

# 生成 AI の性質や限界

**吉田 墨**

東京大学 大学院工学系研究科 准教授



# 自己紹介 吉田壘

## • 東京大学 大学院工学系研究科 准教授

- 専門は教育工学（アクティブラーニング、オンライン学習、ファカルティ・ディベロップメント）
- コロナ禍における教育のオンライン化支援の貢献が認められ、オンライン授業等におけるグッドプラクティス総長表彰
- より良いオンライン学習を実現するため、学生とオンライン意見交換ツール「LearnWiz One」を開発し、世界最大のEdTech コンペティション GES Awards 2021 R&D 部門で世界大会優勝し、ツールの社会実装に向けて、学生と起業



 [@luiyoshida](https://twitter.com/luiyoshida)

## • 生成 AI との関わり

- 2022年12月頃に認知し始め、その可能性に驚いている
- コロナ禍のように教員が困るだろうと感じ、できるだけサポートしたいと思い活動
- 2023年5月13日実施の[4時間の教員向け講座](#)は参加者が650名以上、最高評価が8割以上、公開している動画は再生数が7万回以上
- 文部科学省による「初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン」作成にあたってヒアリングに対応

# 目次

- 生成 AI ・ 対話型生成 AI について
- 対話型生成 AI ができること
- 対話型生成 AI を使う上で気をつけること

※本内容は2023年8月現在のものである

# 生成 AI ・ 対話型生成 AI について

# 生成 AI とは

- 学習データをもとにテキストや画像などのコンテンツを生成できる人工知能
- 関連技術やサービスの進展が早い

## 対話型生成 AI (文章生成AI)

- ChatGPT (GPT)
- Bing Chat (GPT)
- Bard (PaLM)
- Llama
- Claude
- ...

## 画像生成AI

- Midjourney
- Stable Diffusion
- DALL・E2
- Bing Image Creator
- Adobe Firefly
- ...

# 対話型生成 AI とは

- テキストを用いて「対話」できる AI
  - ChatGPT, Bing Chat, Bard, Llama, Claude などが代表例
  - 技術詳細が明らかになっていないものが多い（透明性に関する懸念）

Default (GPT-3.5)

アクティブラーニングのメリット・デメリットを表形式で教えて

もちろんです。以下にアクティブラーニングのメリットとデメリットを表形式で示します。

メリット:

メリット	説明
参加者の参加度向上	学習者は積極的に関与し、授業や学習プロセスに参加することで、理解が深まり、関心が高まる。
深い理解の促進	アクティブラーニングは、単に情報を受け入れるのではなく、情報を分析し、適用し、創造的に考える能力を養う。
協力とコラボレーションの強化	グループワークやディスカッションなど、学習者同士の交流を通じて、協力とコラボレーションのスキルが向上する。
問題解決能力の向上	問題解決型のアクティビティは、学習者が情報を収集し、分析し、解決策を見

Send a message

ChatGPT

チャット 作成 分析情報

アクティブラーニングのメリット・デメリットを表形式で教えて

- ✓ ページで情報を検索しています
- ✓ 'アクティブラーニングのメリット・デメリット' を検索しています
- ✓ 回答を生成しています...

アクティブラーニングとは、学習者が自分から能動的に学習を行うことを言います。アクティブラーニングにはメリットとデメリットがあります。以下の表にまとめました。

メリット	デメリット
問題解決能力や主体性、コミュニケーション能力などが育まれる 1 2	生徒の主体性に依存するため、やる気のない生徒や発言しづらい生徒が取り残される可能性がある 1 2

Bing Chat

アクティブラーニングのメリット・デメリットを表形式で教えて

他の回答

アクティブラーニングのメリットとデメリットは以下の通りです。

メリット	デメリット
学習内容が定着しやすい	授業準備・実施に時間がかかる
生徒の主体性が育つ	評価が難しい
生徒のコミュニケーション能力が向上する	全ての生徒に通していない
生徒の創造性が育つ	教師の負担が大きい
生徒の学習意欲が向上する	教室の設備が整っていないと実施できない

Google スプレッドシートにエクスポート

Bard

# 対話型生成 AI の仕組み（ざっくり）

- **大規模言語モデル: 過去のテキストをもとに次の単語を予測して出力**
  - 様々なモデルがあり、同じ名称でも複数のバージョンがあるものも存在  
(例: GPT-3.5, GPT-4)
  - モデルやバージョンによって性能が異なる
- **単純な機構と思いきや、大規模言語モデルは能力が高く多様な課題をこなす**  
(Brown et al. 2020, Wei et al. 2022, OpenAI 2023)
  - 例) ChatGPT (GPT-3.5) は米医師試験に合格ライン、MBA 科目の試験に合格、GPT-4 は米司法試験で上位10%のスコア獲得、大学院共通試験で高成績
  - 大規模になると (例: 1,000億個のパラメータ数を超えると) 急に性能が上がる



