers, but in some instances, this information could be found in the Supporting Information. Hence, the Supporting Information of each paper was also taken into consideration. In the end, 277 papers and their Supporting Information were selected for data extraction, where only the text information was extracted from the Supporting Information, excluding tables and images.

Evaluation of Text Mining: To evaluate the performance of this method in extracting parameters from OFET literature, precision, recall, and F1 score were primarily employed as the performance metrics, defined as follows

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

In the text mining task for OFET design parameters, each extracted parameter was classified into one of the following three labels: TP (accurately

extracted parameter), FP (incorrectly extracted parameter or irrelevant information extracted), or FN (parameter not successfully extracted). The precision metric measured the accuracy of the method in extracting OFET parameters, the recall metric assessed the completeness of the method in extracting OFET parameters, and the F1 score, which was calculated based on precision and recall, served as an overall representation of the method's performance.

GPT APIs and ML Models: In this study, the GPT API provided by OpenAI was utilized, with all operations conducted in a Python 3.8.8 environment, and the version of the openAI package used was 1.10.0. For image data extraction, GPT-4 V was employed using the gpt-4-1106-vision-preview version, while GPT-4 used for the final data aggregation was the gpt-4-1106-preview version. All parameter settings followed the default configurations of the openAI Python package. XGBoost model was employed using XGBoost package^[50] of 1.7.3 and other ML models were employed using scikit-learn package^[51] of 1.2.2. The models and codes were available on Streamlit cloud (<https://ofet-v1.streamlit.app>) and GitHub (https://github.com/YajingSun-Group/ofet_agent).

Evaluation of ML Models: To assess the performance of the mobility prediction classification model, Equations (1)–(3) were used to calculate the model's precision, recall, and F1 score. In this binary classification model, TPs represented samples correctly predicted as high mobility, FPs represented samples incorrectly predicted as high mobility when they were low mobility, and FNs represented actual high mobility samples that were not correctly predicted by the model. Additionally, the AUROC score was utilized to evaluate the overall performance of the model. The AUROC score was calculated by measuring the model's ability to distinguish between true and false positives across various thresholds, with higher scores indicating stronger discriminative capabilities of the model.

Interpretation of ML Model: After hyperparameter optimization of the XGBoost model, to comprehensively understand the relationship between OFET parameter features and mobility, the SHAP method, developed in the field of game theory, was employed to calculate the SHAP values (i.e., importance values) of each feature in the optimal machine learning model. SHAP values provided an intuitive way to quantify the contribution of each feature, facilitating an understanding of model decisions. For a predictive model with n features, SHAP values calculated the contribution of each feature to the model's prediction by considering all possible combinations of features. The SHAP value for any given feature i was defined as follows

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (n - |S| - 1)!}{n!} [f(S \cup \{i\}) - f(S)] \quad (4)$$

Here, N represents the set of all features, S is a subset of features excluding feature i , and $f(S)$ is the model's prediction given the feature set S . This formula calculated the average marginal contribution of adding feature i to the feature set S , considering all possible combinations of features. SHAP analysis offered a standardized method for explaining individual prediction outcomes and could also be used to interpret the model's overall behavior. The sign of a feature's SHAP value indicated its positive or negative contribution to mobility, with features having larger absolute SHAP values having a more significant impact on high-mobility OFETs.

First-Principle Calculations: The hopping model was employed to describe the carrier transport process, a model widely utilized to elucidate the transport characteristics in organic semiconductors. In this model, charge transport was considered as a nonadiabatic process occurring between two adjacent organic molecules. The carrier hop rate k within a molecular dimer could be expressed as follows

$$k = \frac{2}{\hbar^2} |V|^2 \int_0^\infty dt \exp \left[- \sum_j S_j (2\bar{n}_j + 1) (1 - \cos \omega_j t) \right] \cos \left(\sum_j S_j \sin \omega_j t \right) \quad (5)$$

Here, V represents the transfer integral, indicating the electronic coupling strength between the initial and final states of the electron; ω_j represents the frequency of the j th vibration mode; S_j is the Huang–Rhys factor, a dimensionless factor used to characterize the strength of electron–phonon coupling.

