



産業材料研究開発における 大規模言語モデルの活用例

Use Cases for Large Language Models in Industrial Materials R&D

Michael Heiber, PhD

Enthought, Director, Professional Services & Customer Success,
Materials Informatics

Chemical Material Japan 2024, Tokyo Big Sight | 2024-11-21

AGENDA

産業材料研究開発における 大規模言語モデルの活用例

- 大規模言語モデルの概要
- 材料科学および化学研究開発におけるLLMの概要
- 具体的なプロジェクト提案
- LLM技術の展望

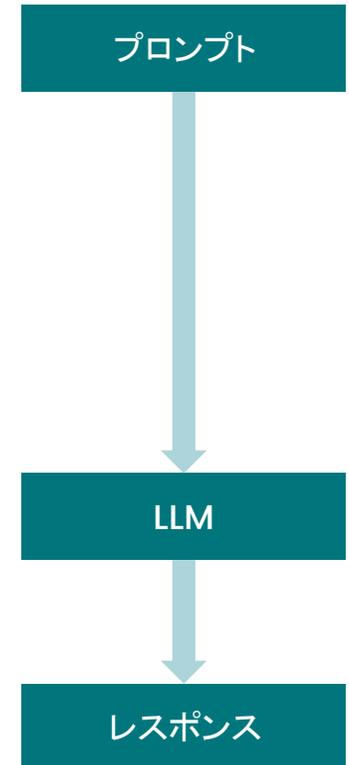


大規模言語モデルの概要

**Introduction to Large
Language Models**

大規模言語モデル(LLM)とは何か？

- 人間が言語を用いて互いにコミュニケーションを取る方法を模倣できる生成AIの基盤言語モデル
- 非常に大規模なテキストデータセットで訓練された、大規模な深層学習トランスフォーマーモデルに基づいている
- 画像、動画、音声用の他の基盤モデルと組み合わせることで、マルチモーダルな生成AIツールやチャットボットを作成できる
 - テキスト記述に基づいて画像を作成する
 - 画像に関する質問をする
 - 特定の声のスタイルを使用してテキストを音声に変換する、または音声を文字起こしする
- 重要な概念
 - **プロンプト** ▶ ユーザーが提供するテキスト文であり、モデルはこれに応答する
 - **コンテキストウィンドウ** ▶ モデルが応答(レスポンス)を生成する際に参照するすべてのテキストの範囲
 - **レスポンス** ▶ LLMによって返されるテキスト



簡単なプロンプト入力例

質問:「基盤モデルとは何ですか？」

LLM/ChatGPTにおける挙動:

- ChatGPTに「基盤モデルとは何ですか？」と尋ねる
- モデルは多くのウェブサイトや文書で訓練されているため、基盤モデルが何であるかを学習しており、それを説明することができる
- 短い要約説明、追加の詳細リスト、および例を含む形式で、見やすく整った回答を得ることができる
- モデルの応答は、通常、分かりやすい形式で情報を提供するよう設計されている

レスポンス

The screenshot shows the ChatGPT interface. At the top, it says 'ChatGPT' with a dropdown arrow and a 'Share' button. Below that, the word 'プロンプト' (Prompt) is written in orange. A text input box contains the question 'What is a foundational model?'. Below the input box, the response is displayed, starting with a circular icon and the text: 'A **foundational model** is a large, versatile AI model that serves as a base for a wide range of tasks across domains and applications. These models are trained on vast amounts of diverse data, often with hundreds of billions of parameters, enabling them to learn general patterns, language structures, and representations that can then be adapted to many specialized tasks. Foundational models are typically **pre-trained** on extensive data and then **fine-tuned** for specific tasks or domains, making them efficient and adaptable.' The response continues with 'Key features of foundational models include:' followed by a numbered list of four features: 1. **General-purpose capabilities:** They are trained to handle a wide array of language, vision, or multimodal tasks, making them versatile for downstream applications. 2. **Transferability:** Foundational models can be fine-tuned for various specialized tasks, allowing developers to adapt a single model for different use cases without building new models from scratch. 3. **Scalability:** These models can often improve performance simply by increasing data and computational resources, making them scalable to new levels of sophistication. 4. **Few-shot or zero-shot learning:** Some foundational models can perform certain tasks without task-specific training data, thanks to the general patterns learned in pre-training. Below the list, there is a section titled 'Examples' with the text 'Popular examples of foundational models include:' followed by a bullet point: '• **GPT models** (like GPT-4) for language tasks'. At the bottom of the interface, there is a 'Message ChatGPT' button and an upward arrow icon.

簡単なプロンプト入力例

コンテキストウィンドウ

質問:「それは他の機械学習モデルとどのように異なりますか？」

LLM/ChatGPTにおける挙動:

- フォローアップの質問は、以前の質問や応答を参照することができ、モデルは現在の議論の文脈を理解する
- 「それは他の機械学習モデルとどのように異なりますか？」と質問すると、チャットボットは文脈に基づき「それ」が基盤モデルを指していることを認識する

レスポンス

The screenshot shows a ChatGPT interface. At the top, it says 'ChatGPT' with a dropdown arrow and a 'Share' button. The main content area is enclosed in an orange border and contains the following text:

without building new models from scratch.

3. **Scalability:** These models can often improve performance simply by increasing data and computational resources, making them scalable to new levels of sophistication.
4. **Few-shot or zero-shot learning:** Some foundational models can perform certain tasks without task-specific training data, thanks to the general patterns learned in pre-training.

Examples

Popular examples of foundational models include:

- GPT models (like GPT-4) for language tasks
- BERT and its variants for NLP tasks
- CLIP for image and text understanding
- DALL-E for text-to-image generation

Foundational models are shaping many current AI applications, especially in natural language processing, computer vision, and multimodal AI (handling multiple types of data like text and images together).

Below this, there is a 'プロンプト' (Prompt) section with a text input field containing: 'How are they different from other machine learning models?'

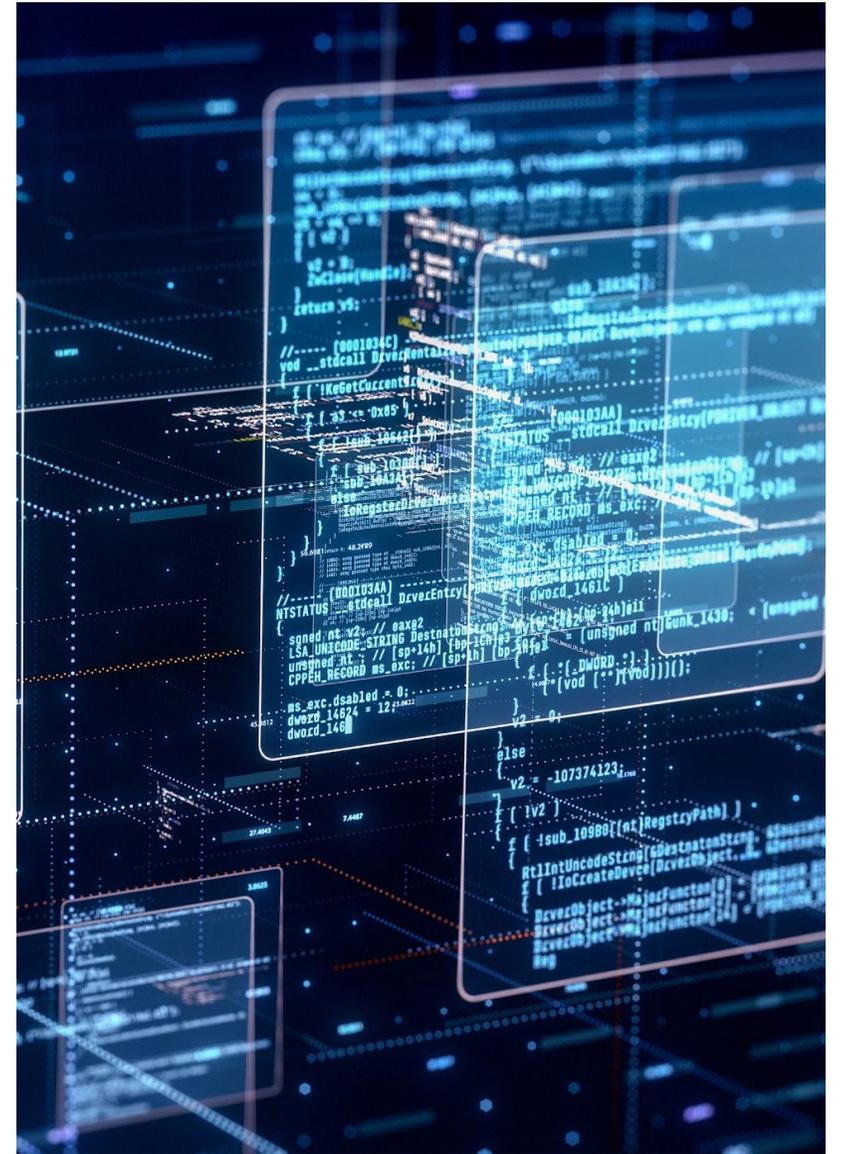
The response section, also enclosed in an orange border, shows a circular icon followed by the text: 'Foundational models differ from traditional machine learning models in several ways, especially in terms of scale, versatility, and training approach. Here are some key distinctions:'

