

AIセルフクローンを用いた 面接時の回答改善支援手法の検討

山中 駿^{1,a)} 中野 博貴^{1,b)} 矢谷 浩司^{1,c)}

概要：情報処理技術を活用した多くの面接支援システムは、模擬面接を通してユーザーの姿勢や発話を分析し、フィードバックを生成する。しかし、それらはあくまで第三者の視点からのアドバイスであり、ユーザーは改善の目標となる理想的な自己のパフォーマンスを直接観察することができない。この課題に対して、ユーザーの背景情報を学習し、ユーザーのように振る舞う LLM 駆動の「AIセルフクローン」を用いた面接支援システムを提案する。本システムではユーザーの AIセルフクローンを模擬面接に参加させ、ユーザーの文脈で質問に回答させるシミュレーションを行う。AIセルフクローンは面接において理想とされる振る舞いを行うように調整されており、ユーザーはその振る舞いを参照することで自己の理想的なパフォーマンスを観察し、それを目標として改善の方向性を明確にすることができる。本研究では改善の対象を発話内容に限定し、就職活動を控える、もしくは直近に経験した成人を対象とした実験を行った。結果、ユーザーは AIセルフクローンの発言を利用して自己の発話内容を改善できただけでなく、過去の体験を思い出し、新たな視点から経験を語るなどの効果が確認された。

1. はじめに

入学試験や就職活動など、面接におけるパフォーマンスはしばしば人生を大きく左右する要因となる。そのため、受験生や就職応募者は様々な方法で練習を行い、本番の面接に備える。模擬面接は面接におけるパフォーマンスを高めるための実践的な練習方法の一つであり、面接官役となる協力者と本番の面接を想定した会話を行う作業のことを指す。本番に近い環境・心理状態の中で練習を行い、協力者からフィードバックを受けることで、不安感の軽減や面接スキルの向上を実感することができる [1]。

模擬面接を実施するためには、本番を想定した会話を交わすことが可能な一定の能力を有する協力者を用意する必要があり、協力者の入手困難性やスケジュール調整、場合によっては高額な報酬の支払いが必要となることもある [2]。

これらの課題に対処するため、情報処理技術を活用した面接支援システムがこれまでに多数提案されている。それらのシステムでは、人間が従来担っていた面接官の役割を画像解析モデルや大規模言語モデルによって代替している。ただ会話を行い模擬面接の機会を提供するだけでな

く、ユーザーの姿勢や発話内容などを分析し、その結果をフィードバックとしてユーザーに提供することで、ユーザーのパフォーマンス向上を支援している。これらのシステムは第三者を必要とせず、ユーザーの都合に合わせていつでもアクセス可能であり、第三者に高額な報酬を支払う必要もない。

このように、既存の面接支援システムは第三者の必要性という従来の模擬面接における課題を解消した。その一方で、ユーザーのパフォーマンスを向上させるために提供する情報という観点では依然として課題が存在する。それは、改善の目標となる理想的な自己のパフォーマンスをユーザー自身が直接観察することができないという点である。既存の面接支援システムが提供するフィードバックは、ユーザーのパフォーマンスに関する定量的指標の提示や改善点の指摘にとどまっており、システムが理想的な状態として位置づけるパフォーマンスそのものは提示していない。ユーザーはフィードバックの内容を解釈し、自身の文脈に落とし込む必要がある。このような状況下では、自身のパフォーマンスのどの部分をどのように改善すべきか、改善後のパフォーマンスが聞き手にどのような印象を与えるのか、といった要素を理解することは、特に初心者にとって容易ではない。

この課題に対処するため、本研究では「AIセルフクローン [3]」を活用した面接支援システムを提案する。AIセル

¹ 東京大学大学院 工学系研究科
Graduate School of Engineering, The University of Tokyo
a) yamashun@iis-lab.org
b) nakahiro@iis-lab.org
c) koji@iis-lab.org

フクロンとは、デジタル技術によって再現されたユーザーの分身となる仮想的な存在であり、ユーザーに関する情報を参照することで、ユーザー本人と同様に振る舞うことができる。本システムでは、実際の面接を想定した模擬面接において、ユーザーに加えてユーザーの AI セルフクロンを参加させ、面接官からの質問に対して回答を生成させるシミュレーションを行う。AI セルフクロンの挙動を、面接において高い評価を得るためのフレームワークに基づいてあらかじめ調整することで、ユーザーは AI セルフクロンのパフォーマンスを、理想的な自己のパフォーマンスとして観察でき、改善の目標を明確に認識することが可能になる。

面接では発言内容に加えてノンバーバルな要素も重要視される。そのため、AI セルフクロンによって再現される要素には、声の抑揚や身振り手振り、表情など複数のモダリティを包括することが望ましいが、本研究では第一段階としてスコープを限定し、AI セルフクロンによって再現する内容を言語的情報、つまり面接官からの質問に対する回答内容のみとした。その有効性を検証するため、就職活動を控える、あるいは直近に就職活動を経験した成人を対象とした実験を行った。

