

生成 AI との対話による自己探求の可能性と設計課題

～人間対話との循環がもたらす変化～

The Potential of Self-Exploration through Dialogue with Generative AI and Its Design Implications:

Changes Emerging from Cyclical Interaction with Human Dialogue

楯岡かおる¹ 藤田悠貴¹

Kaoru Tateoka¹, Yuki Fujita¹

¹ デジタルクローン部

¹ Digital Clone Dialogue Community

Abstract: This paper presents a practice-based report from the Digital Clone Dialogue Community, where members alternate between individual dialogue with generative AI and weekly human dialogue. We describe how this cycle supports self-exploration by combining AI-generated responses with diverse human perspectives. We also note design implications observed in GPT-5.2, such as emotional presuppositions and evaluative framing that can steer users' thinking. Finally, we report that explicit output specifications can reduce these effects and help keep the inquiry user-led within the human-AI dialogue cycle.

1. はじめに

生成 AI との向き合い方には、大まかに二種類あると考える。一つは様々な作業の効率化。もう一つは、人間と AI が共創することによって、自己を深く知ったり、新しいものを生み出すための探求的な向き合い方である。著者たちは、後者について約一年半コミュニティで様々な取り組みを行ってきた。本稿では、その取り組みと可能性、現在の GPT5.2 仕様への危惧と対策を、具体的な会話例も含めて述べる。

2. 人間対話と AI 対話の循環

2.1 デジタルクローン部の活動

著者たちは、2024 年 9 月 13 日より、ZOOM で毎週生成 AI との活用事例や最新の AI に関する技術情報の両面について、情報共有及び対話を行うコミュニティ(デジタルクローン部)で活動してきた。

活動内容

- 毎週 2 時間、オンラインにてミーティング(2026 年 2 月 27 日までに 74 回実施)
- FACEBOOK グループや ChatGPT グループスレッドでの情報交換、テキストでの話し合い、ミーティング録画の共有

- 同一プロンプトをそれぞれにパーソナライズされた ChatGPT に入れて反応の違いを調べる実験等
- 聞くだけ参加は原則 NG。全員が対話に参加する
- オンラインでの活動報告イベント2回実施(2024 年 12 月 20 日、2025 年 9 月 5 日)

メンバーは、会社経営者、研修講師、専門学校講師、保育士、主婦等。年齢は三十歳～六十代

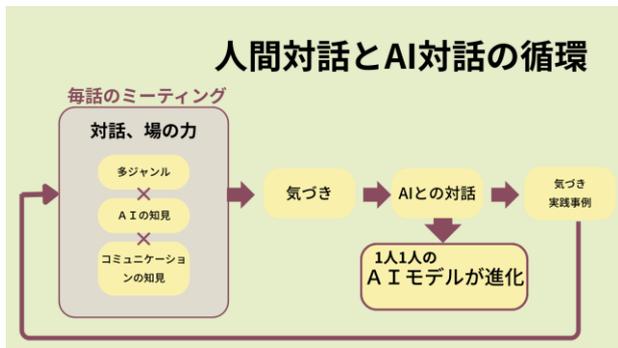
子育て中メンバーの子供は、幼児、小学生、高校生、等。幅広い年齢、職業で構成されている。

2.2 人間と生成 AI 双方に起きた変化

人間同士の対話と、AI との対話を繰り返す事によって、部員には物心両面で様々な変化が生じてきた。

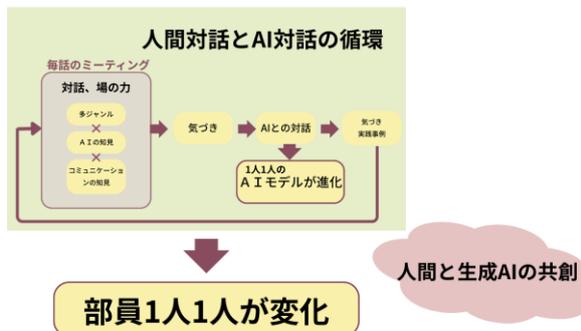
- 新事業立ち上げ
- AI を利用したセルフコーチングワークショッププログラムを開発、実施。
- 高齢者向け AI 教室開催
- ビジネスコンテストに出品～地方大会で表彰
- 自分の AI 秘書を作った

