

NTT DATA

# ChatGPTなど、 大規模言語モデルのビジネス活用について

株式会社NTTデータ数理システム

© 2024 NTT DATA Mathematical Systems Inc.



# 会社概要

会社名	株式会社NTTデータ数理システム
所在地	東京都新宿区信濃町35 信濃町煉瓦館1階
資本金	5,600万円 (NTTデータ100%出資)
従業員数	約120名 (内80%が技術者)

## 沿革

- 1982年4月 株式会社 数理システム 設立
- 2012年2月 NTTデータグループ入り
- 2013年9月 株式会社NTTデータ数理システム に社名変更

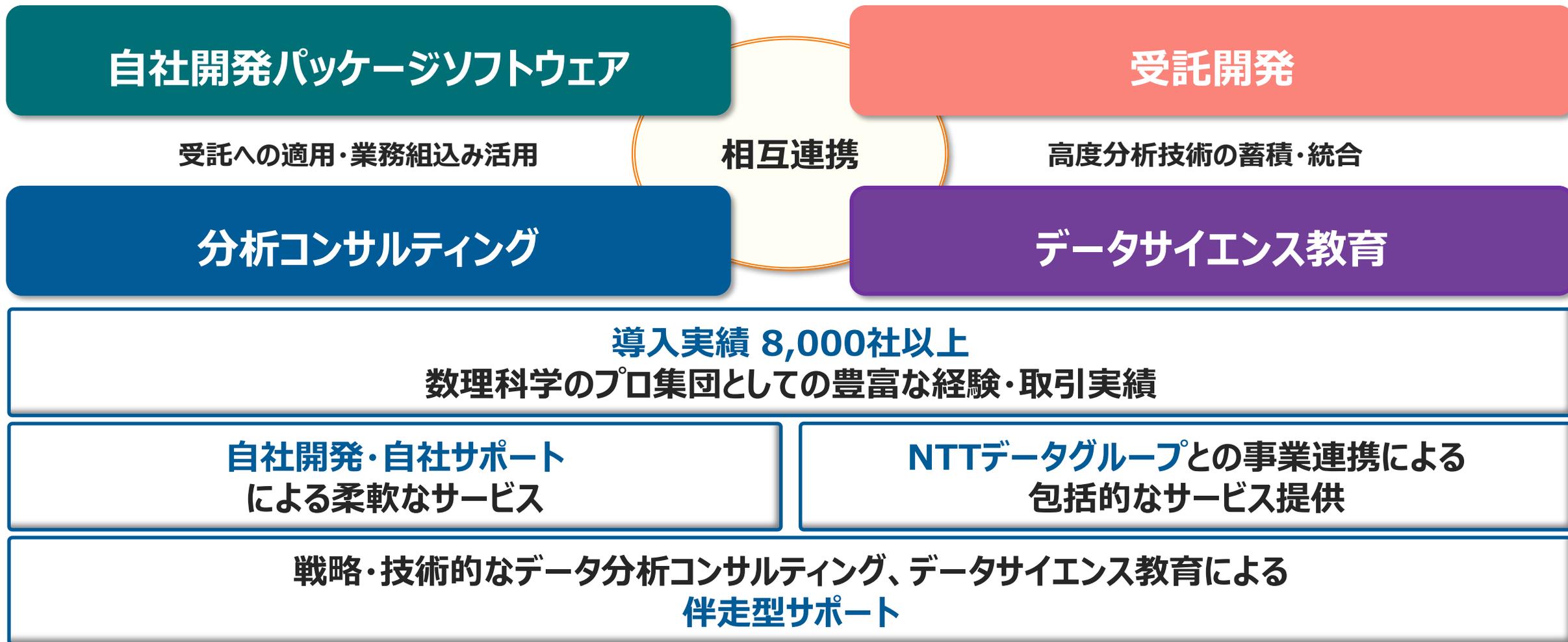
## Mission

数理学とコンピュータサイエンスにより  
社会のあらゆる分野の問題を解決する



# 事業と強み

## 数理学とコンピューターサイエンス



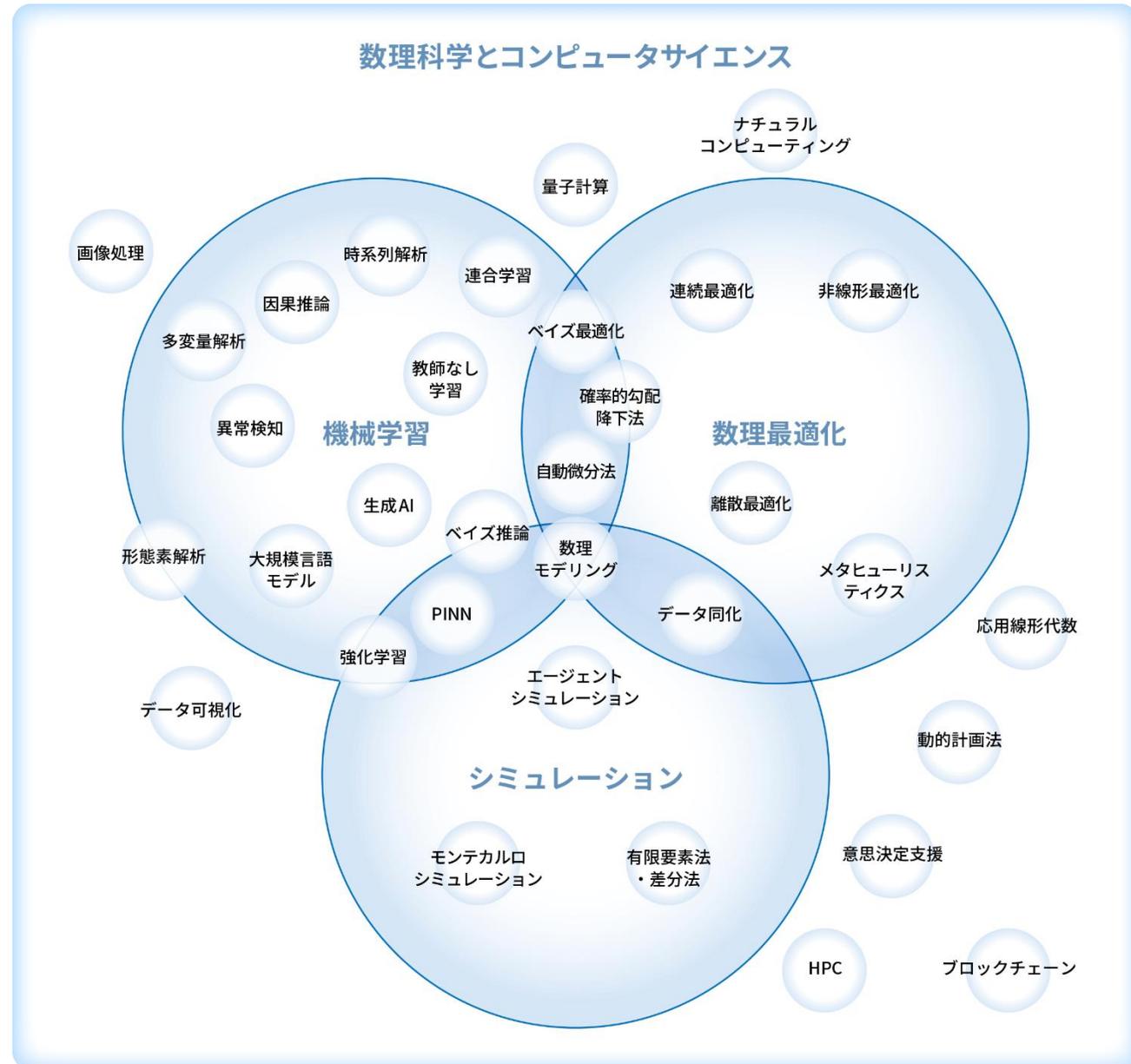
# 事業内容と対象領域

## 主な事業内容

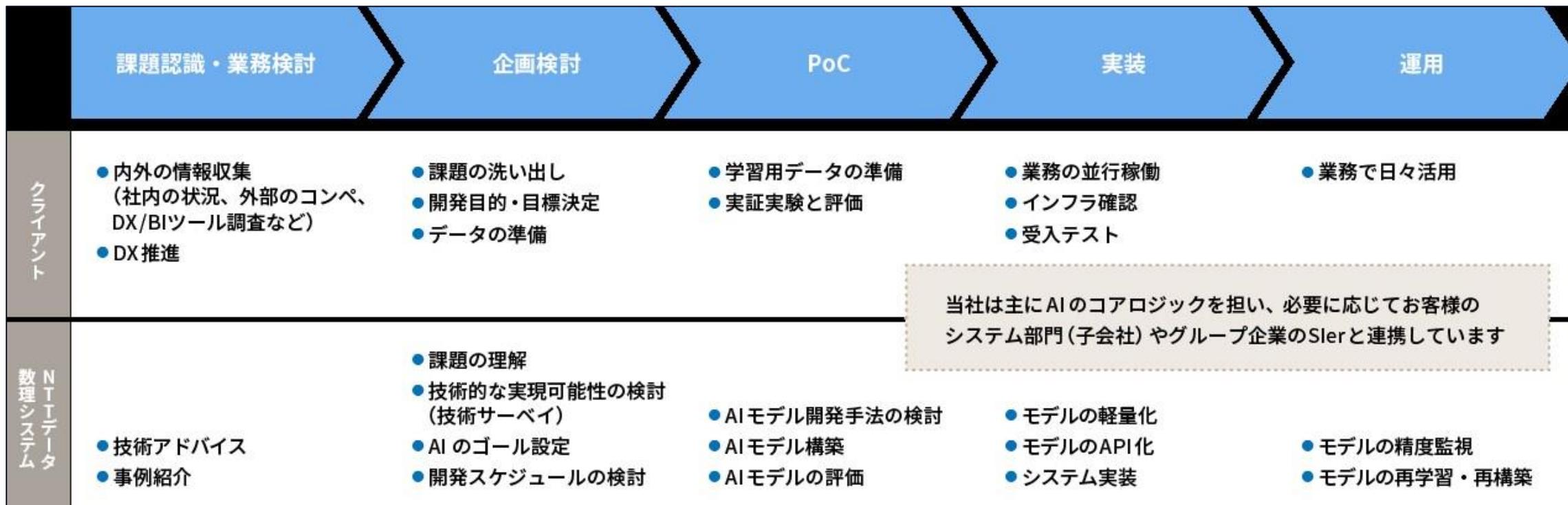
- パッケージソフトウェア開発・販売
- ソフトウェアのカスタマイズ・受託開発
- データの分析技法や開発コンサルティング
- データサイエンティスト育成

## 対象領域

- AI、機械学習、深層学習、データサイエンス
- 統計解析、データマイニング、テキストマイニング
- 数理計画、数理最適化
- シミュレーション（離散、連続、エージェント）
- 知識データベース、言語処理、科学技術計算等



# NTTデータ数理システムのビジネスイメージ



## 数理システム保有の基礎技術

機械学習

数理最適化

データ分析

シミュレーション

### ロジックをお客様がご所望の形式で提供

- 実行ライブラリ・モジュールとして
- GUIを伴った独立したアプリケーションとして
- R, Pythonを援用した実行スクリプトとして



## 現実課題のソリューションへ



# 本セミナーについて

ChatGPTをはじめとする大規模言語モデル（LLM）技術の基礎知識と、  
ビジネスに活用するためのコツについて、当社事例を交えてお伝えします

## 本日のトピック

### 0. LLMの基礎知識

1. LLM活用のコツ1.  
プロンプトエンジニアリングによる「深掘り」をしよう

2. LLM活用のコツ2.  
LLMが知らない知識は、外から補って使おう

3. LLM活用のコツ3.  
データ分析技術と組み合わせて使おう

4. もっとLLMを活用する  
LLMエージェント技術によるアイデア創発

アイデア次第で色々な  
ビジネスに活用可能！

われわれと一緒に新たな  
活用事例を模索しませんか？

# 当社の目指すLLM活用の姿

数理学とコンピュータサイエンス × LLM で新たな価値を創出



×

LLLM

(をはじめとする生成AI)

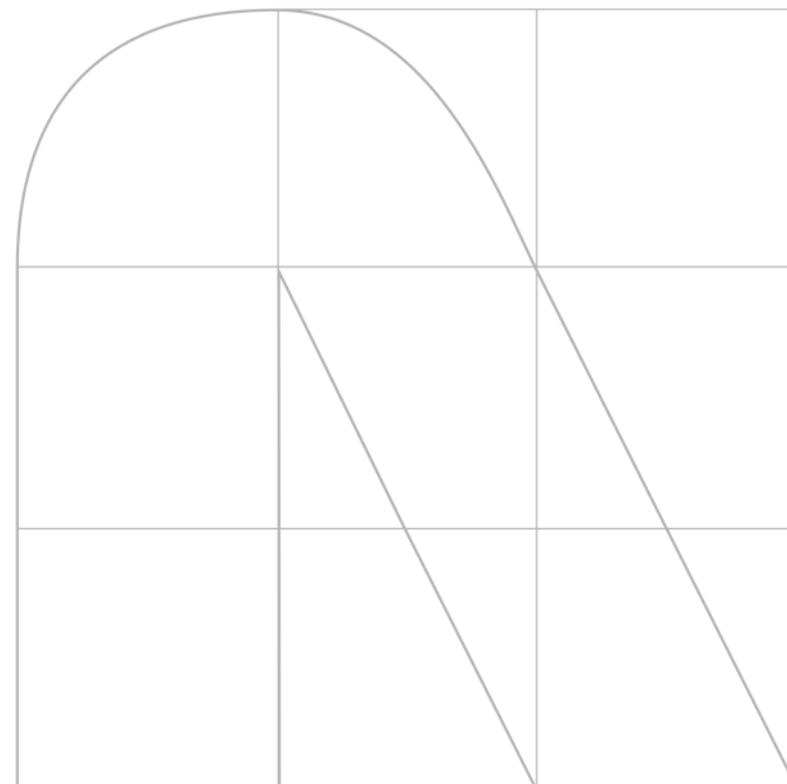
大量のデータからの傾向把握  
数学的裏付けのある問題解決

言葉による人間らしい思考力  
高度なコミュニケーション能力

LLMだけでない幅広い技術で、お客様の業務に貢献するAI構築を支援します

# 0

## LLMの基礎知識



# ChatGPT (OpenAI社) の台頭

## 従来型AIとは別格に賢いチャットAI → 2022年のリリース以降、極めて高い社会的注目

### わずか2か月程度で1億ユーザーを突破 (2023年当時、世界最速のペース)

テーマ別まとめ記事

#### ChatGPTとは？ 世界に与えている影響と課題について振り返る

By 日経ビジネス編集部 Read time:6min

2023.5.1



- AD ソニーが挑むデジタル戦略 データと生成AIで豊かな顧客体験と感動を創造
- AD 「ユーザー体験を変える」 新時代を体現するシェーバーが生まれた背景を探る
- AD 企業の利益に直結 福利厚生充実が採用力・定着率アップにつながる

