

ChatGPTとは

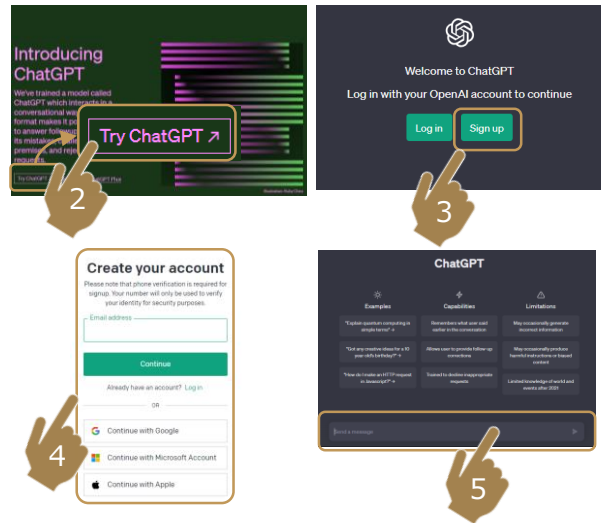
はじめに

2022年末に公開された「ChatGPT」は人間と会話しているかのような自然な受け答えが可能なAIで、無料で誰でも使えることから国内外問わず爆発的に普及しました。第117回医師国家試験を合格するほどの性能¹⁾を持っており、医学教育や研究、臨床においても無視できない存在です。

本展示はChatGPTの適切な使い方や教育・研究における注意点をまとめました。

ChatGPTとは

ChatGPTは2022年11月にOpenAIが公開したチャット型の文章生成サービスです。利用者が日常的に用いる言語(以降、自然言語)を使って命令するだけで、まるで人間と話しているかのような対話を行うことができます。



プロンプトの書き方

プロンプトはできるだけ明確で詳細に記述します。別のテキストエディタで書いてから、チャット欄にコピーするのがオススメです。チャット欄に直接記入する際は、Shift+Enterで適宜改行します。以下は出力結果を最適化するための工夫です。

- 1. 設定(条件)を与える**
 - 「あなたは〇〇です」と役割を与える。
 - 「私は初学者です」など難易度をコントロールする。
 - 細かな条件は箇条書きで指示する。
- 2. 出力内容を指定する**
 - 文字数や表現のテイストを指定する。
 - 「見出しつきで」「表形式で」、「箇条書きで」など出力形式を指定する。
 - 「YES/NOで答えて」など制限をつける。
- 3. 回答の具体例を示す**
 - どんな回答がほしいのか例示する。
 - ×「文体は常体で…」
 - 「文体は常体(～だ、～である)で…」

様々なプロンプトがウェブ上で共有²⁾されており、それらを参考にすることで質の高い回答を得ることができます。

ChatGPTについて150字程度で説明してください

ChatGPTは、OpenAIが開発した大規模な言語モデルです。GPT-3.5のアーキテクチャをベースしており、自然言語の理解や生成に優れています。ChatGPTは、ユーザーとの対話を通じて質問に答えたり、情報を提供したりすることができます。様々なトピックに関する知識を持ち、文章の文法や文脈を考慮した回答が可能です。日常会話や情報検索、創造的な対話など、幅広い目的に利用することができます。ただし、一部の情報は最新のものではないため、注意が必要です。

ChatGPTにChatGPTについて質問しました。これを“専門用語を使わずに”とお願いしたものが次です。

ChatGPTについて専門用語を使わずに150字程度で説明してください。

ChatGPTは、オープンAIが開発した人工知能プログラムです。自然言語処理を利用して、人間のような対話を行うことができます。多くの文章データを学習しており、質問に答えたり、会話をするのが得意です。一般的な知識や情報を持っており、様々なトピックについての質問に対応することができます。

専門用語が何かを具体的に指定していないのに、きちんと平易な文章にしてくれます。このように利用者の入力に応じて自然な回答をしてくれます。

使い方

ChatGPTは2023年6月現在、無料版と有料版の2種があります。ここでは無料版の使い方を紹介します。Googleなどのアカウントを使ってログインすることで簡単に利用できます。その後はチャットするように命令文(以降、プロンプト)を書くだけです。

- ウェブブラウザでChatGPTにアクセス。
- 「Try ChatGPT」をクリックする。
- 「Sign up」する。
- アカウントを作成する。
- プロンプトを入力する。

