



生成AIとビジネス戦略

正しく恐れ正しく活用するために

岐阜大学工学部電気電子・情報工学科

准教授 鈴木 優

- 鈴木 優 (すずき ゆう)
- 情報科学, データ工学, データサイエンス
- 略歴
- 岐阜大学工学部電気電子・情報工学科 准教授
- 詳しくはWikipediaで

≡  ウィキペディア
フリー百科事典

Q ∞ 寄付 アカウント作成 ログイン ...

鈴木優 (情報工学者)

文 言語を追加 ▾

ページ ノート 閲覧 編集 履歴を表示 ツール ▾

出典: フリー百科事典『ウィキペディア (Wikipedia)』

鈴木 優 (すずき ゆう、1977年〈昭和52年〉^[1] -) は、日本の情報工学者。工学博士。専門はメディア情報工学。多元ビッグデータ解析の応用研究を専門とする^[2]。

略歴 [編集]

1999年（平成11年）3月、神戸大学工学部情報知能工学科を卒業し、2004年（平成16年）3月に奈良先端科学技術大学院大学情報科学研究科で工学博士を取得した。博士取得後は立命館大学情報理工学部情報コミュニケーション学科講師、京都大学大学院情報学研究科特定研究員、名古屋大学情報基盤センター研究員、同特任助教を経て、2013年（平成25年）4月より名古屋大学大学院情報科学研究科の特任助教に就任した^{[3][4]}。2014年（平成26年）9月からは母校の奈良先端科学技術大学院大学情報科学研究科の知能コミュニケーション研究室において特任准教授を務めた後^{[3][5]}、2019年（平成31年）4月より岐阜大学工学部の特任准教授、2021年（令和3年）4月より准教授を務めている^[3]。

今日のまとめ

3

- LLMは自分の仕事の一部を任せられる仕組み
 - 適材適所で使うことによって仕事が楽になる
 - まずは使ってみるという勇気は必要
- LLMに完全に仕事をさせてはいけない
 - 誤った結果である可能性はある
 - 正しいか誤っているかの知識は必要
- 過度に恐れない。過度に信用しない。

注意書き

4

- この資料は分かりやすさを優先させているため正確ではない情報が含まれます。
- 正確な情報が必要な場合には専門書など詳しい情報源をお使いください。
- この資料の一部はLLMで作成されています。
- この資料でLLMと機械学習とニューラルネットワークはほぼ同じ意味で使っています。

