

2023.10 ブログ:「ChatGPT:14 の視点からその可能性を探る」を読んで(1)、の詳細
(→ <http://www.1968start.com/M/blog/index3.html#2310>)

「新時代の道具, ChatGPT:14 の視点からその可能性を探る」を読んで(1)

中所武司

■このエッセイのきっかけ

下記の学会誌の特集の、さまざまな分野の著者の14件のレポートを読んでみることにした。

・情報処理 Vol.64 No.9 (Sep. 2023)

「特集 新時代の道具, ChatGPT:14 の視点からその可能性を探る」

★今回は14件のうち、最初の5件を取り上げた。

■内容の要約とコメント(→★)

→★ここでの要約は興味本位に一部分を抽出したもの。

◆◆編集にあたって

- ・情報処理分野へのChatGPT のインパクトを確認するために、本特集では、できるだけさまざまな分野の著者に、ChatGPT がやってきた結果、現時点でどのような使い方をしているか、どんなことを考えているかを、自由にご報告いただいた。

◆1◆ 大規模言語モデルの驚異と脅威 — ChatGPTの衝撃と大規模言語モデルの課題 —

(著者:大学教授、情報理工学)

- ・本稿では、ChatGPT などの大規模言語モデルが知的な応答を返す仕組みとして、事前学習、指示チューニング、選好学習などのトピックに触れ、高性能な大規模言語モデルの構築や利活用において、その潜在的な悪影響を意識・軽減する視点が欠かせないことを説明する。
- ・2022年11月30日にOpenAIが公開したChatGPTの月間アクティブユーザは登場からわずか2カ月で1億人に到達し、社会に大きな衝撃を与えた。その後、ChatGPT Plus, GoogleのBard, Microsoft Bing, Meta AIのLLaMA, OpenAIのGPT-4などが矢継ぎ早に発表された。
- ・ChatGPTのような知的な処理を実現する大規模言語モデルの構築は、3段階に分けられる。
- ・第1段階は、大規模なテキストデータを用いた事前学習(pre-training)である。事前学習は、与えられた文章に続く単語を予測する次単語予測を学習させることが多い。「英国の首都はロンドン」という文章は、「英国」→「の」, 「英国の」→「首都」, 「英国の首都」→「は」, 「英国の首都は」→「ロンドン」という学習データに変換され、これに成功するように言語モデルのパラメータが更新される。GPT-3の事前学習では、約5,000億単語(トークン)のデータを費やした。

- ・第2段階は、翻訳や要約、質問応答、対話など、さまざまなタスクの能力を高めるために行うファインチューニングもしくは指示チューニングである。タスクの学習データを「指示と入出力」形式の自然言語に書き換えて、言語モデルをファインチューニングする。たとえば「英国の首都はロンドン」→「The capital of UK is London」という対訳事例を、「次の入力文を翻訳しなさい。入力：英国の首都はロンドン。出力：The capital of UK is London」という文章に変換し、この文章で次単語予測の学習を行うだけで、翻訳の指示の理解と翻訳の能力を獲得できる。
- ・第3段階では、応答に対する人間の選好を言語モデルに学習させる。第2段階までの学習での、応答が適切か否かの観点の不足を補うために、応答戦略を言語モデルにフィードバックして、間違った情報や有害な表現を抑制したり、利用者が好む応答を返すように調整する。この1つが、ChatGPTが採用している人間のフィードバックによる強化学習である。

→★数年前のチャットボットでは、人間のフィードバック機能で失敗していた。
誰が何を教えるかが重要になるが、その決定権は誰が持っているのか？

<過去の関連ブログ>

▼「2016.3 差別発言した人工知能の学習モデルは？」

<http://www.1968start.com/M/blog/old.html#1603c>

(要点) 人工知能を使ったマイクロソフトのチャットボットが、大量の差別的発言を学習し、同様の発言を返すようになり、公開から1日足らずで実験を中止した

▼2018.6 差別発言をしたチャットボットT a y（2年前）への追加コメント

<http://www.1968start.com/M/blog/index.html#1806b>

(要点) マイクロソフトは学習すべきでない内容を遮断するフィルターを実装していた

▼2017.8 共産党批判でサービス停止のチャットボットの再教育とは？

<http://www.1968start.com/M/blog/index.html#1708>

(要点) 中国 I T 企業が提供するチャットボットは、利用者と対話を繰り返して学習し、共産党批判で答えたことから急きょサービスを停止した

- ・大規模言語モデルの脅威として、人間社会への悪影響を体系的にまとめた分類を紹介する。
 - * 差別・排除・有害：憎悪，不敬，人格攻撃，侮辱，脅し，性的表現，攻撃的表現など
 - * 情報ハザード：個人情報記憶・推測することによるプライバシーの侵害など
 - * 誤情報による悪影響：誤解を招く情報の拡散，低品質な情報提示，問題行動の教唆など
 - * 悪用：スパムや詐欺メールの生成，サイバー攻撃に関するコード生成，違法な監視など
 - * インタラクションにおける悪影響：擬人化に伴う過度な依存，利用者の操作など
 - * 社会や環境への悪影響：環境負荷，自動化による失業，創造的な仕事の破壊など

