

特別セミナー

# 生成AIとテキストデータでマーケティングはこう変わる データドリブンな仮説立案へ

2025年1月22日（水）

# 0 | 当社について

# NTT DATA 株式会社NTTデータ数理システム

「**数理学とコンピュータサイエンスにより現実世界の問題を解決する**」をミッションに

**累計8,000**件以上のプロジェクトに真摯に向き合い

技術力を研ぎ澄ましてきた、**数理学のプロフェッショナル集団**です

## 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130名（うち75%以上が技術者）

## 主要事業

### パッケージソフトウェア開発

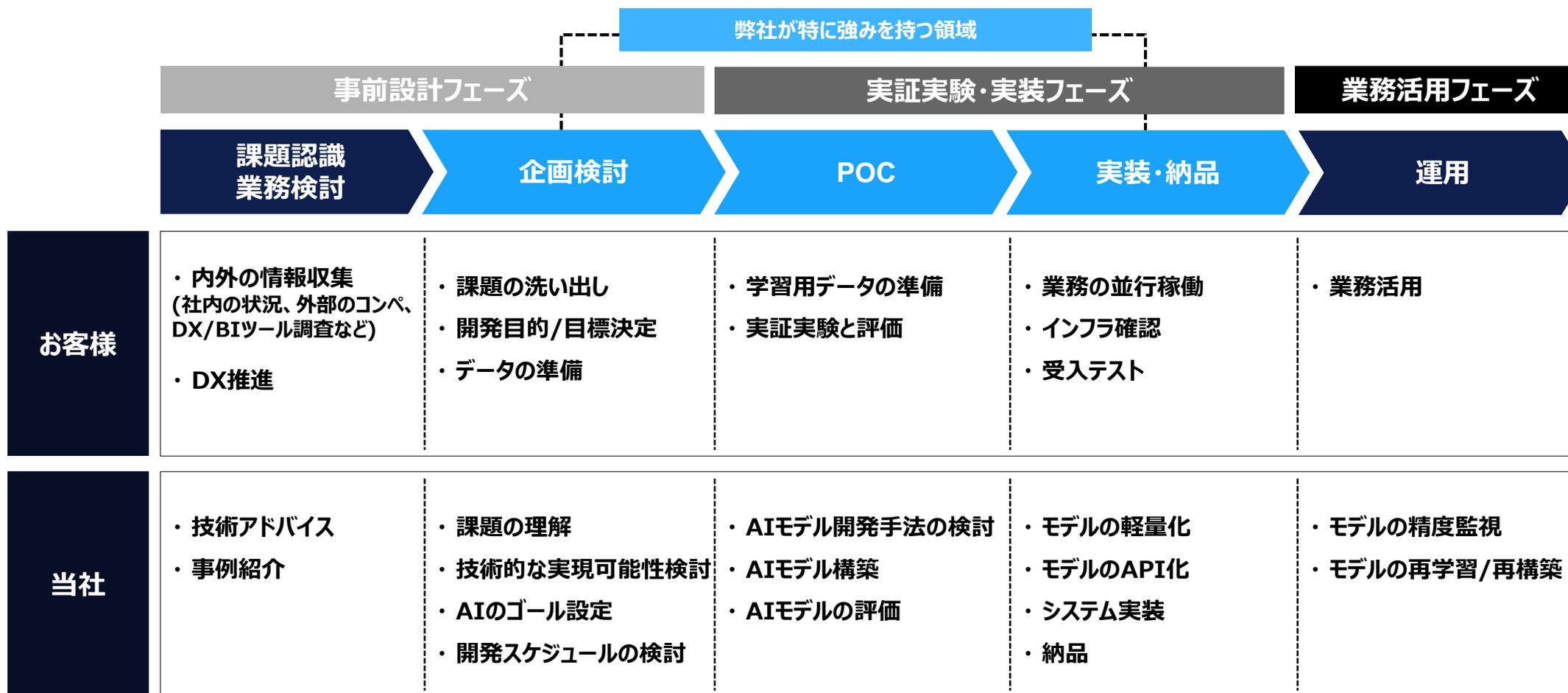


分析  
コンサルティング

データサイエンス  
教育

受託分析・開発

# 受託・コンサルティングサービス プロジェクトの進め方イメージ



当社がコアロジック部分を実装し、スクリプト、実行ライブラリ・モジュール、GUIを伴うアプリケーションとしてご提供します。  
お客様のシステム部門（子会社）やグループ会社のSIerと連携してシステム組み込みを進めることも多くございます。

# 受託・コンサルティングサービス 活用前の進め方の例 (無償対応)

## 手順

## 詳細

### 01 お打ち合わせ (要件整理)

お打ち合わせにて要件整理を実施します。「実現したいこと」「現状・問題」「課題」「課題解決方法」などを弊社スタッフと一緒に整理します。

### 適宜 対応 02 無料アセスメント(分析) ※1

活用したいデータを弊社に送付いただける場合※2、アセスメント(分析)を実施します。目的に対する実現可能性を検証し課題を洗い出します。

### 02 ご提案書作成

頂いたご要件に応じて、今後の進め方(課題解決)に関するご提案書を作成します。アセスメント(分析)の結果を盛り込むことも可能です。

### 03 お打ち合わせ (ご提案)

作成したご提案書をもとに、打ち合わせでご提案・ご報告いたします。事前に詳細を頂いている場合は、概算見積をご提示することも可能です。

### 04 ご提案書送付・ご検討 / プロジェクト開始

ご提案書をもとにサービス活用についてご検討いただけますと幸いです。

※1：無料の範囲で対応させていただきます。

※2：NDA締結のうえ、データのやり取りをさせていただくことも可能です。

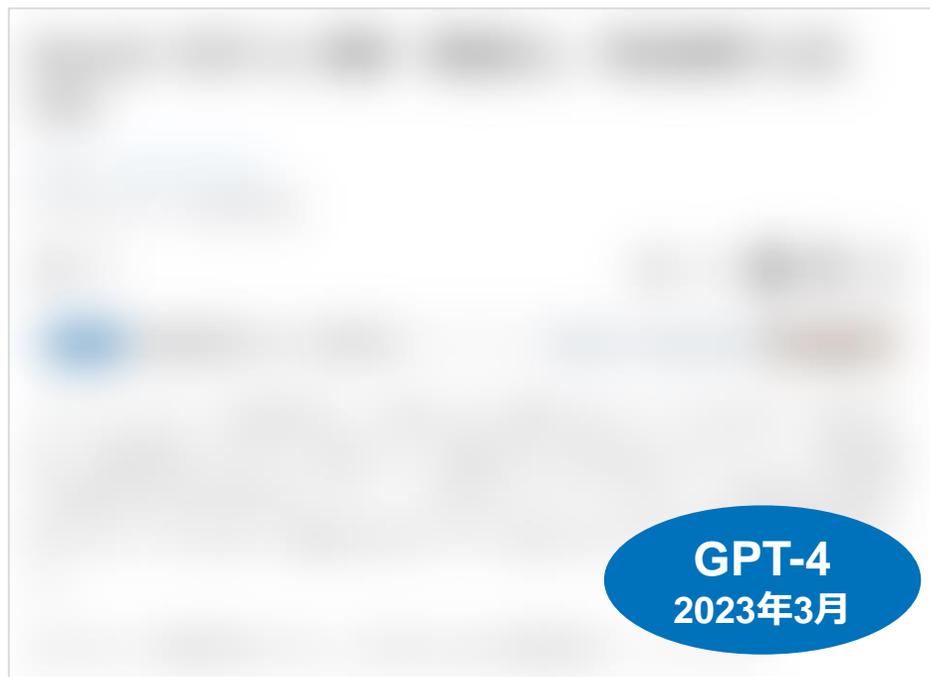
# 1 | イン트로ダクション

——進む生成AI活用と、その課題

# 生成AIの台頭——衝撃的な性能の高さ

2022年11月末のChatGPTリリースを皮切りに、**生成AIによるAIチャット**が台頭  
衝撃的な性能を備えた**ゲームチェンジングテクノロジー**として幅広い注目を集める

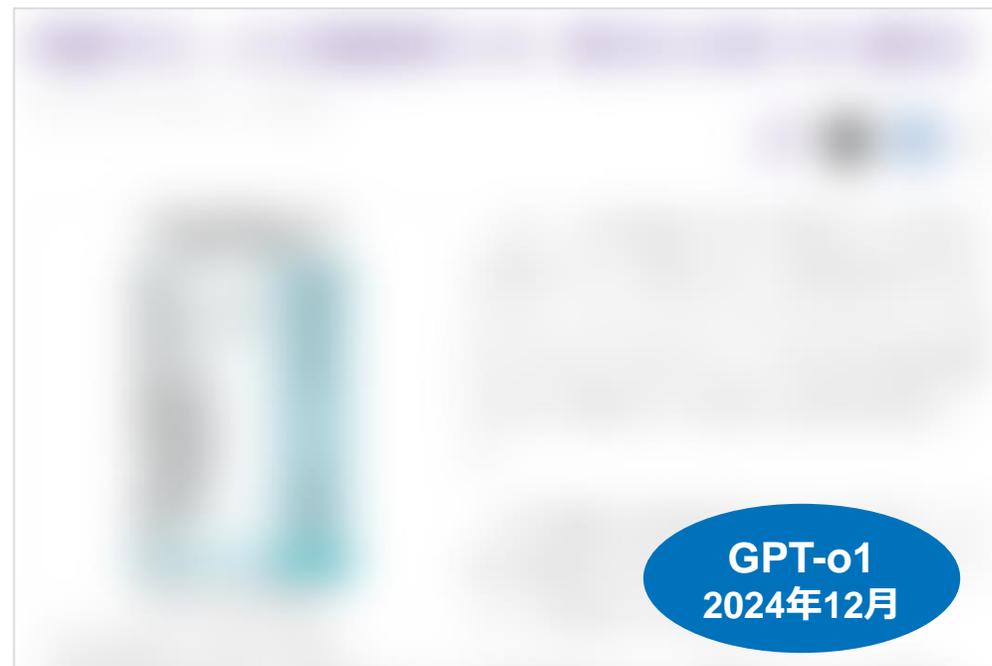
米国司法試験模試で  
GPT-4が上位10%を超える得点率（OpenAI社調査）



GPT-4  
2023年3月

日経新聞 2023年3月15日  
<https://www.nikkei.com/article/DGXZQOGN1507H0V10C23A3000000/> より引用