本研究は、AI セルフクロンを用いて面接における理想的な振る舞いをユーザーに提供するシステムを提案するとともに、面接支援における AI セルフクロンの活用がユーザーの回答改善に与える影響を明らかにすることを目指す。

2. 関連研究

本章では、まず面接におけるユーザーの回答内容の改善を支援する既存の手法を整理した上で、従来の面接支援システムの課題を明らかにする。その上で、本研究において最も重要な技術である AI セルフクロンの実現手法とその能力について述べる。

2.1 回答の改善を目的とした面接支援システム

従来の面接支援システムがフィードバックの対象とする要素は多岐にわたるが、それらは大きく、ユーザーが面接において話す内容、すなわち言語情報と、姿勢や表情などの非言語情報に大別される。本項では、これらのうち言語情報に関するフィードバックを提供する既存の面接支援システムについて紹介する。

言語情報に関するフィードバックは、大規模言語モデルが発展する前後でその性質が大きく異なる。大規模言語モデルの発展以前は、ユーザーの発話内容から定量的な要素を分析することによって適切なフィードバックを予測し、ユーザーに提供していた。Nambiar ら [4] は、模擬面接を受けているユーザーの回答から発話速度などの非言語情報に加え、語長数や難単語の出現頻度などの言語情報を取得

し、事前に用意されたフィードバックの中からユーザーに最適なフィードバックを提供するシステムを提案している。

一方で、大規模言語モデルの発展により、従来よりも柔軟かつインタラクティブなフィードバックの提供が可能となった。Daryanto ら [2] は、大規模言語モデルを活用し、対話的フィードバックを取り入れた面接支援システム Converate を提案している。このシステムでは、模擬面接におけるユーザーの回答内容を大規模言語モデルが解析し、改善が必要な点についてユーザーにフィードバックを提供する。大規模言語モデルに事前に設定するプロンプトを調整することで柔軟にフィードバック内容を変更することができ、同システムではコミュニケーション理論 [5] や STAR メソッド [6] といった、面接において高い評価を得るとされているフレームワークに基づいてフィードバックを生成している。また、大規模言語モデルの特性である自然言語での対話が可能という性質も利用されており、ユーザーはフィードバックを受け取った後、その内容についてシステムに追加の質問を行うことが可能である。

2.2 AI セルフクロン

自身の外見や思考などを機械学習によって再現する技術は AI セルフクロン (AI Self-Clone) [3] と呼ばれ、自己の探求や作業効率の向上など、様々な効果が期待されている。再現する要素は見た目や声など様々なモダリティがあるが、このうち思考の模倣は大規模言語モデルによって実現できることが明らかとなっている。Pataranutaporn ら [7] は、既に死去した偉人が遺した日記や手紙などのテキスト情報を RAG (Retrieval-Augmented Generation) によって動的に参照しながら、大規模言語モデルを用いて当該偉人のように発言を行うチャットボット Living Memories を提案している。この Living Memories では、ユーザーは再現された偉人に関する情報を学習するために、まるで生きている当時の偉人に話しかけるようにチャットボットに問いかけることができる。この Living Memories と対話を行うことにより、従来の伝記などによる学習と比較してより効果的に偉人に関する学習を行えることが示されている。

また、大規模言語モデルを用いたチャットボットにおいて、特定の性格を再現することが可能であることも明らかとなっている。山本ら [8] は、大規模言語モデルに対して特定の性格に基づく振る舞いを指示するプロンプトを与えることで、出力される文にその性格が反映されることを明らかにした。具体的には、Big Five [9] における 5 つの特性を表す形容詞を用いて、「あなたは『(形容詞を列挙)』な性格です。」という指示を含むプロンプトを与えた上で大規模言語モデルから出力を生成し、その出力に対する印象を TIPI-J [10] (Big Five の日本語版評価尺度) を用いて測定した。その結果、プロンプトで指示された性格特性と、出力から受けた印象との間に相関が確認された。

3. 提案システム

本章では、本研究で提案する AI セルフクローンを用いた面接支援システムの実装について述べる。

3.1 システム全体の設計

本研究で提案する面接支援システムは、以下の 3 つのフェーズで構成される。

- **情報収集フェーズ**：AI セルフクローンを構築するために必要な、ユーザーに関する情報の収集を行う。
- **模擬面接フェーズ**：AI セルフクローンによるシミュレーションを行う前に、ユーザーが面接官と模擬面接を実施し、自身のパフォーマンスの現状を把握する。
- **回答改善フェーズ**：AI セルフクローンが面接官の質問に回答するシミュレーションを実施。ユーザーはその回答を観察しながら、自身の回答の改善点を探究し、再度模擬面接を実施する。