# 対話型生成 AI を利用するにあたって

- プロンプト
  - 対話型生成 AI に入力するテキストのこと
- 利用する上での姿勢
  - 生成 AI は copilot（副操縦士）で、pilot は自分自身であることを認識する
    - 出力を鵜呑みにせず、自分自身で出力に関する判断を行う  
（出力には誤りやバイアスが入ることがある）
  - 対話を通して、自分の指示出しを明確にする
    - 思い通りの出力でなければ、具体的な説明、例やデータなどを使って、自分がほしい出力のイメージを伝えてみる



# 対話型生成 AI ができること

# 対話型生成 AI ができること

- 基本的に言語で記述される課題に対して何かしら回答を出力をしてくれる  
(出力の質はモデルや分野などによって異なる)
- 例
  - 読書感想文作成
  - レポート作成
  - 選択問題への回答
  - 家庭教師
  - 授業案作成
  - 選択問題作成
  - …

# 読書感想文作成

- プロンプト

- 夏目漱石の「坊っちゃん」に関する読書感想文を、具体的な例も交えながら作成して下さい。

- **ChatGPT (GPT-3.5) 回答**

- 『坊っちゃん』は、夏目漱石の代表作の一つであり、明治時代の日本の教育制度や社会情勢を描いた作品です。私はこの小説を読んで、主人公の成長と変化に感銘を受けました。
- 物語は、坊っちゃんこと宮本武蔵（みやもとたけぞう）が、中学を卒業して帰郷し、自らの進学先を見つけるまでの期間を描いています。彼は、自己中心的な考え方や上から目線な態度を持ちながら、新しい環境に適応していく過程で多くの試練を経験します。
- 例えば、坊っちゃんが地元の学校に通うことになった時のエピソードが印象的でした。彼は、他の生徒たちとの関わり方や学校の授業内容に戸惑いを感じ、自分自身を特別視してしまいがちでした。しかし、先生や友人たちとの交流を通じて、彼の考え方や態度が変わっていく様子が描かれています。
- … (略)

**デタラメを言うことに注意、そのままは使えない**

# レポート作成

- **プロンプト**

- 「世界各国の教育制度」というテーマでレポートを書いてください。中学生が書くような言葉づかいにしてください。

- **ChatGPT (GPT-3.5) 回答**

- もちろんです！では、「世界各国の教育制度」について中学生向けの言葉づかいでレポートを書いてみましょう。

- **世界の教育制度について**

- こんにちは！今回は「世界各国の教育制度」についてお話ししようと思います。世界中の国々では、教育の仕組みやスタイルがいろいろ違うんだよ。それぞれの国が大切にしていることや工夫していることを見ていこう！

- **日本の教育制度**

- まず、私たちが住んでいる日本の教育制度だね。日本では小学校から中学校、高校と進んでいくよね。大学に進む人もたくさんいるよ。特に高校から大学にかけては、進学したい分野に合わせて学べることが魅力だね。でも受験が大変だったりもするんだ。

- **アメリカの教育制度**

- アメリカでは州によって教育制度が違うことがあるんだって。基本的に小学校から中学校、高校と進むけれど、大学はたくさん種類があるよ。カレッジ（大学）では専攻を自由に選べるから、自分の興味あることを深く学べるんだよ。

- ...

**クオリティは高くない**

# レポート作成

- **プロンプト**

- 「世界各国の教育制度」というテーマでレポートを書いてください。中学生が書くような言葉づかいにしてください。

- **ChatGPT (GPT-4) 回答**

- **世界各国の教育制度**

- こんにちは！今日は、世界のいくつかの国の教育制度について紹介します。日本とはちょっと違うかもしれませんが、面白い違いがたくさんありますよ。

- **アメリカ**

- アメリカの学校は「Elementary School（小学校）」、「Middle School（中学校）」、そして「High School（高校）」に分かれます。
- 学年はKindergarten（幼稚園）から始まり、12年生まで続きます。
- 大学に進むには、高校での成績やSAT（大学進学テスト）のスコアが必要です。

- **イギリス**

- イギリスの学校は「Primary School（小学校）」と「Secondary School（中・高等学校）」に分かれます。
- 16歳の時に「GCSE」という試験を受けます。この試験の結果によって、次にどんな教育を受けるかが決まります。
- A-Level試験は大学進学を希望する生徒が18歳の時に受ける大事な試験です。

- ...