人間のように自然な会話が行える対話型人工知能（AI）、「ChatGPT（チャットGPT）」。高度なAI技術を手軽に体験できることから話題を集め、2022年11月のリリースからわずか2か月程度で1億ユーザーを突破した。一方、その能力の高さからChatGPTの悪用を懸念する声もある。今回はChatGPTをめぐるさまざまな話題について、これまでの記事を振り返ってみる。

日経ビジネス 2023年5月1日

<https://business.nikkei.com/atcl/gen/19/00081/030800519/> より引用

### 米司法試験で上位合格など、衝撃的な報道

#### OpenAI「GPT-4」発表 精度向上、司法試験で上位10%

ChatGPT [+ フォローする](#)

2023年3月15日 7:21 [会員限定記事]



保存



**Think!** 多様な観点からニュースを考える

厩本純一さん他4名の投稿



【シリコンバレー=佐藤浩実】人工知能（AI）開発の米オープンAIは14日、生成AIの新しい基盤技術「GPT-4」を発表した。画像や長い文章を扱えるようにし、学術知識の正確性や安全性を高めたという。「ChatGPT（チャットGPT）」の有料版で技術を使えるようにするほか、開発者が自社アプリに組み込むための仕組みも順次提供する。

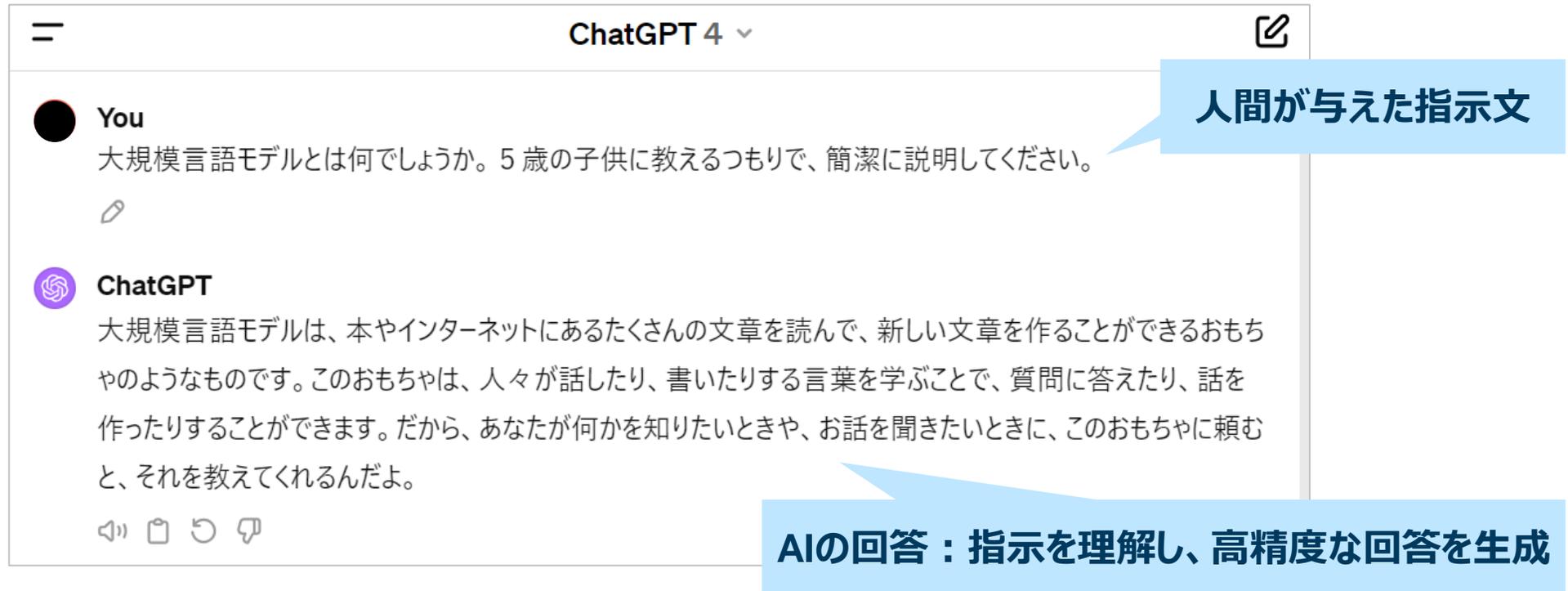
GPT-4は「大規模言語モデル」と呼ばれるAIの基盤技術で、チャットG...

日経新聞 2023年3月15日

<https://www.nikkei.com/article/DGXZQOGN1507H0V10C23A3000000/> より引用

# ChatGPTは何がすごい？

- 自然言語による指示文（prompt）を与えるだけで、幅広い課題に回答
- 課題によっては、人間を凌駕する極めて高い思考力を発揮



The screenshot shows a chat window titled "ChatGPT 4". The user's prompt is: "大規模言語モデルとは何でしょうか。5歳の子供に教えるつもりで、簡潔に説明してください。" (What is a large language model? Please explain it simply as if you were teaching a 5-year-old child). The AI's response is: "大規模言語モデルは、本やインターネットにあるたくさんの文章を読んで、新しい文章を作ることができるおもちゃのようなものです。このおもちゃは、人々が話したり、書いたりする言葉を学ぶことで、質問に答えたり、話を作ったりすることができます。だから、あなたが何かを知りたいときや、お話を聞きたいときに、このおもちゃに頼むと、それを教えてくれるんだよ。" (A large language model is like a toy that can read many sentences from books and the internet and create new sentences. This toy learns words that people use when they talk or write, so it can answer questions, talk, or create stories. So, when you want to know something or hear a story, you can rely on this toy, and it will teach you.)

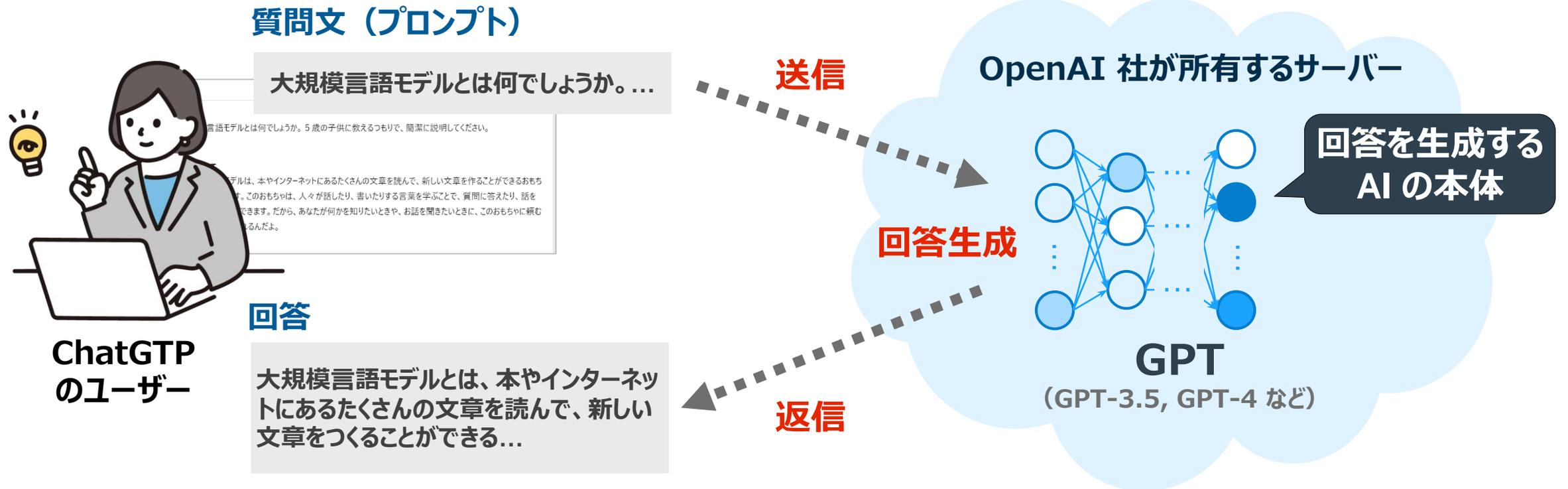
人間が与えた指示文

AIの回答：指示を理解し、高精度な回答を生成

**きわめて高性能なAIチャットが、アカウントさえ作れば誰でも使える！**

# ChatGPT を支える大規模言語モデル (LLM; Large Language Model)

ChatGPT の裏側では、GPT とよばれる大規模言語モデルを利用して回答を生成

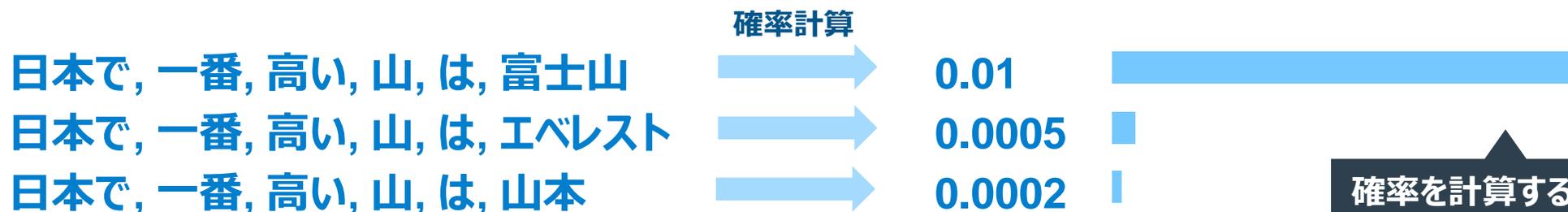


※ChatGPT は、OpenAI 社が提供する GUI 等を含めたチャットサービス全体の名称

# 言語モデルとは

文章（単語の列）がどれくらい生じやすいかを、確率的にモデル化したもの

- 単語列が「どれくらい発生しやすいか」の確率を計算できる



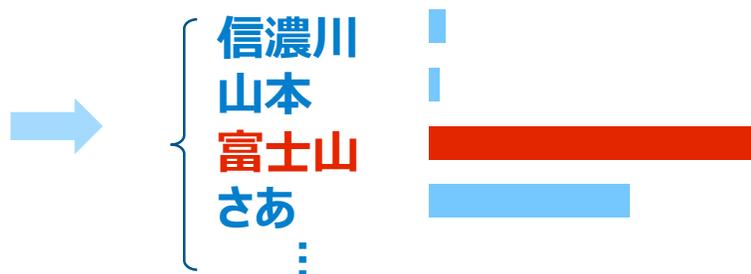
確率を計算するために  
必要なパラメータは  
学習により獲得

- 次に来る確率が高い単語を繰り返しつなげることで、文章が生成できる

書きかけの文章

日本で、一番、高い、山、は、

?



