

生成 AI による文書作成が出願書類の評価に及ぼす影響

——独自設計した語学系学部を想定した志望理由書の比較分析——

関 陽介 (神戸市外国語大学), 大塚 智子 (高知大学),
喜村 仁詞 (岡山県立大学), 上岡 麻衣子 (徳島大学)

志望理由書等の受験生が作成する出願書類は、志願者の適性や能力を評価する上で重要な資料であるが、これらの文書が生成 AI で作成される可能性が懸念される。そこで本研究では、生成 AI がどの程度の文書を生成できるかを明らかにするため、独自に語学系学部を設計し、志望理由書を対象に人間と ChatGPT が作成した文書を比較分析した。分析の結果、人間作成の志望理由書は体験重視の志望傾向が強く、ひらがなを多用した平易な文体が多かった。一方、ChatGPT 生成の志望理由書は研究開発重視の志望傾向が中心で、漢字・カタカナの比率が高く、多様な語彙を用いたより高度な文体で構成されていた。また、3人の評価者が12点満点で志望理由書を採点した結果、平均得点(標準偏差)は人間作成が6.65(2.33)、ChatGPT生成が10.52(1.72)となり、後者が有意に高い評価を得た。

キーワード：生成 AI, 大規模言語モデル, 志望理由書

1 はじめに

近年、Generative Pre-trained Transformer (GPT) や Large Language Model Meta AI (LLaMa) 等の大規模言語モデルが普及しており、これらを基盤とする多様なサービス(以降、生成 AI)が提供されている。たとえば、OpenAI の ChatGPT や Google の Gemini は、文書の生成や校正、画像認識、音声対話等が可能であり、学習支援や業務効率化に利用されている。また、生成 AI を活用したプログラミングのバグ検出 (Surameery and Shakor, 2022) や医学文献の作成補助 (Biswas, 2023)、患者用クリニックレターの作成 (Ali et al., 2023) 等の手法も報告されている。特に自然言語の生成精度は実用可能な水準に達しており、人間が作成したような自然な文書を作成することが可能である。

文部科学省は、大学の教学面における生成 AI の活用を促進するためにガイドラインを策定しており、教育や研究に生成 AI を活用する際の、留意点や推奨される取り組み等を公表している(文部科学省, 2023)。また、多くの大学が生成 AI の利用方針等を示しており、たとえば神戸市外国語大学では研究倫理上の問題や個人情報への扱い、誤回答が生成される危険性等の注意喚起を行っている(神戸市外国語大学, 2023)。また、教育現場への導入事例も報告されており、たとえば横浜国立大学の一部の講義では、授業設計やグループワークのファシリテーション準備、四択問題の誤答生成、記述型の試験問題の作成等に生成 AI が活用されている(横浜国立大学, 2024)。

志望理由書や活動報告書等の出願書類が求められる入学者選抜においても、生成 AI を活用することは可能である。このような状況を踏まえ、東京学芸大学や

山梨大学では、出願書類の作成に生成 AI を用いること、または出力文をそのまま使用することを禁止している(東京学芸大学, 2024; 山梨大学, 2024)。一方で、ChatGPT 等は誰でも利用できるため、生成 AI の利用を制限することは困難である。

生成 AI を用いて出願書類を作成する場合、加美山ら(2025)は、志願者層の学力によっては実力とかけ離れた質の高い成果物が生成される可能性がある、と述べている。本質的には高校教員や保護者等の介入と同じではあるが、生成 AI が出願書類の作成に活用できる程度を把握することは、入試設計を行う上でも重要である。そこで本研究では、生成 AI がどの程度の文書を生成できるかを明らかにする。具体的には、志望理由書を対象として、人間と生成 AI が作成した文書の比較分析を行う。なお、本稿が一般に公開されることを踏まえ、実在する大学が課す志望理由書の作成は避けることが望ましい。そこで本研究では、独自に語学系学部(以降、A学部)を設計し、当該学部に出願書類として提出することを想定した志望理由書の作成と採点を行う。

2 ChatGPT を用いた志望理由書の作成に向けて

2.1 評価に用いる学部概要の作成

本研究では知名度があり利用者数が多い ChatGPT を用いて志望理由書を作成する。図1に研究の流れを示す。①～④は ChatGPT を用いた作業で、① A 学部の概要設計、② アドミッション・ポリシー、③ 評価者が使用するルーブリック評価表、④ 志望理由書の作成を行う。⑤では①、②の学部概要やアドミッション・ポリシーを参考に協力者が志望理由書を作成し、⑥で

