



オープンキャンパス 研究室紹介

---

## 機械学習研究室（樋口研究室）

発表者：

所属： ビジネスデータサイエンス専攻B4,M1,D2

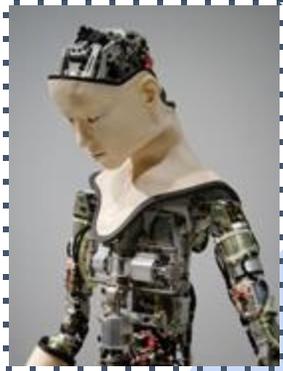
日時： 2024/8/7, 8

1. 機械学習について
2. 研究紹介
3. 研究室の諸情報
4. コラム



# 機械学習について

# 1.1. 機械学習とは



Artificial Intelligence(AI, 人工知能)

広く人間の知能・知性をソフトウェア（プログラミング）で再現する試み

Machine Learning(ML, 機械学習)

AIのうち、特定の問題設定に対して、人間が詳細な指示を出し学習させることで、問題解決が可能なもの

人間のように感情をもつ...?

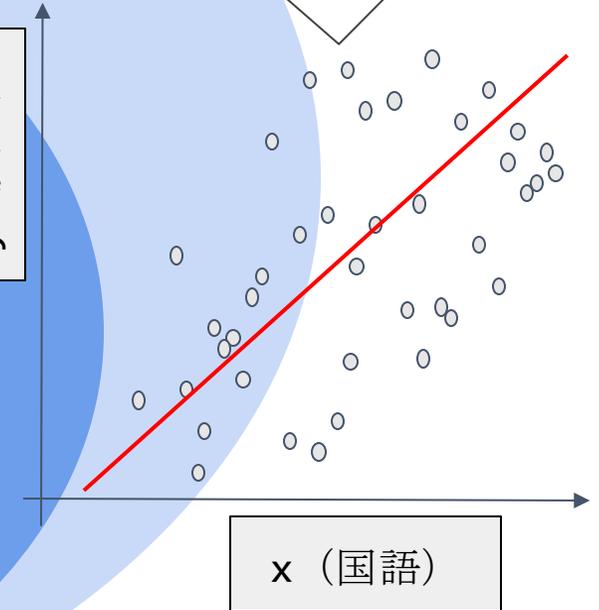
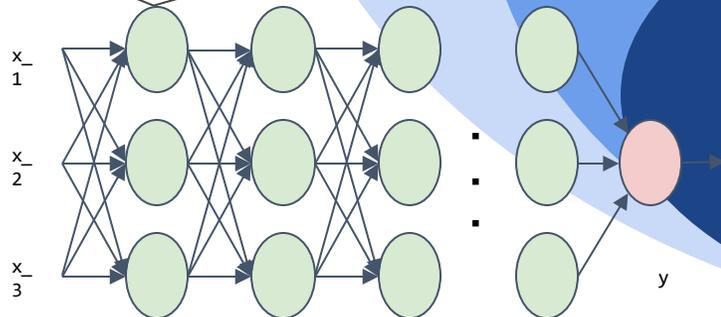
xとyの関係性を、直線(ax+b)で表したい!

y (英語)

xとyの関係性を  
もっと柔軟に表現したい!

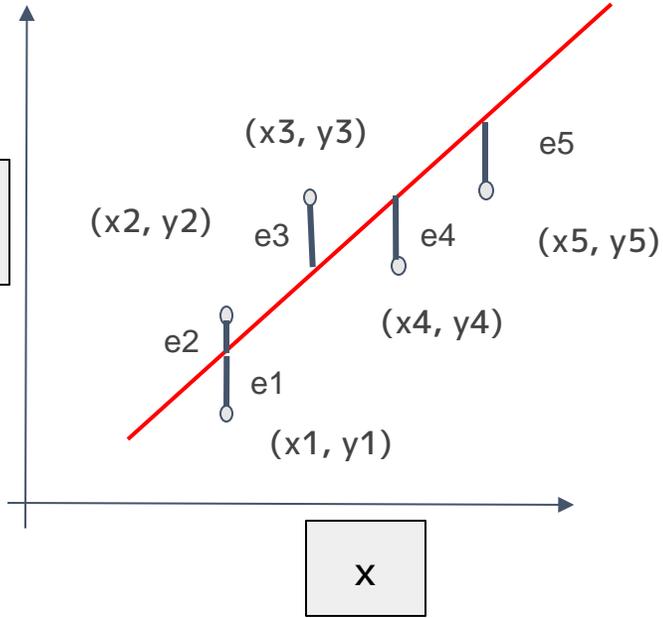
Deep Learning(DL, 深層学習)