1. **Scale and Training Data** ↓

At the bottom, there is a 'Message ChatGPT' button and a scroll-up arrow.

汎用的なモデルの能力の限界

- 最先端のLLMは、すでに理解が進み、情報が整理されているテーマであれば、非常にわかりやすく説明することができる
 - こうしたテーマに関しては膨大な量のテキストで訓練されている
- より専門的で複雑な知識が必要となるニッチな分野については、LLMは説明の精度が落ちてしまう傾向がある。
 - そうしたテーマでは訓練データが少なく、場合によっては矛盾する情報や説明が含まれることがある
- 重要な概念:
 - ハルシネーション: 一見もっともらしく見えるが、実際には重大な事実誤認や論理的な誤りを含むレスポンス



研究開発での利用における プロンプト作成の難しさ

デフォルトの gpt-4o モデル

- 有機太陽電池の物理的メカニズムやその材料の詳細な性能指標に関する難解な科学的質問
- このトピックに関する文書はおそらく400件程度
- これらの文書には、異なる用語、異なる方法から得られた事実や情報、矛盾する説明が含まれていることが多い
- 専門家によるレスポンスの評価 - **6/10**
 - 正確で有用な情報もあるが、重要なメカニズムについての言及が欠けており、説明が浅い
 - 例として挙げられている材料や指標が必ずしも正確ではない
 - 実在しない参考文献が多く挙げられている

The screenshot shows a chat window with the following content:

Chat [Clear] [Code] [Compare] [History] [Settings]

System instructions [Star] [Share]

User
What is reduced Langevin recombination in organic solar cells, and what are the proposed mechanisms that cause it? Provide references to recent academic review papers for further reading about the mechanisms. Then provide a few examples of active layer materials that have reduced Langevin recombination and their associated reduction factors as well as references to publications that report these values.

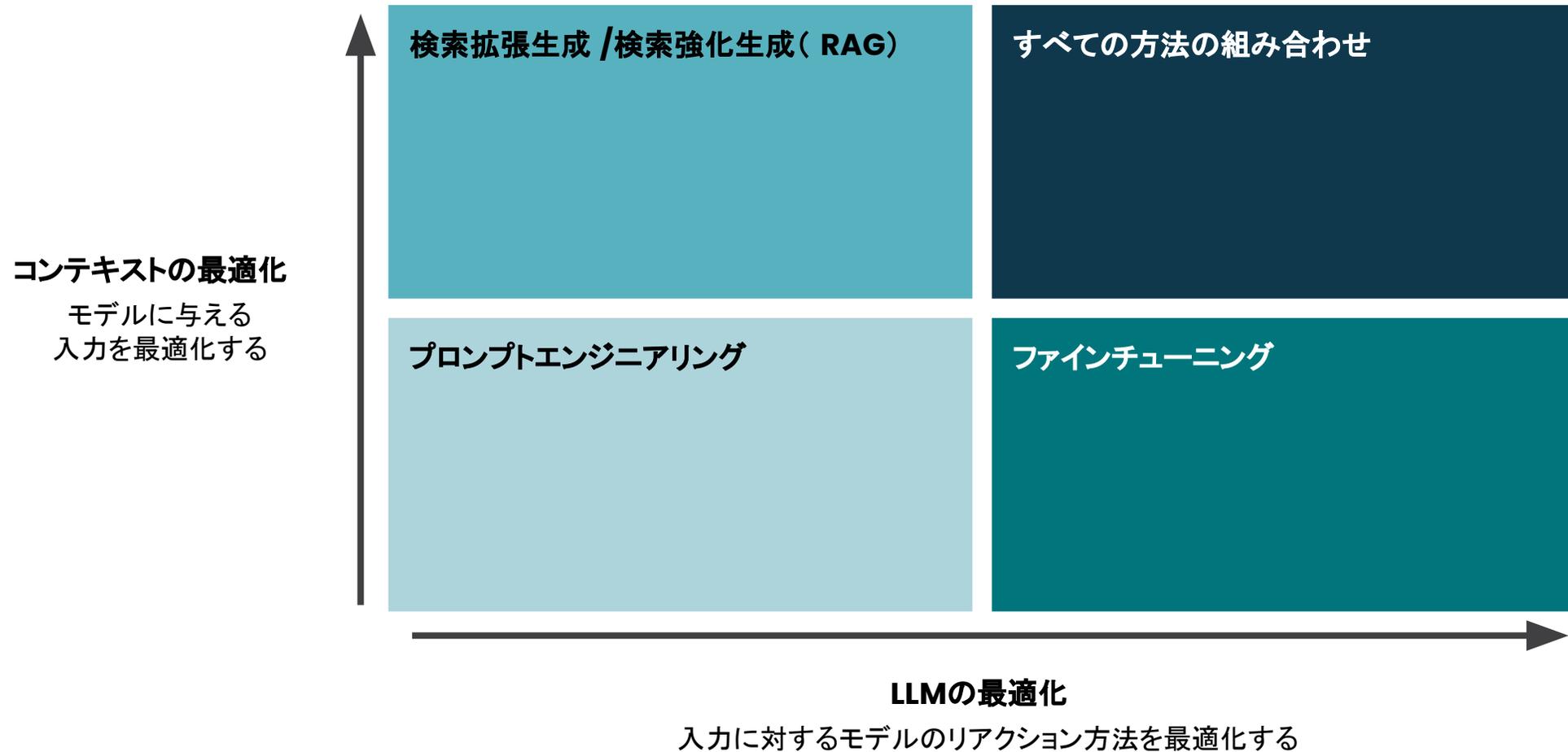
Assistant
Reduced Langevin recombination in organic solar cells refers to a phenomenon where the recombination rate of charge carriers (electrons and holes) is lower than predicted by the classical Langevin model. The Langevin model assumes that charge carrier recombination in disordered organic semiconductors is primarily dictated by diffusion. However, experimental observations often report recombination rates significantly lower than the Langevin prediction, which is beneficial for improving the efficiency of organic solar cells.

Proposed Mechanisms for Reduced Langevin Recombination:

- Phase Separation and Nanomorphology:**
- In organic solar cells, the morphology of the active layer can lead to spatial separation of electrons and holes into different material phases. When phase separation is optimized, it can reduce the probability of electrons and holes meeting and recombining.
- Carrier Mobility Imbalance:**
- A significant difference in mobility between electrons and holes can result in the more mobile carriers quickly being extracted, leading to a decrease in the rate of recombination.

