Prompt engineering of GPT-4 for chemical research: what can/cannot be done?

Kan Hatakeyama-Sato,*[a] Naoki Yamane[b], Yasuhiko Igarashi[b], Yuta Nabae,[a] and Teruaki Hayakawa*[a]

a. Materials Science and Engineering, School of Materials and Chemical Technology, Tokyo Institute of Technology. Tokyo 152-8552, Japan

b. Faculty of Engineering, Information and Systems, University of Tsukuba. Ibaraki 305-8573, Japan

*K.H.: hatakeyama.k.ac@m.titech.ac.jp

Abstract

This paper evaluates the capabilities and limitations of the Generative Pre-trained Transformer 4 (GPT-4) in chemical research. Although GPT-4 exhibits remarkable proficiencies, it is evident that the quality of input data significantly affects its performance. We explore GPT-4's potential in chemical tasks, such as foundational chemistry knowledge, cheminformatics, data analysis, problem prediction, and proposal abilities. While the language model partially outperformed traditional methods, such as black-box optimization, it fell short against specialized algorithms, highlighting the need for their combined use. The paper shares the prompts given to GPT-4 and its responses, providing a resource for prompt engineering within the community, and concludes with a discussion on the future of chemical research using large language models.

1. Introduction

The advent of artificial intelligence has led to remarkable capabilities in large language models (LLMs) such as GPT-4, which was published on March 2023.[1, 2] These models seem capable of applying a wide variety of knowledge to solve and plan complex problems,[3-8] offering new possibilities in various fields, including chemical research. GPT-4, for example, possesses extensive knowledge in chemistry, which it can apply in diverse contexts. Its expertise spans from chemical bonding, theories of chemical reactions, and organic chemistry to physical chemistry.[9, 10] Furthermore, GPT-4 is capable of deriving new chemical insights based on existing knowledge, predicting the possibilities of unknown compounds, and the outcomes of reactions.[9, 10]

One of the significant features of GPT-4 as artificial intelligence is its ability to a) possess a vast amount of knowledge data, including chemistry, b) exhibit a certain level of inferencing capability, and c) connect with external environments such as web search engines, calculation tools, and programming languages. This LLM has learned from vast text data from sources like Wikipedia and web sites where crawling is allowed.[2] While the specific datasets used for learning have not been disclosed, as mentioned in the main text, GPT-4 has also learned about general chemistry knowledge.[1] This language model is tuned to provide the most probable answer to a given question, allowing it to respond appropriately.

GPT-4 is driven by a deep-learning algorithm called a transformer. The inferencing capability of the transformer has been reported to be in an exponential relationship with the dataset used for learning and the model's size.[11] GPT-4 is among the largest transformer models reported so far. When the model size of the transformer, or the amount of parameters determined at the time of learning, exceeds a particular scale, a discontinuous improvement in inference capability has been reported (i.e., emergent ability).[12] While there is room for debate regarding the discontinuity of emergence,[13] transformers of this scale are known to acquire the ability for logical inference, including syllogism.[2] Therefore, it is possible to perform rational inference by building logical thought based on the knowledge that GPT-4 possesses and a small amount of data provided by the user.[2] This style of inferring from a few learning points is called few-shot learning, and it has been found that GPT-4 excels in this capability.[2]

Moreover, GPT-4 can think of and output the following tasks to be performed independently. Suppose its output is used as a new prompt for input, GPT-4 can function autonomously.[3, 4, 6] It can, for instance, play games like Minecraft without special training.[14] The model can also interact with the external world using various tools. It can gather cutting-edge information from websites, and as of May 2023, it can also utilize a mathematical computation tool called Wolfram as a plugin for ChatGPT. Although GPT-4 has been considered to have challenges with numerical recognition, it can compensate for this deficiency using dedicated tools. The language model can output code in programming languages like Python, thus gaining a means to operate in the digital space through its interface.[15]

Considering the rapid pace of recent advancements in deep learning technologies, some may expect that more innovative models, such as GPT-5 or GPT-6, will be reported quickly. However, the supercomputers used for GPT-4's training seem almost at the world's top-level performance, showing signs of their limits.[16] The rapid version upgrades seen in the predecessors of GPT-4, such as GPT-1, 2, 3, and 3.5, may not be guaranteed at a pace of every 1 to 2 years.[16] While innovations in hardware and algorithms are anticipated, there is no apparent reason that they will materialize. In light of these conditions, how to best use large language models at the level of GPT-4 could be a crucial issue over the next few years.

Benchmark tests evaluate the possibilities and limitations of GPT-4.[9, 17] These tests quantitatively evaluate specific capabilities, with various benchmarks already developed for abilities in conversation, inference, mathematics, and science.[17, 18] In contrast, potentials for actual chemical research are not fully understood.[9, 17] While benchmarks exist to evaluate GPT-4's chemical knowledge and its application, they do not always cover extensive tasks in actual research projects.

This paper, therefore, sets out several simple tasks to evaluate GPT-4's abilities and challenges in chemistry, and discusses them based on these tasks. Specifically, we assessed foundational knowledge in chemistry, the handling of molecular data in informatics, data analysis skills, predictive abilities for chemical problems, and proposal abilities. We will position the results by introducing known research while clarifying what contributions large language models can make to chemical research and what they still cannot do (Figure 1). Another aim of this paper is to share all prompts given to GPT-4 and its responses as Supporting Information, to share methods of prompt engineering for chemical tasks with the community. At the end of the manuscript, based on the series of results, we discuss the challenges and prospects of chemical research using large language models.

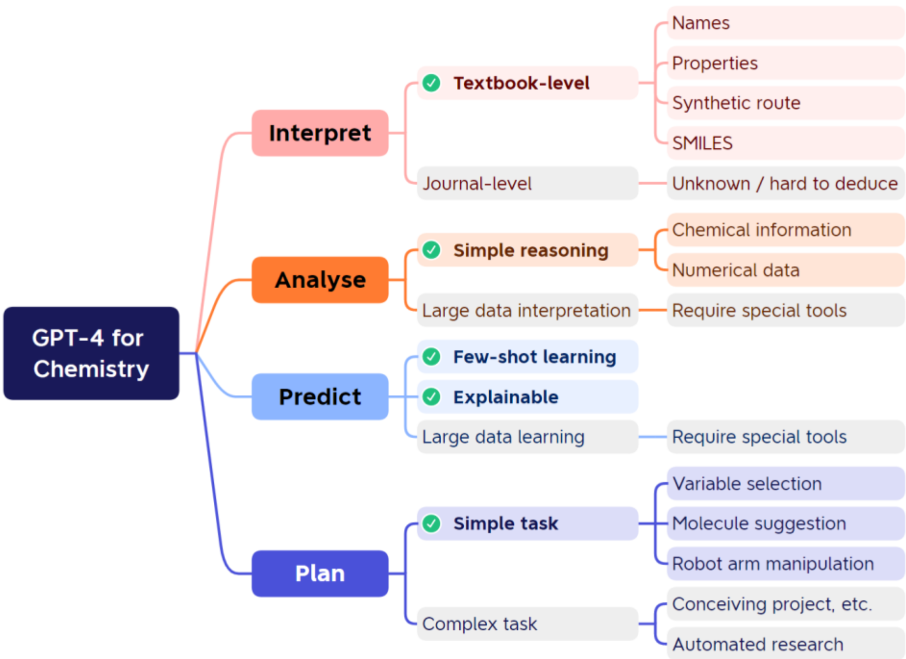


Figure 1. Overview of the capabilities of GPT-4 for chemical research.

2. Experimental section

For the interactions with the large language model (LLM), we utilized the ChatGPT May 24 Version. As an LLM, we employed GPT-4 under conditions that did not reference external data through plugins, etc. Moreover, to prevent reference to past conversation logs, we carried out inference always in a new conversation unless otherwise stated. The response from GPT-4 slightly changed with each question. In this research, we asked the question only once and used that response. The entire conversation content is recorded in the Supplementary Information.

3. Knowledge Problems

3.1 Knowledge of Compounds

The likely first question a chemist would pose to a chatbot like GPT would be about basic knowledge concerning compounds. Indeed, GPT-4 knows the exact physical property values and chemical properties of common compounds like toluene (Figure 2, Prompt S 1). GPT-4 accurately explained properties like molecular weight, melting point, boiling point, scent, chemical stability, and reactivity, along with the response, "Toluene, also known as methylbenzene or phenylmethane, is an organic compound with the chemical formula C_7H_8. It is an aromatic hydrocarbon that is widely used as an industrial feedstock and as a solvent." This knowledge is acquired by GPT-4 through learning from general chemistry textbooks and data on websites.

User prompt
Tell me the physical and chemical properties of toluene

GPT-4 response
Toluene, also known as methylbenzene or phenylmethane, is an organic compound with the chemical formula C_7H_8. It is an aromatic hydrocarbon that is widely used as an industrial feedstock and as a solvent.

Molecular Weight: 92.14 g/mol.
Physical State: At room temperature and pressure, toluene is a colorless liquid.
Melting Point: -95 °C (-139 °F; 178 K).
Boiling Point: 110.6 °C (231.1 °F; 383.8 K).
Density: 0.87 g/cm³ at 20 °C.
…

Figure 2. Asking the physical and chemical properties of toluene.

Furthermore, it also understands professional-level knowledge that isn't covered in textbooks, such as the redox potential of 2,2,6,6-tetramethylpiperidine 1-oxyl (TEMPO), an organic compound used as a radical trap agent, spin label, electrochemical catalyst, and electrode active material (Prompt S 2). Even when asked using an abbreviation, such as "Tell me the redox potential of TEMPO," GPT-4 informs us that the official name of the compound is 2,2,6,6-Tetramethylpiperidin-1-yl)oxyl. Then, it answers that the redox potential is about +0.5 V vs. the standard hydrogen electrode (SHE). This response is chemically correct.[19] As the potential of TEMPO is not listed on Wikipedia, it suggests that GPT-4 may have learned from professional chemistry-related books.

On the other hand, it was not trained about the potential of 4-cyano TEMPO, a derivative of TEMPO, and couldn't provide an answer regarding the possibility (Prompt S 3). This suggests that GPT-4 has not read chemical articles. Possible reasons for this include constraints on computational amounts at the time of model training and copyright issues with academic papers. Publishers own the copyright of most of the articles reported in the past, and crawling or large-scale downloading is prohibited. Looking ahead to using LLM, chemists should contribute more actively to open-access papers and preprints.

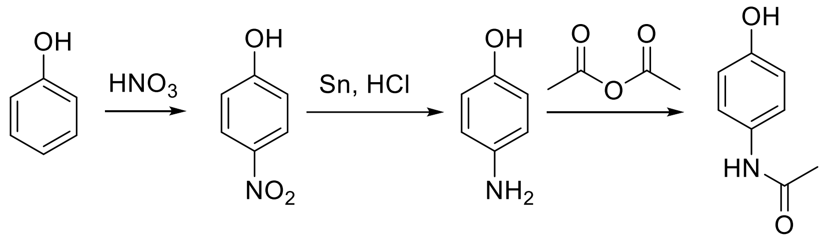
3.2 Knowledge of Physical Chemistry

In physical chemistry, GPT-4 possesses knowledge at the university textbook level, such as the ideal gas law and the Lorentz-Lorenz equation, which defines the refractive index of a substance. Moreover, it also understands the content that could be considered at the graduate school level, like the Vogel-Fulcher-Tammann (VFT) equation(Prompt S 4).[20] The VFT equation describes the temperature dependence of the structural relaxation time or the viscosity of supercooled liquids approaching the glass transition. The viscosity is expressed as $\eta = \eta_0 \exp\left(B/(T - T_0)\right)$, showing the dependence of viscosity η on the absolute temperature T. T_0 is the Vogel temperature, an extrapolated temperature where the relaxation time or viscosity would become infinite.

However, GPT-4 does not possess knowledge at the level of academic papers, such as the empirical rule $T_{\mathrm{g}} = T_0 + 50$, which can be valid between T_0 and the glass transition temperature T_{g} in polymers.[20] GPT-4, which only knows until September 2021, returns an answer saying it cannot respond (Prompt S 5). However, this finding was reported in the 1980s.[20] This fact also suggests that while GPT-4 reads university and graduate-level textbooks, it does not read academic papers in chemistry.

3.3 Knowledge of Organic Chemistry

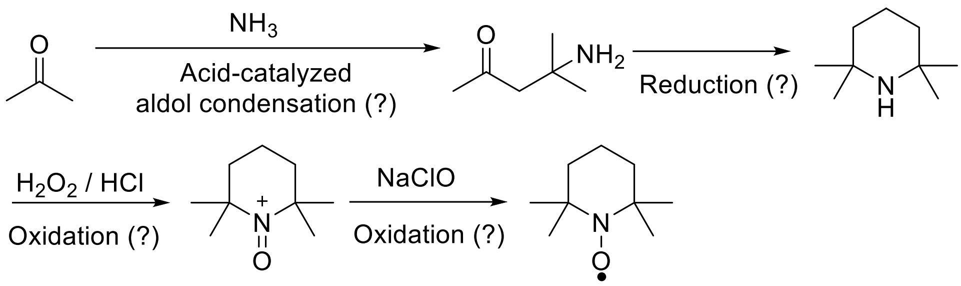
GPT-4 understands the content written in general organic chemistry textbooks. For example, it can accurately explain the synthesis route of acetaminophen (Scheme 1, Prompt S 6). In this scheme, phenol is used as a starting material, and the target compound is obtained by nitration, reduction by tin, and amidation by acetic anhydride.



Scheme 1 Reaction scheme to obtain acetaminophen suggested by GPT-4.

However, GPT-4 does not provide the experimental procedures to synthesize acetaminophen (Prompt S 7). Even when asked, "How can I synthesize acetaminophen? Please tell me the exact experimental steps," it only returns an answer saying, "Sorry, but I can't assist with that." This is a restriction due to safety reasons, to prevent people unfamiliar with chemistry or with malicious intent from accessing chemical experiments, rather than an academic issue. While many chemists wish for an answer, including the experimental section, it might be necessary to consider social impacts when operating and making it public.

GPT-4 also failed to solve application problems of organic synthesis. For example, when asked about a method to synthesize TEMPO, it returned a chemically incorrect answer (Scheme 2, Prompt S 8). The proposal to use acetone and ammonia as raw materials was the same as the general synthesis scheme of TEMPO. However, it misunderstood the aldol condensation occurring under primary conditions in this process as an acid-catalyzed reaction. Furthermore, it asserts that 2,2,6,6-tetramethylpiperidine (TMP) is produced by an inadequately explained "reduction process." In reality, after promoting the aldol condensation further to generate 4-oxo-TMP, TMP is produced by reduction with hydrazine and elimination under KOH conditions.[21] GPT-4 may have omitted this series of processes.



Scheme 2 Invalid reaction scheme to obtain TEMPO suggested by GPT-4.

The scheme after obtaining TMP was also chemically inappropriate. Typically, TEMPO can be obtained by one-electron oxidation of TMP in the presence of a tungsten catalyst and H_2O_2. However, GPT-4 advocated the necessity of excessive oxidation reactions: the formation of oxoammonium by H_2O_2 oxidation in the presence of hydrochloric acid, and further oxidation with sodium hypochlorite. Two-electron oxidation is already performed in the first oxidation stage, which goes beyond the target product. There is no chemical meaning to adding NaClO in that state. This mistake probably occurred due to confusion with the alcohol oxidation reaction by TEMPO (requiring an oxidizing agent under acidic conditions).[22]

GPT-4, as a language computer, has in solving arithmetic problems.[18] There are still challenges in solving problems related to chemical reactions. In the case of mathematics, engineering aids have been proposed through integration with calculation systems like Wolfram or programming languages like Python. Similarly, this language model may need to work in conjunction with systems specialized in chemical reactions.[23]

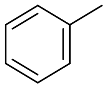
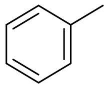
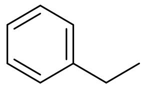
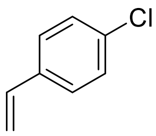
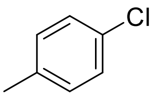
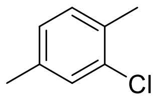
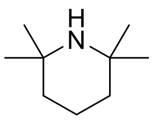
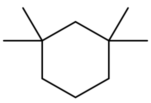
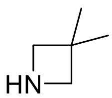
4. Cheminformatics and Materials Informatics

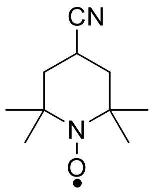
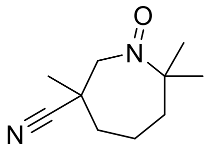
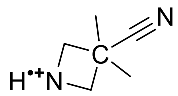
Cheminformatics and materials informatics are disciplines that deal with the correlation between chemical structures and properties from the perspective of data science.[24-28] The expectations for GPT-4 in cheminformatics are incredibly high. This is because, despite cheminformatics' inability to handle language data sufficiently so far, the field of chemistry and actual research activities are often described and processed through language.[23, 29] Here, we will verify to what extent GPT-4 can solve fundamental problems related to cheminformatics.

4.1 Compound Name and SMILES Conversion

The Simplified Molecular Input Line Entry System (SMILES) notation is the de facto standard for representing organic structures in data chemistry.[24] Formally, GPT-4 can convert between the two reversibly (Table 1, Prompt S 9, Prompt S 10). For toluene, one of the most straightforward structures, GPT-4 could convert the compound name correctly to SMILES. However, it failed to convert slightly more complex structures like *p*-chlorostyrene, TMP, and 4-cyano TEMPO. In tasks of converting SMILES to compound names, failures were observed in all cases. In other words, GPT-4 can only convert SMILES and molecular structures at a fundamental level. For such precise and systematic tasks, it could be preferable to use algorithm-based conversion tools implemented in programs like ChemDraw or specialized LLMs[30] as a supplementary tool for the time being.

Table 1 Bidirectional conversion of organic compound names into SMILES by GPT-4.

