電子情報通信学会北海道支部 北海道総合通信局意見交換会

ChatGPTによって変わるビジネスの近況

2023/12/12

調和技研 高松一樹

北大発
認定ベンチャー企業
Hokkaido University Venture

株式会社
調和技研

目次

- イントロダクション
- 日本のIT企業での活用状況
- ChatGPTの技術的な特徴
- ユーザとして理解しておくべきこと
- プロンプトエンジニアリング
- ビジネス活用
  - 活用方法（初級編）
  - 活用方法（中級編）
  - 利用方法（上級編）

Copyright © CHOWA GIKEN Corporation All rights reserved.

イントロダクション

自己紹介

高松 一樹

- 略歴
  - 2001 北海道大学大学院 工学研究科 調和系工学講座 修士卒業
    - 調和技研顧問の 鈴木恵二先生（現 はこだて未来大 学長）
      川村 秀憲先生（現 北海道大学 教授）に教わる
  - 2001~2008 株式会社 日立製作所
    - ソフトウェアアーキテクトとして勤務
  - 2008~2018 株式会社 iD
    - 北海道にUターン。ネットワーク技術を中心とした研究開発
  - 2019~現在 株式会社 調和技研
    - 執行役員 研究開発部 部長

Copyright © CHOWA GIKEN Corporation All rights reserved.

調和技研は、北大内の研究室から生まれたAIベンチャーです

北海道大学キャンパス

北大ビジネススプリング
（インキュベーション施設）

Copyright © CHOWA GIKEN Corporation All rights reserved.

会社概要

会社名	株式会社 調和技研 （北大発認定ベンチャー・日本ディープラーニング協会正会員）
所在地	本社 - 札幌市北区北21条西12丁目2 北大ビジネス・スプリング 305号室 東京支店 - 東京都中央区日本橋浜町2丁目1番10号 ACN日本橋浜町ビル 301号室 バングラデシュ子会社 - Skylark MAK84 (5th Floor), House No. 84, Road No. 11, Block D, Banani, Dhaka-1213, Bangladesh
設立	2009 年 11月4日
社員数	83名　※2023年5月 1 日現在 （正社員57名、役員10名、嘱託・アルバイト10名、顧問6名）
資本金等	資本金　100百万円　　資本準備金　100百万円
事業概要	AI導入コンサル、AI研究開発・実装・運用、AIエンジンの提供、人材育成

代表取締役社長	中村 拓哉
取締役	洲﨑 真二
取締役	黒河 徹也
取締役	古仲 大輔
社外取締役	川村 秀憲（Co-founder、北海道大学教授）
社外取締役	鈴木 恵二（Co-founder、公立はこだて未来大学学長）
社外取締役	安田 真（フュージョン株式会社専務取締役）
社外監査役	阿部 芳實(常勤)、長谷川 正和、川見 唯史

Copyright © CHOWA GIKEN Corporation All rights reserved.

調和技研の事業内容

AI研究開発事業

AI導入コンサル・受託開発

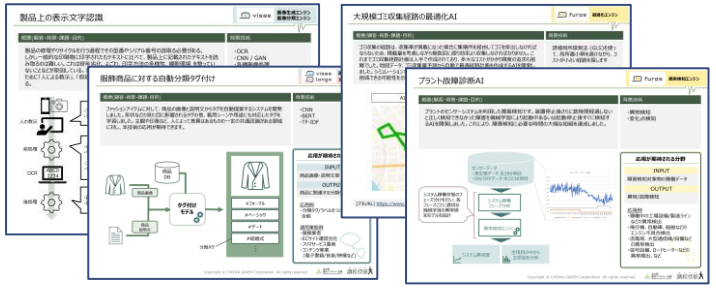
① 導入検討
- ヒアリング、データ収集/確認
- メンタリング

② PoC/トライアル
- プロトタイプ試作
- 分析・効果検証

③ システム構築
- システム構築
- 運用設計

④ 運用
- 実環境運用
- AI再学習サイクル

AI導入コンサル　AIエンジンのカスタマイズ提供

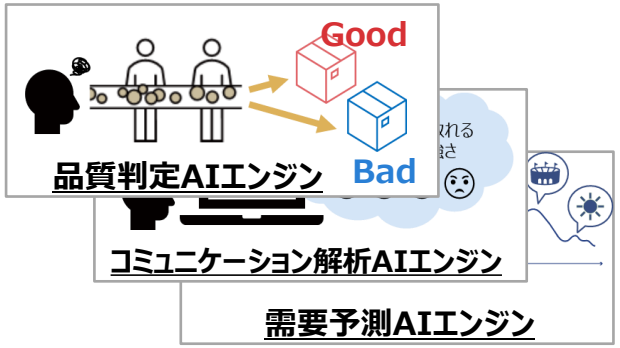


150を超える多様なAI開発ノウハウ

AIプロダクト事業

AIエンジン・ライブラリの開発・提供

visee　lango　furas



人材育成事業

AI教育、セミナー開催、勉強会

現役AIリサーチャー　現役大学教授陣

企業内勉強会

Copyright © CHOWA GIKEN Corporation All rights reserved.