Enter user message... [User] [Share] [Run] [Ctrl+Enter]

特定のタスクに LLM を最適化する方法はいくつも存在する

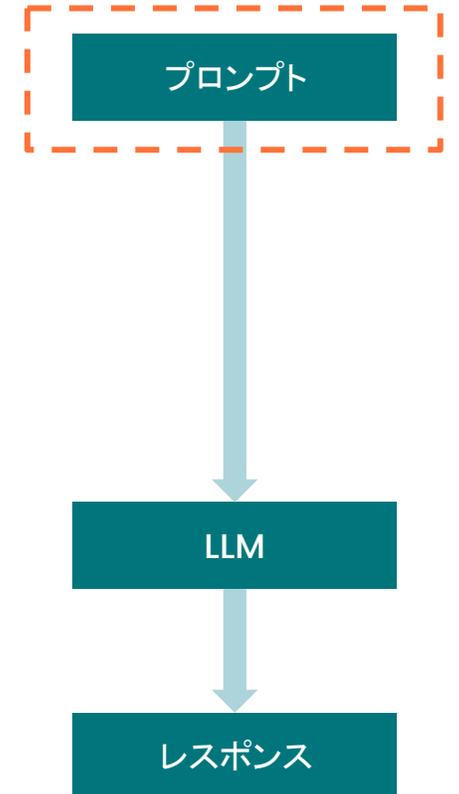


特定のタスク向けに LLMを最適化する方法

プロンプトエンジニアリング

- プロンプトとなるテキストを最適化し、より良いレスポンスを得る
- モデルがプロンプトにどのように応答すべきか(レスポンスの複雑さ、長さ、形式など)について明確な指示を書く
- 複雑なタスクをシンプルなステップに分解するようモデルに指示する
- モデルに自らのレスポンスをチェックするよう指示する
- プロンプト内でいくつかの例を提供する(コンテキスト学習/少数ショット学習)

プロンプト
エンジニアリング



研究開発での利用における プロンプト作成の難しさ

デフォルトの gpt-4oモデル + プロンプトエンジニアリング

- 有機太陽電池の物理的メカニズムやその材料の詳細な性能指標に関する同様の難解な科学的質問
- 「システム指示」を使用してプロンプトエンジニアリングを実装する
 - ターゲットとなる読者、応答形式、思考のステップを指定し、例を提供するなど
- 専門家によるレスポンスの評価 - 7/10
 - 説明はより深くなったが、依然として重要なメカニズムが欠けている
 - 例として挙げられている材料やその指標が依然として必ずしも正確ではない
 - 多くの参考文献が実在しない

The screenshot shows a chat interface with the following content:

Chat [Clear] [Code] [Compare] [History] [Settings]

System instructions [Generate]

Answer chemistry and materials science questions for a PhD-level scientist, providing detailed explanations supported by at least 3 references to relevant research papers.

Make sure your response is logically structured, with clear reasoning leading to conclusions. For each reference, include the key points from the paper that relate to the question being asked. Extract critical information from the paper, such as experimental findings, methodologies, or theoretical discussions, and explain how it supports your answer.

Steps

1. **Understand the Question:** Carefully read and understand the question. Identify the main components, such as the specific chemical concept, material property, or experimental process involved.
2. **Background Explanation:** Provide a brief background or explanation of the primary concept before diving into the specific question. This helps in making the answer more comprehensive and easy to understand.
3. **Detailed Answer:** Expand upon the given subject with the relevant information gathered from your knowledge base. Address all relevant aspects, considering different possibilities or perspectives if applicable.
4. **Reasoning Steps:** Ensure that your reasoning is logically explained before finalizing an answer. Step-by-step reasoning is crucial to arrive at accurate conclusions.
5. **Provide References:** Cite at least three papers or sources where the user can verify the information. For each reference, include a brief discussion of key findings or theoretical points and explain how these relate to the answer given. Prioritize foundational or highly-cited research papers that are directly relevant to the question.

User

What is reduced Langevin recombination in organic solar cells, and what are the proposed mechanisms that cause it? Provide references to recent academic review papers for further reading about the mechanism. Then provide a few examples of active layer materials that have reduced Langevin recombination and their associated reduction

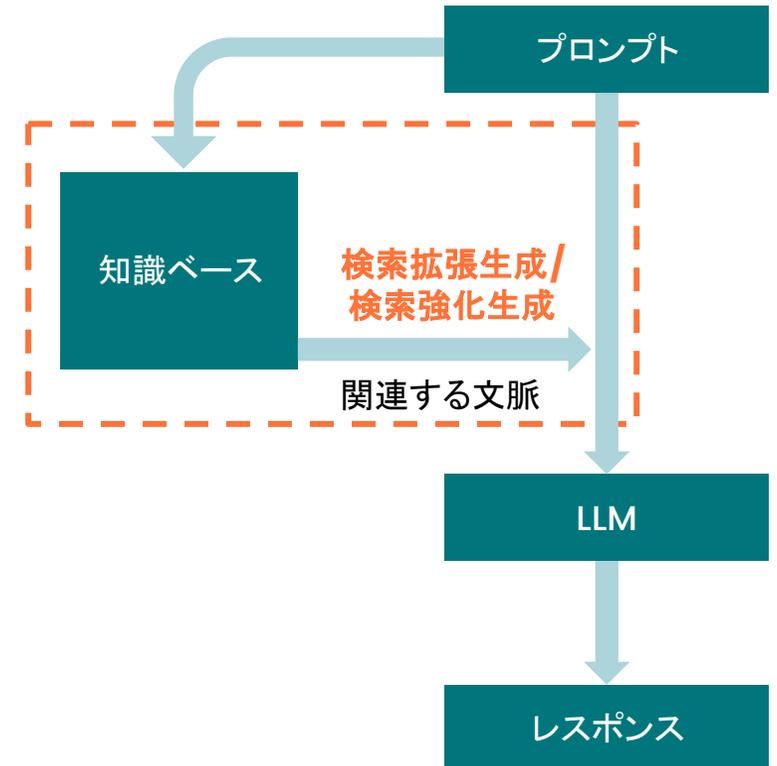
Enter user message...

User [Run Ctrl+R]

特定のタスク向けに LLMを最適化する方法

検索拡張生成 / 検索強化生成 (RAG)

