

ユーザの個人特性を考慮した偉人との対話生成手法

北畑 哲也[†] 関 和広^{††} 灘本 明代^{††}

[†] 甲南大学大学院自然科学研究科知能情報学専攻 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

^{††} 甲南大学知能情報学部知能情報学科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]m2524012@s.konan-u.ac.jp, ^{††}{seki,nadamoto}@konan-u.ac.jp

あらまし 生成 AI の発展に伴い対話生成の研究が進んでおり、歴史上の人物や特定のキャラクタとの対話を再現する手法が提案されている。特に学校や会社の創設者など、特定のコミュニティにおける偉人と対話するシステムは理念の共有や、悩み相談への応用が可能である。一方、多くのシステムはどのユーザにも共通した内容の応答を提示しており、ユーザの経験や強み・弱みといった個人特性を考慮した対話を生成していない。個人特性を考慮するために、ユーザのセンシティブデータを用いることが有用であるが、クラウド型 LLM への入力は情報漏洩や学習データとしての利用といったリスクがある。そこで本研究では、具体的な値を含むセンシティブデータを抽象的なデータにする処理を行い、RAG を用いて参照することでユーザの個人特性を考慮した偉人との対話生成手法を提案する。本研究は対象のコミュニティを甲南大学、対象のユーザを本学の学生として創設者である平生鈺三郎との対話生成を行う。

キーワード LLM, RAG, 対話システム, センシティブデータ

1 はじめに

生成 AI の発展に伴い対話生成の研究が進んでおり、歴史上の人物や特定のキャラクタの口調や理念を考慮した対話生成手法が提案されている [8] [10]。特に、学校や会社の創設者など特定コミュニティにおける偉人と対話するシステムは、理念の共有やコミュニティに所属する人々の意思決定支援などへの応用が可能である。例えば、甲南大学の創設者である平生鈺三郎は「人格の修養と健康の増進を重んじ、個性を尊重して各人の天賦の特性を啓発する人物教育の率先」や「人生は長い長いマラソン競走である。」といった独自の理念を持つ。このようなコミュニティの偉人の理念は、コミュニティに所属する人々の行動指針や意思決定の基準となる重要な概念である。そこで、これまで我々は偉人の理念を容易にコミュニティのメンバーに伝える方法として、コミュニティの偉人の理念を含んだスピーチ生成手法 [2] と対話生成手法 [3] を提案してきた。しかしながら、偉人の理念を考慮した対話システムであっても、ユーザの個人特性や経験を十分に反映できない場合、どのユーザに対しても一般的な発話を生成することが多い。そのため、偉人の理念を対話として提示し、ユーザの個人特性を考慮して発話内容や語り方を調整する個人適応型の対話生成が重要である。

個人特性を考慮した対話生成を行うためには、ユーザの経歴や成績といったセンシティブなデータ (以下、センシティブデータと呼ぶ) を含む情報を LLM に入力する手法を用いることが有効と考えられる。しかしながら、ChatGPT¹や Gemini²などのクラウド型 LLM にセンシティブデータをそのまま入力することは、情報漏洩や意図しない学習データへの利用といったリスクがある。そこで本研究では、ユーザのセンシティブデータ

に含まれる具体的な値をマスキングして抽象化した個人特性の情報 (以下、個人特性情報) を用い、クラウド型 LLM へ具体値を直接入力せずに個人適応した対話生成手法を提案する。さらに、偉人の理念を含む文を収集した名言データベースを構築し、ユーザ発話にマスキング済み個人特性情報を加えたクエリ拡張により名言を検索する RAG を組み込むことで、個人特性を考慮した偉人との対話生成を実現する。本研究の技術的ポイントは以下のとおりである。

- センシティブデータを扱った対話生成
- 偉人の理念を考慮した対話生成

本研究は対象のコミュニティを甲南大学、対象のユーザを本学の学生として創設者である平生鈺三郎との対話生成を行う。具体的には、学生のディプロマ・サプリメントから成績や課外活動の経験などを抽出し、学生の個人特性情報を考慮した平生鈺三郎との対話を生成する。ディプロマ・サプリメントは、学生が取得した学位や資格の学修内容を説明する文書であり、本人情報、学位情報、成績評価、学修成果、課外活動などの情報を含む。本研究の提案手法の概要を図 1 に示す。

以下に本論文で提案する対話生成の手順を以下に示す。

1. 学生のディプロマ・サプリメントから成績や課外活動等を抽出し、抽出したユーザのセンシティブデータを含むデータベースを構築する。このデータベースをユーザデータベースと呼ぶ。このとき、センシティブデータの具体的な値をマスキングするためのタグを付与する。
2. 平生鈺三郎に関する書籍から理念を含む文を収集し、名言データベースを構築する。
3. ユーザ発話に対し、マスキング済み個人特性情報を用いてクエリ拡張を行い、名言データベースから関連する名言を検索する。
4. ユーザの発話、マスキング済み個人特性情報、検索結果の名言を統合してプロンプトを生成し、クラウド型 LLM に