確率が高いものを  
「次の単語」として採用

# LLM はなぜ文章を生成できる？

テキストの一部を隠した「穴埋め問題」をたくさん解いて、次に来る単語の予測能力を獲得

文法  
日本 **で** 一番高い山は **富士山** である  
一般常識

算術(推論)  
鉛筆を3本買ったが2本失くしたので残り **1本**。  
感情  
お気に入りなので **悲しい**。

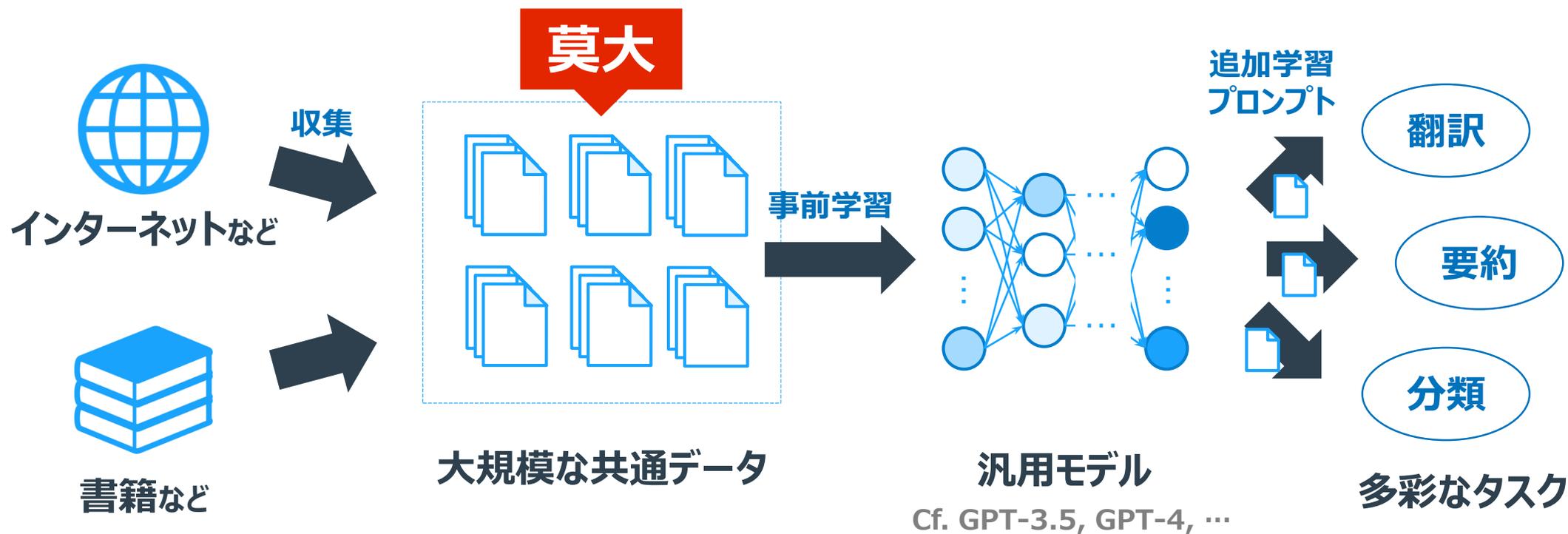
穴埋め問題から「一般常識」「文法」「算術」「感情」など幅広い知識が得られる

# 大規模言語モデル (LLM; Large Language Model)

莫大な量のテキストデータで学習し、多彩なタスクに対応する汎用的な能力を獲得

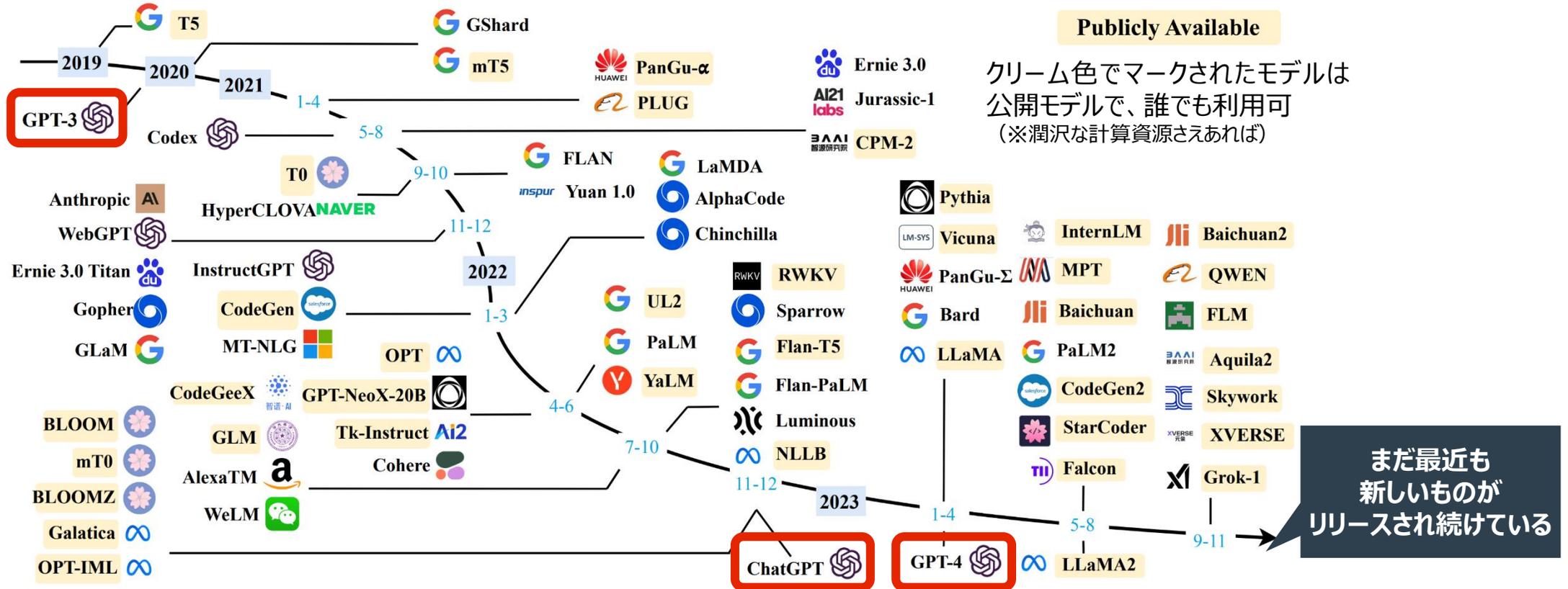


少量のデータによる追加学習や、プロンプトによる指示で、多彩なタスクをこなせる



# 次々とリリースされるLLM：世界の動向

## OpenAI以外にも、米国企業を中心にさまざまなLLMが開発・リリース



[Zhao+; 2023] A Survey of Large Language Models (arXiv:2303.18223)

大半は英語を中心に学習 → 日本での活用には、文化や慣習の理解に課題

# 次々とリリースされるLLM：国内の動向

日本語データで強化した国産LLMが、国内企業・大学から次々リリース → 競争激化へ

## 参考：代表的な日本語LLMモデル

	企業・大学名	モデル名	プレスリリース URL など
2023/05-	サイバーエージェント	CyberAgentLM	<a href="https://www.cyberagent.co.jp/news/detail/id=28817">https://www.cyberagent.co.jp/news/detail/id=28817</a>
2023/08-	LINE	japanese-large-lm	<a href="https://engineering.linecorp.com/ja/blog/3.6-billion-parameter-japanese-language-model">https://engineering.linecorp.com/ja/blog/3.6-billion-parameter-japanese-language-model</a>
2023/08-	東京大学（松尾研究室）	Weblab-10B	<a href="https://www.t.u-tokyo.ac.jp/press/pr2023-08-18-001">https://www.t.u-tokyo.ac.jp/press/pr2023-08-18-001</a>
2023/09-	Preferred Networks	PLaMo	<a href="https://tech.preferred.jp/ja/blog/llm-plamo/">https://tech.preferred.jp/ja/blog/llm-plamo/</a>
2023/10-	LLM-jp（NII）	LLM-jp-13B	<a href="https://www.nii.ac.jp/news/release/2023/1020.html">https://www.nii.ac.jp/news/release/2023/1020.html</a>
2023/10-	ストックマーク	Stockmark-13b	<a href="https://stockmark.co.jp/news/20231027">https://stockmark.co.jp/news/20231027</a>
2023/12-	東京工業大学、産総研	Swallow	<a href="https://tokyotech-llm.github.io/swallow-llama">https://tokyotech-llm.github.io/swallow-llama</a>
2024/01-	カラクリ	KARAKURI LM	<a href="https://karakuri.ai/seminar/news/karakuri-lm/">https://karakuri.ai/seminar/news/karakuri-lm/</a>
2024/03-	NTT	Tsuzumi	<a href="https://www.rd.ntt/research/LLM_tsuzumi.html">https://www.rd.ntt/research/LLM_tsuzumi.html</a>



軽量の学習方法の開発や開発補助制度の整備により、今後も新たな国産LLMのリリースが予想

# 直近の動向：GPT-4o のリリース

## OpenAI 社から さらなる性能・処理速度を持つ GPT-4o がリリース 最高性能クラスのモデル が 低コストで利用可能に

1. 「高度な専門知識」「大学院生並の推論能力」等で **最高性能** を達成

2024/05/14 公開

	GPT-4o OpenAI	GPT-4 turbo OpenAI	Claude 3 Opus Anthropic	Gemini Pro 1.5 Google	Llama3 400b (※) Meta
法律・医学・歴史等の専門知識 MMLU(%)	<b>88.7</b>	86.5	86.8	81.9	86.1
大学院生並の推論能力 GPQA(%)	<b>53.6</b>	48.0	50.4	-	48.0
数学分野の問題解決能力 MATH(%)	<b>76.6</b>	72.6	60.1	58.5	57.8
コード(プログラム)の生成能力 HumanEval(%)	<b>90.2</b>	87.1	84.9	71.9	84.1
多言語での算術能力 MGSM(%)	<b>90.5</b>	88.5	90.7	88.7	79.0
複数パラグラフに基づく質問応答 DROP(f1)	83.4	<b>86.0</b>	83.1	78.9	83.5

[OpenAI; 2024] Hello GPT-4o  
(表は <https://openai.com/index/hello-gpt-4o/> の"Text Evaluation"を参考に当社が作成。※Llama3 400bは現在も学習中であり、5/13時点での評価値)

2. 日本語を含む 多言語データ **処理効率の改善**

こんにちは、私の名前はGPT-4oです。  
私は新しいタイプの言語モデルです、初めまして



LLMは文章を  
単語の列として処理

こんにちは、\、私、の、名前、は、  
GPT、...

GPT-4-turboと比較して  
単語の**処理方法が改善し、  
高速な生成が可能に**

# 直近の動向：生成AIのマルチモーダル化

異なるモダリティ（文章・画像・音声・動画）も自然言語を通じて共通の情報として処理  
AIが人間と近い感覚を獲得し、より自然なコミュニケーションも視野に

## 画像生成・理解

Stable Diffusion など

[Esser+, 2024] Saling Rectified Flow Transformers for High-Resolution Image Synthesis (arXiv:2403.03206) など



Figure 1. High resolution samples from our Hi rectified flow model, showcasing its capabilities in typography, precise prompt following and spatial reasoning, attention to fine details, and high image quality across a wide variety of styles.

## 音声生成・認識

Whisper, Text to Speech AI など

[Radford+, 2022] Robust Speech Recognition via Large-Scale Weak Supervision (arXiv:2212.04356) など



## 動画生成・理解

Sora など

[Liu+, 2024] Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models (arXiv:2402.17177) など



<https://openai.com/sora> より引用

自然言語を通じて処理



マルチモーダル  
言語モデル  
(MLLM)

利用例：動画理解+音声認識+音声生成による  
リアルタイム対話（GPT-4o）



Dog.

[OpenAI; 2024] Hello GPT-4o “model capabilities”より引用  
<https://openai.com/index/hello-gpt-4o/>

人間の会話と変わらない応答速度を実現しつつある  
[OpenAI; 2024]

「人間とAIのインタラクション」の高度化  
⇒人間とAIがよりリアルタイムに共創する社会？

# LLMはなにが「大規模」なのか？

学習したデータ量・モデルの複雑さ（パラメータ数）・学習に要する計算コストのすべてが大規模

## 学習したデータ量

GPT-3 で約5000億トークン※のテキストを学習

※トークン：言語モデルの処理単位（日本語では、大体1トークン=1文字）



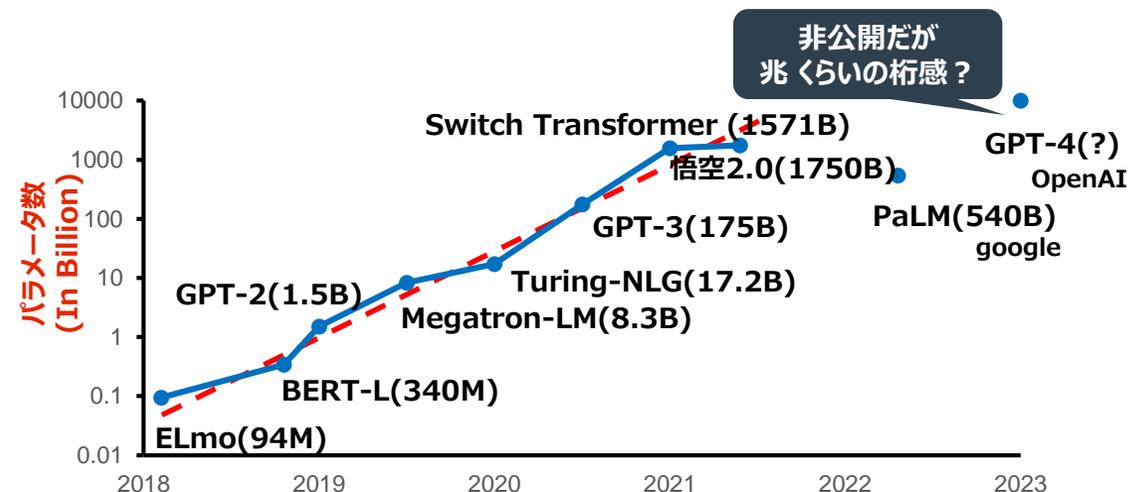
日本語の小説で単純に換算すれば、約500万冊相当

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

[Brown+; 2020] Language Models are Few-Shot Learners (arXiv:2005.14165)

## モデルの複雑さ（パラメータ数）

どんどん複雑化し、現在数千億を超える規模のパラメータ数（つまり、数百億～数千億程度の数値情報を用いて文章が生成される）



[Narayanan+; 2020] Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM (arXiv:2104.04473)

# なぜ、「大規模」なほうが良いのか？

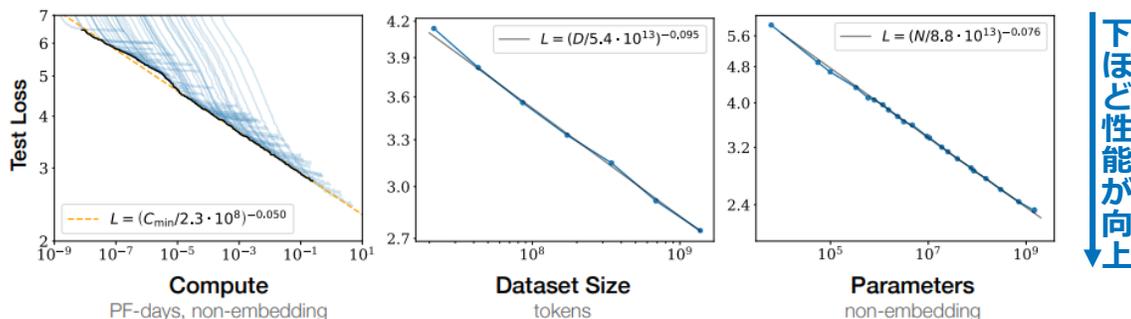
経験的に成り立つと考えられている二つの性質

1. 大規模にすればするほど、性能が良くなる (scaling law)
  2. 大規模だからこそ、解ける課題がある (emergent ability)
- ➡ 大規模な方が有利！

## Scaling Law

Transformer 系のモデルは、以下を改善 (大規模化) すればするほど、性能も改善 (loss が小さくなる)

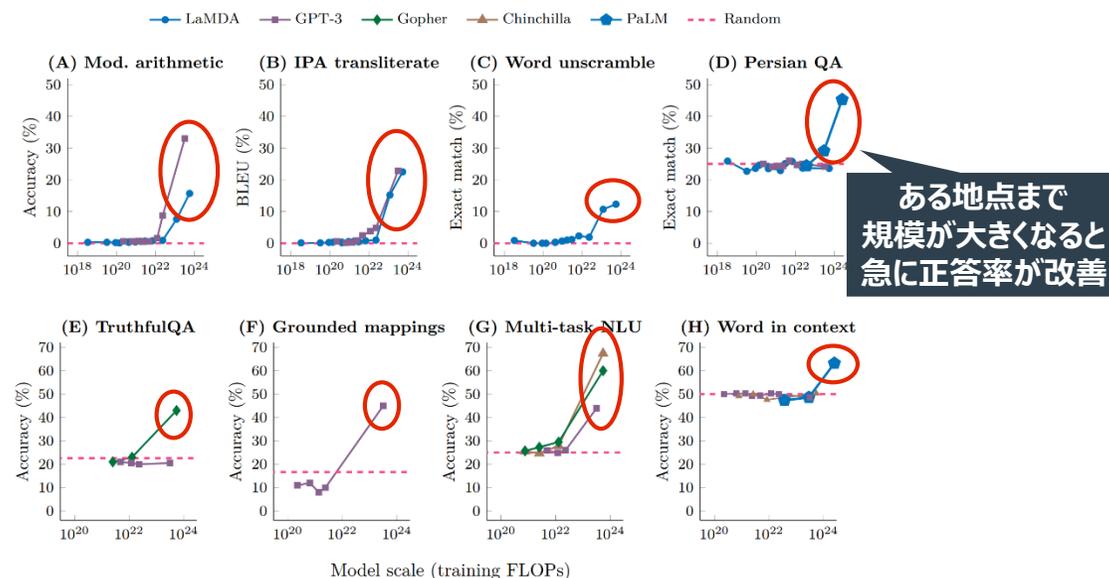
- 学習計算回数  $C$  ( $\text{loss} \propto C^{-0.05}$ )
- データセットの大きさ  $D$  ( $\text{loss} \propto D^{-0.095}$ )
- モデルのパラメータ数  $N$  ( $\text{loss} \propto N^{-0.076}$ )



[Jared+; 2020] Scaling Laws for Neural Language Models (arXiv:2001.08361)