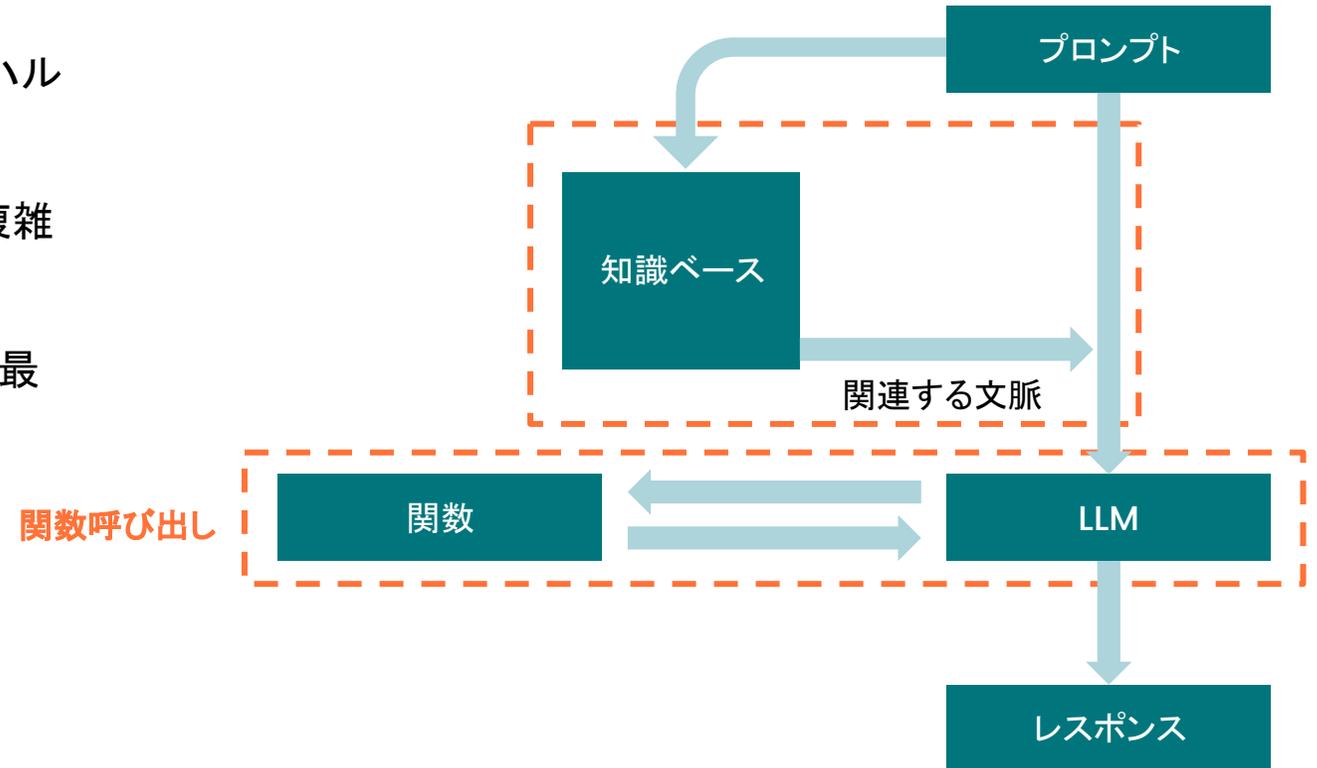
- 大規模な知識ベースを検索し、関連する情報を抽出してプロンプトに適切な文脈を追加する
- 例: 多くの関連する研究論文の内容を質問とともに明示的に文脈として含める
- 重要な概念:
 - **コンテキストウィンドウの長さ**: 文脈として提供できるテキストの量はモデルによって制限されている



特定のタスク向けに LLMを最適化する方法

関数呼び出し

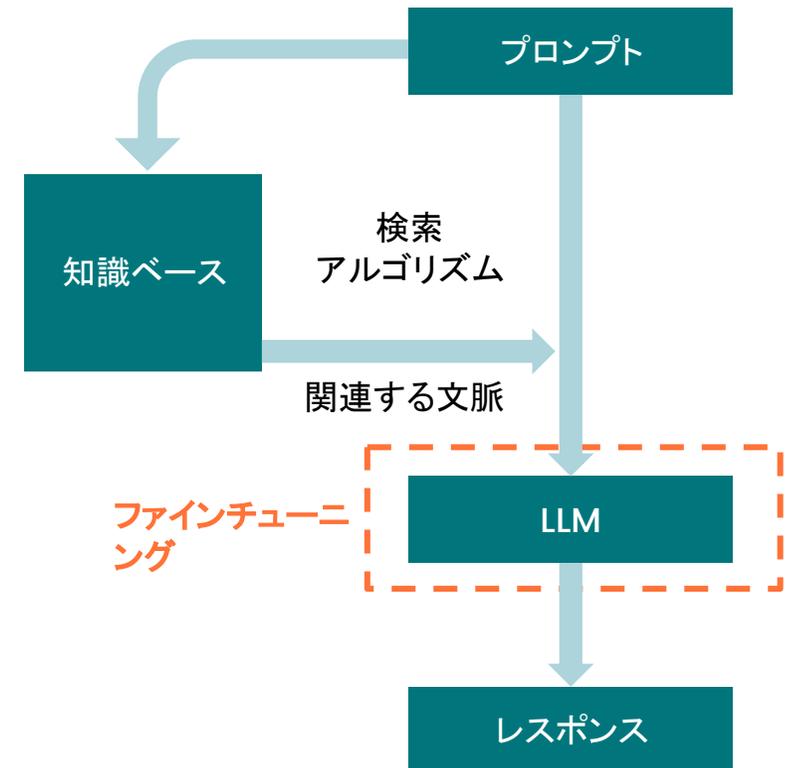
- 特定の関数に困難なタスクを実行させることで、ハルシネーションを減らす
- 関数の例: データベースからデータを取得する、複雑な計算を実行する
- LLMに、いつ関数を呼び出すべきか、どの関数が最適かを教える必要がある



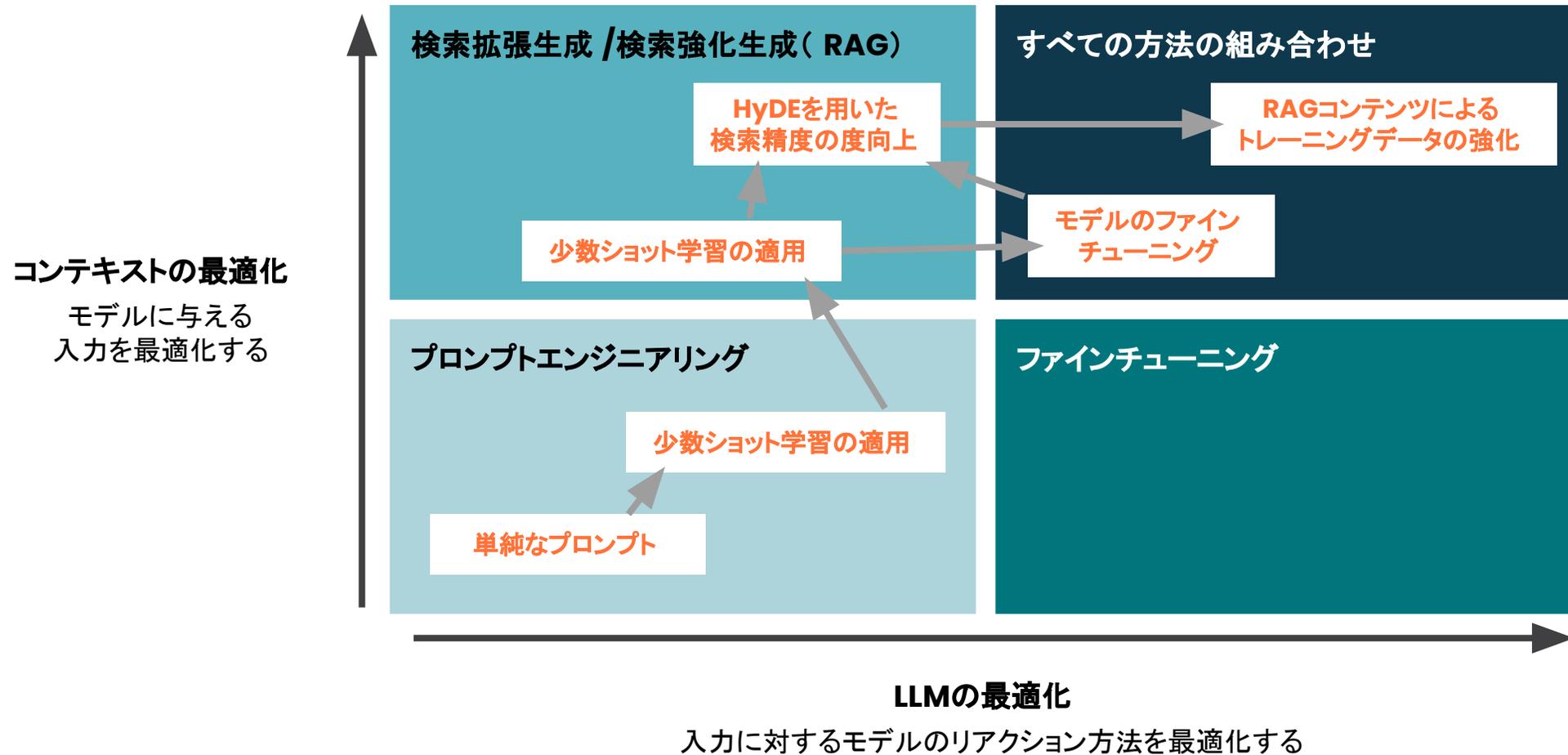
特定のタスク向けに LLMを最適化する方法

ファインチューニング

- 新しい訓練データを提供してモデル出力精度を調整
- 最適化させたいプロンプト(質問)を50~100個ほど用意し、それぞれに対して理想的な回答例を合わせて入力する
- 特定の形式や表現で回答させたい場合に有効
- コンテキストを減らすことで運用コストを削減できる