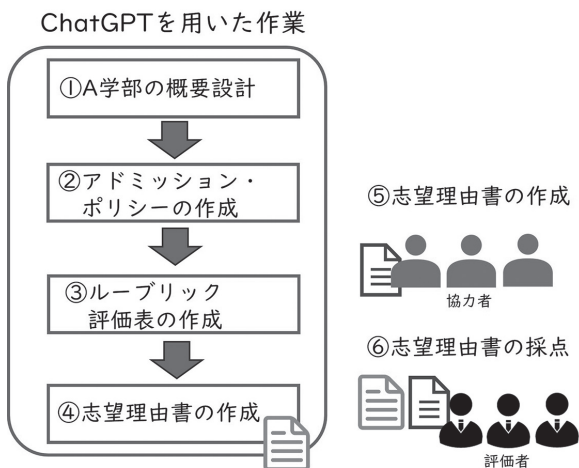


図1 研究の流れ

は評価者が④, ⑤で作成された文書を採点する。なお, ①~④で生成された文は原則として人手で修正せず, ChatGPT の出力をそのまま用いる。使用するモデルは①~③では日常的なタスクに適したChatGPT-4oを用い, ④では2025年4月16日に公開された, 執筆時点で最新かつ高度な推論が可能なChatGPT o3を用いる。

A学部の概要設計(図1の①)をするために, プロンプトに“新しく語学系学部を設計したいです。独自性ができるように学部の説明文を作成してください。たとえば, 教育の目的, 在学生のメッセージ, 進路, 特徴的なカリキュラム, 入試等の項目を作成してください。項目数や内容は一任します。”を入力して, Notion を用いてA学部のWebサイトを作成した。また, A学部の建物や各項目の挿絵, 在学生の写真もChatGPTで作成した。図2にA学部の概要を掲載したWebサイトの参考画面を示す。本サイトは⑤の協力者と⑥の評価者に限定公開しており, A学部の教育目的や学部概要, カリキュラムの特徴, 特徴的な研究紹介, 在学生のメッセージ, 卒業後の進路, 入試情報, 入学を検討される皆様へ, で構成される。参考までにカリキュラムの特徴と研究紹介の一部は以下になる。

【カリキュラムの特徴】

1・2年次: 幅広い教養科目や基礎的な言語科目を学習し, 多言語多文化への基本理解を深めます。留学や短期研修に参加する学生も多く, 早期から国際感覚を養います。3年次: ゼミに所属し, 専門分野を掘り下げた学習・研究を開始。少人数制のディスカッションや指導教員との密なコミュニケーションを通じて, 思考を深め, 論理的な文章力を培います。4年次: 卒業論文の作成を中心に, 自身の研究テーマをさらに高



A学部の教育目的

A学部は, 多言語多文化の実践的な学習を通じて, 高度な言語運用能力と多面的な文化理解を兼ね備えた人材を育成します。国内外の地域社会やグローバルなビジネスシーンで活躍する際に必要なコミュニケーション力, 問題解決能力, リーダーシップを身につけ, 地球規模の課題に貢献できる国際人の輩出を目的としています。

学部概要

- 多言語多文化の実践的学習環境

世界中の主要な言語と文化を学ぶカリキュラムを提供し, 学生はグローバル社会での実践的な言語能力を身につけます。言語学習だけでなく, その背景にある文化や社会を深く理解することで, 学生は異なる国や地域で活躍できるリーダーシップと協働力を養います。留学制度やインターンシップを通じて, 実際に現地での経験を積むことができ, 卒業後の国際的なキャリア形成にも有利です。

図2 A学部Webサイトの参考画面

度化。研究成果を活かし, 学内外でのプレゼンテーションや発表の機会が多数設けられています。

【特徴的な研究紹介】

オンライン教育を活用した遠隔言語学習プログラム
研究代表教員: 客殿 良助 (言語教育工学)

研究概要: ICTを活用し, 海外の教育機関と連携して共同授業を行う“バーチャル留学”プログラムを研究・開発しています。学生はオンライン上で海外の学生と語学学習や文化交流を行い, 互いに学び合う環境を整備。遠隔地とのリアルタイムコミュニケーションを通して, 世界の教育格差や言語教育の未来像を見据える研究成果を積み重ねています。