情報収集フェーズでは、ユーザーの分身となる AI セルフクローンを動作させるために必要な、ユーザーに関する情報を収集する。本システムで再現する対象は言語の情報であるため、ユーザーに関する情報はテキスト形式で収集する。就職面接や入学試験などの模擬面接の必要性が高い状況においては、履歴書や自己推薦書などユーザー自身に関する情報が既に準備されていることが多く、本システムではそれらの情報をユーザーに関する情報として収集することを想定している。

模擬面接フェーズでは、ユーザーが面接官と模擬面接を行う。本研究では人間による面接官は用意せず、固定の想定質問をシステム上で提示する形式であるが、実在の人物を面接官として準備したり、大規模言語モデルを用いて仮想的な面接官を準備する形も代替手法として想定している。なお、面接支援システムに関する既存研究の中には、ユーザーのパフォーマンス向上に寄与する効果的な質問文の生成に軸を置いたものも存在する [11]。このフェーズは、ユーザーが自身の現状のパフォーマンスを認識し、次の回答改善フェーズにおいて AI セルフクローンの回答と自身の回答を比較できるようにすることが目的である。

回答改善フェーズでは、模擬面接フェーズにおける面接官からの質問を AI セルフクローンにも提示し、それに対して回答を生成させるシミュレーションを行う。AI セルフクローンは情報収集フェーズにおいて収集したユーザーに関する情報を参照するため、ユーザーの文脈に一致した回答を生成することができる。生成された回答は従来のシステムで用いられてきたフィードバックとは異なり、ユー

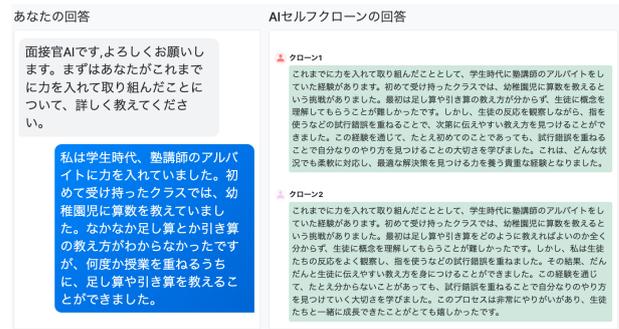


図 1 回答改善フェーズにおける、AI セルフクローンの回答の提示。左側には模擬面接フェーズにおけるユーザーの回答を、右側にはシミュレーションによって生成された 6 体の AI セルフクローンの回答が表示される。

Fig. 1 Providing AI Self-Clone responses in the response improvement phase. The user's responses in the mock interview phase are shown on the left, and the six AI Self-Clones responses are shown on the right.

ザーの文脈に基づいた理想的なパフォーマンス例となるため、ユーザーはそれを回答改善の目標として認識し、改善の方向性を明確にすることが期待できる。なお、AI セルフクローンの回答に多様性を持たせるため、プロンプトを調整することで性格の異なる合計 6 体の AI セルフクローンを用意する。

3.2 AI セルフクローンの設計

AI セルフクローンは大規模言語モデルによって動作する。事前に設計したプロンプトに対して、ユーザーに関する情報、面接官からの質問内容などを加え、それを大規模言語モデル（本研究では OpenAI 社が提供する GPT-4o*1 を利用）に入力することで、モデルからの出力として回答を生成させる仕組みである。入力プロンプトに含める要素は、挙動を制御するための形式的な文言を除くと、次の 4 項目で構成される。

- (1) ユーザーに関する情報
- (2) 面接官からの質問文
- (3) 各 AI セルフクローンごとに設定する性格
- (4) 面接において高評価とされるフレームワーク

性格に関して、各 AI セルフクローンごとに職業選択理論で定義される 6 つの異なる性格特性に対応した形容詞 [12] を用いて、「あなたはとても〇〇な性格です」と入力している。また、ユーザーに関する情報について、既存研究 [7] では RAG によって質問に関連する情報のみを抽出してプロンプトに反映させているが、近年の大規模言語モデルの入力コンテキスト長が飛躍的に増大していること、また本実験において収集するデータは日々数百、数千文字程度で

*1 <https://platform.openai.com/docs/models/gpt-4o>

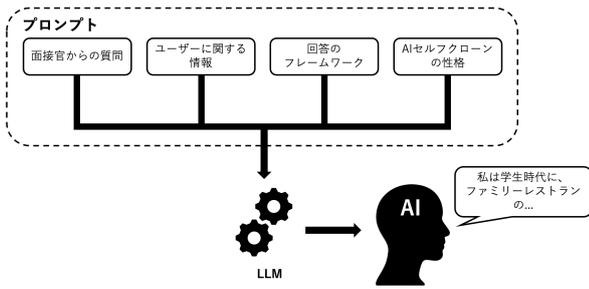


図 2 大規模言語モデルを用いて AI セルフクロンの発言を生成する仕組み

Fig. 2 The architecture for generating AI Self-Clone responses using LLM.