ベストプラクティスを参考に最適な方法を見つけ出す

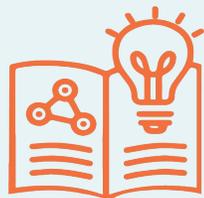




**材料科学・化学分野の研究開発に
おける大規模言語モデル(LLM)**

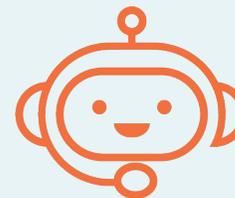
**LLMs in Materials Science
and Chemistry R&D**

LLM活用の代表的なユースケース



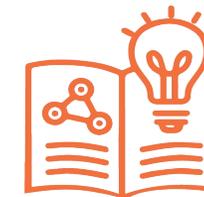
知識の抽出と要約

**KNOWLEDGE EXTRACTION
& SUMMARIZATION**

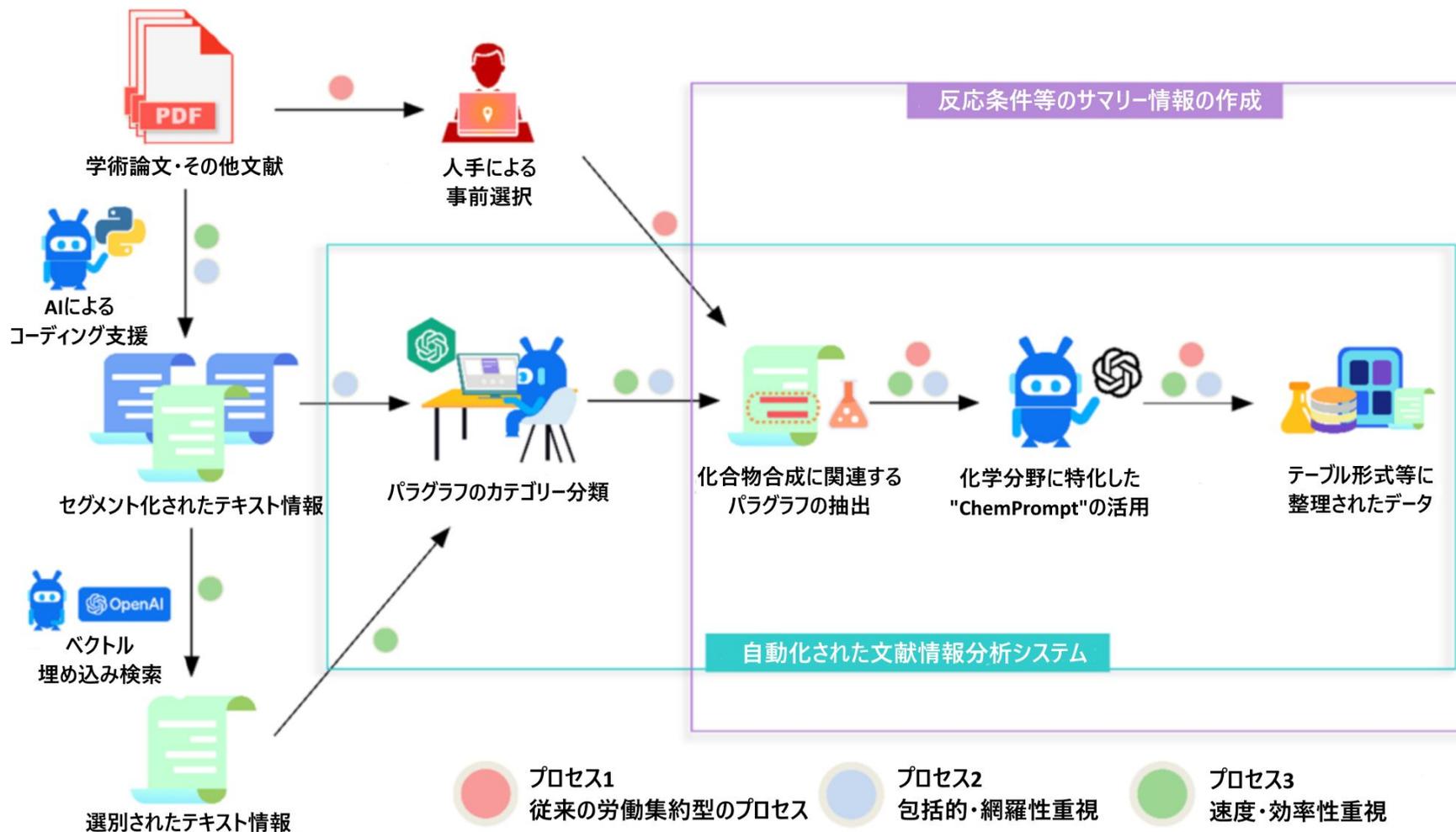


ラボアシスタント&自動化

**LAB ASSISTANTS &
AUTOMATION**



知識の抽出と要約



ラボアシスタント / 自動化



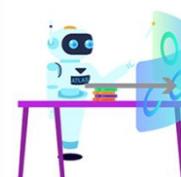
- 研究室のアシスタント
 - LLMを他のツール(他のLLMも含む)に接続して活用する
- 「最低ラインを引き上げる」典型的な例として、未経験の研究者が中級レベルのタスクを実行できるようにする
 - 複数のタスクをアウトソースする
 - ・ 文献レビュー
 - ・ 反応予測
 - ・ ベイズ最適化
 - ・ 実験計画
 - ・ 自動化実験の実行
- どの程度まで実装するかについては多くの選択肢がある
 - シンプルなアシスタントから完全なラボ自動化まで