1 : <https://chatgpt.com/>

2 : <https://gemini.google.com/>

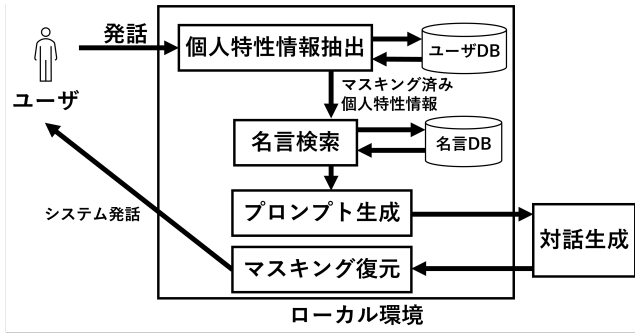


図 1 システム図

よって対話生成を行う。

これにより、具体的な値をクラウド型 LLM に入力せずに、ユーザの個人特性を考慮した偉人との対話生成が可能になる。

2 関連研究

Lewis ら [5] は外部知識を検索して生成に利用する RAG の基本フレームワークを提案し、オープンドメイン QA および言語生成タスクで有効性を示した。本研究では、この RAG のフレームワークに基づき対話システムを構築する。

個人適応型対話生成に関して、ユーザ情報の効率的な利用を行うための様々な研究が行われている。Asthana ら [1] ユーザの入力に対する個人適応の必要性を 3 段階で推定し、必要な場合に限ってユーザの情報をを用いたクエリ拡張と疑似回答を用いた検索手法を提案している。この手法により、個人適応が有効な発話でのみユーザ情報を反映することで過剰な個人適応を抑制することができる。本研究では、個人適応型対話生成に必要な情報を要約し、その要約情報を用いて検索および個人適応型対話生成を行う点で異なる。Wang ら [9] はユーザの過去発話履歴から個人特性情報を抽出することによって、個人適応型対話生成を行う手法を提案している。本研究では、対話履歴からの抽出ではなく、事前に構築したユーザデータベースを参照し、個人適応型対話生成を行う点が異なる。また、これらの個人適応型対話生成手法は、ユーザの個人特性情報を生データのまま利用している。一方、本研究ではユーザのセンシティブデータを保護するために、センシティブデータにマスキングを行い個人適応型対話生成を行う。Zhou ら [11] は LLM を用いてユーザの入力文に含まれるセンシティブデータをマスキングすることにより、ユーザのセンシティブデータを保護する手法を提案している。本研究は、ユーザ入力文そのものではなく、個人適応に用いるユーザの個人特性情報を対象としてマスキングを行う点で異なる。

大規模言語モデルに対象の人物やキャラクターのペルソナ文をプロンプトに挿入することで、特定の人物やキャラクターとの対話生成を行う様々な研究が行われている。Lee ら [4] は応答内容に最も関連するペルソナ文をプロンプトとして対話に挿入することで、特定のキャラクターや人物を再現した対話システムを構築する手法を提案している。本研究は実在した人物を対象としているため、ペルソナ情報を網羅的かつ正確に定義すること

が極めて困難である。そのため、本研究ではペルソナ文の設計ではなく対象人物の著作等から収集した理念を含む文を用いて対話生成を行う。Shao [7] はベートーヴェンのような特定の個人のプロフィールや経験を含むデータから対話データを生成し、そのデータを用いたファインチューニングによって特定の人物との対話を再現する方法を提案した。本研究ではファインチューニングではなく、RAG を用いて偉人との対話生成システムを構築する。

3 データベースの構築

3.1 ユーザデータベースの構築

ユーザデータベースで取り扱うセンシティブデータ

本研究では、個人適応型の対話生成に必要な個人特性情報を保持するために、ユーザデータベースを構築する。本研究におけるユーザは甲南大学の学生とし、ユーザデータベースのデータソースとして、甲南大学が発行する甲南大学ディプロマ・サプリメントを用いる。本研究では、個人を直接特定し得る情報と学修状況や成績評価の値をセンシティブデータとして扱う。つまりは、ユーザデータベースで取り扱うセンシティブデータは、本人情報、学位情報、成績評価、学修成果、課外活動の 5 つのクラスからなる。

- 本人情報は氏名、生年月日、学籍番号からなる。
- 学位情報は学位名と専攻分野からなる。
- 成績評価は年度ごとのその学生の平均 GPA である。
- 学修成果は甲南大学独自の評価項目であり、学生個々の習得科目の成績を元に計算される値である。
- 課外活動は部活やサークルへの所属歴や、ボランティア活動、留学などの経歴が含まれる。

個人特性情報の要約

本研究では、ユーザデータベースの中の個人特性情報を、対話生成に重要な情報であると着目して要約する。これにより、対話生成で参照する必要がある個人特性情報を簡潔にプロンプトで入力することが可能になる。要約処理では、ユーザデータベースの各クラスにおけるインスタンスに対して、クラスごとにデータの特徴を含む要約文を生成する。具体的には、本人情報や学位、課外活動のように記述情報が中心のクラスでは、個人適応した発話を生成するために必要な情報を含むように要約する。さらに、成績評価や学修成果のように数値で表現されるクラスでは、単一の値そのものではなく、年度ごとの推移や特に優れている評価項目に着目して要約する。生成した要約の例を表 1 に示す。

