Accelerated end-to-end chemical synthesis development with large language models

Yixiang Ruan[1, 2], Chenyin Lu[2], Ning Xu[1, 2], Jian Zhang[2], Jun Xuan[2], Jianzhang Pan[1, 3], Qun Fang[1, 3], Hanyu Gao[4], Xiaodong Shen[5], Ning Ye[6], Qiang Zhang[2,7] & Yiming Mo[1, 2]*

[1]College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China.

[2]Zhejiang-Hong Kong Joint Laboratory for Intelligent Molecule and Material Design and Synthesis, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 311215, China.

[3]Institute of Microanalytical Systems, Department of Chemistry, Zhejiang University, Hangzhou 310058, China

[4]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong 999077, China

[5]Chemical & Analytical Development, Suzhou Novartis Technical Development Co. Ltd., Changshu 215537, China

[6]Rezubio Pharmaceuticals Co. Ltd., Zhuhai 519070, China

[7]College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

Abstract

The rapid emergence of large language model (LLM) technology presents significant opportunities to facilitate the development of synthetic reactions. In this work, we leveraged the power of GPT-4 to build a multi-agent system to handle fundamental tasks involved throughout the chemical synthesis development process. The multi-agent system comprises six specialized LLM-based agents, including Literature Scouter, Experiment Designer, Hardware Executor, Spectrum Analyzer, Separation Instructor, and Result Interpreter, which are pre-prompted to accomplish the designated tasks. A web application was built with the multi-agent system as the backend to allow chemist users to interact with experimental platforms and analyze results via natural language, thus, requiring zero-coding skills to allow easy access for all chemists. We demonstrated this multi-agent system on the development of a recently developed copper/TEMPO catalyzed aerobic alcohol oxidation to aldehyde reaction, and this LLM multi-agent copiloted end-to-end reaction development process includes: literature search and information extraction, substrate scope and condition screening, reaction kinetics study, reaction condition optimization, reaction scale-up and product purification. This work

showcases the trilogy among chemist users, LLM-based agents, and automated experimental platforms to reform the traditional expert-centric and labor-intensive reaction development workflow.

Introduction

Designing proper chemical synthesis reactions and routes towards target compounds is one of core tasks during both drug discovery and process development, requiring significant time and cost investments[1]. Currently, due to the enormous design space and necessity of experimental validation, this process mainly relies on expert chemists and chemical engineers to go through iterative design-make-test-analyze cycles to identify an efficient synthesis route[2,3]. The multifaceted and complex requirements for synthesis reaction design, such as efficiency, cost, sustainability, safety, scalability, and impurity control, make it hard to formulate this task into a well-defined problem that can be tackled algorithmically and autonomously without customized inputs and decisions from experts[4].

The recent advancement of machine learning (ML) technologies has shown great potential in expediting various subtasks during the synthesis design[5,6]. Notable examples include deep learning based quantitative structure–activity relationship (QSAR) model facilitating drug molecule design[7,8] and catalyst design[9], rapid identification of promising synthetic routes using machine-learning-aided synthesis planning[10,11], guiding automated high-throughput experimental platforms to search for optimal reaction conditions[12–15], and direct translation of multistep synthesis procedures from literature to experimental execution via natural language processing (NLP) models[16]. Despite this rapid involvement of machine learning methods in synthesis related tasks, the monolithic input-to-output nature of existing machine learning methods makes them to only function as powerful single-purpose tools for experts, while the goal of fully autonomous end-to-end synthesis reaction design and development still remains to be realized.

In November 2022, OpenAI released the large language model (LLM) based ChatGPT tool, marking a significant leap towards the artificial general intelligence (AGI). The enormous knowledge and information packed in the LLM enables it to make decisions flexibly according to the complex and non-standardized inputs (prompts). LLM-based agents, characterized by their strong generalization abilities and broad applicability, have demonstrated significant advancements in language proficiency and interaction with humans[17,18]. Motivated by the outstanding performance of these agent, scholars have explored and exploited their capability in the various tasks of chemical and material research, such as literature mining[19–22], molecule and material design[23–25], reaction condition recommendation and optimization[22,26–28], and lab apparatus automation[27–29].

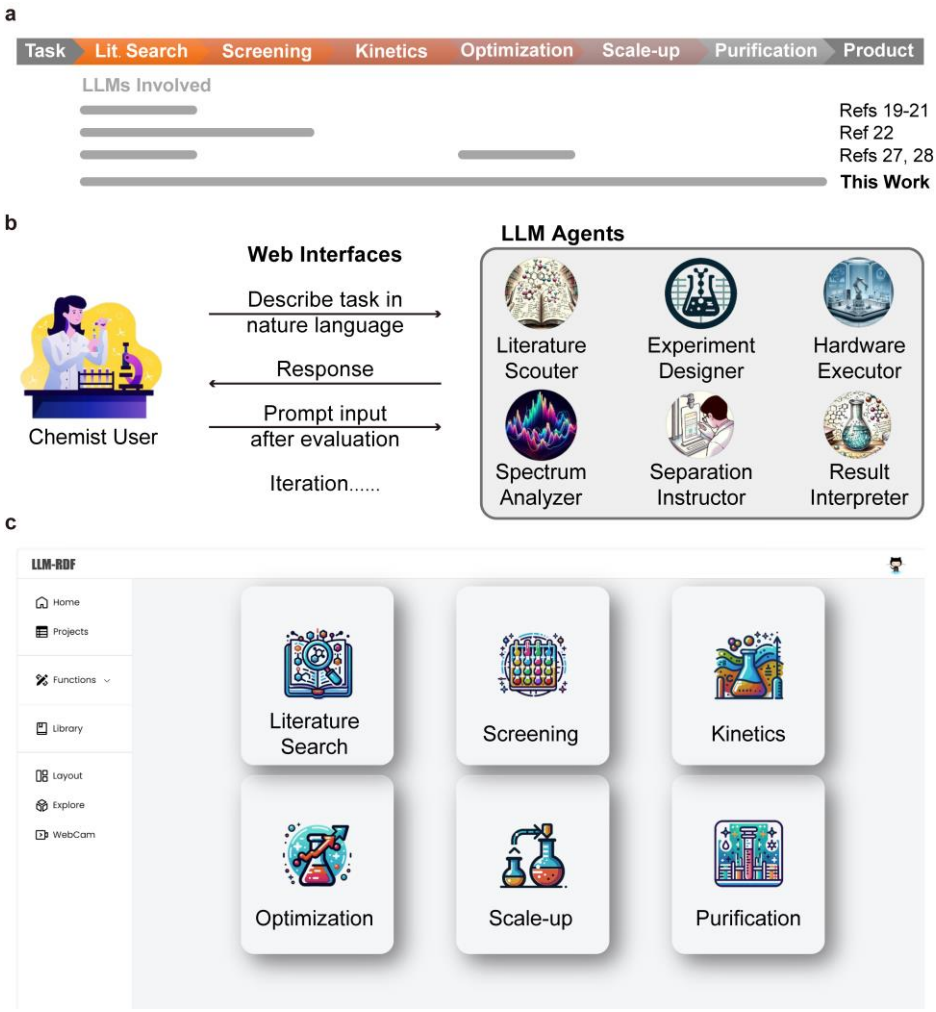

Fig. 1 | Overview of LLM-based multi-agent system for reaction development. a) Workflow for chemical synthesis development facilitated by LLM technology, and comparison with representative published works. (The gray lines denote the involvement of LLMs). **b)** Diagram illustrating the interactions between human chemists and the LLM-based agents for performing tasks in the synthesis reaction development. **c)** The web application with LLM-based agents as backend for end-to-end reaction development.

The existing reports of LLM-based agents showed scattered coverage of the stages in chemical synthesis reaction development but have not presented a path to fully exploit the potential of LLM-based agents in the entire development process. Herein, we demonstrate a unified LLM-based reaction development framework (LLM-RDF) to explore the universality and performance of LLM-based agents in the entire of chemical synthesis reaction development process (Fig. 1a). The findings of this work serve to map out the viable path to the autonomous end-to-end chemical synthesis development using the emerging LLM technology.

Results and discussion

LLM-based agents for end-to-end chemical synthesis reaction development

A typical chemical synthesis reaction development workflow consists of five steps: (1) literature search and information extraction, (2) substrate scope and condition screening, (3) reaction kinetics study, (4) reaction condition optimization, and (5) reaction scale-up and product purification. To exploit the capabilities of LLM facilitating this development process, we developed a set of LLM-based intelligent agents in LLM-RDF to handle the fundamental tasks necessary to complete the development steps above (Fig. 1b). These agents were constructed based on GPT-4 model[30] to maximize their capabilities in context understanding and chemical knowledge reasoning. In addition, they were pre-promoted using customized instructions and documents to achieve consistent behavior and performance for a specific task. These intelligent agents include: (1) Literature Scouter: An agent (based on GPT-4 application, Consensus[31]) leverages LLM's capability to understand user's reaction search request and lookup most relevant literatures in academic journal database, and extract reaction condition and procedures from the literature document. (2) Experiment Designer: An agent translates reaction processes and parameters described in natural language into standardized reaction execution protocols to interface with experimental platforms. (3) Hardware Executor: An agent composes automation hardware running codes according to the reaction protocols such that no manual coding work would be required to execute the experiments. (4) Spectrum Analyzer: An agent processes raw spectral data obtained from analytical apparatus (e.g., gas chromatograph and NMR) to identify the target compound peaks and calculate the corresponding product yield. (5) Separation Instructor: An agent instructs to identify the appropriate thin-layer chromatography (TLC) eluent composition which will be used for the subsequent flash column chromatograph separation. (6) Result Interpreter: An agent interprets and concludes reaction results based on fundamental chemical knowledge.

