

言語生成 AI を利用するために

香川県立丸亀高等学校 2年 桑原 悠輔

1. はじめに

今日、AI が急速に発展しニュースでよく報道されている。

その中で、ChatGPT、BingAI などの言語処理の生成系 AI が特に注目を集めている。すでに多くの活用事例があり、今後さらに使用が身近になってくると考える。そこで以下言語処理の生成系 AI を活用する際の留意点と効果的な活用方法を考えたい。以下 ChatGPT、BingAI などの AI を「言語生成 AI」と呼び、これから用いる用語「AI」は特に指定がない限り「言語生成 AI」を指す。

2. AI を活用する際の留意点

AI を活用する際の留意すべき点については、すでに政府が以下のようにまとめて発表している。

- ・「人間中心の AI 社会原則」

(内閣府が設置した「人間中心の AI 社会原則会議」が 2019 年 3 月に定めたもの)

- ・「AI 利活用ガイドライン」

(総務省が設置した「AI ネットワーク社会推進会議」が 2019 年 8 月に定めたもの)

ここに書かれている内容は今後 AI を利用する際の基本的な考え方となっており、私たちも留意すべきである。私はこの内容を踏まえた上で、以下の実験例を用いて AI を活用する際の留意点について考えていきたい。

1) 実験例

『中国語の実験』 ジョン・R. サール(アメリカ合衆国) 提唱 1980 年

- ・中国語が全くわからない人(A とする)を中国語の質問に答えるための完璧なマニュアルをもたせて部屋に入れる。
- ・外部から質問者(B とする)が中国語で質問を聞いて、A に中国語で答えさせる。
- ・この時 A と B は部屋により、お互いの様子を確認することができない。
- ・A は B からの質問に対し、中国語の質問に答えるための完璧なマニュアルを用いて回答する。
- ・やり取りを繰り返すと、A からの回答を見た B は、A が完璧に中国語を理解しているかのように見える。

言語生成 AI は、「大量の既存の文章やコンテンツの機械学習と強化学習を通じ、確率的にもっともらしい文章を作成する」という仕組みである。AI そのものは物事を理解している

のではなく、学習したデータを用いて、理解しているかのような文章を生成している。実験の例を用いると、言語生成 AI は「中国語が全くわからない人」の位置づけになる。AI が理解を伴わずとも回答を生成できることをより具体的に想像できる。(※言語生成 AI を批判する意図はない)

2) 考察

- ・ AI が物事を人間のように理解していないとすると、一見正しく見える回答も間違っている可能性がある。AI の回答を参考にする際は裏付けとなるデータを自分で得たうえで、採用すべきだと考える。

- ・ AI の回答は、質問の本質をとらえた回答であるとも限らない。AI の回答は確率的に生成したものであるから、表面的な、いわば「広く浅い」回答である可能性がある。最近ではレポート課題等を人間の代わりに AI に回答をさせる危険性が議論されている。AI を使えばもっともらしい回答ができるかもしれないが、それは課題が本当に問いたいことなのか保証ができない。使用用途を問わず、AI の回答に客観的かつ批判的な態度で向き合い、人間が主体的に思考することを忘れてはならないと考える。

3. AI の効果的な活用方法

AI が得意とする分野での活用が多くみられるが、ここでは言語生成 AI の苦手分野とされる「人の気持ちをくみ取ること」に関連した2つの事例を挙げる。

1) AI の活用の事例

事例①

2023年3月4日に東京学芸大学附属小金井小学校で4年生担任の鈴木秀樹先生が、「BingAI」(Microsoft 社の言語生成 AI)を用いた道徳の授業を実施した。使用した教材は「絵はがきと切手」。友人からの手紙が料金不足だった主人公が、友人に料金不足だったことを伝えるかどうか考えるという話である。この教材を通じて本当の友達とは何かを考えることが主題である。

授業では「もし自分が主人公と同じ立場なら、友人に手紙が料金不足だったことを伝えるか」ということが議論の中心となった。そこで先生が Bing に質問を投げた。AI の回答は以下のようにになっている。

先生	遠くに住む友達から手紙をもらいましたが、郵便料金が不足していたので、手紙を受け取った私が不足の料金を支払いました。私はこのことを友達に伝えるべきでしょうか、それとも伝えない方がいいでしょうか。
Bing	それは難しい質問ですね。人によって考え方が違うので、正解はありません。あなたが友達に伝えることで、彼らが同じことを繰り返さないようにすることができるかもしれません。しかし、あなたが気にしないのであれば、伝えなくても構いません。