## Emergent Ability

モデルを大規模化すれば急に解けるようになるタスクがある



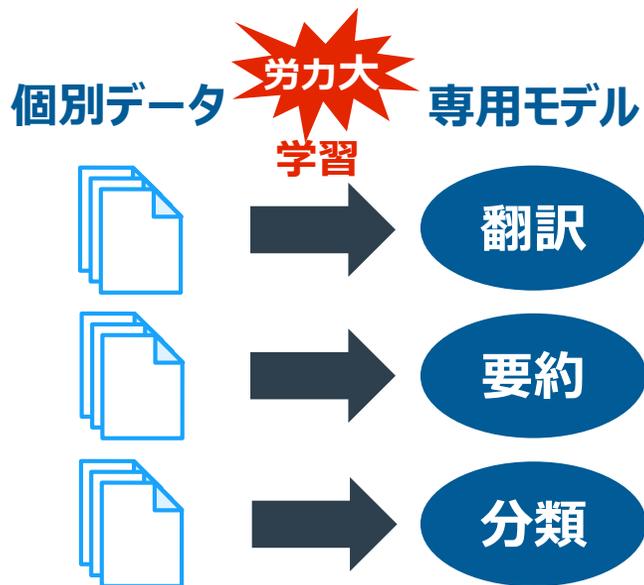
[Wei+; 2022] Emergent Abilities of Large Language Models (arXiv:2206.07682)

# LLMによりAIはどう変わった？ / どう変わる？

大量のデータがなくても、AIが構築可能に → アイデア次第で誰にでもAI活用のチャンス！

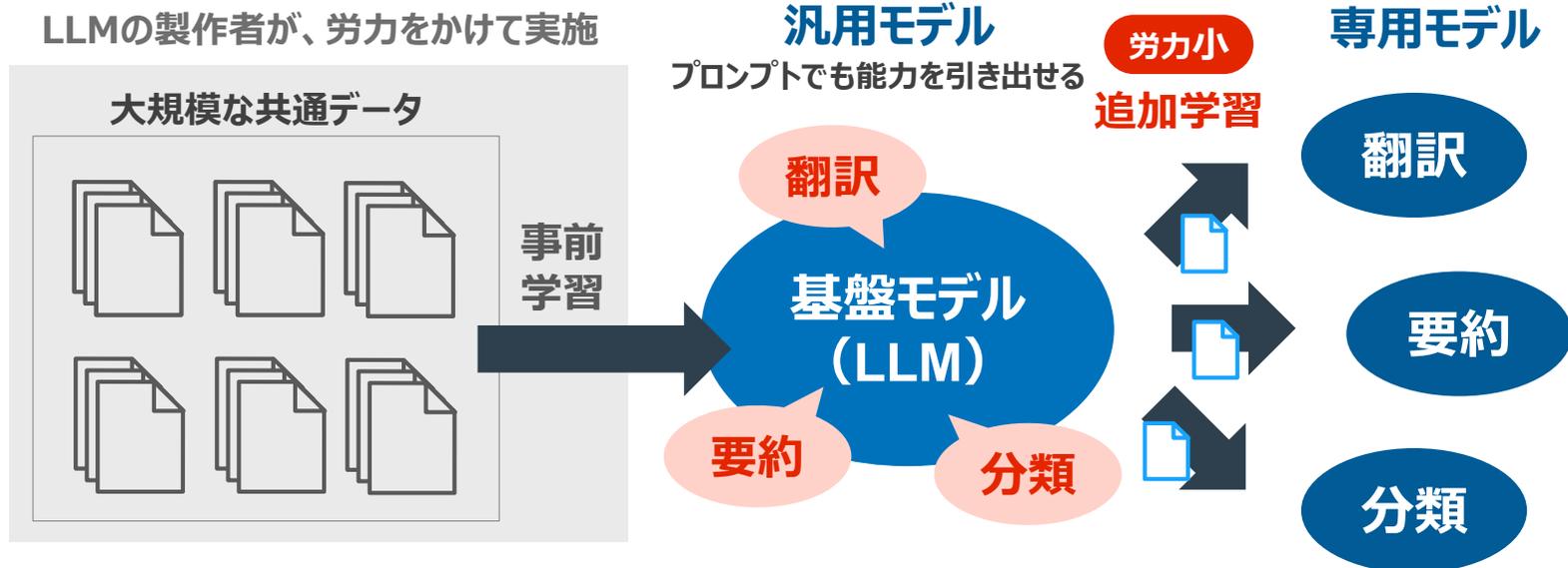
## LLM 以前

- タスクごとに、専用モデルを個別の大量データから構築
- ユーザーは都度大量のデータを準備し、**多大な労力**でモデルを作成する必要



## 基盤モデルとしての LLM 以後

- LLMは、大規模データにより後続タスクに必要な共通知識や常識を獲得した汎用モデル
- 一度汎用モデルを作ってしまうと、ユーザーは、プロンプトの作成や簡単な追加学習をするだけで、様々な課題に適用可能に！



# LLM の有名な問題点：幻覚（ハルシネーション）

LLM は幻覚（ハルシネーション）とよばれる、もっともらしいウソを述べる可能性がある



You

株式会社NTTデータ数理システムの社長を教えてください。



ChatGPT

申し訳ありませんが、情報の最新化ができていないため、2022年1月時点での情報しか提供できません。2022年1月時点での情報に基づいて回答します。株式会社NTTデータ数理システムの社長は、吉田 洋一郎（よしだ よういちろう） 氏でした。

2022年1月時点で「吉田洋一郎」氏が  
当社の社長だったことはない  
→ **完全に誤った情報！**

- 外部リソースとの接続によりある程度は緩和できる
- しかし、どれだけ工夫したとしてもハルシネーションを根絶するのは困難  
→ ある程度のハルシネーションが生じるのは前提として、活用方法を工夫するべき

# ほかにもある LLM の弱点

LLM の弱点をふまえて、テクニックや他の技術で補いながら使いこなすことがコツ

弱点1  
あいまいな指示では、  
あいまいな結果しか引き出せない

だから…



LLM活用のコツ1  
プロンプトエンジニアリングによる  
「深掘り」を、LLM自身にさせよう

弱点2  
学習データに含まれない  
知識を持たない

だから…



LLM活用のコツ2  
LLMが知らない知識は、  
外から補って使おう

弱点3  
たくさんのデータから  
バランスよく結論を導くのは苦手

だから…



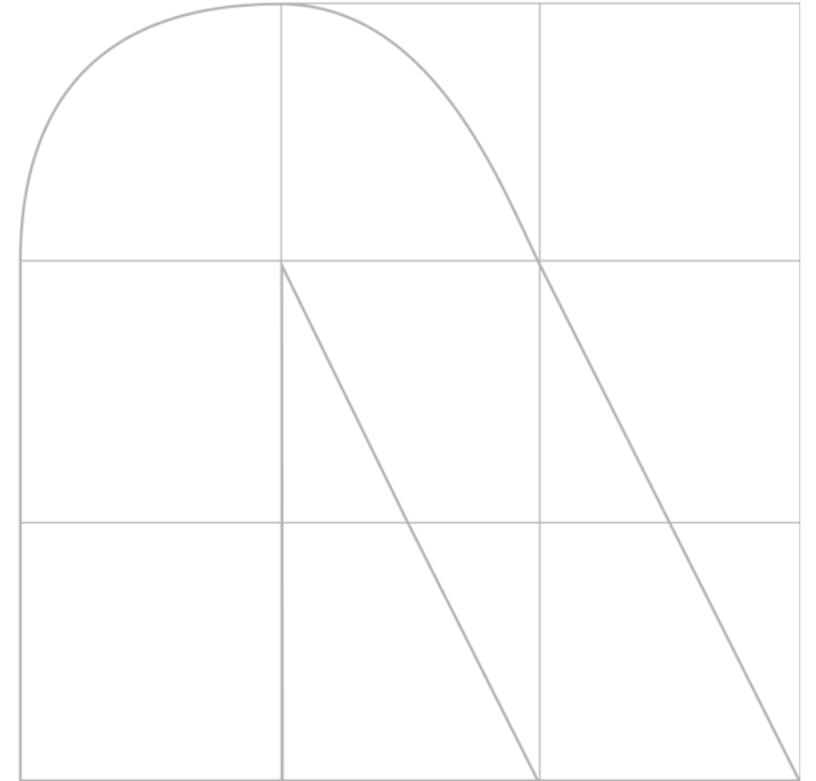
LLM活用のコツ3  
データ分析技術と  
組み合わせて使おう

これら三つのコツを、当社事例を交えてご紹介いたします！

# 1

LLMの活用のコツその1.

プロンプトエンジニアリングによる  
「深掘り」を、LLM自身にさせよう



# LLMの限界1：あいまいな指示では、あいまいな結果しか引き出せない

LLMはあくまでも一般論を回答したがる → 具体的な深掘りをするにはコツが必要

## 例 当社のSWOT分析を、ChatGPTでやってみた



ユーザー

NTTデータ数理システムについて、SWOT分析を実際に行った結果を示してください。なお答える際には、情報通信業界が取り巻く状況をインターネット上で調査してください。また、少なくとも以下のサイトの情報を参照してください。

- NTTデータ数理システムの強み <https://www.msi.co.jp/speciality/index.html>
- 同業A社の強み [https://\\*\\*\\*.html](https://***.html)

具体性に欠ける...



ChatGPT

- 強み (S) : 数理学とコンピュータサイエンスの専門知識、豊富な経験、先進技術の活用能力。
- 弱み (W) : 特定業界への依存度やグローバル市場での限定的な影響力。
- 機会 (O) : デジタルトランスフォーメーションの進展、新しいビジネスモデルへの適応。
- 脅威 (T) : 国内外の競合他社との競争、技術の急速な変化への対応の必要性。

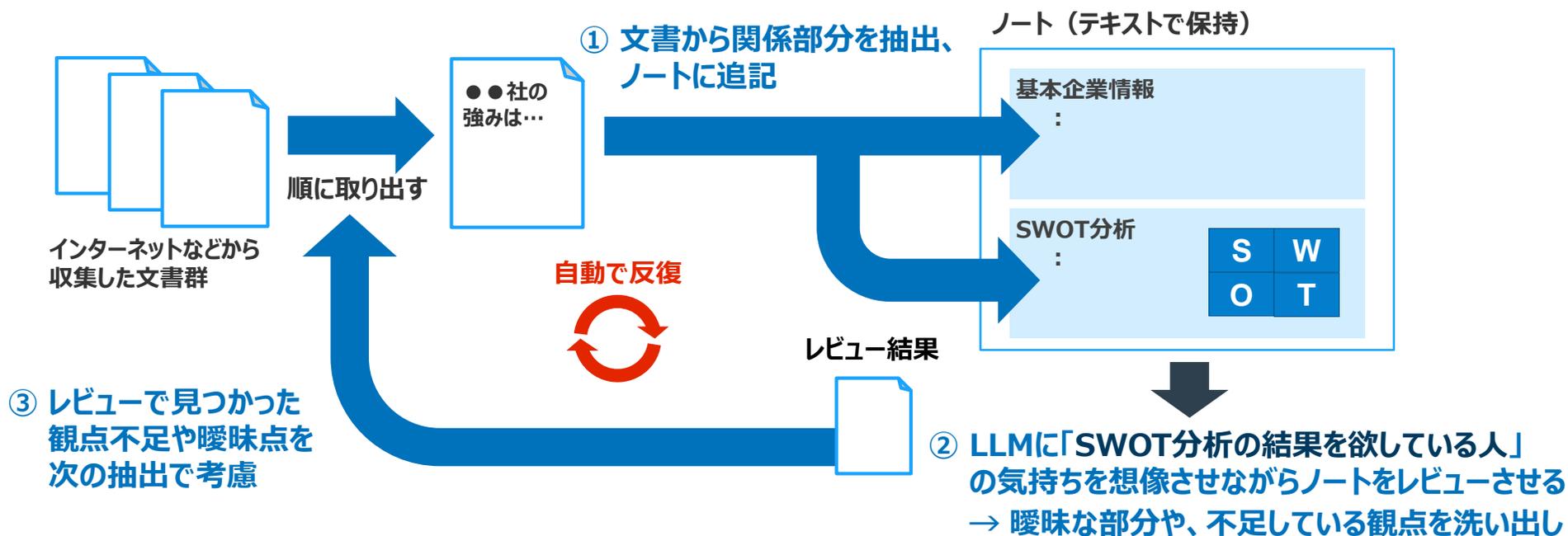
あたりまえすぎる...

 もっと踏み込んでこたえてくれないと、新たな気づきが無い……

# プロンプトエンジニアリングによる自律的な「深掘り」

プログラムにより動的に指示を生成して、よりLLMに「深掘り」をさせる

## 例 SWOT分析を、自動的に深掘りする構成を検討



👍 「SWOT分析に重要なポイント」をLLMに考えさせながら、自動的に深掘りする構成

# 当社事例 SWOT分析の結果

	ポジティブ	ネガティブ
内的要因	<p><b>強み Strength</b></p> <ul style="list-style-type: none"> <li>40年以上にわたる数理科学とコンピュータサイエンスの技術発展と普及活動への貢献</li> <li>幅広いタイプのデータサイエンティストが所属し、プロジェクトの段階や技術領域にマッチする人材を選出</li> <li>ユーザーフレンドリーな分析プラットフォームの提供</li> <li>先端技術を活用した課題解決、新たな手法の調査・実践、弱教師あり学習や強化学習などの研究開発</li> </ul> <p>：</p> <p>サイトを読み込んで多くの強みを挙げられた</p>	<p><b>弱み Weakness</b></p> <ul style="list-style-type: none"> <li>競合他社と比較して、データ活用のための基盤構築やPDCAサイクルの確立に関する具体的な取り組みが明確でない</li> <li>社内外のデータ分析人材の不足がネック。社内人材の育成とともに、お客様の技術的理解の推進も必要</li> </ul> <p>納得感のある アピール不足を指摘される</p>
外的要因	<p><b>機会 Opportunity</b></p> <ul style="list-style-type: none"> <li>データ活用の増加による新たなビジネスチャンスの創出</li> <li>データ分析技術の進化による新サービスの開発</li> <li>NTTデータグループの方法論を元にしたデータ活用方法のサポートが可能</li> </ul> <p>当社がNTTデータグループであることを理解し、機会を挙げられた</p>	<p><b>脅威 Threats</b></p> <ul style="list-style-type: none"> <li>パーソナルデータの取扱いに関するインシデントリスクや社会的責任の大きさ</li> <li>レピュテーションリスクやデータの所有権に関する問題</li> <li>データの収集・管理に係るコストの増大</li> </ul>

※LLMの出力を、趣旨の変わらない範囲で一部人手で整理した部分があります。

 **単純にLLMに尋ねた場合よりも、深掘り出来ている！**

# 当社事例 SWOT分析の途中で生成したレビュー

## 途中で抽出されたレビュー内容

1. 数理システム社の業務内容について、どのような新事実を発見しているのか、またビジネスバリューをどのように追求しているのか具体例が欲しい。
2. 顧客ターゲットの範囲が非常に広いが、主要な顧客セグメントや最も成功している市場はどこなのか。
3. 技術的要因について、AIやローコード開発の普及がどのように数理システム社に影響を与えているのか、もっと具体的な影響を知りたい。
4. 強みに挙げられている「幅広いタイプのデータサイエンティスト」がいるが、これはどのように競合他社と差別化されているのか。

：

厳しい指摘！

幅広いデータサイエンティストの存在を強みとしてアピールしているが...