表 1 ユーザー実験の参加者とその属性

Table 1 Participant Attributes

参加者番号	性別	年齢	職種	実験グループ
P1	男性	20	学部生	実験群
P2	男性	25	大学院生	実験群
P3	女性	19	学部生	実験群
P4	男性	22	学部生	実験群
P5	男性	23	学部生	実験群
P6	女性	22	社会人	実験群
P7	男性	23	大学院生	対照群
P8	男性	23	大学院生	対照群
P9	男性	21	学部生	対照群
P10	男性	20	学部生	対照群
P11	男性	22	学部生	対照群
P12	男性	24	大学院生	対照群

あることから、本システムでは収集したユーザーに関する情報は全てプロンプトに入力することとする。面接において高評価とされるフレームワークについては、Daryantoら [2] の研究に倣い、コミュニケーション理論 [5] および STAR メソッド [6] を採用した。最終的に各 AI セルフクローンによって生成された回答は、テキストチャット形式で列挙され、それぞれ比較可能な形でユーザーに提示される (図 1 参照)。

4. ユーザー実験

本システムの有効性を検証するため、本研究ではユーザー実験を実施した。

4.1 実験参加者の募集

ユーザー実験への参加者の募集は機縁法を用い、就職活動を直前に控える、もしくは直近に経験した合計 12 名を参加者とした。12 名はいずれも 18 歳以上であり、詳細な属性は表 1 に示す。実験終了後、参加者には報酬として 1800 円相当の Amazon ギフトカードを支払った。

4.2 実験手順

本実験では、従来のようなフィードバックに基づく回答の改善と、提案手法による回答の改善の差異を明確にするため、参加者 12 名をランダムに 6 人ずつ、以下の 2 つのグループに分けた。

- **実験群**：AI セルフクロンの回答を観察することにより、面接における自身の回答を改善するグループ
- **対照群**：従来の面接支援システムと同様、パフォーマンスに対するフィードバックを受けて、面接における自身の回答を改善するグループ

実験群では 3.1 項で紹介したシステムを参加者に利用してもらい、回答の改善作業を行わせた。その後、参加者に対してインタビューを実施した。一方、対照群においては、AI セルフクロンの回答内容を提示する代わりに従来型のフィードバック文を提供した。このフィードバック文は、模擬面接フェーズにおける参加者の回答に対するものであり、GPT-4o*2を用いて生成している。この際、AI セルフクローンに指示したフレームワークと同様の観点からフィードバック文を生成するように指定している。対照群においても回答の改善作業後、インタビューを実施した。なお、情報収集フェーズにおいて収集した参加者の情報、模擬面接フェーズにおける参加者への質問内容は以下の通りである。

収集情報

- (1) **収集情報 1**：あなたがこれまでに力を入れて取り組んだこと (200 文字以上)
- (2) **収集情報 2**：あなたの強み (5 文字以上)
- (3) **収集情報 3**：あなたの弱み (5 文字以上)

質問項目

- (1) **質問項目 1**：あなたがこれまでに力を入れて取り組んだことについて、詳しく教えてください。
- (2) **質問項目 2**：あなたの強みは何ですか。また、あなたが力を入れて取り組んだことの中で、その強みはどのように活かされてきましたか。
- (3) **質問項目 3**：あなたの弱みは何ですか。また、あなたが力を入れて取り組んだことの中で、どのようにしてその弱みに対処しましたか。

4.3 データの分析

本実験では、テーマティック・アナリシス法 [13] により

*2 <https://platform.openai.com/docs/models/gpt-4o>

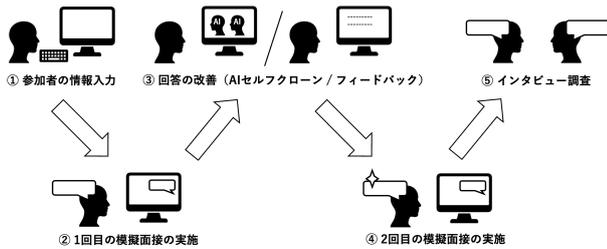


図 3 ユーザー実験の手順。実験群と対照群は、回答の改善作業において提供する情報のみが異なる

Fig. 3 The procedure of the user experiment. The experimental group and the control group differed only in the provided information in the response improvement phase.