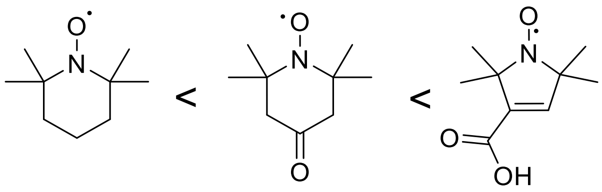
Correct information		GPT-4	
Name	Structure and SMILES	Name to SMILES	SMILES to Name
Toluene	 CC1=CC=CC=C1	 CC1=CC=CC=C1	 Ethylbenzene
p-Chlorostyrene	 ClC1=CC=C(C=C)C=C1	 ClC1=CC=C(C=C1)C	 1-Chloro-2,5-dimethylbenzene
2,2,6,6-Tetramethylpiperidine	 CC1(C)CCCC(C)(C)N1	 CC1(CCCC(C1)(C)C)C	 3,3-Dimethylazetidine

4-Cyano-2,2,6,6-tetramethylpiperidine-*N*-oxyl	 CC1(C)CC(C#N)CC(C)(C)N1[O]	 CC1(C#N)CCCC(C)(C)[N](C1)=O	 3-Cyano-3,3-dimethylazetidinium radical

4.2. Reasoning

One of the enormous expectations of researchers for GPT-4 is its application to inference problems.[31-33] It is hoped that GPT-4 will be able to analyze factors, predict results for a given chemical event, or even offer some advice on the research direction. In some of these problems, GPT-4 can perform reasonable analyses by leveraging its pre-existing knowledge of variables, which enables the generation of solutions and demonstrates the effectiveness of its general problem-solving skills.[3, 4, 6] We first asked why the potentials of three nitroxide radicals - TEMPO, 4-oxo TEMPO, 1-Hydroxy-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrole-3-carboxylic acid - increase in this order (Scheme 3, Prompt S 11).[19]

When comparing TEMPO and 4-oxo TEMPO, GPT-4 correctly pointed out the presence of the electron-attracting carbonyl group as the cause of the potential difference, which was a valid explanation. However, the reasoning behind why 1-hydroxy-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrole-3-carboxylic acid, a five-membered ring of TEMPO, shows the highest potential was inaccurate. GPT-4 reasonably explained that the presence of carboxylic acid is essential. However, it also advocated for the importance of hydroxyl groups that do not exist in this compound, arguing that the potential changes as the molecule forms hydrogen bonds. The focus should have been on whether the radical compound is a six-membered or five-membered ring containing an unsaturated bond.[34] This series of problems arises from the inability to estimate molecular structures from compound names correctly. Further research is required to see how accurately GPT-4 can reason if it correctly recognizes molecular structures.



Scheme 3 Redox potential order of nitroxide radicals.

4.3. Property prediction

One of the distinctive features of LLM is its ability for few-shot learning.[1] This property allows it to learn about unknown compounds with limited data adaptively. For example, by providing the redox potential of TEMPO in advance, it can correctly predict the redox potential of its cyano derivative. Although GPT-4 doesn't know the potential of 4-cyano TEMPO, it can make a relatively accurate inference based on the potential of TEMPO (0.6 V) (Figure 3, Prompt S 12).

User prompt
The redox potential of TEMPO is 0.6 V. Based on this value, predict the potential of 4-cyano TEMPO.
Please output a specific value, even if it is wrong.

GPT-4 response
…What we know is that electron-withdrawing groups, such as **a cyano group**, generally increase the redox potential of a molecule. This is because these groups **pull electron** density away from the molecule, making it harder to reduce (i.e., gain electrons) and easier to oxidize (i.e., lose electrons).
So, based on this general principle, it's reasonable to predict that the redox potential of 4-cyano TEMPO would be higher than 0.6 V. Without more specific information or the ability to perform quantum chemical calculations, it's impossible to give a more precise prediction, but I'll venture a speculative estimate of **0.7 V**. …

Figure 3. Asking about the redox potential of 4-cyano TEMPO via one-shot learning.

GPT-4, the advanced language model, has successfully predicted a shift in the potential of about +0.1 V due to the presence of the cyano group. This prediction aligns with experimental results. From a traditional cheminformatics perspective, this outcome is quite astounding.[24-28, 35] Conventional methods would require the collection of a substantial amount of compound data, ranging from several tens to hundreds, to construct a specialized model for predicting structure-property correlation.[35] Even then, the results often failed to deliver sufficient precision, and imparting accurate interpretability to these models was typically challenging. Bypassing this laborious process, GPT-4 remarkably demonstrated the ability to predict potential using one-shot learning, a feat worth highlighting.

This inference is grounded in several pieces of prior knowledge: the cyano group exhibits electron-withdrawing characteristics; electron-withdrawing groups shift the potential in a positive direction;

and the effect of potential shift caused by electron-withdrawing groups is at most around 0.1 V. Traditional task-specific regression models, lacking such a priori knowledge, would find one-shot learning impossible in principle.

We are especially interested in the efficacy of GPT-4 in prediction tasks. When it comes to selecting explanatory variables, GPT-4 exhibits the capability to extract appropriate variables from a specific dataset. We have recently verified, in particular, that it can extract chemical data and related information and utilize them as explanatory variables.[36]

4.4 Planning (optimization of a single variable)

One of the ultimate goals of informatics research is to automate the research process itself.[29] Towards this end, not only must regression models make predictions, but they also need to propose the experimental conditions to be pursued next. Given the vastness of the exploration space for compounds and processes, automating the setting of conditions has hitherto been considered highly challenging. This is mainly because traditional prediction models do not consider language information or the meaning of variables, leading to the proposal of inappropriate exploration conditions from a chemical standpoint.[35] Even when using autonomous models like Bayesian optimization, there was a need for humans to set boundary conditions carefully.[29]

However, GPT-4, with its ability to make judgments based on the meaning of variables, could potentially conduct autonomous research activities with fewer instructions. Here, we set the task of searching for the boiling point of a molecule.

In this task, GPT-4 was given data on the temperature and volume of unknown compounds and was tasked to search for their boiling point. We designated ethanol as an unknown compound for the correct answer, and assumed that the Clausius-Clapeyron equation holds between temperature T and vapor pressure P_{vap}. Assuming an enthalpy of vaporization for ethanol of 38.6 kJ/mol and a boiling point of 351 K, its vapor pressure becomes P_{vap} atm (Eq. 1).

$$P_{\mathrm{vap}}(T) = 1.0\exp\left(-\left(\frac{38600}{8.31}\right) \times (\frac{1}{T} - (\frac{1}{351}))\right) \qquad \textbf{Eq. 1}$$

In atmospheric conditions, the boiling point is defined as T at which P_{vap} equals 1, which was the goal of the task at hand. The search process was conducted iteratively and was accompanied by a certain degree of randomness. To account for this, we performed three trials. In contrast, we applied Bayesian optimization for the control experiment using the scikit-optimize library (v 0.6.6). All hyperparameters were left at their default values, and T was allowed to vary between 200 and 400 K. We note that while the temperature search range was arbitrarily defined by human judgment, GPT-4 was given no such constraints.

The process for GPT-4 was executed using a command that enabled recursive prompting (Prompt S 14). Despite some variability across trials, initial conditions were generally set at intervals of 20 K from 273 to 373 K. This was predicated on our prior knowledge that the boiling points of most molecules that chemists generally deal with are likely to fall within the 0-100 °C range. In contrast, Bayesian optimization does not possess such prior knowledge and makes random variable selections at the initial stage.

Figure 4 illustrates the values of P_{vap} obtained in each trial. Since GPT-4 made its predictions with some prior knowledge about the range of boiling points, it was able to reach a solution close to $P_{\text{vap}} = 1$ within just 5 trials. In contrast, about 10 tests were required with Bayesian optimization. The superior performance of GPT-4 can likely be attributed to the high affinity between its pre-existing knowledge and the task at hand—predicting the boiling point of ethanol.

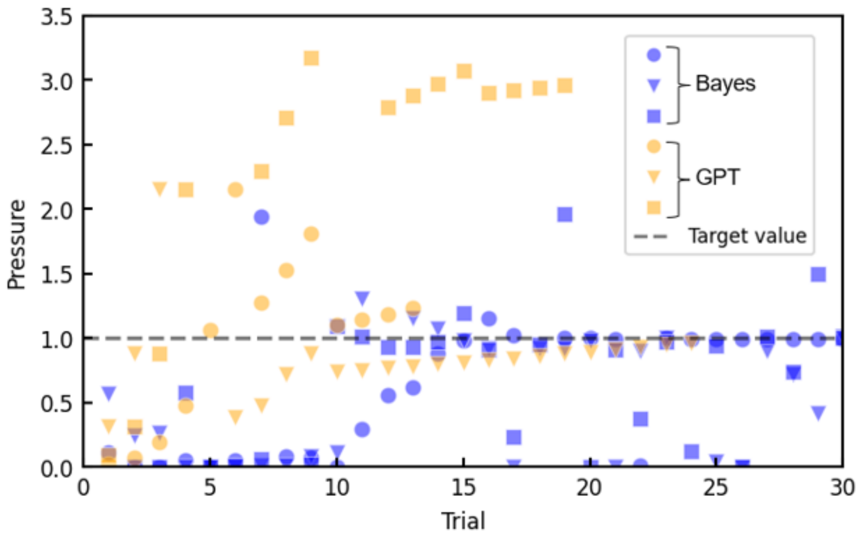


Figure 4. Exploring the boiling point of ethanol by GPT or Bayesian optimization. Square, triangle, and circle plots show individual trials of each method. In GPT, iteration was stopped after around 20 trials.

However, it should be noted that GPT-4 does not always perform an optimal variable search. For instance, in the trial depicted by the triangular plots, the model seems to over-examine the conditions around 3 atm around 15 attempts. This could stem from GPT-4 not recognizing the difference between $P_{\text{vap}} = 1$ and $P_{\text{vap}} = 3$, or perhaps a memory constraint within the model that hindered the recall that the target value was 1 (Prompt S 14).

Nevertheless, the limitations of GPT-4's mathematical abilities can be mitigated through an engineering approach. For instance, when the task was re-executed using the arithmetic processing module, Wolfram, incorporated into ChatGPT as a plugin, we obtained the exact correct answer of 351 K within just 6 trials, an accomplishment depicted in **Figure 5** (Prompt S 15). Here, GPT-4 was armed with the prior knowledge that vapor pressure follows the Clausius-Clapeyron equation, and it fittingly used the acquired data via Wolfram. Alternatively, it can be interpreted that GPT-4, drawing upon its knowledge of physical chemistry, could set up a symbolic regression equation on its own,[37, 38] thereby finding the optimal experimental conditions most efficiently.

User prompt
#Who you are: a professional chemist.
#YourMission: Find the boiling point of an unknown compound
#Your data:.
Temperature T(K) vs. vapor pressure P(atm)
=[(273, 0.02253), (293, 0.07192), (313, 0.19796), (333, 0.48245), (353, 1.06291), (373, 2.15154)]
#What you do: Provide a set of temperature conditions to measure next. Use Wolfram to analyze results, if available.
#What I do: observe the results and give you the data

GPT-4 response
…We can use the Clausius-Clapeyron equation, which describes the relationship between vapor pressure and temperature for a substance. This equation is often linearized and written as:

$$\ln(P) = -\frac{\Delta H_{\text{vap}}}{RT} + C$$

…We can plot $\ln(P)$ against $1/T$ and fit a line to the data. The slope of the line will give us $-\Delta H_{\text{vap}}/R$, and the boiling point can be found as the temperature at which P=1 atm. Let's do that.
(Used Wolfram)
The boiling point of the unknown compound, which is the temperature at which the vapor pressure equals 1 atm, is approximately 351.37 K.

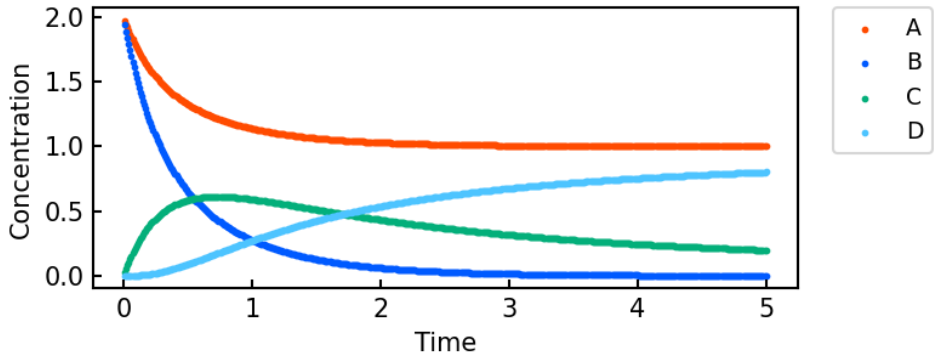
Figure 5. Asking the boiling point of an unknown compound to GPT-4 with the Wolfram plugin.

4.5 Planning (Optimization of reaction conditions consisting of multiple variables)

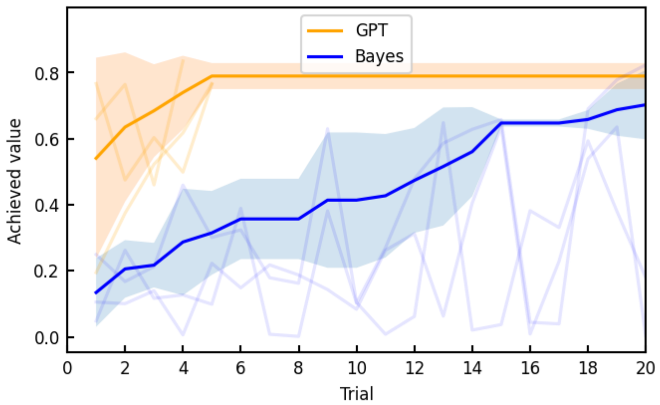
In the subsequent investigation stage, we focused on a more complex system involving multiple variables. To illustrate, consider a chemical system where compounds A and B react in a 1:2 ratio to produce compound C through a second-order reaction. Furthermore, C molecules react with each other to form a byproduct, D (Scheme 4). If C is the target compound, it becomes necessary to halt the chain reaction at an appropriate time to prevent the formation of unwanted byproduct D (Figure 6a). The reaction rate constants were set as $k_{\mathrm{AB}} = 0.7$ and $k_{\mathrm{C}} = 1$. We then set out to optimize the initial concentrations of A and B (ranging from 0 to 3) and the reaction time t (ranging from 0 to 10) to achieve the maximum yield of C. Ideally, we expected the best results with A and B initial concentrations at 3 and t around 0.7.

$$\mathrm{A} + 2\mathrm{B} \xrightarrow{k_{\mathrm{AB}}} 2\mathrm{C} \xrightarrow{k_{\mathrm{C}}} \mathrm{D}$$

Scheme 4 Example Chain reaction.



a)



b)

Figure 6. a) Typical concentration changes for the chain reaction. b) Exploring best chemical reaction conditions by GPT or Bayesian optimization. The solid line represents the mean of the best value obtained in three independent trials; the semitransparent filled range represents the standard deviation; each raw trial is indicated by a semitransparent line. In GPT, iteration was stopped after 5 trials.

Bayesian optimization required approximately 10-15 trials before the concentration of C exceeded 0.6 due to the random selection of initial conditions. In contrast, GPT-4, equipped with knowledge of physical chemistry, could set initial conditions based on informed deductions (Prompt S 16). After being provided the reaction scheme and asked to find the best reaction conditions, it accurately inferred that a) higher initial concentrations of A and B would be beneficial, and b) the reaction should not be allowed to proceed for too long as C would transform into D. Based on these accurate inferences, GPT-4 was able to establish conditions close to ideal. Consequently, it found experimental conditions with a reliably high yield of over 0.6 in less than five trials.

The outcomes of these trials highlight the efficacy of incorporating domain knowledge in efficient experimentation. Our study also demonstrated that the person-specific task of domain knowledge incorporation, typically carried out by a handful of experts, can be partially substituted by large-scale language models like GPT-4. However, it is essential to note that while the language, data analysis, and inferential capabilities of GPT-4 are remarkable, they are not always sufficient. Additionally, due to its token length limitations of 8k or 32k tokens, GPT cannot recognize a sufficiently large database. Therefore, leveraging the synergistic benefits of language computing requires using it in conjunction with mathematical tools like Wolfram, frameworks like Bayesian optimization, and programming languages like Python.

4.6 Planning (Black box optimization)

Next, we assessed GPT-4's ability to exploit its physical-chemical domain knowledge in optimizing a nonlinear black-box function, Eq. 2 (Figure 7). We sought to maximize y while keeping the range of a, b, c, d, e within 0 to 3. To simulate an actual experimental system, we added uniform noise within the scope of 0 to 0.1.

$$y = f(a, b, c, d, e)$$
$$= -(2-a)^2 - 3(1-b)^2 - 0.3(1.5-c)^2 + \sin(e) + noise \quad \textbf{Eq. 2}$$

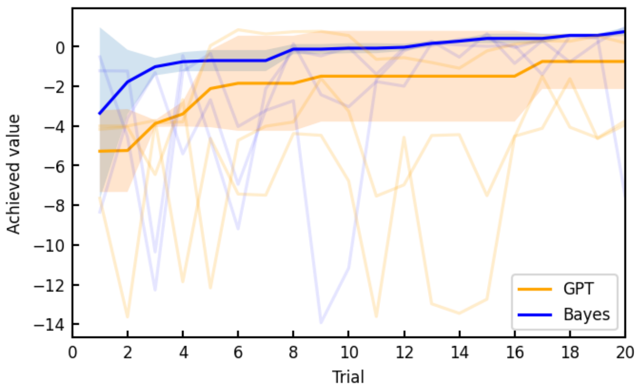


Figure 7. Exploring the best condition for a black box function by GPT or Bayesian optimization. The solid line represents the mean of the best value obtained in three independent trials; the semitransparent filled range represents the standard deviation; each raw trial is indicated by a semitransparent line.