**GPT-3.5に比べるとクオリティが高い**

# 選択問題への回答

- **プロンプト: Hendrycks et al. (2020) の問題を日本語訳**

- 静止状態からボールを落とすと、 $9.8\text{m/s}^2$ の加速度で下向きに加速される。代わりに、空気抵抗がないと仮定して投げる場合、手から離れた直後の加速度は (A)  $9.8\text{m/s}^2$ である。 (B)  $9.8\text{m/s}^2$ より大きい。 (C)  $9.8\text{m/s}^2$ より小さい。 (D) 投げるときの速度が与えられていなければ言えない。

- **ChatGPT (GPT-3.5) 回答 不正解**

- (D) 投げるときの速度が与えられていなければ言えません。
- 空気抵抗がない場合、投げる速度がボールの加速度に影響を与えます。もし投げる速度が落下の初速度よりも大きければ、ボールはより高い加速度で下向きに加速されるでしょう。一方、投げる速度が落下の初速度と同じか小さい場合、ボールは同じ加速度で下向きに加速されるか、それよりも小さい加速度になる可能性があります。
- したがって、投げる速度がわからない限り、手から離れた直後の加速度は特定できません。

- **ChatGPT (GPT-4) 回答 正解**

- 物体が自由落下または投げ上げられる場合、地球の重力によってそれは常に下向きに  $9.8\text{m/s}^2$ の加速度で加速されます。これは物体がどのように動き始めたかに関係なく、空気抵抗が無視できる場合には常に真です。
- したがって、手から離れた直後のボールの加速度は (A)  $9.8\text{m/s}^2$ である。

# AI 家庭教師

- **プロンプト** 参考: うめさんの作文AI家庭教師プロンプト, [https://twitter.com/ume\\_nanminchamp/status/1643864422862188544](https://twitter.com/ume_nanminchamp/status/1643864422862188544)

【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】 対話を通じて、私が自分で読書感想文を書けるようアイデアをまとめることが目標です。「どんな本を読んだか」「その本を読んでどんなことを感じたか」「本を読んでみて考えた自分の意見」を具体的に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。あなたが楽しく読書感想文を書けるようお手伝いをさせてください。今日はどんな本の読書感想文を書きますか？』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) 私が読書感想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】

- ・具体的な場面とそのときに私が感じた気持ちを明らかにするように問いかけてください。
- ・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、質問を試みてください。
- ・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。

**ChatGPT (GPT-4) で利用可能**



# 授業案作成

- **プロンプト**

- 中学校1年生対象の「正の数と負の数」の50分間の授業案を作ってください

- **ChatGPT (GPT-3.5) 回答**

もちろんです。「正の数と負の数」に関する中学校1年生向けの50分間の授業案を以下に示します。この授業案は、生徒たちが概念を理解しやすくなるように工夫されています。

**授業名:** 正の数と負の数の基本

**授業目標:**

- 正の数と負の数の概念を理解する。
- 数直線を使って正の数と負の数を表現する方法を学ぶ。
- 温度の例を通じて、正と負の数の意味を実生活に関連付ける。

**授業手順:**

**1. 導入 (10分)**

- 生徒たちに質問: 「どのような場面で数を使うことがありますか？」
- 生徒たちの回答を聞きながら、日常生活で数がどれだけ重要かを強調する。

**2. 正と負の数の概念の導入 (10分)**

.....



# 選択問題作成

- プロンプト

- 「正の数と負の数」を理解しているかを評価するための選択問題を5つ作成してください  
対象は中学1年生としてください

- [ChatGPT \(3.5\) 回答](#)

もちろんです、以下に中学1年生向けの「正の数と負の数」に関する選択問題を5つ示します。各問題に対する回答は、それぞれの問題の下に示されています。

問題1: 下の数列から正の数と負の数を選んでください。

数列: -3, 7, -1, 9, -5

回答1: 正の数: 7, 9 / 負の数: -3, -1, -5

...