In this work, the initial structures were derived from experimental data. The computational analysis was carried out using the MOMAP 2022B1-3 software^[52] suite in conjunction with the Gaussian 16 Revision C.014 program. To accurately calculate the reorganization energy, a supercell was constructed, and the own n-layered integrated molecular orbital and molecular mechanics (ONIOM) method was employed. The supercell generation was facilitated by molecular material property prediction package (MOMAP) software, which ensured that each unit contained every molecule up to the second nearest neighbor through constraints. The calculation of transfer integrals was based on the M06-2X functional and 6-31G(d,p) basis set, while the reorganization energy calculations utilized the PBE1PBE functional combined with 6-31G(d,p) basis set, methods known to yield reliable results. The carrier mobility was obtained through kinetic Monte Carlo simulations. The energies of the frontier orbitals were calculated at the Perdew–Burke–Ernzerhof hybrid with 25% Hartree–Fock exchange (PBE1PBE)/6-31G(d,p) level.

Si/SiO₂ Substrate Preparation: A p-type Si wafer, heavily doped and layered with 300 nm of SiO₂, underwent an O₂ plasma treatment at a power setting of 80 W for a duration of 5 min. A 5 μL aliquot of octadecyltrichlorosilane (OTS) was applied around the wafer. The assembly was then heated to 120 °C and held under vacuum conditions for 2 h. Following this, the wafers were meticulously cleansed through ultrasonic bathing in chloroform, n-hexane, and isopropanol, each for 30 min.

Single Crystal OFET Construction: DP-DTT single crystals were grown in a physical vapor transport technique within a dual-zone horizontal furnace and evacuated by mechanical pumping vacuum. The DP-DTT was placed in a quartz boat in the high temperature region of the furnace, utilizing high purity N₂ as the delivery medium. DP-DTT single crystals were obtained on the OTS-treated SiO₂/Si substrate in the low-temperature zone. The process parameters included maintaining the high-temperature zone at 130 °C for 150 min with an argon flow rate maintained at 30 sccm. Gold was used as the source/drain electrodes on an individual single crystal and the channel width and channel length were determined via optical microscopy.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Q.Z. and Y.H. contributed equally to this work. This work was supported as part of A Multi-Scale and High-Efficiency Computing Platform for Advanced Functional Materials, funded by the Haihe Laboratory in Tianjin (Grant No. 22HHXCJ00007). The authors acknowledge financial support from the National Key R&D Program (Grant Nos. 2022YFB3603800, 2022YFA1204401, 2021YFB3600700), the National Natural Science Foundation of China (Grant Nos. 52121002, U21A6002), and the GuangDong Basic and Applied Basic Research Foundation (Grant No. 2023A1515110356). The calculations were performed on the National Supercomputer Center in Tianjin (Tianhe 3F) and the Scientific Computing Center of CIC, Tianjin University. The authors also thank “the Fundamental Research Funds for the Central Universities.”

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords

accelerated design, large language models, machine learning, organic field-effect transistors