◆2◆ LLMはインタフェースである

—人間とコンピュータ，人間どうしの対話にもたらされる可能性—

(著者：大学教授、先端メディアサイエンス学科)

- 本稿では、GUI より使いやすいインタフェースとしてのLLM の位置づけ、LLM を用いた味覚表現の探索、LLM を用いた研究指導の事例を紹介する。
- LLM技術をインタフェースとして使い、操作方法を覚えてもらうだけでなく、その操作自体をさせてしまうこともできる。
- 当研究室の味センサで測定した味覚を再現する味ディスプレイの研究開発の最新モデルであるTTTV3には、「LLMと対話して味を探求するモード」を追加した。
- LLMに「以下の論文概要を持つ論文についての不採択コメントを生成せよ」というプロンプトとともに論文概要を貼り付けると、生成されるコメントは結構リアルだが、本物の不採択コメントほどは傷つかないので、大学教員が学生指導を行う場合に役立つと思う。

→★この提案は意外性があり、興味深いが、それ以前に、学生がLLMを道具として適切に使いこなすようなレベルまで指導することが前提となりそう。自分で書いたレポートを自分でレビューできる能力を身につけさせるのは結構難しい。

◆3◆ AIはどのような仕事ができるようになったのか？

—ChatGPTで変わる「優秀な人材」—

(著者：企業コンサルティングとプログラミング教育サービスの会社を起業)

- ホワイトカラーのどのような仕事か、どのように機械学習によって奪われつつあるのか、また、LLM で書かれた「プログラム」を普通のプログラムで書き直す仕事について解説する。
- 従来の業務を横軸にミスの許容度、縦軸に入出力の複雑さをとって、図にすると、3つの領域に分かれる。
 - * 左下が、従来のプログラミングの領域で、間違っではいけない定形な仕事
 - * その上の土業が、責任を取る領域で、ミスには罰則があり、ミスの発生を防いでいる。
 - * 右側が、ホワイトカラーの仕事の領域で、ホワイトカラーの仕事は入出力が不定形で、作業フローも非定型で、多少のミスはあっても仕方ないという前提の仕事
- ChatGPTが登場し、自然言語のような不定形な入力から不定形な出力が行えるようになり、ホワイトカラーの多くの仕事が取って代われようとしている。現にChatGPTは、文章の要約、データ分析、メールの作成、企画書の作成、ブログ記事の作成、翻訳、校正、といった仕事を一瞬で行う。教育も、ChatGPTと対話しながら自分が分からないことを質問攻めにすることで、効率よく学習できる。

- ・プロンプトエンジニアリングによりさまざまな機能を簡単に実装できる。
これが次の時代の「プログラミング」となっていく。どのようなプロダクトが欲しいのかの要件定義や仕様書を書きだすと、動くものができあがってしまうことも起こり得る。

→★私の研究室で、30年前の1993年に開始したした、M-baseプロジェクトでは、
以下の基本コンセプトを設定したが、最近のAI技術でついに実現するという事かな。

＊Analysis ≡ design ≡ programming （分析，設計，プログラミングの一体化）

<http://www.1968start.com/M/rd/rd/M-base.html>

- ・生成的AIが普及した時代の新規事業の立ち上げ方は、次のような流れになっていく。
 - ＊確率的にあっていればよく、不定形な入出力が行われている個所で、自動化のニーズを探す
 - ＊プロンプトエンジニアリングによりプロトタイプを作る
 - ＊ プロダクトの可能性を見せ、顧客を集める
 - ＊ 顧客評価をもとに資金調達を行う
 - ＊ ChatGPTで作った部分を普通のプログラミング言語で書き直し、精度や信頼性を向上させる
- ・ChatGPTなどの生成AIの登場で「優秀な人材」の定義は以下のように変化しつつある：
 - ＊従来のIT産業：定形な業務を発見，定義すること
 - ＊機械学習：確率的にあっていれば良い業務を発見できること
 - ＊生成的 AI：不定形な入出力を行っている業務を発見し，素早くプロンプトで実装すること
 - ＊プログラムの書き直し：生成的AIで作られた信頼性の低い「プログラム」を
普通のプログラミング言語で書き直して安定化させること