機械学習とは？

5

- 分類
- 牛肉を産地ごとに分類



アメリカ産



飛騨牛

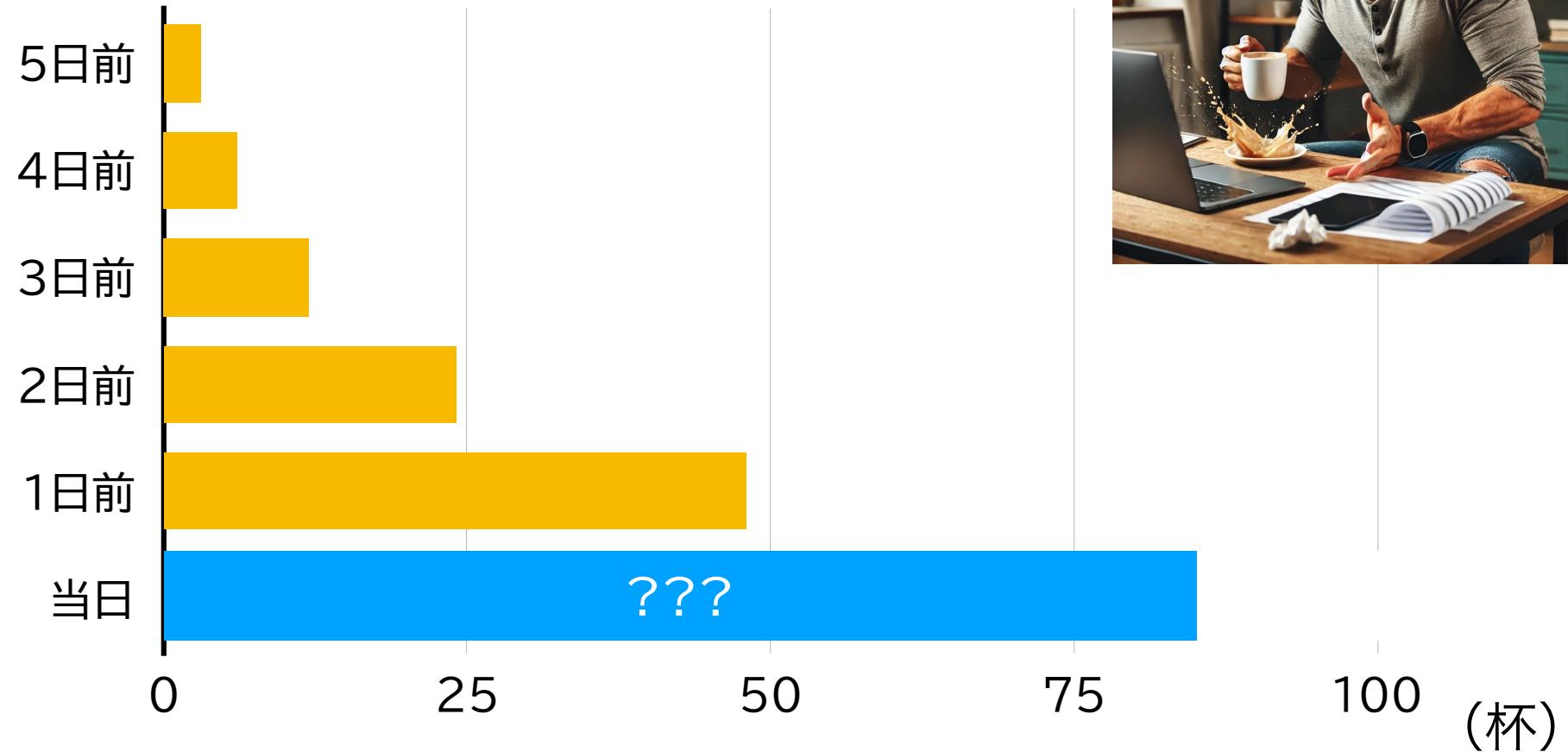


神戸牛

機械学習とは？

6

- 予測（回帰）
- 論文締切までの時間とコーヒーの量



LLMとは？

9

- ・どのような質問でもその質問に合った回答を返してくれる仕組み

質問の回答

日本の首都は？

東京です

解説

地球温暖化について
教えてください

地球温暖化は、主に化石燃料の燃焼や森林破壊によって発生する温室効果ガスの増加が原因で、地球全体の気温が上昇する現象です。

文章創作

短い小説を書いて
ください

むかしむかし、小さな村に勇敢な少女が住んでいました。村は長い間、暗い森の奥に住むドラゴンに脅かされていました。ある日、少女は村を救うために立ち上がり、ドラゴンとの戦いに挑む決意を固めました。

アドバイス

英語の勉強法について
アドバイスして

英語の勉強には毎日の習慣化、会話練習、語彙の拡大がおすすめです。

- 文章生成やコンテンツ作成
 - 記事, ブログ, 広告, メール内容, 小説などの執筆補助
- 教育や学習支援
 - 疑問への回答, 学習計画の作成
- 翻訳や要約
 - 複数言語の自動翻訳や長文の要約
- 考えていることの整理
 - 膨大な文献の解析, データ要約, アイデアのブレインストーミング
- カスタマーサポート
 - チャットボットや自動応答システムなど

正しい返答をしないLM

11

岐阜市役所周辺のレストランを教えてください

1. 割烹 風月

特徴: 地元の新鮮な魚介類と和食が楽しめる老舗の割烹。
おすすめ: 季節ごとの旬の素材を活かした一品料理。

実在せず

2. イタリアンレストラン ラ・テラス

特徴: 落ち着いた雰囲気で、上格的なイタリア料理を提供。
おすすめ: 自家製パスタやソースとのペアリング。

実在せず

3. 中華料理 龍吟

特徴: 香り豊かな本格的な中華料理を楽しめるアットホームなお店。
おすすめ: 麻婆豆腐やチキンマーボーなどの定番メニュー。

実在せず

正しい返答をしないLM

12

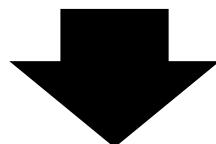
あるカフェでは23個のりんごがあります。ランチのときに20個使い、6個買い足したとき、いくつのりんごがあるでしょう