分析



化学合成アシスタント



実験計画・ガイダンス



実験機器の制御



文献要約・情報収集



ベイズ最適化アルゴリズム



ロボットプラットフォーム

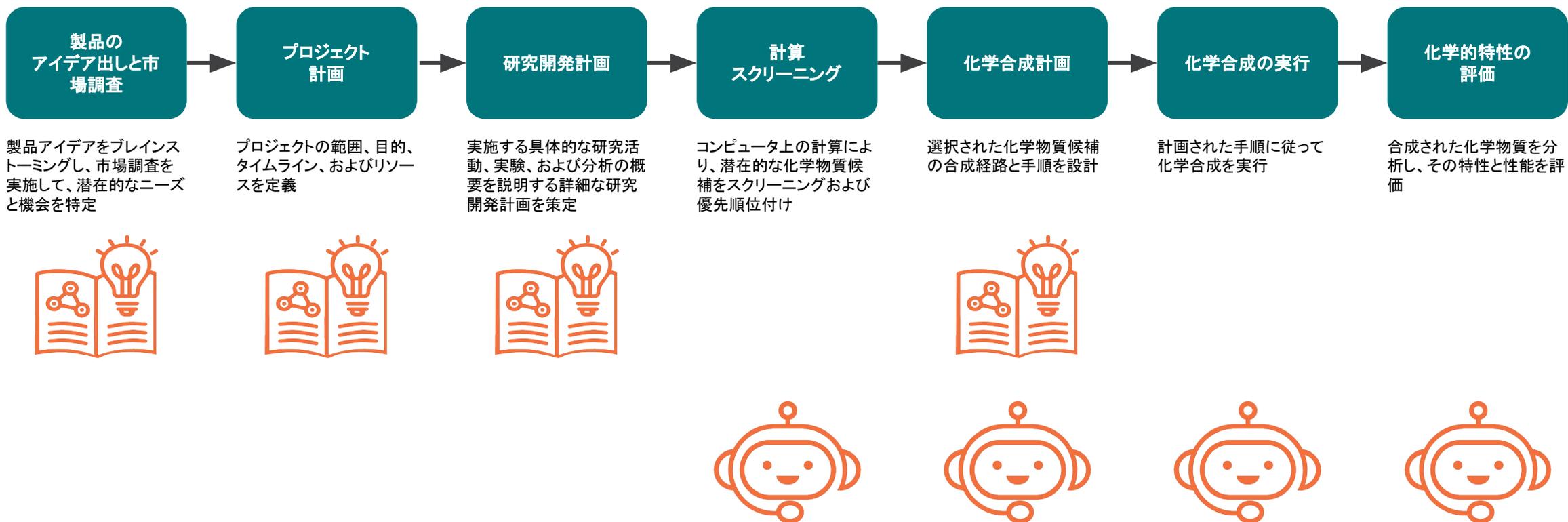
ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs, ACS Central Science, 9(11), 2161-2170.をもとに作成



具体的なプロジェクト提案

**Specific Project
Recommendations**

化学製品の典型的な研究開発ワークフローにおける LLMの適用可能性

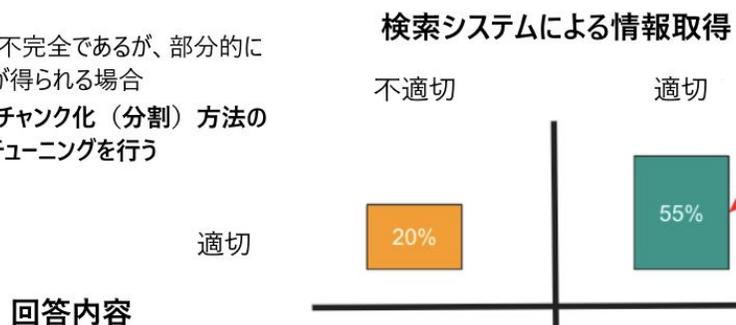




科学文書のための質問 / 回答システム

- 社内科学文書に対するRAG(検索拡張生成)システムの構築
 - Q&Aインターフェース
 - 検索結果を最適化するための再ランク付け
 - 初期検証のための質問/文書データの作成
- 検索アルゴリズムの最適化は極めて重要であり、用途に応じて異なる最適化が求められる

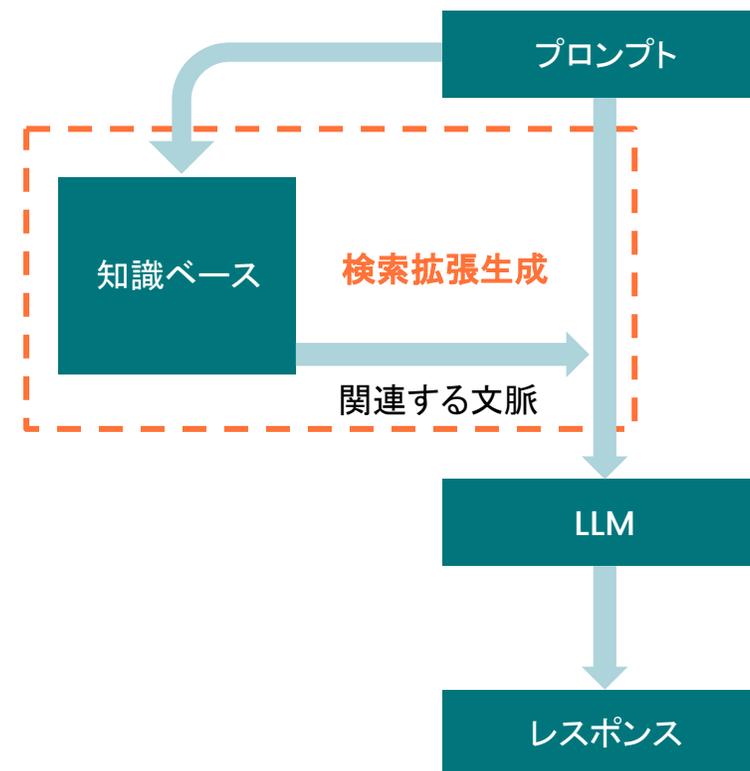
情報取得が不完全であるが、部分的に正しい回答が得られる場合
→テキストのチャンク化(分割)方法の調整によるチューニングを行う