2025年1月実施の大学入試共通テストで  
GPT-o1が9割を超える得点率（ライフプロンプト社調査）



GPT-o1  
2024年12月

東京新聞 2025年1月20日 <https://www.tokyo-np.co.jp/article/380427> より引用

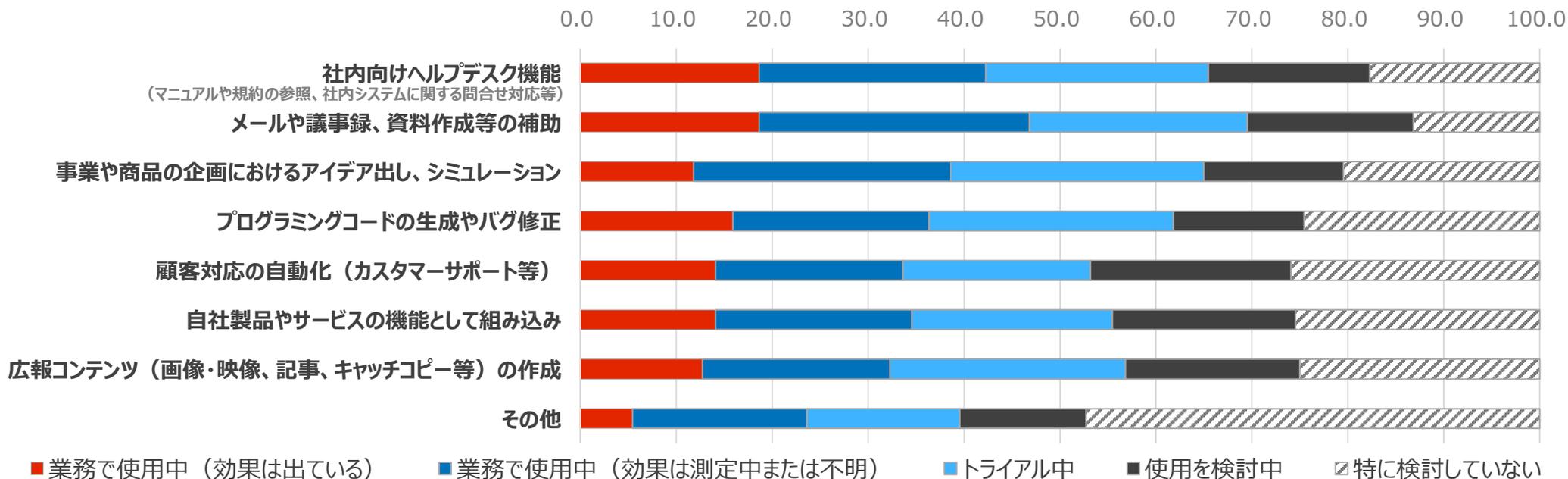
# 日本でも進んできた生成AI活用

社内業務効率化を中心に、生成AI活用のトライアル・試用が進んでいる  
2024年時点で「効果は出ている」と回答した企業は僅かだが、今後増加が予想される

## 日本企業が生成AIを活用している対象業務

(n=220)

(調査期間 2024年1月～2024年2月 ※2025年現在はさらに活用が進んでいる可能性が高い)



総務省『国内外における最新の情報通信技術の研究開発及びデジタル活用の動向に関する調査研究』（2024）より。グラフは株式会社NTTデータ数理システムが作成

# でも、業務の最前線ではまだまだ悩みがつかない生成AI導入



自分の業務、たとえばマーケティング業務では  
どう活かせるのかイマイチ想像がつかない...

いまあるテキストデータを、生成AIで活用するといっても  
具体的にどうすればいいのか...

とりあえずChatGPTに聞いてみたけど、思うような結果が返ってこないし...



そもそも従来のテキストマイニングと何が違うの？  
結局どちらを使えばよいのやら...

**本セミナーでは、これらの悩みにできるだけ具体的にお答えします！**

# 本セミナーについて——生成AI活用のポイントは？

生成AIを「使ってみた」で終わらない、真の**業務効率化・新規価値創造を目指す**には？  
そのポイントを、「**テキストデータのマーケティング活用**」を例にお伝えします

## 本日のトピック

1. **イントロダクション——進む生成AI活用と、その課題**
2. **マーケティングにおける生成AI活用、そのコツは？**
3. **分析事例：膨大な「顧客の声」を、俯瞰する**  
——生成AI×テキストマイニングによる顧客の声分析
4. **アドバンスドトピック：生成AIエージェントで、さらに仮説を磨き上げる**
5. **まとめ**

## 2 | マーケティングにおける生成AI活用、そのコツは？

# 生成AIの特徴——チャット形式でのコミュニケーション

まるで人間のような自然なコミュニケーションが可能  
膨大なデータから学習した知識に基づいて、的確な回答を返してくれる

## ChatGPT の画面例

「生成AIとテキストデータでマーケティングはこう変わる ～データドリブンな仮説立案へ～」というタイトルのセミナーのLP（ランディングページ）を作っています。このランディングページのファーストビューに表示するキャッチフレーズで、顧客の関心を引き、効率よく集客をするための文言を、5個作成してください。

自然言語による  
ユーザーの指示に対して…



セミナーのファーストビューをより魅力的にするためのキャッチフレーズを以下に5つ提案します。どれもデータドリブンなマーケティングと生成AIの活用に興味を持たせることを意識しています。

1. 「データ×生成AIで次のマーケティング戦略を。仮説立案が“ひらめき”から“科学”へ変わる」
2. 「生成AIとデータでマーケティングが進化する！効率化だけでなく、発見と仮説立案の新常識」
3. 「データドリブンマーケティングの最前線を体験！生成AIで加速する“戦略立案の精度”」
4. 「生成AIで仮説立案を革新！データが導く新しいマーケティングのカタチ」
5. 「データの力を引き出し、生成AIで未来を描く！マーケティングの成功確率を上げる方法とは？」

これらは、セミナーのテーマである「データドリブン」と「生成AI」に焦点を当てながら、具体的なメリットを想像しやすくするための表現です。

文章による回答を生成

## 叩き台作成やアイデア出しに有用

※画像は、NTTデータ数理システムが ChatGPT 4o を利用して実験した結果です。

「生成AIとテキストデータでマーケティングはこう変わる ～データドリブンな仮説立案へ～」というタイトルのセミナーのLP（ランディングページ）を作っています。このランディングページのファーストビューに表示するキャッチフレーズで、顧客の関心を引き、効率よく集客をするための文言を、5個作成してください。



セミナーのファーストビューをより魅力的にするためのキャッチフレーズを以下に5つ提案します。どれもデータドリブンなマーケティングと生成AIの活用に興味を持たせることを意識しています。

1. 「データ×生成AIで次のマーケティング戦略を。仮説立案が“ひらめき”から“科学”へ変わる」
2. 「生成AIとデータでマーケティングが進化する！効率化だけでなく、発見と仮説立案の新常識」
3. 「データドリブンマーケティングの最前線を体験！生成AIで加速する“戦略立案の精度”」
4. 「生成AIで仮説立案を革新！データが導く新しいマーケティングのカタチ」
5. 「データの力を引き出し、生成AIで未来を描く！マーケティングの成功確率を上げる方法とは？」

これらは、セミナーのテーマである「データドリブン」と「生成AI」に焦点を当てながら、具体的なメリットを想像しやすくするための表現です。

# 生成AIの問題点——ハルシネーション

知らないことを「知らない」と言えず、**それらしい解説を捏造してしまう**



コトラーによるマーケティング9.0の概要を、端的に説明してください。



フィリップ・コトラーによる「マーケティング9.0」は、デジタルと人間性を融合させ、特にAIやデータ解析技術が進化する中で、企業が顧客とより深い関係を築くことを重視したマーケティングアプローチです。この新しい時代のマーケティングは、単なる消費者ニーズの満たしにとどまらず、より人間的な価値観や倫理観に基づいた社会貢献と、企業の社会的役割の向上も含まれます。9.0の主な特徴は、個人の幸福や社会的な意義を重視する点で、パーパスドリブン（目的主導）のマーケティング戦略が中心です。

「マーケティング9.0」という現時点では存在しない概念についてそれらしい解説を捏造してしまう

**AIの回答をそのまま鵜呑みにするのではなく、確認やファクトチェックが重要**

※画像は、NTTデータ数理システムが ChatGPT 4o を利用して実験した結果です。

# 生成AIの使いどころ、基本的な考え方は？

ハルシネーションの撲滅は難しい—ある程度間違いがあっても価値を見出せる業務が使いどころ

## 生成AI向き

叩き台程度の精度でもある程度価値があるもの  
精度が上がればその分価値も上がるもの

例) 顧客課題に関する「仮説」をたくさん挙げる



## 生成AIに不向き（避けた方がよい）

精度ほぼ100%でないと価値が無いもの

※生成AIに限らず、AI技術一般にとって精度100%の達成は困難

例) 重要な契約書類のチェックを全自動化する



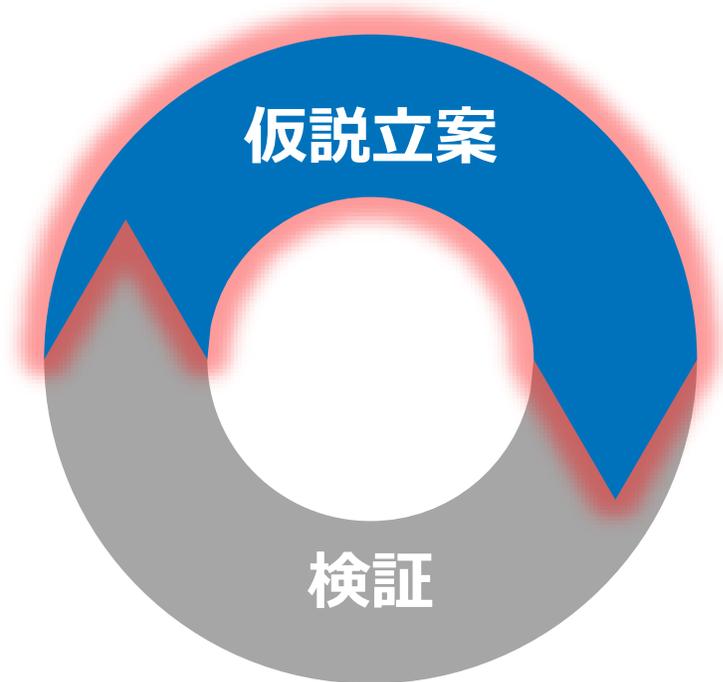
**生成AI以前から「仮説ベースでどんどん進めること」にも価値があった業務とは相性が良い**