機械学習のうち、人間の詳細な指示を一部不要とするもの  
複雑な構造を一部自動で学習できるもの



# 1.1. 機械学習とは

どのように直線（傾きa,切片b）を決めるか？



①2点間のyの値の誤差( の2乗)

$$e_i^2 = (y_i - (ax_i + b))^2$$

②誤差の和

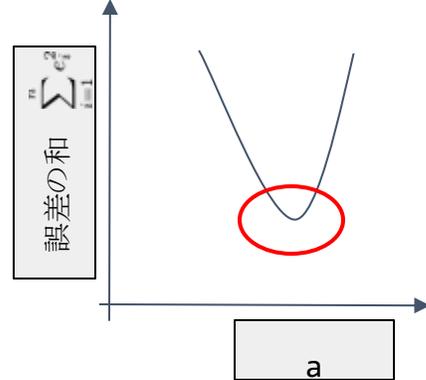
$$\sum_{i=1}^n e_i^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$$

③誤差が小さくするように直線を決めたい(傾きaと切片bを決めたい)

$$\sum_{i=1}^n e_i^2 = \bigcirc a^2 + \triangle a + \text{定数}$$

※②に①を代入し、式変形しただけ  
※bについて同様

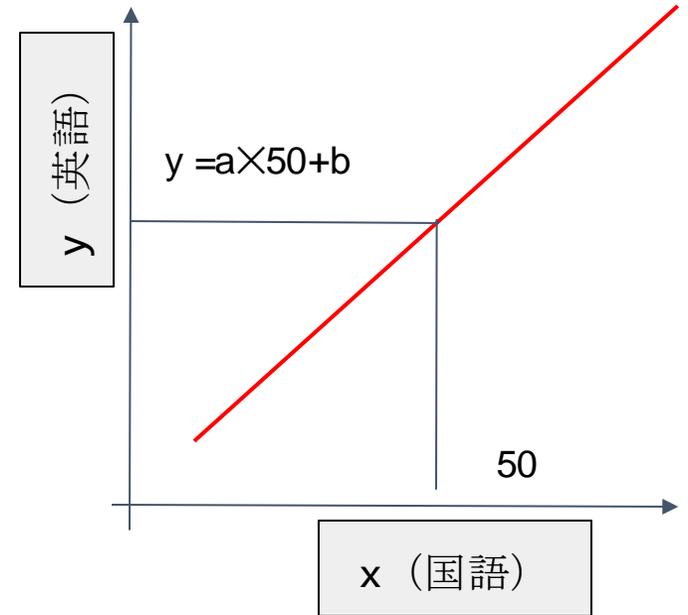
誤差を最小にするaを見つけた！



この直線で何が出来るか？



Q. ある生徒Aの国語の点数が50点の時、英語の点数を、過去のデータから予測したい！



A. 英語の点数はa × 50 + b (a=1, b=1なら51点)



# 研究紹介

### 去年の学部生の研究

- Tリーグにおける入場者数予測
- いちごの葉の病気判別
- 暗号資産の価格予測
- 競馬場の入場者予測
- ゴルフ場の衰退予測

etc..

### 過去の院生の研究

- 各世帯の消費電力推移データ分析によるコロナ禍前後での生活パターンの比較
- 自己教師あり学習手法のCutPasteを用いた製造物品に関する異常検知の性能向上
- 家庭用スマートメータへの適用を目的とする世帯の在住推定手法の開発
- 生産ラインにおける異常検知・非定常サイクル同定のオンラインシステムについて

etc..