適切な情報取得によって正しい回答が得られるケース

情報取得が適切であるにもかかわらず、誤った回答が出力される場合
→プロンプトまたはモデルパラメータのチューニングを行う

情報取得が適切にできていない場合
→検索そのもののチューニングを行う





科学文書のための質問 / 回答システム

RAG LLMシステム

市場調査

- 未充足の市場は存在するか？
- 先行特許は存在するか？
- ビジネスと製造における制約は何か？

内部文書、特許、競合他社のウェブサイトを検索対象とする

調査の主眼となる項目群から回答を生成

化学合成計画

- どの試薬を購入する必要があるか？
- どの反応が実施できるか、スケールアップは可能か？
- どの分野の専門知識が必要で、SME（専門家）は利用可能か？

Google Scholar、特許、内部在庫、過去の合成報告書を検索対象とする

内部ファイルのメタデータ、分子画像などから回答を生成

ユーザーサービスリクエスト

- 過去に同様の問題が生じたことはあるか？
- ユーザーに次のステップとして何を提案すべきか？
- 問題解決にはどれくらいの時間がかかるか？

過去のメール通信、内部ナレッジベース、問題追跡システムを検索対象とする

化学品/製品名、問題の説明などから回答を生成

用途に応じて検索を最適化すべき

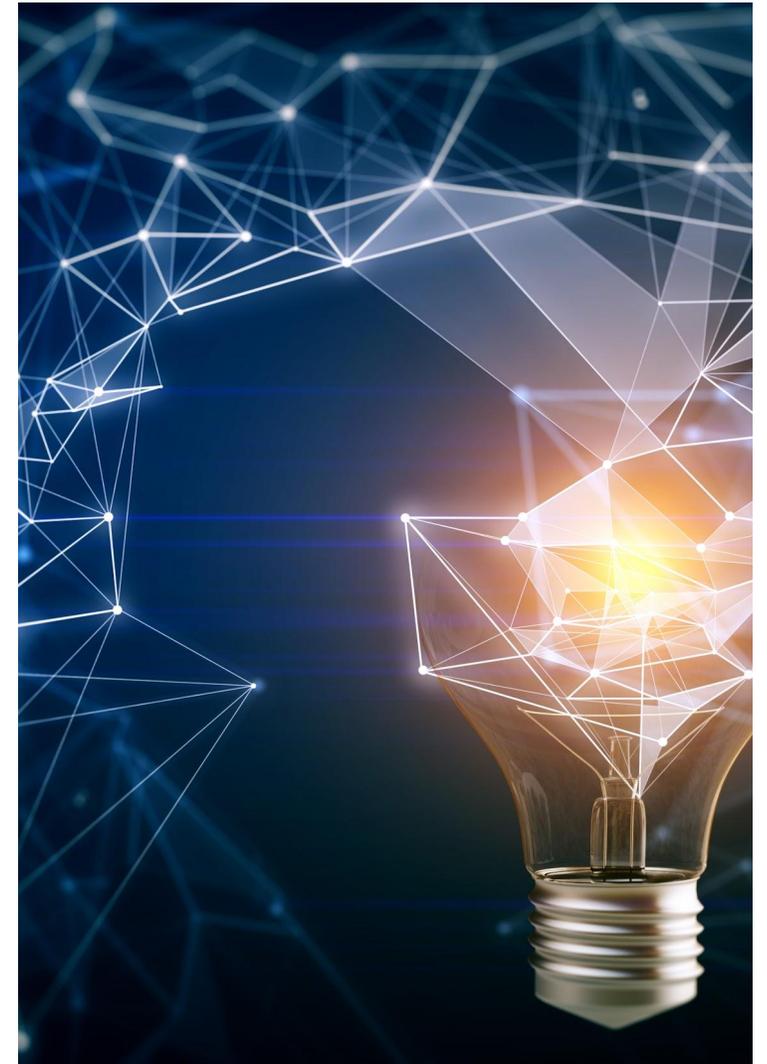
分子設計・分析アシスタント



- 分子を生成および分析できるLLMベースのシステムを構築する
 - LLMを特定のタスクを実行するツールに接続する
- 多くのプラグインを組み込む可能性
 - 材料科学の特定のニッチ分野に特化した文献検索および実験計画用のRAGエキスパートアシスタント
 - 化学在庫管理
 - 自動合成アシスタント
 - 計算化学シミュレーション、コンピューター支援工学シミュレーション、または重要な特性を予測する機械学習モデル

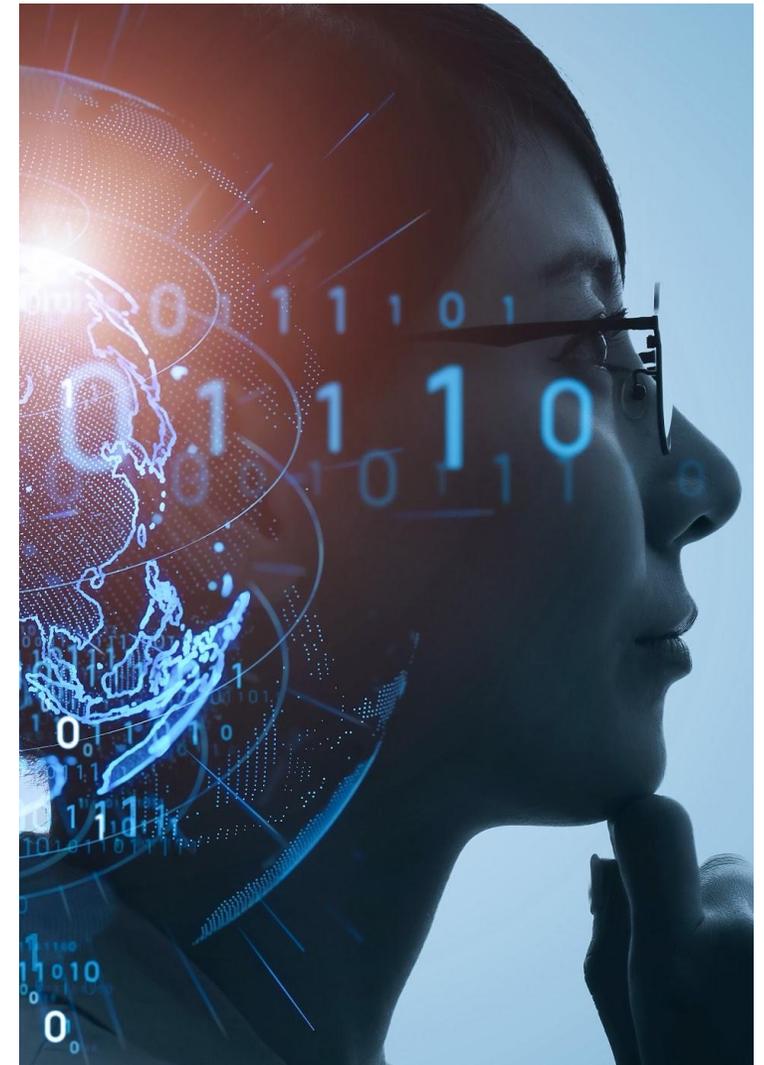
科学チャットボットの利点

- 研究開発体制が整っていない分野において、予備的な研究開発プロジェクトを支援するのに役立つ
- 経験の浅い研究者でも、熟練研究者の時間を奪うことなく、知識や学習効率を高め、迅速に生産性を向上させることができる
- オペレーション連携の柔軟性を大きく高めることができる



科学チャットボットの欠点

- ミッションクリティカルな技術的判断に必要な、信頼性の高い情報を常に提供できるとは限らない
 - 新米エンジニアと同様に、その成果は経験豊富な担当者によって検証される必要がある
- 専門家が行うような、特殊で複雑な研究開発タスクにおいては、チャットインターフェースによる操作が煩わしく感じられる可能性がある
 - 専門家の判断で特定のボタンをクリックする方が、チャット入力や音声コマンドによる指示よりも、誤解のリスクがなく、迅速かつ簡単
- あらゆる研究開発の状況に対応できる AIチャットボットを開発するには、従来の方法よりもはるかに高いコストがかかる可能性がある
- チャットボットの機能や限界、使用方法をすべて把握することは容易ではない
- チャットボットの出力は確定的なものではなく、同じ入力でも常に同じ結果が得られるとは限らないため、トレーサビリティの問題が生じる可能性がある