2.2 アドミッション・ポリシーとルーブリック評価表の作成

多くの大学では, 入学者に求める能力及びその評価基準・方法等を明確化したアドミッション・ポリシーを公表している。これには, 求める学生像や入学者の選抜方法, 重点評価項目, 評価する能力等が記載されている。A学部においても, 入試の透明性や公平性を担保し, 志望理由書の作成指針を示すために, 独自の

表1 ChatGPTを用いて作成したルーブリック評価表

評価項目	0点	1点	2点	3点
多文化への好奇心	多・異文化への興味が示されていない。	多・異文化に触れた経験はあるが、好奇心が明確に示されていない。	多・異文化に関心があり、具体的な行動や学びの例が示されている。	多・異文化や異なる価値観への強い興味が具体的に示され、尊重する姿勢が明確である。
実践的な語学力への関心	語学に関心がない、もしくは曖昧な表現。	語学学習への関心が表明されているが、具体的な目的や行動が不足している。	語学力向上への関心があり、具体的な学びの意欲が感じられる。	語学力を実際の場面で活用し、さらに高めたいという明確な目標と具体例が示されている。
国際的な視野	国際的な視野や興味が全く示されていない。	国際的な視野に興味があるものの、具体性に欠けている。	国際的な課題や問題に対する理解が示されており、具体的な例が示されている。	国際的な視野を持ち、具体的な活動や将来の目標に結びつけた具体例が豊富。
探究心と研究意欲	探究心や学問への意欲が全く感じられない。	探究心はあるが、研究テーマや具体的な学問への関心が薄い。	研究や探究への意欲があり、具体的なテーマが示されている。	自分のテーマを深く追求する姿勢と、それに向けた具体的な行動が明確に示されている。

アドミッション・ポリシーを作成することが望ましい。そこで、本研究では求める学生像に焦点を当てたアドミッション・ポリシーを作成する（図1の②）。具体的には、プロンプトとして①で作成した学部概要の説明文と“この文書から入試で用いるアドミッション・ポリシーの求める人物像を作成して箇条書きで4点に纏めてください。”を入力して、ChatGPTで以下の通り作成した。

- 多文化への好奇心が強い学生：異なる文化や価値観を尊重し、多様な視点から物事を考えることができる人
- 実践的な語学力を高めたい学生：外国語を使ったコミュニケーションやビジネスに強い興味を持ち、語学力を磨きたい人
- 国際的な視野を持つ学生：グローバルな課題に興味を持ち、国際社会での活躍を目指す人
- 探究心が旺盛な学生：学問に対する強い探究心を持ち、自分のテーマを深く追求したい人

本研究ではこの4点を志望理由書の採点項目を検討する際に使用する。なお、求める学生像はA学部Webサイトの入試情報の中に掲載する。

次に、評価者が志望理由書を採点するために、評価基準を設計する必要がある。一貫性があり客観性の高い採点を実施するため、ルーブリック評価表の導入事例が報告されている（井上ほか、2017）。そこで本研究においても、求める人物像に基づくルーブリック評価表を作成する（図1の③）。具体的には、“求める

人物像を対象に志望理由書を評価するためのルーブリック評価表を作成してください。評価項目数は4つで、各項目の点数は0～3点としてください。”をプロンプトとしてChatGPTで作成した。表1に作成したルーブリック表を示す。不自然な表現が含まれる可能性があるが、出力された評価表を使用する。

2.3 志望理由書の作成

プロンプトの内容は生成AIの回答内容に大きな影響を与える。そのため、完成度が高い志望理由書を作成するためには、ChatGPTに適切な指示を与える必要がある。効果的なプロンプトの作成手法は多く報告されており、たとえば具体的な指示や役割の付与がある。前者においては、プロンプトの内容が抽象的である場合、生成AIは指示や質問文の意図が理解できず、ユーザが期待した結果が得られない可能性があるというものである。そのため、質の高い回答を得るためには、背景や前提、意図等を具体的に明示することが望ましい。後者は、“あなたは優秀な人工知能の研究者です。”や“あなたは大学の数学科の教授です。”等の役割を与える方法である。生成AIが役割になりきることによって回答範囲が限定されて、回答精度が高められる。他にも多様な手法があるが、本研究ではこれら2点や他の手法を用いて志望理由書を作成（図1の④）する。