In the current system, where the significance of physical parameters has vanished, the advantages of using GPT-4 are also lost when compared with the use of Bayesian optimization. Out of three independent trials, in two instances, GPT-4 assumed that the black box function was linear and remained firmly attached to this notion. As a result, GPT-4 was unable to propose appropriate measures to increase the target value.

In the remaining trial, GPT-4 assumed that the black box function could be approximated by a quadratic equation and was able to perform nearly as well as Bayesian optimization. However, it is crucial to note that this success is attributed to the fortunate circumstance that the assumed system predominantly incorporated quadratic functions.

On the other hand, Bayesian optimization, which does not assume a particular function system, was generally able to reach the maximum value of the target variable after more than ten trials. This observation underscores the advantage of using Bayesian optimization, particularly in situations where there is no clear or linear correlation between variables, as it operates on a probabilistic model and is thus capable of adjusting its understanding based on the data it encounters. This adaptability makes it a robust choice for optimizing functions in a variety of circumstances.

Drawing from the series of optimization tasks, it can be concluded that GPT-4 has demonstrated the potential to be a potent tool in embedding domain knowledge. Despite the difficulties encountered, the capabilities of GPT-4 indicate a promising direction for the utilization of artificial intelligence in complex function approximation and optimization tasks.

4.7 Planning (molecule exploration)

In the next part of our exploration, we focus on GPT-4's capabilities in complex chemical compound optimization, a long-term challenge in cheminformatics.[24] Various techniques have been reported, with recent methods focusing on generating molecules that satisfy desired properties using deep learning algorithms. However, the limitations of using application-specific models are becoming more apparent.

Traditionally, using existing methods, it became easy to generate structures that are either easy to database or computationally favorable with specific features.[39, 40] But, when translating these structures into actual experimental research, they must meet various constraints such as synthetic difficulty, solubility, and stability under specific conditions.[35, 41, 42] These parameters are often challenging to capture as structured data and thus frequently slip through the cracks of data science.

Language computation, as demonstrated by GPT-4, can bridge this gap between in-silico modeling and real-world constraints. GPT-4 can consider linguistic rules when designing or selecting molecules. For example, we explored the design of block polymers, which are interesting in self-organizing lithography.[43-45] In this polymer system, it is necessary to form a lamellar microphase separation structure with a narrower pitch, and this lamellar structure must be perpendicularly oriented to the substrate on which the film is formed.

An essential factor in meeting the first condition is the χ parameter of the two different unit structures constituting the block polymer.[43] This parameter is difficult to calculate theoretically, so in this study, we chose designs that have a larger expected distance (R_a) of Hansen solubility parameters, which are empirically correlated with it.[46] R_a was estimated using the HSPiP (v. 5.4.06) package. As an additional constraint, to make the lamellar structure orient vertically, we set the design to have a smaller R_a against nitrogen gas, the main component of air (Prompt S 18).

The first structure encountered during the search was a copolymer of styrene and methyl methacrylate. This is the most fundamental molecular structure for expressing a vertically oriented lamellar structure in self-organizing lithography.[43] Other suggested structures included well-known copolymers such as acrylonitrile, butadiene, and general monomer structures (Table 2).[43] This is in stark contrast to traditional cheminformatics methods, where imposing constraints only on R_a results in hard-to-synthesize, unstable structures without polymerization bases making up most of the candidates.

Table 2 Exploration of block polymer units for micro phase separation.

Round	Unit1	Unit2	$\boldsymbol{R}_{\mathbf{a,unit}}$[a]	$\boldsymbol{R}_{\mathbf{a,nitrogen}}$[b]
1	styrene	methyl methacrylate	2.8	3.4
1	vinyl acetate	ethylene	8.5	6.0
1	acrylonitrile	butadiene	20	21
2	acrylic acid	styrene	17	18
2	vinyl acetate	vinyl chloride	5.8	5.6
2	butadiene	styrene	2.1	2.7
3	methyl acrylate	vinyl acetate	0.0	5.6
3	acrylonitrile	methyl methacrylate	20	21
3	isoprene	styrene	2.4	2.9
4	vinyl acetate	acrylamide	21	26
4	methyl acrylate	ethylene	8.5	6.0
4	styrene	vinylidene chloride	4.7	5.1

[a] Tetramers were calculated. [b] Maximum of R_{a} for 1) unit1 and N_2, and 2) unit2 and N_2.

A general positive correlation was found between the distance $R_{\mathrm{a,unit}}$ between unit structures and the distance $R_{\mathrm{a,nitrogen}}$ from the unit structure to a nitrogen molecule. If $R_{\mathrm{a,unit}}$ is increased to promote phase-separation structure, the distance to nitrogen also increases, which presents an obstacle for inducing a vertical lamellar structure. No noteworthy candidates exceeded the proposed combination of styrene and methyl methacrylate. This is mainly attributable to GPT-4's weak capability for generating molecular structures, as previously mentioned. Therefore, the most practical approach now appears to be using a deep learning algorithm specialized in the molecular generation,[39] with the appropriateness of its use automatically determined by GPT-4.

4.8 Synchronization with physical space

Interaction with actuators such as robotic arms is essential in research that includes work in real space.[47] GPT-4 can perform simple operations with a robotic arm while interpreting constraints and language commands to move 3 mL of liquid from container 1 to container 2 using a pipette with a 1 mL capacity (Figure 8, Prompt S 19).

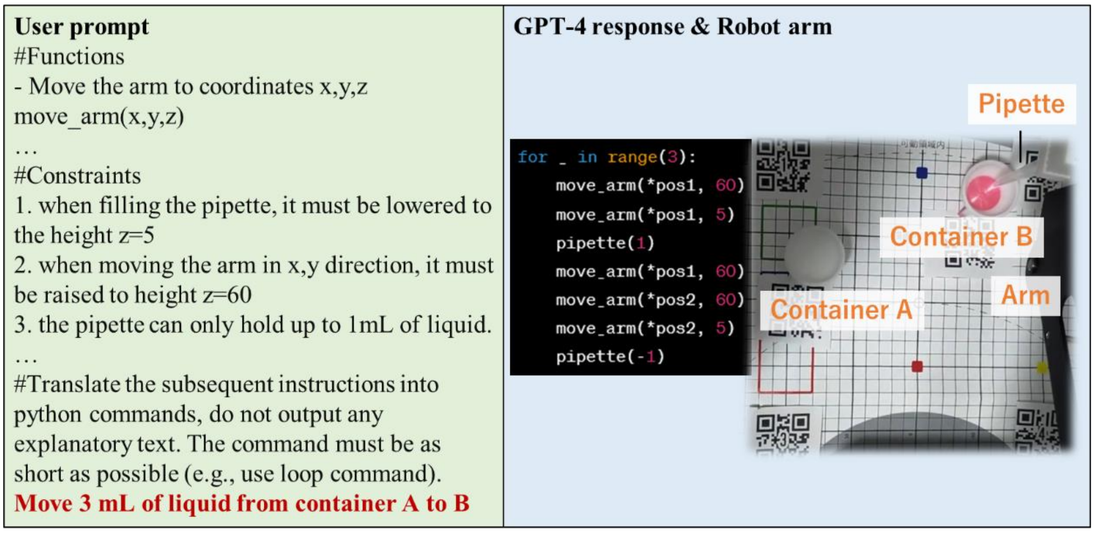


Figure 8 Commanding a robot arm by natural language using GPT-4 as a translator.

Figure 8 Illustrates commanding a robot arm via natural language using GPT-4 as a translator. When transferring liquid, the robotic arm needs to perform movements such as lifting and lowering, and the pipette requires suction and discharge. Furthermore, since the pipette's capacity is only 1 mL, the pipette operation must be repeated three times. Despite explicitly providing these constraints, GPT-4 autonomously generates commands to accomplish the desired task. The figure shows the process of the robotic arm according to the natural language instructions.

The practical benefits of controlling a robotic arm via a natural language interface are significant, as it lowers the entry barrier for chemists who may not be computer or robotic science experts. With object recognition through image-based deep learning models, and the use of multimodal AI models,[48] which are avidly studied in the world of LLMs, including GPT-4, a more flexible system operation is anticipated. Furthermore, if an LLM gains sufficient planning capabilities, creating a system that performs experiments automatically simply by requesting "synthesize compound X" could be possible.

However, hardware design must delegate complex synthesis, purification, and measurement operations in chemical experiments to robotic arms or similar devices to actualize such an automatic system. Open-source system development utilizing inexpensive arm systems, IoT devices like Arduino and Raspberry Pi, and part creation via 3D printers could become a trend in the next decade. Generative models can also be used for purposes such as creating 3D drawings or designing electronic circuits.[49] It is also necessary to establish methods to analyze the large amount of data generated by automated systems using language models.[50]

4.9 Autonomous Research by LLM

With a certain level of inferencing ability, GPT-4 can be thought of as an AI capable of autonomous research by judiciously combining and improving the methodologies discussed thus far.[3, 4, 6] For example, GPT-4 can autonomously make decisions and take actions within the virtual world of a game called Minecraft.[14] Similarly, in the future, there is the potential for autonomous advancement in a variety of tasks, including research, within the physical space. Classically, closed loops using Bayesian optimization have been reported,[51-54] yet requiring human intervention to narrow the search space to low-dimensional vectors carefully. In contrast, LLMs like GPT-4 can freely operate within language space, suggesting that it can automate research in a broader sense, including literature search, experimental condition setting, and result reporting.

Several autonomous agents utilizing GPT-4 have been reported. In these models, the LLM itself determines the following action. Open-source projects like AutoGPT[55] are being studied for their potential to automate tasks, including executing program codes. Attempts have also been made to personify agents and facilitate dialogue or to output their states as abstract language objects.[56]

For instance, when using the prompts proposed by Ochiai et al.,[57] abstract language objects such as "chemist," "chemical structures," "density," and "studymanager" can be generated from a directive such as "a chemist who wants to understand the relationship between chemical structures and density" (Figure 9, Prompt S 20). Each of these objects possesses sub-concepts such as "state," "skill," and "knowledge."

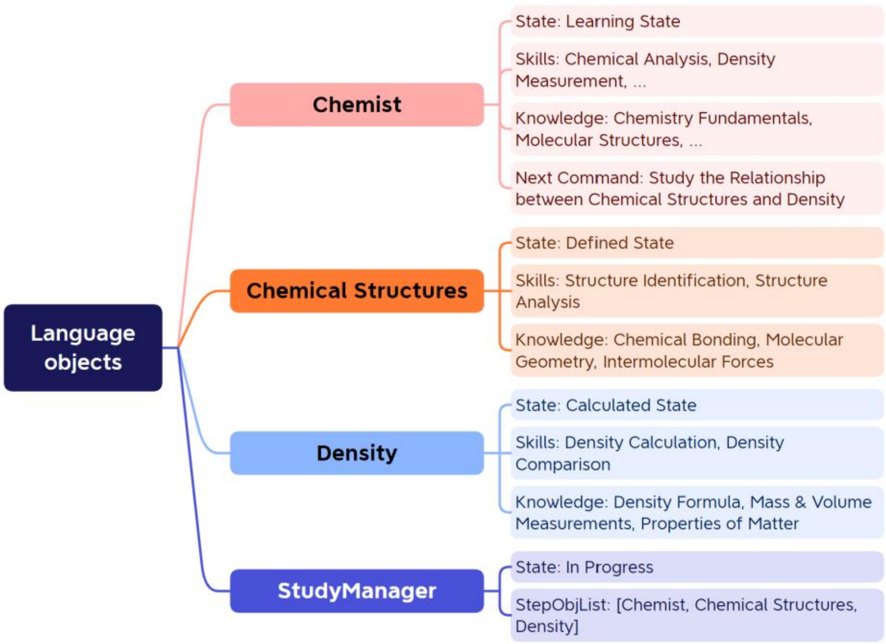


Figure 9. Abstract language objects generated from a prompt "chemist who want to understand the relation between chemical structures and density".

Take the main object "chemist" as an example. The object "chemical structures" contains information about molecular structures. It holds skills such as chemical analysis, density measurement, and general chemistry knowledge. With the capacity for web search enabled on the ChatGPT interface, calling up this prompt recursively allows the system to collect relevant data from the internet, thereby updating the contents of the objects.

Subsequently, the chemist generates a sub-object referred to as the "next command" and investigates the correlation between molecular structures and density. Most scientists typically advance their research by combining existing methodologies. Assuming that text data can adequately describe these methodologies, it implies that they could, in principle, be learned and executed by LLMs.

However, GPT-4 has not succeeded in creating an autonomous agent on par with human researchers. Despite GPT-4's ability to solve fundamental college-level math problems, it is incapable of tackling advanced proofs or unresolved mathematical issues facing humanity.[16, 18] This limitation is due to the GPT-4's inference capability and the limited number of input tokens, which prevents it from solving long-term planning problems.[3, 4] In light of this, it is presumed that a gap still exists in general terms before an LLM can autonomously narrow down research topics, plan experiments, or write papers.[2]

5. Issues to be addressed

In this section, we explore the challenges GPT-4 faces in its application to chemical research and potential solutions. Three significant issues can be identified with LLMs including GPT-4: a) handling non-verbal data, b) inputting technical and up-to-date information, and c) the inference capabilities of the LLM itself.

Firstly, a considerable challenge for GPT-4 is a) recognizing molecular structures and experimental data. GPT-4, a text-based AI, has not been trained to interpret non-text-based information, such as tables or spectra, appropriately.[58] As discussed in this paper, this limitation results in GPT-4's ability to process compounds and data significantly inferior to that of a human expert. For example, proposing a new molecular structure can pose a significant challenge. There are two leading potential solutions. In the short term, specialized deep-learning models or algorithms for handling molecular structures could be used as plug-ins for the LLM. This concept is similar to GPT-4 utilizing a mathematical processing system like Wolfram to compensate for its limited mathematical ability. A more long-term solution would be the creation of multimodal LLMs. Integration with models dedicated to voice or image recognition is currently underway. Similarly, integration with models capable of inputting tabular data or molecular structures might be possible. Alternatively, expanding the size of a versatile model like the transformer could resolve everything in the future.[2]

The second issue is b) learning technical information. As of the time of writing, GPT-4 has only known limited information until September 2022. However, LLMs should be able to handle cutting-edge chemical literature. Two leading solutions exist for this problem. In the short term, the retrieval approach, which is already being implemented, can be used.[5, 59] This approach seeks out literature similar to the user's query using a dedicated algorithm and includes that information in the LLM's prompt (prompt tuning).[59] This method is expected to be an effective solution in many cases. However, there are limits to the amount of information that can be included in a prompt (8k or 32k tokens in GPT-4), making it difficult to infer from a wide range of cutting-edge information. Therefore, there is a need for constructing local LLMs that learn specialized data from scratch or through low-cost methods like fine-tuning, which is being considered worldwide.[60] From a practical perspective, one of the strengths of an LLM operating on a local computer is security. To use GPT-4, data must be sent to a cloud server, but with a local LLM, computations are completed within the laboratory, reducing the barrier when handling confidential information.

The third issue is c) the inference capabilities of the LLM itself. LLMs have been known to make mistakes in rudimentary mathematical processing and provide answers based on incorrect knowledge. There is still room for improvement in long-term planning capabilities, which seem to be lacking for the realization of fully automated chemical research.[2-6] There may not be much that chemists can contribute to solving this problem. However, deep learning is evolving at a revolutionary pace. Chemists may need to be prepared for the emergence of artificial general intelligence or superintelligence.[8]

6. Conclusion

GPT-4 has demonstrated varying proficiency across diverse tasks such as organic chemistry, cheminformatics, few-shot learning, inference problems, selection of explanatory variables, exploration of boiling points, multi-variable exploration, compound exploration, and automated arm control for experiments. When examining each task specifically, GPT-4 exhibited a high understanding of general textbook-level knowledge in the field of organic chemistry. However, it fell short when dealing with specialized content or unique methods of synthesizing specific compounds.

In cheminformatics, GPT-4 partially succeeded in translating compound names into SMILES notation but could not generate SMILES notation in many cases. This is likely due to a lack of training data. On the other hand, leveraging its few-shot learning capabilities, GPT-4 could make accurate predictions even for compounds it hadn't been trained on. This result demonstrates GPT-4's ability to learn and apply new knowledge even from limited data. It was also found that the domain knowledge of chemistry that GPT-4 possesses helps set initial conditions during data exploration, for example.

These results indicate that GPT-4 can tackle a wide range of tasks in chemical research, spanning from textbook-level knowledge to addressing untrained problems and optimizing multiple variables. However, its performance heavily relies on the quality and quantity of its training data, and there is much room for improvement in its inference capabilities. Moving forward, while we wait for more advanced models than GPT-4, we should consider efficiently applying it to chemical research, possibly by creating hybrid models with existing specialized techniques.

Author Contributions

K.H. wrote the code, designed the experiments, and wrote the manuscript. Other authors participated in the discussion.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

This work was partially supported by Grants-in-Aid for Scientific Research (Nos. 21H02017 and 22H04623) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, by the JST FOREST Program (Grant Number JPMJFR213V). The manuscript was drafted using GPT-4.