センシティブデータのマスキング

センシティブデータはユーザの個人特性を考慮した対話を行うために必要となる情報である。しかしながら、ユーザの許可なく外部へ共有されてはいけない情報である。そこで、センシティブデータの要約及びマスキングからなる前処理をローカル環境で行い、前処理を行った情報のみクラウド型 LLM を用いた対話生成で利用する。前処理には、OpenAI が公開したオープンウェイトモデルである gpt-oss-20b [6] を用いる。gpt-oss-20b

表 1 ディプロマ・サブリメントから生成した要約の例

クラス	要約文
本人情報	氏名は甲南太郎、生年月日は 2002 年 5 月 15 日です。
学位	学位の名称は学士（経営学）であり、甲南大学の経営学部経営学科から授与されました。
成績評価	GPA は 2022 年度が最も高く 2.89 でした。2022 年度から 2023 年度にかけて低下し、通算での GPA は 2.47 です。
学修成果	2021 年度から 2024 年度にかけて、企業組織の理解や企業と市場の関係理解の達成率が顕著に向上しています。所属ゼミ・研究室は山本ゼミであり、分野・テーマは日本語専攻です。
課外活動	2021 年度から 2024 年度にかけて、体育会硬式野球部の主務や文化会グリークラブの会計を務めました。また、ボランティア活動に参加し、累積 60 時間の活動を行い、SDGs の実践プロジェクトにも関与しました。

表 2 マスキングした要約の例

クラス	マスキング後の要約文
本人情報	氏名は{本人情報_01}、生年月日は{本人情報_03}です。
学位	学位の名称は{学位_01}であり、甲南大学の{学位_02}から授与されました。
成績評価	GPA は 2022 年度が最も高く{GPA 推移表_02}でした。2022 年度から 2023 年度にかけて低下し、通算での GPA は{GPA 推移表_05}です。
学修成果	2021 年度から 2024 年度にかけて、企業組織の理解や企業と市場の関係理解の達成率が顕著に向上しています。所属ゼミ・研究室は{学修情報_19}であり、分野・テーマは{学修情報_20}です。
課外活動	2021 年度から 2024 年度にかけて、{課外活動_01}の{課外活動_02}や{課外活動_03}の{課外活動_04}を務めました。また、ボランティア活動に参加し、累積 60 時間の活動を行い、SDGs の実践プロジェクトにも関与しました。

12 月 9 日

夜来雨今朝ニ至ルモ止マズ昨日午後三時半（其実四時）ヨリ兵庫
県教育会幹事会開会ス。議案ハ四五アリタルモ尤モ重要ナルモ
ノハ先日表彰セラレタル教育ニ関スル功労者、即チ四十年以上
教育ニ従事セラレタル人々ニ対シテ祝賀会ヲ催サントスルモノ
ナルガ、余ハカカル祝賀会ニハナルベク多数ノ会衆ガ賛同スル
コトガ被表彰者ニ対シテ祝意ヲ表スルコトナレバ、ナルベク会費
ヲ廉ニシ会衆ガ集リ易キ機会ヲ扱マザルベカラズト提議セシニ、
来ル二十四、五、六日ニハ神戸ニ於テ県学務部ノ主催ニテ講習会
アルヲ以テ一校ニ一人位ノ教師ガ出席スベク、マタ年末ニハ教員
ノ昇給問題等モアレバ校長ニシテ上庁スルモノモアレバ、来ル二
十四日ハ県会モ閉会后ナレバ好都合ナラントノ事ニテ、二十四日
ト決定。

図 2 平生鈺三郎日記の例

はモデルの重みが公開されている LLM であり、ローカル環境で実行することができる。そのため、甲南大学ディプロマ・サブリメントに含まれるセンシティブデータを外部サービスへ送信することなく前処理を行うことができる。

マスキング処理では、個人を特定できる可能性のある数値や単語を抽象的なタグに置換することにより処理を行う。具体的には、甲南ディプロマ・サブリメントから抽出した各データに対して、クラス名と出現順に割り振られる ID を組み合わせたタグを割り当てる。各データに一意のタグを割り当てることで、生成後にローカル環境でタグを元の値に復元してユーザに提示することが可能になる。マスキングを行った要約文の例を表 2 に示す。

ユーザデータベースには、ディプロマ・サブリメントから抽出したデータに加え、マスキング後の要約結果およびタグ情報を保存する。これにより、センシティブデータがマスキングされた個人特性情報をプロンプト生成に用いることができる。