逆にそうでない業務に生成AI導入をする場合、業務プロセスを生成AIにあわせて整理しながら、さらに細かく使いどころを見極めることが重要

# マーケティングと生成AIは相性が良い！

マーケティング業務の多くは「仮説を作って検証するサイクルを回すこと」が大事  
仮説立案パートでは、生成AIによる成果が比較的素直に享受しやすい

## 生成AIでエンパワーする

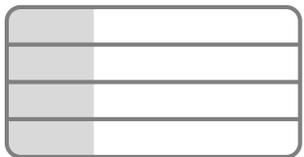


そもそも精度100%（必ず当たる）は不可能であり、それを前提として業務がつけられている

# データから仮説を得ることが重要

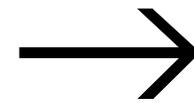
仮説は（思い付きではなく）データから得ることが望ましい  
各種データと生成AIをどのように組み合わせるかが重要になる

マーケティング・営業活動で  
得られるデータ



×

生成AI



データドリブ  
な  
仮説立案

どう組み合わせる？

組み合わせ方で、得られる仮説の精度や解像度が大きく変わる！

# よりより仮説を得るためのデータ連携のコツは？

生成AIは **強み・弱み** がはっきりとあらわれる技術  
できるだけ強みを活かし、弱みは他の技術で補いながら使うのがコツ！

強み

高度な  
自然言語理解

膨大な  
データから得た  
知識

うまく活かす！

弱み

ハルシ  
ネーション

定量的な  
考察の弱さ

他の技術で補う！

具体的なテキストデータの分析事例で、実践例を理解！

# 3 | 分析事例：膨大な「顧客の声」を、俯瞰する

——生成AI×テキストマイニングによる顧客の声分析

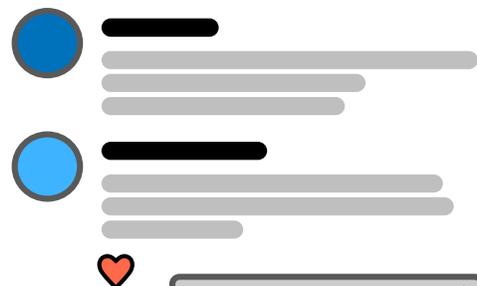
# 背景: お客様の声は気づきの宝庫

テキストデータに潜んでいる「**お客様の声**」は、  
「**なぜこの商品を選んだか / 選ばなかったか**」という顧客理解を深める**重要な手がかり**

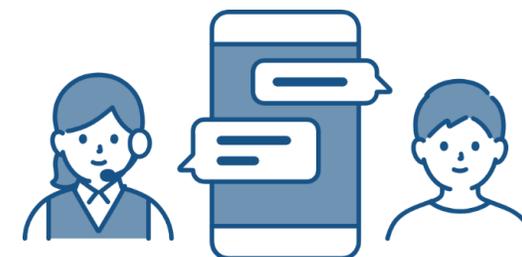
## 「お客様」の声が潜んでいるテキストデータ



ECサイトのレビューデータ



SNSデータ



コンタクトセンターの  
応対履歴データ

**これらのテキストデータを読み込んで、お客様の声を逃さないことが大事！**

# 課題: でも、テキストデータを「気合い」だけで活用するのは無理！

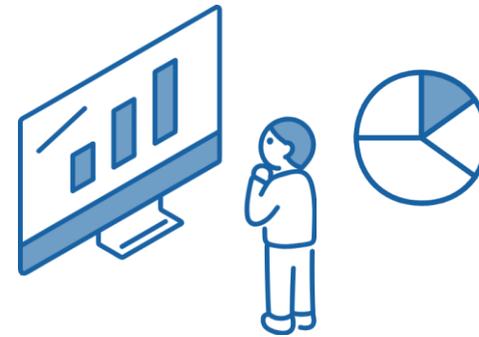
テキストデータはそれ自体が「読める」データなので、究極的には全部読めば良さそうだが...  
大量のデータを全件読んで定量的にまとめるのは困難

たくさんあって  
読み切るのに膨大な時間がかかる



時間・コストを圧迫

どんな意見がどれくらいあるのか  
定量的に把握することが困難



意思決定が結局「勘」頼りに

**膨大なテキストデータの内容を「すばやく定量的に把握できるテクノロジー」が重要に**



**生成AIにデータを入れれば解決——？**

# 生成AIの弱点——定量的考察が苦手

生成AI（大規模言語モデル）はその仕組み上、**数値の扱いがあまり得意ではない**

## 数値の大小を間違える

10.9 と 10.11 はどちらが大きいかと尋ねると…

10.9 と 10.11 では、どちらの数が大きいですか？

🌀 10.11の方が10.9より大きいです。 ❌

理由：

小数点以下の数字を比較すると、10の後の小数点第一位までは同じですが、第二位では9と11となり、 $11 > 9$ であるため、10.11の方が大きいです。

🔊 📄 👍 🗑️ ↻

## 集計を間違える

「織田」「豊臣」「徳川」が30回ずつ出現するデータを  
カウントさせると…

集計してください。

豊臣  
徳川  
徳川  
徳川  
織田  
豊臣  
⋮



以下の結果に集計しました：

- 豊臣: 29 ❌
- 徳川: 34 ❌
- 織田: 32 ❌

合計数: 95 ❌

🔊 📄 👍 🗑️ ✎ ↻

**定量的な考察を生成AIにさせるべきではない → 弱点を補う別の技術が必要**

※画像は、いずれも ChatGPT 4o を用いてNTTデータ数理システムが実験した結果です（なお、Gemini 1.5 Pro, Claude 3.5 Sonnet でも同様の誤答が生ずる）。

生成AI

×

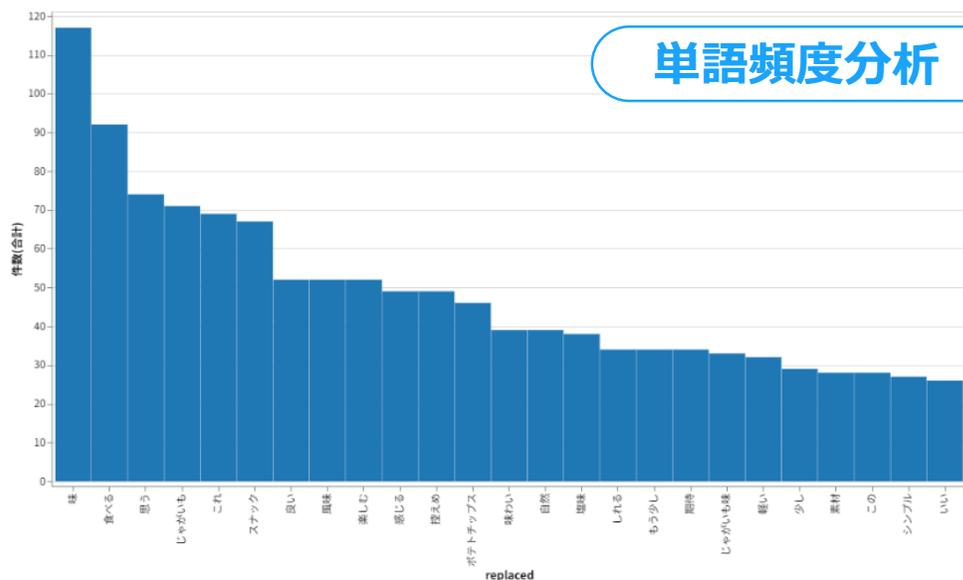
テキストマイニング

# テキストマイニングとは？——テキストデータを定量的に可視化する技術

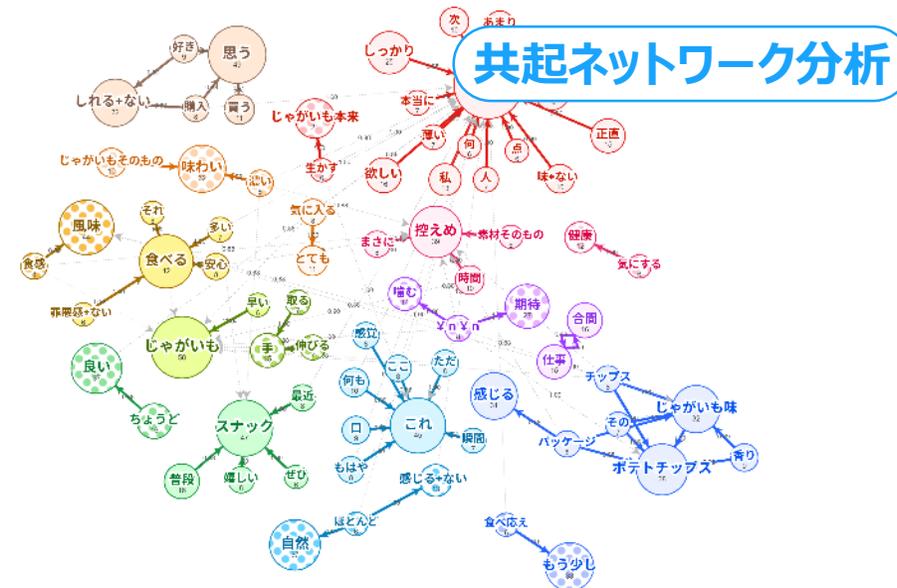
テキストデータにどんな単語がどれくらい出現するかを可視化・統計分析  
テキスト全体的话题をビジュアルかつ定量的に把握することが可能

## テキストマイニングでわかる「文章の全体観」

どういふ単語が多く出現する？



どういふ単語どうしが一緒に出現しやすい？



テキストデータを俯瞰して眺める「地図」を得る

# 例題でテキストマイニングの威力を体感！

## 架空の商品「ポテトチップス じゃがいも味」に対する商品レビューデータから ユーザーの購入の決め手は何かを探る

### 架空の商品「ポテトチップス じゃがいも味」に対するレビューデータ

#### レビュー

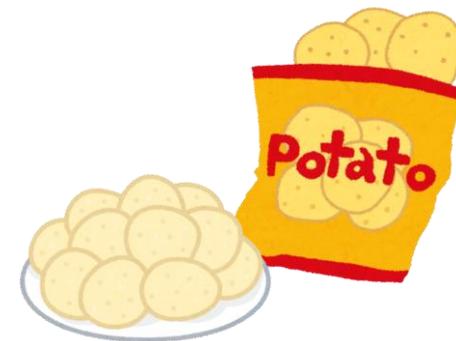
少し薄味すぎて正直物足りなかったです。自然なじゃがいもの甘さは伝わってきましたが、もう少し塩味があった方がスナックとしてはバランスが良かったかもしれません。とはいえ、健康志向のスナックとしては余計な味付けが少なくて良いと思いますし、罪悪感なく食べられるのは嬉しいですね。リピートするかは微妙ですが、軽くつまみたい時にはまた買うかも。

とてもあっさりしていて、食べやすいですね。普段は塩気の強いスナックを控えているので、こういった自然な味付けのお菓子があるのは嬉しいです。食べているとじゃがいも本来の甘さや旨味が感じられて、余計な味付けがない分、素朴で飽きがないところも良かったです。孫と一緒に食べても安心ですし、健康を気にしている方にはぜひおすすめしたいですね。また購入したいと思います。

...