### 2024年度学部生

- 旅行記の感情分析とトレンド分析
- 清涼飲料水のまとめ買い・ネット購入の顧客分析
- トラック輸送量の要因分析・時系列予測
- ウェブ・アプリニュースに対する消費行動の分析
- 後樂園キャンパス電力使用量の状態空間モデリングにおける気温効果モデルの改善
- 砂防ダムにおける水位・浮遊物・堆積物の検出および異常検知モデルの構築
- Quasi-periodic Anomaly Detectionシステムを用いた異常検知の有用性の検証



# 旅行記の感情分析とトレンド分析



背景：観光地でオーバーツーリズムにより  
様々な問題発生  
→トレンドを分析して観光客の分散をしたい!



使用するデータ：スケジュールが付属して  
いる旅行記(2007年11月-2022年2月)  
※地球の歩き方旅行記データセットを使用



分析方法：  
1.感情分析 → 旅行記を3つの感情に分割  
2.クラスタリング → 似た旅行記をまとめる  
3.トレンド分析 → 旅行のトレンドをまとめる



# 清涼飲料水の顧客分析

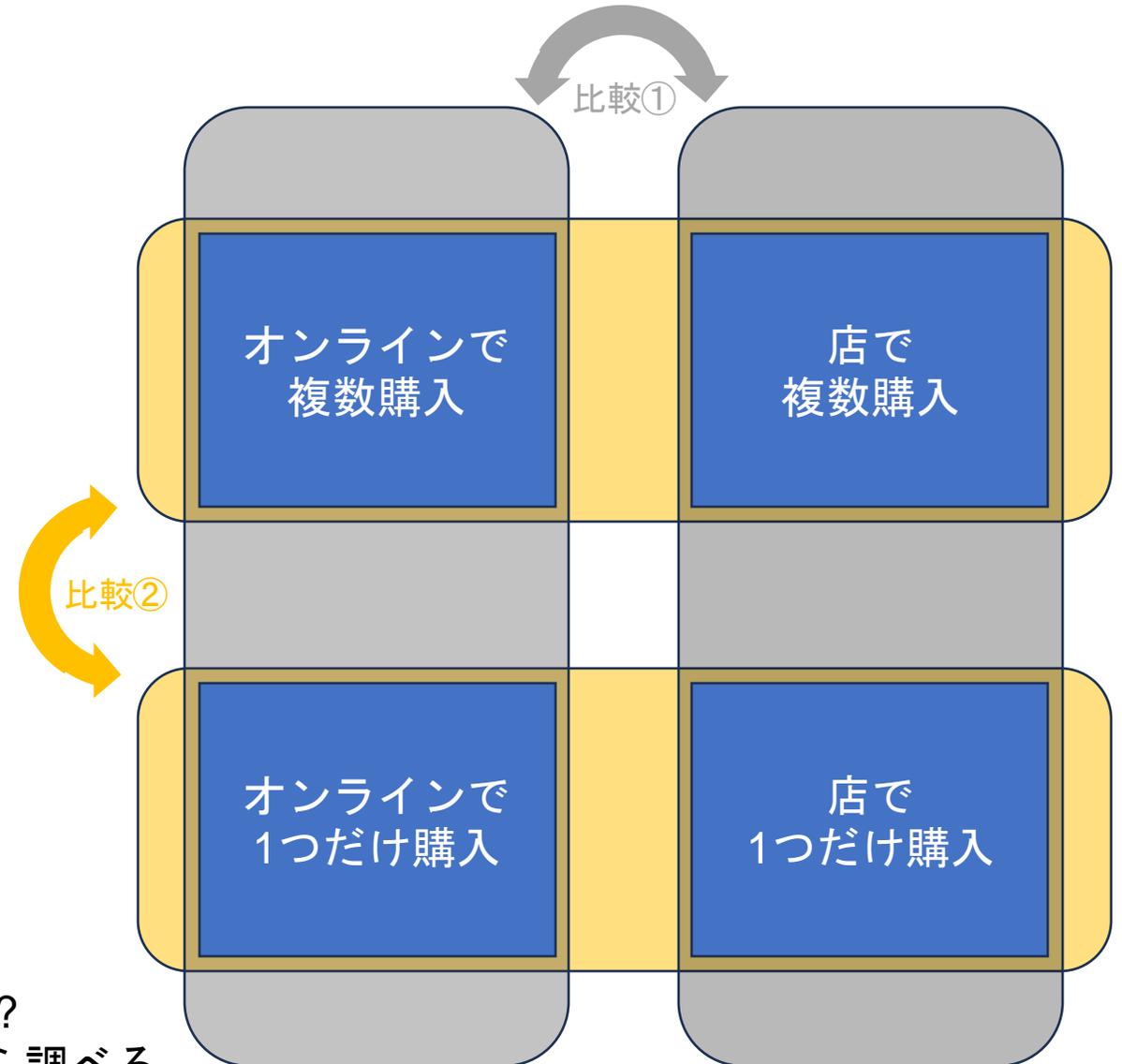
お茶・コーヒーなどの清涼飲料水が  
どんな風に使われているのか分析!