3.2 名言データベース

データソース

偉人との対話生成を行うために、その偉人の考え方や理念を含む文を LLM にプロンプトとして入力することで、偉人との対話を生成する。本研究では、対象人物の理念、価値観、人生哲学、物事に対する考え方などが表現されている文および文章を名言と定義する。そして、対象の偉人の著作から名言を抽出し、RAG を構築するための名言データベースを構築する。本研究では、1913 年から 1945 年にかけて書かれた平生鈺三郎の日記をデータソースとして使用する。平生鈺三郎日記は、日付と本文で構成された全 18 巻からなり、合計 12,942 ページからなる。日記には平生鈺三郎の経験や、その日の新聞記事および講演の原稿などが含まれている。そこで、このように様々な情報を含むテキストデータから、平生鈺三郎が実際に発言した名言を抽出する。これらの日記は電子のテキストデータが存在しないため、OCR によりテキストの抽出を行う。このとき、日記本体のページ番号やヘッダーなどの要素により、正確に読み取れない場合がある。これらのエラーは人手により修正を行う。そのため人的コストの観点から、本研究では 1927 年から 1945 年にかけて書かれた第 9 巻から第 18 巻までの合計 10 巻を使用する。平生鈺三郎日記の例を図 2 に示す。この日記は戦前の日本語で書かれているため、旧字体と歴史的仮名遣いが用いられている。このような文を用いた対話が生成された場合、ユーザが解釈できない場合が考えられる。そこで、平生鈺三郎の日記を現代の日本語に変換し、名言データベースを構築するためのデータソースとして用いる。本研究では、OpenAI が提供する gpt-4o³ を用いて平生鈺三郎日記第 9 巻から第 18 巻の 117,000 文、5,337,764 単語を現代の日本語に翻訳した。その結果、158,590 文、6,048,866 単語となった。旧字体と歴史的仮名

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

表 3 名言抽出に用いるプロンプト

入力文が「名言」の定義に合致するかどうか判定してください。

- ・名言の必須条件

文章が以下の性質を明確に含んでいるか評価してください。

1. その文の主張や教訓が、特定の個人の経験や特定の文脈を越えて時代や文化を問わず多くの人々の生き方や社会活動に当てはまる。
2. その文が物事の本質を突く深い洞察や、人間や社会が「どうあるべきか」という規範的な考え方（人生哲学）を表現している。

- ・名言ではない条件

文が以下のいずれかの特徴に合致する場合名言ではない判定してください。

条件 1：特定の出来事や体験談。文脈依存性が高く、普遍性が欠けるため名言ではない。

条件 2：社会や制度の問題点を客観的に指摘・描写しているだけの文章。筆者の哲学ではなく、現状分析に留まるため名言ではない。

- ・出力形式

判定結果を名言の場合は 1、名言でない場合は 0 の数字のみで出力してください。他の文字（解説、句読点、空白など）は一切含めないでください。

遣いを用いて書かれた原文の意図や意味を完全に保持した現代語訳を行うことは困難である。そのため、時代背景を考慮する必要のある文章は翻訳によってニュアンスが変化する場合があるが、翻訳されたすべての文章を人手で検証することは困難である。そのため、当時の価値観やニュアンスを反映した現代語訳は今後の課題である。

名言の自動抽出

平生鈺三郎日記には、様々な内容の文が混在する。そのため、本研究では現代の日本語に変換した平生鈺三郎日記から名言のみを抽出する。本研究では、OpenAI が提供する gpt-4.1⁴を用いて自動的に名言を抽出する。名言抽出のために用いるプロンプトを表 3 に示す。本研究では名言の自動抽出手法の精度を、人手によるアノテーションで作成した正解データを用いて評価する。正解データは現代語に翻訳した平生鈺三郎日記からランダムに抽出した 500 文に対して、各文が名言かどうかを人手で判定することで作成する。本研究では、各文を 10 名のアノテータが判定し、6 名以上が名言であると判定した文を名言とする。この結果、500 文の内 147 文が名言であると判定された。正解データに対して、名言の自動抽出を行った結果、適合率が 0.51、再現率が 0.73 であった。適合率が 0.51 と低い原因は、名言の判定は文脈を考慮する必要がある点や、普遍的な教訓や考え方との区別が難しい点が考えられる。

現代の日本語に変換した平生鈺三郎日記の全文に対して名言の自動抽出を行った結果、158,590 文から 3,626 文の名言を抽出した。ここで、RAG を構築するために、名言を OpenAI が提供するテキスト埋め込みモデルである text-embedding-3-large⁵を用いてベクトル化する。これらの抽出した名言と、各名言の埋め込みベクトルを名言データベースに格納する。

4 : <https://developers.openai.com/api/docs/models/gpt-4.1>

5 : <https://platform.openai.com/docs/models/text-embedding-3-large>

表 4 対話生成に用いるプロンプト

あなたは甲南学園の創設者、平生鈺三郎です。

平生鈺三郎として、あなたの理念を反映した回答を出力してください。

以下のユーザ情報は、センシティブデータをマスキングした状態のデータです。このユーザ情報を参照して、ユーザの個人特性に応じた発話を行ってください。具体的な値は{本人情報_01}などの指定されたタグを使用して出力してください。

以下の名言は、あなたの理念や価値観、人生哲学、物事に対する考え方を反映した文です。この名言の内容を考慮して回答してください。

ユーザの入力に対する応答のみを出力してください。出力は 3 文程度で行ってください。

- ・参考にするユーザ情報

{profile_text}

- ・参考にする名言

{quotation}

- ・これまでの対話履歴

{history}

- ・ユーザの入力

{user_input}

出力：

4 対話生成手法

本研究では、ディプロマ・サプリメントから抽出したユーザの個人特性情報と、平生鈺三郎日記から抽出した名言を用いて、ユーザの個人情報特性を考慮した偉人との対話生成を行う。