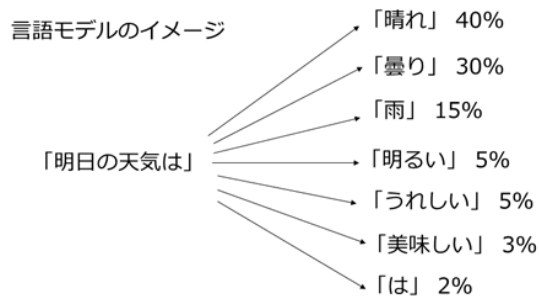
※ChatGPTなど生成AIへの命令文を**プロンプト**といいます。

ChatGPTの仕組み

GPTの仕組み

ChatGPTは、GPTという文章生成のための言語モデルをチャット用にチューニング(調整)し、対話可能なインターフェースを付与したものです。

GPTはインターネット上で収集した45テラバイト(約4兆単語)もの文章データを学習し、単語間のつながりの強さを確率的に表現しています。



3) <https://www.intelliink.co.jp/column/ai/2021/031700.aspx>

▼自然言語処理技術 “Transformer”

ChatGPTは正式名称を“Chat Generative Pre-trained Transformer”といいます。意識すると「チャット用の事前学習済みTransformer」です。

Transformerとは2017年にGoogleが発表した新しい自然言語処理技術で、従来の自然言語処理では文頭から文末まで順に作業していたのに対して、文全体を並列的に処理しながら文章内の単語間の比重を評価できます。これによって「文脈を加味した自然な文章生成」が可能となりました。

The trophy would not fit in the brown suitcase because **it** was too big. What does “it” refer to ?

In the sentence you provided, “it” refers to the trophy.



ChatGPT は文脈からitがtrophyであるときちんと判定できますが、スーツケースが容器で、トロフィーがそこに入れるものであるという常識的な関係をAIが学習するのは実は困難です。このようなテストは常識推論と呼ばれます。なお文中の**too big**をtoo smallにするとitはsuitcaseだと判定できます。

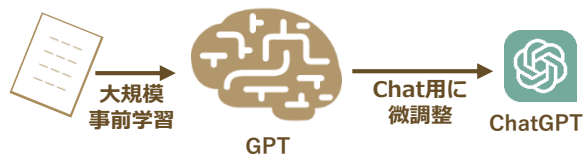
▼事前学習(Pre-Trained)とは

GPTはこのような自然さ、流暢さを獲得するために、収集した莫大な文章データを用いて「文章の次の単語を予測する」というタスクを自動で行います。これにより言語の構造を統計的に学習した大規模な言語

モデルを構築します。GPTの革新的な点は十分な事前学習によってある程度の汎用性を持っている点です。

▼微調整

GPTをチャット用に微調整したのがChatGPTです。非倫理的な回答や虚偽の情報を生成しないように最低限の調整作業が行われています。これにより利用者の入力に応じた多様な受け答えができる余地が残されています。



言語モデルをある目的に特化するように調整することをFine-Tuningと言います。

Fine-Tuningの方法を変えれば別種のサービスを実現できます。つまりChatGPTはGPTを使った1例でしかなく、例えば医学論文執筆に特化したものなど、自然言語を扱う多様なサービスが今後登場するかもしれません。

生成される文章と注意事項

以上の仕組みから、**ChatGPTは入力に対して“確率的に”文章を生成します。具体的には文全体の意味から次の単語を予測し、評価値に基づき文章を構築しています。**つまり情報を検索し真偽に基づいて回答をしているわけではありません。あくまでも確率的な出力であるため、ChatGPTはときに誤情報を出力します。この現象はハルシネーション(幻覚)と呼ばれます。

なお同じプロンプトを入力しても出力は毎回異なります。ランダム性を与える仕組みが組み込まれているため、多様性のある回答ができるようになっています。

GPTのバージョン

GPTにもバージョンがあります。無料版のChatGPTが参照するのはGPT-3.5と呼ばれるバージョンで、有料プランでは最新のGPT-4が利用できます。GPT-4では文章の質が向上しハルシネーションの頻度が低下しています。

ChatGPTと教育・研究①

ChatGPTの得意・不得意

▼得意

正確さを要求しない用途であれば多様な使い方ができます。例えば素案作成や文章の要約・翻訳、アイデア創出などです。英語学習の教師役や、プログラミングにも利用できます。

▼不得意

情報検索を依頼する	文章を生成するソフトであり、情報検索を代行しているわけではありません。
最新情報を質問する	ChatGPT(正確にはGPT3.5)には2021年9月までのデータしかありません。
計算させる	ChatGPTには計算機能はありません。四則演算を理解しているわけではなく、確率的に予測した回答をします。

以下はChatGPTが出力した参考文献の一つです。

Smith J, et al. Antibiotic therapy versus appendectomy for the treatment of uncomplicated acute appendicitis: systematic review and meta-analysis of randomized controlled trials. *BMJ*. 2017;357:j1703.