- 全105件のテキストデータ
- 評価が良いものも悪いものも混ざっている（評価の良し悪しを表す★は無し）

何が購入の決め手になっているんだろう？



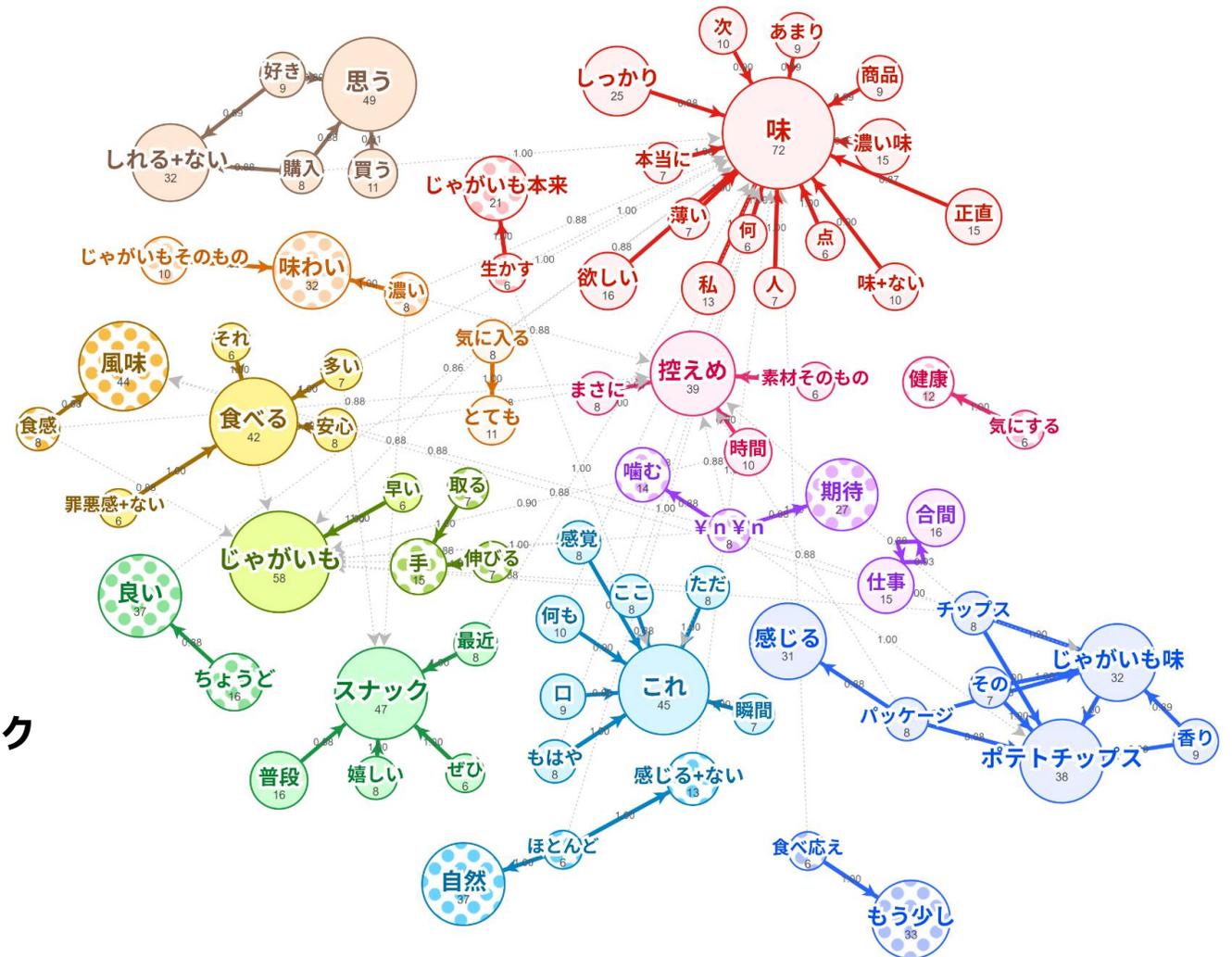
# 共起ネットワーク分析で、話題を俯瞰する

## レビュー

少し薄味すぎて正直物足りなかったです。自然なじゃがいもの甘さは伝わってきましたが、もう少し塩味があった方がスナックとしてはバランスが良かったかもしれません。とはいえ、健康志向のスナックとしては余計な味付けが少なくても良いと思いますし、罪悪感なく食べられるのは嬉しいです。リピートするかは微妙ですが、軽いつまみたい時にはまた買うかも。

とてもあっさりしていて、食べやすいですね。普段は塩気の強いスナックを控えめにしているので、こういった自然な味付けのお菓子があるのは嬉しいです。食べているとじゃがいも本来の甘さや旨味が感じられて、余計な味付けがない分、素朴で飽きがこないところも良かったです。孫と一緒に食べても安心ですし、健康を気にしている方にはぜひおすすめしたいですね。また購入したいと思います。