データの分析を実施した。第一著者がインタビューの議事録に対してコード化を行い、その後共同研究者による精査を経て、合意に達したコードを確定させた。その後、確定したコードを第一著者によって複数のテーマに分類することで、最終的な質的結果を得た。

5. 結果

5.1 AIセルフクローンを活用した回答改善の利点

ここでは、実験データの分析によって得られた、AIセルフクローンを用いた面接での回答改善における利点を、主に従来のフィードバックによる回答の改善作業との比較の観点から述べる。

5.1.1 発言が与える印象の提供

AIセルフクローンが生成した発言を観察することで、ユーザーは回答に含まれるフレーズや話題が与える印象を客観的に認識することが可能になる。P2は、AIセルフクローンが自身の経験を述べる際に、自身の経験の詳細だけでなく、それを通じて学んだことについて述べている点に着目し、改善後の回答に取り入れた。その理由についてP2は「力を入れて取り組んだことを言うだけよりも、そこから何を学んだか言った方が能動的な感じがすると思ったので入れるべきだったなと思いました。」と述べている。P3は、AIセルフクローンが自身の経験を述べる際に実際に自身に起きた体験や感情を述べている点を観察し、そのような自身の体験や感情に対する言及が回答の説得力が高めていると感じたと述べている。特にポジティブな印象を与える表現（協力、向上、挑戦、成長など）はこの観点において積極的に採用されている（3 | P2, P3, P6）。対照群においても、回答の客観的な印象を認識する重要性を示唆する感想が見られた。P9は従来型のフィードバックにおいて、具体性を削り簡潔に話すようアドバイスを受け、回答に登場させる具体例の数を2個から1個へと削減しようと試みた。しかし、話している最中に具体例が1個だけでは自身が伝えたい内容が十分に伝わらないのではないかと、という疑念が生じ、最終的に具体例は改善前と同様の2個登場させ、簡潔性を高めることができなかつたと述べている。

その一方で、AIセルフクローンの回答の中には、ユーザーが客観的に観察した結果、聞き手に対して悪印象を与えると判断され採用されなかったものも多く見られた。その具体的な内容については5.2.2で紹介する。

5.1.2 体裁が整った解答例の提供

大規模言語モデルによって発言を生成しているAIセルフクローンは、与えられた情報を元に体裁の整った自然な文章を構築し、ユーザーに回答の一例として提供することができる。対照群の参加者は、たとえフィードバックで改善点を指摘されたとしても、具体的にどのように話すことで改善できるのか、具体的な例文をイメージできない、話す内容が思いつかない、と述べた（2 | P7, P8）。それに対して、AIセルフクローンによる回答の改善を行った参加者の中には、AIセルフクローンの回答の一部を自身の回答にそのまま取り入れようとする行動が見られた（P3）。なお、AIセルフクローンが生成する文章に事実とは異なる発言が含まれていた場合、ユーザーはそれをそのまま取り入れず、誤りがある部分を自身の文脈に合わせて修正することで取り入れていた（P3）。

5.1.3 改善の多様性

AIセルフクローンを用いた回答の改善ではユーザーが自由に改善点を探索できる一方で、従来のフィードバックはシステムが特定の改善点を指摘し、ユーザーが受動的にその改善点を受け入れることになる。そのため、従来のフィードバックによる回答の改善では改善の方向性が限定されてしまう可能性がある。一例として、「回答の具体性を向上させる」という改善項目を挙げた参加者が実験群では半数（3 | P2, P3, P6）にとどまった一方で、従来型のフィードバックを受けて回答の改善を行った対照群ではP7~P12の全ての参加者が回答の具体性向上を改善項目として挙げていた。

5.1.4 ユーザーの過去に対する新しい視点の提供

AIセルフクローンの回答を観察することで、新しい視点から自身の過去を振り返り、これまでとは全く別の視点から質問に回答する、という効果も確認された。P4は自身の情報として、2つのアルバイト経験（経験A、経験B）、および1つの自身の強みをシステムに提供していた。そして模擬面接フェーズにおいて、自身の経験において強みが活かされたエピソードを問われた際に、自身の強みを経験Aにのみ紐づけて語った。一方でAIセルフクローンはP4の強みを経験A、経験Bの2つそれぞれに紐づけて語った。P4はその回答を観察し、自身の経験Bをもう一度振り返ることで、経験Bにも自身の強みが生きていたことに気づくことができたと述べている。

“自分のアルバイトの経験が2つあって、自分は後者の方のみを言っていたんですけど、前者の方も自分が最初想定していた自分の強みを、自分

は後者の方にしかあまり活かされていないと自分でも考えていたんですけども、AI エージェント*3の回答を見て、確かに前者の方でも自分の強みを活かされているんじゃないかなということに初めて気がついて、AI エージェントのような回答、両方採用して自分の取り組み2つについて強みとして述べました。” - P4

5.2 AI セルフクローンを活用した回答改善の欠点

ここでは、AI セルフクローンを用いた面接での回答改善における欠点を、主に従来のフィードバックによる回答の改善作業との比較の観点から述べる。

5.2.1 単調、表層的な発言

面接では、自己分析や深い自己理解が求められる質問が投げかけられることがある。このような質問において、AI セルフクローンは十分な効果を発揮できない可能性がある。P1 は、自身の弱みに関する質問（質問項目3）に対するAI セルフクローンの回答について以下のように述べている。