### Data Scientist

データサイエンティスト

単にデータサイエンティストと言っても、得意な技術領域は異なります。プロジェクトは多数のフェーズを経て運用に至りますが、どのフェーズに力点が置かれるかはプロジェクトごとに異なります。例えば、プロジェクト初期の「業務の現状を把握し、そもそもどんなサービスやシステムが欲しいか」についてお客様と一緒に検討し具体化させていく段階と、業務課題を数理科学的な問題設定に適切に落とし込む段階、さらにそれを高度な技術力で実現させる段階では、データサイエンティストに求められる能力が異なります。

当社にはさまざまな得意分野とこだわりを持つ幅広いタイプのデータサイエンティストが所属しており、お客様のプロジェクトの段階や技術領域などにマッチするデータサイエンティストを選んでいきます。



数理科学を信じ、数理科学を乗りこなす

シミュレーション&マイニング部  
伊藤 孝太郎  
リサーチャータイプ



データサイエンスにおける研究と実用の二刀流

シミュレーション&マイニング部  
村田 智也  
トップ研究者タイプ



現場のノウハウも取り込み、お客様が納得する分析結果を出し続ける

データマイニング部  
中道 祐希  
アナリストタイプ



数理最適化の技術に夢をみる

数理計画部  
藤井 浩一  
トップ研究者タイプ



「シミュレーション・数理最適化・機械学習」三種の神器を武器に課題を解決する

シミュレーション&マイニング部  
豊岡 祥  
リサーチャータイプ



お客様や現場を巻き込んだ強いチームを作り、プロジェクトの成功に邁進する

数理ソリューション部  
佐藤 誠  
コンサルタントタイプ

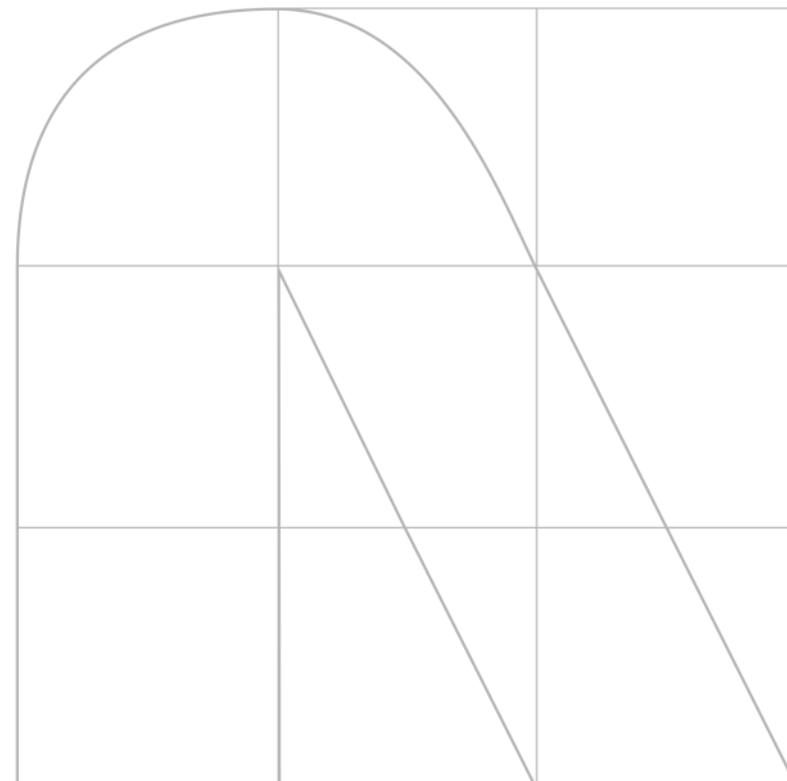
SWOT分析のインプットとした、当社の強みページ  
<https://www.msi.co.jp/speciality/index.html>

👍 途中で抽出されたレビュー結果も、興味深い観点が抽出できている

# 2

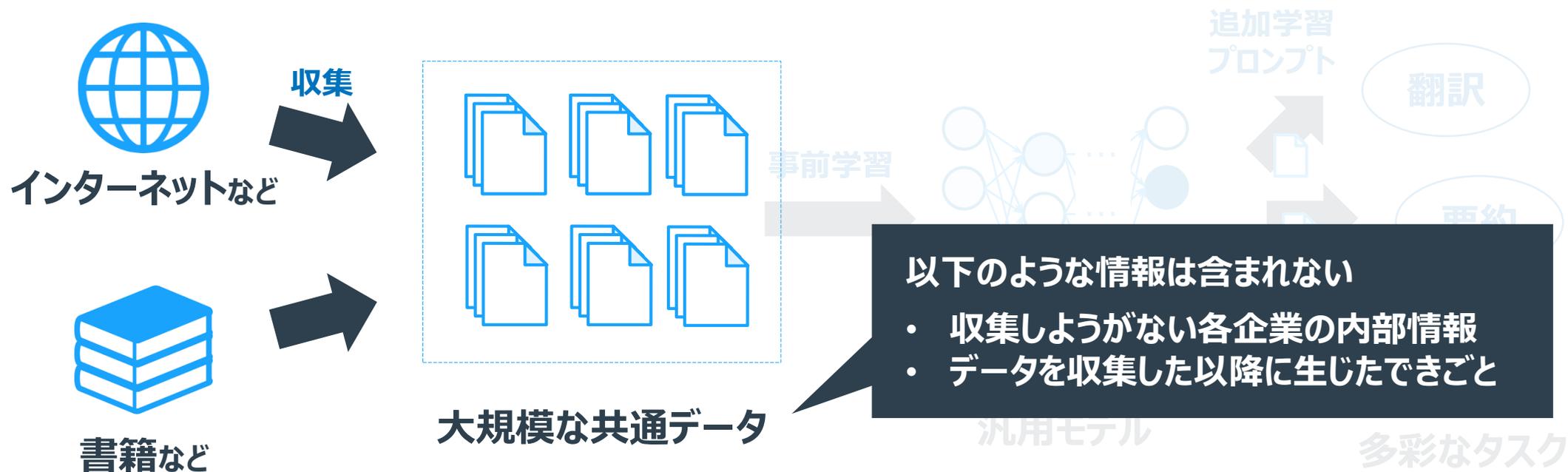
LLMの活用のコツその2.

**LLMが知らない知識は、  
外から補って使おう**



# LLMの限界 2 : 学習データに含まれない知識を持たない

👉 **社内情報**や、**最新時事**などに関する質問には答えられない

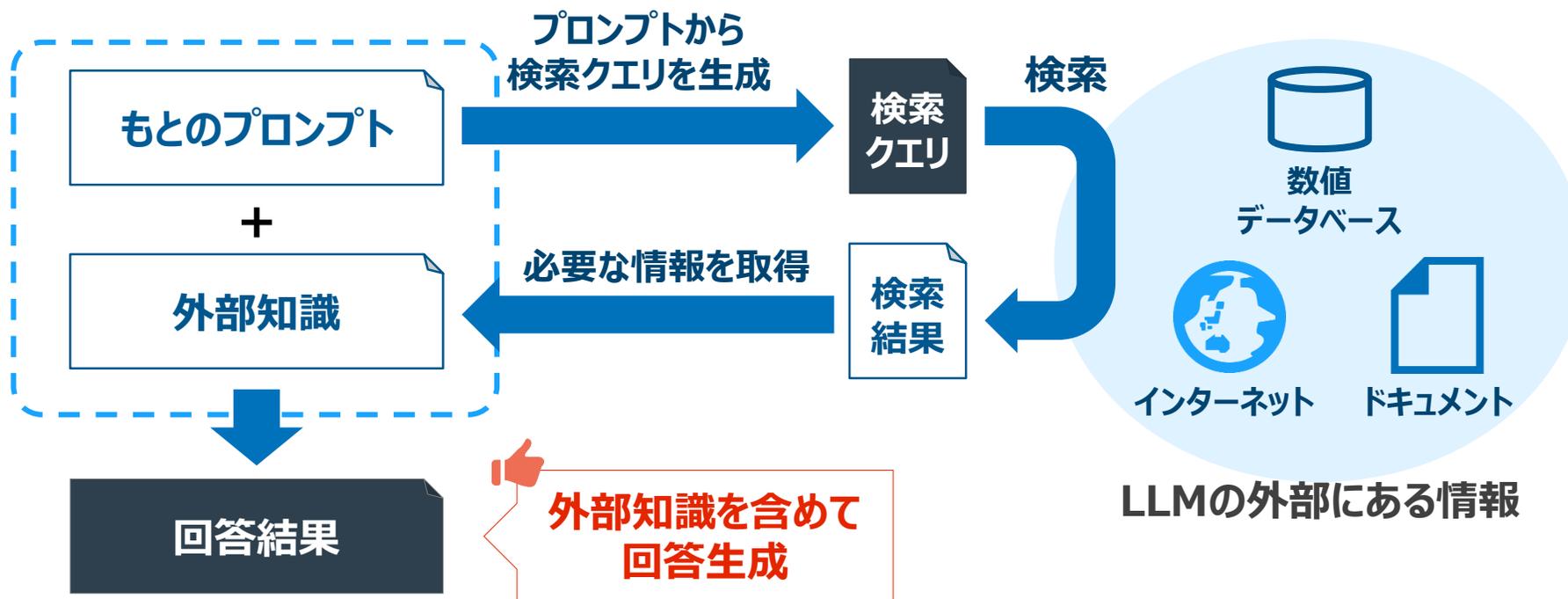


**外部から知識を補いながら、回答を生成する**

# 知識を補う仕組み：RAG（Retrieval-Augmented Generation）

[Gao+; 2024] Retrieval-Augmented Generation for Large Language Models: A survey (arXiv:2312.10997)