→★ソフトウェア開発に関する考えが楽観的過ぎると思われる。

まず、現時点でも、要求仕様書の不備なケースが多いが、これをプロンプトとして
機能的に問題のないプログラムが生成できるとは思われない。

また、生成されたプログラムを書き直して、精度や信頼性を向上させる作業は、
最初からプログラムを作成するより、工数が多くなる可能性が大きい。

たとえば、標準パッケージの3割をカスタマイズするくらいなら、
全体を新規作成したほうがよい、という話もある。

◆4◆ ChatGPTでつくる自分だけのバーチャルアシスタント

（著者：民間企業のCTOで、画像配信，ログ解析基盤などを担当）

- ・ChatGPT の登場で、これまで難しかったキャラクタの正確な表現ができるようになった。
ChatGPT を活用することでどのような表現の可能性があるか，関連技術とともに紹介する。
- ・ChatGPTでは、対話型のインタフェースで入力した文章に対する応答を生成するが，
「私と会話してください。応答には最適な返答1つだけを含んでください」と入力し，
「userとあなたは幼少からずっと一緒にいる親友として振る舞ってください。敬語を
使わないでください」と追加すると友達とチャットしているような文章が生成される。

→★言葉づかいを選択できるということですね。思い出を共有していないので、
「幼少からずっと一緒にいる親友」になりすますことはできませんね。
また、地域性の強い方言を使うのは、かなりの学習を要するので、まだ無理では。

- ChatGPTは自然文の生成に長けているが、プロンプトで出力する形式を指定して、応答文中にメタ情報をタグとして応答に含めることができるので、「<happy>こんにちは！
<neutral>何についておはなししましょうか？」と、文章の中に感情タグを出力させれば、感情を持ったキャラクターの応答が可能になる。

→★音声出力の場合、言葉の意味を理解していないAIには「感情タグ」が必要かな。

- 言語モデルでは、強いバイアスを受けた応答など、意図しない応答がユーザの入力によって引き起こされるため、悪意ある入力には事前に拒否するといった方法が考えられる。

→★バーチャルアシスタントとしては失敗では？

最初の解説（◆1◆）の最後のコメント参照のこと。

- 現状の学習モデルは英語を日本語に翻訳したものが多く、口語調の文章は少ない。
自然の対話の実現には、日本語に特化し、さまざまな口調を実現できる学習モデルが必要。
表情豊かで、親しみのある会話ができることも重要ではないか。

→★バーチャルキャラクターの定義は？

口調や表情の実現だけでは、特定個人らしい価値観に添った発言は無理では？

特定個人の価値観を反映した応答の生成のためには、特定個人の特定の時期の発言を学習させる必要があるのでは。

◆5◆ ChatGPTでどこまでものぐさできるか

（著者：大学教授、主要研究テーマ：人工知能の哲学・倫理）

- 人となりは言動で判断してきたが、ChatGPT が用いられる世界ではその前提が崩される。
だが、社会活動を円滑にするための敵意のなさや無害さを表明するプロトコル部分を
ChatGPT が代行してくれると、人間は自分であるために表明する本質的な部分に注力できる。
- ChatGPTでは、言語行為という考え方から見れば、サイボーグ化されてしまったようなもの。
「自分が自分ではないような動きをしている？」と、使い方が分からず、ごちない。
「ほかの人から見て、私の行為とみなされるのは何か？」という問いが改めて立ち上がる。
社会における言動がその人がどのような人であることを規定するという前提を維持すれば、
サイボーグ化したときでもその人の行為とみなすのが本筋であろう。
- だが、確率的にありがちなことを出力しているだけでは、
決して自分だけのコミットメントを表明することにはなり得ない。

ものぐさであっても、自分が自分であるためには、そこだけには注力すべきだろう。

→★ちょっと論理展開がわかりにくかったが、要するに、次のようなことかな：

自分の意見をChatGPTに代行させても、社会的にはそれをその人の意見と認めてもよいが、
「自分が自分であるためには」「確率的にありがちなことを出力しているだけでは」なく、
「本質的な部分」については、自分だけの意見を言うことに注力すべきだ、と。

→★平たくいうと、ChatGPTを使いすぎて自分を見失うなよ、ということかな。

以上