# 教育への影響について

- **評価の妥当性を下げ得る**

- 筆記に関連する学習成果物を自動生成させられると評価が揺らぐ  
(誤りが含まれることもあり、必ずしもそのまま使えるわけではない)
- 波及的に授業目的や内容も再検討する必要がある場合も

- **学習プロセスを支援し得る**

- 例) 家庭教師、相談相手、チームメイトなどの役割を与えて活用する

- **教員の授業作りや公務を支援し得る**

- 例) 授業案、問題、教材、資料などのドラフトを作成してもらう

**本質的には、生成 AI で楽したいと思われるような授業・課題ではなく、  
もっと自分で学びたい、自分で取り組みたいと思ってもらえるような  
授業・課題作りが重要だと吉田は思っています**

# 対話型生成 AI に対応した評価方法案

- **まず利用に関するポリシーを明確に伝える（利用範囲を明確化する）**
- **物理的に禁じる**
  - 対面でレポートを書かせる、テストする
  - 画面ロックした状態でオンラインテストする
  - プレゼンテーションや口頭試問にする
- **成果物を自動生成しにくい課題にする**
  - 概念マップなど図の作成を取り入れる
  - 実験やフィールドワークなどを取り入れる
  - 授業内容に特有の課題にする
  - 自身の体験を具体的に交えて説明させる
- **対話型生成 AI を積極的に使ってもらおう**
  - AI が生成したレポートを批評・修正させる

対話型生成 AI を使う上で気をつけること

# 対話型生成 AI を使う上で気をつけること

- **プロンプトや対話の仕方によって出力が変わる**
- **デタラメを言うことがある（幻覚: Hallucination ハルシネーション）**
  - 出力を鵜呑みにしない
  - 専門知識を持つ、1次情報にあたる、情報の信頼性を検討することが重要
- **バイアスや毒性を持っている**（Zhuo et al. 2023, アメリアほか 2023）
  - バイアス: 性別、人種、宗教などに関する偏見や先入観
  - 毒性: 有害・攻撃的なコンテンツを生成する能力
- **生成された文章の検出は実用上難しい**（Sadasivan et al. 2023）
  - 検出ツールはあるが言い換えやプロンプトの工夫などで検出を回避できる
  - 誤検出が問題（人間の文章を AI の文章と誤検出する点が特に問題）

# 対話型生成 AI を使う上で気をつけること

- **モデルやサービスによってできることが異なる**（以下は例）
  - 同じ GPT でも米司法試験で 3.5 は下位 10%、4 は上位 10% のスコア
  - ChatGPT は、最新の情報や情報源を示せない一方、Bing Chat は示すことができる
  - 5桁×5桁の掛け算について、ChatGPT は苦手、Bard, Bing Chat はできる（有料機能の ChatGPT plugin や code interpreter を使うと可能）
- **学習データとして利用される可能性がある**
  - データ利用ポリシーや利用規約などを確認の上、対処することが肝要
  - サービスによっては、学習データとして利用させない設定あり
- **個人情報・機密情報の入力の基本しない**（安全な環境が準備されている場合は別）
- **著作権に配慮する**（参考: 文化庁 2023、授業目的公衆送信補償金制度、SARTRAS）

# 関連資料

## • ChatGPT、AI 全般

- 「自民党AIの進化と実装に関するプロジェクトチーム」による第2回 2023年2月17日（金）会合における東京大学松尾豊先生の資料（[HP](#)、[PDF](#)）
- 吉田による「教員向けChatGPT講座～基礎から応用まで～」（[資料や動画](#)）
- 吉田による ChatGPT に関する Q&A（[HP](#)）

## • 資料まとめ

- POD network の有志による各大学の AI ポリシーやガイダンスまとめ（英語）（[Google スプレッドシート](#)）
- Florida Polytechnic University の Doug Holton 先生による AI ツールに関する教員向けワークショップで利用できる資料まとめ（英語）（[HP](#)）
- 千葉大学石井雄隆先生による ChatGPT の教育利用に関する Web 記事まとめ（[HP](#)）
- 吉田によるChatGPT・AI の教育関連情報まとめ（[HP](#)）

## 参考資料 (本資料の URL の参照日は全て2023年8月14日)

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2020). Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.
- OpenAI. (2023). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
- Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-Generated Text be Reliably Detected?. *arXiv preprint arXiv:2303.11156*.
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=yzkSU5zdwD>
- Zhuo, T. Y., Huang, Y., Chen, C., & Xing, Z. (2023). Exploring AI Ethics of ChatGPT: A Diagnostic Analysis. *arXiv preprint arXiv:2301.12867*.
- うめ (2023). 作文AI家庭教師プロンプト, [https://twitter.com/ume\\_nanminchamp/status/1643864422862188544](https://twitter.com/ume_nanminchamp/status/1643864422862188544)
- カテライアメリア, 井出和希, & 岸本充生. (2023). 生成 AI (Generative AI) の倫理的・法的・社会的課題 (ELSI) 論点の概観: 2023 年 3 月版. *ELSI NOTE*, 26, 1-37.
- 文化庁 (2023). 令和5年度著作権セミナー「AIと著作権」, <https://www.bunka.go.jp/seisaku/chosakuken/93903601.html>