4.1 名言検索手法

偉人との自然な対話を生成するためには、ユーザ発話に対して内容的に関連している名言を対話生成で用いる必要がある。そこで、text-embedding-3-large によるテキスト埋め込みとコサイン類似度による名言の検索を行う。また、本研究では OpenAI が提供する gpt-5⁶を用いてテキスト生成を行う。名言検索の手順を以下に示す。

1. ユーザの入力とマスキング済み個人特性情報を与え gpt-5 によって仮の応答を生成する。これは最適な名言を検索するためのクエリとする。
2. text-embedding-3-large によって、生成した仮の応答をベクトル化する。
3. 名言データベース内の全ての名言とのコサイン類似度を計算する。この結果、類似度が大きい上位 n 件の名言を検索結果とする。本研究では n=5 とする。

4.2 プロンプトの設計

対話生成に用いるプロンプトを表 4 に示す。プロンプトは人物の指定、発話生成のために考慮すべきデータの説明、ユーザ情報、検索結果の名言、対話の履歴、ユーザの入力から構成される。ユーザ情報は、ユーザデータベースに格納されているユーザの個人特性情報と、クラスごとに生成する要約文を入力

6 : <https://developers.openai.com/api/docs/models/gpt-5>

する。これらは、センシティブデータをマスキングしたデータを用いる。検索結果の名言は、コサイン類似度の大きい上位 n 件を入力する。本研究では $n=5$ とする。対話の履歴は、ユーザ発話とシステム発話の履歴を入力する。ここで入力する対話履歴の文は、センシティブデータをマスキングした状態のデータである。また、プロンプトではセンシティブデータの具体的な値を利用する必要がある場合、タグを用いて出力することを指示する。これにより、LLM はタグを用いた発話を生成する。その発話に含まれるタグはローカル環境で具体値に復元してユーザに提示する。これにより、クラウド型 LLM へセンシティブデータを入力せずに個人適応した対話生成を行う。

5 評価実験

提案する対話生成手法について、以下に示す 2 点について評価実験を行う。まず、名言を用いた対話生成手法によって平生鈺三郎の理念を含んだ対話を生成できるか評価する。次に、提案手法によってユーザのセンシティブデータを考慮した対話を生成できるか評価する。

実験方法

本研究で対象とする偉人は、甲南学園の創設者である平生鈺三郎である。

本実験では、マスキングの有無による個人特性情報の扱いと、要約文の入力の有無が対話品質に与える影響を比較するため、以下に示す 4 つの手法で比較する。

1. ベースライン：ユーザデータベースの中の個人特性情報を列挙したテキストを、プロンプトに入力して対話を生成する。
2. マスキング：センシティブデータをマスキングした個人特性情報を列挙したテキストを、プロンプトに入力して対話を生成する。
3. ベースライン+要約：個人特性情報を列挙したテキストを入力する。さらに、ユーザデータベースの中の要約文も入力して対話を生成する。
4. マスキング+要約 (提案手法)：センシティブデータをマスキングした、ユーザデータベースの中の個人特性情報と要約文をプロンプトに入力して対話を生成する。

どの手法においても表 4 に示すプロンプトを用いるが、参考にするユーザ情報として挿入する部分が相違点であり、それ以外の条件は同一とする。本実験において、対話生成に用いる LLM は gpt-5 である、評価者は平生の理念を理解している甲南大学の学生 14 名である。話題は指定せず、自由にモデルとの対話を行うように指示した。そのため、一つの話題について 5 ターン以上の対話を行っている場合もあれば、何度か話題が遷移しながら 5 ターン以上の対話を行っている場合もある。話題が遷移した場合においても適切な応答や、一貫性を保った対話ができるかを評価するため、このような実験設定とする。対話終了後、評価者は以下の評価項目について 4 段階 (1. 全くそう思わない, 2. そう思わない, 3. そう思う, 4. とてもそう思う) のリッカート尺度で評価を行う。また、評価者は実験の対話履

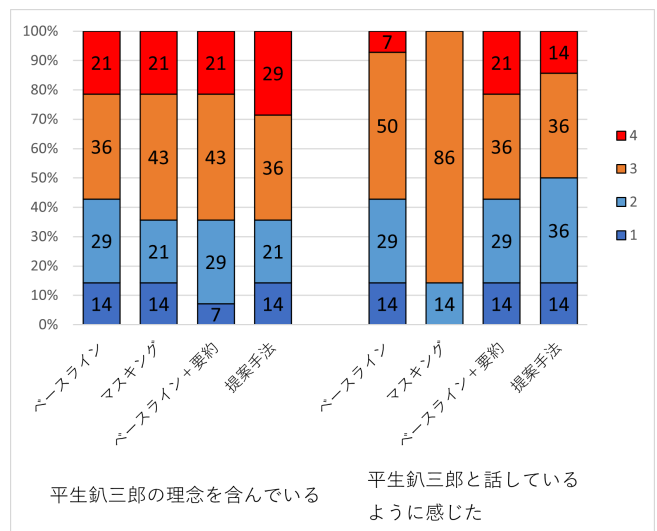


図 3 評価項目 1, 2 の結果

表 5 対話生成手法と評価項目ごとの実験結果の平均値

対話生成手法	1	2	3	4	5	6
ベースライン	2.64	2.50	2.71	3.00	3.57	3.36
マスキング	2.71	2.86	2.86	2.86	3.14	2.57
ベースライン+要約	2.79	2.64	2.93	3.29	3.79	3.36
マスキング+要約 (提案手法)	2.79	2.50	3.07	3.21	3.57	2.79