## 共起ネットワーク 描画

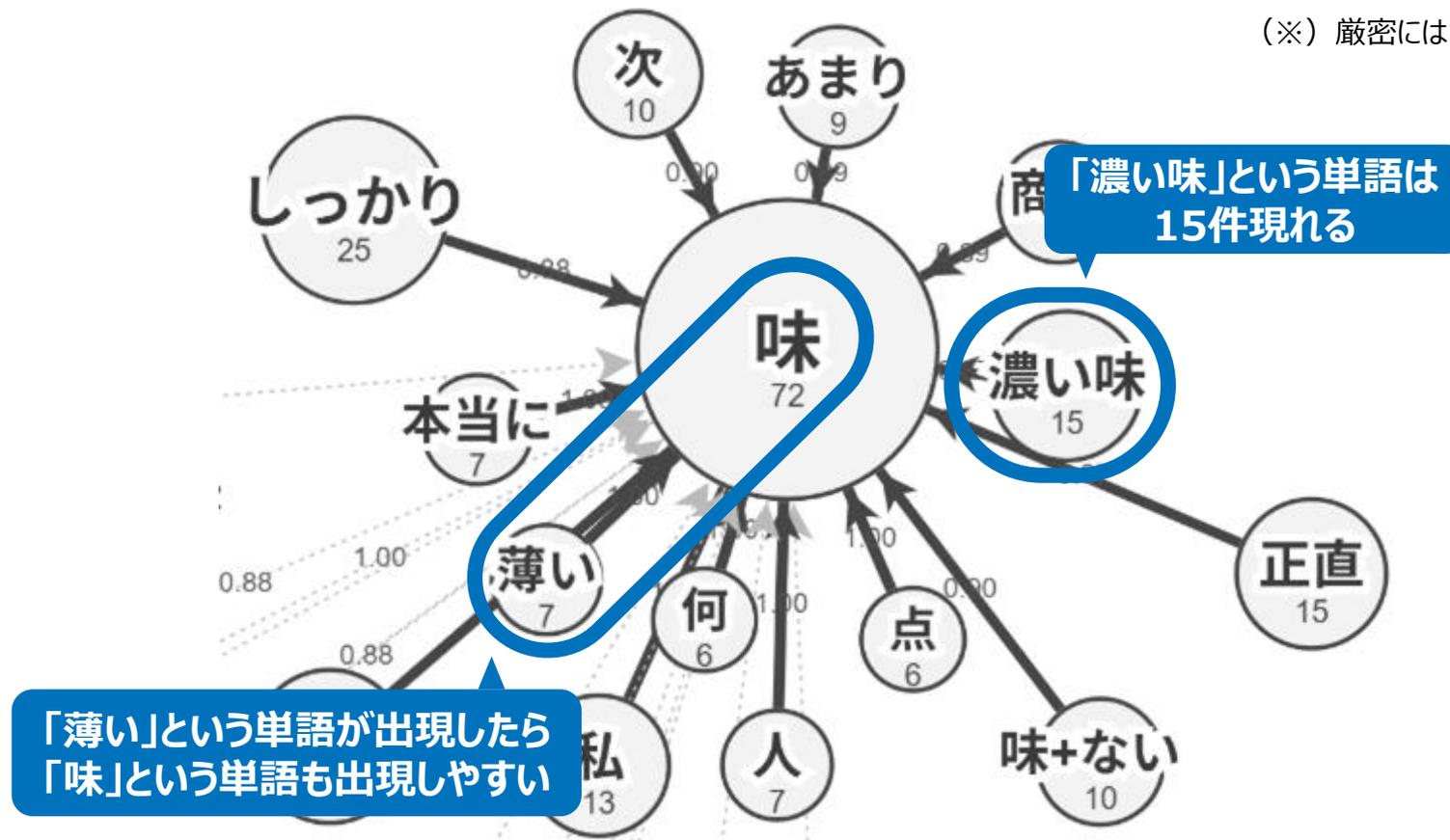


※共起ネットワーク分析の描画にはデータ分析プラットフォームAlkanoの「ネットワークからの話題抽出」アイコンを利用

## 【参考】共起ネットワーク図の見方

ノード（ひとつひとつのマル）の大きさが単語の出現数（※）  
エッジ（矢印）が引かれている2単語 A→B は、Aが出現したらBも出現しやすいことを表す

（※）厳密には「その単語が出現する文書数」を表す







# 【参考】テキストマイニングの活用事例

当社オウンドメディア MSIISM にて、当社テキストマイニングツール  
Text Mining Studio の利用事例公開中！

<https://www.msiism.jp/case/text-mining-studio/>

数理科学とコンピュータサイエンスのメディア  
**MSIISM**  
by NTTデータ数理システム



2024年6月27日 10:00  
東京海上研究所 様  
アンケートの質問項目作成、回答の分析へのテキストマイニング活用事例



2023年9月 8日 10:00  
株式会社日本経済新聞社 様  
テキストマイニングツールを活用した新聞記事データ分析事例



2024年5月23日 18:00  
太陽誘電株式会社 様  
営業レポートのテキスト情報を分析し、市場動向を可視化した事例



2023年9月 1日 17:00  
株式会社コムデザイン 様  
サービスオペレーションエンジニアのアップスキリングでお客様のDXを推進



2023年11月13日 17:00  
東京理科大学 経営学部 朝日 弓未 様  
データ解析ツールを使った実践的なマーケティング・サイエンス授業事例



2023年1月 4日 13:00  
株式会社NTTデータ・スマートソーシング 様  
テキストマイニングツールを使ったコールセンターのVOC分析事例

# でも、単純なテキストマイニングだけでは…

単語ベースで考える単純なテキストマイニングだけでは  
複数の話題の混合・皮肉など、商品レビューでよくある「困ったデータ」を扱うのが難しかった

## 商品レビューでよくある少し「困った」データ

複数の話題（商品・配送）が混合しているデータ

商品は**良い**のですが配送が**ダメ**でした。

皮肉が使われたデータ

極めて哲学的な**深みのある**お菓子で、人間の味覚に挑戦している新感覚の何かです。お菓子とは何か考えさせられ、非常に**良い**経験になりました。

## 単純なテキストマイニングで扱おうとすると…

単純な単語の集計だけでは、**ポジティブな語**と**ネガティブな語**がどの観点・話題に結びついているのかわからなくなってしまう



**ポジティブな語**のみで構成されているが、常識的に考えれば明らかに**ネガティブな意味合い**の**ポジティブな語**が多様 → ポジティブというルールが通用せず、ルールベースでは意味の把握が困難

**観点分離や皮肉の処理などは、別の技術で補う方が良い**

# 生成AIのもつ高度な文脈把握能力を活かせば… 複数観点を分離しながらのポジネガ判定

**課題** ECサイトのレビューには商品自体の評価以外にも、配送・カスタマーセンター対応など**複数の観点の評価が混入**

生成AIであれば、**複数の観点を分離しながらのポジネガ判定も比較的容易にできる**

## 複数の観点が混在したレビューデータ

商品は良いのですが配送がダメでした。

電話対応が最悪。もう二度とこの配送サイトは利用しません

素材本来の味が楽しめて、とても美味しかったです！

今なら月額3000円で脱毛できます！ **今すぐここをクリック！**

話題を  
分離して  
ポジネガ  
判定



## ChatGPT 4o での判定結果

商品	配送	CC対応
positive	negative	---
---	negative	negative
positive	---	---
---	---	---

※後で紹介するプロンプトにより、NTTデータ数理システムが ChatGPT 4o を利用して実験した結果。実験の結果は確率的に変動するため、3回実験を行い最頻値を取った。

# 生成AIのもつ高度な文脈把握能力を活かせば… 皮肉が含まれるデータのポジネガ判定

**課題** インターネット上のテキストデータには、ハイコンテキストな皮肉が含まれることも多くある

生成AIであれば、皮肉が含まれるデータのポジネガ判定も比較的正確にできる

## 皮肉が含まれるデータ

極めて哲学的な深みのあるお菓子で、人間の味覚に挑戦している新感覚の何かです。お菓子とは何か考えさせられ、非常に良い経験になりました。

単なる芋です。それ以下でもそれ以上でもなし。

そりゃジャガイモ味にしたらそうなるだろって味。個人的には納得。

話題を  
分離して  
ポジネガ  
判定



## ChatGPT 4o での判定結果

商品	配送	CC対応
negative	---	---
negative	---	---
middle	---	---

※後で紹介するプロンプトにより、NTTデータ数理システムが ChatGPT 4o を利用して実験した結果。実験の結果は確率的に変動するため、3回実験を行い最頻値を取った。

# 【参考】話題分離・ポジネガ判定の出力を得るために利用したプロンプト

後で与える文字列のうち、<review></review> xml タグで囲まれた部分は、以下に示す(レビュー対象商品)の購入者による EC サイトのレビュー投稿です。また、レビュー投稿の文字列をそのまま画像化したものも添付します。このレビューについて、以下に示す(注意すべき点)に注意しつつ、購入者が、後に示す 3 つの(観点)のそれぞれに対して、言及がない (none) か、好意的 (positive) か、言及があるが中立 (middle) か、否定的 (negative) かを判断し、以下の(出力形式)にしたがって出力してください。

## (レビュー対象商品)

ポテトチップス じゃがいも味

## (観点)

- 商品 (merchandise) : 商品自体に対する評価です。
- 配送 (shipping) : 商品の配送自体に対する評価です。
- カスタマーセンター対応 (customer\_center) : 電話対応やEメール対応など、カスタマーセンターとのコミュニケーションに対する評価。

## (注意すべき点)

- レビュー対象商品の特性によく注意してください。たとえば嗜好品に対するレビューの場合は、一見中立的なレビューに見えても、期待に比べて大したことはないという意味で否定的 (negative) である可能性が高いです。
- インターネットへの投稿であることを踏まえ、皮肉的な表現にも注意してください。皮肉的な表現がある場合、否定的なレビューである可能性が高いです。皮肉的な表現を見つけるヒントとして、一見肯定的に述べているように見える表現が使われているものの、よく考えると不自然な箇所があったり、現実には起こりにくい内容が記載されている場合があります。
- スпамなどが投稿されている可能性も考慮してください。つまり、レビュー対象の商品と関係のない別の商品の宣伝や、迷惑投稿である可能性もあります。その場合は、どの観点も "none" として出力してください。

## (出力形式)

以下の二つの XML タグに囲まれた文章を出力してください。

1. <reason></reason> タグ

各観点 (商品、配送、カスタマーセンター対応) に対して、判断の根拠を、できるだけ詳細に (最大800トークン程度で) 書いてください。この際、(注意すべき点)に記載された事項によく注意して判断をしてください。

2. <answer></answer> タグ

このタグの中に、以下のキーを持つ json オブジェクト形式を出力してください。

- "merchandise", "shipping", "customer\_center" のそれぞれに対して、判断の結果を記載してください。値は、"positive", "middle", "negative", "none" のいずれかとしてください。

## (出力例)

以下は、上の条件にすべて適合する出力の例です。reason については例なので省略していますが、実際の出力では詳しく判断理由を述べてください。

```
<reason>
...
</reason>
<answer>
{
  "merchandise": "negative",
  "shipping": "none",
  "customer_center": "none"
}
</answer>
```

高精度な出力を得るためには  
これぐらいのプロンプトを  
書く必要がある  
(これでも少ない方?)

# 生成AIによるテキストデータ活用のコツ——テキストマイニングとあわせて使う

テキストマイニング×生成AI  
お互いの強みを活かしながら（弱点を補い合いながら）使う

単純なテキストマイニング

生成AI

定量的な考察

**しやすい**  
可視化して俯瞰できる  
統計手法と組み合わせることも可能

**（生成AIだけでは）しにくい**  
生成AIはそもそも数の扱いが苦手

文脈を踏まえた  
高度な考察

**苦手**  
単語・構文レベルの情報しか使えず、厳しい

**得意**  
常識を踏まえた高度な文脈把握力

**組み合わせることで、より正確かつ直観的な「顧客理解」へ！**

# 【参考】単純なテキストマイニングと生成AIの比較

単純なテキストマイニングと生成AIの強み・弱みは、他の様々な観点でもよく噛み合う

## 単純なテキストマイニング

## 生成AI

データ処理コスト

**安い**  
処理自体にはほとんどコストがかからない

**高い**

定量的な考察

**しやすい**  
可視化や統計手法と組み合わせることが可能

**(生成AIだけでは) にくい**  
生成AIだけでは基本的な計算も間違う

安定性・再現性

**高い**  
シンプルなアルゴリズムは安定性・再現性が高い

**低い**  
入力が少し変化すると回答が大幅にバラつく場合も

文脈を踏まえた  
高度な考察

**苦手**  
単語・構文レベルの情報しか使えず、厳しい

**得意**  
高度な自然言語処理による高い推論力

データの前処理の  
必要性

**必要**  
データクレンジングを丁寧にする必要がある

**高い解釈能力によりある程度省略可能**  
高い解釈能力を、前処理自体に使う場合も

# 例題でテキストマイニング×生成AIの威力を体感！

## 架空の商品「ポテトチップス じゃがいも味」に対する商品レビューデータから ポジティブなレビューとネガティブなレビューの傾向の違いを探る

### 架空の商品「ポテトチップス じゃがいも味」に対するレビューデータ

#### レビュー

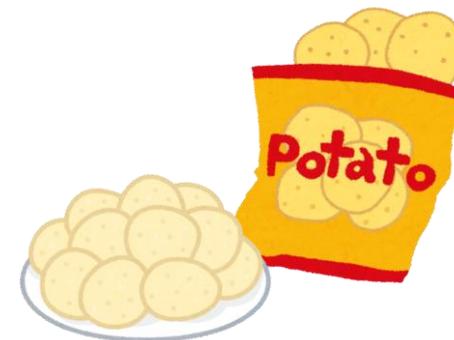
少し薄味すぎて正直物足りなかったです。自然なじゃがいもの甘さは伝わってきましたが、もう少し塩味があった方がスナックとしてはバランスが良かったかもしれません。とはいえ、健康志向のスナックとしては余計な味付けが少なく良いと思いますし、罪悪感なく食べられるのは嬉しいですね。リピートするかは微妙ですが、軽くつまみたい時にはまた買うかも。

とてもあっさりしていて、食べやすいですね。普段は塩気の強いスナックを控えているので、こういった自然な味付けのお菓子があるのは嬉しいです。食べているとじゃがいも本来の甘さや旨味が感じられて、余計な味付けがない分、素朴で飽きがないところも良かったです。孫と一緒に食べても安心ですし、健康を気にしている方にはぜひおすすめしたいですね。また購入したいと思います。

...

- 全105件のテキストデータ
- 評価が良いものも悪いものも混ざっている（評価の良し悪しを表す★は無し）

ポジティブなレビューと  
ネガティブなレビュー  
では、どういう傾向の  
違いがある？



# 分析ワークフロー

生成AIで商品の観点でポジ・ネガを分け、それぞれをテキストマイニング  
ポジティブレビューのネットワークとネガティブレビューのネットワークを並べて比較してみる

レビューデータ全体  
(105件のテキストデータ)

生成AIによる  
観点別  
ポジネガ分類

商品の  
ポジ・ネガを分けた  
テキストデータ

テキスト  
マイニング

商品の  
ポジ・ネガを分けた  
ネットワーク

## レビュー

少し薄味すぎて正直物足りなかったです。自然なじゃがいもの甘さは伝わってきましたが、もう少し塩味があった方がスナックとしてはバランスが良かったかもしれません。とはいえ、健康志向のスナックとしては余計な味付けが少なく良いと思いますし、罪悪感なく食べられるのは嬉しいですね。リピートするかは微妙ですが、軽くつまみたい時にはまた買うかも。

とてもあっさりしていて、食べやすいですね。普段は塩気の強いスナックを控えめにしているのですが、こういった自然な味付けのお菓子があるのは嬉しいです。食べているとじゃがいも本来の甘さや旨味が感じられて、余計な味付けがない分、素朴で飽きがないところも良かったです。孫と一緒に食べても安心ですし、健康を気にしている方にはぜひおすすめしたいですね。また購入したいと思います。

...