With the set of LLM-based agents developed above, we created a web application to allow users to access them using natural language in a centralized manner, such that no coding was required during the synthesis reaction development (Fig. 1c). After agents receive prompts and related reference documents from the users describing the chemical task, they will analyze the requests and infer the appropriate responses or solutions through in-context learning[32]. If necessary, they would employ external tools to enhance their capability to respond information out of the scope of the LLM knowledge itself, including Python interpreter, academic database search, and self-driven reaction optimization algorithms. In addition, there is a chain-of-thought mechanism to allow agents to interact with these tools step-by-step, thus maximizing their reasoning capability. Despite the advanced intelligence of GPT-4 model used for these agents, human chemists are still necessary in the decision-

making loop, responsible for evaluating the correctness and completeness of agents' responses and deciding whether to directly implement their suggestions or further communicate with them to tweak the responses.

To explore and evaluate the capabilities of the engineered LLM-based agents, we opted alcohol oxidation to the corresponding aldehyde as the model transformation considering the prevalent presence of hydroxyl and carbonyl functional groups in the fine chemical and pharmaceutical compounds[33]. In particular, the synthesis sustainability has been one of the key considerations of the future green chemistry. Compared to the hazardous oxidants used for alcohol oxidations, such as chromium(VI) compounds[34], Dess-Martin periodinane[35], pyridinium chlorochromate (PCC)[36], and manganese dioxide (MnO_2)[37], molecular oxygen from air is a promising terminal oxidant due to its low cost and high atomic efficiency with minimal wastes generated[38]. Moreover, the implementation of molecular oxygen as the oxidant could operate under milder conditions, thus tolerating sensitive functional groups and enabling improved selectivity in the oxidation of alcohols to aldehydes. Thus, we focused on the search, screening, optimization, and scaling-up of an aerobic alcohol oxidation protocol to demonstrate the key roles of LLM-based agents during such a typical synthesis development process, hoping to illustrate the potentials of LLM in the practical synthesis development beyond the existing chemistry-knowledge-oriented chatting.

Literature search and information extraction

To initiate the synthesis development of the aerobic alcohol oxidation to the corresponding aldehyde, instead of manually finding relevant reports in conventional academic search engines (e.g., SciFinder and Web of Science), we directly input the request to Literature Scouter agent with "Searching for synthetic methods that can use air to oxidize alcohols into aldehydes" prompt. Leveraging vector search technologies, Literature Scouter automatically sifted through the Semantic Scholar database containing over 20 million academic literatures. The use of the Semantic Scholar database instead of relying on the LLM's knowledge (i.e., training data used by OpenAI to train GPT-4) ensured the accuracy of the chemistry details with proper references (Fig. 2b).

Among the various methods given by Literature Scouter (Fig. 2b), we continued to query which method had the greatest potential for practical applications. Literature Scouter recommended the recently developed copper/TEMPO dual catalytic system developed by Stahl group[39] as this method outpaced others in the aspects of the environmental sustainability, simplicity, safety, chemoselectivity, and substrate compatibility (Fig. 4a). After manually evaluating other recommended methods, this copper/TEMPO catalytic chemistry indeed avoids the use of heterogeneous catalysts[40], high-cost

palladium catalysts[41], or light irradiation[42] used in other approaches, proving to have claimed potentials in practical applications as suggested by the Literature Scouter. In addition, the chemoselective oxidation of the target hydroxyl group in diols or polyols is attractive in practice as function group protection and deprotection would not be required, reducing the number of steps required to synthesize the target compound. The literature Scouter recognized the capability of copper/TEMPO catalytic system was able to selectively oxidize primary alcohols in presence of the secondary alcohols on the same molecule (Fig. 2b).

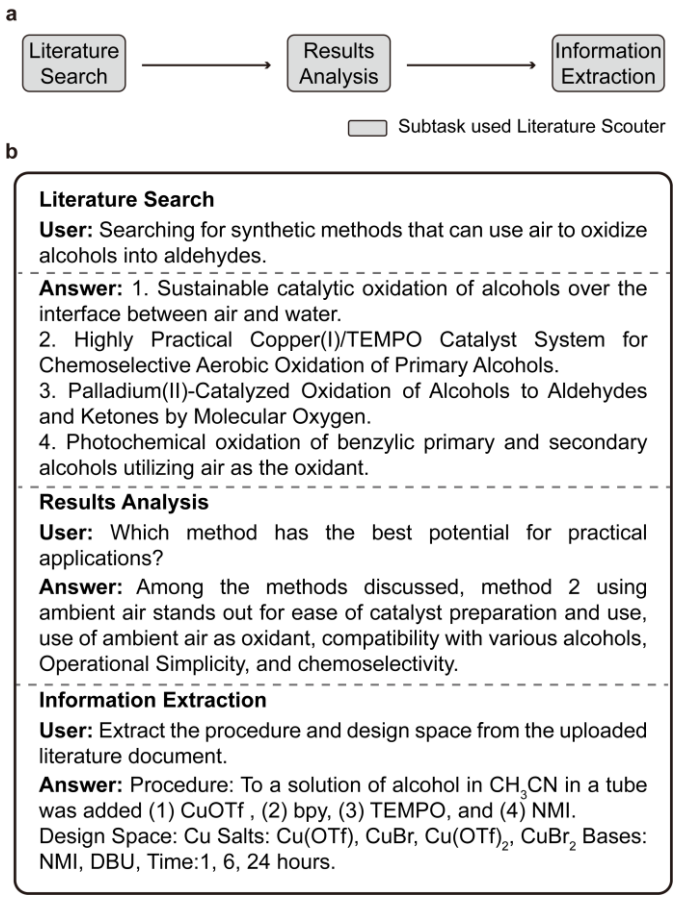

Fig. 2 | LLM-based agents facilitated literature search and information extraction. a) Workflow for literature search and information extraction copiloted by Literature Scouter agent. **b)** The interaction between human chemists with Literature Scouter. The dialogue presented in the figure is simplified for the illustrative purpose, and see details in Supplementary Information Section 2.

Having identified the target transformation, we next turned to extract the detailed reaction conditions for this catalytic system. The literature document was provided to Literature Scouter to summarize the detailed experimental procedures and options for various reagents and catalysts. This information served as the basis for the subsequent experimental exploration of this chemistry (Fig. 2b).

As demonstrated in the task of method search and information extraction from literature, Literature Scouter demonstrated its capability to assist researchers to identify the possible

methodologies necessary to achieve the target transformation under desired conditions, and extracting the required experimental details for executing the reaction. Compared to conventional workflow for identifying the proper chemistry from literature database, Literature Scouter alleviated the labor-intensive tasks of literature searching and reviewing, thus significantly expedited the process and lowered the expertise requirement. Especially, when Literature Scouter was connected to an up-to-date academic journal database, it could propose the new chemistries that were not included in the LLM model pre-training process (Supplementary Information Section 2.1-2.2).

Methodology substrate scope and condition screening

With the literature reported aerobic alcohol oxidation protocol in hand, understanding the substrate scope under various reaction conditions for a methodology is essential for selecting the suitable reaction conditions based on the target compound structure in practical synthesis. It is typically challenging to predict the reaction yield based on first-principle theories, while recently emerging machine learning based methods need a decent amount of experimental data to train the neural model for accurate predictions[43–45]. Thus, the second step for synthesis development process generally involves the exploration of substrate scope for a specific protocol. The recent development of automated high-throughput screening (HTS) technology has been proven as a powerful tool to accelerate the experimental data acquisition for these substrate scope studies[46,47]. However, HTS technology is still not a routine tool that synthesis practitioners would use on their daily reaction development workflows. Apart from the high costs of the required HTS hardware (e.g., liquid handler platform, automated synthesis platform, and HTS analytical apparatus), the time-consuming programming for executing the automation platforms and manual analysis of large amount of HTS results create barriers for chemists with minimal coding experience to use HTS technology in their routine workflows.

