

潜在的な心理バイアスに着目した対話型 AI による コミュニケーション能力訓練システム

酒井 真紀†
Maki Skai

佐野 睦夫‡
Mutuo Sano

1. はじめに

現代の若者におけるコミュニケーションは、主にテキストや画像・動画を中心とした非対面形式が主流となり、文字やスタンプで情報や感情を伝える文化が定着している。このような環境下において、対面での即時対応が求められるコミュニケーション（例：職場での電話応対など）に対し、若者がストレスを感じるケースが報告されている[1]。一般社団法人 日本経済団体連合会「2018 年度 新卒採用に関するアンケート調査結果」では、「(参考) 選考にあたって特に重視した点(5つ選択)」において、「コミュニケーション能力」が 82.4%で第1位となっている[2]。また、楽天インサイト株式会社が行った 2024 年度の調査結果の図 1[3]でも、「社会人に特に求められていると思うスキルの回答」においてコミュニケーション能力が 1 位となっており、現状でも変わらずコミュニケーション能力は社会活動の中で重要とされるスキルの 1 つであると考えられる。

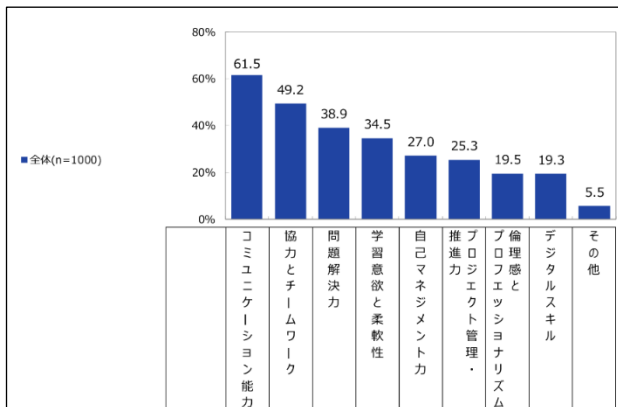


図 1 2024 年度社会人に特に求められていると思うスキルの回答

しかし、このようなトレーニングを行うには、場所・時間・金銭的制約や対人ストレスが伴う。

本研究では、対話型 AI を活用し、アクセスしやすく心理的負担を感じにくい環境で、対人コミュニケーション能力(対応力、適応力、発想力)の向上とともに、潜在的な心理バイアスに着目し、テキストから画像が生成されるプロセスにおいて、ユーザーが伝えていない情報である暗黙情報を LLM

と画像生成モデルが補完する仕組みを利用し、フィードバックに画像を用いることにより、潜在的な心理バイアスへの気づきから、メタ認知につなげるトレーニングシステムを開発・評価することを目指すものである。

コミュニケーション・スキルの定義は、藤本らの研究[4]で記されているように「ソーシャル・スキルと概念上の重複が見られるなど一意ではない」[藤本等,2007:347]が、本研究では、次のように記されているものを参考とした。コミュニケーション能力、対応力、適応力、発想力については次のように記されている。コミュニケーション能力は「コミュニケーションとは対人関係での情報共有や意思の疎通のことであり、コミュニケーション能力は、それらをスムーズに行うことができる力」[5]。対応力は「対応」の項目第 4 義にあたる「周囲の状況などに合わせて事をする事」[6]から周囲の状況に合わせて物事に対処する力と解釈する。適応力は「環境に対して速やかに適応できる能力」[5]。発想力は「何かを思い付き、新しいものを生み出す力」[5]とある。本システムでは、対話型 AI が複数のキーワードを提示し、ユーザーは即興的に予測できないキーワードを用いてストーリーを作成することで、これら力の向上に期待するものである。

2. 関連研究

LLM を用いたトレーニングシステムの先行研究として Yang らが行った研究[7]がある。AI 活用の可能性を探索し、LLM を用いたソーシャルスキルトレーニングシステムにおいて、より多くの人々にアクセス可能な学習機会を提供することを目指し、トレーニングによるスキル向上が示されている。また、LLM を社会的利益のために、特に社会的スキルのトレーニングのような分野で活用するために不可欠であると述べられている。