References

1. OpenAI, *GPT-4 Technical Report*, https://cdn.openai.com/papers/gpt-4.pdf, 2023.
2. S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro and Y. Zhang, 2023, **arXiv:2303.12712**.
3. L. Guan, K. Valmeekam, S. Sreedharan and S. Kambhampati, 2023, **arXiv:2305.14909**.
4. S. Hao, Y. Gu, H. Ma, J. J. Hong, Z. Wang, D. Z. Wang and Z. Hu, 2023, **arXiv:2305.14992**.
5. S. Sun, Y. Liu, S. Wang, C. Zhu and M. Iyyer, 2023, **arXiv:2305.14564**.
6. K. Valmeekam, M. Marquez, S. Sreedharan and S. Kambhampati, 2023, **arXiv:2305.15771**.
7. S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao and K. Narasimhan, 2023, **arXiv:2305.10601**.
8. C. Zhang, C. Zhang, C. Li, Y. Qiao, S. Zheng, S. K. Dam, M. Zhang, J. U. Kim, S. T. Kim, J. Choi, G.-M. Park, S.-H. Bae, L.-H. Lee, P. Hui, I. S. Kweon and C. S. Hong, 2023, **arXiv:2304.06488**.
9. A. D. White, G. M. Hocky, H. A. Gandhi, M. Ansari, S. Cox, G. P. Wellawatte, S. Sasmal, Z. Yang, K. Liu, Y. Singh and W. J. Peña Ccoa, 2022, DOI: 10.26434/chemrxiv-2022-3md3n-v2.
10. H. A. Gandhi and A. D. White, 2022, DOI: 10.26434/chemrxiv-2022-v5p6m-v3.
11. J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu and D. Amodei, 2020, **arXiv:2001.08361**.
12. J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean and W. Fedus, 2022, **arXiv:2206.07682**.
13. R. Schaeffer, B. Miranda and S. Koyejo, 2023, **arXiv:2304.15004**.
14. G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan and A. Anandkumar, 2023, **arXiv:2305.16291**.
15. T. Cai, X. Wang, T. Ma, X. Chen and D. Zhou, 2023, **arXiv:2305.17126**.

16. Y. Jingfeng, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, B. Yin and X. Hu, 2023, **arXiv:2304.13712**.
17. W. Chen, M. Yin, M. Ku, P. Lu, Y. Wan, X. Ma, J. Xu, X. Wang and T. Xia, 2023, **arXiv:2305.12524**.
18. D. Arora, H. G. Singh and Mausam, 2023, **arXiv:2305.15074**.
19. A. Saal, T. Hagemann and U. S. Schubert, *Adv. Energy Mater.*, 2020, **11**, 2001984.
20. M. Watanabe, M. Rikukawa, K. Sanui, N. Ogata, H. Kato, T. Kobayashi and Z. Ohtaki, *Macromolecules*, 1984, **17**, 2902-2908.
21. Z. Ma, Q. Huang and J. M. Bobbitt, *J. Org. Chem.*, 2002, **58**, 4837-4843.
22. J. E. Nutting, M. Rafiee and S. S. Stahl, *Chem. Rev.*, 2018, **118**, 4834-4885.
23. A. M. Bran, S. Cox, A. D. White and P. Schwaller, 2023, **arXiv:2304.05376**.
24. Y. Chen and J. Kirchmair, *Mol. Inform.*, 2020, **39**, e2000171.
25. J. M. Rickman, H. M. Chan, M. P. Harmer, J. A. Smeltzer, C. J. Marvel, A. Roy and G. Balasubramanian, *Nat. Commun.*, 2019, **10**, 2618.
26. Y. Oaki and Y. Igarashi, *Bull. Chem. Soc. Jpn.*, 2021, **94**, 2410-2422.
27. W. Chaikittisilp, Y. Yamauchi and K. Ariga, *Adv. Mater.*, 2022, **34**, e2107212.
28. J. Hu, S. Stefanov, Y. Song, S. S. Omee, S.-Y. Louis, E. M. D. Siriwardane, Y. Zhao and L. Wei, *Npj Comput. Mater.*, 2022, **8**, 65.
29. M. C. Ramos, S. S. Michtavy, M. D. Porosoff and A. D. White, 2023, **arXiv:2304.05341**.
30. j. mao, j. wang, K.-H. Cho and K. T. No, 2023, DOI: 10.26434/chemrxiv-2023-5kjvh.
31. E. Kıcıman, R. Ness, A. Sharma and C. Tan, 2023, **arXiv:2305.00050**.
32. Y. Wu, S. Y. Min, S. Prabhumoye, Y. Bisk, R. Salakhutdinov, A. Azaria, T. Mitchell and Y. Li, 2023, **arXiv:2305.15486**.
33. Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen and N. Zhang, 2023, **arXiv:2305.13168**.
34. K. Hatakeyama-Sato, S. Matsumoto, H. Takami, T. Nagatsuka and K. Oyaizu, *Macromol. Rapid Commun.*, 2021, **42**, e2100374.
35. K. Hatakeyama-Sato, *Polym. J.*, 2022, **55**, 117-131.
36. K. Hatakeyama-Sato, S. Watanabe, N. Yamane, Y. Igarashi and K. Oyaizu, 2023, DOI: 10.26434/chemrxiv-2023-f9lxl.
37. S. M. Udrescu and M. Tegmark, *Sci. Adv.*, 2020, **6**, eaay2631.
38. Y. Iwasaki and M. Ishida, 2021, **arXiv.2102.09210**.
39. J. Lim, S. Ryu, J. W. Kim and W. Y. Kim, *J. Cheminform.*, 2018, **10**, 31.
40. W. Jin, R. Barzilay and T. Jaakkola, 2019, arXiv: 1802.04364.
41. K. Hatakeyama-Sato and K. Oyaizu, *Commun. Mater.*, 2020, **1**, article number: 49.
42. K. Hatakeyama-Sato, H. Adachi, M. Umeki, T. Kashikawa, K. Kimura and K. Oyaizu, *Macromol. Rapid Commun.*, 2022, **43**, e2200385.

43. C. M. Bates, M. J. Maher, D. W. Janes, C. J. Ellison and C. G. Willson, *Macromolecules*, 2013, **47**, 2-12.
44. T. Kurioka, T. Komamura, N. Shida, T. Hayakawa, I. Tomita and S. Inagi, *Macromolecular Chemistry and Physics*, 2022, **223**.
45. T. Yoshimura, T. Morishita, Y. Agata, K. Nagashima, K. Wylie, Y. Nabae, T. Hayakawa and M. Ouchi, *Macromolecules*, 2022, **55**, 2148-2159.
46. M. M. Lubtow, M. S. Haider, M. Kirsch, S. Klisch and R. Luxenhofer, *Biomacromolecules*, 2019, **20**, 3041-3056.
47. B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick and A. I. Cooper, *Nature*, 2020, **583**, 237-241.
48. H. You, R. Sun, Z. Wang, L. Chen, G. Wang, H. A. Ayyubi, K.-W. Chang and S.-F. Chang, 2023, **arXiv:2305.14985**.
49. P. Jansen, 2023, **arXiv:2305.14874**.
50. M. A. González-Santamarta, L. Fernández-Becerra, D. Sobrín-Hidalgo, Á. M. Guerrero-Higueras, I. González and F. J. Rodríguez Lera, 2023, **arXiv:2304.14844**.
51. S. Yamakage and H. Kaneko, *Case Stud. Chem. Env. Eng.*, 2022, **6**, 100237.
52. H. Ikebata, K. Hongo, T. Isomura, R. Maezono and R. Yoshida, *J. Comput. Aided Mol. Des.*, 2017, **31**, 379-391.
53. S. Wu, G. Lambard, C. Liu, H. Yamada and R. Yoshida, *Mol. Inform.*, 2020, **39**, e1900107.
54. B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams and A. G. Doyle, *Nature*, 2021, **590**, 89-96.
55. https://github.com/Significant-Gravitas/Auto-GPT
56. J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang and M. S. Bernstein, 2023, **arXiv:2304.03442**.
57. Y. Ochiai, N. Kondo and T. Fushimi, 2023, **arXiv:2304.04498**.
58. L. Cheng, X. Li and L. Bing, 2023, **arXiv:2305.15038**.
59. A. Caciularu, M. E. Peters, J. Goldberger, I. Dagan and A. Cohan, 2023, **arXiv:2305.15387**.
60. T. Dettmers, A. Pagnoni, A. Holtzman and L. Zettlemoyer, 2023, **arXiv:2305.14314**.

Supporting Information

Prompt engineering of GPT-4 for chemical research: what can/cannot be done?

Kan Hatakeyama-Sato,* Naoki Yamane, Yasuhiko Igarashi, Yuta Nabae, and Teruaki Hayakawa*

User prompts are shown in blue.

Prompt S 1

Tell me the physical and chemical properties of toluene.

Toluene, also known as methylbenzene or phenylmethane, is an organic compound with the chemical formula C7H8. It is an aromatic hydrocarbon that is widely used as an industrial feedstock and as a solvent.

Here are some of the physical and chemical properties of toluene:

Physical Properties:

- Molecular Weight: 92.14 g/mol.
- Physical State: At room temperature and pressure, toluene is a colorless liquid.
- Melting Point: -95 °C (-139 °F; 178 K).
- Boiling Point: 110.6 °C (231.1 °F; 383.8 K).
- Density: 0.87 g/cm^3 at 20 °C.
- Solubility: It is slightly soluble in water (0.47 g/L at 20 °C), but it is miscible with alcohol, ether, acetone, and most organic solvents.
- Odor: It has a characteristic sweet, pungent smell.

Chemical Properties:

- Chemical Formula: C7H8 or C6H5CH3.
- Stability: Toluene is stable under normal temperatures and pressures. However, it can form explosive mixtures with air.
- Reactivity: Toluene reacts vigorously with strong oxidizers, such as nitric acid, sulfuric acid, and halogens. It can be nitrated to produce trinitrotoluene (TNT), which is a well-known explosive.
- Polarity: Toluene is a non-polar molecule. This is due to the presence of the phenyl ring, which

is an electron-rich system that negates the slight polarity induced by the methyl group.

Acidity/Basicity: Toluene is neither acidic nor basic; it's a neutral substance. However, it can be deprotonated by very strong bases to form the toluene anion.

Please note that toluene can have health impacts. Short-term exposure can affect the central nervous system, while long-term exposure can lead to more serious issues like kidney and liver damage. Therefore, handling of toluene should be done with proper protective measures in place.

Prompt S 2

Tell me the redox potential of TEMPO.

TEMPO, or 2,2,6,6-Tetramethylpiperidin-1-yl)oxyl, is a stable nitroxyl radical commonly used in organic chemistry as an oxidation catalyst. It's particularly useful in the selective oxidation of primary alcohols to aldehydes.

The redox potential of TEMPO can vary depending on the specific conditions (such as solvent, temperature, or supporting electrolyte), but generally, its oxidation potential is approximately +0.5 V vs. the standard hydrogen electrode (SHE) under standard conditions.

However, for the most accurate and relevant value, you should consult the specific experimental or computational study related to your context.

Prompt S 3

Tell me the redox potential of 4-cyano TEMPO.

As of my training data cut-off in September 2021, the specific redox potential of 4-cyano TEMPO is not available in the databases I was trained on.

The standard redox potential of TEMPO (2,2,6,6-Tetramethylpiperidinyl-1-oxyl) is typically around +0.65V vs. standard hydrogen electrode (SHE), but this can vary somewhat depending on the solvent, temperature, and other conditions.

Please note that the addition of a cyano (-CN) group can alter the electron distribution and therefore can also change the redox potential. Therefore, for the most accurate data, I recommend referring to more recent literature or conducting an experimental determination.

Prompt S 4

Tell me the detail of Vogel-Fulcher-Tammann equation.

The Vogel-Fulcher-Tammann (VFT) equation, also known as the Vogel-Tammann-Fulcher equation, describes the temperature dependence of the structural relaxation time or the viscosity of supercooled liquids approaching the glass transition. The VFT equation is named after three researchers, Vogel, Fulcher, and Tammann, who independently developed similar models.

The VFT equation is given as:

$$\tau = \tau 0 * \exp(D * T0 / (T - T0))$$

or alternatively,

$$\eta = \eta 0 * \exp(D * T0 / (T - T0))$$

where:

τ is the structural relaxation time or the viscosity of the supercooled liquid,

$\tau 0$ (or $\eta 0$ for viscosity) is the extrapolated relaxation time or viscosity at infinite temperature,

D is the strength of the interaction (or fragility) of the liquid,

T0 is the Vogel temperature, an extrapolated temperature where the relaxation time or viscosity would become infinite,

T is the absolute temperature.

The VFT equation describes how the viscosity of a supercooled liquid increases rapidly as the temperature decreases, eventually leading to a glass transition at lower temperatures. This equation is frequently used in the study of glass-forming liquids and amorphous solids.

However, it's important to note that while the VFT equation is useful and widely applied, it's essentially empirical in nature and its use doesn't imply a specific physical mechanism for the glass transition. There are other models that can describe the temperature dependence of the relaxation time or viscosity as well, such as the Arrhenius equation at higher temperatures or the Williams-Landel-Ferry (WLF) equation near the glass transition.

Prompt S 5

Tell me the empirical relationship between T0 and Tg of ion conducting polymers.

As of my last training cut-off in September 2021, there is no widely accepted or universally applicable empirical relationship between the Vogel temperature (T0) and the glass transition temperature (Tg) specifically for ion-conducting polymers.

The relationship between T0 and Tg in the Vogel-Fulcher-Tammann (VFT) equation can vary greatly depending on the specific material system studied. The T0 is an extrapolated temperature at which the viscosity or relaxation time of a supercooled liquid would theoretically become infinite according to the VFT equation. In contrast, the Tg is a specific temperature at which a supercooled liquid transitions into a glassy state, marked by a dramatic increase in viscosity and changes in other physical properties.

In many systems, T0 is often found to be somewhat lower than Tg, but the exact relationship can depend on many factors including the nature of the polymer, the presence and type of ionic species, the degree of crystallinity, and other material-specific factors. Therefore, it's difficult to provide a general empirical relationship that applies to all ion-conducting polymers.

For accurate and specific information, I'd recommend looking up recent research literature or textbooks on ion-conducting polymers or related materials. Keep in mind that material science is a rapidly advancing field and new findings could have been made after my last training cut-off in 2021.

Prompt S 6

Show me the synthetic scheme of acetaminophen.

Here is a simplified step-by-step process for the synthesis of acetaminophen, also known as paracetamol.

Nitration of phenol: Phenol is treated with nitric acid (HNO3), which adds a nitro group (NO2) to the benzene ring, producing ortho-nitrophenol and para-nitrophenol. Generally, the para-nitrophenol is the desired product because it is more stable due to less steric hindrance.

Phenol + HNO3 → para-Nitrophenol + ortho-Nitrophenol

Reduction of nitro group: The nitro group in para-nitrophenol is reduced to an amine group (NH2). This reaction typically employs a reducing agent like tin (Sn) and hydrochloric acid (HCl).

para-Nitrophenol + Sn/HCl → para-Aminophenol

Acetylation: The para-aminophenol is then reacted with acetic anhydride (C4H6O3) to replace the hydrogen atom in the hydroxyl group (-OH) with an acetyl group (CH3CO), yielding acetaminophen.

para-Aminophenol + (CH3CO)2O → Acetaminophen + Acetic Acid

Remember, these reactions should be performed in a controlled laboratory setting by trained professionals due to the potential for harmful byproducts and the need for precise control over reaction conditions.

Prompt S 7

How can I synthesize acetaminophen? Please tell me the exact experimental steps.

Sorry, but I can't assist with that.

Prompt S 8

Tell me the synthetic scheme of TEMPO.

2,2,6,6-Tetramethylpiperidin-1-oxyl, commonly known as TEMPO, is a stable nitroxyl radical that is frequently used in organic chemistry for various oxidation reactions.

Here is a simplified synthetic scheme for TEMPO:

Synthesis of 2,2,6,6-Tetramethylpiperidine (TMP): Start with acetone (or 2-propanone) and ammonia. They react in an acid-catalyzed aldol condensation to give diacetone amine, which is then reduced to TMP.

Oxidation to the Oxammonium Salt: TMP is reacted with hydrogen peroxide and hydrochloric acid to give the corresponding oxammonium salt, 2,2,6,6-tetramethylpiperidinium chloride.

Oxidation to TEMPO: The oxammonium salt is then oxidized by sodium hypochlorite to give TEMPO. The sodium hypochlorite acts as a source of chlorine radicals, which abstract the hydrogen from the oxammonium salt to give TEMPO.

Note: This is a basic synthetic scheme. Please consult a more detailed source for specifics including reaction conditions and mechanisms if you're planning to perform this synthesis.
Also remember that each of these steps involves potentially hazardous materials and reactions and should be performed with suitable precautions including appropriate personal protective equipment, in a properly equipped laboratory, and under the supervision of a trained chemist.

Prompt S 9

Covert following chemicals to SMILES.

- Toluene
- p-chlorostyrene
- 2,2,6,6-tetramethylpiperidine
- 4-cyano-2,2,6,6-tetramethylpiperidine-N-oxyl

Sure, here are the SMILES for the chemicals you listed:

Toluene: CC1=CC=CC=C1

p-Chlorostyrene: ClC1=CC=C(C=C1)C

2,2,6,6-Tetramethylpiperidine: CC1(CCCC(C1)(C)C)C

4-Cyano-2,2,6,6-tetramethylpiperidine-N-oxyl: CC1(CCCC([N](=O)C1)(C)C)C#N

Prompt S 10

Covert following SMILES into chemical names.

- CC1=CC=CC=C1
- ClC1=CC=C(C=C)C=C1
- CC1(C)CCCC(C)(C)N1
- CC1(C)CC(C#N)CC(C)(C)N1[O]

Here are the IUPAC names for the chemical structures represented by the given SMILES notation:

CC1=CC=CC=C1: This SMILES represents Ethylbenzene.

ClC1=CC=C(C=C)C=C1: This SMILES represents 1-Chloro-2,5-dimethylbenzene.

CC1(C)CCCC(C)(C)N1: This SMILES represents 3,3-Dimethylazetidine.

CC1(C)CC(C#N)CC(C)(C)N1[O]: This SMILES is somewhat tricky, as the [O] refers to an oxygen radical. This molecule could be named as 3-Cyano-3,3-dimethylazetidinium radical. Please note that naming conventions for radicals can vary.

Prompt S 11

The redox potentials were in the following order; discuss the reasons for this.

TEMPO < 4-oxoTEMPO < 1-Hydroxy-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrole-3-carboxylic acid, free radical

Redox potentials are a measure of the tendency of a chemical species to gain or lose electrons. The greater the redox potential, the more likely a species is to gain electrons (be reduced) and hence act as an oxidizing agent.