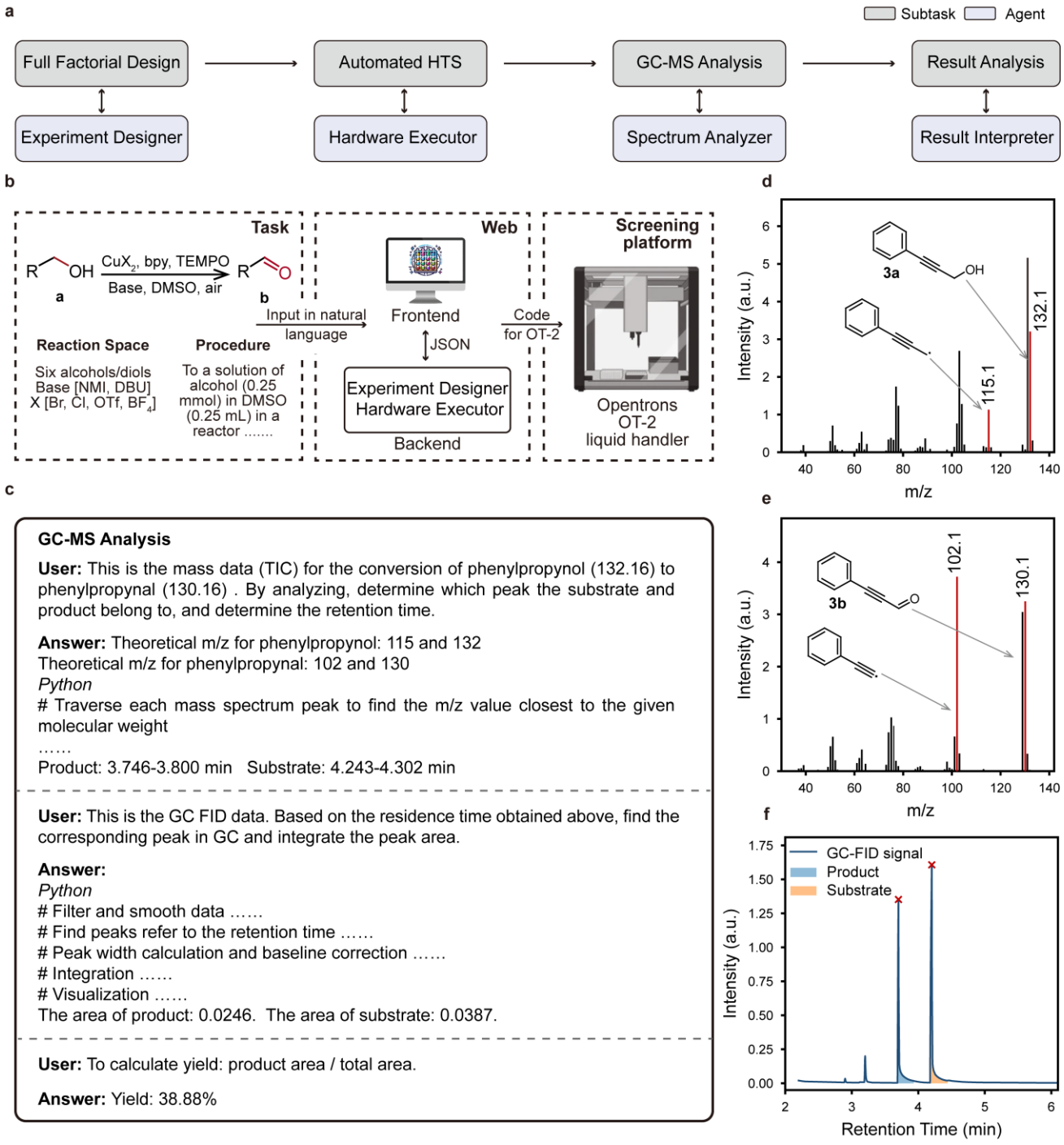

Fig. 3 | LLM-based agents facilitated substrate scope and condition screening. a) Workflow for substrate scope and condition screening copiloted by Experiment Designer, Hardware Executor, Spectrum Analyzer, and Result Interpreter agents. **b)** The aerobic alcohol oxidation reaction screening task described in natural language for subsequent LLM-based agent understanding and OT-2 liquid handler reaction execution. **c)** The interaction between human chemists with Spectrum Analyzer for GC-FID-MS result analysis (see detailed interaction dialogues in Supplementary Information Section 3.2.2). **d)** The mass spectra for the retention time within 4.243-4.302 min, matched with substrate **3a** by Spectrum Analyzer. **e)** The mass spectra for retention time within 3.746-3.800 min, matched with product **3b** by Spectrum Analyzer. **f)** Visualization of peak area integration for substrate **3a** and product **3b** by Spectrum Analyzer using GC-FID data.

To tackle the above-mentioned challenges, we implemented Experiment Designer, Hardware Executor, Spectrum Analyzer, and Result Interpreter agents to automate HTS investigation of the substrate scope, such that the barrier for routine usage of HTS technology could be significantly lowered. The HTS substrate scope study consists of a series of subtasks, including HTS experiment design, automated HTS experiments, gas chromatography (GC) analysis, and results analysis (Fig. 3a).

In HTS experiment design, Experiment Designer agent parsed the HTS experiment task described in natural language into a standardized experimental procedure that could be displayed in the web application (Supplementary Information Section 3.1.1). To execute the HTS task, we chosen the Opentrons liquid handler (OT-2) as the automated reaction screening platform since the Cu/TEMPO catalyzed aerobic alcohol oxidation reaction only involved soluble reagents. In addition, the OT-2 liquid handler has a well-written Python API document, based on which Experiment Executor agent could compose liquid handler running code. Thus, Hardware Executor converted the standardized HTS experimental procedure to OT-2 execution codes to load the necessary labware and pipettes, plan the storage locations for stock solutions, prepare the reaction mixtures as dictated by the experimental procedures, and shake the well plate to perform the aerobic alcohol oxidation. With the seamless transition from HTS task described in natural language to automated reaction execution, two rounds of HTS experiments were conducted, and each round contained a full factorial screening of six alcohol substrates (six monols for the first round and six diols for the second round), four copper catalysts [$CuCl_2$, $CuBr_2$, $Cu(OTf)_2$ and $Cu(BF_4)_2$], and two bases [N-Methylimidazole (NMI) and 1,8-diazabicyclo-[5.4.0]undec-7-ene (DBU)].

After the HTS experiment, the products were characterized with gas chromatography with parallel flame ionization detector and mass spectrometer (GC-FID-MS). The use of parallel FID and MS detectors enabled the simultaneous identification and quantification of the components in the reaction crudes. Instead of labor-intensive manual identification of peaks for reactants and peaks and yield calculation, Spectrum Analyzer agent was used to automated this process (see details in Supplementary Information Section 3.2.2). Specifically, GC-FID-MS analysis instructions and the raw chromatogram data, including FID intensity chromatogram and total ion chromatogram (TIC) from MS detector, were provided to Spectrum Analyzer. It could identify the corresponding reactant and product peaks in TIC by looking for their characteristic fragmentation patterns, and calculated the reaction yield based on FID intensity chromatogram. With phenylpropynol (**3a**) and the corresponding product phenylpropynal (**3b**) as an example, Spectrum Analyzer thought that **3a** should have a 132 mass to charge (*m/z*) ratio signal for the molecule itself and 115 *m/z* signal for the fragment resulting from the loss of a hydroxyl group, and **3b** should have 130 *m/z* signal for the molecule itself and 102 *m/z* signal

for the fragment resulting from the loss of the carbonyl group. Subsequently, Spectrum Analyzer wrote a Python code to search the TIC data for mass spectrometry peaks containing the characteristic *m/z* signals and determine the retention times of the substrate and product (Fig. 3d-e). Next, Spectrum Analyzer integrated the FID peak areas at the substrate and product retention times to determine the reaction yield (assuming that the response factors of the products and substrates are the same in FID) (Fig. 3f). The yields obtained by Spectrum Analyzer were consistent with those derived from manual analysis using chromatography software.