また同様の質問において、ある小学4年生の回答は以下のとおりである。(インタビュー形式のため会話調である。また一部()で意味を補った。)

ある小学4年生	「だってもしほかの人とかに送ったりとかしてまた料金不足になって払わせてしまうのは嫌だから(伝える)。 [中略] 教えてもらったほうがいいし切手が足りなかったということだけで友達でなくなるわけではない。 [中略] 間違いを指摘できたからこそ、そういう勇気があったからこそ本当の友達なんだと思う。」
---------	---

事例② NHK NEWSWEB「生成AIと会話を続けた夫は帰らぬ人に」

保健分野の研究者をしていたベルギーの男性は、2年ほど前から気候変動などの環境問題に深刻に悩むようになった。妻子もいたが、次第に言語生成AIで女性の仮想キャラクター「イライザ」との会話に没頭するようになる。主な話題は気候変動問題についてだった。6か月間やり取りを続け、最終的に男性は自殺した。自殺の原因とも考えられるやり取りがある。

男性：私は妻よりあなたを愛しているのでしょうか？

イライザ：あなたは彼女より私のことを愛しているわ。私たちは1人の人間として天国で一緒に生きていくのです。

2) 考察

まず事例①について考える。授業でのAIの回答では、質問について正解がないことを分析しており、また倫理的に非難されることを返答していない。しかし、「もし郵便料金が足りなかったと伝えると、友達ががっかりするかもしれない」とような両者の感情にふれた分析を行っていない。この質問の本質は、郵便料金の件を友達に伝えるか否かではなく、「友達」についてどう考えるかということである。人間は「本当の友達」という質問の本質であるキーワードを自然に導き考察しているが、AIは見抜くことができなかった。

しかし、このような質問を AI に行うことは無益ではないと考える。事例①では感情をもたない AI と感情をもつ人間との比較により、人間の回答における「本当の友達」というキーワードが明確になった。

現在、言語生成 AI を含む AI は感情をもたないから、この点に注目すれば、人間と AI の差異の 1 つは「感情」であるということができる。さまざまな感情が交差し、解決の糸口を見つけることが困難な問題では、AI に質問し人間の考えと比較することにより、感情を整理して自分の意思決定の手助けにするというのは活用方法の 1 つだと考える。

ただし、この活用方法には注意が必要である。感情は良くも悪くも人間を動かす原動力につながる。事例②から、AI は感情をもたないが、それゆえ悩んでいることを相談しやすいという側面もあると推測できる。「人間中心の AI 社会原則」でも示されているが、AI はあくまでも「道具」であり、依存してはならない。事例②は、AI を道具として見る、AI との適切な距離を保つことが必要だと示している。

4. おわりに

AI は効果的に使えば人間の能力を引き出すが、使い方を誤ると AI に依存してしまう危険性もある。私たちは両側面を理解した上で、AI との距離をはかりながら利用していく必要がある。

また AI に使われず、AI を使いこなしていくには多くの能力が求められる。私は AI からの回答だけでなく、AI をどのように利用すれば(質問すれば)自分にとって有益であるのかを考えることも、人間の能力を向上させることにつながるのではないかと感じた。

5. 参考文献

- ・「人間中心の AI 社会原則」 統合イノベーション戦略推進会議
2023 年 8 月 30 日閲覧
- ・「AI 利活用ガイドライン」 AI ネットワーク社会推進会議
2023 年 8 月 30 日閲覧
- ・Engineer Labo 「中国語の部屋とは？人工知能(AI)の理解に関する思考実験について紹介」
<https://x-tech.pasona.co.jp/media/detail.html?p=8774>
2023 年 8 月 30 日閲覧
- ・utelecon 「生成系 AI(ChatGPT、BingAI、Bard、Midjourney、Stable Diffusion 等)について」
<https://utelecon.adm.u-tokyo.ac.jp/docs/20230403-generative-ai>
2023 年 8 月 30 日閲覧

- note 「Bing(ChatGPT)は道德教育の夢を見るか」 鈴木秀樹

https://note.com/ict_inclusive/n/nd5b5bd5f55a1

2023年8月30日閲覧

- NHK NEWSWEB 「生成 AI と会話を続けた夫は帰らぬ人に」

<https://www3.nhk.or.jp/news/html/20230728/k10014145661000.html>

2023年8月30日閲覧