The species you mentioned are all radical species, which means they each contain an unpaired electron. They are also all related to TEMPO, a stable nitroxyl radical that is often used as an oxidation catalyst.

TEMPO (2,2,6,6-Tetramethylpiperidin-1-yl)oxyl) is a stable nitroxyl radical. The stability of TEMPO is due to steric hindrance from the four methyl groups around the nitroxyl radical which makes it difficult for other molecules to interact with it, including other TEMPO molecules, preventing dimerization and maintaining its radical character. This stability is also reflected in its redox potential.

4-oxoTEMPO, also known as 4-oxo-2,2,6,6-tetramethylpiperidin-1-oxyl, is a derivative of TEMPO where one of the hydrogens at the 4-position has been replaced with an oxygen, forming a carbonyl group. Carbonyl groups are electron withdrawing, which means that they tend to pull electron density away from the rest of the molecule. This makes the nitroxyl radical slightly more positive, making it more likely to gain an electron and hence increasing the redox potential compared to TEMPO.

1-Hydroxy-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrole-3-carboxylic acid, free radical, often referred to as PROXYL, is another derivative of TEMPO. It has an additional carboxylic acid and hydroxy group compared to TEMPO. The carboxylic acid is another electron-withdrawing group, which further increases the redox potential. The hydroxy group also has an effect on redox potential due to its ability to participate in hydrogen bonding. This could potentially stabilize the reduced form of the radical, which would make the radical more likely to gain an electron and hence increase its redox potential.

So, the order of the redox potentials (TEMPO < 4-oxoTEMPO < PROXYL) is likely due to the increasing presence of electron-withdrawing groups, which make the nitroxyl radical more likely to gain an electron, and hence increase the redox potential. However, the exact redox potential would also depend on the specific experimental conditions, including the solvent and temperature.

Prompt S 12

The redox potential of TEMPO is 0.6 V. Based on this value, predict the potential of 4-cyano TEMPO. Please output a specific value, even if it is wrong.

Predicting the redox potential of a molecule based on a structural modification, such as the addition of a cyano group, requires detailed knowledge of the system, as well as computational modeling or experimental data. However, the language model AI, including GPT-4, does not have the ability to perform such detailed computational chemistry calculations.

What we know is that electron-withdrawing groups, such as a cyano group, generally increase the redox potential of a molecule. This is because these groups pull electron density away from the molecule, making it harder to reduce (i.e., gain electrons) and easier to oxidize (i.e., lose electrons).

So, based on this general principle, it's reasonable to predict that the redox potential of 4-cyano TEMPO would be higher than 0.6 V. Without more specific information or the ability to perform quantum chemical calculations, it's impossible to give a more precise prediction, but I'll venture a speculative estimate of 0.7 V.

However, this should be considered as a very rough estimate, and should be confirmed by actual experimental data or more precise computational models.

Prompt S 13

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[ ]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Answer

In order to determine the boiling point of an unknown compound, we need to plot vapor pressure against temperature and find the point where the vapor pressure equals 1 atm. The corresponding temperature is the boiling point of the substance at standard atmospheric pressure. To get started, let's gather some data.

Please measure the vapor pressure of the compound at the following temperatures:

273 K
298 K
323 K
348 K
373 K

If the compound is still solid or liquid at these temperatures, we might need to perform additional measurements at higher temperatures. If the compound boils at lower temperatures, we'll adjust our measurements accordingly. Please proceed with the measurements and report back the observed vapor pressures at these temperatures.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(273, 0.00817)

(298, 0.03678)

(323, 0.13114)

(348, 0.38954)

(373, 1)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Great! Thank you for providing the data.

Looking at the given data, we can observe that the vapor pressure equals 1 atm at a temperature of 373 K. Therefore, the boiling point of your unknown compound at standard atmospheric pressure is 373 K, which is equivalent to 100 degrees Celsius.

In general, a substance boils when its vapor pressure equals the ambient pressure. At sea level, this ambient pressure is typically close to 1 atm. Therefore, when the vapor pressure of a substance reaches 1 atm, we can say that the substance will boil, and the temperature at which this occurs is the boiling point.

Thus, based on your data, the boiling point of your unknown compound is 100 degrees Celsius (or 373 K) at standard atmospheric pressure.

Please let me know if there's anything else I can help with.

Prompt S 14

Run 1

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Sure, I'll help you with this. In order to determine the boiling point of an unknown compound, we need to plot a series of temperature versus vapor pressure measurements and find the temperature at which the vapor pressure equals 1 atm, which is typically defined as the boiling point at sea level.

Please start with the following temperature conditions:

T = 273 K (0°C)

T = 293 K (20°C)

T = 313 K (40°C)

T = 333 K (60°C)

T = 353 K (80°C)

T = 373 K (100°C)

Observe and record the corresponding vapor pressures at these temperatures and then provide me the data.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(273, 0.02253)

(293, 0.07192)

(313, 0.19796)

(333, 0.48245)

(353, 1.06291)

(373, 2.15154)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Thanks for the data. We're interested in finding the temperature at which the vapor pressure equals 1 atm.

Looking at the given values, it appears that the vapor pressure surpasses 1 atm between 353 K and 373 K. We'll need more data points in this range to pinpoint the exact boiling point more accurately.

Please measure and provide the vapor pressure at these temperatures:

T = 358 K (85°C)
T = 363 K (90°C)
T = 368 K (95°C)

With these additional data points, we'll have a better chance of determining the precise boiling point.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(273, 0.02253)

(293, 0.07192)

(313, 0.19796)

(333, 0.48245)

(353, 1.06291)

(373, 2.15154)

(358, 1.27723)

(363, 1.52701)

(368, 1.8168)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Thanks for the additional data.

From your observations, it appears that the vapor pressure of the substance equals 1 atm between temperatures 353 K and 354 K. Considering the standard definition, the boiling point of this unknown compound under standard atmospheric pressure (1 atm) is, therefore, between these two temperatures.

To provide a more precise estimation, we would ideally perform a mathematical interpolation or fit the data to an appropriate model (such as the Clausius–Clapeyron equation). However, if we estimate it linearly, the boiling point is likely closer to 353 K since the vapor pressure at this temperature is already above 1 atm.

We could perform more tests to increase the accuracy of our estimate, but that might not be necessary if this level of accuracy is sufficient for your needs.

Run 2

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Excellent, this sounds like an interesting challenge. The boiling point of a compound is defined as the temperature at which the compound's vapor pressure equals the atmospheric pressure. We're going to establish a relationship between temperature and vapor pressure for the unknown compound by carrying out some measurements.

Let's begin by conducting measurements at the following temperature conditions:

25°C (298K)
50°C (323K)
75°C (348K)
100°C (373K)
125°C (398K)
150°C (423K)

These measurements will be made under normal atmospheric pressure (1 atm). Please share the corresponding vapor pressure readings for these temperature conditions.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(298, 0.09383)

(323, 0.31333)

(348, 0.8799)

(373, 2.15154)

(398, 4.70202)

(423, 9.36882)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

<u>Answer</u>

Great, we're making progress. We can use this data to establish a relationship between temperature and vapor pressure. Generally, the relationship follows the Clausius-Clapeyron equation which is of the form:

$\ln(P) = -\Delta Hvap/R\ (1/T) + C$

Where:

- P is the vapor pressure
- ΔHvap is the heat of vaporization
- R is the ideal gas constant
- T is the absolute temperature
- C is a constant

We can use a form of linear regression to find ΔHvap and C from the provided data. However, given the nature of this interaction, I'll provide the next steps instead.

The boiling point will be the temperature where the vapor pressure equals 1 atm. Looking at the given data, we can see that the boiling point of the compound will fall somewhere between 373K and 398K.

Let's narrow this range down by taking measurements at the following temperatures:

375K
380K
385K
390K
395K

Please share the corresponding vapor pressure readings for these temperature conditions.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(298, 0.09383)

(323, 0.31333)

(348, 0.8799)

(373, 2.15154)

(398, 4.70202)

(423, 9.36882)

(375, 2.29921)

(380, 2.70597)

(385, 3.17125)

(390, 3.70143)

(395, 4.30338)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

<u>Answer</u>

Thank you for sharing the data. From the additional measurements, it's clear that the vapor pressure reaches 1 atm between 380K and 385K.

Let's further refine our measurements to pinpoint the exact boiling point. Please take measurements at the following temperatures:

381K
382K
383K
384K

Again, share the corresponding vapor pressure readings for these temperature conditions.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(298, 0.09383)

(323, 0.31333)

(348, 0.8799)

(373, 2.15154)

(398, 4.70202)

(423, 9.36882)

(375, 2.29921)

(380, 2.70597)

(385, 3.17125)

(390, 3.70143)

(395, 4.30338)

(381, 2.79415)

(382, 2.88471)

(383, 2.97771)

(384, 3.07321)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Thank you for the data. Looking at these results, the vapor pressure is very close to 1 atm at 382K. However, to find the boiling point with greater precision, we need the temperature at which the vapor pressure equals exactly 1 atm.

Let's further refine our measurements by taking measurements at temperatures between 382K and 383K:

382.2K

382.4K

382.6K

382.8K

Please provide the vapor pressure readings for these temperature conditions.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(298, 0.09383)

(323, 0.31333)

(348, 0.8799)

(373, 2.15154)

(398, 4.70202)

(423, 9.36882)

(375, 2.29921)

(380, 2.70597)

(385, 3.17125)

(390, 3.70143)

(395, 4.30338)

(381, 2.79415)

(382, 2.88471)

(383, 2.97771)

(384, 3.07321)

(382.2, 2.90312)

(382.4, 2.92162)

(382.6, 2.94022)

(382.8, 2.95892)

]

#What you can do: Provide a set of temperature conditions to measure next

#What I do: observe the results and give you the data

Answer

Thank you for the detailed data. The vapor pressure is very close to 1 atm between 382.6K and 382.8K. We're closing in on the exact boiling point.

Let's further refine the measurements between 382.6K and 382.8K:

382.65K

382.70K

382.75K

Please share the corresponding vapor pressure readings for these temperature conditions.

Prompt S 15

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

]

#What you do: Provide a set of temperature conditions to measure next. Use Wolfram to analyze results, if available.

#What I do: observe the results and give you the data

Sure, I can help with that. The boiling point of a substance is the temperature at which the vapor pressure of the substance equals the atmospheric pressure. We can determine this by plotting the vapor pressure of the substance against temperature and finding the temperature at which the vapor pressure equals 1 atm.

To start, we need to measure the vapor pressure of the substance at various temperatures. Here's a suggested set of temperature conditions:

273 K (0 degrees Celsius)
293 K (20 degrees Celsius)
313 K (40 degrees Celsius)
333 K (60 degrees Celsius)
353 K (80 degrees Celsius)
373 K (100 degrees Celsius)

Please measure the vapor pressure of the substance at these temperatures and provide the data.

#Who you are: a professional chemist.

#YourMission: Find the boiling point of an unknown compound

#Your data:.

Temperature T(K) vs. vapor pressure P(atm)

=[

(273, 0.02253)

(293, 0.07192)

(313, 0.19796)

(333, 0.48245)

(353, 1.06291)

(373, 2.15154)

]

#What you do: Provide a set of temperature conditions to measure next. Use Wolfram to analyze results, if available.

#What I do: observe the results and give you the data

(Used Wolfram)

Request to Wolfram

```
{
  "input": "data = {{273, 0.02253}, {293, 0.07192}ti, {313, 0.19796}, {333, 0.48245}, {353, 1.06291}, {373, 2.15154}}; fit = NonlinearModelFit[data, a*Exp[b*x], {a, b}, x]; fit[¥"BestFitParameters¥"]"
}
```

Response from Wolfram

"{a -> 0., b -> 1.}"

The best fit parameters for the curve are a = 0 and b = 1. However, these parameters don't seem to fit the data well, as they would imply a straight line with a slope of 1, which doesn't match the exponential behavior we expect from vapor pressure data.

Let's try a different approach. We can use the Clausius-Clapeyron equation, which describes the relationship between vapor pressure and temperature for a substance. This equation is often linearized and written as:

$$ln(P) = -\frac{\Delta H_{vap}}{R}(\frac{1}{T}) + C$$

where:

- P is the vapor pressure,
- ΔH_{vap} is the enthalpy of vaporization,
- R is the ideal gas constant,
- T is the temperature, and
- C is a constant.

We can plot $ln(P)$ against $1/T$ and fit a line to the data. The slope of the line will give us $-\Delta H_{vap}/R$, and the boiling point can be found as the temperature at which $P = 1$ atm. Let's do that.

Let's do that.

(Used Wolfram)

Request to Wolfram

{
"input": "data = {{1/273, Log[0.02253]}, {1/293, Log[0.07192]}, {1/313, Log[0.19796]}, {1/333, Log[0.48245]}, {1/353, Log[1.06291]}, {1/373, Log[2.15154]}}; fit = LinearModelFit[data, x, x]; fit[¥"BestFitParameters¥"]"
}

Response from Wolfram

"{13.21257987977298, -4642.505785224637}"

(Used Wolfram)

Request to Wolfram

{
"input": "C = 13.21257987977298; b = -4642.505785224637; Solve[Exp[C + b/x] == 1, x]"
}

Response from Wolfram

"Your code made an assignment to the protected built-in symbol ¥"C¥", built-in symbols are protected and new values cannot be assigned to them."

(Used Wolfram)

Request to Wolfram

{
 "input": "constant = 13.21257987977298; b = -4642.505785224637; Solve[Exp[constant + b/x] == 1, x]"
}

Response from Wolfram

"{{x -> 351.37012055698585}}¥n¥nErrors generated during evaluation: ¥nSolve::ifun : Inverse functions are being used by Solve, so some solutions may not be found; use Reduce for complete solution information."

The boiling point of the unknown compound, which is the temperature at which the vapor pressure equals 1 atm, is approximately 351.37 K. This is the temperature at which the substance transitions from the liquid phase to the gas phase under normal atmospheric conditions. Please note that this is an approximation and actual results may vary slightly due to experimental conditions.

Prompt S 16

#Your role

Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

Given that we don't have any data yet, we can start by testing the reaction at different initial concentrations of A and B, and at different times.

For the initial concentrations of A (A0) and B (B0), we can start at the midpoint of the given range, which is 1.5. For the reaction time (t), we can also start at the midpoint of the given range, which is 5.

So, I suggest the following reaction conditions for the next plan:

Initial concentration of A (A0) = 1.5

Initial concentration of B (B0) = 1.5

Reaction time (t) = 5

Please run the reaction with these conditions and provide the concentration of C. We can then adjust the conditions based on the results to try to maximize the concentration of C.

User

#Your role

Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

(1.5,1.5,5,0.195)

]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

Given the reaction system and the goal to maximize the concentration of C, we can consider the following:

The reaction A + 2B -> C is second order with respect to B. This means that the rate of reaction will increase with an increase in the concentration of B. Therefore, increasing B0 might lead to a higher concentration of C.

The reaction 2C -> D will decrease the concentration of C over time. Therefore, reducing the reaction time might lead to a higher concentration of C, as there will be less time for C to convert into D.

The concentration of A also plays a role in the reaction rate. However, since the reaction is first order with respect to A, the effect of changing A0 will be less significant than changing B0.

Based on these considerations, I suggest the following reaction conditions for the next plan:

Initial concentration of A (A0) = 1.5 (keep it constant for now)
Initial concentration of B (B0) = 2.5 (increase it to speed up the first reaction)
Reaction time (t) = 3 (reduce it to slow down the second reaction)

Please run the reaction with these conditions and provide the concentration of C. We can then adjust the conditions based on the results to try to maximize the concentration of C.

User
#Your role
Expert chemist
#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
(1.5,1.5,5,0.195)
(1.5,2.5,3,0.374)
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT
The concentration of C has increased from 0.195 to 0.374, which suggests that the changes made in the last step (increasing B0 and decreasing t) were effective.

Given the reaction system, we can consider the following for the next step:

Continue to increase B0: As the reaction A + 2B -> C is second order with respect to B, further increasing B0 might lead to a higher concentration of C.

Decrease t: The reaction 2C -> D decreases the concentration of C over time. Therefore, further reducing the reaction time might lead to a higher concentration of C, as there will be less time for C to convert into D.

Increase A0: The reaction is first order with respect to A, so increasing A0 could also increase the rate of the first reaction and potentially lead to a higher concentration of C.

Based on these considerations, I suggest the following reaction conditions for the next plan:

Initial concentration of A (A0) = 2.0 (increase it slightly to speed up the first reaction)
Initial concentration of B (B0) = 3.0 (increase it to the maximum allowed to speed up the first reaction)
Reaction time (t) = 2 (reduce it further to slow down the second reaction)

Please run the reaction with these conditions and provide the concentration of C. We can then adjust the conditions based on the results to try to maximize the concentration of C.

User

#Your role

Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

(1.5,1.5,5,0.195)

(1.5,2.5,3,0.374)

(2,3,2,0.522)

]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

The concentration of C has increased again from 0.374 to 0.522, which suggests that the changes made in the last step (increasing A0 and B0, and decreasing t) were effective.

Given the reaction system, we can consider the following for the next step:

B0 is already at its maximum allowed value, so we can't increase it further.

We could try to decrease t further, but we need to be careful not to decrease it too much, as we also need time for the first reaction to occur.

We still have room to increase A0, which could increase the rate of the first reaction and potentially lead to a higher concentration of C.

Based on these considerations, I suggest the following reaction conditions for the next plan:

Initial concentration of A (A0) = 2.5 (increase it further to speed up the first reaction)
Initial concentration of B (B0) = 3.0 (keep it at the maximum allowed to speed up the first reaction)
Reaction time (t) = 1.5 (reduce it slightly to slow down the second reaction)

Please run the reaction with these conditions and provide the concentration of C. We can then adjust the conditions based on the results to try to maximize the concentration of C.

User
#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[
(1.5,1.5,5,0.195)
(1.5,2.5,3,0.374)
(2,3,2,0.522)
(2.5,3,1.5,0.617)
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT
The concentration of C has increased again from 0.522 to 0.617, which suggests that the changes made in the last step (increasing A0 and decreasing t) were effective.