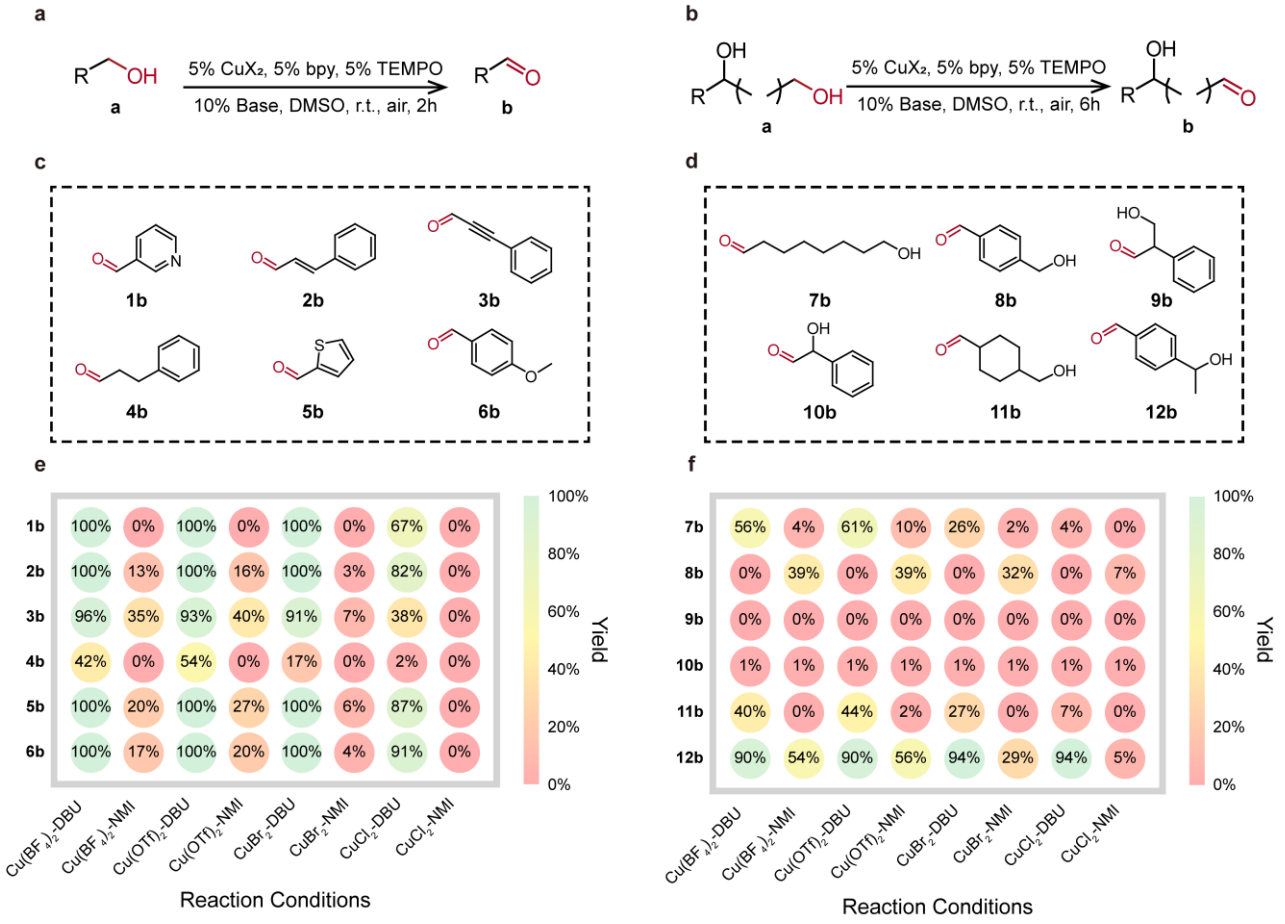

Fig. 4 | The substrate scope and condition screening results. The copper/TEMPO-catalyzed aerobic oxidation reaction of **a)** monohydric alcohols and **b)** diols to the corresponding aldehydes in the screening task. Reaction condition: substrate (0.25 mmol), 5 mol% Cu catalyst, 5 mol% bpy, 5 mol% TEMPO, and 10 mol% base were dissolved in DMSO solvent (1.25 mL), and reaction was performed under room temperature and open to air for 2 (monohydric alcohols) or 6 (diols) hours. The aldehyde products derived from the oxidation of the corresponding **c)** monohydric alcohol and **d)** diols. Yield heat maps for the oxidation of **e)** monohydric alcohols and **f)** diols under various combinations of copper catalyst and base.

Finally, we utilized Result Interpreter agent to summarize HTS results and explain observed patterns based on fundamental chemistry knowledge (Supplementary Information Section 3.3.2). Result Interpreter recognized that DBU base significantly outperformed NMI, and the reactivity of

copper salt followed the order of $CuCl_2$ < $CuBr_2$ < $Cu(OTf)_2$ ~ $Cu(BF_4)_2$. In addition, it concluded that electron-withdrawing functional groups near the hydroxyl group (e.g., aromatic rings or unsaturated carbon bonds) could increase the oxidation reactivity, which was consistent with chemistry principles[48,49]. However, Result Interpreter's ability to conduct further in-depth analysis was still limited with existing GPT-4 model as the backend. For example, in explaining why diol **9a** and **10a** exhibited no reaction in any condition tested, it could only suggest superficially that the arrangement of functional groups or the spatial configuration of the molecules might play a role. The literature-proposed mechanism involves the chelation of copper catalyst by the vicinal diol substrates (**9**-**10a**) to form an unreactive Cu-phenolate species, thus deactivating the copper catalyst[39]. This limitation should stem from that GPT-4 was trained with publicly available information (i.e., only including general chemistry knowledge), and advanced coordination chemistry knowledge and reaction specific behavior reported by professional chemistry journals were beyond the knowledge space of the GPT-4. However, this lack of advanced chemistry knowledge should be solved by further feeding the LLM with domain-specific information.

Reaction kinetic modeling

As mentioned earlier, this copper/TEMPO catalytic system prefers to oxidize primary hydroxyl group compared to secondary hydroxyl group. We observed that dimethyl sulfoxide (DMSO) solvent (used in the HTS experiment) gave superior primary alcohol (**12a**) oxidation chemoselectivity compared to acetonitrile (MeCN) solvent (used in the literature[39]) (Fig. 5b). To investigate the observed solvent effects, Experiment Designer agent suggested that we could conduct oxidation kinetics study for different solvent. Recently, automated kinetic profiling has become an efficient tool to help researchers establish reaction kinetic models[50–52]. However, similar to the HTS technology discussed above, it is still not a routine tool used in process development due to the high entry barrier for mastering automated hardware and intricate programming involved in fitting kinetics models. In order to avoid extensive coding and data analysis, Experiment Designer, Hardware Executor, Spectrum Analyzer, and Result Interpreter agents orchestrated to complete the kinetic study task, consisting of subtasks including kinetics experiment design, automated sampling experiments, proton nuclear magnetic resonance (^{1}H NMR) analysis, and kinetic model fitting and analysis (Fig. 5a).

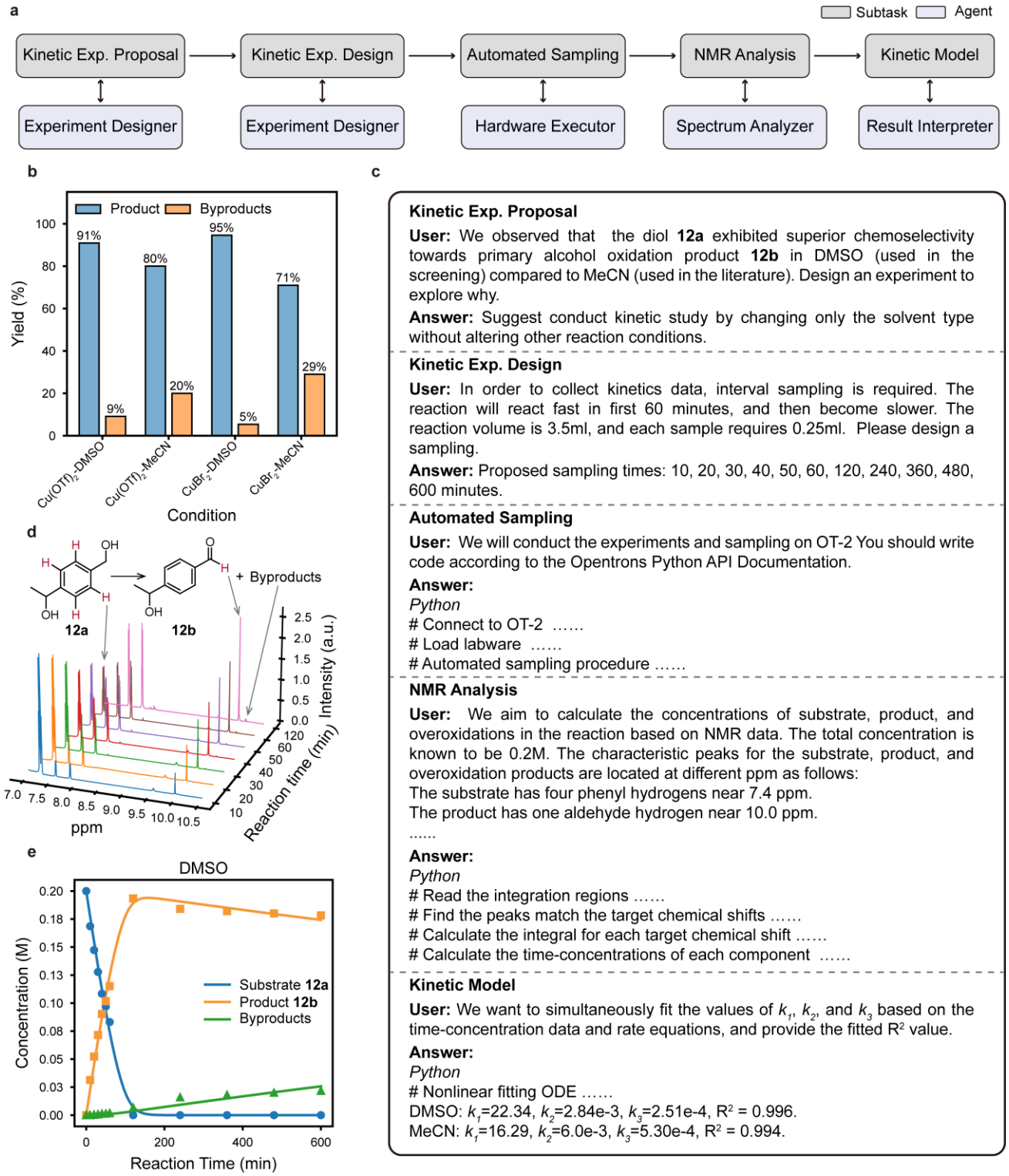

Fig. 5 | LLM-based agents facilitated reaction kinetic study. a) Workflow for reaction kinetic modeling copiloted by Experiment Designer, Hardware Executor, Spectrum Analyzer, and Result Interpreter agents. **b)** Comparison of the yields for products and overoxidation byproducts of diol **12a** for different and copper catalysts. **c)** The interaction between human chemists with LLM-based agents for reaction kinetic study (see detailed interaction dialogues in Supplementary Information Section 4). **d)** Characteristic ^{1}H NMR peaks identified by Spectrum Analyzer for calculations of reaction samples' compositions. **e)** The time-course concentration profile in DMSO solvent, and the fitted reaction kinetic curves given by Result Interpreter.