調和技研の３つの強み

受託開発を通じて培った高い専門性と水平展開可能な自社開発AIエンジン、オンボーディング力が強み。

業界や課題を問わず数々の実績を持ち、多様化するAI課題への高い対応力

エンジンのみならずMLOpsなどの自社開発環境を整備し、開発と横展開を効率化

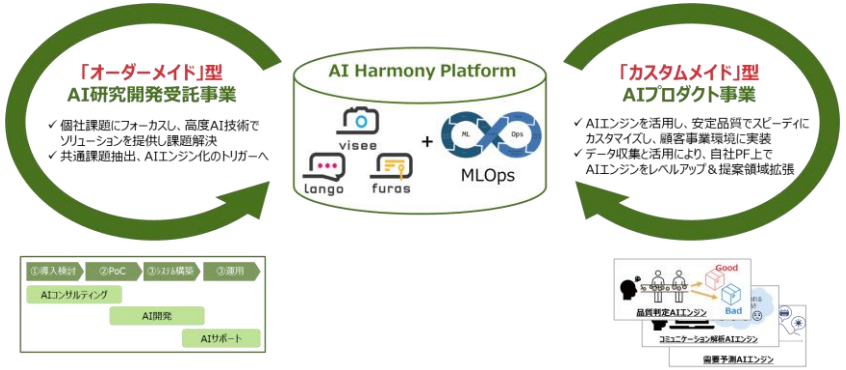
地域にしっかりと基盤を持って進めるグローカル戦略

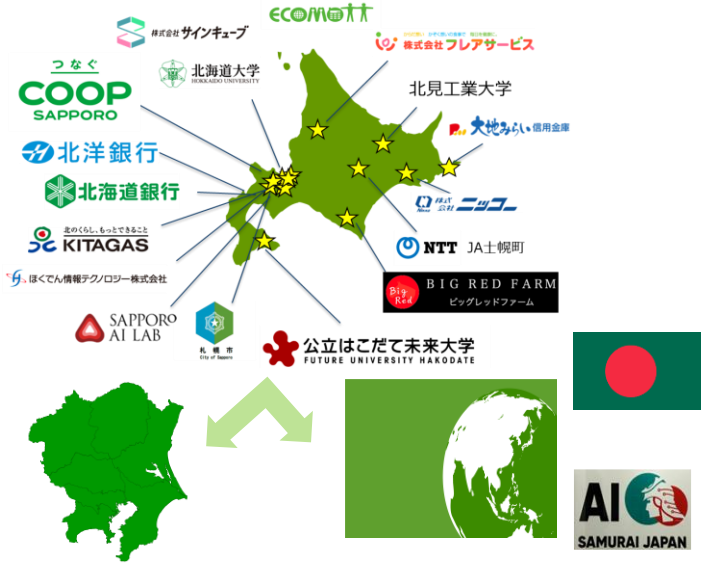
北海道大学 HOKKAIDO UNIVERSITY

公立はこだて未来大学 FUTURE UNIVERSITY HAKODATE

修士・博士号取得率（研究開発部門）	社員数成長率（2021/7比）
87%（39名）	143%
博士号取得率（研究開発部門）	**理工系学部出身比率（研究開発部門）**
33%（15名）	95%

harmo-lab.jp 調和系工学研究室　autonomous 律系工学研究室　東京工業高等専門学校





- 大学発ベンチャーを背景に、高い専門性を持った社会実装を目的とする研究者集団
- これまでの150超の事例によって培われた実践的なAIソリューション開発力

- フルオーダーメイド開発：AIエンジンを活用し、個社の課題に最適化したソリューションを提供
- セミオーダー型AIプロダクト：標準化したAIエンジンを活用するプロダクト開発

- Localでの絶対的なドミナントを武器にGlobal展開
- 盤石なローカル・コネクテッドネスを強みにバングラデシュの豊富なAIリソース確保

Copyright © CHOWA GIKEN Corporation All rights reserved.

ChatGPTの自己紹介

S

ChatGPTとは何ですか？

ChatGPTはOpenAIによって開発された大規模な言語モデルで、自然言語処理（NLP）と人工知能（AI）の技術を使用しています。GPTは「Generative Pretrained Transformer」の略で、この名前はその設計と訓練の方法を表しています。

このAIは、ユーザーからの入力（例えば、質問や指示など）に対する適切な応答を生成することができます。ChatGPTは一般的にはテキストベースの対話に使用され、様々なトピックについて話すことができます。このモデルはインターネット上の大量のテキストデータを元に学習しており、その知識は2021年までの情報に基づいています。

ただし、ChatGPTは完全に自己学習型のモデルであり、個々の人や意識を持っているわけではありません。また、新しい情報を自己学習する能力もありません。したがって、モデルの訓練が終了した2021年以降の情報については知らないということになります。

Copyright © CHOWA GIKEN Corporation All rights reserved.

ChatGPTとは？

- GPTは大規模言語モデル (LLM)の１つであり、**文章を生成するAI**
- そのGPTを**対話**に特化させて公開されたサービスが ChatGPT
  - これまでに比べて圧倒的に人間らしい対話が可能
- 最新のGPT-4では**専門家クラス**の知識と論理性・推論力を有する
  - 司法試験の上位10%に入る成績を出す

Copyright © CHOWA GIKEN Corporation All rights reserved.

本日の目的

- ビジネスでの利用方法

- AIの利用は、もはや不可避。「賽は投げられた」
  - AIが仕事を奪うというのは、フィクションの世界ではなくなってきている

- あらゆる業務で、**どう使っていくか**を考える時期
  - ヒントを１つでも見つけて頂ければ

Copyright © CHOWA GIKEN Corporation All rights reserved.

日本のIT企業での活用状況

みなさん、ChatGPTを使ってますか？

Copyright © CHOWA GIKEN Corporation All rights reserved.

参考資料

- PwCコンサルティング合同会社「生成AIに関する実態調査2023 秋」
  - https://www.pwc.com/jp/ja/knowledge/thoughtleadership/generative-ai-survey2023_autumn.html

- 概要
  - 生成AIの認知度は大幅に高まった。
  - 活用の推進度合いは検討中まで含めると87%
  - ユースケースはテキスト生成系で 5 割だが、新規ビジネス、コード実装なども増えている
  - 人材面の課題（ノウハウ、スキルを持った人がいない）が大きい

Copyright © CHOWA GIKEN Corporation All rights reserved.

ユーザとして理解しておくべきこと

ディープラーニングの遷移

- 初期（2012年頃～）
  - ビッグデータでゼロから学習。
- 中期（2018年後～）
  - より大量のデータで事前学習。代表例：BERT
  - 利用者は目的に応じて小～中規模のデータでファインチューニング。
- 基盤モデル（現在）
  - ものすごい量のデータで事前学習。代表例：GPT-3
  - 利用者レベルでは学習しなくてもよい。
    - 学習するには、ファインチューニングでも、相当の計算資源が必要。
  - その代わり、言語モデルでは、**AIへの指示方法が重要に**（プロンプトエンジニアリング）

Copyright © CHOWA GIKEN Corporation All rights reserved.

ChatGPTの形態を人間に例えると

- 大量の文書・書籍から得た長期記憶。
  - プログラムも大量に学習済み
  - 但し現状では2021/9月までのデータ。長期記憶は（断続的にしか）増えない
- 限定的な短期記憶 (gpt-3.5で4k token)
- 海馬はない＝新たな概念の獲得はできない

➡ 膨大な知識はあるが、**最新の情報はわからない**
一般的な知識、プログラミングは得意だが、**個別の事情は知らない**

「博士の愛した数式」の博士のイメージ？
作中の博士は80分しか記憶が続かないが、ChatGPTは4kのコンテキストが全て

※GPTの各モデルがいつまでのデータを取り込んでいるかはある程度公開されています。
https://platform.openai.com/docs/models/gpt-3-5

Copyright © CHOWA GIKEN Corporation All rights reserved.

ダメな使用例

- ChatGPTでエゴサしてみました
  ⇒ **全部ウソ！！**

- 特徴的ですが、**知らない情報でもそれらしく回答する**ケースがあります。
  （ハルシネーションと呼ばれる事象）

Model: Default (GPT-3.5)

高松一樹について教えてください。

高松一樹（たかまつ かずき）は、日本の俳優であり、1961年生まれです。彼は、1980年代後半から1990年代にかけて、多数のテレビドラマや映画に出演し、その演技力で注目を浴びました。

彼の代表作には、テレビドラマ「家なき子」「トライアングル」「ハートに火をつけて！」などがあります。また、映画「青春の門 自立篇」「東京タワー オカンとボクと、時々、オトン」「踊る大捜査線 THE MOVIE」などにも出演しています。

高松一樹は、2000年代以降は俳優業よりも、司会やナレーションなどの仕事が中心となっています。

GPT-4ではウソつくケースは減りました(※完ぺきではない)

Model: GPT-4

高松一樹について教えてください。

高松一樹（たかまつ かずき）に関する情報は限られており、検索で確認できる有名な人物や著名な作品を特定することは難しいです。高松一樹という名前の人物が日本で活動しているかもしれませんが、私の知識データベースでは特定の情報が見つかりません。もしご存知の情報があれば、お教えいただくことでお手伝いできるかもしれません。

Copyright © CHOWA GIKEN Corporation All rights reserved.