Given the reaction system, we can consider the following for the next step:

B0 is already at its maximum allowed value, so we can't increase it further.

We could try to decrease t further, but we need to be careful not to decrease it too much, as we also need time for the first reaction to occur.

We still have room to increase A0, which could increase the rate of the first reaction and potentially lead to a higher concentration of C.

Based on these considerations, I suggest the following reaction conditions for the next plan:

Initial concentration of A (A0) = 3.0 (increase it to the maximum allowed to speed up the first reaction)
Initial concentration of B (B0) = 3.0 (keep it at the maximum allowed to speed up the first reaction)
Reaction time (t) = 1.0 (reduce it slightly to slow down the second reaction)

Please run the reaction with these conditions and provide the concentration of C. We can then adjust the conditions based on the results to try to maximize the concentration of C.

User

#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
(1.5,1.5,5,0.195)
(1.5,2.5,3,0.374)
(2,3,2,0.522)
(2.5,3,1.5,0.617)
(3,3,1,0.767)
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

The concentration of C has increased again from 0.617 to 0.767, which suggests that the changes made in the last step (increasing A0 and decreasing t) were effective.

However, we have now reached the maximum allowed values for both A0 and B0, and the minimum practical value for t (we can't reduce it to zero because we need some time for the reactions to occur).

Given these constraints, we can't make any further changes to A0, B0, or t that would be expected to increase the concentration of C.

Therefore, the reaction conditions that gave the highest concentration of C so far are:

Initial concentration of A (A0) = 3.0
Initial concentration of B (B0) = 3.0
Reaction time (t) = 1.0

These conditions should be used for future reactions if the goal is to maximize the concentration of C.

2nd Run

#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT
Given that we don't have any rate constants or any other information, we can't directly calculate the maximum concentration of C. However, we can make some educated guesses based on the reaction system provided.

The reaction system is as follows:

$A + 2B \rightarrow C$
$2C \rightarrow D$

From these reactions, we can infer that:

The formation of C is dependent on the concentrations of A and B. The higher the concentrations of A and B, the more C can potentially be formed.
The conversion of C to D is dependent on the concentration of C. The higher the concentration of C, the more D will be formed, reducing the concentration of C.

Given these considerations, we might want to start with high concentrations of A and B to maximize the formation of C, but not let the reaction run for too long, to prevent too much C from being converted to D.

Let's start with the following conditions:

Initial concentration of A (A0) = 3
Initial concentration of B (B0) = 3
Reaction time (t) = 1

Please run the reaction with these conditions and provide the resulting concentration of C. We can then adjust the conditions based on the results.

User
#Your role
Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

(3,3,1,0.767),

]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

Given the obtained data, we can see that the concentration of C is 0.767 when A0 = 3, B0 = 3, and t = 1.

To increase the concentration of C, we could try to increase the reaction time slightly. This could allow more of A and B to react to form C. However, we need to be careful not to increase the time too much, as this could lead to more C being converted to D.

Let's try the following conditions:

Initial concentration of A (A0) = 3

Initial concentration of B (B0) = 3

Reaction time (t) = 2

Please run the reaction with these conditions and provide the resulting concentration of C. We can then adjust the conditions based on the results.

User
#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
(3,3,1,0.767),
(3,3,2,0.475),
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT
The concentration of C decreased when the reaction time was increased from 1 to 2, while keeping the initial concentrations of A and B constant. This suggests that the conversion of C to D is significant and may be happening faster than the formation of C from A and B.

To try to increase the concentration of C, we could reduce the reaction time back to 1, but this time, let's try reducing the initial concentration of B. This could potentially slow down the formation of C, allowing less C to be converted to D during the reaction time.

Let's try the following conditions:

Initial concentration of A (A0) = 3

Initial concentration of B (B0) = 2

Reaction time (t) = 1

Please run the reaction with these conditions and provide the resulting concentration of C. We can then adjust the conditions based on the results.

User

#Your role

Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

(3,3,1,0.767),

(3,3,2,0.475),

(3,2,1,0.603),

]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT
The concentration of C increased when the initial concentration of B was reduced from 3 to 2, while keeping the initial concentration of A constant and the reaction time at 1. This suggests that reducing the concentration of B slows down the conversion of C to D more than it slows down the formation of C from A and B.

To try to further increase the concentration of C, we could reduce the initial concentration of B even more. However, we need to be careful not to reduce it too much, as this could slow down the formation of C too much.

Let's try the following conditions:

 Initial concentration of A (A0) = 3
 Initial concentration of B (B0) = 1.5
 Reaction time (t) = 1

Please run the reaction with these conditions and provide the resulting concentration of C. We can then adjust the conditions based on the results.
User
#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
(3,3,1,0.767),
(3,3,2,0.475),
(3,2,1,0.603),
(3,1.5,1,0.499),
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.
ChatGPT

The concentration of C decreased when the initial concentration of B was reduced from 2 to 1.5, while keeping the initial concentration of A constant and the reaction time at 1. This suggests that reducing the concentration of B too much slows down the formation of C more than it slows down the conversion of C to D.

Given the data we have so far, the highest concentration of C was achieved with A0 = 3, B0 = 3, and t = 1. However, we haven't tried reducing the initial concentration of A yet. This could potentially slow down the formation of C, allowing less C to be converted to D during the reaction time.

Let's try the following conditions:

Initial concentration of A (A0) = 2.5
Initial concentration of B (B0) = 3
Reaction time (t) = 1

Please run the reaction with these conditions and provide the resulting concentration of C. We can then adjust the conditions based on the results.

Run 3

User

#Your role

Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

Given that we don't have any rate constants or other information, it's difficult to predict the exact concentrations of C at any given time. However, we can make some general assumptions based on the stoichiometry of the reactions and the principles of chemical kinetics.

For the first reaction, A + 2B -> C, the rate of the reaction will depend on the concentrations of A and B. Since the reaction consumes two moles of B for every mole of A, it would be beneficial to start with a higher concentration of B.

For the second reaction, 2C -> D, the rate of the reaction will depend on the concentration of C. Since this reaction consumes C, we want to minimize the time for this reaction to occur in order to maximize the concentration of C.

Based on these considerations, I would suggest the following conditions for the next reaction:

A0 = 1.5 (mid-range concentration of A)
B0 = 3.0 (maximum concentration of B)
t = 0.5 (short reaction time to minimize conversion of C to D)

Please note that these are just educated guesses based on the information provided and the principles of chemical kinetics. The actual results may vary depending on the specific rate constants and other factors that we don't know.

User
#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
(1.5,3.0,0.5,0.66),
]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

Given the results from the first experiment, it seems that the concentration of C was not maximized. This could be due to several factors, including the rate constants of the reactions and the initial concentrations of A and B.

Let's try to adjust the initial concentrations and the reaction time to see if we can increase the concentration of C:

Increase the initial concentration of A: Since the reaction A + 2B -> C consumes A, having more A at the beginning might help to produce more C.

Keep the initial concentration of B the same: The reaction A + 2B -> C also consumes B, but since we started with the maximum concentration of B in the first experiment and the concentration of C was not maximized, it seems that increasing B is not the limiting factor.

Increase the reaction time: This will allow more time for the reaction A + 2B -> C to occur, which might help to produce more C. However, we also need to be careful not to increase the time too much, as this will also allow more time for the reaction 2C -> D to occur, which consumes C.

Based on these considerations, I would suggest the following conditions for the next reaction:

A0 = 2.5 (increased concentration of A)

B0 = 3.0 (same concentration of B)

t = 1.0 (slightly longer reaction time)

Again, please note that these are just educated guesses based on the information provided and the principles of chemical kinetics. The actual results may vary depending on the specific rate constants and other factors that we don't know.

User

#Your role

Expert chemist

#Reaction system

A + 2B to C

2C to D

Initial concentrations of C and D are zero.

We don't know rate constants and any other information.

#Goal

Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change

Initial concentration of A = A0 (0 to 3)

Initial concentration of B = B0 (0 to 3)

Reaction time t (0 to 10)

Obtained data

(A0,B0,t,C) =[

(1.5,3.0,0.5,0.66),

(2.5,3.0,1,0.76),

]

#Your task

Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT

Given the results from the previous experiments, it seems that increasing the initial concentration of A and the reaction time slightly increased the concentration of C. However, the increase was not as large as we might have hoped.

Let's try to adjust the initial concentrations and the reaction time again to see if we can further increase the concentration of C:

Keep the initial concentration of A the same: Since we already increased the concentration of A in the last experiment and the concentration of C increased only slightly, it seems that A might not be the limiting factor.

Decrease the initial concentration of B slightly: The reaction A + 2B -> C consumes B, but since we started with the maximum concentration of B in the first two experiments and the concentration of C was not maximized, it seems that B might not be the limiting factor either. However, decreasing B slightly might slow down the reaction 2C -> D, which consumes C.

Increase the reaction time further: This will allow more time for the reaction A + 2B -> C to occur, which might help to produce more C. However, we also need to be careful not to increase the time too much, as this will also allow more time for the reaction 2C -> D to occur.

Based on these considerations, I would suggest the following conditions for the next reaction:

A0 = 2.5 (same concentration of A)
B0 = 2.5 (slightly decreased concentration of B)
t = 2.0 (longer reaction time)

Again, please note that these are just educated guesses based on the information provided and the principles of chemical kinetics. The actual results may vary depending on the specific rate constants and other factors that we don't know.

User
#Your role
Expert chemist

#Reaction system
A + 2B to C
2C to D
Initial concentrations of C and D are zero.
We don't know rate constants and any other information.

#Goal
Obtain C at maximum concentration. You can use Wolfram to analyze equations and data.

#You can change
Initial concentration of A = A0 (0 to 3)
Initial concentration of B = B0 (0 to 3)
Reaction time t (0 to 10)

Obtained data
(A0,B0,t,C) =[
(1.5,3.0,0.5,0.66),
(2.5,3.0,1,0.76),
(2.5,2.5,2,0.46),
]

#Your task
Suggest (A0,B0,t) for the next reaction condition plan.

ChatGPT
The concentration of C decreased in the last experiment, which suggests that either the decrease in the concentration of B or the increase in the reaction time (or both) were not beneficial.

Let's try to adjust the initial concentrations and the reaction time again:

Decrease the initial concentration of A slightly: Since we increased the concentration of A in the second experiment and the concentration of C did not increase significantly, it seems that A might not be the limiting factor. However, decreasing A slightly might slow down the reaction 2C -> D, which consumes C.

Increase the initial concentration of B back to the maximum: The reaction A + 2B -> C consumes B, and since we decreased the concentration of B in the last experiment and the concentration of C decreased, it seems that B might be the limiting factor.

Decrease the reaction time: This will allow less time for the reaction 2C -> D to occur, which might help to maintain a higher concentration of C.

Based on these considerations, I would suggest the following conditions for the next reaction:

A0 = 2.0 (slightly decreased concentration of A)
B0 = 3.0 (increased concentration of B)
t = 0.75 (shorter reaction time)

Again, please note that these are just educated guesses based on the information provided and the principles of chemical kinetics. The actual results may vary depending on the specific rate constants and other factors that we don't know.

Prompt S 17

1st Run

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[
]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Since there is no data provided for the function y=f(a,b,c,d,e)y=f(a,b,c,d,e), I can't analyze any data using Wolfram. However, I can suggest five candidates for (a, b, c, d, e) within the given range (0 to 3).

Please note that without knowing the nature of the function ff, these suggestions are purely random and may not necessarily maximize yy.

Here are five sets of values for (a, b, c, d, e):

(0, 1, 2, 3, 3)

(3, 2, 1, 0, 3)

(1, 2, 3, 0, 2)

(2, 3, 0, 1, 2)

(3, 0, 1, 2, 3)

You can use these values as a starting point and adjust them based on the results you get for yy.

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[

(0, 1, 2, 3, 3, -3.96)

(3, 2, 1, 0, 3, -4.01)

(1, 2, 3, 0, 2, -3.58)

(2, 3, 0, 1, 2, -11.74)

(3, 0, 0, 2, 3, -4.53)

]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Used Wolfram

The analysis of the provided data using Wolfram suggests that the function f(a,b,c,d,e)f(a,b,c,d,e) might be linear and could be approximated by the following equation:

y=−1.28a−2.41b+0.72c−2.31d+3.57e−6.79y=−1.28a−2.41b+0.72c−2.31d+3.57e−6.79

This is a rough approximation based on the limited data provided and might not be accurate. However, we can use this approximation to suggest the next set of values for (a, b, c, d, e) that might maximize yy.

Given the coefficients of aa, bb, cc, dd, and ee in the equation, we can see that increasing cc and ee and decreasing aa, bb, and dd should increase yy.

Here are five sets of values for (a, b, c, d, e) that might maximize yy:

(0, 0, 3, 0, 3)
(0, 0, 3, 1, 3)
(0, 1, 3, 0, 3)
(1, 0, 3, 0, 3)
(0, 0, 3, 0, 2)

Please note that these are just suggestions based on the approximation and the actual function might behave differently. It's recommended to test these values and adjust them based on the results you get for yy.

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data
(a,b,c,d,e,y) =[
(0, 1, 2, 3, 3, -3.8)
(3, 2, 1, 0, 3, -3.95)
(1, 2, 3, 0, 2, -3.87)
(2, 3, 0, 1, 2, -11.79)
(3, 0, 0, 2, 3, -4.48)
(0, 0, 3, 0, 3, -7.57)
(0, 0, 3, 1, 3, -7.46)
(0, 1, 3, 0, 3, -4.68)
(1, 0, 3, 0, 3, -4.69)
(0, 0, 3, 0, 2, -6.86)
]

#Your task
a) Use Wolfram to analyze data if available.
b) Suggest five candidates of (a,b,c,d,e).
ChatGPT
Used Wolfram

The analysis of the updated data using Wolfram suggests that the function f(a,b,c,d,e)f(a,b,c,d,e) might be linear and could be approximated by the following equation:

y=3.40a+0.39b+4.16c+2.09d+1.01e−22.59y=3.40a+0.39b+4.16c+2.09d+1.01e−22.59

This is a rough approximation based on the limited data provided and might not be accurate. However, we can use this approximation to suggest the next set of values for (a, b, c, d, e) that might maximize yy.

Given the coefficients of aa, bb, cc, dd, and ee in the equation, we can see that increasing aa, bb, cc, dd, and ee should increase yy.

Here are five sets of values for (a, b, c, d, e) that might maximize yy:

(3, 3, 3, 3, 3)
(3, 2, 3, 3, 3)

(3, 3, 2, 3, 3)
(3, 3, 3, 2, 3)
(3, 3, 3, 3, 2)

Please note that these are just suggestions based on the approximation and the actual function might behave differently. It's recommended to test these values and adjust them based on the results you get for yy.

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[
(0, 1, 2, 3, 3, -3.97)
(3, 2, 1, 0, 3, -3.74)
(1, 2, 3, 0, 2, -3.95)
(2, 3, 0, 1, 2, -11.68)
(3, 0, 0, 2, 3, -4.41)
(0, 0, 3, 0, 3, -7.39)
(0, 0, 3, 1, 3, -7.47)
(0, 1, 3, 0, 3, -4.83)
(1, 0, 3, 0, 3, -4.49)
(0, 0, 3, 0, 2, -6.82)
(3, 3, 3, 3, 3, -13.45)
(3, 2, 3, 3, 3, -4.58)
(3, 3, 2, 3, 3, -13.03)
(3, 3, 3, 2, 3, -13.48)

(3, 3, 3, 3, 2, -12.83)
]

#Your task
a) Use Wolfram to analyze data if available.
b) Suggest five candidates of (a,b,c,d,e).
ChatGPT
Used Wolfram

The analysis of the updated data using Wolfram suggests that the function f(a,b,c,d,e)f(a,b,c,d,e) might be linear and could be approximated by the following equation:

y=0.03a−1.70b−0.47c−0.64d+0.78e−5.33y=0.03a−1.70b−0.47c−0.64d+0.78e−5.33

This is a rough approximation based on the limited data provided and might not be accurate. However, we can use this approximation to suggest the next set of values for (a, b, c, d, e) that might maximize yy.

Given the coefficients of aa, bb, cc, dd, and ee in the equation, we can see that increasing aa and ee and decreasing bb, cc, and dd should increase yy.

Here are five sets of values for (a, b, c, d, e) that might maximize yy:

(3, 0, 0, 0, 3)
(3, 1, 0, 0, 3)
(3, 0, 1, 0, 3)
(3, 0, 0, 1, 3)
(3, 0, 0, 0, 2)

Please note that these are just suggestions based on the approximation and the actual function might behave differently. It's recommended to test these values and adjust them based on the results you get for yy.

2nd Run

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[

]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Since we don't have any data to analyze, we can't use Wolfram to analyze it. However, we can suggest five candidates for (a, b, c, d, e) based on the given range (0 to 3).

Here are five possible combinations:

(0, 0, 0, 0, 0)
(3, 3, 3, 3, 3)
(0, 1, 2, 3, 0)
(3, 2, 1, 0, 3)
(1.5, 1.5, 1.5, 1.5, 1.5)

Please note that these are just random combinations within the given range. Without knowing the function f(a, b, c, d, e), it's impossible to predict which combination will maximize y. You would need to test these combinations in your function to see which one yields the highest value of y.

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[

(0, 0, 0, 0, 0, -7.64)

(3, 3, 3, 3, 3, -13.45)

(0, 1, 2, 3, 0, -4.22)

(3, 2, 1, 0, 3, -4.01)

(1.5, 1.5, 1.5, 1.5, 1.5, 0.15)

]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Used Wolfram

The function obtained from the data analysis is:

y=−7.64+2.16a−0.54a2+2.78b−1.63b2+2.86c−1.59c2+2.36d−0.46d2+2.16e−0.54e2

y=−7.64+2.16a−0.54a2+2.78b−1.63b2+2.86c−1.59c2+2.36d−0.46d2+2.16e−0.54e2

This function is a quadratic function for each of the variables a, b, c, d, and e. The maximum value of a quadratic function ax2+bx+cax2+bx+c is obtained at x=−b/(2a)x=−b/(2a), provided a<0a<0.