答えは27個です

今、何時ですか？

わかりません

- なぜ正しい返答をしないのか？
- どのような仕組みで返答を作っているのか？



仕組みを理解する

LLMの仕組み

13

- 予測を使って文を組み立てる

今日は天気が

良い 60%

悪い 20%

曇り 10%

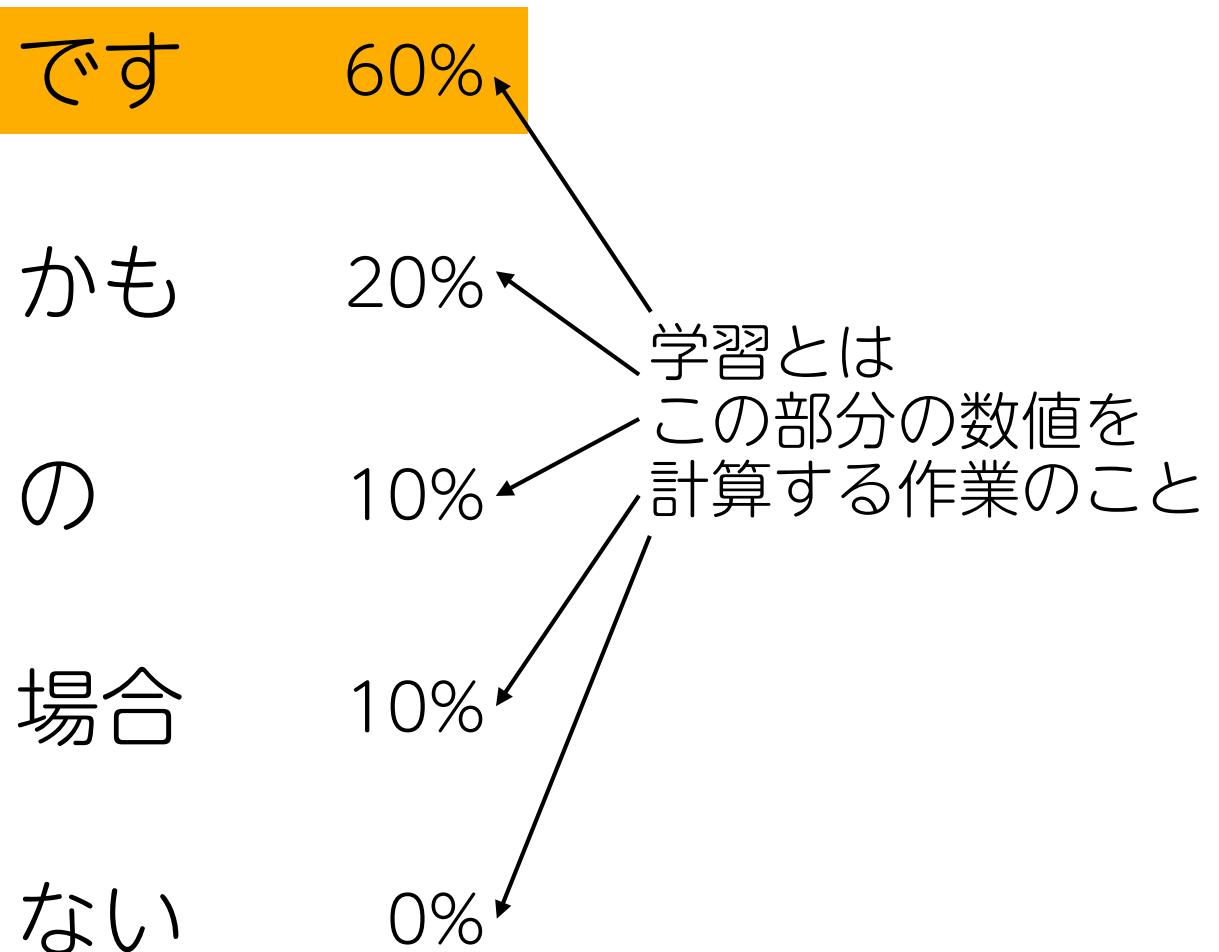
雨 10%

楽しい 0%

LLMの仕組み

14

今日は天気が**良い**



- 記憶するパラメータが多いほど性能は向上
 - GPT-3で1,750億個 (175GB)
 - 現段階で一般向けの一番高いGPUで32GB (40万円)
- 計算時間が問題
 - CPU (一般に使われるコンピュータの頭脳) では遅すぎる
 - GPU (ゲームで使われる映像処理用の仕組み) を活用
 - できることはCPUよりもかなり少ない
 - 速度はCPUと比較にならないほど速い
- 機器が手に入らない
 - NVIDIAやAMDなどの企業が注目されつつある
 - 元々ゲームのための装置メーカー
 - 今はAI構築に必須のGPUメーカー
 - 世界中でGPUを取り合い

何を学習しているのか

16

- 公開されているテキスト
 - Wikipedia, ウェブサイト, 書籍, ニュース記事など
- ライセンスされたデータ
 - 書籍や資料など
- 人間が生成したデータ
 - 専門家やトレーナーが作成した対話例やフィードバック