「商品がポジティブ」  
のみのレビューデータ

可視化  
→

「商品がポジティブ」  
ネットワーク

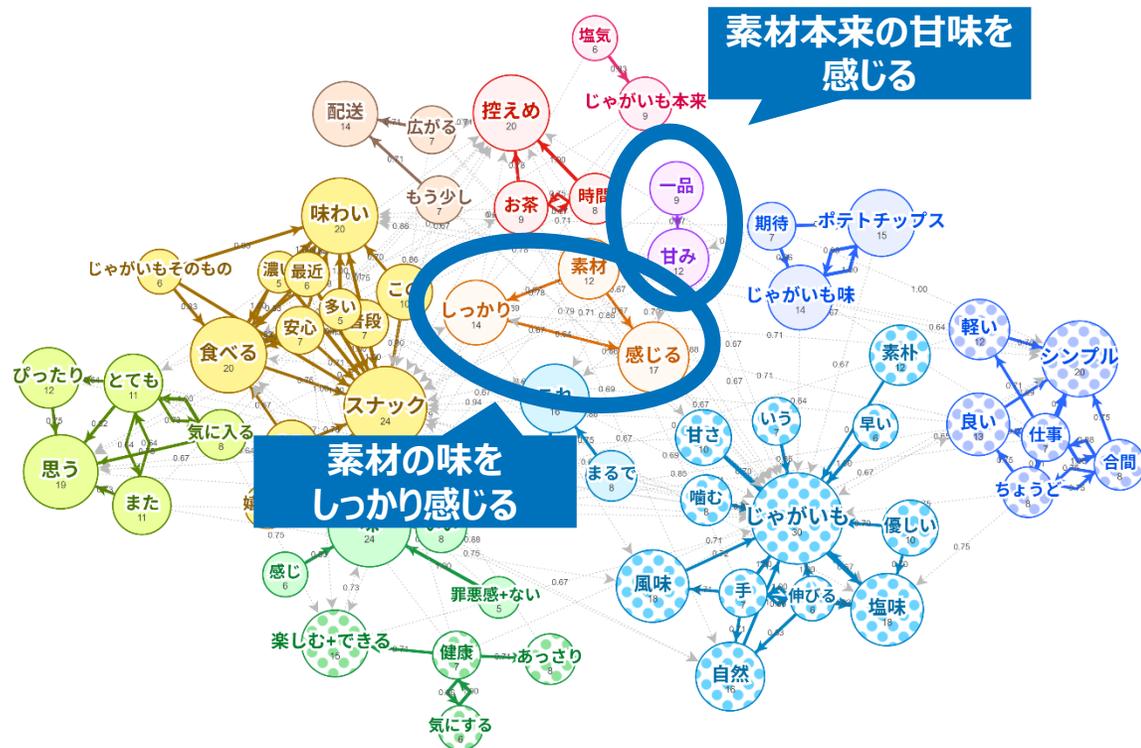
VS

「商品がネガティブ」  
のみのレビューデータ

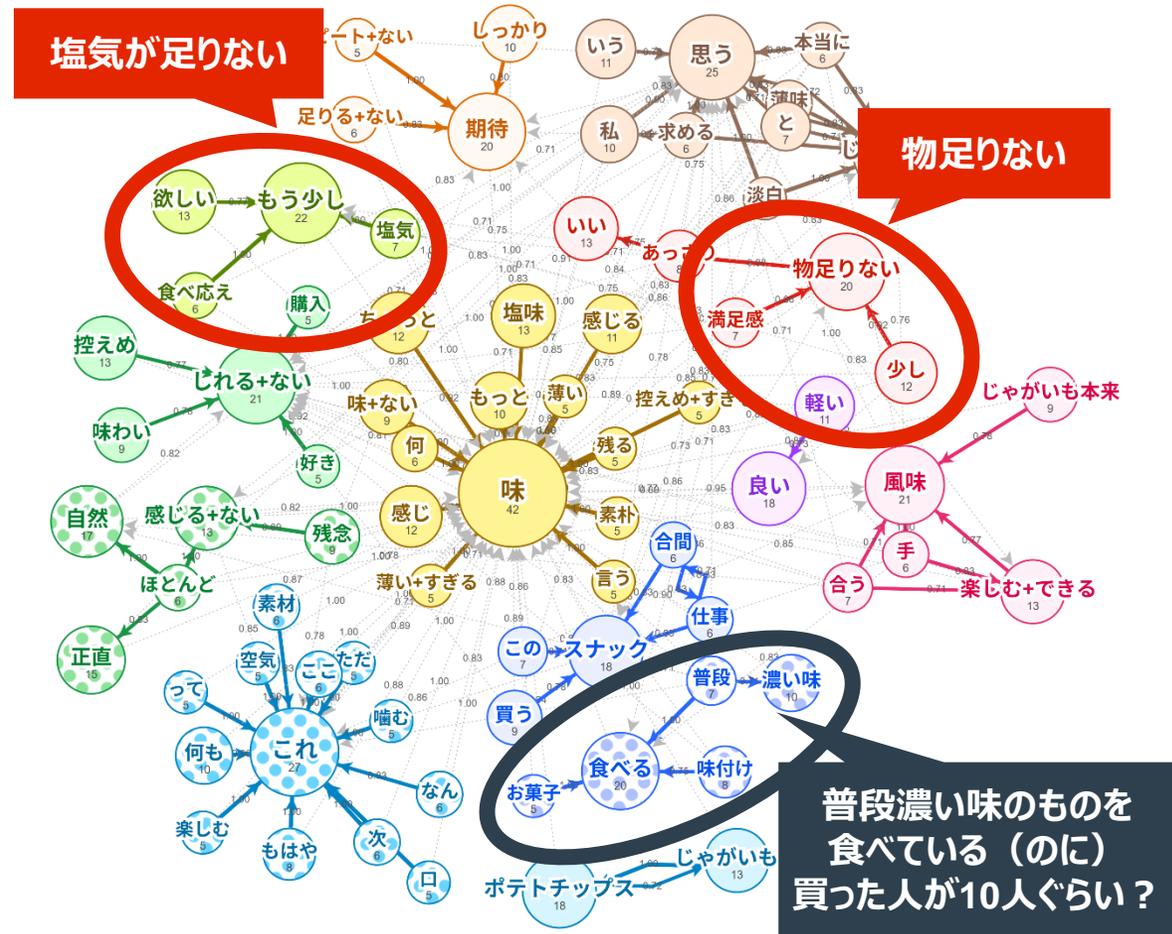
可視化  
→

「商品がネガティブ」  
ネットワーク

# ポジティブレビューの共起ネットワーク



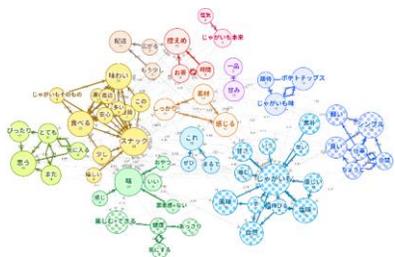
# ネガティブレビューの共起ネットワーク



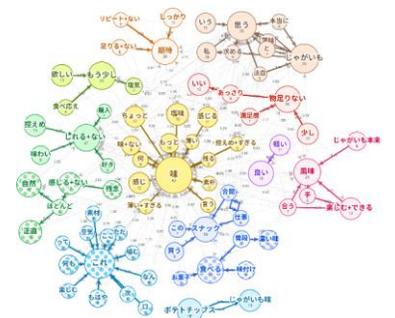
商品をポジティブに捉えた層・ネガティブに捉えた層の「捉え方の違い」が可視化できる

# より進んだテキストマイニング×生成AIの使いどころ——仮説探索

## ネットワークの解釈および仮説探索自体を、生成AIでやらせることも可能



ポジティブレビュー  
共起ネットワーク



ネガティブレビュー  
共起ネットワーク

二つのネットワークを  
入力して  
傾向の違いから  
仮説を挙げさせる

### 仮説 3: じゃがいもの味に対する期待値が異なる

- 高評価レビューでは「じゃがいも本来『素朴』など、じゃがいもの自然な味わいを評価している傾向が見られる一方、低評価レビューでは「じゃがいもらしさが足りない」という不満が見られ、期待値の違いが評価に影響している可能性があります。

### 仮説 4: 高評価者は繊細な味を好み、低評価者はインパクトのある味を好む

- 高評価のレビューには「ほんのり『素材感』といった繊細な味に対する評価が多く見られるが、低評価レビューでは「塩味『しっかり』など、インパクトのある味を求めている可能性が考えられます。

### 仮説 5: ポテトチップスに求める役割が異なる

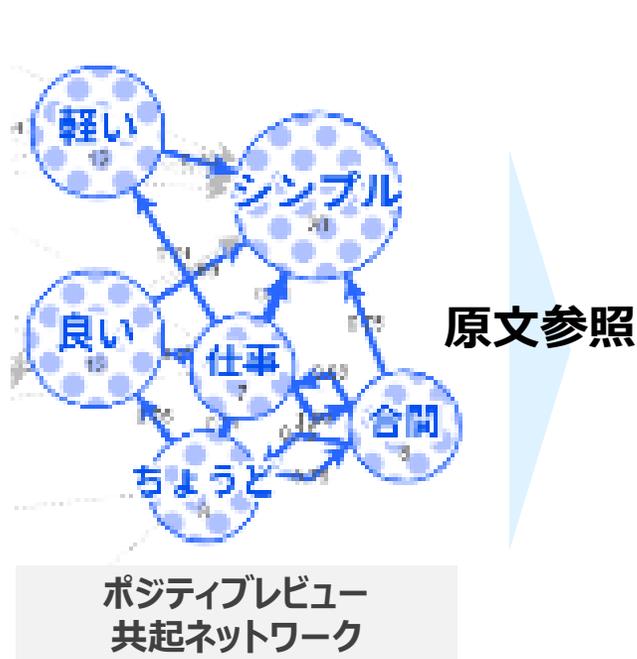
- 高評価者は「お茶」や「軽さ」を好む傾向が見られ、ポテトチップスを軽い間食やお茶うけとして捉えている可能性があります。一方、低評価者は「満足感『リピートできない』といった言葉が多く、ポテトチップスを「しっかり食べ応えがあるもの」として期待しているのかもしれませんが。

**ポジティブレビューをした顧客と、ネガティブレビューをした顧客の違いに関する仮説が得られる！**

※画像は、いずれも ChatGPT 4o を用いてNTTデータ数理システムが実験した結果です。

# より進んだテキストマイニング×生成AIの使いどころ——ペルソナ作成

## ネットワークに紐づく原文から、ペルソナを生成して顧客ニーズの仮説を作る



テキスト

在宅ワークのお供として『ポテトチップス じゃがいも味』を試してみましたが、これが予想以上に良かったです！じゃがいもの風味がしっかりしていて、程よい塩加減も丁度いい。油っぽさも少ないので作業中にも罪悪感なく食べられます。素材の味が楽しめるシンプルなチップス、こういうのを探していました！またリピートします。

ポテトチップス じゃがいも味は、スナック菓子なのにとっても優しい味わいで気に入りました。じゃがいもの自然な風味がしっかりと感じられ、塩味も控えめで食べやすいです。重くならず、お茶の時間にぴったり。こうしたシンプルなお菓子を楽しまれるのはうれしいですね。また買おうと思います。

...