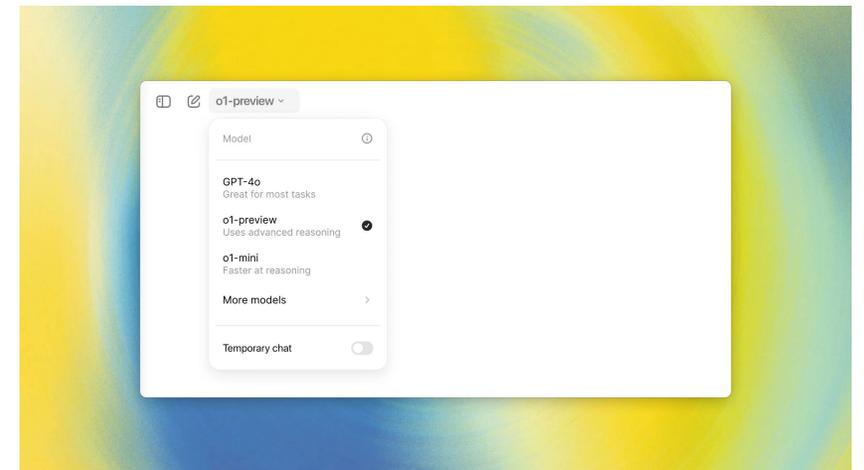


LLMテクノロジーの展望

LLM Technology Outlook

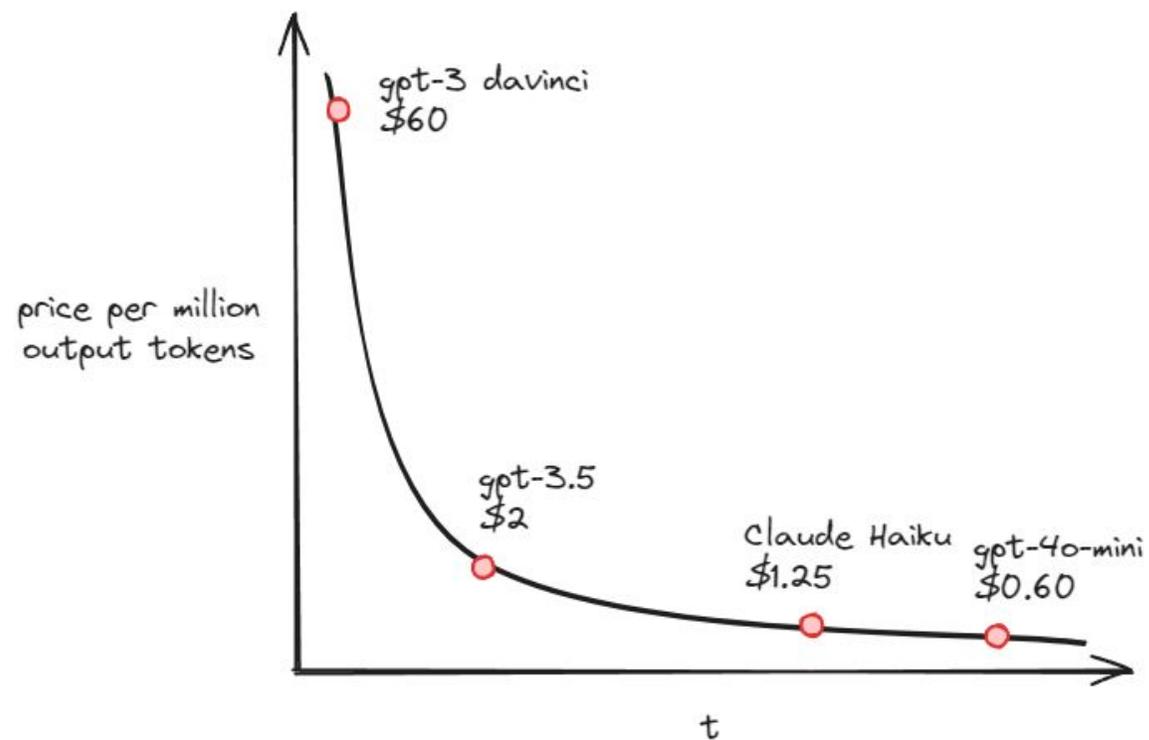
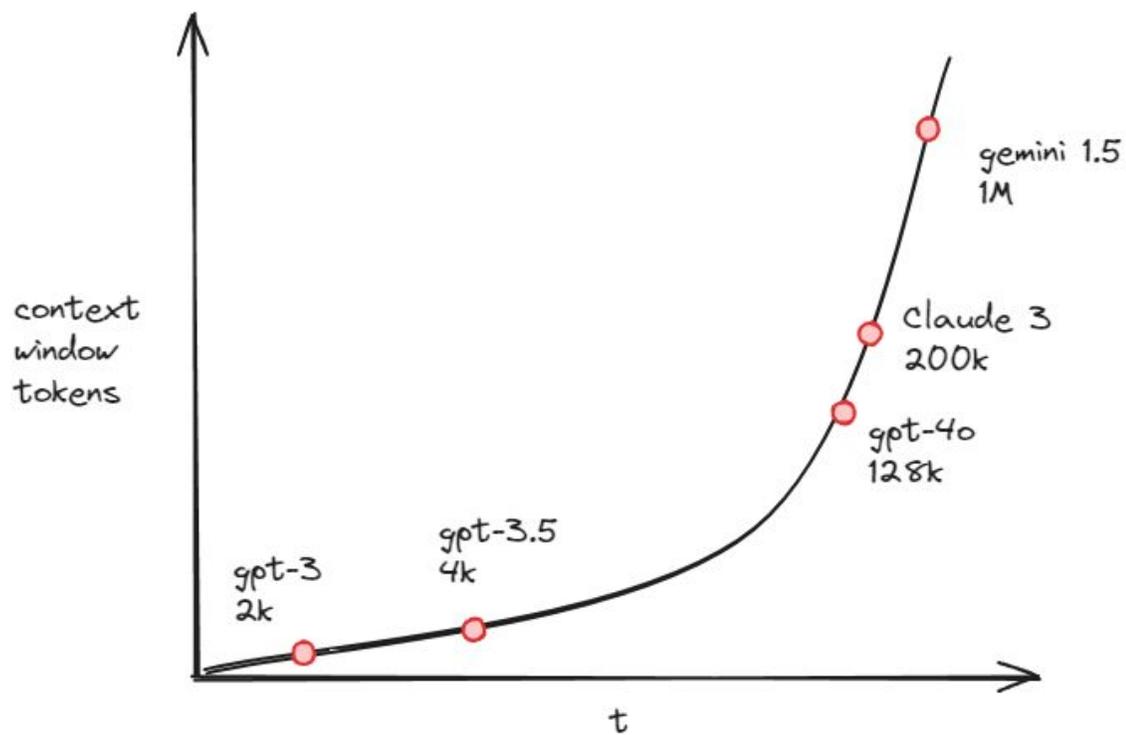
困難な技術的タスクの継続的な最適化

- 2024年9月、OpenAIは連鎖的思考(chain of thought)と強化学習技術を使用して、複雑な推論タスクでより優れた性能を発揮する最適化推論モデル(o1)を導入した
- 推論モデルo1は、コーディング、データ分析、数学、生物学、化学、物理学のベンチマークテストにおいて、4oと比較して著しい性能向上を示した



パフォーマンスとコストのトレンド

テクノロジーは今も急速に進化している





本プレゼンテーションのまとめ
Summary

キーメッセージ

1. 大規模言語モデル(LLM)は、多様なタスクをこなす基盤となる言語モデルです。
2. LLMは、材料科学や化学といった専門性の高い分野でも、その性能を最大限に引き出すための最適化が可能です。
3. 学術機関や政府の研究者は、LLMを知識の抽出や要約、情報収集や分析、研究補助などに活用できることを示しています。
4. 適切に設計・最適化されたLLMは、研究開発の課題解決に役立ち、イノベーションを加速させるための実用的なツールとなります。
5. LLMの技術は、性能が飛躍的に向上し、コストが低下するなど、急速に進化しています。
6. 研究開発におけるLLMの活用方法をまだ見出せていないなら、今すぐその可能性を探るべきです。



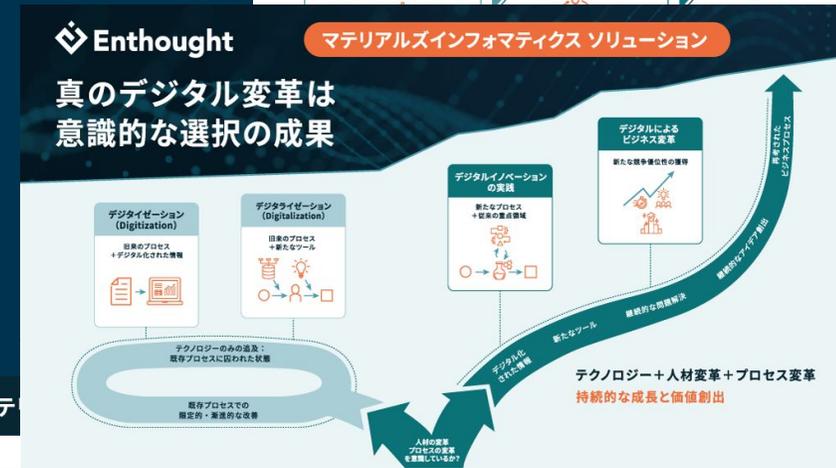
次世代材料開発に関する書籍を含む
賞品を獲得するチャンス！

(価値：¥71,500)

Enter to win prizes, including a
book on Next-generation
materials development!

(Valued at ¥71,500)

#D-16 にぜひお越
しく下さい





Enthought