購買データを

- ①商品「1つだけ購入した」「複数購入した」
  - ②商品「店で買った」「オンラインで買った」
- などの軸から分割し、比較する

マーケティングでは、このように  
「行動特性から顧客をグループ分けする」  
“セグメンテーション”という操作が重要です

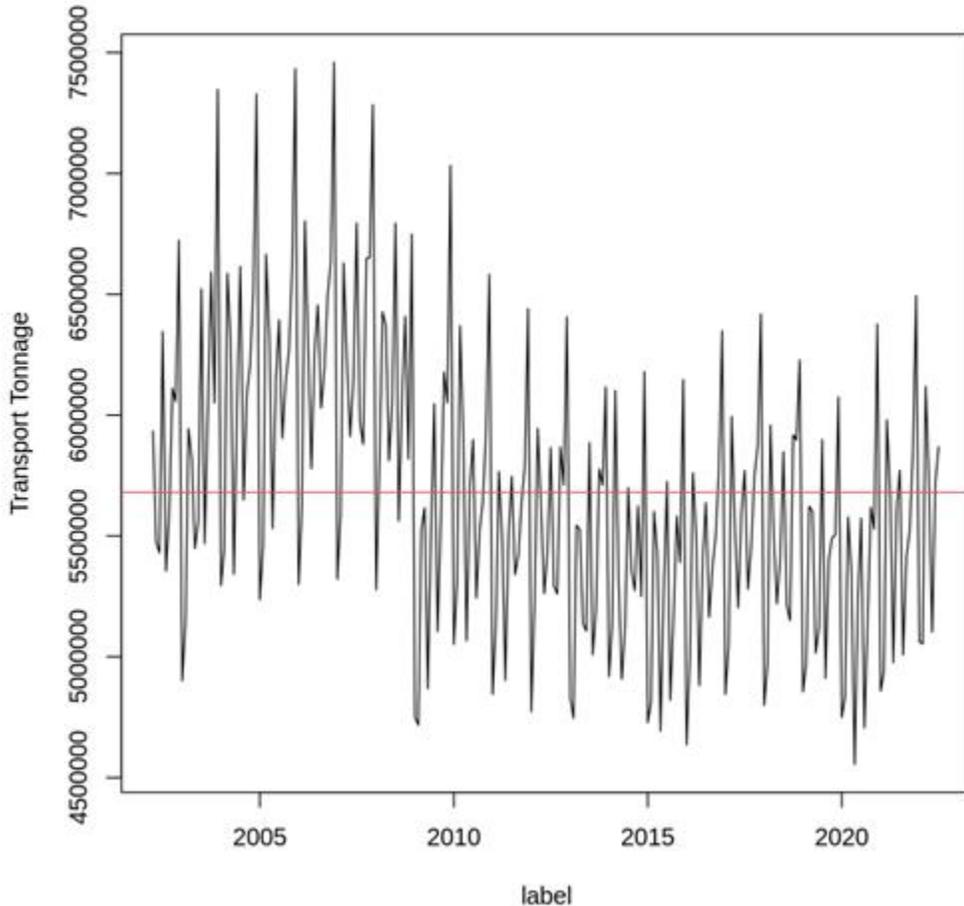
商品をよく買ってくれるのはどんな人なのか?  
オンラインで商品を買ってくれるのはどんな人なのか?  
購買回数や顧客情報(年齢・職業・ライフスタイル)から調べる  
→より効果的なマーケティング&プロモーションに繋がる



# 日本のトラック輸送量の分析と予測



Transport Tonnage Over Years



## 背景

2024年問題・ラストワンマイル問題  
→物流の人手不足をデータサイエンスで解決したい!