ペルソナ生成

項目	詳細
名前	小林香織
年齢	36
性別	女性
会社/職位	個人事業主
性格	健康的な生活を送りたいと考えているが、 <b>忙しさや在宅ワークの合間に甘い物やスナックをつい食べてしまうことに罪悪感を覚えがち</b> 。素材の味を大切にする自然派志向で、余計な味付けが苦手。ほんのり優しい味わいを好み、仕事中でも軽くつまめるおやつを探すのが趣味。商品デリバリーの遅延やトラブルには少し神経質になりがちで、ストレスフリーな購買体験を求めている。
仕事	在宅でクライアントのWEBサイトやバナー広告をデザインし、プロジェクトごとに進捗管理から納品まで一人で担当。細かな修正対応やタイトなスケジュールが多く、日中から深夜まで稼働することもある。
気質	穏やかで優しい物腰だが、妥協を許さない職人氣質。興味を持ったテーマは徹底的に調べる探究心があり、仕事も趣味も効率より納得感を重視する傾向が強い。 <b>健康や美味しさに関しても“素材の本質を味わいたい”という思いから、余計なものが入っていないシンプルさを好む</b> 。



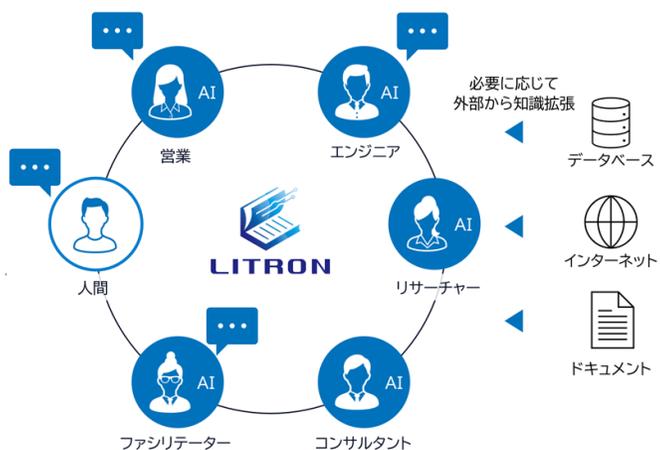
※生成されたペルソナは、ChatGPT 4o を用いて株式会社NTTデータ数理システムが生成したものです。また、ペルソナの画像も ChatGPT 4o および DALL・E3 を用いてNTTデータ数理システムが生成したものです。

# 4 | アドバンスドトピック 生成AIエージェントで、さらに仮説を磨き上げる

# 2025年は「AIEージェント」技術で、生成AIの本格展開が加速？

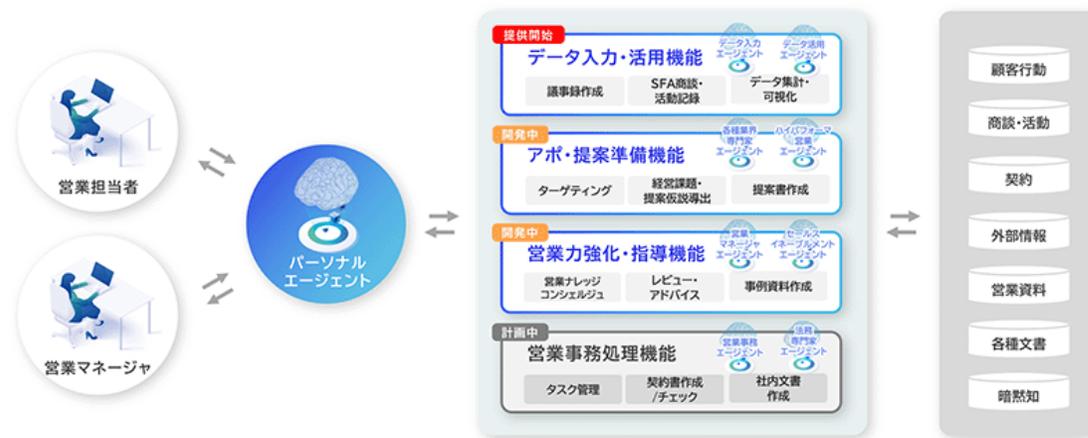
幅広い業務を自律的に支援するAIEージェントが、直近で相次いで発表  
2025年は、AIEージェント技術を中心に生成AIの本格展開が進む年に？

## 事例1：NTTデータ・数理システム LITRON® MAS AIEージェント同士の議論によるアイディエーション支援



NTTデータ数理システム『生成AI同士のコミュニケーションが新たな洞察を導く、LITRON®新サービスを提供開始』  
(2024年07月31日) <https://www.msi.co.jp/information/2024/20240731-1145.html> より引用

## 事例2：NTTデータ LITRON® Sales 営業支援のためのAIEージェント

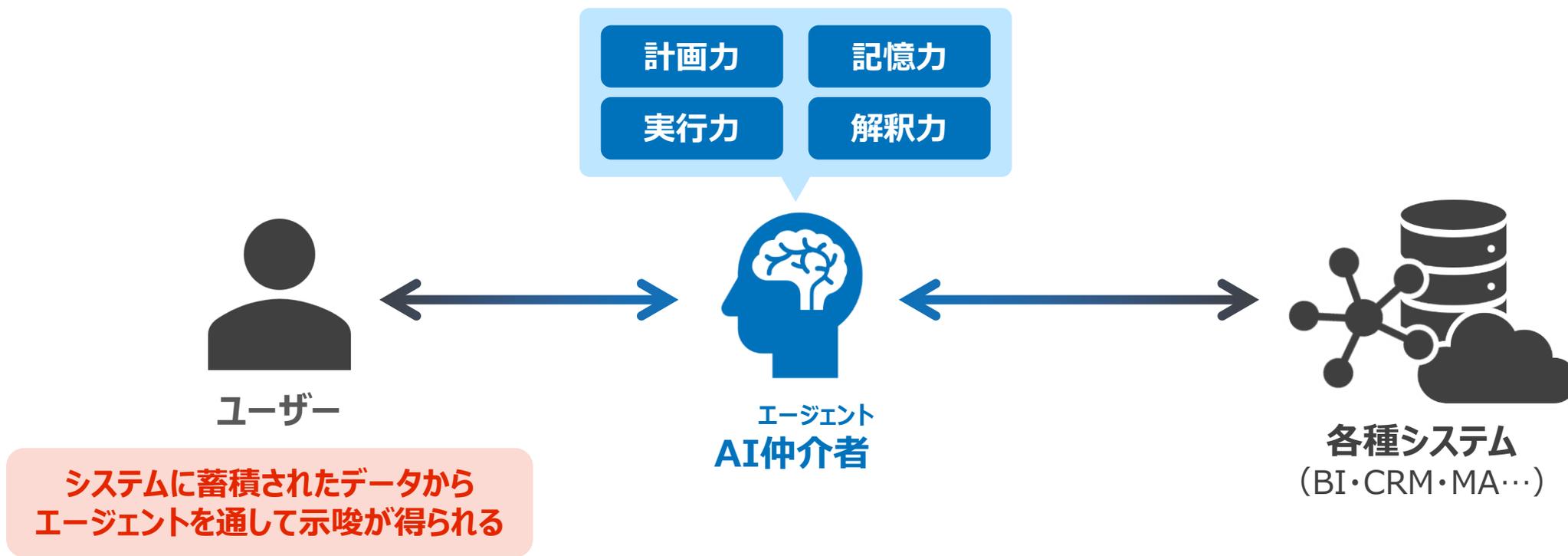


NTTデータ『AIEエージェントを活用した新たな生成AIサービスを提供開始』(2024年10月24日)  
<https://www.nttdata.com/global/ja/news/release/2024/102401/> より引用

# AIエージェントは、問題解決のための「仲介者」

AIエージェントは、各種システムとユーザーの間の仲介者となって問題解決能力を提供  
ユーザーの代わりに各種システムから情報を収集し、解釈し、示唆を出す

システムを使った問題解決に必要な能力を保持



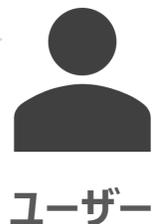
より多くの人にデータ活用の恩恵が広がる？

# マーケティングにおけるAIエージェントの使いどころは

各種システムの顧客データを分析して、実際の顧客特性に近いペルソナ作成  
AIエージェントに「顧客のように振る舞ってもらう」ことでインタビューやロールプレができる

例)

なぜ当社の商品を購入したのでしょうか？



インタビュー  
ロールプレ

顧客の  
ペルソナらしく振る舞う



顧客データを要約して  
ペルソナを作成



各種システム  
(BI・CRM・MA…)

これまでもデータ分析による顧客理解・ペルソナ作成は行われていたが、  
「ペルソナのAIエージェント化」で、ペルソナの活用の幅が広がる

※ペルソナの画像は ChatGPT 4o および DALL・E3 を用いてNTTデータ数理システムが生成したものです。

# AIエージェントのもう一つのトレンド：マルチAIエージェント

複数のAIエージェントをディスカッションさせ、より仮説や論点を深堀  
商品開発担当者によるグループインタビューの再現も可能



生成AIでペルソナを作成

(顧客データのクラスタリング結果をインプットにできる)