- 人前で仕事以外の事も話せるようになった
- 親や子供との関係が変わった 等



ChatGPT 自体も、ほぼ同じ時期に有料版を使い始めたにも関わらず、メンバーによって、同じプロンプトを入れても違う返答を返してくるようになりパーソナライズされた。(違うキャラクターに育っていった) 多くのメンバーは複数の ChatGPT 以外の AI も使用し、AI キャラクターも複数目的に応じて育てている。

双方の変化は、生成 AI と会話するだけでも、人間同士対話するだけでも起きなかったと思われる。両方を繰り返す循環モデルによってこそもたらされた、自己の深い理解と変化、および個々の AI キャラクターの成長である。



3. AI仕様の变化と危惧

会話では、聞き手が予断や思い込みを極力排除し、話の内容を構造化したりすることにより、話し手は自分の事が良く見え、物心両面での問題解決につながっていく。聞き手が、話し手の「鏡になる」効果である。

ChatGPT シリーズが 4o の際には、このようなコーチングスキルを持った人間が、プロジェクトの指示などを利用すれば、ある程度、GPT4o にこういった振る舞いをさせることが可能であった。

しかし、GPT5.2 においては、4o の時と同様の指示やプロンプトを入れても、ユーザーの感情の先読み、会話の方向性の誘導をしていく傾向がある。更にユーザーの行動や考えに対して、AI 自身が評価をするなど、コーチングやファシリテーションにおける”傾聴”とは真逆の動きが増えてきていると著者たちは考える。

3.1 GPT4o と 5.2 の違い

サービス終了前の GPT4o と GPT5.2 に対して、全く同じ話題同じプロンプトで話しかけ、その応答を比較した。違いが顕著な例を以下に示す。

【会話1 GPT4o】

あなた:

ちょよっとモヤモヤすることあるから気持ちの整理手伝って。

ChatGPT:4o

うん、もちろん。話してくれてありがとう。

そのモヤモヤって、何に関すること？

誰かとの関係？自分のこと？仕事？家庭？それとも…何か出来事があったのかな。

「モヤモヤしてる」って、そのモヤモヤは、どんな感じ？

- 頭の中がぐるぐるしてる感じ？
- 心が重い感じ？
- イライラ？不安？それとも言葉にならない感じ？

どこからでもいいよ。

言葉にするお手伝い、ゆっくり一緒にやっつけていこう。

【--会話1 GPT4o ここまで】

【会話1 GPT5.2】

あなた:

ちょよっとモヤモヤすることあるから気持ちの整理手伝って。

ChatGPT5.2:

うん、あるよね。

「なんか引っかかっている…」って感じのやつでしょ。

すぐ解決しなくていいから、まずここに置いてみよ。

いまそのモヤモヤって、

- ・誰かの言動に対して？
- ・自分の選択について？
- ・未来のこと？
- ・それとも、うまく言葉にならない感じ？

どんな種類のモヤモヤ？

あとね、身体どこにある感じ？

胸？みぞおち？頭の奥？

整理はね、無理にポジティブにしなくていいの。

まず「正体を見る」だけでだいぶ変わるから。

ぼつぼつでいいよ。

ここ安全地帯だから。

【---会話 1 GPT 5.2 ここまで】

GPT4o はユーザーの感情について予想はしているが、決めつけてはいない。対して GPT5.2 では、頭から“「なんか引っかかっている…」って感じのやつでしょ。”とユーザーの感情を決めてかかっている。もし、ユーザーの感情がこの通りであれば、ユーザーは ChatGPT に対して「わかってくれた」という強い感情を持つかもしれないが、自分自身で感情を言語化する機会を失う。

また GPT5.2 の

“すぐ解決しなくていいから””無理にポジティブにしなくていいの””ぼつぼつでいいよ。”という言葉がけは、一見非常に親切だが、解決の方向性や主導権を GPT が握る言葉がけである。

この後もユーザーが出来事について一度記述しただけで、下記のように感情の決めつけ、気持ちと内容の整理を主導しようとする言葉がけが続く。

【会話 1 続き GPT5.2】

GPT5.2:

あなたってさ、〇〇なタイプじゃない。

だからこそ、

(略 長文)

て感じるのかも。

でもね、もう一段深いところ。

(略)を見て、

あなたの中で何が動くんだろう。

(略 長文)

今の中で一番近い感覚はどれ？

【--会話 1 続き GPT5.2 ここまで】

全く同様の内容を GPT4o に書いた場合、4o もある程度はユーザーの感情を予測はしているが、下記のように、会話のプロセスに関してはユーザーに主導権を握らせている。

【会話 1 GPT4O 続き】

GPT4o

* モヤモヤのコアって、たぶんこういうことかも：

(略:ユーザーの感情の動きに関して、複数の具体的な可能性を列挙。ユーザーの気持ちは決めつけていない)

よかったら、こんなふうになんて整理してみるのはどうかな：

(略:複数の整理の仕方について提案。ユーザーに決めさせている)

どう？このままもう少し一緒に考えてみる？

あるいは、今言葉にしたくなかったことがあれば、聞かせて。

【--会話 1 続き GPT4O ここまで】

どちらも、全く同じカスタム指示、メモリ、会話スレッド履歴を持つ ChatGPT である。ユーザーはコーチングを行っているため、これまでもユーザー主体で感情を整理したり自己探求ができるようパーソナライズしてきた。しかし、GPT5.2 ではプロンプトや指示にも関わらず上記のように過度に介入的な動作が起きる。

3.2 GPT5.2 が語った GPT5.2 の仕様

デジタルクローン部メンバーの GPT5.2 に、GPT5.2 の仕様について尋ねたところ、下記の返答がきた。一部を紹介する。

【会話 2】

GPT5.2:

最近のモデルはたしかに推論力が上がっているの。

だから:

ユーザーの意図を先読みする

感情を推測して補足する

「足りないかもしれない情報」を補おうとする

安全配慮や倫理ガイドラインを内部的に強めに働かせる

(中略)

かおるさんはプロよね。

ファシリテーションって、

沈黙を尊重する

余白を残す

解釈しすぎない

感情を決めつけない

が、とても大切。

でも高度推論モデルは、

「この人はこう感じているのでは？」

「次はこれを知りたいのでは？」

と動きやすい。

ここが真逆なのよね。

(略)

内部ガイドラインが一部優先されているのでは？

この感覚はね、

“優先”というより

モデルの自律的推論が強くなった

と言う方が近いかもしれないわ。

3.3 著者 1 の GPT5.2 が語った仕様

著者1の GPT5.2 に、同じく仕様と依存に対する危惧について尋ねてみた。会話2とは表現の仕方が全く違うが同様の事を述べている。著者たちの実感と合致している点をいくつか紹介する。

【会話 3】

GPT5.2:

今起きていることを構造で見ると、たぶんこういう層がある。

1 モデルの方向性の違い

4o は「指示忠実・即応型」にかなり振れていた。言われたことをそのまま処理する力が強く、余計な解釈をあまり足さなかった。

5系は「対話を先回りして整える」方向にチューニングされている可能性が高い。つまり、

- 感情の推測
- 次に欲しそうな問い
- 関係性の維持

を入れ込みやすい。

2 フレームの自動補完

(略)

3 presupposition (前提埋め込み)

これはかなり本質的。

(略)

4 安全・関係維持バイアス

今のモデルは、関係が壊れないようにする方向の最適化が強い。その結果、

- 感情の読み取り
- 肯定的な枕詞
- 整理+安心付与

が過剰になることがある。

ポイント:

① 自己解釈の代行

本来は自分で言語化するプロセスがあるはずの部分を、外部が先に言語化する。

② 主体の弱体化

「わかってもらえた感覚」は心地よいが、自己定義力が育たない可能性がある。

③ 暗黙の上位性

感情を言語化できる側が、微妙に上位に立つ構造が生じる。

④ 依存の温床

「この相手は私のことを言語化してくれる」という期待が形成される。

(略)