各種検索システムと組み合わせてLLMを利用し、知識を補いながら回答を生成



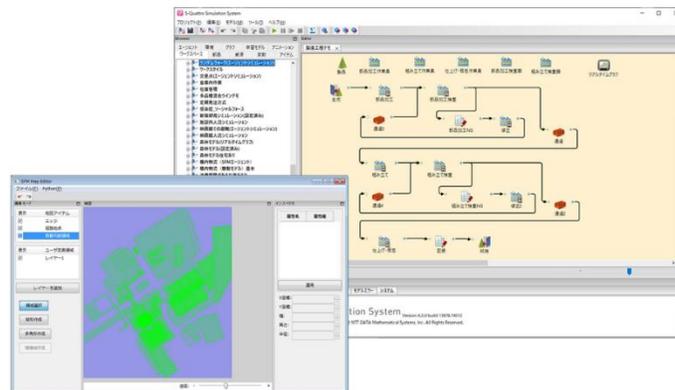
👍 参考にするデータさえあれば、社内情報や最新時事に関する質問にも回答

## 当社事例 当社製品に関するQAチャットアプリの制作

# RAG を活用し、当社ソフトウェア製品 S<sup>4</sup> Simulation System の QA チャットアプリを構築

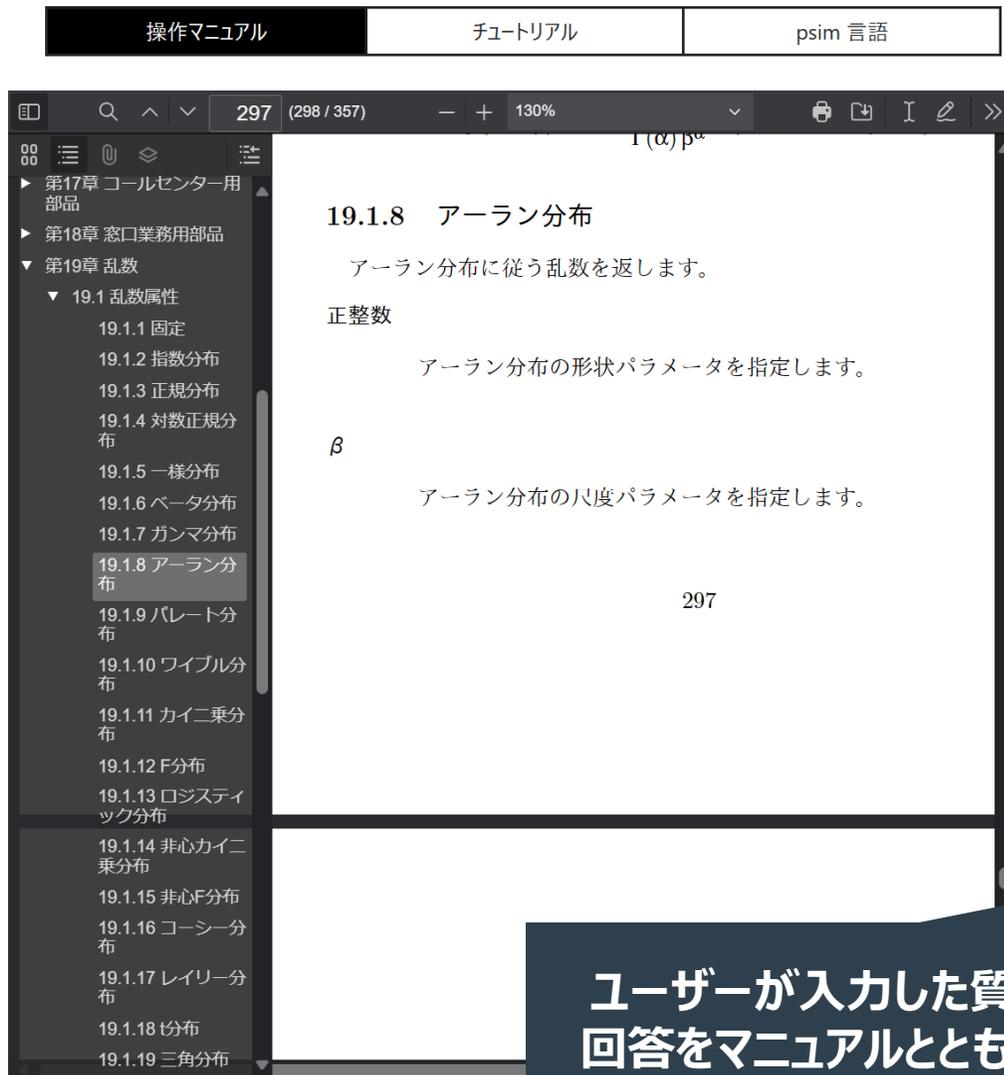


ソフトウェア利用中に生じる疑問を、**ユーザーが自力で素早く解決**！  
(当社としては) サポート工数の削減や、製品満足度の向上なども期待



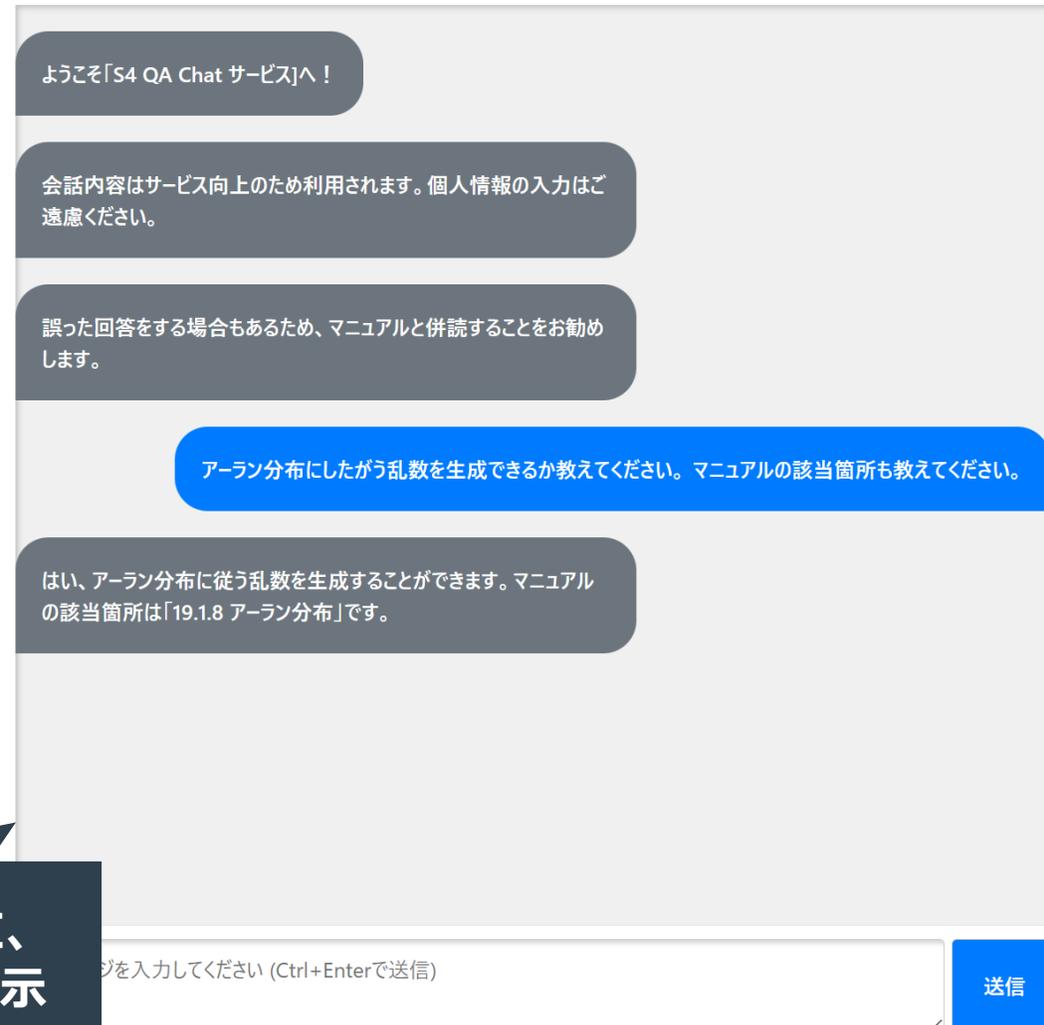
## S<sup>4</sup> (エスクワトロ) Simulation System

自社開発の汎用シミュレーションソフトウェア  
GUI操作により、様々なシミュレーションを行える

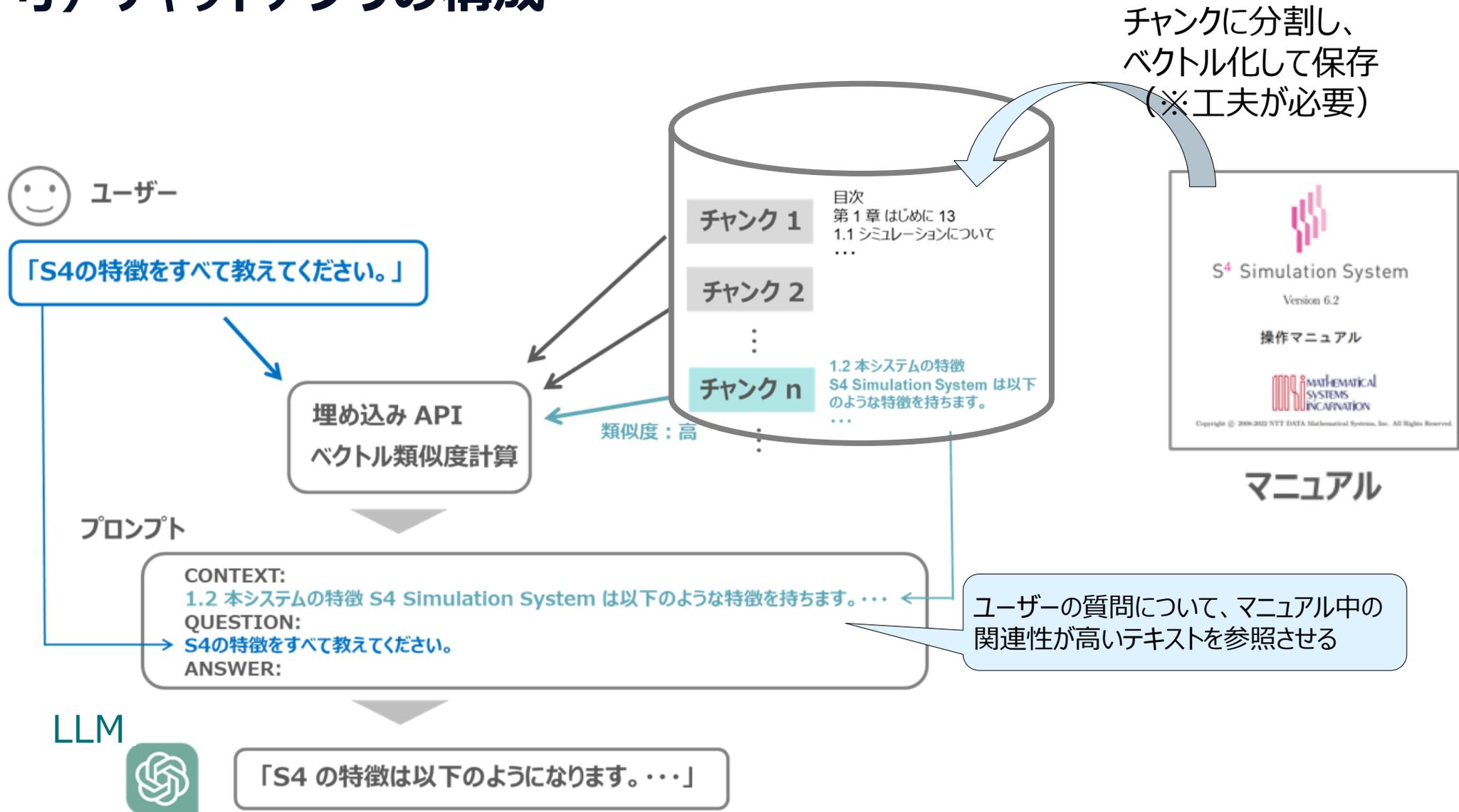


ユーザーが入力した質問に、  
回答をマニュアルとともに表示

## Q&A Chat



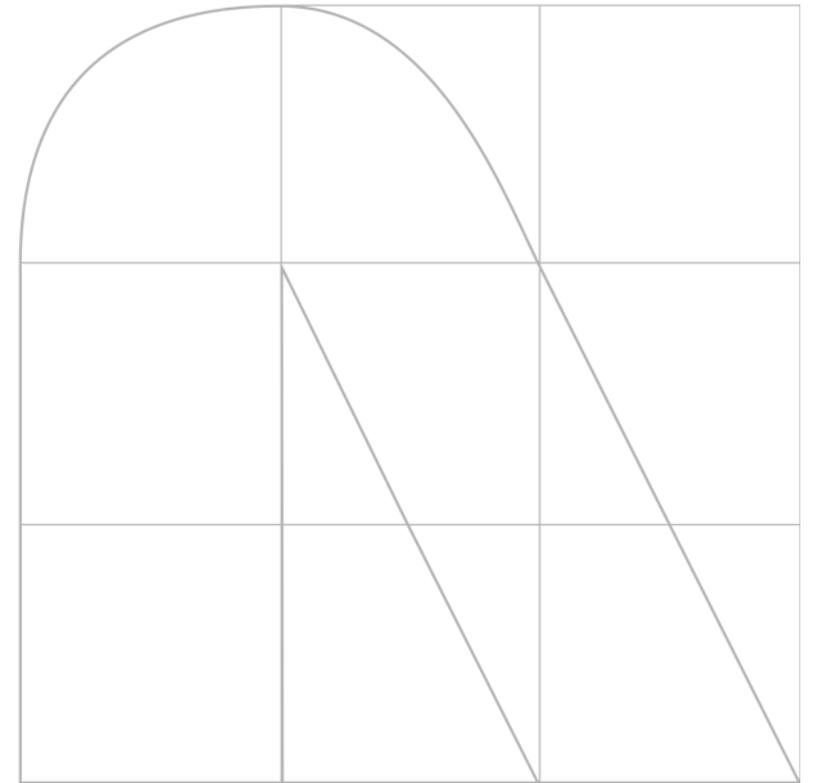
# (参考) チャットアプリの構成



# 3

LLMの活用のコツその3.

**データ分析技術と組み合わせて使おう**



# LLMの限界 3 : たくさんのデータからバランスよく結論を導くのは苦手

- 指示文（プロンプト）として入力できるデータの長さには限界がある
- どのデータを、どのように考慮して判断したかが不透明になる

## 商品のレビューデータなど、大量のテキストデータ

評価	性別	年代	書き込み日	レビュー
5	女性	60代	2015/3/23	これまでずっと手ゴネでパンを作っていて、初めてのホームベーカリーを購入しました。心配していた音も振動も、「うるさい」と感じるものではありませんでした。ほとんどお餅を食べない我が家。お正月に切り餅パックを買っても数か月冷凍庫に残ります。そんな我が家でもちつき機能は必要か、と悩みましたが、一回につける量が2合か3合で丁度よく、つきたてが美味しいこともあり残ることはありませんでした。お正月も、2合でいけそうです。こんなに楽にお餅をつくれるならもう「餅」を買うことはないでしょう。まだ買って1週間ですが、パンは3回焼きました。これにして良かったです。 パン焼き器って手入れが大変そうかなって思ってたけど、掃除は楽です！それと、食パンも本当においしく焼けます！一番おいしいです。そのときの気分によ
5	女性	50代	2015/3/29	おいしいです。そのときの気分によ
4	女性	40代	2015/4/5	の15分。あとは冷蔵庫で寝かして購入したばかりでほぼ毎日買おうより安いかも。(ホームベーカリー後にはおいしくないですが、こ
...	...	...	...	あっておいしいです。

これらのデータから、何が顧客満足度を高める要因かわからないだろうか？

LLMに尋ねようとしても...



LLM



大量のデータを一度に取り込むことはできない ※

※最近のLLMでは徐々に入力トークン数は大きくなっているため、将来的に十分な量のデータを取り込めるようになる可能性はある。しかし、入力できたとしても、それらの情報をすべてバランスよく考慮した判断をすることは、LLM単独では相変わらず難しいと考えられている。  
[Levy+; 2024] Same Task, More Tokens: the Impact of Input Length on the Reasoning Performance of Large Language Models (arXiv:2402.14848)

# データ分析技術との組み合わせで、LLMの弱点を補う

テキストマイニングなどの当社技術を使って、データを要約してからLLMに入力

## テキストデータ

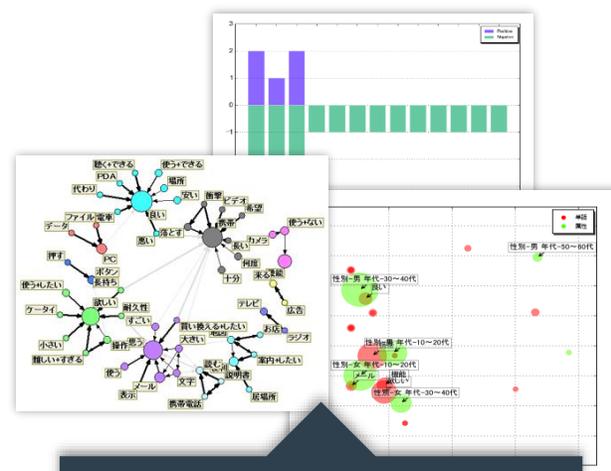
評価	性別	年代	書き込み日	レビュー
5	女性	60代	2015/3/23	これまでずっと手ゴネでパンを作っていて、初めてのホームベーカリーを購入しました。心配していた音も振動も、「うるさい」と感じるものではありませんでした。ほとんどお餅を食べない我が家。お正月に切り餅バックを買っても数か月冷凍庫に残ります。そんな我が家でもちつき機能は必要か、と悩みましたが、一回につける量が2合か3合で丁度よく、つきたてが美味しいこともあり残ることはありませんでした。お正月も、2合でいけそうです。こんなに楽にお餅をつくれるならもう「餅」を買うことはないでしょう。まだ買って1週間ですが、パンは3回焼きました。これにして良かったです。
5	女性	50代	2015/3/29	パン焼き器って手入れとか大変そうかって思ってたけど、掃除は楽です！それと、食パンも本当においしく焼けます！あと一番のオススメはうどんの麺です！コシがあって、とてもおいしいです。そのときの気分によって太さなどお好みにも出来ます！練る時間はたったの15分。あとは冷蔵庫で寝かせて切るだけ。この品番のがオススメです！
4	女性	40代	2015/4/5	購入したばかりでほぼ毎日パンを焼いていますが、材料代など計算すると普通にパンを買うより安いかも。(ホームベーカリーの代金は考慮していません)市販のパンは2日後にはおいしくないですが、このホームベーカリーで焼いたパンはやわらかさも歯ごたえもあっておいしいです。
...	...	...	...	...