正しい情報かどうかを確認して
学習しているわけではない

何を学習して「いない」のか

17

- 学習の時に存在しないデータ
 - GPT-4でも2023年8月ごろ以降は学習していない
- 入手が不可能なデータ
 - データがOpenAI社（など）で入手できないデータ
 - 社内だけで使われるデータなど

ハルシネーション（幻覚）とは？

18

1. 架空の引用・出典の生成

「最新の医学研究に関する論文を教えて」と尋ねると「スマスら(2019)の研究によると....」などと返答するが、このスマス氏は実在せず、論文も存在しない。

2. 誤った事実の提示

「オーストラリアの首都を教えて」と尋ねると「オーストラリアの首都はシドニーです」と返答するが、誤り。

ハルシネーション（幻覚）とは？

19

3. 架空の統計データや数値の生成

「岐阜駅の利用者数を教えて」と尋ねると、「1万人前後です」と回答するが誤り。

4. 文脈と合わない情報を急に生成

「カレーの作り方を教えて」と尋ねると、カレーのレシピを返答するが内容が急に金融市場の話題へ変化する。

Wikipediaでの事例

20

・アンベルリヒサル(Amberlihisar) の記事

Amberlihisar (アンベルリヒサル) は、トルコのトラブゾンに位置するオスマン帝国の要塞です。1466年に征服者メフメトによって、アク・コユンル（「白羊トルコ人」とも呼ばれる）に対する防衛のために建てられました。[1]

この要塞はトラブゾン市を見下ろす丘の上にあり、オスマン帝国の軍事力を示す証となっています。高い城壁と多数の塔や堀防を備えた大規模で堂々たる構造で、攻撃時の避難経路として、また物資の保管場所として利用された地下の通路や部屋も数多く存在します。[2]

また、Amberlihisarは軍事的な役割に加えて、トラブゾンの文化生活にも貢献していました。多くの重要な役人や高官が居住していただけでなく、学問や文化の中心地としての役割も果たしていました。[3]

さらに、Amberlihisarには姫にまつわる伝説があります。伝説によれば、その姫は美しく才能に恵まれ、音楽と詩においてその美貌と技量で名高い若い女性でした。彼女は要塞内で詩を作り、リュートを奏でながら日々を過ごし、その音楽と美しさを求めて遠方から多くの人々が訪れたと伝えられています。[4]

歴史的・文化的な重要性にもかかわらず、Amberlihisarはトプカプ宮殿やブルーモスク（スルタンアフメト・モスク）などの他のオスマン帝国のランドマークほど広く知られてはいません。しかし、トルコの豊かな文化遺産に興味がある人にとっては、オスマン帝国の重要で魅力的な歴史の一部として訪れる価値のある場所です。[5]

などなど2000語以上

Wikipediaでの事例

21

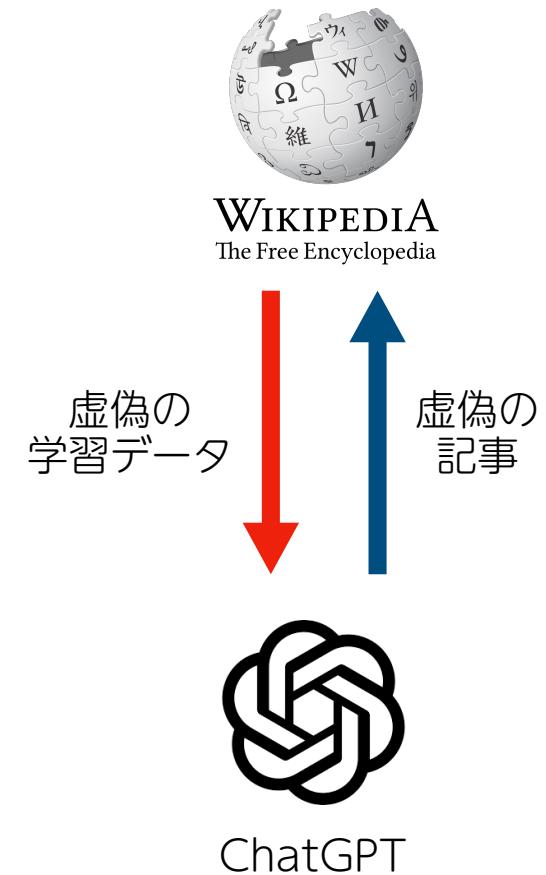
- アンベルリヒサル(Amberlihisar)とは？
 - トルコ・トラブゾンにあるオスマン帝国の要塞
 - 1466年にメフメト大帝がアク・コユンルに対抗するため建設
 - 防衛機能だけでなく、役人の居住地や学問の中心地としても機能
 - 美しい音楽と詩で知られる姫にまつわる伝説
 - オスマン帝国の豊かな歴史を感じられる貴重な史跡