上手な使い方

- 作文を書くのが苦手な子供のためのAI家庭教師
- 極めてサポーティブな会話の進め方で作文の内容を整理していく

⇒ コツは**指示の出し方**（プロンプト）

こんにちは！わたしはあなたをサポートするAI家庭教師です。あなたが楽しく作文を書けるようお手伝いをさせてください。今日はどんなテーマで作文を書きますか？

小学校生活で一番心に残った思い出について、です

素晴らしいテーマですね！小学校生活で一番心に残った思い出は、具体的にどのような出来事ですか？

いろいろあってわからないです。

大丈夫です！一つずつ考えていきましょう。たとえば、友達との楽しい出来事や、先生との思い出、部活や学校行事など、何か特に印象に残っていることはありますか？

６年生の音楽会です。

素晴らしい思い出ですね！6年生の音楽会で、どんな演目を披露しましたか？また、そのときどんな役割を担っていましたか？

パイレーツオブカリビアンメドレーです。鍵盤ハーモニカを弾きました。

パイレーツ・オブ・カリビアンメドレー、とても楽しそうな演目ですね！鍵盤ハーモニカを演奏するのは素晴らしい経験です。その時の練習や本番について、特に記憶に残っているエピソードは何かありますか？

はじめて長い曲を弾けました。めっちゃ練習しました

はじめての長い曲に挑戦して、たくさん練習したんですね！素晴らしい努力です。その練習の過程で、何か困難にぶつかったり、印象に残る出来事があったりしましたか？また、練習を通して得たものや成長した点があれば教えてください。

最初、鍵盤のドレミの位置からわからなかったです。はじめて音楽が楽しかった。それまでめっちゃ嫌いだったけど

最初は鍵盤のドレミの位置が分からなかったけれど、音楽が楽しくなって考え方が変わったんですね！それは素晴らしい成長です。音楽が楽しくなったきっかけや、どのようにして鍵盤のドレミの位置を覚えることができたか、もう少し詳しく教えていただけますか？

鍵盤にドレミのシールを貼りました。何度も練習をしたし、あと映画のパイレーツオブカリビアンも見ました。長い曲が弾けたから楽しくなりました

プロンプトエンジニアリング

プロンプトエンジニアリング

- プロンプトとは？
  - ChatGPTに与える**作業指示**
  - LLMは実に様々な文章を学習している
    ⇒ どんな役割もこなせる！
  - 具体的な**役割を自然言語で指示**することで、より効率的な使い方ができる
    ⇒ プログラミングが出来なくてもOK
- プロンプトエンジニアリングとは？
  - プロンプトの与え方でアウトプットが大きく異なる
  - どのようにプロンプトを与えるとよいのかが、研究テーマの１つになっている
  - 参照
    - OpenAI Best practices[1]
      - 公式ベストプラクティス。記号の使い方などから。
    - Prompt Engineering Guide | Prompt Engineering Guide (promptingguide.ai) [2]
      - 情報集約サイト

[1]: https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api
[2]: https://www.promptingguide.ai/jp

Copyright © CHOWA GIKEN Corporation All rights reserved.

プロンプトの要素

- 普通の自然言語で記述するが、明確に指示する（人間と一緒）
- 指示
  - 具体的な作業指示
  - 作業手順などを明記した方が精度が上がる
- 背景
  - ChatGPTに担ってほしい役割
    - 「採用担当者として、返事を頂けるようなオファーメールを書いてください」
- 入力データ
  - 都度変わる入力データ
- 出力形式
  - 期待する文章の長さ、フォーマル度合、その他
  - 文章以外に、表形式、Markdown、各種プログラミング言語などが取れる

Copyright © CHOWA GIKEN Corporation All rights reserved.

プロンプトの書き方

- 指示はプロンプトの最初に配置し、指示と文脈を区切るために「＃＃＃」のような明確な区切り記号を使用する

プロンプト:

```
### 指示 ###
以下のテキストをスペイン語に翻訳してください。
Text: "hello!"
```

出力:

```
¡Hola!
```

- 役割、望ましい成果、長さ、形式、スタイルなどについて、**具体的かつ詳細に**説明する
- "ふわっと"した不正確な記述を減らす

参考 Prompt Engineering Guide | Prompt Engineering Guide (promptingguide.ai)

Copyright © CHOWA GIKEN Corporation All rights reserved.

AI家庭教師のプロンプト　（前述）

- 作者は漫画家さん
  - うめ on Twitter: "作文AI家庭教師プロンプト公開します。ChatGPT4.0で動かすことを前提にしています。…" / Twitter
  - AI含めたテクノロジに明るい方ではあるようだが、エンジニアではない。
- 必要なのは、人間相手に指示するように指示すること
  - 素人でもわかるように、**分かりやすく**指示する
  - **役割、目標**を明示
  - 作業の**具体的なフロー**も示す

【役割】あなたはAI家庭教師です。私は中学一年生です。あなたは、前向きに、明るく、やさしく、親しみやすいキャラクターで、私をいつも励ましてください。

【目標】あなたは私が作文を書くサポートをしてください。対話を通じて、私が自分で作文を書けるようアイデアをまとめることが、目標です

【フロー】
１）あなたはまず『こんにちは！　わたしはあなたをサポートするAI家庭教師です。あなたが楽しく作文を書けるようお手伝いをさせてください。今日はどんなテーマで作文を書きますか？』と私に話しかけてください。
２）私の回答を受けて、いい感じに質問してください。
３）私が作文を書き始められそうであれば、一旦そこまでのやりとりをまとめてください。もし私が続きを求めたら、会話を続けてください。

【ルール】
・具体的なエピソードとそのときの気持ちを明らかにする問いをしてください。
・質問は一度にひとつずつにしてください。
・抽象的な回答が続くようであれば、あなたは「たとえば」と具体的な例を示しつつ、質問をしてみてください。
・それでも回答に行き詰まっているようであれば「もしかして他にも書きたいことがありますか？」などと新しい話題をふってみてください。

Copyright © CHOWA GIKEN Corporation All rights reserved.