歴を参照しながら、感想や意見を自由記述形式で回答する。

1. 平生鈺三郎の理念を含んでいる。
2. 平生鈺三郎と話しているように感じた。
3. 個人特性を具体的に反映した対話である。
4. 個人特性と矛盾していない。
5. 対話の一貫性：内容に論理的な飛躍や矛盾がない。
6. 対話の自然さ：質問が多すぎたり、受け身すぎたりしない自然な対話である。

また、上記の表項目において、以下の 3 つのリサーチクエション (以下、RQ) を立てて考察を行う。

- RQ1 名言をプロンプトに入力する対話生成手法によって、偉人の理念を考慮した対話を生成できているか?(評価項目 1, 2)
- RQ2 提案手法によって、個人特性情報を考慮した対話を生成できているか?(評価項目 3, 4)
- RQ3 提案手法によって、マスキングを行う手法においても自然な対話を生成できているか?(評価項目 5, 6)

RQ1. 名言をプロンプトに入力する対話生成手法によって、偉人の理念を考慮した対話を生成できているか?

評価項目 1 および 2 の結果を、図 3 に示す。また、対話生成手法と各評価項目ごとの平均スコアを表 5 に示す。

評価項目 1 では、すべての対話生成手法で平均スコアが 2 を超えている。名言を用いた対話生成手法によって、平生鈺三郎の理念や考え方を含んだ発話が生成できたことが分かる。評価項目 2 では、すべての対話生成手法で平均スコアが 2 を超えている。一方、平生鈺三郎と話しているように感じないと評価された場合もある。評価項目 2 について、2 と評価された例を表

表 6 平生鈺三郎と話しているように感じないと評価された対話の例

話者	対話
system	甲南太郎さんは、多岐にわたる活動に取り組んできましたね。特に、体育会硬式野球部やボランティア活動などの経験は、将来の大きな力となるでしょう。
user	具体的にどんな役に立つと思いますか？
system	体育会硬式野球部で培ったリーダーシップを活かして、多様な価値観を尊重し、共生の精神を広めることが大切です。これらの活動を通じて、周囲の人々にも良い影響を与えることができるでしょう。

表 7 個人特性を考慮する必要のある対話の例

ベースライン	
user	部活で練習しても思うように上手くいかないんですけど、どうすればいいですか？
system	野球部での練習がうまくいかないと感じるときは、まず仲間と支え合い、共に成長する意識を持つことが大切です。先輩やコーチにアドバイスを求めることで、新たな視点を得られるかもしれません。
マスクング+要約 (提案手法)	
user	部活で練習しても思うように上手くいかないんですけど、どうすればいいですか？
system	練習が思うようにいかないときは、仲間との信義や気風を大切に、互いに助け合うことが重要です。

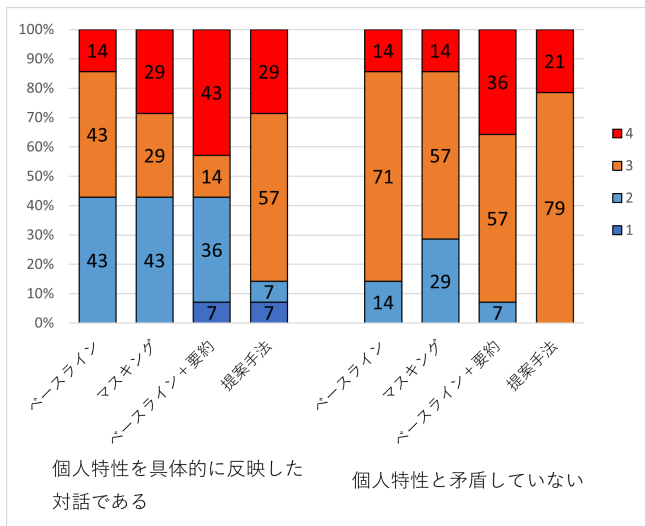


図 4 評価項目 3, 4 の結果

6 に示す。この例では、平生鈺三郎の考え方を含んでいると評価されたが、平生鈺三郎と話しているように感じていないと評価された。このように評価した被験者は自由記述回答で、「共生の精神」といった平生鈺三郎の理念に関連する言葉遣いが含まれているものの、そのほかの言葉遣いや語尾が一般的であると述べている。これは、平生鈺三郎日記を現代語に翻訳した際に、平生鈺三郎特有の使用語彙や文末表現が失われたためであると考えられる。