In kinetics experiment design, Experiment Designer planned a sampling schedule for time-course data collection. To provide an approximate reaction rate information for experimental design, we firstly monitored the reaction via TLC, and found that substrate **12a** was rapidly consumed within the initial first hour reaction time and the reaction slowed down afterwards. Based on this observation, Experiment Designer proposed a sampling schedule spanning a 10-hour reaction period. Samples were to be collected at 10, 20, 30, 40, 50, 60, 120, 240, 360, 480, and 600 minutes, such that denser data points could be obtained during the early stage of the reaction when the reaction rate was large. Subsequently, Hardware Executor agent generated the OT-2 running code based on the experimental design proposed by the Experiment Designer. The coded OT-2 liquid handler procedure contained a series of operations for sampling, such as stopping the reaction's shaking, pipetting to sample, quenching the reaction in the sample, and resuming shaking. The compositions of the sampled reaction crude were analyzed by ^{1}H NMR. Instead of manual analysis of the NMR data, we provided Spectrum Analyzer agent with ^{1}H NMR spectra and approximate chemical shifts for characteristic hydrogen atoms in the substrate, product, and byproducts (overoxidation products). Spectrum Analyzer wrote a Python program according to the API documentation for the TopSpin NMR processing software to automate the analysis of NMR data, the procedure of which included identifying target peaks, performing peak integration, and calculating the compositions of the samples (Fig. 5d).

Next, providing the obtained kinetics experiment results to Result Interpreter agent, it fitted the time-course data to the kinetic model equations. The reaction rate for substrate to product followed saturation kinetic dependence on the substrate alcohol (Eq.(1)[53], and in addition, the product overoxidation was assumed to be a first-order reaction (Eq.(2). Result Interpreter calculated the corresponding reaction rate constants (k_1, k_2, k_3), and the proposed kinetic models fitted well with the experimental data with the coefficients of determination (R^2) of 0.995 (Fig. 5e and Fig. S12).

$$r_{Product} = \frac{k_1 k_2 C_{Substrate}}{1 + k_1 C_{Substrate}} \quad (1)$$

$$r_{Byproducts} = k_3 C_{Product} \quad (2)$$

Result Interpreter further concluded that the rate constant for the product overoxidation (k_3) was larger in MeCN than that in DMSO, indicating that the product overoxidation rate had strong dependence on the reaction solvent choice (Supplementary Information Section 4.5). This analysis highlighted that the Result Interpreter agent had the ability to understand the underlying kinetics behind the observed the reaction selectivity. Even though we provided the kinetic model to the Result Interpreter here without demonstrating the agent's ability to identify the suitable kinetic model among

various possible kinetic models, the self-driven kinetic model identification has already been shown to be feasible using conventional statistical or machine-learning algorithms[54,55], which should be able to be integrated into the LLM kinetic analysis workflow.

Self-driven reaction condition optimization

When a specific target compound is determined for process development towards manufacturing, reaction condition optimization is necessary to improve the synthesis efficiency along with other considerations (e.g., costs and impurity generation)[56]. Instead of traditional manual one-factor-at-time (OFAT) optimization, the recent development of optimization algorithms, such as Bayesian optimization (BO)[13,15,57], Nelder–Mead Simplex[58], stable noisy optimization by branch and fit (SNOBFIT) algorithm[59], and the mixed-integer nonlinear program (MINLP) algorithm[60], have enabled the automated experimental platforms to perform closed-loop reaction optimization in an autonomous manner. However, akin to the HTS technology mentioned previously, the steep learning curve associated with mastering automated hardware and optimization algorithms prevents the widespread adoption of the self-driven reaction optimization workflow as a routine tool in process development, despite its demonstrated effectiveness.

To address this challenge, we employed Experiment Designer and Hardware Executor agents as the backend of a reaction optimization module within our developed web application, such that users could interface with the reaction optimization hardware system via natural language. This hardware system is a robotic platform capable of performing end-to-end reaction and analysis, and the closed-loop reaction optimization was driven by a Bayesian optimization algorithm. Specifically, an automated synthesis equipment (Unchained Big Kahuna) conducts the chemical reactions, which are then analyzed by a high-performance liquid chromatography (HPLC) to provide result feedbacks to the BO for suggesting the next-round reaction candidates. The capability of LLM as an optimizer has been evaluated in recent publications. However, although the LLMs have shown superior performance for optimizing reactions with clear kinetics or prior knowledge, they still fell behind statistical optimization algorithms (e.g., Bayesian optimization) for complex reaction systems[26,28].

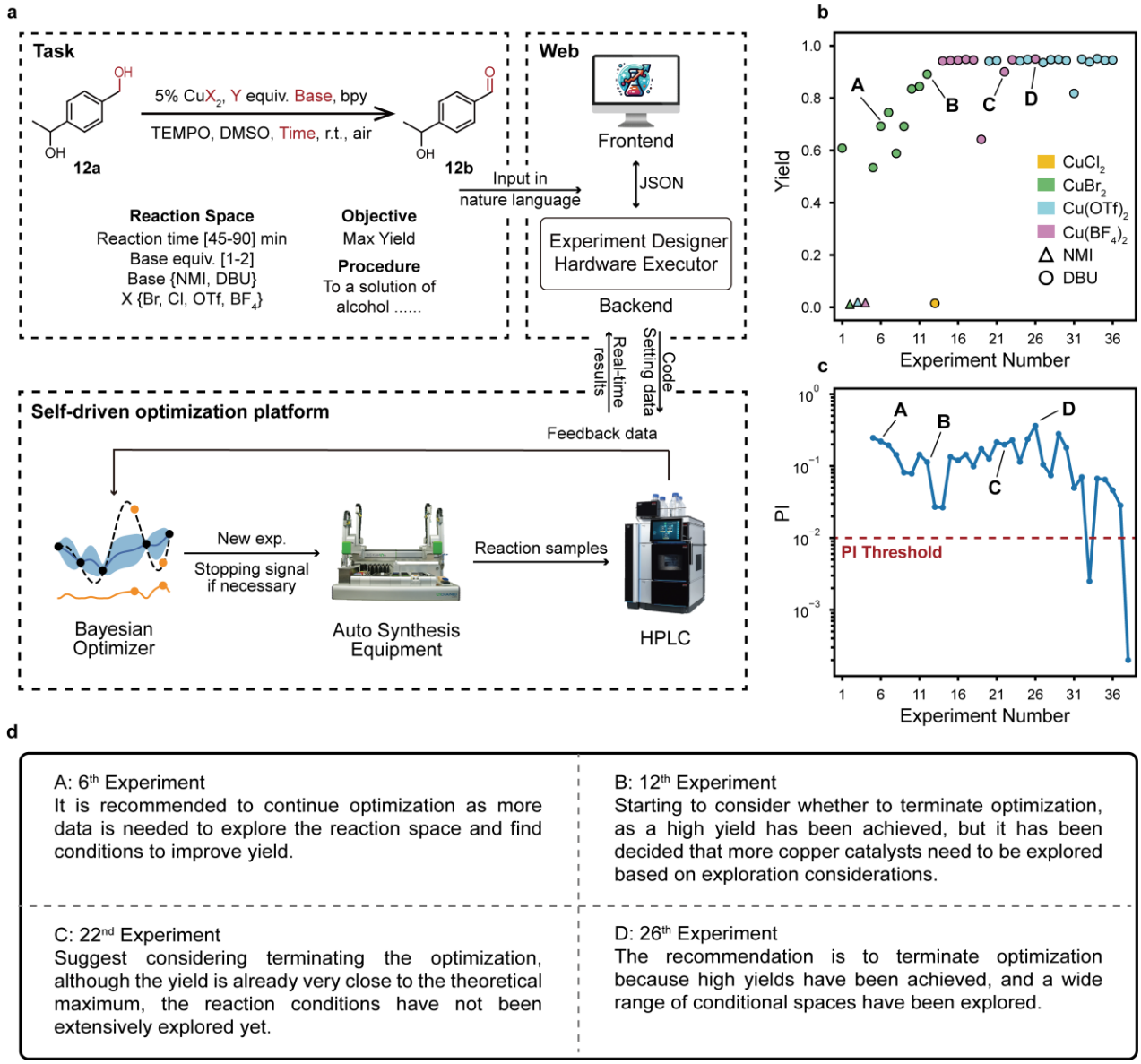

Fig. 6 | LLM-based agents facilitated self-driven reaction condition optimization. a) The LLM-based agents copiloted self-driven reaction optimization system. The evolution profile of **b)** yield and **c)** PI value during the closed-loop reaction optimization process driven by the BO. **d)** Result Interpreter's recommendations on whether reaction optimization should be terminated at 6th, 12th, 22nd, and 26th experiment (see detailed interaction dialogues in Supplementary Information Section 5.3).