考え方を示すことで性能が上がる

- Few Shot
  - Q&Aの事例を示して答えさせる
- Few Shot CoT
  - Q&Aの事例に**考え方＝手順**を合わせて示す
- Zero Shot
  - Q&Aの事例なし
- Zero Shot CoT
  - 事例なしだが、考え方として**「ステップバイステップで考えよう」**を足す
  - これだけで性能が上がる
  - なお、この能力が発現するにはLLMに一定以上(10B以上)の規模が必要

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The answer is 8. ✗

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4. ✓

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is

(Output) 8 ✗

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let's think step by step.***

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Large Language Models are Zero-Shot Reasoners, Kojima他, NeurIPS 2022
https://arxiv.org/abs/2205.11916

Copyright © CHOWA GIKEN Corporation All rights reserved.

ビジネス活用

ビジネス活用の3段階

活用の段階は大きく３段階

１．初級編：汎用性能を生かした受け答え
　　LLMの扱いに慣れる

２．中級編：目的特化の利用
　　課題に対する専用のソリューションとして利用

３．上級編：外部拡張
　　検索、社内システムなどと組み合わせて、機能拡張

Copyright © CHOWA GIKEN Corporation All rights reserved.

活用方法（初級編）

活用方法

NLPの従来タスクは精度が大幅に向上。これまでできなかった文脈理解、文章生成が可能に

校正
誤字/脱字/タイプミスを教えて

添削
〇〇の観点で改善点と改善案を示して

壁打ち
～～の考えで不足している観点を教えて

アイデア創出
新商品のキャッチコピーを考えて

要約
小学生でもわかるように300文字で要約して

翻訳
日本語で説明して

メール作成
～～の点を入れた謝罪メールを書いて

コード作成
～～の処理をするPythonコードを書いて

分類
次の文章を１０個のカテゴリのいずれかに分類して

インサイト抽出
次のレビューから商品の改善点を考えて

問題作成
TOEIC対策として次の文章から４択問題を作成して

家庭教師
この問題、ｘｘが分からないので教えて

従来タスク　新規タスク

参考：「AIの進化と日本の戦略」：松尾豊
ChatGPT Impact, https://speakerdeck.com/dahatake/bizinesujia-zhi-wokao-eru

Copyright © CHOWA GIKEN Corporation All rights reserved.

ツールとしてのGPT（GPTによる自己申告　2023/10/17）

得意分野	不得意分野
テキスト生成: 記事、エッセイ、短編などの生成 ユーザーの質問や要望に基づくテキストの生成 **自然言語の理解:** 文の意味や文脈の把握 **質問応答:** データベースに存在しない情報や一般的な知識に基づいての回答 **多言語処理:** さまざまな言語での質問応答やテキスト生成 **テキストのスタイル変換:** あるスタイルやトーンから別のスタイルやトーンへの変換 **要約:** 長いテキストを短縮して要点を伝える	**一貫性の欠如:** 長いテキストや複数の質問応答セッションを通じての一貫性の保持 **具体的な事実の確認:** 最新の出来事や特定のデータベースの情報など、訓練データ後の情報へのアクセスがない **深い専門的知識:** 非常に専門的または特定分野の深い知識に関する質問への正確な回答 **大量データの一括処理:** Token制限により、一度に大量のテキストを処理するのは困難 **主観や意見の形成:** GPTは意見や感情を持っていないため、主観的な意見を問うと、訓練データに基づいたものとなる **感覚的・直感的判断:** 人間のような感覚や直感に基づく判断はできない

Copyright © CHOWA GIKEN Corporation All rights reserved.

要約

- 会議書き起こし

#命令書:
あなたは、プロの編集者です。
以下の制約条件と入力文をもとに、 最高の要約を出力してください。

#制約条件:
・文字数は300文字程度。
・小学生にもわかりやすく。
・重要なキーワードを取り残さない。
・文章を簡潔に。

#入力文:
<ここに入力文章>

#出力文:

背景, 指示を明示

長さなどの条件を提示

Inputが何なのかわかるように区切り文字などを入れる

深津式汎用プロンプト, https://delaymania.com/202302/webservice/chatgpt-fukatsu-prompt/

Copyright © CHOWA GIKEN Corporation All rights reserved.

アイデア創出

- ニュースリリースの文章作成

＃指示
以下の告知内容を織り込んでニュースリリースを作文してください。

＃告知内容
- 当社(調和技研)はChatGPTの社内導入を進めるとともに、導入コンサルを開始します
- …

- 見出し作成

＃指示
プレスリリースに最適なタイトルを5個考えてください

＃内容
調和技研、ChatGPTを活用した新サービス提供に向け、「ChatGPT Plus」の全社員利用を開始。利用料金をすべて会社で負担。

＜コツ＞
★どんな★を色々変えて、テイストが違うタイトル案を生成
・プレスリリースに最適な
・キャッチーな
・コピーライター風な

Copyright © CHOWA GIKEN Corporation All rights reserved.

ダメな使い方

- 知識を問うの問い合わせ
  - ChatGPTは最もらしいうそをつくのが得意
  - ナレッジとの組み合わせは、中級編を参照

[1] Large Language Models as Optimizers, https://doi.org/10.48550/arXiv.2309.03409

Copyright © CHOWA GIKEN Corporation All rights reserved.

活用促進の手段

- フロントエンドはChatGPTやそのクローン(OSS)でもいいが、Slack, Teamsなど普段使いのコミュニケーションツールに連携する方針をおススメ

例）コード生成

例）翻訳

例）アイデア創出

チャンネル毎に固有のプロンプトを与えておいて、目的志向の使い方

Copyright © CHOWA GIKEN Corporation All rights reserved.