Received: April 10, 2024

Revised: May 23, 2024

Published online: June 6, 2024

- [1] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C. P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T.-Y. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Veličković, M. Welling, L. Zhang, C. W. Coley, Y. Bengio, M. Zitnik, *Nature* **2023**, 620, 47.
- [2] K. Sanderson, *Nature* **2023**, 615, 773.
- [3] D. A. Boiko, R. MacKnight, B. Kline, G. Gomes, *Nature* **2023**, 624, 570.
- [4] Z. Zheng, Z. Rong, N. Rampal, C. Borgs, J. T. Chayes, O. M. Yaghi, *Angew. Chem., Int. Ed.* **2023**, 62, 202311983.
- [5] K. M. Jablonka, P. Schwaller, A. Ortega-Guerrero, B. Smit, *Nat. Mach. Intell.* **2024**, 6, 161.
- [6] Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, O. M. Yaghi, *J. Am. Chem. Soc.* **2023**, 145, 18048.
- [7] X. Zhang, Z. Zhou, C. Ming, Y.-Y. Sun, *J. Phys. Chem. Lett.* **2023**, 14, 11342.
- [8] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, G. Ceder, *Nature* **2023**, 624, 86.
- [9] J. Choi, B. Lee, *Commun. Mater.* **2024**, 5, 13.
- [10] O. N. Oliveira jr, L. Christino, M. C. F. Oliveira, F. V. Paulovich, *J. Chem. Inf. Model.* **2023**, 63, 7605.
- [11] M. P. Polak, D. Morgan, *Nat. Commun.* **2024**, 15, 1569.
- [12] Z. Hong, *Energy Mater. Adv.* **2023**, 4, 0026.
- [13] Z. Xiao, W. Li, H. Moon, G. W. Roell, Y. Chen, Y. J. Tang, *ACS Synth. Biol.* **2023**, 12, 2973.
- [14] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmuely, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, *Science* **2023**, 379, 1123.
- [15] R. Riveland, A. Pouget, *Nat. Neurosci.* **2024**, 27, 988.
- [16] H. Cui, C. Wang, H. Maan, K. Pang, F. Luo, N. Duan, B. Wang, *Nat. Methods* **2024**, <https://doi.org/10.1038/s41592-024-02201-0>.
- [17] W. Hou, Z. Ji, *Nat. Methods* **2024**, <https://doi.org/10.1038/s41592-024-02235-4>.
- [18] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, D. S. W. Ting, *Nat. Med.* **2023**, 29, 1930.
- [19] R. Wang, H. Feng, G.-W. Wei, *J. Chem. Inf. Model.* **2023**, 63, 7189.
- [20] M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, P. Rajpurkar, *Nature* **2023**, 616, 259.
- [21] F. Wong, C. de la Fuente-Nunez, J. J. Collins, *Science* **2023**, 381, 164.
- [22] D. Van Veen, C. Van Uden, L. Blankemeier, J.-B. Delbrouck, A. Aali, C. Bluethgen, A. Pareek, M. Polacin, E. P. Reis, A. Seehofnerová, N. Rohatgi, P. Hosamani, W. Collins, N. Ahuja, C. P. Langlotz, J. Hom, S. Gatidis, J. Pauly, A. S. Chaudhari, *Nat. Med.* **2024**, 30, 1134.
- [23] K. M. Merz, G.-W. Wei, F. Zhu, *J. Chem. Inf. Model.* **2023**, 63, 5395.
- [24] C. Stokel-Walker, R. Van Noorden, *Nature* **2023**, 614, 214.
- [25] Z. Zhang, W. Wang, Y. Jiang, Y.-X. Wang, Y. Wu, J.-C. Lai, S. Niu, C. Xu, C.-C. Shih, C. Wang, H. Yan, L. Galuska, N. Prine, H.-C. Wu, D. Zhong,