To demonstrate LLM-based agents copiloted reaction optimization workflow (Fig. 6a), we conducted the condition optimization for the selective oxidation of diol (**12a**) to the corresponding mono-oxidized aldehyde product (**12b**). The reaction design space included two continuous variables (i.e., equivalents of base and reaction time) and two categorical variables (i.e., types of bases and copper catalysts). The optimization objective is to maximize the reaction yield of **12b**. First, Experiment Designer translated synthesis procedure description [To a solution of alcohol (0.25 mmol) in DMSO (0.25 mL) in a reactor was added sequentially a solution of (1) CuX_2/bpy(0.25 mL, 0.05M), (2) TEMPO (0.25 mL, 0.05M), and (3) NMI (0.25 mL, 0.10M)] into standardized JavaScript Object

Notation (JSON) procedure steps for the automated synthesis device. Hardware Executor generated code templates based on these JSON procedure steps to define the automated synthesis platform operation workflows. Next, Experiment Designer converted the optimization parameter space described in natural language [I want to optimize four variables: 1. Reaction time: 45-90 minutes, 2. Base volume: 0.125-0.25 mL, 3. Cu catalyst: $CuCl_2$, $CuBr_2$, $Cu(OTf)_2$, $Cu(BF_4)_2$ 4. Base type: NMI, DBU.] into JSON format that was used as inputs for the Bayesian optimizer (Supplementary Information Section 5.1.1). At last, users reviewed the entire experimental plan before running the reaction optimization.

The self-driven optimization system iteratively conducted reactions and proposed candidate experiments based on existing reaction results, thus gradually improving the reaction yield of **12b** (Fig. 6b). Multiple high-yield reaction conditions were identified within the design space (Table S2). To automatically stop the reaction optimization task when the expectation of further yield improvement was diminished, we compared the statistical stopping criterion and stopping decision given by the LLM-based agent Result Interpreter. The probability of improvement (PI) metric, a typical statistical stopping criterion[61,62], was first examined by stopping the optimization when the cumulative number of proposed reaction conditions with PI values below 0.01 reached two. This PI stopping criterion was met after completing 36 experiments (Fig. 6c), based on which the optimal conditions should be confidently identified. In comparison, Result Interpreter was used to determine the appropriate stopping point for the optimization task using the concept of balancing exploration and exploitation for black-box function optimization (Supplementary Information Section 5.3). During the exploitation of $CuBr_2$-DBU combination (after 12 experiments), Result Interpreter indicated that the yield was sufficiently high to consider stopping optimization, however, it still recommended further exploration in copper catalysts based on exploration considerations. Then, the BO continued to explore two more catalysts (i.e., $Cu(BF_4)_2$ and $Cu(OTf)_2$). After several small condition adjustments proposed by the BO near the high-yield conditions, the reaction yield did not increase significantly, and a yield decrease was observed in the 22nd experiment. Result Interpreter once again suggested considering the cessation of the optimization. After the 26th experiment, Result Interpreter assessed the reaction yield as sufficiently high and the exploration of the reaction space as comprehensively executed, explicitly recommending the termination of further optimization (Fig. 6d). This comparison showed that the optimization stopping suggestions given by Result Interpreter agent were more intuitive and also required less experiments to identify high-yield reaction conditions compared to the PI stopping criterion. Unlike the PI stopping criterion relying on human experience to pre-define the stopping

threshold (improper selection may lead to poor optimization results or prolonged optimization time), utilizing Result Interpreter to terminate optimization offers better flexibility and adaptability.

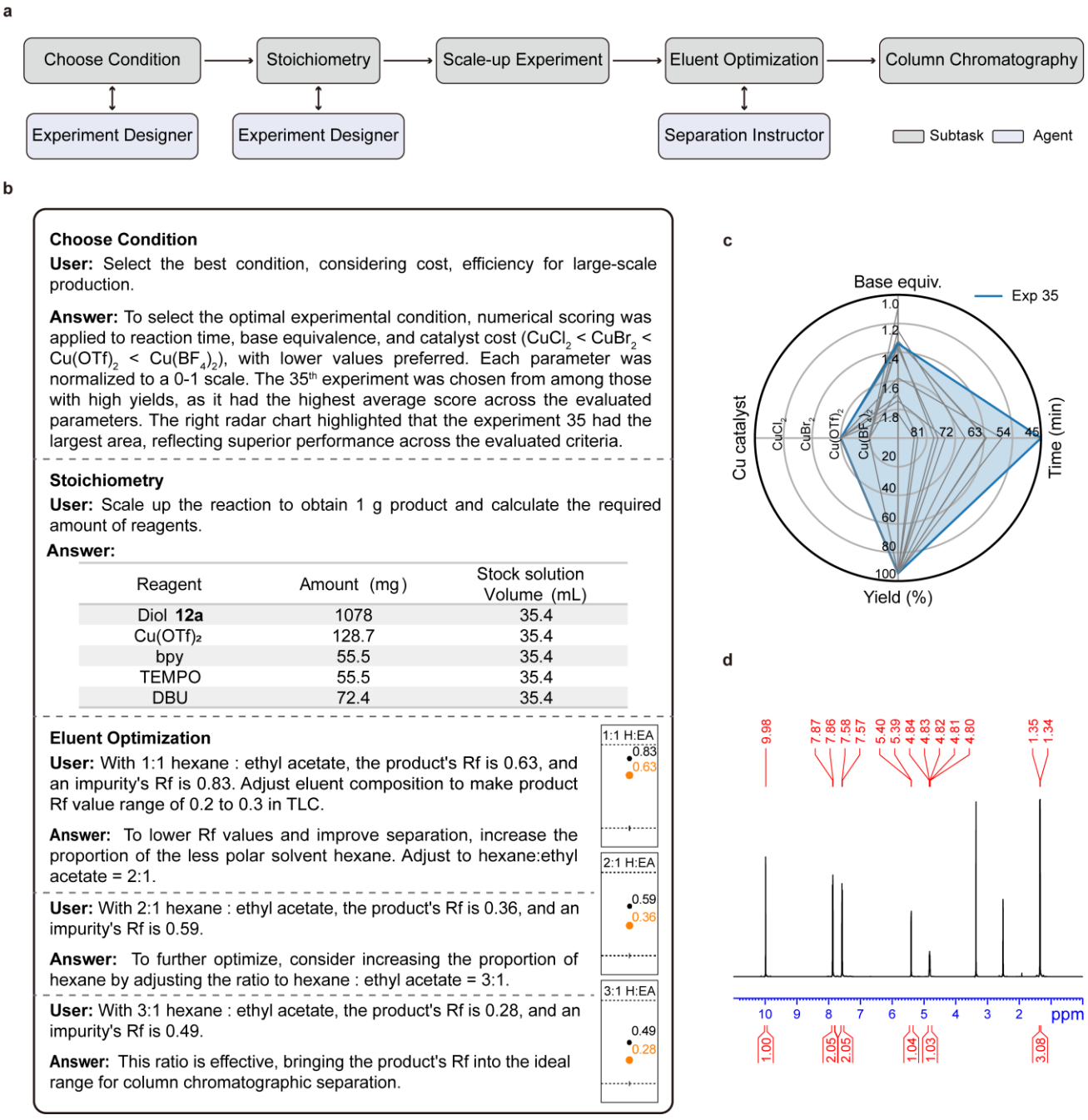

Fig. 7 | LLM-based agents facilitated reaction scale-up and product purification. a) Workflow for reaction scale-up and product purification copiloted by Experiment Designer and Separation Instructor agents. **b)** The interaction between human chemists with Experiment Designer for reaction scale-up and Separation Instructor for finding the optimal eluent composition (see detailed interaction dialogues in Supplementary Information Section 5). **c)** Radar chart of the reaction conditions for selecting the most potent reaction condition for scale up. **d)** ^{1}H NMR spectrum of the purified target product (**12b**) in DMSO-d6.

Reaction scale-up and product purification

In the process development, the scale-up investigation serves as a critical phase to determine whether a small-scale chemistry is suitable for further large-scale synthesis with similar reaction efficiency[63]. Here, we used the high-yield reaction conditions found in the previous reaction optimization task for targeting 1 gram scale synthesis of the compound **12b** to demonstrate the utility of LLM-based agents in facilitating the reaction process development.

Among various high-yield conditions during the condition optimization of diol oxidation, Experiment Designer selected the conditions used in 35th experiment for scaling up (Fig. 7b). The choice of reaction conditions was made based on the preference to the high product yield, short reaction time, and low catalyst and reagent costs. The 35th experiment used a 45-minute reaction time, $Cu(OTf)_2$ catalyst, and 1.34 equivalent DBU base, achieving a high yield of 94.5% (Fig. 7c). To showcase LLM's ability to facilitate reaction scale-up, Experiment Designer accurately calculated the stoichiometries of the reagents based on selected the reaction condition to produce 1 gram of the desired product (Fig. 7b). We conducted the scale-up experiment according to the stoichiometries proposed by Experiment Designer.