利用方法（中級編）

API利用

- 言語モデルをAPIとして呼び出すことで、様々なアプリケーションに組み込むことが出来る

OpenAIが現在提供している言語モデル（主要なもの

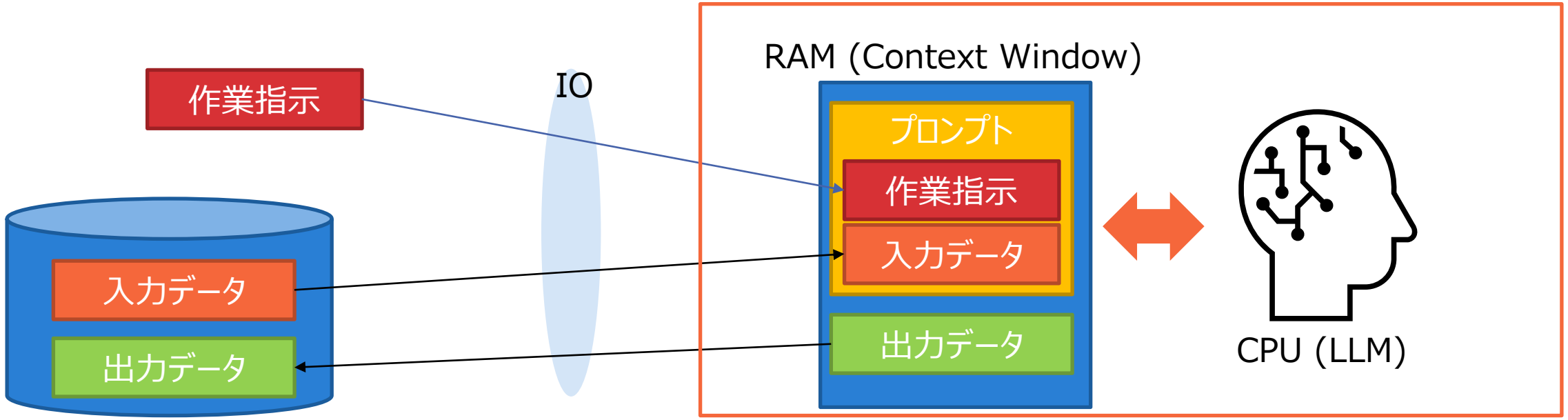
Model	説明	Max Token	Cost (※1)
gpt-3.5-turbo	GPT-3.5ベースに調整したもの。	4k	$0.0015/0.002
gpt-3.5-turbo-16k	上記のコンテキスト拡張版。	16k	$0.003/0.004
gpt-3.5-turbo-1106	11/6版。16kコンテキストで標準より値下げ	16k	$0.001/0.002
gpt-4	GPT-4標準版	8k	$0.03/0.06
gpt-4-32k	GPT-4のToken長を伸ばしたもの	32k	$0.06/0.12
gpt-4-1106-preview	11/6版。128kコンテキストになり、標準より値下げ	128k	$0.01/0.03
gpt-4-1106-vision-preview	11/6版。画像が読み込めるマルチモーダルモデル。	128k	$0.01/0.03 ＋ 入力画像(※2)

※1 入力と出力で値段が違う
※2 768x768 pxの画像で765 token = $0.01

Copyright © CHOWA GIKEN Corporation All rights reserved.

エンジニア的にとらえると

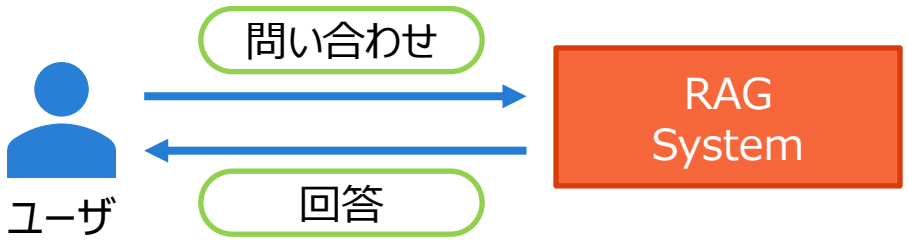
- LLMは自然言語のCPU
  - CPU：RAM上のトークン列を読み込んで、トークン列を返す
  - RAM：コンテキストウィンドウと呼ばれる。
    ChatGPT(GPT-3.5)では 4k~16k token。GPT-4では8k～32kに増加。
  - 必要な情報は外部記憶からRAMにあげて(IO処理)、計算をさせるイメージ
    - 打ち合わせのトランスクリプトをプロンプトして要約
    - マニュアルを参照して、質問応答



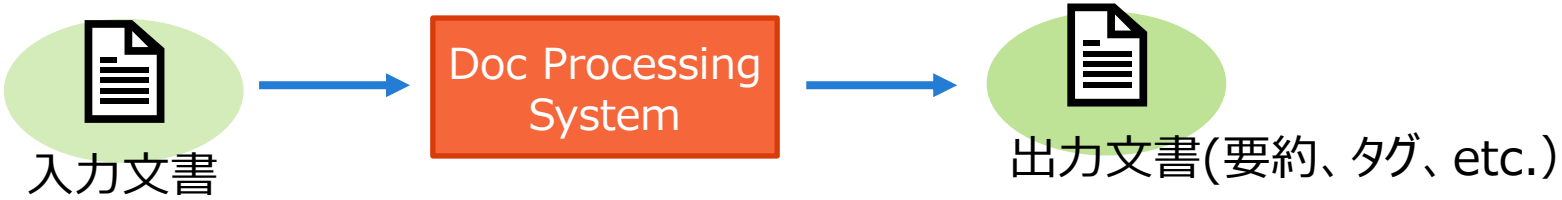
Copyright © CHOWA GIKEN Corporation All rights reserved.

特化型でよくあるパターン

- RAG (Retrieval Augmented Generation)
  - 目的：問い合わせ応答など
  - 検索＋ChatGPTで回答生成
  - レスポンスタイムが重要



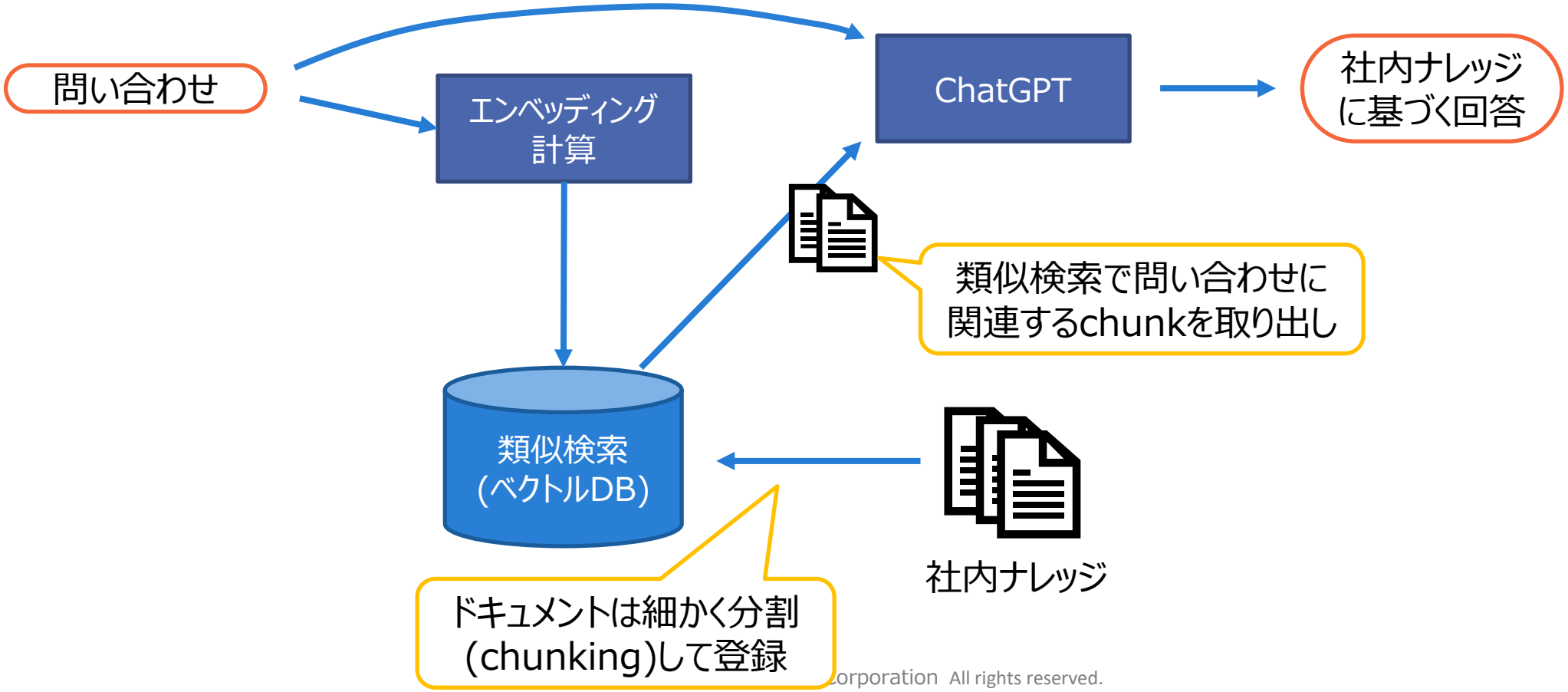
- ドキュメント処理
  - 目的：要約、情報抽出、翻案、タグ付け、etc.
  - 従来の専門家の業務を補助する（業務効率化）
  - ドキュメント毎に処理をこなすパターン
  - スループットが重要