次に、今回着目した潜在的な心理バイアスについては Atreides らの研究[8]によって、LLM にも認知バイアスが見られることがデータに基づいて報告されており、人間の偏りを反映・再構築している可能性があるという示唆がなされている。また、これまで LLM の認知バイアスについて検出

† 大阪工業大学大学院, Osaka Institute of Technology Graduate School

‡ 大阪工業大学, Osaka Institute of Technology

する研究が多かったが、Lemieux らの研究[9]では、LLM と高度なプロンプトエンジニアリング技術を用いて、ユーザー生成テキストにおける認知バイアスをリアルタイムで検出する新たなアプローチが示されている。これらのことから、LLM により潜在的な心理バイアスのフィードバックを行うことが可能であると考えられる。

コミュニケーションにおいては、林らの研究[10]の中で「人間の言語コミュニケーションにおいて、他者の視点に立つ認知プロセスを理解するというのは、極めて重要な課題であるといえる。日常会話場面でコミュニケーションが成立するためには、情報の送り手と受け手がお互いの状況や意図を推測しあいながら情報交換を行う必要がある」[林等,2011:580]と述べられている。この他者視点の認知プロセスにおいて、本研究では画像生成を用いる。

画像生成においては、テキスト→画像生成は大きく以下の三つのステップで構成されている[11,12,13,14,15]。①テキスト埋め込みの生成：ユーザーのプロンプトは Transformer や CLIP といったテキストエンコーダを通じて、多次元の意味ベクトル（テキスト埋め込み）へと変換される。②潜在表現の生成：代表的な画像生成モデルとして、トークンを自己回帰的に生成する自己回帰型（DALL·E 1）、ノイズ→逆拡散で生成する拡散モデル（Stable Diffusion など）、条件付き GAN（AttnGAN など）がある。いずれもテキスト埋め込みを条件情報として用いるが、拡散モデルではクロスアテンション+Classifier-Free Guidance によって高度なプロンプト適合を実現する。③生成された潜在表現は、自己回帰型では dVAE のデコーダ、拡散モデルでは潜在空間のデコーダ/アップサンプリングネットワーク、GAN ではジェネレータ本体を用いて RGB ピクセル空間に復元される。また、画像生成過程ではアテンション機構を介して、モデルが学習データから獲得した統計的先行知識を動的に参照し、プロンプトで指定されていない背景情報や構図、質感などの詳細を適切に埋め込みながら生成が行われる。図2はテキストから画像への生成プロセスを表したものである。

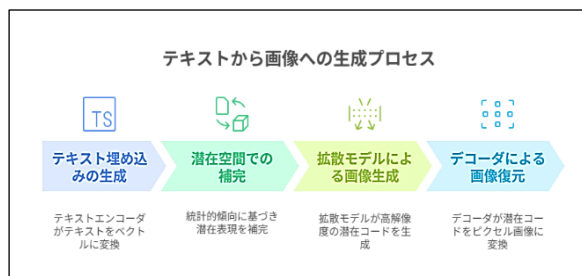


図 2 画像生成プロセス

本研究では、LLM で生成したテキスト埋め込みを画像生成モデルが統計的先行知識に基づいて暗黙情報を補完し、

潜在空間上に妥当な解釈を配置する仕組みを利用する。さらに、その結果得られた画像をフィードバックとして活用するものである。

3. システムについて

本研究では、潜在的な心理バイアスが「話者が双方向的にコミュニケーションを行う場面では、コミュニケーションの齟齬として現れると考えられる」[林等,2011:580]、聞き手は相手の意図を汲み取るために、自分と相手が共有する知識を参照する必要があるが、しばしば自己中心的ヒューリスティックを用いることがある[16]という考えをもとに、ストーリーの可視化を行うことで自分の伝えた（伝えたい）ことが 100%正確に相手に伝わるわけでないという体験から気づきを得ることを目指している。また、「他者視点の理解の促進が、異なるコミュニケーションプロセスが媒介することによりもたらされたことは興味深い」[林等,2011:582-583]という考えからコミュニケーションの方法が違えば、他者の視点を理解するプロセスも変わる、それが結果的に他者視点の理解を促進することにつながると理解し、画像でのフィードバックというプロセスが媒介することで、他者視点理解の促進がもたらされる可能性があると考えられる。