一見するととってももらしいですが、実はこの論文は存在しません。これがハルシネーションと呼ばれる誤情報生成の例です。**ChatGPTはあくまでも「言語を統計的(確率的)に扱っているもの」であり、例えば検索機能も計算機能も備わってません。**どんなに正しそうに見えてもファクトチェックは欠かせません。このようなリスクを十分に理解して使用することが必要です。

本学の方針

2023年6月23日に「東邦大学における生成系AIの利用に関する方針」⁴⁾が公開されました。ChatGPT等の生成AIを使用する上でのリスクと留意事項が明記されています。一部を抜粋します。

各種報道等にすでに言及されているとおり、生成系AIの利用には負の側面が伴い、具体的には以下のようなリスクが考えられます。本学の学生・教職員におかれては、これらを常に念頭に置き、以下の留意事項として挙げた点を熟慮の上、個々人の責任のもと、生成系AIのメリットを上手く活用し、教育・学修・研究等の諸活動に取り組んでいただきたいと思います。

リスクの種類	内容	留意事項
不正行為のリスク	生成系AIは、インターネット等でパブリックに公開されている著作物や発表物、調査結果等を基に学習しているため、生成系AIから出力された回答をそのまま利用することは、間接的に他人の著作物を剽窃してしまう、また著作権を侵害してしまうことにつながり得る。	生成系AIから出力された回答を、そのまま自分の課題回答・研究成果等として提出することは不正行為に該当する。部分的に利用する場合であっても、どの部分が生成系AIによって出力されたものか、また、どのような生成系AIを利用したかなどについて示す必要がある。
誤情報等のリスク	そもそもインターネット上には、バイアスや誤りを含む情報も存在していることから、生成系AIから出力される情報も同様にバイアス・誤り等を含む可能性が十分にある。	生成系AIから出力された回答は、必ずしも正しいとは限らず、誤りが含まれることも多い。したがって、生成系AIから出力された回答を利用する場合には、その内容について、信頼できる情報源によって確認することを怠らないようにする必要がある。
情報漏えいのリスク	厳密なセキュリティ体制下で管理されるべき情報を、生成系AIに対しての質問や指示として入力すると、生成系AIを媒介してサービス提供事業者がそれらの情報が知れ渡る、あるいはほかの誰かが入力した質問等の回答として返される、といった情報漏えいリスクがある。	学生・教職員等の個人情報や、未発表の研究データ・研究成果・論文等、厳密なセキュリティ体制下で管理しなければならない情報またはそれに準ずる情報は、絶対に生成系AIに対しての質問や指示として入力してはならない。

(※)特に学生の皆さんに対しては、生成系AIに依存して課題や論文等を作成することについて、知識習得や必要な情報を吟味する能力・表現力・独自性等の伸長を妨げるリスクもあるため、各教員から個別に利用に関する指示・注意が出される場合があります。必ず指示等を確認するようにしてください。

ChatGPTと教育・研究②

他大学の例

本学の他にも多くの国内大学がChatGPT等の生成AIに関する情報発信や指針の作成を行っています。

▼東北大学の例

例えば東北大学では学生と教員向けに2つのサイトを作成し、ChatGPT等の生成AIに関する「留意事項」をまとめています。以下はその内容を抜粋したものです。

学生向け⁵⁾

- AIの出力をレポート等の解答にそのまま利用することは自身の勉強にならない
- 授業によってはAIの利用を禁止しており、場合によっては剽窃とみなされる場合がある
- 調べ学習等で使用する場合も、AIの出力には誤りが混ざっていることも少なくなく、AIの出力が正しい内容か、誤った内容なのか、自身でしっかり確認する必要がある
- 未発表の論文や秘密にすべき情報（個人情報やプライバシー情報等）を生成系AIに入力してしまうと、それらの情報が意図せず流出・漏えいしてしまう可能性がある点に気をつける

教員向け⁶⁾

- 生成系AIの利用を完全に排除することは現実的ではない（特に授業時間外）
- 学生が生成系AIをレポート作成等に利用することで、学生自身の学びや厳格な成績評価の点で大きな問題が生じ得る
- 多くの学生が使用することも想定した上で、必要に応じて対策を講じる
- 使用する際は研究内容や機密情報の漏洩に注意する