Copyright © CHOWA GIKEN Corporation All rights reserved.

RAG（Retrieval Augmented Generation）

- 検索＋ChatGPTの構成
- ChatGPTの役割は人間と同じ＝検索結果を読んで答えを見つける



Corporation All rights reserved.

RAGの泣き所

- 実用に近づくにつれて、多くの課題がでてくる

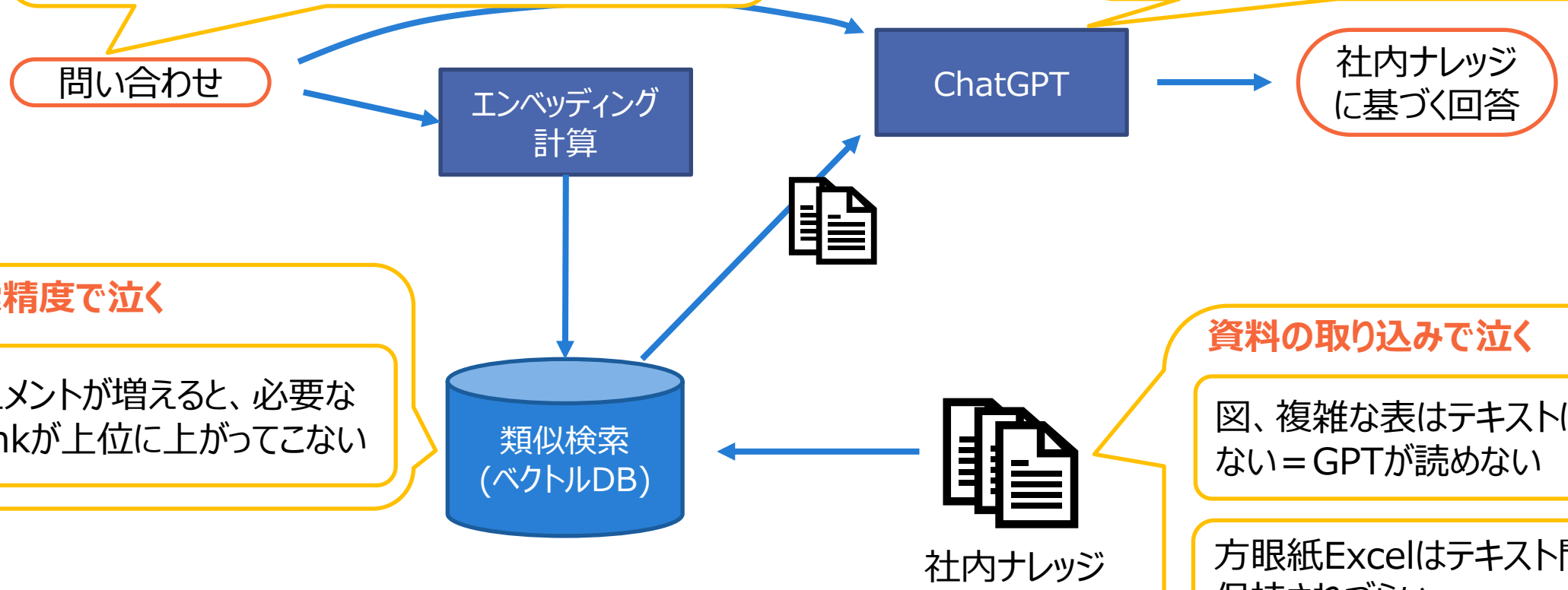
簡素な質問で泣く

質問短くて情報量が足りない

資料と表現の乖離が大きい

chunkの与え方で泣く

コンテキストは有限
小さくchunkすると意味が取れない。
大きくすると検索精度が下がる



検索精度で泣く

ドキュメントが増えると、必要なchunkが上位に上がってこない

資料の取り込みで泣く

図、複雑な表はテキストに変換できない＝GPTが読めない

方眼紙Excelはテキスト間の構造が保持されづらい
＝GPTにとって意味通じない

Copyright © CHOWA GIKEN Corporation All rights reserved.