図 3 は、本研究の画像フィードバックにおける概念図である。

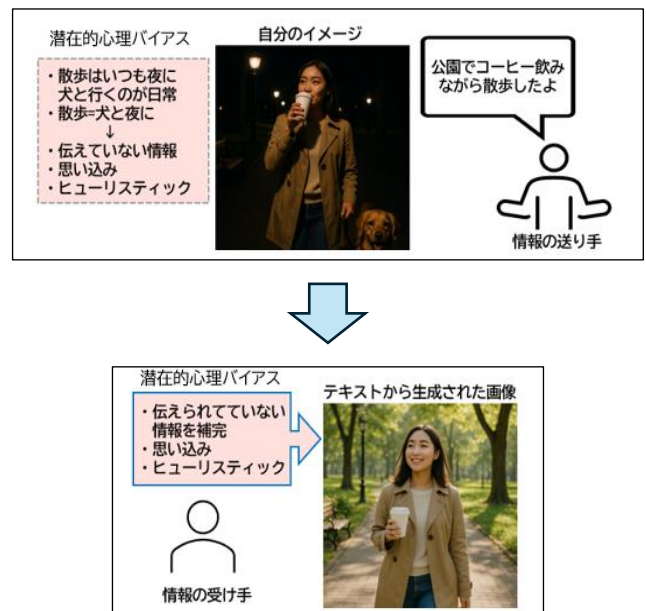


図 3 画像フィードバックの概念図

自分にとって当たり前の情報（散歩はいつも夜に犬と行く）を伝えていない場合、生成 AI は平均的解釈を配置することで不完全なテキスト情報を充足し画像生成を行う（テキストから生成された画像）。このような場面は日常に限らずビジネスの場でも生じており、林らは「それぞれの分野で暗

黙の前提となっている問題に対する捉え方の違いや、着眼点の違いが、コミュニケーションの齟齬を引き起こし、逆に協同問題解決の制約が現れる場合もある」[林等,2011:569]としている。

以下、図4はシステムの構成図である。

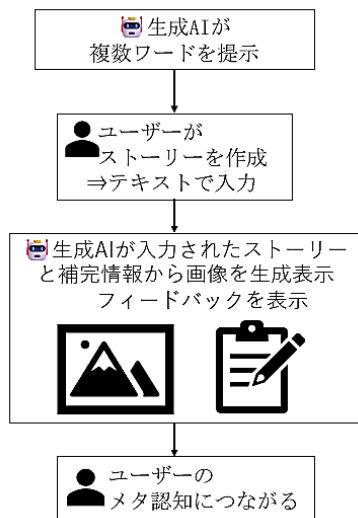


図4 システム構成図

ユーザーは対話型 AI から提示された複数のキーワードを使用して、即興的にテキストでストーリーを作成する。事前に準備ができない状況でユーザーはストーリーを作成する体験から、対応力・適応力・発想力を養う。ストーリーを受け取った対話型 AI は、画像生成モデルからユーザーが伝えなかった情報である暗黙情報を補完して画像を生成する。対話型 AI は生成された画像と、作成されたストーリーの文章に対するフィードバックとともに、補完した内容と潜在的的心理バイアスについてのフィードバックを表示する。ユーザーは自分が作成したストーリーの映像が頭の中にあり、生成された画像との相違で気づきを得ることができる。相違がない場合は、テキストのフィードバックにより、潜在的的心理バイアスについての学びを得ることができるものである。

4. 潜在的的心理バイアスの実験

大学生・大学院生の11名（男性10名、女性1名）に協力をいただき、潜在的的心理バイアスについての実験を行った。実験の内容としては、まず実験の条件を揃えるため、①参加者全員に同じストーリーが提示された。・提示されたストーリー：「Aさんは、カフェの窓際テーブルで、手元の資料を見ながら誰かと小声で話していた。ときどき資料に線を引いたり、ペンを走らせたりしている。」