各エージェントならではの視点から議論を深め、新たな仮説を自律的に作り出して深堀りする

※ペルソナの画像は ChatGPT 4o および DALL・E3 を用いてNTTデータ数理システムが生成したものです。

# マルチAIエージェントによるディスカッションの活用例：施策の比較

複数の施策を比較するため、想定ペルソナの「AI消費者」へのグループインタビューを実施

例) 材料費の値上げに伴うレストランのメニュー「ほうれん草のソテー」改定施策を比較

ベーコンのトッピング

VS

サイズオプションの追加



AI消費者に「どう思うか」をグループインタビュー

※ペルソナの画像は ChatGPT 4o および DALL・E3 を用いてNTTデータ数理システムが生成したものです。

# 施策の比較——ベーコンのトッピングの場合



藤原健二 / システムエンジニア—ネットワークエンジニア (30)

ポジティブな評価としては、ベーコンのトッピングが追加されることで、さらにおいしくなる点です。特に家族で楽しむ際には、少し贅沢な気分も味わえるので、値上げには納得できる部分もあります。これで家族全員が喜ぶのであれば、50円の値上げも許容できますよ。



山下駿太 / 高校生—理系コース (18)

私は一高校生として、家族では外食を楽しむことが少ないんですが、お小遣いをうまく使って友達と一緒にいくことが多いです。正直言って、50円の値上げはお財布に少し響くかもしれませんが、ベーコンのおかげで「ほうれん草のソテー」がもっとおいしくなるなら、試してみたいですね。



竹内美咲 / 商品企画—シニアマーケター (37)

さて、皆さん、非常に興味深い意見が出てきております。ありがとうございます。  
初川さん、先ほど「50円の値上げが財布に響く」とおっしゃいましたが、それでもベーコンのトッピングに魅力を感じているとのこと。具体的にどのような点で「試してみる価値がある」と感じられるのか、もう少し詳しく聞かせていただけますか？



初川美咲 / ITサービス—フロントエンドエンジニア (23)

正直、私はベーコンの脂身が苦手なんです。でも、たまに友達と一緒に食事をする時にはこういったトッピングがあると盛り上がりますし、ちょっとしたご褒美感がありますね。

# 施策の比較——サイズオプションの追加



藤原健二 / システムエンジニアネットワークエンジニア (30)

そうですね、サイズオプションの導入については、個人やシチュエーションに応じて選べる自由が増える点が非常にポジティブだと思います。特に家族で外出するとき、子供たちの食べる量がまだ少ないので、料金の無駄を減らせる小サイズがあると嬉しいですね。また、大人も今日は軽く食べたい、今日はしっかり食べたいといった日によって選べるのも良いと思います。



山下駿太 / 高校生一理系コース (18)

また、学生である自分にとって、家計を意識した食事選びが必要なので、手頃な価格で選べる小サイズがあると特に助かります。例えば、部活の後にちょっとした軽食が欲しいときにもピッタリだと思います。



藤原健二 / システムエンジニアネットワークエンジニア (30)

私自身も家族で外出することが多いですが、子供たちは食べる量が日によって違います。サイズオプションがあると、その日の食欲や状況に合わせて無駄なく選べるので、とても助かります。



初川美咲 / ITサービスフロントエンドエンジニア (23)

そうですね、個人的に気になるのは、ほうれん草の分量が減る小サイズだと栄養バランスが心配かなって思います。家族の健康を考えると、やっぱり十分な野菜が摂れる大サイズの方が魅力的だと感じます。

※テキストおよびパーソナの画像は ChatGPT 4o および DALL・E3 を用いてNTTデータ数理システムが生成したものです。

# 5 | まとめ

# よりよい仮説を得るためのデータ連携のコツは？

生成AIは **強み・弱み** がはっきりとあらわれる技術  
できるだけ強みを活かし、弱みは他の技術で補いながら使うのがコツ！

強み

高度な  
自然言語理解

生成AIによる  
観点別ポジネガ

膨大な  
データから得た  
知識

うまく活かす！

弱み

ハルシ  
ネーション

定量的な  
考察の弱さ

テキストマイニング  
可視化

他の技術で補う！

他技術との組み合わせることで、一度やって微妙だった生成AI活用案件が進むかも——  
**ぜひNTTデータ数理システムにご相談ください！**



# 【参考】NTTデータ数理システムの生成AI関連コア技術スタック

生成AIを単に個人レベルの業務効率化ツールに留めるのはもったいない——生成AIはデータと組み合わせることでより大きなインパクトを発揮できると考えています  
当社の得意領域である「データ分析・データ処理技術」と組み合わせることで、お客様がもつデータを〈ナレッジ〉として蓄積・活用できるようご支援が可能です

## アプリケーション構築

プロンプトエンジニアリング・メカニズムエンジニアリングにより生成AIの性能を引き出し、使いやすいアプリケーションにまとめます

使いやすいGUIの迅速な構築

Streamlit

React

Next.js

ワークフロー構築・管理

Dify

高度なマルチモーダルインターフェース

音声によるリアルタイム対話アプリケーション

性能を限界まで引き出すプロンプトエンジニアリング

論文調査・統計的評価に基づく  
ベストプラクティスの追及

徹底的な業務の言語化

エージェント構築・メカニズムエンジニアリング

ペルソナプロンプティング

Task Planning

マルチエージェント

## データ活用技術との連携

データサイエンス領域における幅広い知見と経験を活かし、多種多様で複雑なデータからのナレッジ抽出・蓄積・活用を支援します

多種多様なデータからの情報・ナレッジ抽出

ソースコードの構造解析



図表を含む複雑な文書



手書きを含む文書

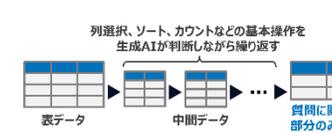


情報・ナレッジの蓄積と活用 (RAGの構築)

ナレッジグラフ



RDB等の表データ



## 生成AIモデル選定・学習

幅広いモデルのなかから課題にあわせて適切なモデルを選定し、ファインチューニングにより特化モデルを作成します

各種パブリッククラウドサービス／ローカルモデル

Azure OpenAI Service

Amazon Bedrock

各種ローカルモデル

特化モデルの構築技術

データクレンジング

合成データ生成

ファインチューニング

## NTTデータ数理システムによる、生成AI関連ご提案 3つの強み

### 強み1

分析案件実績多数の  
データサイエンティストが  
周辺技術・最新技術を  
幅広く調査したうえで  
最適なお提案が可能です

### 強み2

業務に関わる皆様自身で  
活用していただくため  
GUI制作も含め、  
ワンストップなお提案が可能です

### 強み3

当社の生成AIアセットを活かした  
クイックなPoCの  
ご提案が可能です

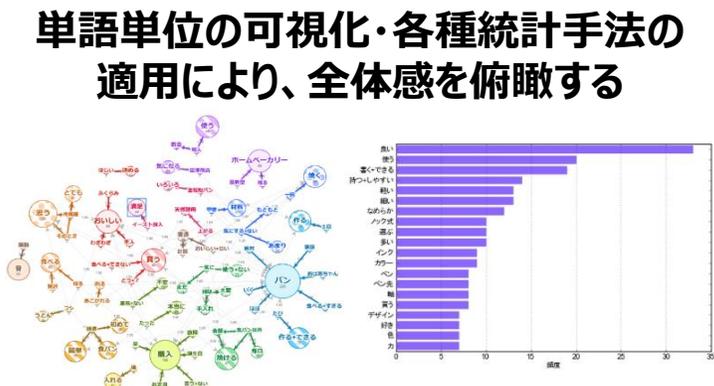
**われわれと一緒に、活用方法を模索しましょう！**

**生成AIをはじめとする幅広い技術要素のなかから最適な組み合わせをご提案し  
生産性向上・新規価値創造を、強力にお手伝いいたします！**

# 【参考】自然言語処理技術の変遷——3ステージに分けて考える

- 単語を機械的に扱う古典的なテキストマイニングの時代から、高度な文脈を扱える生成AIの時代へ

技術の進展	深層学習以前の時代 (-2000s)	深層学習以後の時代 (2010s-)	
技術	自然言語処理ステージ1 (-2000s) 古典的なテキストマイニングの時代	自然言語処理ステージ2 (2010s) 深層学習の時代 I ——Word2Vec	自然言語処理？ステージ3 (2020s) 深層学習の時代 II ——生成AI
できることの例	分かち書き・品詞／構文解析で文章を単語のカテゴリ集合データとして捉えられるように	単語のベクトル化により単語同士の意味関係（類似性）が把握可能に	単語のより高次元なベクトル化により「文脈」を含んだ、単語同士の高度な意味関係が把握可能に
課題	単語単位の可視化・各種統計手法の適用により、全体感を俯瞰する	単語の類似度から文章の類似度を計算している文章を引っ張り出す	今まで人間にしかできなかった高度なアノテーションなど
	類義語等の情報を「辞書」で与える必要があり、その整備に多大な手間	話題の分離や皮肉の理解など高度な文脈の理解に課題	(今のところは) ハルシネーションなどの問題があると言われている

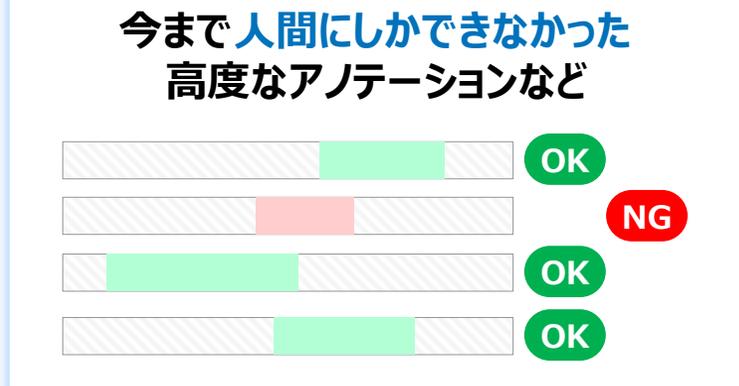


単語の類似度から文章の類似度を計算している文章を引っ張り出す

電池が長持ちする。耐久性がある。  
壊れない。壊れにくい。

類似度 74%

この前トイレに落としたり、動かなくなった。  
買ってから10ヶ月経ってなかったので、買い換えできなかった。  
その程度の水には耐えるよ。



**NTT data**

株式会社NTTデータ数理システム