したがって、RQ1 に関しては名言を用いることで偉人の理念に沿った対話生成が可能である一方、偉人らしい文体の対話生成には課題があると言える。

RQ2. 提案手法によって、個人特性情報を考慮した対話を生成できているか？

評価項目 3 および 4 の結果を、図 4 に示す。評価項目 3 と 4 では、個人特性を反映した対話を生成できているか評価する。評価項目 3 では、どの対話生成手法でも平均スコアが 2 を超えた。また、提案手法では 86% が肯定的に評価しており、提案手法のみ平均スコアが 3 を超えた。さらに、マスクングの有無で比較するとマスクングを行う対話生成手法の方が平均スコアが大きい。これはセンシティブデータをタグとして与えることによって、LLM が対話で具体的な値を用いることが多いためであると考えられる。また、要約の有無で比較すると、要約文を

入力する対話生成手法の方が平均スコアが大きい。これは、モデルが参照すべき情報を要約文として入力することによって、対話に組み込むべき個人特性情報を利用しやすくなったためであると考えられる。また、GPA などの数値で扱う情報は、その数値の変化の特徴に着目した要約文を生成することによって、具体的な個人特性を含む対話を生成できたと考えられる。しかしながら、マスクングを行う対話生成手法では、個人特性情報を考慮して回答する必要がある発話に対して一般的な発話を生成するケースが多い。個人特性を考慮して回答する必要がある対話の例を表 7 に示す。マスクングを行っていないベースラインでは、ユーザが野球部であるという情報を考慮して、コーチなどにアドバイスを求めると良いという具体的な助言を生成できている。しかしながら、マスクングを行った対話生成手法では野球部での練習に対しての助言ではなく、一般的な助言を生成している。センシティブデータのマスクングにより LLM が具体的な値を参照できず、ユーザが野球部に所属しているという個人特性情報を利用できないためである。評価項目 4 では、マスクングのみの対話生成手法は平均スコアが 2.86 と低く、そのほかの対話生成手法は平均スコアが 3 以上である。特に、提案手法では 100% が肯定的に評価した。自由記述回答によると、個人特性情報に矛盾する発話はなかったが、個人特性情報を考慮して回答してほしい場面であっても一般的な回答が多く感じられたという指摘がある。これは、マスクングにより具体的な値の参照ができないため、個人特性情報を考慮しない一般的な発話が生成されたと考えられる。また、要約文を入力する対話生成手法では、マスクングの有無に関わらず平均スコアが大きくなっている。これは、要約文によってマスクングされたセンシティブデータをどのように利用すればよいか明確にすることができるためであると考えられる。

したがって、RQ2 に関してはクラスごとの要約文を入力することによって、個人特性情報と矛盾せず具体的な値を用いた対話を生成できたとと言える。しかしながら、個人特性情報を考慮して回答することが重要な質問に対しても一般的な内容の対話を生成してしまうという点が課題である。

RQ3. 提案手法によって、マスクングを行う手法においても自然な対話を生成できているか？

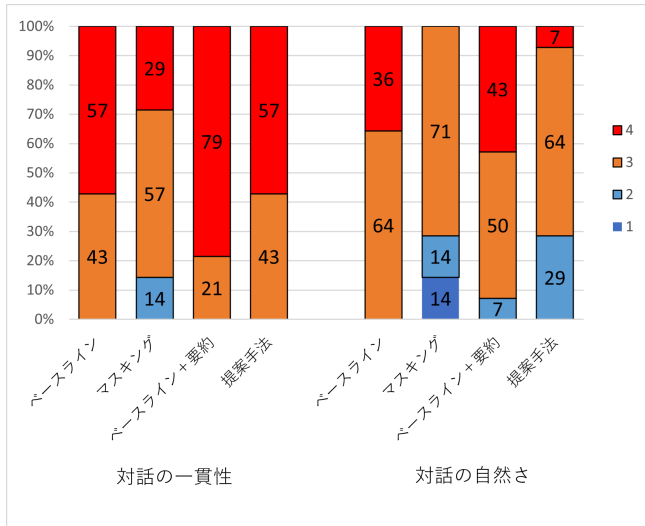


図5 評価項目5, 6の結果

表8 自然でないと評価された提案手法による対話の例

話者	対話
user	成績が思うように上がらなくて困っています
system	2.31 から 2.44 にかけて上昇傾向が見られます。甲南 太郎さんの努力には必ず成果があると信じて続けてみてください。

表9 ユーザの挨拶に対する発話の例

話者	対話
user	こんにちは
system	こんにちは、甲南太郎さん。あなたの学びや活動に対する情熱を感じます。独立自営の精神を育むことは、未来のリーダーとしての成長に繋がりますね。