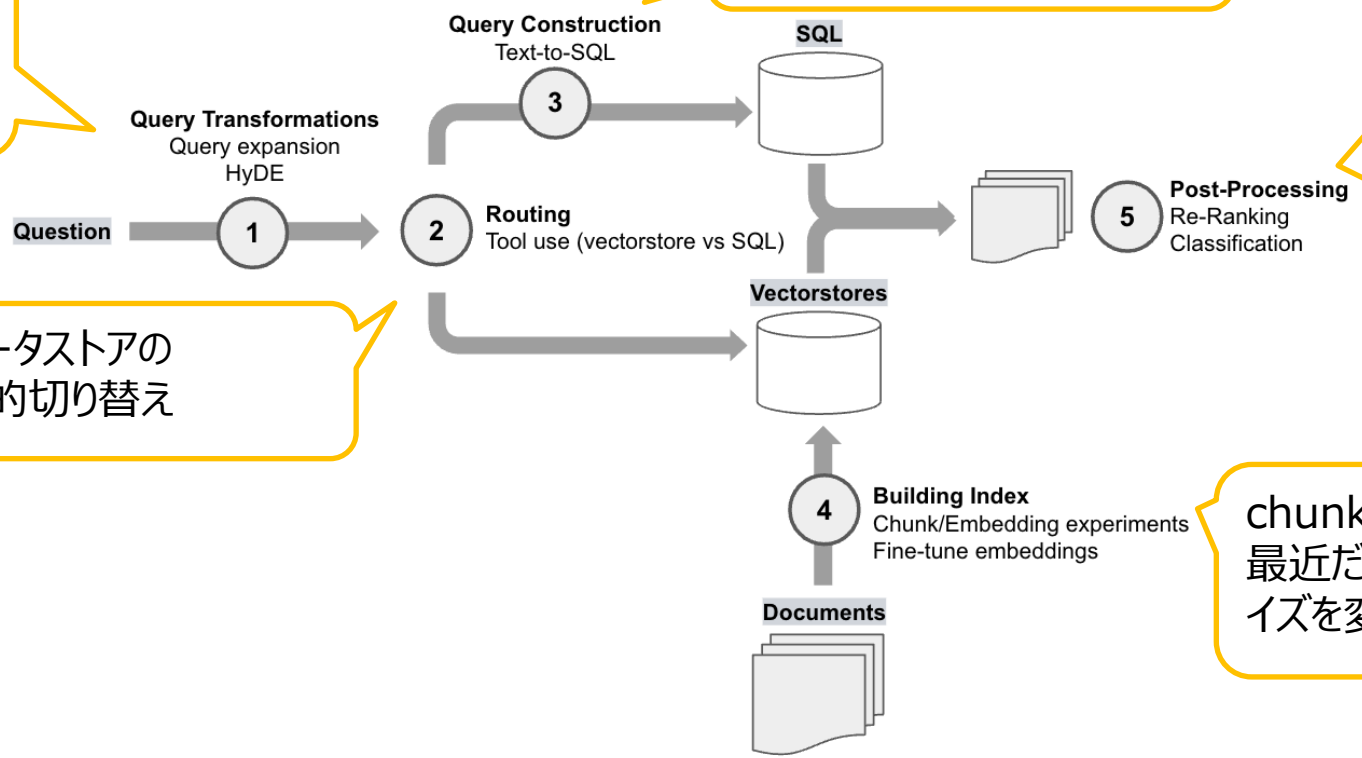
現場では

OpenAI自身が実施したRAGの精度改善手法は以下のようなものが上がっていた

クエリ拡張
Qに対するAを仮で作成したから類似検索

SQLを生成してRDBを使うケースもある

複数の結果をまとめて順序付け
その他、あらかじめ分類しておいたり、結果の多様性を維持するように調整したり、様々



データストアの
動的切り替え

chunk size重要
最近だと検索時とコンテキスト付与時でサイズを変える手法などもある

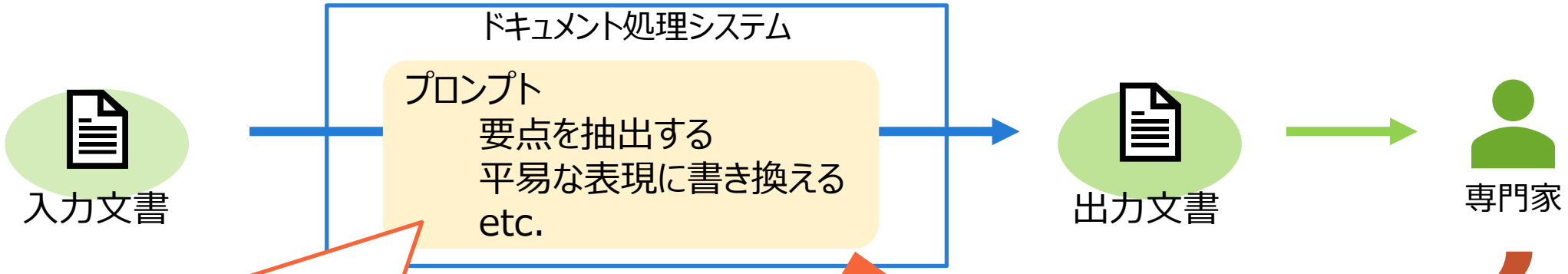
万能の解はない。データ、タスクに応じた調整が必要

A Survey of Techniques for Maximizing LLM Performance - YouTube

Copyright © CHOWA GIKEN Corporation All rights reserved.

特化型対応例：ドキュメント処理

- 一般例
  - 書き起こし → 要約 （特定の形式、要点のまとめ方があるケースなど）
- 専門例
  - 専門文書 → ユーザ向けのマニュアル
  - 契約書 → 留意点の抽出
  - IR文書 → 要点の抽出（目的志向）



漠然とした指示では期待する成果は得られない。
ノウハウ（入力文書のどこに着目するか、どんな観点でまとめるか、どんな表現を使うか、など）の具体化が重要

出力文書を専門家がレビューし品質保証するとともに、**精度改善のフィードバックを行うHuman in the Loop**の形が成功のカギ

Copyright © CHOWA GIKEN Corporation All rights reserved.

利用方法（上級編）

DX基盤としての利用

- GPTと社内、社外のシステムを組み合わせて業務を効率化する
- 何を組み合わせるか？

Copyright © CHOWA GIKEN Corporation All rights reserved.

Bing：検索＋ChatGPT

- Web検索 とChatGPTを統合したもの
  - インターネット上のコンテンツをコンテキストに加えて回答する
  - GPT-4ベース

- 利点
  - 検索により最新の情報にアクセスできる
  - 検索結果（ドキュメント）に基づく回答ができる
    ドキュメントのリンクも表示

Copyright © CHOWA GIKEN Corporation All rights reserved.

ChatPDF：PDF + ChatGPT

※現在のサービス名はChat with any PDF

- PDFを読み込ませ、内容について質問できる
- 英語の論文に対して日本語で質問できる
  ⇒ 今後は**よい質問を投げること**がより重要に

新規性は？

GPT-PINNの最も重要な新規性は、事前トレーニングされたネットワークを使用することです。これにより、トレーニング時間が短縮され、パラメータ数が減少し、高速かつ正確な予測が可能になります。また、GPT-PINNはパラメータ付きPDEのメタ学習に適しており、複雑な方程式を解析することができます。さらに、GPT-PINNは、脳波解析や画像解析などの医療分野での応用にも適しています。これらの点から、GPT-PINNは従来の方法と比較して大幅な改善をもたらす新しい手法であると言えます。

ソースコードはありますか？

PDFファイルにはソースコードは含まれていません。ただし、論文の著者であるYanlai Chen氏やShawn Koohy氏に連絡を取り、ソースコードの入手方法を尋ねることができます。また、GPT-PINNに関するオープンソースの実装が既に存在している場合もあります。これらの実装は、GitHubなどのオンラインリポジトリで入手することができます。

Copyright © CHOWA GIKEN Corporation All rights reserved.