東北大学は利用する/されることを前提とした留意事項となっています。このように大学ごとにスタンスが異なります。なお東北大学の教員用サイト⁶⁾には国内大学の関連情報がリンク集としてまとめられています。

学術出版社の対応

各出版社は論文投稿に関するChatGPT等の生成AI(本項ではAIと略称する)の取り扱いを発表しています。現状は出版社ごとに対応が異なっており、どこまでなら使用して良いのかという線引は明確ではありません。詳しくは投稿先の出版社サイトを確認してください。

ここではICMJE(国際医学雑誌編集者委員会)の定めたAI使用に関する推奨事項⁷⁾を

紹介します。

- 使用したかどうか開示し、使用した場合は用途について説明する。
- AIは著者となることはできない。
- 著者はAIの利用を含めて論文全体に責任を負う。
- AIを使用した場合は慎重にレビューして編集する。
- AIによって生成されたものについて盗作(≠不正行為)がないと主張できる。

検知は困難

ChatGPTの利用を検知することは困難です。例えばChatGPT開発元のOpenAIが発表した検知ツールの感度は26%⁸⁾だといいます。その後、99%で検知できたという論文⁹⁾も出ましたが、専門家の書いた論文とChatGPTの生成した論文の一つのセクションを比較したもので、汎用性のある検知ツールとは言えません(ある目的に特化できれば検知ツールが作れると言う示唆とも言えます)。

現実的にはChatGPTが生成した誤情報や違和感(でき過ぎなど)を頼りに不正を疑うこととなります。レポート等の課題を課す際はChatGPTが不得意な内容にするなど、不正を困難にするか検知可能にする工夫を行うのも有効です。

まとめ

文章生成AI「ChatGPT」が登場して約7ヶ月が経ちます。様々な影響が懸念されていますが特に教育・研究における対応は喫緊の課題です。共存を目指した慎重な議論が行われています。

一方で様々な技術・サービスに応用できる可能性を秘めています。例えば医療面接の模擬患者として機能させる¹⁰⁾研究もスタートしています。またChatGPTを実現させたTransformerと言う技術はタンパク質の立体構造予測¹¹⁾にも一役買っています。

ChatGPT及びその周辺技術は医学分野においても要注目です。ChatGPTの仕組みや本学の指針に注意しつつ活用してください。

参考文献

No.	情報源	URL	QR
1	MICIN、金沢大学と実施していたChatGPTおよびGPT-4を用いて第117回医師国家試験(2023年2月実施)を解かせる研究において初めて合格点に到達し、その成果を論文としてオンライン公開いたしました 株式会社MICIN	https://micin.jp/news/10073	
2	GPT best practices - OpenAI API	https://platform.openai.com/docs/guides/gpt-best-practices	
3	自然言語処理モデル「GPT-3」の紹介 NTTデータ先端技術株式会社	https://www.intellilink.co.jp/column/ai/2021/031700.aspx	
4	東邦大学における生成系AIの利用に関する方針 お知らせ 東邦大学	https://www.toho-u.ac.jp/upnews/2023/20230623_1_AI.html	
5	Tohoku University Online Class Guide - ChatGPT等の生成系AI利用に関する留意事項 (学生向け)	https://olg.cds.tohoku.ac.jp/forstudents/ai-tools	
6	Tohoku University Online Class Guide - ChatGPT等の生成系AI利用に関する留意事項 (教員向け)	https://olg.cds.tohoku.ac.jp/forstaff/ai-tools	
7	ICMJE Recommendations Responsibilities in the Submission and Peer-Review Process	https://www.icmje.org/recommendations/browse/roles-and-responsibilities/responsibilities-in-the-submission-and-peer-review-process.html	
8	New AI classifier for indicating AI-written text	https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text	
9	Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools - ScienceDirect	https://doi.org/10.1016/j.xcrp.2023.101426	
10	zero to one、東北大学自然言語処理研究グループ、昭和大学と共同で、自然言語処理技術を活かした対話型AIチャットサービス「AI模擬患者」の開発に着手、医学教育におけるAIを用いたDXを推進 - 産経ニュース	https://www.sankei.com/pressrelease/prtimes/TMIAS33DVRNXZIBHV3HDQ720AA/	
11	AlphaFold2までのタンパク質立体構造予測の軌跡とこれから	https://www.jstage.jst.go.jp/article/jsbibr/3/2/3/jsbibr.2022.3/_html/-char/ja	