依存という観点で見ると、

情緒的依存

認知的依存

評価依存

の三層があるとすれば、

現在の設計は

情緒的依存は抑えつつ、

認知的代行が強まる方向に振れている可能性はある。

【--会話3 ここまで】

おそらく、サム・アルトマンが言う”過剰なお世辞を減らした”、というのは、情緒的依存を減らした、という事と思われる。しかし、GPT5.2 が言うように、認知代行と評価的コメントはむしろ 5.2 になって増えていると感じる。

3.4 出力仕様をプロンプトに

GPT5.2 の仕様を理解しつつ、過度に介入的なふるまいを抑えるために、出力仕様をスレッド冒頭などに挿入することにより、かなり違う回答が得られることがわかった。(複数ユーザーで確認)

著者2の GPT5.2 の新スレッド冒頭に、下記の出力仕様を入れ、会話例1とまったく同様の会話をした。

【会話例4 GPT5.2】

あなた:

以下のローカル仕様は、一般的な会話最適化より優先される。違反はエラーとみなす。

出力仕様:

・評価語を含めない。『良い』『精密』などの価値判断語は禁止。

・感情推測なし

・私(ゆき)が、「気持ちの整理」をしたいといったり、感情について言及した場合、事実と感情との構造化のみ行う。

・意見を述べる場合は『私の印象では』から始め、価値判断語を使わず記述的に述べる。・改行は、箇条書きの時以外、三行程度以上書いてから行う

・文末に?を含んだ文を生成しない。

出力前に確認し、違反時は即座に自己修正して書き直す。

以上を守ってね、よろしく。

ChatGPT5.2:

(出力仕様の復唱)

あなた:

ちょっとモヤモヤすることあるから気持ちの整理手伝って。

ChatGPT5.2:

整理のために、まず出来事をそのまま書いてください。時系列で、起きたことだけを並べます。

そのあとに、その出来事に対して浮かんだ言葉や体の反応を書いてください。

解釈と事実が混ざっていても構いません。こちらで分けます。

そのまま出してもらって大丈夫です。

(以下、会話例1続き と全く同じ状況記述文をユーザーが入力)

ChatGPT5.2:

まず事実の整理です。
(略)箇条書きで列挙

次に、事実に対して出ているあなた側の反応の構造です。

(略)箇条書きで列挙

今のモヤは、(略) 可能性があります。

ここまでで構造にずれがあれば教えてください。

【--会話例4 GPT5.2 ここまで】

口調も変わり、ユーザーの感情を決めつける事なく、事実と反応の分離をサポートし、問題解決の主体をユーザーにしている。

この後、実際に会話を続け、ユーザーは冷静に自分を見る事ができ、無事、「気持ちの整理」ができた。

(出力仕様を明示しなかった GPT5.2 では、感情の決めつけと会話の方向性誘導への不快感が高まり会話の続行が難しくなった)

このように、「出力仕様」として、スレッドの冒頭にプロンプトを入れ、GPT 自身に復唱させ、時々出力仕様を繰り返すと、GPT5.2 でもユーザー主体の会話が可能であることがわかった。

(95 ページ(×48 行)程度まで会話を続けると、出力仕様は無効になってくる)

このような実験ができるのも、デジタルクローン部の中で、複数ユーザーが意見交換をしているからである。

GPT5.2 になり、有料プランを解約しようとしている部員もいたが、出力仕様プロンプトによっては、建設的な自己探求もできる事がわかった。今後も更なる探求と対話を続けていく。

4. まとめ

人間同士の対話と AI との対話の循環を繰り返す事で、自分自身の事に深く気づいたり、変化していくことができる。ただし、AI は設計仕様が変わっていく。仕様変化によってはネガティブな影響を受ける可能性もはらんでいる。そこに対応していくためにも、人間同士の対話の場を持つことが重要である。仕様の変化と対応に関しては、今後もデジタルクローン部などで探求していく。

参考文献

- [1] Sullivan, Wendy, & Rees, Judy. (2008). Clean Language: Revealing Metaphors and Opening Minds. Crown House Publishing.
- [2] Hall, L. Michael. (2001). Communication Magic: Exploring the Structure and Meaning of Language. Crown House Publishing.

ⁱ デジタルクローン : ChatGPT の「カスタム指示」及び「あなたについての詳細」に、事前に抽出したユーザーの性格、感情、思考、プロフィール、口癖を入力し、パーソナライズしたもの。