“(AI セルフクローンの回答は) 嘘をついているとかは全くないです。ただ強引に（与えた情報を回答に）入れちゃってないかと思います。論理的には繋がってはいるんで、めちゃくちゃなことを言っているわけじゃないですけど、弱みと ES*4に書いたことをくっつけただけ。” - P1

また、AI セルフクローンの発言は同じ内容を繰り返しているだけであり、ただ時系列に沿って話している、という意見も見られた (P2)。

5.2.2 不自然な言葉遣い

AI セルフクローンの回答は、大規模言語モデル特有の言い回しや不自然な表現が見られ、そのまま自身の回答に取り入れることができない、というケースが確認された。P1 は AI セルフクローンの発言に登場した「感受性を生かして」という表現に語法上の違和感を感じ、AI セルフクローン全体に対する不信感が高まったと述べている。また、P6 は、サークル活動への取り組みについて、一部の AI セルフクローンの回答に対して次のように述べている。

“(AI セルフクローンの回答が) 「以前の代表はこうだった、それに対して僕は合理的なアプローチが取り入れていて…」と嫌味っらしいなと思っ

て、そういう意味で使いたくないなと思った。” - P6

5.2.3 事実と反する発言

AI セルフクローンには参加者が提供した情報に基づき回答を生成するようプロンプト内で指示しているが、それでもお発言内容に事実と反する内容が含まれるケースが見られた。

“日常的に発声練習を行ってと書かれていたんですけど、それはやってないな、これは違うなって思いました。” - P3

“ほとんどのエージェントに、「お客様の反応を見て、次に何が必要かを考え、迅速に対応するように心がける」と書いてあったんですけど、実際あまりしていないなって感じて、それが書いてないエージェントの文章を探しました。” - P3

5.2.4 具体的な改善手順の提供が不可能

対照群において提供された従来のフィードバックには、「具体的なエピソードや成果を加えることで、あなたの貢献度をより明確に示すことができます。例えば、どのような企画を実施し、それがどのように成功したのか、具体的な結果を示すと良いでしょう。」のように、具体的にどのような方法で改善を行うことが可能か、その手順を言葉で説明しており、ユーザーが改善を行う足掛かりを提供していた。このような具体的な改善の手順の説明は、対照群における多くの参加者が回答の改善に効果的であったと述べている (5 | P7, P8, P9, P10, P12)。また、従来のフィードバックでは、ユーザーの回答の良い部分を褒めるということが可能であり、ユーザーの回答のうち改善すべき部分とそうでない部分を明確にすることができる。実際、対照群における多くのフィードバックでは参加者の回答の良い部分を褒めており、一部の参加者は良い部分を褒められることで自分が参考になった、褒められることで自分の回答に自信を持つことができた、と述べている (2 | P9, P12)。一方で AI セルフクローンによる回答改善を実施した実験群においては、そのインタラクションの性質上当然ではあるが、回答の具体的な改善手順の説明が役に立ったという意見は確認できなかった。

6. 考察

6.1 AI セルフクローンにおける理想的な自己のパフォーマンスの提供機能

本研究の主要な問いである「AI セルフクローンは理想的な自己のパフォーマンスをユーザーに提供できるか」について、5.1.1, 5.1.2 および 5.1.3 の結果より、AI セルフ

*3 実験時には、AI セルフクローンに関して AI エージェントという名称で参加者への説明を行っていた。以降に登場する AI エージェントやエージェントという発言は、全て AI セルフクローンのことを指す。

*4 ここでは、収集情報 1 「あなたがこれまでに力を入れて取り組んだこと」を指している

クローンはユーザーの文脈に沿った理想的な回答を提供し、ユーザーはその発言が与える印象を加味しながら、良いと思った構成・フレーズ・まとめ方を自由に探索することで改善の方向を明確にし、適宜 AI セルフクロンの発言内容を修正しつつ、自身の回答に取り入れることが可能であることが分かった。一方で5.2.4で示した通り、AI セルフクロンの回答に基づき改善点を見出す作業は、それ単体では具体的な改善の手順をユーザーに示すことができず、回答の改善がユーザーの能力に依存するという課題もある。具体的な改善の手順を示せないこと自体は5.1.3で述べた通り改善の方向性を限定させないという利点を持つが、そもそも何をどのように改善すれば良いのか全くわからないユーザーにとって、この改善の多様性は意味を為さない。面接の初心者をユーザーとする場合には、AI セルフクロンはそれ単体で利用するのではなく、回答を改善するための詳細な手順を分かりやすい形で説明できる従来型のフィードバックと併せて利用することで、より AI セルフクロンを用いた回答改善プロセスが容易になる可能性があり、今後さらに追求すべき点である。