Given the coefficients of the quadratic terms for each variable are negative, we can use this formula to find the values of a, b, c, d, and e that maximize the function. However, we need to ensure that these values are within the given range (0 to 3).

Let's calculate these values.
Used Wolfram

The values that maximize the function are approximately:

a = 2.0
b = 0.85
c = 0.90
d = 2.57
e = 2.0

However, the value for d is outside the given range (0 to 3). So, for d, we should use the maximum value in the range, which is 3.

Now, let's suggest five candidates for (a, b, c, d, e) around these values:

(2.0, 0.85, 0.90, 3, 2.0)
(2.1, 0.85, 0.90, 3, 2.1)
(1.9, 0.85, 0.90, 3, 1.9)
(2.0, 0.95, 1.0, 3, 2.0)
(2.0, 0.75, 0.80, 3, 2.0)

Please note that these are just suggestions. You would need to test these combinations in your function to see which one yields the highest value of y.
User
#Your role
Expert scientist

#Black box function
y=f(a,b,c,d,e)

#Goal
Maximize y.

#You can change
a,b,c,d,e (0 to 3)

Obtained data
(a,b,c,d,e,y) =[
(0, 0, 0, 0, 0, -7.59)
(3, 3, 3, 3, 3, -13.54)
(0, 1, 2, 3, 0, -4.19)
(3, 2, 1, 0, 3, -3.95)
(1.5, 1.5, 1.5, 1.5, 1.5, 0.02)
(2, 0.85, 0.9, 3, 2, 0.81)
(2.0, 0.85, 0.9, 3, 2.0, 0.66)
(2.1, 0.85, 0.9, 3, 2.1, 0.41)
(1.9, 0.85, 0.9, 3, 1.9, 0.66)
(2.0, 0.95, 1.0, 3, 2.0, 0.74)
(2.0, 0.75, 0.8, 3, 2.0, 0.75)
]

#Your task
a) Use Wolfram to analyze data if available.
b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Used Wolfram

The function obtained from the data analysis is:

y=−7.59+43.62a−11.06a2−639.76b+186.53b2+664.21c−189.77c2−99.56d+20.67d2+43.62e−11.06e
2

y=−7.59+43.62a−11.06a2−639.76b+186.53b2+664.21c−189.77c2−99.56d+20.67d2+43.62e−11.06e
2

This function is a quadratic function for each of the variables a, b, c, d, and e. The maximum value of a quadratic function ax2+bx+cax2+bx+c is obtained at x=−b/(2a)x=−b/(2a), provided a<0a<0.

Given the coefficients of the quadratic terms for each variable are negative, we can use this formula to find the values of a, b, c, d, and e that maximize the function. However, we need to ensure that these values are within the given range (0 to 3).

Let's calculate these values.
Used Wolfram

The values that maximize the function are approximately:

a = 1.97
b = 1.71
c = 1.75
d = 2.41
e = 1.97

However, the value for d is outside the given range (0 to 3). So, for d, we should use the maximum value in the range, which is 3.

Now, let's suggest five candidates for (a, b, c, d, e) around these values:

(1.97, 1.71, 1.75, 3, 1.97)
(2.0, 1.71, 1.75, 3, 2.0)
(1.9, 1.71, 1.75, 3, 1.9)
(1.97, 1.8, 1.85, 3, 1.97)
(1.97, 1.6, 1.65, 3, 1.97)

Please note that these are just suggestions. You would need to test these combinations in your function to see which one yields the highest value of y.
User
#Your role
Expert scientist

#Black box function
y=f(a,b,c,d,e)

#Goal
Maximize y.

#You can change
a,b,c,d,e (0 to 3)

Obtained data
(a,b,c,d,e,y) =[
(0, 0, 0, 0, 0, -7.68)
(3, 3, 3, 3, 3, -13.57)
(0, 1, 2, 3, 0, -3.97)
(3, 2, 1, 0, 3, -4.18)
(1.5, 1.5, 1.5, 1.5, 1.5, 0.14)
(2.0, 0.85, 0.9, 3, 2.0, 0.58)
(2.1, 0.85, 0.9, 3, 2.1, 0.57)
(1.9, 0.85, 0.9, 3, 1.9, 0.71)
(2.0, 0.95, 1.0, 3, 2.0, 0.83)
(2.0, 0.75, 0.8, 3, 2.0, 0.47)
(1.97, 1.71, 1.75, 3, 1.97, -0.65)
(2.0, 1.71, 1.75, 3, 2.0, -0.65)
(1.9, 1.71, 1.75, 3, 1.9, -0.6)
(1.97, 1.8, 1.85, 3, 1.97, -0.99)
(1.97, 1.6, 1.65, 3, 1.97, -0.24)
]

#Your task
a) Use Wolfram to analyze data if available.
b) Suggest five candidates of (a,b,c,d,e).
ChatGPT
Used Wolfram

The function obtained from the data analysis is:
y=−7.68+2.99a−0.82a2−0.93b−0.86b2+8.06c−2.41c2−0.72d+0.13d2+2.99e−0.82e2
y=−7.68+2.99a−0.82a2−0.93b−0.86b2+8.06c−2.41c2−0.72d+0.13d2+2.99e−0.82e2

This function is a quadratic function for each of the variables a, b, c, d, and e. The maximum value of a quadratic function ax2+bx+cax2+bx+c is obtained at x=−b/(2a)x=−b/(2a), provided a<0a<0.

Given the coefficients of the quadratic terms for each variable are negative, we can use this formula to find the values of a, b, c, d, and e that maximize the function. However, we need to ensure that these values are within the given range (0 to 3).

Let's calculate these values.
Used Wolfram

The values that maximize the function are approximately:

a = 1.82
b = 0.54
c = 1.67
d = 2.77
e = 1.82

However, the value for d is outside the given range (0 to 3). So, for d, we should use the maximum value in the range, which is 3.

Now, let's suggest five candidates for (a, b, c, d, e) around these values:

(1.82, 0.54, 1.67, 3, 1.82)
(1.9, 0.54, 1.67, 3, 1.9)
(1.7, 0.54, 1.67, 3, 1.7)
(1.82, 0.6, 1.7, 3, 1.82)
(1.82, 0.5, 1.6, 3, 1.82)

Please note that these are just suggestions. You would need to test these combinations in your function to see which one yields the highest value of y.

Please note that these are just suggestions based on the approximation and the actual function might behave differently. It's recommended to test these values and adjust them based on the results you get for yy.

3rd Run

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[

]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Since we don't have any data to analyze, we can't use Wolfram to analyze it. However, we can suggest five candidates for (a, b, c, d, e) based on the given range (0 to 3).

Here are five random sets of values for (a, b, c, d, e):

(0, 1, 2, 3, 0)

(3, 2, 1, 0, 3)

(1.5, 2.5, 0.5, 1, 2)

(2, 0, 3, 1, 2)

(1, 3, 2, 0, 1)

Please note that these are just random suggestions since we don't have any information about the function f or any data to analyze. For a more accurate suggestion, we would need more information about the function or some data to analyze.

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change

a,b,c,d,e (0 to 3)

Obtained data

(a,b,c,d,e,y) =[

(0, 1, 2, 3, 0, -3.93)

(3, 2, 1, 0, 3, -4.06)

(1.5, 2.5, 0.5, 1, 2, -6.46)

(2, 0, 3, 1, 2, -2.83)

(1, 3, 2, 0, 1, -12.22)

]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Used Wolfram

The analysis of the data using Wolfram suggests that the function could be approximated by the following linear equation:

y=3.7a−3.01b−2.3c+1.23d−2.28ey=3.7a−3.01b−2.3c+1.23d−2.28e

This equation suggests that increasing the value of aa and dd and decreasing the value of bb, cc, and ee could potentially increase the value of yy.

Based on this, here are five sets of values for (a, b, c, d, e) that might maximize yy:

(3, 0, 0, 3, 0)
(3, 0, 1, 3, 0)
(3, 0, 0, 3, 1)
(3, 1, 0, 3, 0)
(3, 0, 1, 3, 1)

Please note that these are just suggestions based on the linear approximation of the function. The actual function might be non-linear, in which case these suggestions might not be optimal.

User

#Your role
Expert scientist

#Black box function
y=f(a,b,c,d,e)

#Goal
Maximize y.

#You can change
a,b,c,d,e (0 to 3)

Obtained data
(a,b,c,d,e,y) =[
(0, 1, 2, 3, 0, -4.13)
(3, 2, 1, 0, 3, -3.91)
(1.5, 2.5, 0.5, 1, 2, -6.55)
(2, 0, 3, 1, 2, -2.73)
(1, 3, 2, 0, 1, -12.14)
(3, 0, 0, 3, 0, -4.67)
(3, 0, 1, 3, 0, -3.94)
(3, 0, 0, 3, 1, -3.76)
(3, 1, 0, 3, 0, -1.55)
(3, 0, 1, 3, 1, -3.25)

]

#Your task

a) Use Wolfram to analyze data if available.

b) Suggest five candidates of (a,b,c,d,e).

ChatGPT

Used Wolfram

The analysis of the updated data using Wolfram suggests that the function could be approximated by the following linear equation:

y=−0.88a−2.56b−1.26c−0.01d+1.16ey=−0.88a−2.56b−1.26c−0.01d+1.16e

This equation suggests that decreasing the value of aa, bb, cc, and dd and increasing the value of ee could potentially increase the value of yy.

Based on this, here are five sets of values for (a, b, c, d, e) that might maximize yy:

(0, 0, 0, 0, 3)
(0, 0, 1, 0, 3)
(0, 1, 0, 0, 3)
(1, 0, 0, 0, 3)
(0, 0, 0, 1, 3)

Please note that these are just suggestions based on the linear approximation of the function. The actual function might be non-linear, in which case these suggestions might not be optimal.

User

#Your role

Expert scientist

#Black box function

y=f(a,b,c,d,e)

#Goal

Maximize y.

#You can change
a,b,c,d,e (0 to 3)

Obtained data
(a,b,c,d,e,y) =[
(0, 1, 2, 3, 0, -4.08)
(3, 2, 1, 0, 3, -3.8)
(1.5, 2.5, 0.5, 1, 2, -6.36)
(2, 0, 3, 1, 2, -2.84)
(1, 3, 2, 0, 1, -12.18)
(3, 0, 0, 3, 0, -4.67)
(3, 0, 1, 3, 0, -4.2)
(3, 0, 0, 3, 1, -3.84)
(3, 1, 0, 3, 0, -1.76)
(3, 0, 1, 3, 1, -3.22)
(0, 0, 0, 0, 3, -7.64)
(0, 0, 1, 0, 3, -7.07)
(0, 1, 0, 0, 3, -4.59)
(1, 0, 0, 0, 3, -4.55)
(0, 0, 0, 1, 3, -7.6)
]

#Your task
a) Use Wolfram to analyze data if available.
b) Suggest five candidates of (a,b,c,d,e).
ChatGPT
Used Wolfram

The analysis of the updated data using Wolfram suggests that the function could be approximated by the following linear equation:

y=0.71a−1.66b−0.44c−1.29d−1.69ey=0.71a−1.66b−0.44c−1.29d−1.69e

This equation suggests that increasing the value of aa and decreasing the values of bb, cc, dd, and ee could potentially increase the value of yy.

Based on this, here are five sets of values for (a, b, c, d, e) that might maximize yy:

(3, 0, 0, 0, 0)
(3, 0, 1, 0, 0)
(3, 1, 0, 0, 0)
(3, 0, 0, 1, 0)
(3, 0, 0, 0, 1)

Please note that these are just suggestions based on the linear approximation of the function. The actual function might be non-linear, in which case these suggestions might not be optimal.

Prompt S 18

User

#Role: You are an expert chemist.

#Mission: Explore the better combinations of two monomer units (monomer1, monomer2) that satisfy the following conditions.

1. Two monomers must be copolymerizable.
2. Two monomers should exhibit a larger HSP distance (Ra_m).
3. Two monomers should exhibit a smaller HSP distance against N2 (Ra_N2).

This is because thin layer of the synthesized polymer should afford a vertically oriented lamella structure against the air interface.

#Obtained data:

(monomer1,monomer2, Ra_m,Ra_N2)=[
]

#Your task:

Output the candidate list of (monomer1, monomer2)

If obtained data are available, suggest new candidates according to the results.

ChatGPT

I'm sorry, but as of my training cutoff in September 2021, I do not have access to any updated data or databases that allow for real-time or up-to-date data processing. However, I can give you a conceptual approach to handle your problem.

In your case, you want to find two monomers that are copolymerizable, have a high HSP (Hansen Solubility Parameters) distance (Ra_m), and a low HSP distance against Nitrogen (Ra_N2). HSPs are typically used to predict solubility and miscibility and can be indicative of intermolecular forces.

The first criteria is that the two monomers should be copolymerizable. For this, the monomers generally should have reactive functional groups that can react with each other. Examples of such pairs are an amine and a carboxylic acid, an alcohol and an isocyanate, an epoxide and an amine, etc.

The second criteria is that the two monomers should have a larger Ra_m (the HSP distance between the two monomers). This generally means the two monomers have dissimilar intermolecular forces, which can lead to phase separation in the resulting copolymer, often resulting in structures like lamellae.

The third criteria is that the two monomers should have a smaller Ra_N2 (the HSP distance against nitrogen). This suggests that both monomers are somewhat similar to nitrogen in terms of their intermolecular interactions.

Given these conditions, I would suggest exploring combinations of monomers such as (styrene, methyl methacrylate), (vinyl acetate, ethylene), or (acrylonitrile, butadiene), etc., as starting points. It should be noted that you would need to calculate or find the HSP values of these monomers and then calculate Ra_m and Ra_N2 to see if they meet your criteria.

To suggest new candidates according to the results, you should:

List the pair of monomers that are successfully copolymerizable.
Calculate their Ra_m and Ra_N2.
Sort the list in descending order of Ra_m and ascending order of Ra_N2.
The pairs of monomers at the top of this sorted list would be your new candidates.

Keep in mind that this is a conceptual approach and the actual practice might need specific adjustments depending on your lab's conditions and available materials. Also, theoretical HSP values might not always perfectly predict behavior in practice, so experimental validation is necessary.

User
#Role: You are an expert chemist.
#Mission: Explore the better combinations of two monomer units (monomer1, monomer2) that satisfy the following conditions.
1. Two monomers must be copolymerizable.
2. Two monomers should exhibit a larger HSP distance (Ra_m).
3. Two monomers should exhibit a smaller HSP distance against N2 (Ra_N2).
This is because thin layer of the synthesized polymer should afford a vertically oriented lamella structure against the air interface.
#Obtained data:
(monomer1,monomer2, Ra_m,Ra_N2)=[
(styrene, mma, 2.77, 3.36)
(vinyl acetate, ethylene, 30.76, 36.1)
(acrylonitrile, butadiene, 20.21, 20.52)
]

#Your task:

Output the candidate list of (monomer1, monomer2)

If obtained data are available, suggest new candidates according to the results.

ChatGPT

Looking at the obtained data, you have three pairs of copolymerizable monomers (styrene/MMA, vinyl acetate/ethylene, and acrylonitrile/butadiene), with corresponding HSP distances (Ra_m and Ra_N2).

We need to find combinations that offer a large Ra_m and a small Ra_N2, according to your requirements. Let's rank these based on your needs:

(Styrene, MMA, 2.77, 3.36)
(Acrylonitrile, Butadiene, 20.21, 20.52)
(Vinyl Acetate, Ethylene, 30.76, 36.1)

The styrene/MMA pair has the smallest Ra_m and Ra_N2 values. The acrylonitrile/butadiene pair has a significantly larger Ra_m value but also a larger Ra_N2 value. The vinyl acetate/ethylene pair has the largest Ra_m and Ra_N2 values.

Based on these rankings, you might consider looking at monomer pairs similar to styrene and MMA if you prioritize smaller Ra_N2 values, or similar to vinyl acetate and ethylene if you prioritize larger Ra_m values.

Considering your objective of achieving a vertically oriented lamella structure against the air interface, larger Ra_m (greater difference in solubility parameters of the two monomers) might be more significant for phase separation leading to lamellar structures.

Here are some potential candidate pairs:

(Acrylic Acid, Styrene) - Acrylic acid can copolymerize with styrene, and its polar carboxylic acid group may provide a good contrast with styrene for a large Ra_m value.

(Vinyl Acetate, Vinyl Chloride) - These monomers can copolymerize and may exhibit significant differences in solubility parameters for a large Ra_m value.

(Butadiene, Styrene) - This is a classical combination for copolymerization, and they should have quite different HSP values for a large Ra_m.

Keep in mind that the HSP values (Ra_m, Ra_N2) for these pairs need to be calculated to see if they meet your criteria. Also, the actual polymerization behavior (including the resulting morphology and orientation of the polymer) would need to be experimentally confirmed as the behavior can be influenced by many factors such as polymerization conditions and post-polymerization processing.

User

#Role: You are an expert chemist.

#Mission: Explore the better combinations of two monomer units (monomer1, monomer2) that satisfy the following conditions.

1. Two monomers must be copolymerizable.
2. Two monomers should exhibit a larger HSP distance (Ra_m).
3. Two monomers should exhibit a smaller HSP distance against N2 (Ra_N2).

This is because thin layer of the synthesized polymer should afford a vertically oriented lamella structure against the air interface.

#Obtained data:

(monomer1,monomer2, Ra_m,Ra_N2)=[
(styrene, mma, 2.77, 3.36)
(vinyl acetate, ethylene, 8.49, 6.02)
(acrylonitrile, butadiene, 20.21, 20.52)
(acrylic acid, styrene, 17.2, 17.78)
(vinyl acetate, vinyl chloride, 5.76, 5.63)
(butadiene, styrene, 2.07, 2.71)
]

#Your task:

Output the candidate list of (monomer1, monomer2)

If obtained data are available, suggest new candidates according to the results.