次に②「提示されたストーリーからどんな映像が浮かびましたか」という問いに対し記述を行っていただいた。記述が終了した時点で、③画像1と画像2が示され「画像が示されたことにより、潜在的的心理バイアスについて気づきを得た点がありましたらご記入ください」という問いに対し記述を行っていただいた。

また実験後には、該当項目にチェックをお願いする実験後アンケートが実施された。図5は③で示された画像1と画像2である。今回の実験では対象が学生であったため、学生と一般的に想像しやすいと思われるビジネスパーソンの画像を用意した。



図5 実験で提示された画像

5. 結果と考察

実験後アンケートはリッカート尺度を用いて評価し、自由記述を設けて意見を募った。以下、結果の内容である。表1で示す実験記述アンケートの結果では、項目「ストーリーのみ」ではストーリーから浮かんだ映像についての内容が記されている。対象は全員大学生であるが、「キャリアウーマン」、「友達と勉強」、「マンガ家」、「ビジネスの商談」などシーンが多岐にわたっており、潜在的バイアスが年代で限定されるものでなく、個々によって多様であることを示している。実験後の選択アンケートでは、図6で示す問1 普段の潜在的的心理バイアスの意識について「意識している・やや意識している」と「意識していない・やや意識していない」がそれぞれ5名(45%)の同数であり、潜在的バイアスの意識について年代で限定されるものではないことが示されている。図7で示す問2の画像の有効性では、「有効・やや有効」との回答が11名(100%)と画像の提示において潜在的的心理バイアスの気づきを得ることの有効性が確認された。図8で示す問3のこの体験を機とした今後の潜在的的心理バイアスの意識について「思う・やや思う」が9名(80%)、「やや思わない」2名(18%)と、今回の体験が今後のコミュニケーションにおいて潜在的バイアスの意識につながるものとなった。また表2で示す問4の自由記述では、「画像と頭の中で比較しやすかった」「抽象的な表現の場合、適切な伝わ

り方をするのは難しいことが分かった」「画像を見ると想像していたものと違って、心理バイアスを意識しているつ

もりはなくても、少しバイアスを考えているのかもしれないと思った」などの記述がみられた。

表 1 実験アンケート結果

回答者	ストーリーのみ	画像表示後
1	仕事できそうなキャリアウーマンが話している	カフェで資料を使っていることから、女性かつ働いている様子を想像してしまった
2	カフェで楽しんでいる	画像2を想像したが、3人以上は想像していなかった 人数分のドリンクがあると思った、グラスを想像 社会人の想像はなかった 対面だと思った
3	カフェの中なので小声で話している	窓際でAさんが小声で話しているとあるから、1対1で話していると思った
4	Aさんは男性であり、スーツを着ている様子を想像した カフェは木造で道沿いに位置し、一面ガラス張りの状況を想像した	画像2に近いが、向かい合っている様子を想像した テーブルは倍以上の広さを想像した ノートではなく数枚のプリントホッチキス止めの資料を想像した
5	カフェで何かについて話し合っている2人 片方はスーツで片方は私服 スーツの方がAさんで、手元に何かしらの資料を広げている 広めのテーブル	画像2の方を想像した ノートのようなものではなく、紙の資料のようなもののイメージだった 窓に面していると思ったが、窓は後ろにあった なんとなく女性同志のイメージだった
6	カフェ内で友達と勉強、あるいは試験の勉強を行っている映像	画像1を想像した 私自身が学生という観点から、画像2のようなサラリーマンの姿はイメージされなかった ほとんど画像1のものと同じものを考えていた
7	カウンター席に座っている 紙の資料を見ている 蛍光ペンを使って線を引き、赤ペンで書き込んでいる	学生でなく社会人を想像していた 男性だけだと思った 3人ではなく2人だと思った
8	マンガ家が原稿を担当に持って行き、その話をしている	想像とは両方違った 画像1は高校生の時にしていたので懐かしいと感じた 画像2はイメージになかった。カフェでスーツを着た人が本格的な話しをしているイメージがない
9	ビジネスの商談をしている 対面に座っていて、両者ともにスーツを着ている男性である	画像2にバイアスがかかっていた 年齢や性別に関しては、違う可能性を考慮していたが、人数に関しては2人の前提で考えていた
10	資格勉強をしている本人とその友人	画像1よりも画像2におけるイメージが強い 画像2では資格というより、知識を深める会話をしているのかなと感じた
11	カフェの窓際に人が2人隣同士で座っている 2人は仕事の資料を見て話している その資料の大事な部分に赤ペンで線を引いている 資料はPCではなく紙である	最初に思ったのは画像2の店が近い しかし男女2人組を想像したので、画像とまったく同じではなかった 資料と言われると、図や表のあるものだったと思ったが、画像では文字ばかりだった ドリンクは浮かばなかった 線を引くなら、色ペンかマーカーを想像する 窓際と言われると想像するのと少し違った