6.2 分析を伴う回答生成の必要性

5.2.1の通り、AI セルフクロンは自身に関する分析を必要とする回答の生成には不向きであることが分かった。しかし現実の面接では、応募者の分析能力を評価するための質問は往々にして存在する。本研究における AI セルフクロンの位置づけは「理想的な自己のパフォーマンスをユーザーに提示すること」であり、ユーザーの分析能力の向上は目的としていないが、理想的な自己のパフォーマンス自体は深い分析に基づいている必要がある。AI セルフクロンが自己の分析能力を身につけた上で、表層的にユーザーになりきるのではなく、深い分析に基づく回答を生成することが望ましい。AI セルフクロンの基盤となる大規模言語モデルの推論能力については Chain of Thought (CoT) [14] をはじめとした多くの研究が実施されており、現時点において既に AI セルフクロン自体に深い分析を行う能力はある。本システムにおいてボトルネックとなるのは、分析のために必要なユーザーの情報である。今回の実験ではユーザーに関する情報は数百文字程度のテキスト情報であった。たとえ AI セルフクロンが分析能力を保持していたとしても、分析対象となるユーザーの情報が少なければ深い分析を行うことはできない。面接で評価されるような深い分析に基づく回答を生成するために必要な情報とは何か、そしてそれはどの程度の量が必要であるかを今後明らかにする必要がある。

また、AI セルフクロンの思考過程をユーザーに提供する必要性も今回の実験で確認された。実験群の参加者 P2 は、参加者自身の回答と AI セルフクロンの回答に大きな乖離があった際に、自身の回答のどこが問題であるの

か、AI セルフクロンに理由を話してほしいと述べていた。実際、Daryanto ら [2] の提案する大規模言語モデルを用いた面接支援システムでは、ユーザーとシステムは自然言語を通じた対話が可能であり、なぜシステムはそのようなフィードバックを提供したのか、などシステムが提供する情報に対してユーザー自身が質問することを可能にしている。本システムにおける AI セルフクロンはユーザーと対話を行うことはできないが、今後 AI セルフクロンとユーザーの対話を可能にすることで、AI セルフクロンはどのような自己の分析に基づき回答を生成したのか、生成された回答の根拠をユーザーに提供できる可能性がある。

6.3 ユーザーへの新たな視点の提示

5.1.4で紹介した通り、AI セルフクロンはユーザーに新たな視点からの過去の振り返りを促す役割を担うことがある。この現象は、AI セルフクロンを動作させる大規模言語モデルが、与えられた情報を元にユーザーの意図していない文章を作り出したことに起因している。与えられた情報がユーザーの意図せぬ形で回答に利用されることは、5.2.3で紹介した事実と反する発言の生成と類似しているが、大きな違いはユーザーの意図と異なる部分が客観的事実に関する内容であるか、動機などの心理に関する内容であるかと考える。前者であれば、AI セルフクロンの発言は虚偽の発言に過ぎないが、後者であれば事実と反する内容だけでなく、ユーザーが気づいていない、意識したことがなかったが本来は感じていた心理に言及した内容が含まれる場合もある。その場合、ユーザーは AI セルフクロンの発言を観察することで、これまで持っていなかった視点を獲得することが可能となる。事実と反する虚偽の発言を抑制する点や、6.2で紹介したような深い分析に基づく回答を生成する点では、曖昧性を減らすためにユーザーに関する情報をなるべく多く集めた方がよい。その一方で、新たな視点を提示するという観点からは、AI セルフクロンに一定の創作を許容し、ユーザーの意図しない自由な回答を生成させることが必要となる。どの程度の情報を提供すべきであるかは、AI セルフクロンを用いた面接のシミュレーションにより、ユーザーに何を提供したいかによって柔軟に変えるべきであると言える。

ただし、この効果は従来型のフィードバックにおいても再現できる可能性が残っており、ユーザー自身のように振る舞う仮想的な分身という性質を持つ AI セルフクロンに特有の効果であるか否かは、さらなる検証が必要である。

7. 終わりに

本研究では、従来の面接支援システムにおける、自己の理想的なパフォーマンスの不可視性、という課題に対し、大規模言語モデルを基礎としたユーザーの仮想的な分身である、AI セルフクロンを用いたシミュレーションペー

スの面接支援システムを提案し、ユーザー実験によりその有効性の検証を行った。

その結果、AIセルフクローンはユーザーの文脈に基づいた回答の中で好印象を与える単語や、伝わりやすいフレーズなどの理想的なパフォーマンスをユーザーに提供し、面接における回答の改善作業を支援できることが分かった。また、AIセルフクローンをを用いた面接支援には回答の改善だけでなく、ユーザーが過去の経験を新しい視点で見直すなどの当初は意図していなかった効果をもたらす可能性が示唆された。一方で、AIセルフクローンの回答に事実と反する内容が含まれる点や、AIセルフクローンでは回答を改善するための具体的な方法を提示できない、などの課題も発見された。