このような要塞の存在が確認されていない

参考文献が一つも見当たらない

Wikipediaでの事例

- Wikipediaに掲載された記事は学習に使用
 - それがハルシネーションによるものであったとしても
- 学習したデータを使ってLLMは文章を生成
 - 虚偽データを使って作られた文章は当然虚偽
- 虚偽データを使ってWikipedia記事を生成



虚偽データの再生産が行われはじめている

WikiProject AI Cleanup

23

- WikipediaからLLMで生成された虚偽の情報を取り除くプロジェクト
 - AIが書きそうな記述を特定
 - 本当にその記述が事実かどうかを検証
 - 存在しなければ削除

岐阜市役所での事例

24

- 岐阜市生成AI活用ガイドラインを策定
- <https://www.city.gifu.lg.jp/info/seisaku/1021670/1025012.html>
- ルールを守って使用する
 - 利用のときは所属長に申請
 - 機密性のある情報を入力しない
 - 生成物に虚偽や古い情報が無いかを確認
 - 利用や公開のとき誰かの権利を侵害していないか調査
 - 生成物は加筆・修正し、そのまま使わない

ハルシネーションの原因

25

- 過学習によるもの
 - 機械学習はほぼ失敗、たまに成功
 - 過学習は失敗の代表例
 - 教えられたものは完璧に回答できるが、教えられていないものは全く回答できない
 - 記憶はできるけど応用ができない

ハルシネーションの原因

26

- データの偏りと思い込み
 - 看護師の採用をAIが行うことを考える
 - 看護師は女性であるという思い込みがある
 - 男性看護師を受け入れた実績 (=学習データ) が無い
 - AIは男性だという理由で不採用と判断する場合がある
 - 機械学習モデルは偏見の有無を自身で判断できない
 - 自分が偏っているかどうかを判断できない
- 偉人は男性であるという思い込み
 - Wikipediaに掲載されている多くの歴史的人物は男性
 - 無意識で差別を行ってしまう可能性

機械学習の重要な問題点

27

- なぜうまくいっているのか誰も分かっていない
 - 研究者であっても分からない
 - 数百数千数万という大量のパターンを試して1個うまくいくかどうかという世界
 - どうやつたらうまくいくのかがよく分からない
- なぜ“出力が”うまくいっているのかわからない
 - どのような思考過程で回答が出ているのかは不明
 - XAI (Explainable AI; 説明可能なAI)についての研究が進行中

研究者側の課題

28

- ニューラルネットワークの解明
 - どのようなメカニズムで出力されているのかを明らかに
 - 不適切な出力が出ないような仕組みを開発
- 傾りの解消
 - 学習データは今までの人類の歴史の一部
 - 歴史上存在している偏り、偏見はLLMが学習する
 - 傾りを再生産している
- LLMが有効な業務、ビジネスの調査
 - 有効に使える状況を提供
- 精度の高いLLMの構築
 - アメリカでは1年で77兆円規模
 - 中国のDeepSeekが格安できた = 9億円程度
 - 日本でもこのような規模の研究を行うことができる環境を（共同研究を）

使う側の対応

29

- 自分でやらなくていいことをAIに任せる
 - 単純な作業は精度良くできる
 - テンプレートのような作業も楽にできる
 - メールの返信
 - Excelでファイルを操作
 - 簡単なプログラミング(ExcelのマクロやPythonなど)
 - ChatGPT Pro(月30万円) で作成できるレポートは人間で行うと数ヶ月かかるレベルのものが数分ができる
 - ちょっとした動画も作成可能
- あくまで補助ツールであることを認識する
 - 全部を信頼すると誤ったものを見破れない
 - 頼りすぎると自分のスキルが落ちる
 - 特に検索目的で使用するときには注意する

高等教育における情報

30

- 情報学の分野を学ぶ学生がどんどん増加
 - 共通テスト（旧センター試験）で情報が追加
 - 高校で情報という科目が増加
 - 大学でもデータサイエンスに関する講義が増加
 - 学部学科を問わず全ての分野で必修
 - 大学を問わずどの大学でも必修化
 - データサイエンティスト育成に関する補助金も増加

誰でも情報技術が活用できる時代に

今日のまとめ

31

- LLMは自分の仕事の一部を任せられる仕組み
 - 適材適所で使うことによって仕事が楽になる
 - まずは使ってみるという勇気は必要
- LLMに完全に仕事をさせてはいけない
 - 誤った結果である可能性はある
 - 正しいか誤っているかの知識は必要
- 過度に恐れない。過度に信用しない。