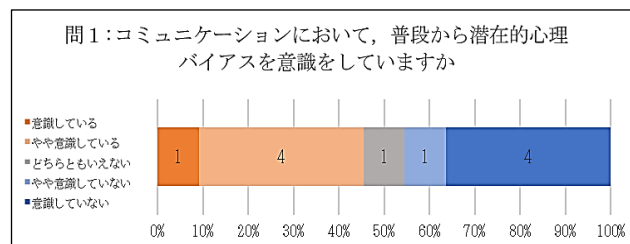


図 6 潜在バイアスへの意識 実験後アンケート結果

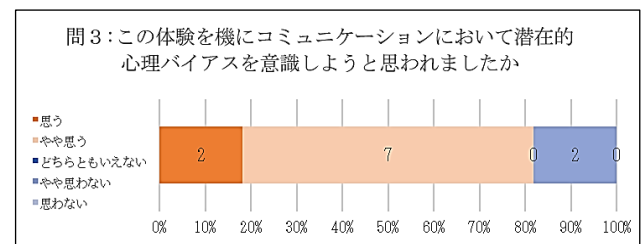


図 8 意識の変化 実験後アンケート結果

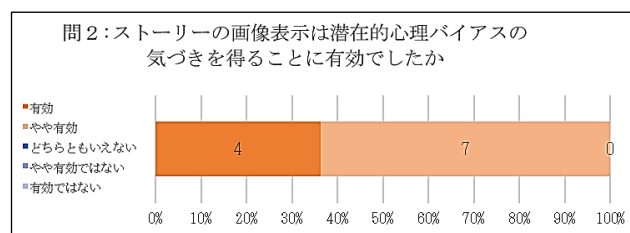


図 7 画像の有効性 実験後アンケート結果

表 2 自由記述 実験後アンケート結果

問4:その他（自由記述）

画像と頭の中で比較しやすかった
頭で考えた情景と画像を照らし合わせると、バイアスに気づきやす いと感じた
とても簡単で非常に興味深い実験だった
2枚の画像表示が自分が想像していたものとどちらも少し違っていた ので、自分が想像していたか少し悩んだ
抽象的な表現の場合、適切な伝わり方をするのは難しいことが分 かった
今回のストーリーで凄く今自分がやっていることを書いた感覚だっ た
画像を見ると想像していたものと違って、心理バイアスを意識して いるつもりなくても、少しバイアスを考えているのかもしれないと 思った

また、実験後アンケートの問1(潜在的な心理バイアスへの事前意識)と問2(画像提示による効果実感)の関連性を検討するため、カイ2乗検定を実施した。検定の結果、カイ2乗値は3.44、p値は0.487であり、有意水準5%を超えており統計的に有意な関係は認められなかった。