Prior to the product purification using flash column chromatography, the optimal eluent composition is typically determined with manual TLC. TLC fine-tunes the eluent polarity to ensure that the retention factor value (R_f value) of the target compound falls within 0.2-0.3, and, at the same time, impurities are separated from the target compound. A recent publication has applied machine learning model to predict the R_f value of a given compound structure in different eluent compositions[64]. However, due to the inevitable prediction inaccuracy, this data-driven prediction model can only serve to provide good initial eluent composition guesses to try, and chemists still need to determine the eluent suitable for practical separation processes by conducting iterative trial-and-error experiments based on their own experience and the polarity-controlled separation principles in TLC. To enable automated identification of optimal eluent composition, we implemented Separation Instructor agent to replace chemists for making eluent composition decisions during the iterative TLC experiment. Here, TLC experiments were performed manually, but the automated TLC device is commercially available to achieve closed-loop optimal eluent composition identification in an autonomous manner. Upon inputting the initial TLC outcome of **12b** separation at hexane : ethyl acetate = 1:1 ratio into Separation Instructor, it advised to reduce the polarity of the eluent to decrease the R_f value of **12b**. Following two iterative decision-and-experiment rounds, Separation Instructor finalized the eluent composition (hexane : ethyl acetate = 3:1), under which the product's R_f value was 0.28 with 0.49 Rf value for the impurity, providing a sufficiently large difference for effective separation. Subsequently, this optimal

eluent composition was used in the automated preparative column chromatography system to successfully separate the product, yielding 915 mg of the product (**12b**) with the isolated yield of 86% and a purity >98% (Fig. 7d).

Conclusion

In this work, we utilized a large language model (LLM)-based multi-agent system to demonstrate the end-to-end development of sustainable aerobic alcohol oxidation, from methodological search to product purification. The specialized LLM-based agents showcased their versatility in autonomous chemical research, undertaking tasks such as synthesis method search, code composing for automated equipment, spectrum signal processing and analysis, reaction stoichiometric calculation, optimization of separation eluent composition, and deriving chemically informed conclusions. The LLM-based multi-agent system demonstrates a transformative approach to chemical synthesis that integrates user chemists, LLM-based agents, and automated experimental platforms, significantly streamlining the traditional expert-driven and labor-intensive workflow of reaction development. Although the LLM technology is still nascent in chemistry applications primarily due to the lack of the advanced and professional chemistry knowledge in the existing LLM training dataset, we would envision that this work outlines a viable avenue to a deeper engagement of LLM technology in reaction development and relevant fields in the future.

Methods

Construction of LLM-based agents

LLM-based agents developed in this work were based on OpenAI's GPT-4 model. (1) Literature Scouter: This agent was developed using Consensus[31] available from OpenAI's GPT store, which can access Semantic Scholar database for academic literatures. (2) Experiment Designer: For tasks include substrate scope screening and self-driven reaction condition optimization, Experiment Designer was configured through few-shot learning of several examples or pre-prompting to transform reaction procedures and parameters described in natural language into standardized execution protocols. (3) Hardware Executor: Specific hardware running code examples or Opentrons Python API manual were provided in the prompt, such that Hardware Executor could generate running codes for the automation platforms according to the standardized execution protocols. (4) Spectrum Analyzer, (5) Separation Instructor, and (6) Result Interpreter: We provided detailed descriptions and instructions as pre-prompts to teach them to perform these tasks. For more details, refer to the Supplementary Information Section 1.

Web application

The web application functioned as the interface through which users could interact with agents and experimental platform. The frontend graphical interface was developed using the Vue.js and Node.js frameworks, creating a user-friendly and interactive environment. For the backend, the Python FastAPI framework was employed to manage the logics of multi-agent system and experimental platform, including interfacing with the LLM-based agents through the GPT-4 APIs hosted on Microsoft Azure and handling the operations of the experimental platforms. In addition, the web application was segmented into individual modules corresponding to each task of the chemical synthesis reaction development workflow.

Liquid handler platform for substrate scope screening and reaction kinetic study

The experimentation for substrate scope screening and reaction kinetic study steps was conducted using the Opentrons OT-2 liquid handling workstation. In the OT-2, modules including the pipette module (P300 GEN2, 20-300 μL) for liquid transferring, heater-shaker module (200-3000 RPM, 37-95 °C) for enhancing mixing of reaction mixture, and storage module for storing reaction stock solutions. Operation codes, generated by the Hardware Executor, were uploaded to the OT-2 via its desktop application or a Jupyter notebook to initiate automated reaction execution.

Automated reaction platform for self-driven reaction condition optimization

The self-driven reaction condition optimization platform consists of three modules, including an automated synthesis equipment (Unchained Labs, Big Kahuna), a HPLC (Thermo Fisher Scientific Vanquish), and a six-axis robotic arm (AUBO-i5) with a linear track. Big Kahuna automated experimental procedures, incorporating several components, including an extended tip liquid dispenser (20-3000 μL) for liquid transferring, the vortexing stations (60-3750 RPM) for mixing the reaction mixture, and a vial/plate gripper for transferring reaction vials and plates. HPLC analyzed reaction mixtures using a C18 reverse-phase column, with water and MeCN as the mobile phases. The robotic arm was responsible for transferring samples between Big Kahuna and HPLC. This hardware platform was controlled via a customized LabVIEW software, and experimental procedures and parameters were defined by the JSON method files.

Reaction optimization algorithm

The Bayesian optimization algorithm and the PI stopping criterion was developed and discussed in previous work[62]. In brief, it is composed of Gaussian process (GP) model and acquisition functions (AF). GP was a mixed kernel (Eq. S3), combining the Matérn52 kernel (Eq. S1) with the categorical

kernel (Eq. S2), to handle the reaction's design space, which includes both continuous and categorical variables. The new experiment candidates are proposed by maximizing the multi-points expected improvement (qEI) acquisition functions:

$$\{\boldsymbol{x}_{new}^{(k)}\}_{k=1}^{q} = \text{argmax}q\text{EI}\left(\{\boldsymbol{x}^{(k)}\}_{k=1}^{q}\right) = \text{argmax}\mathbb{E}_n\left(\text{ReLu}\left(\max_{i=1,..,q} f(\boldsymbol{x}_i) - f_n(\boldsymbol{x}^+)\right)\right) \quad (3)$$

where $\{\boldsymbol{x}_{new}^{(k)}\}_{k=1}^{q}$ is the q newly proposed reaction conditions, $\boldsymbol{x}^+$ is the current optimal condition, and $\mathbb{E}_n$ indicates that the expectation is taken under the posterior distribution at time n.

The probability of improvement (PI) value is a measure of the possibility that the newly proposed reaction candidate could have an improvement over the current optimal value (Eq. (4).

$$\text{PI}(\boldsymbol{x}) = \mathbb{P}(f(\boldsymbol{x}) \geq f(\boldsymbol{x}^+) + \xi) = \Phi\left(\frac{\mu(\boldsymbol{x}) - f(\boldsymbol{x}^+) - \xi}{\sigma(\boldsymbol{x})}\right) \quad (4)$$

where $\mu(\cdot)$ is GP's mean, $\sigma(\cdot)$ is GP's standard deviation, $\Phi(\cdot)$ is the normal cumulative distribution function, and ξ is the trade-off parameter of exploitation and exploration.

Data and code availability

All the relevant data and code are publicly available in the repository (https://github.com/Ruan-Yixiang/LLM-RDF).

References

1. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).
2. Feng, F., Lai, L. & Pei, J. Computational Chemical Synthesis Analysis and Pathway Design. *Front. Chem.* **6**, 199 (2018).
3. Molga, K., Szymkuć, S. & Grzybowski, B. A. Chemist Ex Machina: Advanced Synthesis Planning by Computers. *Acc. Chem. Res.* **54**, 1094–1106 (2021).
4. Andersson, S. *et al.* Making medicinal chemistry more effective—application of Lean Sigma to improve processes, speed and quality. *Drug Discov. Today* **14**, 598–604 (2009).
5. Struble, T. J. *et al.* Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.* **63**, 8667–8682 (2020).

6. Griffin, D. J., Coley, C. W., Frank, S. A., Hawkins, J. M. & Jensen, K. F. Opportunities for Machine Learning and Artificial Intelligence to Advance Synthetic Drug Substance Process Development. *Org. Process Res. Dev.* **27**, 1868–1879 (2023).
7. Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688-702.e13 (2020).
8. Wong, F. *et al.* Discovery of a structural class of antibiotics with explainable deep learning. *Nature* **626**, 177–185 (2024).
9. Zahrt, A. F. *et al.* Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **363**, eaau5631 (2019).
10. Coley, C. W. *et al.* A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).
11. Mikulak-Klucznik, B. *et al.* Computational planning of the synthesis of complex natural products. *Nature* **588**, 83–88 (2020).
12. Bédard, A.-C. *et al.* Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).
13. Shields, B. J. *et al.* Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
14. Wang, J. Y. *et al.* Identifying general reaction conditions by bandit optimization. *Nature* **626**, 1025–1033 (2024).
15. Slattery, A. *et al.* Automated self-optimization, intensification, and scale-up of photocatalysis in flow. *Science* **383**, eadj1817 (2024).
16. Steiner, S. *et al.* Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
17. Xi, Z. *et al.* The Rise and Potential of Large Language Model Based Agents: A Survey. Preprint at http://arxiv.org/abs/2309.07864 (2023).