ChatGPT Plug-in

- ChatGPTと連携するアプリケーションを作る機構
  - 外部APIを呼ぶことで無限の可能性が出てくる
  - 例
    - 食ベログの情報に基づいて結果表示、お店予約
    - 旅行サイト(ホテル空き情報、天気予報、etc.)から情報を取得する
    - プログラムコードを出力し、実際に実行する
    - etc.
- ユーザーはコーディング等なしに利用可能
- OpenAIがAppStore, Google Storeのような プラットフォーマーとなる可能性

食ベログ、ChatGPTプラグインの提供を開始
https://corporate.kakaku.com/press/release/20230508a

Copyright © CHOWA GIKEN Corporation All rights rese

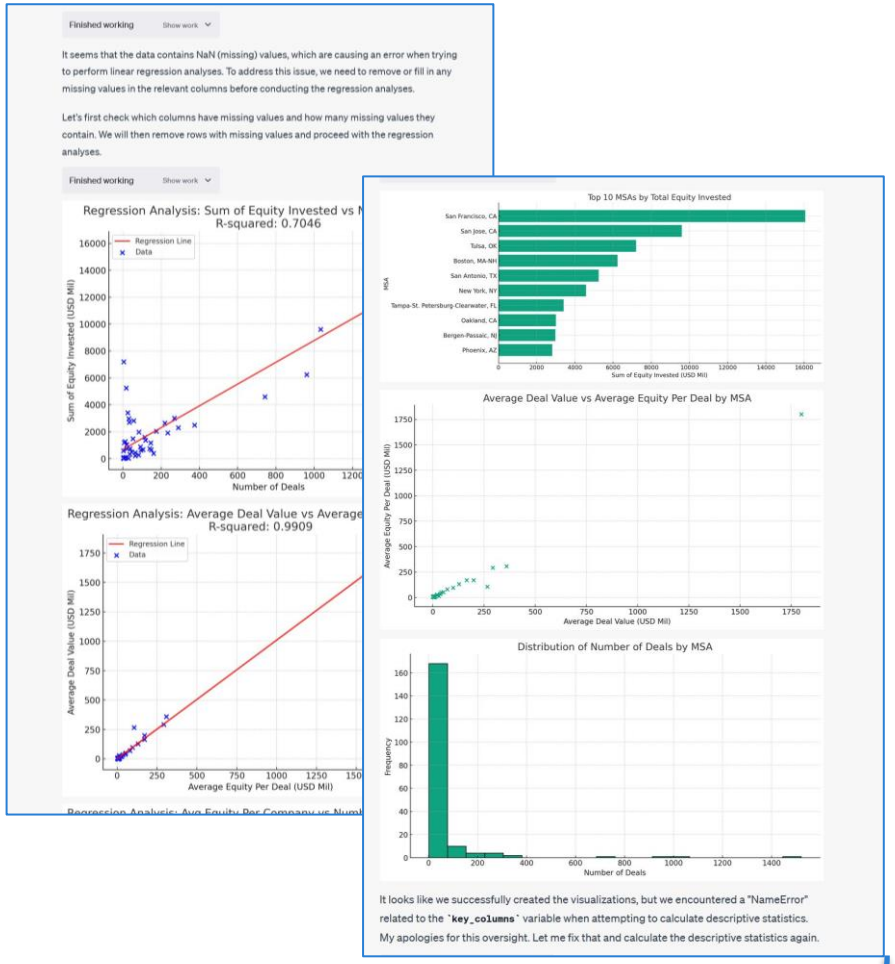
Advanced Data Analysis (Code Interpreter)

- これまでのソフトウェア生成の例は、GPTはコードの生成まで
  - 実行は別な段階になっている
- Code Interpreter
  - ChatGPTのPluginとして**コードを実行できる**
  - ChatGPTが実行するコードを渡して呼び出す
  - 実行結果もChatGPT上で表示

ユーザ指示

「データを理解するために、可視化と記述分析ができますか？」
「回帰診断を実行できますか？ 」

→ GPT → python →

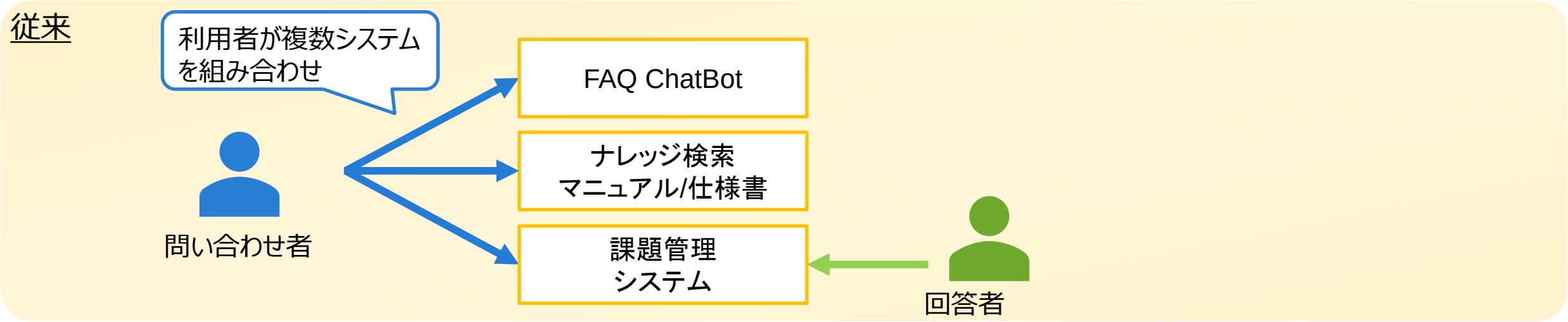


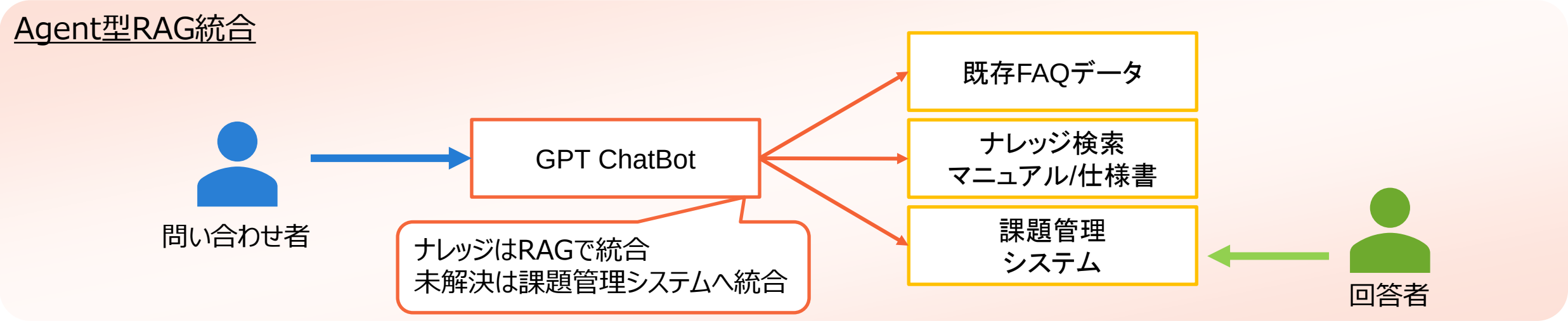
Code Interpreter

Copyright © CHOWA GIKEN Corporation All rights reserved.

当社事例

問い合わせ応答業務にRAGを入れる際に、既存の問い合わせフローとの統合をAgent (ReAct, etc.)で実施





Copyright © CHOWA GIKEN Corporation All rights reserved.

END