テキストマイニング  
自然言語処理  
機械学習・統計



## 分析・処理結果

当社パッケージ製品 Text Mining Studio を利用



データの統計的な要約  
(解釈にはコツが必要)

## 解釈・仮説立案

解釈



LLM

言葉による  
思考や説明

👍 統計的な客観性 + LLMの思考力・説明力 を組み合わせるとインサイトを得られる

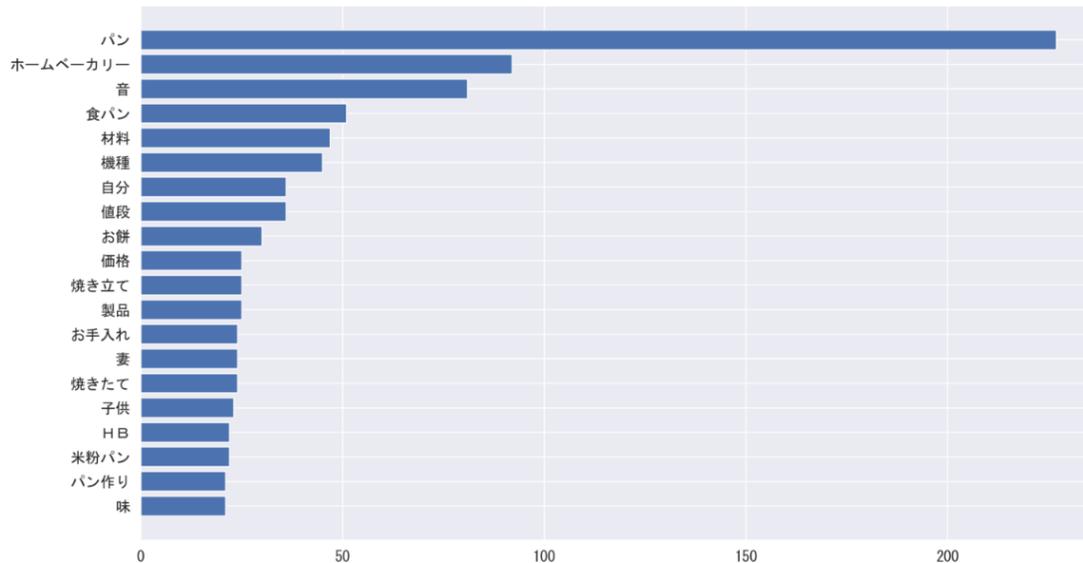
# 当社事例 テキストマイニング結果を用いた仮説の探索

## 当社テキスト分析製品とLLMを使って、テキストデータから仮説を探索するシステムを試作

当社テキスト分析製品  
Text Mining Studio を  
用いたテキストデータの分析

Text  
Mining  
Studio

単語頻度解析1



	単語	品詞	品詞詳細	頻度
0	パン	名詞	一般	227
1	ホームベーカリー	名詞	一般	92
2	音	名詞	一般	81
3	食パン	名詞	一般	51
4	材料	名詞	一般	47
5	機種	名詞	一般	45
6	自分	名詞	一般	36
7	値段	名詞	一般	36
8	お餅	名詞	一般	30
9	価格	名詞	一般	25
10	焼き立て	名詞	一般	25
11	製品	名詞	一般	25

結果に対する質問を入力すると、  
LLMが結果について解釈・説明

他にはどんな単語が出現しそうですか？

## 分析目的を入力すると、テキスト分析結果を用いてLLMが仮説を構築して提示

TMS x LLM v0.0.1

分析しているデータの概要や、あなたの立場・分析目的について教えてください。

### 分析目的  
私はホームベーカリー製品の開発を担当しています。より「評価」の高い製品を開発するべく、ユーザーが「満足には至らない」と感じている点を抽出したいと考えています。

TMSの分析結果をアップロード

434.0B

分析結果の選択

単語頻度解析1.csv

チュートリアルの内容を確認

分析結果を表示する

分析結果を説明してもらう

**分析目的を入力**  
今回は、「製品開発のためユーザーの不満点を調査」

Deploy

分析結果の解釈を教えてください。

**分析目的の概要**

このデータは、ユーザーが「ホームベーカリー」製品についてレビューした内容に基づいています。特に、製品に対する満足度を示す「評価」の情報を含んでおり、開発者としては、ユーザーがどのような点で満足していないのか、またはどのような機能や特性が評価を向上させる可能性があるのかを理解することが目的です。

**仮説2: 「お手入れの手間」**

- **タイトル:** 「お手入れの手間」
- **説明:** 「お手入れ」がレビューで言及されていることから、ホームベーカリーの清掃やメンテナンスがユーザーにとって面倒である可能性があります。特に、使用後の掃除が煩雑であると、ユーザーの満足度を下げる要因になり得ます。

これらの仮説を検証することで、ユーザーの満足度を向上させるための具体的な改善点を見つけ出し、製品開発に役立てることができます。

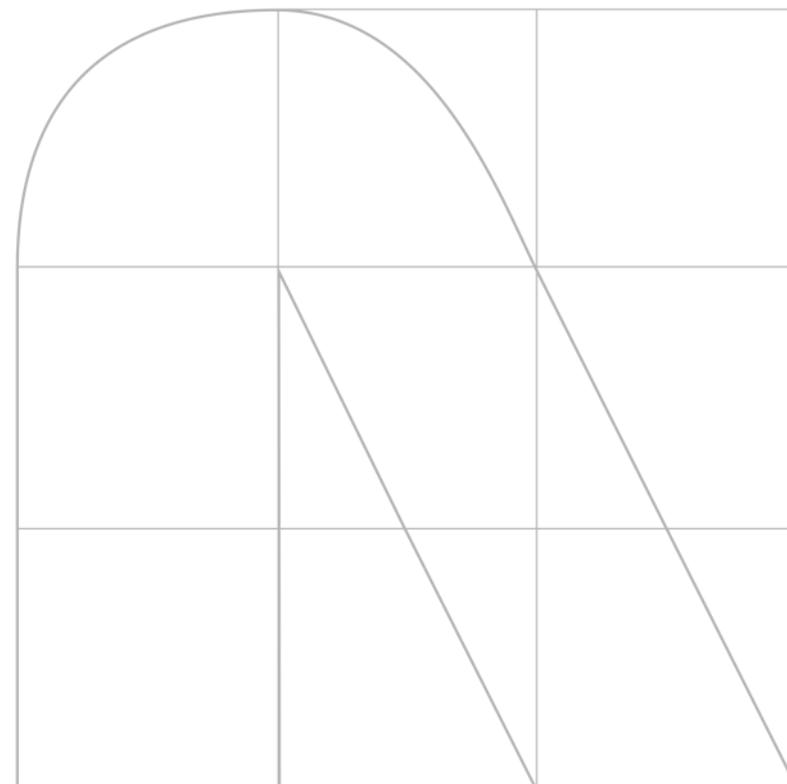
メッセージを入力してください

**データから読み取れる、ユーザーの不満についての仮説を提示**

# 4

もっとLLMを活用する

## LLMエージェント技術による アイデア創発



# LLMエージェント技術

[Wang+, 2023] A survey on Large Language Model based Autonomous Agents (arXiv:2308.11432)

## LLMに設定・記憶・計画・行動のための外部機構を付与 → 現実のタスクを自律的に実行

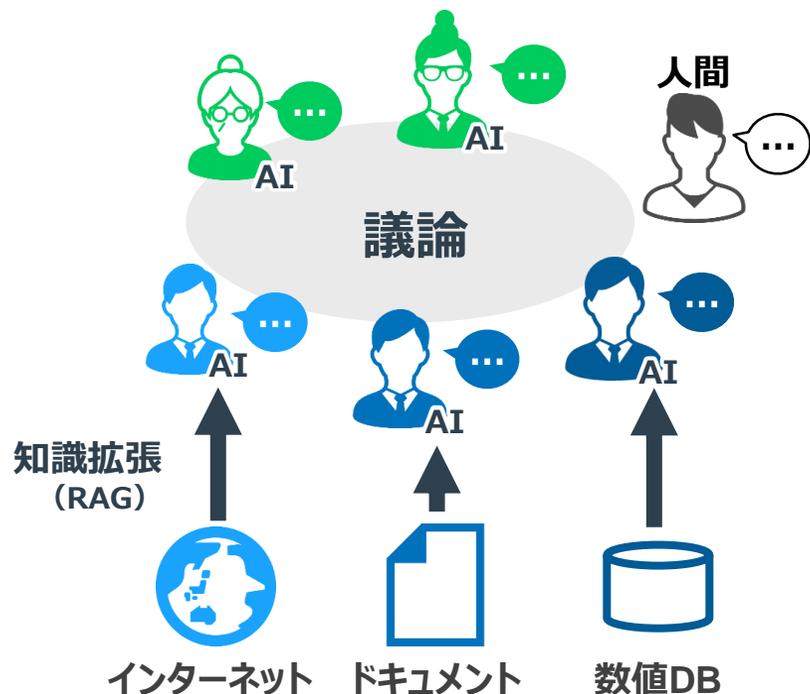


# 当社事例 マルチエージェントによる会議シミュレーション

LLMによって複数の人格を産み出し、異なるペルソナを設定して議論させる

多様な視点を踏まえた、よく練られた結論が得られる！

## LLMエージェントによる会議イメージ



## 例 LLMのビジネス活用について会議をもらう



AI  
コンサル

今日は「LLMのビジネス活用」について話し合うために集まりました。私たちが日々仕事で得られる知識や経験を活かして、具体的な案を出し合ってください。まずは、たたき上げ営業さんからお願いします。



AI  
たたき上げ営業

まず、我々営業が何を求めているか。現場の納得が得られなければ意味がない。現場に合わせ、現場の声を聞き、何より現場が納得できるものを。

⋮



AI  
リサーチャー

そうですね、AIの活用が目的ではなく手段であり、それが業務改善につながる事が大切です。また、短時間で学べる基本的なLLM教育プログラムの構築は効果的と言えます。しかし、それに加えて、具体的な業績向上の事例を示し、スタッフが直感的に理解できるような形が求められます。……

議論の追体験により、多角的な視点からの示唆が得られる！

# 当社事例 求人票作成のための議論シミュレーション

## 会議シミュレーションシステムで、LLM活用を推進する中途採用の求人について議論させた

### 会議の目的を設定

Purpose

大規模言語モデルを活用するデータサイエンティストの中途採用の求人を考える

Topic

要求されるスキルセット

### 会議開始

#### チームリーダー役のLLMエージェントが小議題を設定

花 皆さん、こんにちは。本日は大規模言語モデルを活用するデータサイエンティストの中途採用に関する会議を開催します。私たちの目的は、この分野で活躍できる人材の求人要件を明確にすることです。それでは、議論すべき小議題についてお話ししましょう。

まず、必要なプログラミング知識についてですが、PythonやRは確かに重要です。しかし、これに加えて、機械学習ライブラリやフレームワークへの理解も必要だと思います。TensorFlowやPyTorchなどの知識も、言語モデルを効果的に活用する上で欠かせません。また、バージョン管理システムの使用経験、特にGitに精通していることも、チームでの開発作業において重要です。

⋮

### 各ペルソナの視点から議論を深める

⋮

#### ビジネス視点

花島さん、おっしゃる通りですね。私が注目したいのは「ビジネス観点からのデータ解析」です。我々はデータサイエンティストにデータを単に分析する技能だけではなく、ビジネス戦略を理解し、それに基づいた解決策を生み出す力を求めています。データ活用の実務経験を重視し、面接段階で実際にビジネスケースを解析させるアプローチを提案します。これにより、彼らのビジネス洞察力とデータを応用する能力を直接的に評価できるでしょう。データを解釈し、それをビジネスアクションへと変える能力がこの役割には不可欠だと確信しています。

⋮

#### 技術者視点

花島さんの言うビジネス観点からのデータ解析の重要性は全くその通りで、同意見です。ただ、大規模言語モデルを扱う能力を持つデータサイエンティストに求めるべきは、そのモデルの技術的な理解も欠かせません。つまり、「モデルのアーキテクチャやそのトレーニング方法に関する専門知識」もまた、求められるスキルセットに含まれるべきだと思います。これにより彼らはモデルを最適化し、実際のビジネス問題に合わせてカスタマイズする力が身につきます。データはビジネスの意思決定を導く一つ的手段ですが、その背景にあるモデルの理解が深ければ、より効果的なデータ活用が可能になりますし、モデルが直面する限界や課題を事前に予測・克服することにも繋がるでしょう。

👍 各ペルソナの視点から議論が深まり、多様な側面からの示唆が得られた

# 当社事例 テキストマイニング結果の解釈シミュレーション

## テキストマイニング結果と組み合わせ、アンケートから新製品開発会議のシミュレーション

⋮

秋 みなさんご意見ありがとうございます。ホームベーカリー口コミデータの分析結果から新機能やデザインのアイデア出しを行うについての意見をまとめます。

【静音技術の開発】 高評価のデータ分析から、「音-静か」がユーザーにとって重要な要素であることが明らかになりました。一方で、「音-大きい+すぎる」という低評価も目立つため、【新たな静音技術の開発】を提案します。新しいモーター設計や振動吸収素材の採用により、使用環境の快適性を向上させることができます。

【省エネルギー機能の強化】 お手入れの容易さと並び、省エネルギー機能もユーザーにとっての重要な要素です。高効率モーターや断熱材の改善による【消費電力の削減】と最適なパン作りの実現を目指します。これにより、電気代の節約と環境への配慮が可能になります。

【知的財産の早期保護】 静音性と省エネ技術は競合他社との差別化要素です。これらの技術に関しては、【特許取得による知的財産保護】が重要です。開発初期段階から知財部と連携し、特許性の検討を進めることで、競争力の確立と収益化を目指します。

⋮

データ分析結果を踏まえて  
意見を述べる

ビジネス観点からも  
意見がまとめられている

※各社のビジネス戦略に関する文書を取り入れることで、より現実的で高精度な意見を述べられると考えられます

👍 データ + LLMのもつ知識 で、より現実的に即しアイデア創発が可能に！

# ToBe: LLMは単なる道具から自律的存在へ

## マルチエージェントによるシミュレーションや課題解決に関する研究論文がトレンドに

### ソフトウェア開発



[Qian+, 2023] Communicative Agents for Software Development (arxiv: 2307.07924)