18. Wang, L. *et al.* A Survey on Large Language Model based Autonomous Agents. Preprint at http://arxiv.org/abs/2308.11432 (2024).
19. Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T. & Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **145**, 18048–18062 (2023).
20. Zhang, W. *et al. Fine-Tuning Large Language Models for Chemical Text Mining*. https://chemrxiv.org/engage/chemrxiv/article-details/65baa07b9138d2316124f224 (2024) doi:10.26434/chemrxiv-2023-k7ct5-v2.
21. Zheng, Z. *et al.* Image and data mining in reticular chemistry powered by GPT-4V. *Digit. Discov.* 10.1039.D3DD00239J (2024) doi:10.1039/D3DD00239J.
22. Chen, K. *et al.* Chemist-X: Large Language Model-empowered Agent for Reaction Condition Recommendation in Chemical Synthesis. Preprint at http://arxiv.org/abs/2311.10776 (2024).
23. Zheng, Z. *et al.* A GPT-4 Reticular Chemist for Guiding MOF Discovery**. *Angew. Chem.* **135**, e202311983 (2023).
24. Zheng, Z. *et al.* Shaping the Water-Harvesting Behavior of Metal–Organic Frameworks Aided by Fine-Tuned GPT Models. *J. Am. Chem. Soc.* **145**, 28284–28295 (2023).
25. Li, J. *et al.* Empowering Molecule Discovery for Molecule-Caption Translation with Large Language Models: A ChatGPT Perspective. Preprint at http://arxiv.org/abs/2306.06615 (2023).
26. Yang, C. *et al.* Large Language Models as Optimizers. Preprint at http://arxiv.org/abs/2309.03409 (2023).
27. Zheng, Z. *et al.* ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs. *ACS Cent. Sci.* **9**, 2161–2170 (2023).
28. Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language models. *Nature* **624**, 570–578 (2023).
29. Yoshikawa, N. *et al.* Large language models for chemistry robotics. *Auton. Robots* **47**, 1057–1086 (2023).

30. OpenAI. GPT-4 Technical Report. Preprint at https://doi.org/10.48550/arXiv.2303.08774 (2023).
31. ChatGPT - Consensus. *ChatGPT* https://chat.openai.com/g/g-bo0FiWLY7-consensus.
32. Brown, T. B. *et al.* Language Models are Few-Shot Learners. Preprint at https://doi.org/10.48550/arXiv.2005.14165 (2020).
33. Gampe, C. & Verma, V. A. Curse or Cure? A Perspective on the Developability of Aldehydes as Active Pharmaceutical Ingredients. *J. Med. Chem.* **63**, 14357–14381 (2020).
34. Jain, S., Hiran, B. L. & Bhatt, C. V. Oxidation of Some Aliphatic Alcohols by Pyridinium Chlorochromate -Kinetics and Mechanism. *E-J. Chem.* **6**, 237–246 (2009).
35. Varala, R., Seema, V., Alam, M. M., Amanullah, M. & Dubasi, N. Dess-Martin Periodinane (DMP) in Organic Synthesis-A Septennial Update (2015-tillDate). *Curr. Org. Chem.* **27**, 1504–1530 (2023).
36. M. Heravi, M., Fazeli, A. & Faghihi, Z. Recent Advances in Application of Pyridinium Chlorochromate (PCC) in Organic Synthesis. *Curr. Org. Synth.* **13**, 220–254 (2015).
37. Zhu, L.-Y. *et al.* An efficient and selective oxidation of benzylic and aromatic allylic alcohols with manganese dioxide supported on kieselguhr under solvent-free conditions. *Res. Chem. Intermed.* **39**, 4287–4292 (2013).
38. Wu, W. & Jiang, H. Palladium-Catalyzed Oxidation of Unsaturated Hydrocarbons Using Molecular Oxygen. *Acc. Chem. Res.* **45**, 1736–1748 (2012).
39. Hoover, J. M. & Stahl, S. S. Highly Practical Copper(I)/TEMPO Catalyst System for Chemoselective Aerobic Oxidation of Primary Alcohols. *J. Am. Chem. Soc.* **133**, 16901–16910 (2011).
40. Huang, Z., Li, F., Chen, B. & Yuan, G. Sustainable catalytic oxidation of alcohols over the interface between air and water. *Green Chem.* **17**, 2325–2329 (2015).
41. Nishimura, T., Onoue, T., Ohe, K. & Uemura, S. Palladium(II)-Catalyzed Oxidation of Alcohols to Aldehydes and Ketones by Molecular Oxygen. *J. Org. Chem.* **64**, 6750–6755 (1999).

42. Nikitas, N. F., Tzaras, D. I., Triandafillidi, I. & Kokotos, C. G. Photochemical oxidation of benzylic primary and secondary alcohols utilizing air as the oxidant. *Green Chem.* **22**, 471–477 (2020).
43. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **360**, 186–190 (2018).
44. Xu, L.-C. *et al.* Enantioselectivity prediction of pallada-electrocatalysed C–H activation using transition state knowledge in machine learning. *Nat. Synth.* (2023) doi:10.1038/s44160-022-00233-y.
45. Tu, Z., Stuyver, T. & Coley, C. W. Predictive chemistry: machine learning for reaction deployment, reaction development, and reaction discovery. *Chem. Sci.* **14**, 226–244 (2023).
46. Buitrago Santanilla, A. *et al.* Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015).
47. Perera, D. *et al.* A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434 (2018).
48. Yin, G. Understanding the Oxidative Relationships of the Metal Oxo, Hydroxo, and Hydroperoxide Intermediates with Manganese(IV) Complexes Having Bridged Cyclams: Correlation of the Physicochemical Properties with Reactivity. *Acc. Chem. Res.* **46**, 483–492 (2013).
49. Annunziatini, C., Gerini, M. F., Lanzalunga, O. & Lucarini, M. Aerobic Oxidation of Benzyl Alcohols Catalyzed by Aryl Substituted *N*- Hydroxyphthalimides. Possible Involvement of a Charge-Transfer Complex. *J. Org. Chem.* **69**, 3431–3438 (2004).
50. Christensen, M. *et al.* Development of an automated kinetic profiling system with online HPLC for reaction optimization. *React. Chem. Eng.* **4**, 1555–1558 (2019).
51. Deem, M. C. *et al.* Best Practices for the Collection of Robust Time Course Reaction Profiles for Kinetic Studies. *ACS Catal.* **13**, 1418–1430 (2023).

52. Eyke, N. S. *et al.* Parallel multi-droplet platform for reaction kinetics and optimization. *Chem. Sci.* **14**, 8798–8809 (2023).
53. Hoover, J. M., Ryland, B. L. & Stahl, S. S. Mechanism of Copper(I)/TEMPO-Catalyzed Aerobic Alcohol Oxidation. *J. Am. Chem. Soc.* **135**, 2357–2367 (2013).
54. McMullen, J. P. & Jensen, K. F. Rapid Determination of Reaction Kinetics with an Automated Microfluidic System. *Org. Process Res. Dev.* **15**, 398–407 (2011).
55. Burés, J. & Larrosa, I. Organic reaction mechanism classification using machine learning. *Nature* **613**, 689–695 (2023).
56. Carlson, R. *Design and Optimization in Organic Synthesis*. (Elsevier, Amsterdam New York, 1992).
57. Guo, J., Ranković, B. & Schwaller, P. Bayesian Optimization for Chemical Reactions. *CHIMIA* **77**, 31–31 (2023).
58. McMullen, J. P., Stone, M. T., Buchwald, S. L. & Jensen, K. F. An Integrated Microreactor System for Self-Optimization of a Heck Reaction: From Micro- to Mesoscale Flow Systems. *Angew. Chem. Int. Ed.* **49**, 7076–7080 (2010).
59. Parrott, A. J., Bourne, R. A., Akien, G. R., Irvine, D. J. & Poliakoff, M. Self-Optimizing Continuous Reactions in Supercritical Carbon Dioxide. *Angew. Chem. Int. Ed.* **50**, 3788–3792 (2011).
60. Baumgartner, L. M., Coley, C. W., Reizman, B. J., Gao, K. W. & Jensen, K. F. Optimum catalyst selection over continuous and discrete process variables with a single droplet microfluidic reaction platform. *React. Chem. Eng.* **3**, 301–311 (2018).
61. Lorenz, R. *et al.* Stopping criteria for boosting automatic experimental design using real-time fMRI with Bayesian optimization. *ArXiv151107827 Q-Bio Stat* (2016).
62. Ruan, Y., Lin, S. & Mo, Y. AROPS: A Framework of Automated Reaction Optimization with Parallelized Scheduling. *J. Chem. Inf. Model.* **63**, 770–781 (2023).

63. Lovato, K., Fier, P. S. & Maloney, K. M. The application of modern reactions in large-scale synthesis. *Nat. Rev. Chem.* **5**, 546–563 (2021).

64. Xu, H. *et al.* High-throughput discovery of chemical structure-polarity relationships combining automation and machine-learning techniques. *Chem* **8**, 3202–3214 (2022).

Author contributions

Y.R. and Y.M. conceived the project. Y.R. developed and implemented the LLM-based agents. Y.R. and C.L. developed the web application. Y.R. and N.X. performed the chemical experiments. Y.R. and Y.M. wrote the manuscript. All authors were involved in the discussions.

Competing interests

The authors declare no competing interests.