## データ

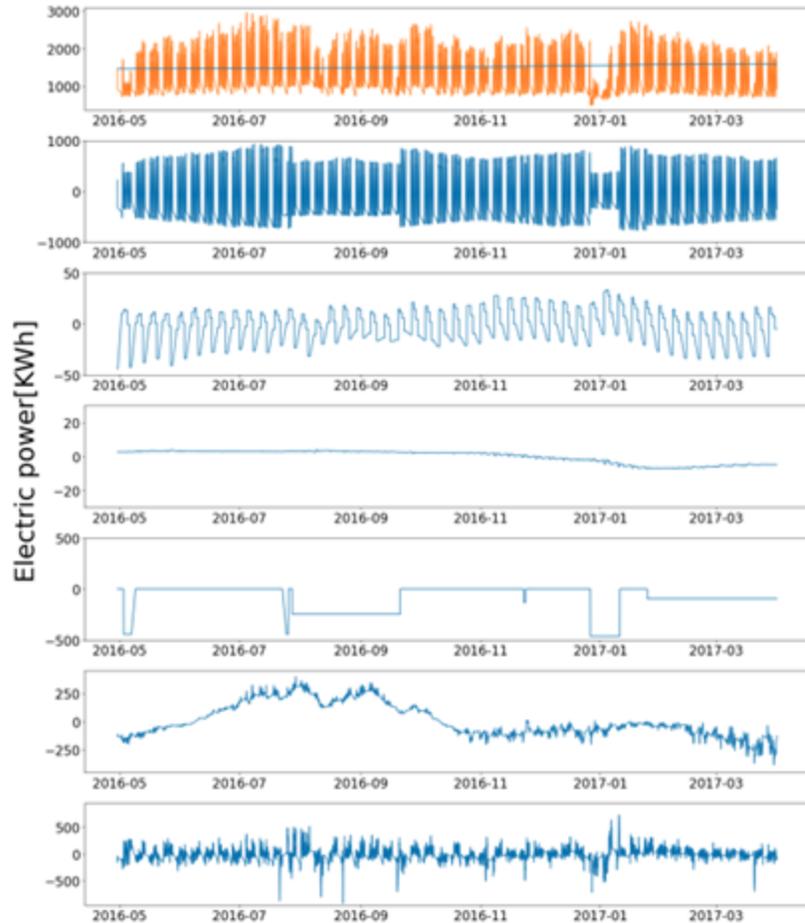
2002年4月から2022年7月までの  
日本全体の特別積合せ貨物のトラック輸送トン数



## 手法

**SARIMA**モデルで予測→仮説を立ててデータ整理を繰り返すことで、モデルの精度を上げていく

# 後樂園キャンパス電力使用量の状態空間モデリング における気温効果モデルの改善



## 背景

現在でも節電が求められる中、後樂園キャンパスの節電対策に役立ちたい！



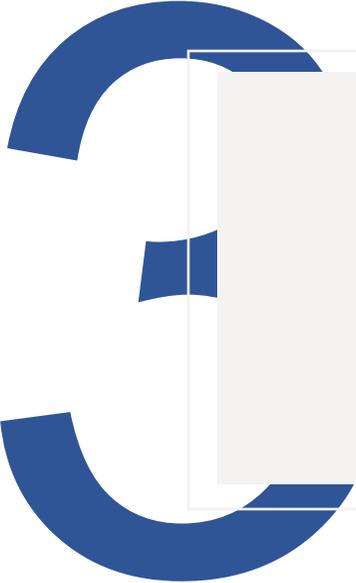
## データ

東京電力様から後樂園キャンパスの電力データを提供していただき使用しています。



## 手法

状態空間モデルを使って電力使用量をモデル化して、具体的な効果などを可視化していきます！



# 研究室の諸情報

### 3.1 必要な知識とスキル

機械学習に必要な知識・スキルは大きく分けて3つ  
(最初から全て持っている必要はなし！)