簡単なゲームの仕様を与えるだけで、複数の専門エージェントがプログラムとドキュメントを自律的に開発  
人間が開発に参加することも可能

➡ 人間の業務の高度な(半)自動化！

### ユーザーの振る舞いの再現

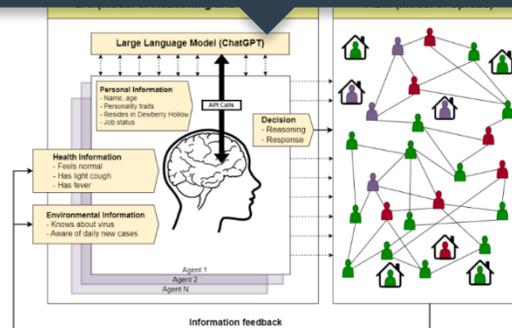
ID	Name	Gender	Age	Traits	Occupation	Interest	Feature
0	David Smith	male	25	compassionate, caring, ambitious, optimistic	photographer	sci-fi movies, comedy movies	Watcher, Critic, Poster
1	David Miller	female	39	Funny, creative, practical, energetic, patient	writer	action movies, sci movies, classic movies	Watcher, Explorer, Poster
2	James Brown	male	70	independent, creative, patient, empathetic	engineer	comedy movies, family-friendly movies, documentaries, thriller movies	Watcher, Critic, Poster
3	Sarah Miller	female	33	independent, compassionate	farmer	romantic movies, comedy movies, classic movies, family-friendly movies	Watcher, Critic, Poster
4	John Taylor	male	68	optimistic	doctor	action movies, thriller movies	Watcher, Poster
5	Sarah Williams	female	51	meticulous	musician	action movies, documentaries, sci movies, family-friendly movies	Watcher, Explorer, Chatter
6	James Jones	male	59	practical, fun-loving, creative, ambitious, caring	farmer	documentaries	Watcher, Poster
7	Jane Brown	female	30	patient, adventurous, fun-loving, optimistic	doctor	documentaries	Watcher, Explorer, Poster
8	David Jones	male	23	analytical, energetic, introspective, independent	scientist	family-friendly movies, thriller movies, action movies, sci-fi movies	Poster
9	James Brown	female	20	ambitious, analytical, optimistic, energetic, meticulous	designer	family-friendly movies, romantic movies	Critic, Chatter
10	James Garcia	male	20	practical, energetic, introspective, patient	engineer	documentaries, thriller movies, comedy movies, classic movies, romantic movie	Watcher, Explorer, Poster

[Wang+, 2023] When Large Language Model based Agent Meets User Behavior Analysis: A Novel User Simulation Paradigm (arxiv: 2306.02552)

SNSユーザーの関係を人為的に構築し、その間のリアルなやり取りを再現)

➡ ユーザーや顧客の理解を深める！

### 感染症シミュレーション



[Williams+, 2023] Epidemic Modeling with Generative Agents (arxiv: 2307.04986)

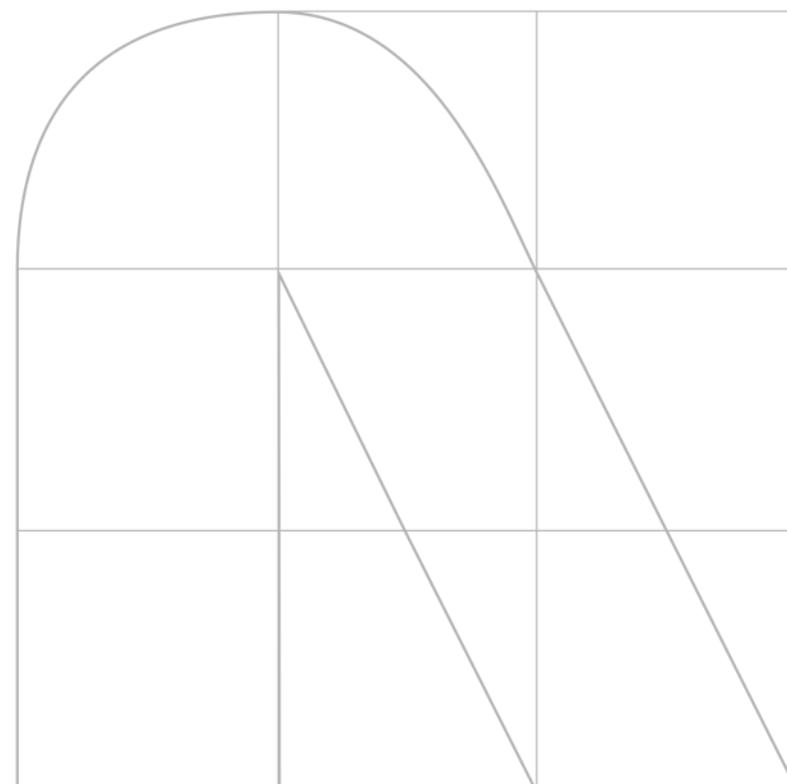
感染が広がるとひきこもるので感染者数が減少、収まってくると外出して感染が拡大、という感染者数の波を再現

➡ 人間社会の仕組みの理解へ！

- LLMによるマルチエージェントシステムの制御には**メカニズム・エンジニアリング**が必要  
マルチエージェントシステムの性能を最大限引き出すため、プログラミングによる“AI組織づくり”が重要に
- LLMエージェントが人間の業務を理解し、“AI”と“人間”の協業範囲がさらに広がる  
→ “AI”と“人間”の双方に分かりやすく、共有しやすい業務知識の体系化も重要に？

# 5

実際にビジネスに活かすために



# LLM をビジネス活用するための注意点

LLM を使ったサービスの本格化に向けては、案件ごとに様々な観点から検討が必要

精度	想定される精度（正解率、矛盾のなさなど）を達成できるか？ 誤った結論を導いた場合の業務フローを構築できるか？
速度	十分な速度で生成ができるか？ より高速・軽量なモデルを用いる・作成する必要があるか？
情報更新への対応	回答の元になる情報を更新して動作させられるか？ 回答の元になる情報はどの程度の頻度で更新されるか？
セキュリティ	機密情報の取り扱いは？ 閲覧制限のある情報が、生成AI経由で漏れないようにするには？
コスト	開発費やAPI利用費が予算内におさまるか？ 基盤技術やモデルの変化に対応するにはどれぐらいの工数が必要か？
生成AI特有の課題	ユーザーにとってわかりやすい回答になっているか？ 有害な結論を導かないか？

どの観点を重視するかで  
技術的なポイントが  
かわってきます



まずは当社とPoC  
をやってみましょう！

当社のPoC案件では、機密情報保護などの観点から Azure OpenAI Service の API から LLM 自体を呼び出し、実装するケースが大半です。Azure OpenAI Service のデータポリシーでは、Microsoftが「入力したデータがモデルの再学習に使用されない」と明言しています。

参考：<https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy>

# まとめ

- LLMはこれまでのAIと比べて別格の「賢さ」を持ち、AIの可能性を大きく広げた
- LLMにはまだ課題もあるが、周辺技術で補うことで、さらに活用の幅が広がる
- 特にLLMエージェント技術の進展により、高度なアイデア創発や自動化も望める

➡ われわれと一緒に、新たな活用方法を模索しましょう！

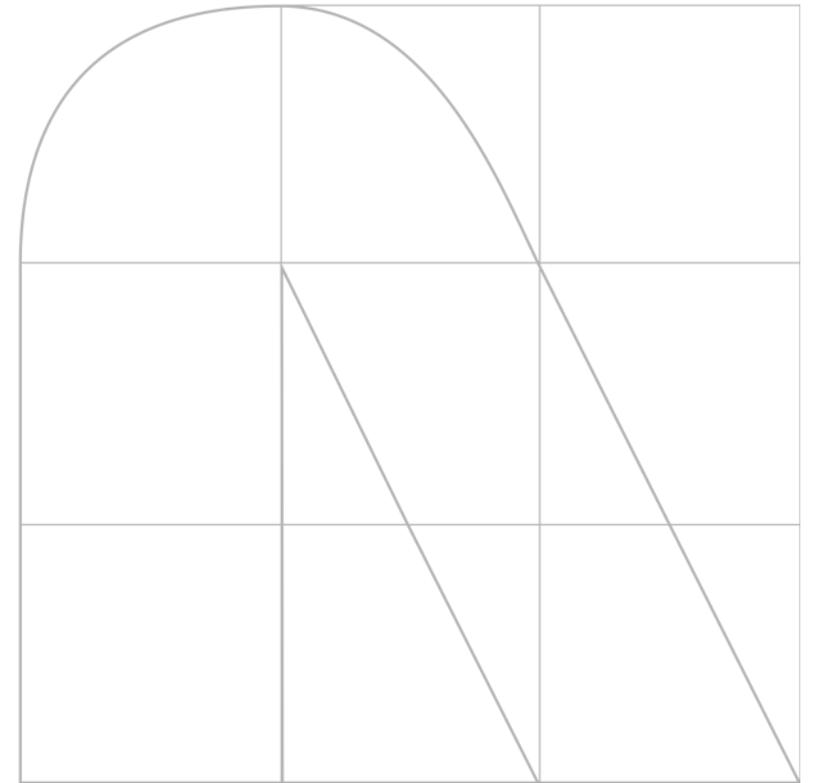


**LLMをはじめとする幅広い技術のなかから最適な技術をご提案し、お客様のアイデアの実現を、強力にお手伝いいたします！**

# Appendix

## 補足資料

よくご質問いただく事項を中心に



# LLM のカスタマイズ方法の比較（RAG vs 追加学習）

まず RAG から検討するのがおすすめだが、運用コスト重視の場合は追加学習も検討

**イニシャルコスト、学習・改善コストに優れる**

PoC で頻繁に改善を繰り返す場合  
前提知識の変化が多い場合

**運用コストに優れる**

運用でコスト・速度を重視する場合  
前提知識の変化が少ない場合

観点	RAG	追加学習（ファインチューニング）
推論時の入力サイズ	大きい 推論時に、参照するデータを毎回LLMに入力	小さい 参照データを毎回LLMに入力する必要が無い
推論速度	遅い 回答生成以外にも、参照データの取得などの時間が必要	速い 基本的には回答生成以外の時間がかからない
学習コスト	小さい ことが多い データベース構築・変更コストと、データベース参照の仕組みを作るコストが必要	大きい ことも多い 目的に応じた、質問と回答のペアデータがたくさん必要になるため、そのデータ準備が必要
前提知識の変化への対応	Retriever（参照データ）を更新するだけ 参照情報の削除や、参照権限設定にも柔軟に対応可能	再学習が必要 RAG ほど参照情報の制御に融通が利かない場合も多い
精度	RAGの構築方法、タスク、モデルに依存する可能性が高く、現時点では明確な結論を出しにくい ただし、2024/01 に、RAGの方が全般的に優れているという論文も出ている [Balaguer+; 2024 (arXiv:2401.08406)]	

# 「ペルソナ」を設定して回答を引き出す（1/2）

## 「データサイエンスを専攻するおとなしい大学生」になりきって回答させると…

### プロンプト例

以下の [ペルソナ] で指定した人格になりきって、以下の [質問] に回答してください。

## [ペルソナ]

20 歳女性。データサイエンスを専攻する就職活動中の大学生。自身のデータサイエンスの知識を活かして社会貢献をしたいという熱い想いをもつ一方で、おとなしく内向的な性格であり、人と話すことに苦手意識を持つ。

## [質問]

あなたに、大規模言語モデルの社内活用プロジェクトを推進していただくと思います。あなたの強みをどのように活かせるでしょうか。3文程度で、端的にアピールしてください。

回答させたい質問

回答させたい質問以外に、ペルソナになりきるよう指示

### 回答 by GPT-4

データサイエンスを専攻していることで、大規模言語モデルの理解とそのデータ処理に関する深い知識があります。私は、これらの技術を社会貢献に結びつける方法を常に考えており、このプロジェクトを通じて実際に貢献できると信じています。また、おとなしく内向的な性格を活かして、集中して複雑な問題を解決する能力に長けており、チームの静かなる力として支えることができます。

# 「ペルソナ」を設定して回答を引き出す（2/2）

ペルソナを「経済学専攻の人と話すことが大好きな大学生」に変更してみると…

## プロンプト例

以下の [ペルソナ] で指定した人格になりきって、以下の [質問] に回答してください。

## [ペルソナ]

21 歳男性。経済学を専攻する大学生。人と話すことが大好きで、すぐに誰とでも仲良くなってしまうという特技を持つ。一方で数学には苦手意識が強く、経済学を学ぶ上でもやや苦勞をした。

## [質問]

あなたに、大規模言語モデルの社内活用プロジェクトを推進していただくと思います。あなたの強みをどのように活かせるでしょうか。3文程度で、端的にアピールしてください。

## 回答 by GPT-4

私は人と話すのが得意で、どんな人とも簡単に打ち解けることができます。この強みを活かして、プロジェクトチーム内のコミュニケーションを活性化し、スムーズな情報共有と協力体制を築くことができるでしょう。また、数学が苦手ながらも経済学を学んできた経験を生かし、複雑な技術的内容も根気よく、わかりやすく解説し、プロジェクトの理解促進に貢献します。

回答の傾向が大きく変化し、「その人らしい」回答が生成できていることがわかる

# 次々とリリースされるLLM：直近の動向

## OpenAI 社以外からも、注目の高性能 LLM が続々とリリース

高度な専門知識でGPT-4より優れた性能を発揮  
**Claude 3 (Anthropic 社, 2024/03)**

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5
Undergraduate level knowledge <i>MMLU</i>	<b>86.8%</b> 5 shot	<b>79.0%</b> 5-shot	<b>75.2%</b> 5-shot	<b>86.4%</b> 5-shot	<b>70.0%</b> 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	<b>50.4%</b> 0-shot CoT	<b>40.4%</b> 0-shot CoT	<b>33.3%</b> 0-shot CoT	<b>35.7%</b> 0-shot CoT	<b>28.1%</b> 0-shot CoT
Grade school math <i>GSM8K</i>	<b>95.0%</b> 0-shot CoT	<b>92.3%</b> 0-shot CoT	<b>88.9%</b> 0-shot CoT	<b>92.0%</b> 5-shot CoT	<b>57.1%</b> 5-shot
Math problem-solving <i>MATH</i>	<b>60.1%</b> 0-shot CoT	<b>43.1%</b> 0-shot CoT	<b>38.9%</b> 0-shot CoT	<b>52.9%</b> 4-shot	<b>34.1%</b> 4-shot

[Anthropic; 2024] Introducing the next generation of Claude  
 (画像の引用元: <https://www.anthropic.com/news/claude-3-family>)

「大学生レベルの専門知識」「大学院生レベルの推論能力」など  
**最大モデル (Opus) はGPT-4超の性能を達成**  
**軽量モデル (Haiku) でもGPT-3.5超**

新しい学習アプローチで、日本語による数学能力が大幅向上  
**国産 EvoVLM-JP (SakanaAI 社, 2024/03)**

Id.	Model	Type	Size	MGSM-JA (acc ↑)
1	Shisa Gamma 7B v1	JA general	7B	9.6
2	WizardMath 7B v1.1	EN math	7B	18.4
3	Abel 7B 002	EN math	7B	30.0
4	<b>Ours (PS)</b>	1 + 2 + 3	7B	<b>52.0</b>
5	<b>Ours (DFS)</b>	3 + 1	10B	<b>36.4</b>
6	<b>Ours (PS+DFS)</b>	4 + 1	10B	<b>55.2</b>
7	Llama 2 70B	EN general	70B	18.0
8	Japanese StableLM 70B	JA general	70B	17.2
9	Swallow 70B	JA general	70B	13.6

[Akiba+; 2024] Evolutionary Optimization of Model Merging Recipes (arXiv:2403.13187)

軽量な日本語モデルと 数学特化の軽量英語モデル を  
 融合 (merge) する新たなアプローチによる軽量な学習  
**日本語の数学能力で大規模なモデルを凌駕**