- G. Chen, N. Matsuhisa, Y. Zheng, Z. Yu, Y. Wang, R. Dauskardt, X. Gu, J. B.-H. Tok, Z. Bao, *Nature* **2022**, *603*, 624.
- [26] Y. Jiang, Z. Zhang, Y.-X. Wang, D. Li, C.-T. Coen, E. Hwaun, G. Chen, H.-C. Wu, D. Zhong, S. Niu, W. Wang, A. Saberi, J.-C. Lai, Y. Wu, Y. Wang, A. A. Trotsyuk, K. Y. Loh, C.-C. Shih, W. Xu, K. Liang, K. Zhang, Y. Bai, G. Gurusankar, W. Hu, W. Jia, Z. Cheng, R. H. Dauskardt, G. C. Gurtner, J. B.-H. Tok, K. Deisseroth, et al., *Science* **2022**, *375*, 1411.
- [27] Y. Chen, L. Chen, B. Geng, F. Chen, Y. Yuan, D. Li, Y. Wang, W. Jia, W. Hu, *SmartMat* **2023**, <https://doi.org/10.1002/smm.2.1229>.
- [28] Z. Qin, T. Wang, H. Gao, Y. Li, H. Dong, W. Hu, *Adv. Mater.* **2023**, *35*, 2301955.
- [29] D. Zhong, C. Wu, Y. Jiang, Y. Yuan, M. Kim, Y. Nishio, C.-C. Shih, W. Wang, J.-C. Lai, X. Ji, T. Z. Gao, Y.-X. Wang, C. Xu, Y. Zheng, Z. Yu, H. Gong, N. Matsuhisa, C. Zhao, Y. Lei, D. Liu, S. Zhang, Y. Ochiai, S. Liu, S. Wei, J. B.-H. Tok, Z. Bao, *Nature* **2024**, *627*, 313.
- [30] A. Dance, *Nature* **2009**, *458*, 664.
- [31] L. Hawizy, D. M. Jessop, N. Adams, P. Murray-Rust, *J. Cheminf.* **2011**, *3*, 17.
- [32] D. M. Jessop, S. E. Adams, E. L. Willighagen, L. Hawizy, P. Murray-Rust, *J. Cheminf.* **2011**, *3*, 41.
- [33] J. Mavračić, C. J. Court, T. Isazawa, S. R. Elliott, J. M. Cole, *J. Chem. Inf. Model.* **2021**, *61*, 4280.
- [34] Z. Yang, L. Li, K. Lin, J. Wang, C.-C. Lin, Z. Liu, L. Wang, arXiv: 2309.17421, V2, unpublished, **2023**.
- [35] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, *Nat. Mach. Intell.* **2020**, *2*, 56.
- [36] K. Rajan, H. O. Brinkhaus, M. I. Agea, A. Zielesny, C. Steinbeck, *Nat. Commun.* **2023**, *14*, 5045.
- [37] D. Weininger, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
- [38] Y. Zhang, Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen, L. Wang, A. T. Luu, W. Bi, F. Shi, S. Shi, arXiv: 2309.01219, V2, unpublished, **2023**.
- [39] C. Wang, H. Dong, W. Hu, Y. Liu, D. Zhu, *Chem. Rev.* **2012**, *112*, 2208.
- [40] L. Wang, X. Chen, X. Deng, H. Wen, M. You, W. Liu, Q. Li, J. Li, *NPJ Digital Med.* **2024**, *7*, 41.
- [41] U. Zschieschang, F. Ante, D. Käblein, T. Yamamoto, K. Takimiya, H. Kuwabara, M. Ikeda, T. Sekitani, T. Someya, J. B. Nimoth, H. Klauk, *Org. Electron.* **2011**, *12*, 1370.
- [42] J. Liu, H. Zhang, H. Dong, L. Meng, L. Jiang, L. Jiang, Y. Wang, J. Yu, Y. Sun, W. Hu, A. J. Heeger, *Nat. Commun.* **2015**, *6*, 10032.
- [43] K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, H. Hosono, *Nature* **2004**, *432*, 488.
- [44] H. Sirringhaus, *Adv. Mater.* **2014**, *26*, 1319.
- [45] J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273.
- [46] H. L. Morgan, *J. Chem. Doc.* **1965**, *5*, 107.
- [47] R. Li, L. Jiang, Q. Meng, J. Gao, H. Li, Q. Tang, M. He, W. Hu, Y. Liu, D. Zhu, *Adv. Mater.* **2009**, *21*, 4492.
- [48] Y. M. Sun, Y. Q. Ma, Y. Q. Liu, Y. Y. Lin, Z. Y. Wang, Y. Wang, C. A. Di, K. Xiao, X. M. Chen, W. F. Qiu, B. Zhang, G. Yu, W. P. Hu, D. B. Zhu, *Adv. Funct. Mater.* **2006**, *16*, 426.
- [49] M.-C. Chen, S. Vegiraju, C.-M. Huang, P.-Y. Huang, K. Prabakaran, S. L. Yau, W.-C. Chen, W.-T. Peng, I. Chao, C. Kim, Y.-T. Tao, *J. Mater. Chem.* **2014**, *2*, 8892.
- [50] T. Chen, C. Guestrin, in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining, ACM, New York, USA, **2016**, pp. 785–794.
- [51] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
- [52] Y. Niu, W. Li, Q. Peng, H. Geng, Y. Yi, L. Wang, G. Nan, D. Wang, Z. Shuai, *Mol. Phys.* **2018**, *116*, 1078.