- ・工学系大学レベルの数学  
(微積・線形代数・確率統計・最適化)
- ・機械学習、深層学習のアルゴリズムへの理解

機械学習の全てのアルゴリズム  
・手法は数学で記述されるため、数学の知識・応用力は欠かせない！

課題背景を理解した上で、ビジネス課題を理解し、解決する力

ビジネス分野

- ・論理的思考力
- ・コミュニケーション能力
- ・語学能力
- ・特定のドメインへの知識
- ・クリエイティビティ

データ  
サイエンス  
分野

データ  
エンジニアリング  
分野

情報処理、人工知能、統計学などの情報科学系の知恵を理解し、使う力

- ・プログラミング
- ・データベースの知識
- ・コンピューターサイエンスの知識
- ・コンピューターの知識

データサイエンスを意味のある形に使えるようにし、実装、運用できるようにする力

## 3.2 研究室で身につく5つの能力

---

- **問題分析力**  
→社会や企業の問題を正確に把握し、その原因を分析する能力
- **数理的定式化能力**  
→特定された課題をDS技術で解決可能な形に変換する能力
- **データ分析力**  
→収集したデータを用いてモデルを構築し、検証する能力
- **エンジニアリング力**  
→開発したモデルを実用的なシステムとして実装する能力
- **マネジメント・コミュニケーション能力**  
→開発したシステムを現場に導入し、効率的に運用する能力

## 教授紹介

### 樋口 知之 教授

(ひぐち ともゆき)

- 東京大学大学院理学系研究科博士課程修了、理学博士号取得
- 中央大学 AI・データサイエンスセンター前所長
- (株)ブリジストン デジタルソリューション本部 特別技術顧問
- 産業技術総合研究所 人工知能研究センター 研究顧問
- 大学共同利用機関法人 情報・システム研究機構 統計数理研究所 名誉教授
- 専門分野はベイジアンモデリング、深層学習と状態空間モデルを統合する深層確率計算



### □ 主な構成メンバー

- 教員：1名， 助教：1名
- 博士課程学生：1名， 修士課程学生：3名， 学部学生：8名（1名休学）

### □ 共同研究先企業・研究での繋がりがある企業

- 東京電力ホールディングス株式会社
- 沖電気工業株式会社

### □ 卒業論文の指導方針

- 研究テーマは自由に決定できる
- 週1のゼミ(1時限分に相当)参加が求められるが、その他は基本的に自由に計画可能

### □ 先輩の就職先（今年で6年目の新しい研究室です！）

- 株式会社博報堂、株式会社サイバーエージェント、アクセンチュア株式会社、株式会社NTTデータなど(今年の学部生の内定先を知りたい方は個別に質問してください)



# ChatGPT : Generative Pre-trained Transformer

- OpenAIが開発した人工知能の一種
- 人間のように自然に対話を行うAI

テキストを入力すると、それに対応するテキストを出力。（有料版は画像データも入力可能）

- 無料でサービスを利用することができる



### ChatGPTについて教えて！

ChatGPTはOpenAIが開発した自然言語処理のためのAIモデルの一つです。ChatGPTは、大規模なトレーニングデータを用いて学習された言語モデルであり、人間のように対話を行うことができます。

ChatGPTは、ユーザーの質問や文脈に基づいて回答を生成する能力を持っています。それにより、一般的な知識の提供、質問への回答、説明の提供など、多くの対話タスクに対応でき、自然な対話を行うために、文脈を考慮した応答を生成することも特徴の一つです。

ChatGPTは、インターネット上の大量のテキストデータを学習しており、様々なトピックに関する情報を持っています。ただし、注意点として、ChatGPTは情報の正確性を保証するものではなく、訓練データに含まれるバイアスや誤った情報を反映する可能性があることを覚えておきましょう。

OpenAIは、ChatGPTの利用者が倫理的な使用を守ることを推奨しており、モデルが偏見のある内容を生成することを防ぐための対策を行っています。

最後に、ChatGPTは常に進化しており、新しいアップデートや改善が行われています。そのため、最新の情報や機能についてはOpenAIの公式発表やドキュメントを確認することをおすすめします。

# ChatGPT の活用例

- 文章の要約
- 翻訳
- 計算
- プログラミング
- 文章作成
- 文章の添削



# ChatGPT の注意点



- 必ずしもすべての回答が正確であるとは限らない
- 安全性を確保するために、個人を特定できる情報や機密情報は入力しない
- 長い会話や複雑な質問の場合、誤解を招くことがある
- 簡単な推論を間違える可能性がある
- バイアスを含んだ回答をしてしまう可能性がある



---

ご清聴ありがとうございました！

質問がある方は気軽にメンバーへ話しかけてください！