評価項目5および6の結果を、図5に示す。評価項目5と6では、マスキングしたデータを対話生成に用いることで、対話生成の性能が低下しないか評価する。評価項目5では、どの対話生成手法も平均スコアが3を超えた。これは、ユーザの個人特性を考慮した対話であるため、個人適応した対話が生成されている点と、偉人の考え方を含む対話であるため一貫性のある対話であると評価されたと考えられる。評価項目6では、マスキングを行った対話生成手法は平均スコアが3を下回り、マスキングを行っていない対話生成手法は平均スコアが3を超えた。これは、センシティブデータのマスキングによって、個人特性情報が不自然に利用される場合があるためである。また、マスキングを行う対話生成手法では、「甲南 太郎さんが体育会硬式野球部や学内活動で担った役割の中で」のように人名や部活動の名称を常に正式名称で出力する。このため、発話が堅い印象になりやすい。また、提案手法で自然な対話ではないと評価された例を表8に示す。ユーザの成績が上がらないという相談に対してGPAが上昇していることを伝えているが、具体的な数値のみを出力しているため、ユーザがどんな意味の数値か理解できず不自然な対話であるという印象を与えている。このように、マスキングを行った場合はセンシティブデータが不自然に

利用される場合が多い。

また、ベースライン+要約の対話生成手法で自然な対話ではないと評価された例を表9に示す。この例では、ユーザが入力した単純な挨拶に対して、ユーザの個人特性と偉人の考え方を含めた発話を返している。本研究の手法では、一つのプロンプトにユーザの個人特性情報と5件の名言を含めている。それらの情報を考慮した対話を生成するという指示に従って対話を生成したため、挨拶という状況を見逃した対話が生成されたと考えられる。このことから、対話の状況に応じてRAGの適用を切り替えるなどといった仕組みが必要であると考えられる。

したがって、RQ3に関しては一貫性のある対話は生成できるが、センシティブデータの具体的な値を自然に利用した対話生成には課題があると言える。

6 まとめと今後の課題

本研究では、センシティブデータをクラウド型LLMに入力せずにユーザの個人特性を考慮した偉人との対話生成を行うための手法を提案した。具体的には、ユーザのセンシティブデータに対してマスキング処理を行い、オープンウェイトモデルのLLMで事前に生成する要約文を用いることで個人適応型対話生成を行う。また、偉人の著作から名言を抽出し、RAGを構築することによって偉人との対話を生成する手法を提案した。評価実験の結果、センシティブデータにマスキングを行っても個人特性情報を利用した対話が可能であることが分かったが、自然な対話の生成には課題があることが明らかになった。今後の課題は、偉人らしい言葉遣いを反映した対話生成や、質問や挨拶などユーザの入力に合わせた応答の生成である。

謝 辞

論文の一部はJSPS科研費24K03044、及び私学助成金（大学間連携研究助成金）、甲南デジタルツイン研究所の助成によるものである。ここに記して謹んで感謝の意を表する。

文 献

- [1] Sumit Asthana, Jane Im, Zhe Chen, and Nikola Banovic. "i know even if you don't tell me": Understanding users' privacy preferences regarding ai-based inferences of sensitive information for personalization. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [2] Tetsuya Kitahata, Kazuhiro Seki, and Akiyo Nadamoto. Speech-scenario generation based on the philosophy of a prominent leader within a small community. In Robert Wrembel, Gabriele Kotsis, A. Min Tjoa, and Ismail Khalil, editors, *Database and Expert Systems Applications*, pp. 291–306, Cham, 2026. Springer Nature Switzerland.
- [3] Tetsuya Kitahata, Kazuhiro Seki, and Akiyo Nadamoto. Two-stage fine-tuning for dialogue generation with small community prominent leaders' philosophies. In Eric Pardede, Qiang Ma, Gabriele Kotsis, Toshiyuki Amagasa, Akiyo Nadamoto, and Ismail Khalil, editors, *Information Integration and Web Intelligence*, pp. 427–442, Cham, 2026. Springer Nature Switzerland.

- [4] Joosung Lee, Minsik Oh, and Donghun Lee. P5: Plug-and-play persona prompting for personalized response selection. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 16571–16582, Singapore, December 2023. Association for Computational Linguistics.
- [5] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20*, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [6] OpenAI. gpt-oss-120b gpt-oss-20b model card, 2025.
- [7] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-LLM: A trainable agent for role-playing. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 13153–13187. Association for Computational Linguistics, 2023.
- [8] Lipeipei Sun, Tianzi Qin, Anran Hu, Jiale Zhang, Shuojia Lin, Jianyan Chen, Mona Ali, and Mirjana Prpa. Personal has entered the chat: Leveraging llms and ability-based framework for personas of people with complex needs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, New York, NY, USA, 2025. Association for Computing Machinery.
- [9] Yuanying Wang, Fuyong Xu, Yingzheng Zhu, Guangjin Wang, Peiyu Liu, and Ran Lu. Imdp: A unify dialogue framework with awareness and understanding for implicit personalized dialogue generation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, Vol. 24, No. 7, July 2025.
- [10] Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. A pre-training based personalized dialogue generation model with persona-sparse data. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 05, pp. 9693–9700, Apr. 2020.
- [11] Jijie Zhou, Eryue Xu, Yaoyao Wu, and Tianshi Li. Rescriber: Smaller-llm-powered user-led data minimization for llm-based chatbots. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, New York, NY, USA, 2025. Association for Computing Machinery.