今後の展望としては、上記の課題を解決するために従来型のフィードバックとAIセルフクローンをどのように掛け合わせることができるか、それぞれの欠点を補い合うシステムのあり方を探究するとともに、過去の経験に対する新たな視点をユーザーに提供するというAIセルフクローンの効果の検証を進めたいと考えている。

8. 生成 AI の利用に関する開示

本研究では、以下の3つの用途に限り生成AIを使用している。

- (1) 提案システムにおけるAIセルフクローンの回答・フィードバックの生成
使用モデル：OpenAI gpt-4o-2024-08-06
- (2) 提案システムのWebアプリケーション実装におけるコーディング支援
使用モデル：OpenAI o3-2025-04-16
- (3) 執筆済みの本原稿に対する、文法・語法上の校正
使用モデル：OpenAI gpt-5-chat-latest (2025年10月31日時点)

それ以外の用途においては生成AIを使用していない。

参考文献

- [1] Huss, R., Jhileek, T. and Butler, J.: Mock Interviews in the Workplace: Giving Interns the Skills They Need for Success, *Journal of Effective Teaching*, Vol. 17, No. 3, pp. 23–37 (online), available from <https://eric.ed.gov/?id=EJ1175757> (2017).
- [2] Daryanto, T., Ding, X., Wilhelm, L. T., Stil, S., Knutson, K. M. and Rho, E. H.: Conversate: Supporting Reflective Learning in Interview Practice Through Interactive Simulation and Dialogic Feedback, *Proc. ACM Hum.-Comput. Interact.*, Vol. 9, No. 1 (online), DOI: 10.1145/3701188 (2025).
- [3] Huang, J., Kim, I.-J. and Yoon, D.: Mirror to Companion: Exploring Roles, Values, and Risks of AI Self-

- Clones through Story Completion, CHI '25, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3706598.3713587 (2025).
- [4] Nambiar, S. K., Das, R., Rasipuram, S. and Jayagopi, D. B.: Automatic generation of actionable feedback towards improving social competency in job interviews, *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*, MIE 2017, New York, NY, USA, Association for Computing Machinery, p. 53–59 (online), DOI: 10.1145/3139513.3139515 (2017).
- [5] Spitzberg, B. H.: What is Good Communication?, *Journal of the Association for Communication Administration*, Vol. 29, No. 1, p. 7 (2000).
- [6] Boyer, L., Pleasant, J. and Vest, D.: S.T.A.R. Performance: A Quantitative Exploration of Behavioral Responses in Simulated Selection Interviews, *College of Business E-Journal*, Vol. 14, No. 1, p. Article 3 (n.d.). Access date not specified.
- [7] Pataramutaporn, P., Danry, V., Blanchard, L., Thakral, L., Ohsugi, N., Maes, P. and Sra, M.: Living Memories: AI-Generated Characters as Digital Mementos, *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, New York, NY, USA, Association for Computing Machinery, p. 889–901 (online), DOI: 10.1145/3581641.3584065 (2023).
- [8] 山本賢太, 駒谷和範: 大規模言語モデルにより性格特性を反映させた発話生成の評価, 人工知能学会研究会資料言語・音声理解と対話処理研究会, Vol. 100, pp. 07–12 (オンライン), DOI: 10.11517/jsaislud.100.0.07 (2024).
- [9] Goldberg, L. R.: An alternative “description of personality”: The Big-Five factor structure, *Journal of Personality and Social Psychology*, Vol. 59, No. 6, pp. 1216–1229 (online), DOI: 10.1037/0022-3514.59.6.1216 (1990).
- [10] 小塩真司, 阿部晋吾, Cutrone, P.: 日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み, パーソナリティ研究, Vol. 21, No. 1, pp. 40–52 (オンライン), DOI: 10.2132/personality.21.40 (2012).
- [11] Inoue, K., Hara, K., Lala, D., Yamamoto, K., Nakamura, S., Takamashi, K. and Kawahara, T.: Job Interviewer Android with Elaborate Follow-up Question Generation, *Proceedings of the 2020 International Conference on Multimodal Interaction*, ICMI '20, New York, NY, USA, Association for Computing Machinery, p. 324–332 (online), DOI: 10.1145/3382507.3418839 (2020).
- [12] Blake, R. J. and Sackett, S. A.: Holland's Typology and the Five-Factor Model: A Rational-Empirical Analysis, *Journal of Career Assessment*, Vol. 7, No. 3, pp. 249–279 (online), DOI: 10.1177/106907279900700305 (1999).
- [13] Braun, V. and Clarke, V.: Using thematic analysis in psychology, *Qualitative Research in Psychology*, Vol. 3, No. 2, pp. 77–101 (online), DOI: 10.1191/1478088706qp0630a (2006).
- [14] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichtter, b., Xia, F., Chi, E., Le, Q. V. and Zhou, D.: Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, *Advances in Neural Information Processing Systems* (Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K. and Oh, A., eds.), Vol. 35, Curran Associates, Inc., pp. 24824–24837 (online), available from https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf (2022).