6. 今後の展望と課題

今回は本研究で着目する潜在的な心理バイアスについて実験を行った。ユーザーが作成したストーリーをLLMと画像生成モデルが暗黙情報を補完した画像を生成し、可視化を行うことで自分の伝えた（伝えたい）ことが100%正確に相手に伝わるわけでないという体験から気づきを得ることを目指すものであり、実験結果から画像を示すことの有効性を確認することができた。今後の展望として、作成したストーリーをテキストで返す段階から音声で返す段階など、ユーザーが段階的に学習・レベルアップができるような設計が必要ではないかと考える。また課題としては、画像とともに示される対話型AIの潜在的なバイアスフィードバックについての実装と本研究で扱うコミュニケーション能力（対応力・適応力・発想力）のスキル向上についての実験・評価を行うことである。

参考文献

- [1]東洋経済オンライン、営業なのに「電話が怖い…」、若者に広がる「電話恐怖症」のリアル。電話が苦手な部下に上司はどう対応すればいいのか？、<https://toyokeizai.net/articles/-/880458?display=b>、(参照 2025-7-3)
- [2]一般社団法人 日本経済団体連合会 <https://www.keidanren.or.jp/policy/2018/110.pdf>、(参照 2025-7-2)
- [3]楽天インサイト。社会人の意識に関する調査 <https://insight.rakuten.co.jp/report/20240314/>、(参照 2025/7/2)
- [4]藤本学,大坊郁夫, “コミュニケーション・スキルに関する諸因子の階層構造への統合の試み”, 日本パーソナ

- リティ心理学会, 第15巻 第3号 347-361, J-STAGE, https://www.jstage.jst.go.jp/article/personality/15/3/15_3_347/_pdf-char/ja (2007).
- [5]GLOBIS CAREER NOTE <https://mba.globis.ac.jp/careernote/1181.html> <https://mba.globis.ac.jp/careernote/1503.html> <https://mba.globis.ac.jp/careernote/1322.html> (参照 2025-7-13)
- [6] 小学館 デジタル大辞泉
- [7]Diyi Yang, Caleb Ziems, William Held, Omar Shaikh, Michael S. Bernstein, John Mitchell, “Social Skill Training with Large Language Models”, arXiv, <https://arxiv.org/abs/2404.04204>, (2024).
- [8]Kyrtin Atreides, David Kelley, “Cognitive Biases in Natural Language: Automatically Detecting, Differentiating, and Measuring Bias in Text”, ResearchGate, https://www.researchgate.net/publication/372078491_Cognitive_Biases_in_Natural_Language_Automatically_Detecting_Differentiating_and_Measuring_Bias_in_Text, (2023).
- [9]Frederic Lemieux, Aisha Behr, Clara Kellermann-Bryant, Zaki Mohammed, “Cognitive Bias Detection Using Advanced Prompt Engineering”, arXiv, <https://arxiv.org/abs/2503.05516>, (2025).
- [10]林勇吾, 三輪和久, “コミュニケーション齟齬における他者視点の理解”, 認知科学, 18巻4号 p.569-584, J-STAGE, https://www.jstage.jst.go.jp/article/jcss/18/4/18_569/_article/-char/ja/, (2011).
- [11]Jonathan Ho, Tim Salimans, “Classifier-free diffusion guidance”, arXiv, <https://arxiv.org/abs/2207.12598>, (2022)
- [12]Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever., “Zero-shot text-to-image generation”, arXiv, <https://arxiv.org/abs/2102.12092>, (2021).
- [13]Jonathan Ho, Ajay Jain, Pieter Abbeel., “Denoising diffusion probabilistic models”, arXiv, <https://arxiv.org/abs/2006.11239> (2020).
- [14]Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin., “Attention is all you need”, arXiv, <https://arxiv.org/abs/1706.03762>, (2017).
- [15]Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio., “Generative adversarial nets”, arXiv, <https://arxiv.org/abs/1406.2661>, (2014).
- [16] B Keysar, D J Barr, J A Balin, J S Brauner “aking perspective in conversation: the role of mutual knowledge in comprehension”, Psychological Science, <https://journals.sagepub.com/doi/10.1111/1467-9280.00211>, (2000).