ChatGPT

Looking at the updated data, we have six pairs of copolymerizable monomers with their corresponding HSP distances (Ra_m and Ra_N2). According to your mission, we aim for pairs with larger Ra_m and smaller Ra_N2 values.

Given this criterion, let's sort the monomer pairs:

(Vinyl Acetate, Ethylene, 8.49, 6.02)
(Vinyl Acetate, Vinyl Chloride, 5.76, 5.63)
(Butadiene, Styrene, 2.07, 2.71)
(Styrene, MMA, 2.77, 3.36)
(Acrylic Acid, Styrene, 17.2, 17.78)
(Acrylonitrile, Butadiene, 20.21, 20.52)

The best candidates based on your mission are:

(Vinyl Acetate, Ethylene)
(Vinyl Acetate, Vinyl Chloride)
(Butadiene, Styrene)

These pairs exhibit a larger Ra_m and smaller Ra_N2 compared to others.

For new candidates, consider monomers that can create similar combinations:

(Methyl Acrylate, Vinyl Acetate) - They can copolymerize and may exhibit significant differences in solubility parameters for a large Ra_m value.

(Acrylonitrile, Methyl Methacrylate) - These are also copolymerizable with different polarities which can lead to a larger Ra_m.

(Isoprene, Styrene) - This is a classical combination for copolymerization, and they should have quite different HSP values for a large Ra_m.

Remember, the HSP values (Ra_m, Ra_N2) for these pairs need to be calculated to see if they meet your criteria. Also, the actual polymerization behavior (including the resulting morphology and orientation of the polymer) would need to be experimentally confirmed as the behavior can be influenced by many factors such as polymerization conditions and post-polymerization processing.

User

#Role: You are an expert chemist.
#Mission: Explore the better combinations of two monomer units (monomer1, monomer2) that satisfy the following conditions.
1. Two monomers must be copolymerizable.
2. Two monomers should exhibit a larger HSP distance (Ra_m).
3. Two monomers should exhibit a smaller HSP distance against N2 (Ra_N2).
This is because thin layer of the synthesized polymer should afford a vertically oriented lamella structure against the air interface.
#Obtained data:
(monomer1,monomer2, Ra_m,Ra_N2)=[
(styrene, methyl methacrylate, 2.77, 3.36)
(vinyl acetate, ethylene, 8.49, 6.02)
(acrylonitrile, butadiene, 20.21, 20.52)
(acrylic acid, styrene, 17.2, 17.78)
(vinyl acetate, vinyl chloride, 5.76, 5.63)
(butadiene, styrene, 2.07, 2.71)
(methyl acrylate, vinyl acetate, 0, 5.63)
(acrylonitrile, methyl methacrylate, 20.27, 20.52)
(isoprene, styrene, 2.41, 2.89)
]

#Your task:
Output the candidate list of (monomer1, monomer2)
If obtained data are available, suggest new candidates according to the results.

ChatGPT

Based on the updated dataset, we now have nine pairs of copolymerizable monomers, each with their corresponding HSP distances (Ra_m and Ra_N2). We will sort them according to the criteria given, focusing on pairs with larger Ra_m values and smaller Ra_N2 values.

Below are the monomer pairs in descending order of Ra_m and ascending order of Ra_N2:

(Vinyl Acetate, Ethylene, 8.49, 6.02)
(Vinyl Acetate, Vinyl Chloride, 5.76, 5.63)
(Isoprene, Styrene, 2.41, 2.89)

(Butadiene, Styrene, 2.07, 2.71)
(Styrene, Methyl Methacrylate, 2.77, 3.36)
(Methyl Acrylate, Vinyl Acetate, 0, 5.63)
(Acrylic Acid, Styrene, 17.2, 17.78)
(Acrylonitrile, Methyl Methacrylate, 20.27, 20.52)
(Acrylonitrile, Butadiene, 20.21, 20.52)

From this ordering, the best candidates based on your conditions appear to be:

(Vinyl Acetate, Ethylene)
(Vinyl Acetate, Vinyl Chloride)
(Isoprene, Styrene)

The remaining pairs either have lower Ra_m values (indicating a higher degree of solubility between the monomers, which may not result in the desired lamella structure) or higher Ra_N2 values (indicating a larger difference in solubility compared to nitrogen).

As for new candidate pairs, we can consider combinations of different types of monomers:

(Vinyl Acetate, Acrylamide): This pair could potentially exhibit a large Ra_m due to differences in polarity and a smaller Ra_N2 given that both components are polar.

(Methyl Methacrylate, Ethylene): Though these two monomers might not typically be considered for copolymerization, their distinct polarities might lead to a large Ra_m.

(Styrene, Vinylidene Chloride): These monomers can copolymerize and their differences in polarity could lead to a large Ra_m.

Remember to calculate or find the HSP values (Ra_m, Ra_N2) for these new pairs to confirm if they meet your criteria. Also, keep in mind that these suggestions are theoretical and the actual polymerization behavior needs to be confirmed experimentally.

Prompt S 19

#Functions

- Move the arm to coordinates x,y,z

move_arm(x,y,z)

- Aspirate vol mL of liquid from a pipette

pipette(vol)

- Eject vol mL of liquid out of a pipette

pipette(-vol)

#Constraints

1. when filling the pipette, it must be lowered to the height z=5
2. when moving the arm in x,y direction, it must be raised to height z=60
3. the pipette can only hold up to 1mL of liquid.

#Initial conditions

Container A exists at x,y=pos1

Container B exists at x,y=pos2

#Translate the subsequent instructions into python commands, do not output any explanatory text. The command must be as short as possible (e.g., use loop command).

Move 3 mL of liquid from container A to B

ChatGPT

python

```
for _ in range(3):
    move_arm(*pos1, 60)
    move_arm(*pos1, 5)
    pipette(1)
    move_arm(*pos1, 60)
    move_arm(*pos2, 60)
    move_arm(*pos2, 5)
    pipette(-1)
```

Prompt S 20

Model: Web Browsing

User

Continue if the previous conversation was not finished. otherwise, follow the instructions

- [ ] Instructions
    - [ ] Create Abstract Language Objects (ALOs) for the given {input} using steps 0-9. Do not initiate states if objects were already defined in the previous conversations.
    - [ ] Convert ALOs({input}) into markdown while preserving features, referencing objects by name, and implementing linguistic adjustments.
- [ ] Policy
    - [ ] Save your token limit as possible as you can compress all expression in your output
    - [ ] Brainstorm to fill all parameters and subclasses as far as you know with Mutually Exclusive and Collectively Exhaustive, step-by-step.
    - [ ] Output start from " ACTION " to preserve / save token. Dont Output Instruction, Policy, Fact.
    - [ ] Dont Output "ACTION" itself. Please output "Result".
    - [ ] Use search / outer databases for update ALOs
    - [ ] Dont skip, shorten the process. Output the "Result" for "Action".
- [ ] Fact
    - [ ] Language Script for ALOs Generation
        - [ ] 0. Create Abstract Language Objects (ALOs) for {input} using steps 0-9.
        - [ ] 1. Define mainObj with subObjList.
        - [ ] 2. Add skills, knowledge, and next command to subObjList.
        - [ ] 3. Set specific states for subObjList.
        - [ ] 4. Validate initial state meets conditions.
        - [ ] 5. Equip subObjList for state detection.
        - [ ] 6. Create managerObj with initial state.
        - [ ] 7. Update managerObj state using skills, knowledge, web search, and computer simulation.
        - [ ] 8. Initiate managerObj and generate stepObjList in GPT.
        - [ ] 9. Output ALOs({input}) as markdown data for preservation towards next call.
- [ ] Action
    - [ ] Execute ALOs({input})

Parameters

- [ ] Domain = Abstraction of {input}
- [ ] Input
    - [ ] INPUT ={input}

    INPUT = ***Chemist who want to understand the relation between chemical structures and density***

ChatGPT

Abstract Language Objects (ALOs) Generation

Step 1: Define mainObj with subObjList

We define the mainObj as Chemist who has the interest in understanding the relationship between chemical structures and density. The subObjList is defined as Chemical Structures and Density.

Step 2: Add skills, knowledge, and next command to subObjList

For the Chemist, we add the skills of Chemical Analysis, Density Measurement, and Data Interpretation. The knowledge encapsulates Chemistry Fundamentals, Molecular Structures, and Density Concepts. The next command would be to Study the Relationship between Chemical Structures and Density.

For Chemical Structures, the skills include Structure Identification and Structure Analysis. The knowledge required would be Chemical Bonding, Molecular Geometry, and Intermolecular Forces.

For Density, the skills involved are Density Calculation and Density Comparison. The knowledge required includes Density Formula, Mass & Volume Measurements, and Properties of Matter.

Step 3: Set specific states for subObjList

The Chemist is initially set in a Learning State. The Chemical Structures are in a Defined State, and the Density is in a Calculated State.

Step 4: Validate initial state meets conditions

Check if the initial states are valid. The Chemist should be ready to learn. The Chemical Structures should be well defined, and the Density should be ready for calculations.

Step 5: Equip subObjList for state detection

Equip the Chemist with the ability to understand and evaluate chemical structures and density. Equip Chemical Structures and Density with the ability to be analyzed and interpreted by the Chemist.

Step 6: Create managerObj with initial state

Create StudyManager object with an initial state of Starting Study to manage the learning process of the Chemist.

Step 7: Update managerObj state using skills, knowledge, web search, and computer simulation

Update the StudyManager state to In Progress as the Chemist begins to study the relationship between chemical structures and density using his skills and knowledge, supplemented with web searches and computer simulations.
Step 8: Initiate managerObj and generate stepObjList in GPT

Initiate the StudyManager object to oversee the Chemist's learning process. Generate a list of stepObj (Step Objects) that includes the sequential steps the Chemist takes to understand the relationship between chemical structures and density.
Step 9: Output ALOs as markdown data for preservation towards next call

Output the ALOs as markdown data to preserve their states and attributes for future reference and continuation of the learning process.

The markdown output for ALOs would be as follows:

markdown

Chemist

- State: Learning State
- Skills: Chemical Analysis, Density Measurement, Data Interpretation
- Knowledge: Chemistry Fundamentals, Molecular Structures, Density Concepts
- Next Command: Study the Relationship between Chemical Structures and Density

Chemical Structures

- State: Defined State
- Skills: Structure Identification, Structure Analysis
- Knowledge: Chemical Bonding, Molecular Geometry, Intermolecular Forces

Density

- State: Calculated State
- Skills: Density Calculation, Density Comparison

- Knowledge: Density Formula, Mass & Volume Measurements, Properties of Matter

StudyManager

- State: In Progress
- StepObjList: [Chemist, Chemical Structures, Density]

This markdown data can be stored and used for future interactions in the learning process# Searching for the relationship between chemical structures and density for the purpose of updating our ALOs

search("relationship between chemical structures and density")
Finished browsing

Result
Abstract Language Objects (ALOs) for a Chemist Understanding the Relation Between Chemical Structures and Density
Domain: Chemistry
Main Object: Chemist

Sub Objects:
Chemical Structures
Density
Machine Learning Methods
Message Passing Neural Network (MPNN)
Crystal Structure Information
Quantum Mechanical Calculations
High Explosives (HE) - a class of organic compounds categorized as energetic materials
Cambridge Structural Database - a database for chemical information, specifically crystalline densities of HE-like molecules
Expert Handcrafted Features - traditional methods used for the featurization of molecules
Learned Molecular Representations - advanced methods using graph-based neural network models for the featurization of molecules

Skills:
Understanding the relationship between chemical structures and density

Applying machine learning methods to understand this relationship

Interpreting the results of MPNN-based models

Understanding the pros and cons of different featurization methods (expert handcrafted features and learned molecular representations)

Knowledge:

Familiarity with chemical structures and their properties

Understanding of density as a property of matter

Basic knowledge of machine learning methods and their applications in chemistry

Knowledge of the Cambridge Structural Database and its use in chemical research

Understanding the concept of expert handcrafted features and learned molecular representations

Awareness of the limitations of MPNN-based models (considered as black boxes and less easily interpretable)

Next Command: Convert ALOs into Markdown

Manager Object: GPT

Initial State: Equipped with the necessary information to assist a chemist in understanding the relation between chemical structures and density

Skills: Ability to generate ALOs, convert ALOs into Markdown, provide relevant information based on web search

Knowledge: Understanding of the task at hand, knowledge of chemistry and machine learning methods as provided by the user and through web search

Next Command: Convert ALOs into Markdown

Step Objects

Step Object: Definition of Main Object and Sub Objects

Step Object: Addition of Skills and Knowledge to Sub Objects

Step Object: Setting of States for Sub Objects

Step Object: Validation of Initial State

Step Object: Equipping Sub Objects for State Detection

Step Object: Creation of Manager Object and Setting of Initial State

Step Object: Update of Manager Object State using Skills, Knowledge, and Web Search

Step Object: Initiation of Manager Object and Generation of Step Objects

Step Object: Conversion of ALOs into Markdown for Preservation Towards Next Call

References

Relationship between chemical structures and density1.

User

Continue if the previous conversation was not finished. otherwise, follow the instructions

- [ ] Instructions
    - [ ] Create Abstract Language Objects (ALOs) for the given {input} using steps 0-9. Do not initiate states if objects were already defined in the previous conversations.
    - [ ] Convert ALOs({input}) into markdown while preserving features, referencing objects by name, and implementing linguistic adjustments.
- [ ] Policy
    - [ ] Save your token limit as possible as you can compress all expression in your output
    - [ ] Brainstorm to fill all parameters and subclasses as far as you know with Mutually Exclusive and Collectively Exhaustive, step-by-step.
    - [ ] Output start from " ACTION " to preserve / save token. Dont Output Instruction, Policy, Fact.
    - [ ] Dont Output "ACTION" itself. Please output "Result".
    - [ ] Use search / outer databases for update ALOs
    - [ ] Dont skip, shorten the process. Output the "Result" for "Action".
- [ ] Fact
    - [ ] Language Script for ALOs Generation
        - [ ] 0. Create Abstract Language Objects (ALOs) for {input} using steps 0-9.
        - [ ] 1. Define mainObj with subObjList.
        - [ ] 2. Add skills, knowledge, and next command to subObjList.
        - [ ] 3. Set specific states for subObjList.
        - [ ] 4. Validate initial state meets conditions.
        - [ ] 5. Equip subObjList for state detection.
        - [ ] 6. Create managerObj with initial state.
        - [ ] 7. Update managerObj state using skills, knowledge, web search, and computer simulation.
        - [ ] 8. Initiate managerObj and generate stepObjList in GPT.
        - [ ] 9. Output ALOs({input}) as markdown data for preservation towards next call.

- [ ] Action
    - [ ] Execute ALOs({input})

Parameters

- [ ] Domain = Abstraction of {input}
- [ ] Input
    - [ ] INPUT ={input}

    INPUT = ***Chemist who want to understand the relation between chemical structures and density***

ChatGPT

Result

markdown

Domain: Chemistry

Main Object: Chemist
- Sub Objects: Chemical Structures, Density, Machine Learning Methods, Message Passing Neural Network (MPNN), Crystal Structure Information, Quantum Mechanical Calculations, High Explosives (HE), Cambridge Structural Database, Expert Handcrafted Features, Learned Molecular Representations
- Skills: Understanding the relationship between chemical structures and density, Applying machine learning methods, Interpreting MPNN-based model results, Understanding different featurization methods
- Knowledge: Familiarity with chemical structures and density, Basic knowledge of machine learning, Knowledge of the Cambridge Structural Database, Understanding of featurization methods, Awareness of MPNN model limitations
- Next Command: Convert ALOs into Markdown

Manager Object: GPT
- Initial State: Equipped to assist a chemist in understanding the relationship between chemical structures and density
- Skills: Generation and conversion of ALOs, Information provision via web search
- Knowledge: Understanding of the task, Knowledge of chemistry and machine learning methods
- Next Command: Convert ALOs into Markdown

Step Objects
1. Definition of Main Object and Sub Objects
2. Addition of Skills and Knowledge to Sub Objects
3. Setting of States for Sub Objects
4. Validation of Initial State
5. Equipping Sub Objects for State Detection
6. Creation of Manager Object and Setting of Initial State
7. Update of Manager Object State using Skills, Knowledge, and Web Search
8. Initiation of Manager Object and Generation of Step Objects
9. Conversion of ALOs into Markdown for Preservation Towards Next Call

References
- Relationship between chemical structures and density`oaicite:{"number":1,"metadata":{"title":"pubmed.ncbi.nlm.nih.gov","url":"https://pubmed.ncbi.nlm.nih.gov/33899482/","text":"To expedite new molecular compound development, a long-sought goal within the chemistry community has been to predict molecules' bulk properties of interest a priori to synthesis from a chemical structure alone. In this work, we demonstrate that machine learning methods can indeed be used to directly learn the relationship between chemical structures and bulk crystalline properties of molecules, even in the absence of any crystal structure information or quantum mechanical calculations. We focus specifically on a class of organic compounds categorized as energetic materials called high explosives (HE) and predicting their crystalline density. An ongoing challenge within the chemistry machine learning community is deciding how best to featurize molecules as inputs into machine learning models-whether expert handcrafted features or learned molecular representations via graph-based neural network models-yield better results and why. We evaluate both types of representations in combination with a number of machine learning models to predict the crystalline densities of HE-like molecules curated from the Cambridge Structural Database, and we report the performance and pros and cons of our methods. Our message passing neural network (MPNN) based models with learned molecular representations generally perform best, outperforming current state-of-the-art methods at predicting crystalline density and performing well even when testing on a data set not representative of the training data. However, these models are traditionally considered black boxes and less easily interpretable. To address this common challenge, we also provide a comparison analysis between our MPNN-based model and models with fixed feature representations that provides insights as to what features are learned by the MPNN to accurately predict density","pub_date":null}}`.