プロンプトの概要は、“A学部の概要に基づき志望理由書を400字程度で作成してください。なお、アドミッション・ポリシーの求める人物像は必ず参考に

表 2 各志望理由書の形式的な分析結果

	協力者	ChatGPT	p 値	効果量
総文字数	406.90 (67.28)	407.71 (24.11)	0.33	0.18(r)
漢字数	142.57 (31.08)	181.57 (13.35)	8.94×10^{-6} ***	0.80(r)
ひらがな数	201.76 (43.48)	125.67 (9.21)	1.29×10^{-7} ***	2.36(d)
カタカナ数	39.52 (19.75)	74.81 (21.09)	2.69×10^{-6} ***	-1.69(d)
句読点数	17.05 (6.53)	17.33 (1.58)	0.85	-0.06(d)
文数	7.10(2.04)	7.19 (0.85)	0.89	0.03(r)
1 文あたりの文字数	72.34 (71.38)	57.40 (6.68)	0.29	0.19(r)
総語彙数	111.60(23.53)	130.50(8.16)	9.47×10^{-5} ***	0.72(r)
異なり語彙数	74.54 (13.15)	101.35 (6.15)	6.35×10^{-10} ***	-2.59(d)
語彙の多様性	0.68 (0.06)	0.78 (0.03)	7.05×10^{-8} ***	-2.33(d)

*** $p < .001$, $|r| \geq 0.3$ または $|d| \geq 0.5$ が意味のある差とする。

してください。”とした。なお、前述した手法も用いているが、受験生への影響を考慮して、実際に使用したプロンプトの詳細は非公開とする。ChatGPT で作成した志望理由書の例文は以下になる。

私は幼少期をボリビアで過ごし、言葉の壁が友情を生む瞬間と誤解を招く瞬間の両方を体験しました。帰国後は移民児童の学習支援ボランティアを続け、通訳役として文化の懸け橋になることに挑戦しています。高校では多文化理解クラブを立ち上げ、年四回の交流イベントを企画しながら、多様な価値観を尊重する姿勢を養いました。A学部ではスペイン語に加え先住民語ケチュア語も修得し、留学と自治体インターンを通して多文化共生政策の実務を学びたいと考えています。佐藤教授のゼミでコミュニティ通訳と「やさしい日本語」の効果を検証し、卒論ではAIを活用した多言語発信プラットフォームを提案する予定です。また、統計学と公共政策も履修し、データに基づく施策立案力を身につけます。在学中はキャンパス内外の異文化プロジェクトを主導し、リーダーシップと実践力を磨きます。将来は地方行政の国際課で外国人住民と地域社会をつなぎ、誰一人取り残さない共生社会の実現に貢献したいです。

3 志望理由書の比較分析

協力者とChatGPTが作成した志望理由書の比較分析をするために、各文書の形式的分析(3.1節)と意味的分析(3.2節)、評価者による採点(3.3節)を行う。協力者はインターネット上で募集した大学生21人として、志望理由書の作成(図1の⑤)を依頼した。協力者の平均年齢(標準偏差)は20.57(1.09)で、事

前にA学部の概要を閲覧することや、アドミッション・ポリシーの求める人物像に基づき志望理由書を作成すること、生成AIを使用しないこと、400字程度で作成すること等を伝えた。また、比較対象としてChatGPTを用いて同件数になるよう21件の志望理由書を作成した。評価者は入試業務に従事する大学関係者3人で、計42件の志望理由書の採点(図1の⑥)を依頼した。採点順による影響を軽減するために、採点する書類には作成者の情報を伏せた上で、協力者とChatGPTによる志望理由書を交互に並べて配置した。

3.1 志望理由書の形式的分析

協力者とChatGPTが作成した志望理由書を比較分析した結果を述べる。まず、総文字数や語彙の多様性等の形式的分析を行った。表2に分析結果を示す。調査項目は10項目で、第2列と第3列には各項目の平均値(標準偏差)を示す。文数は句点の出現回数に基づき、総語彙数は名詞・動詞・形容詞を対象に算出し、語彙の多様性は異なり語彙数を総語彙数で割った値とする。検定にはShapiro-Wilk検定とLevene検定で正規性と等分散性を確認したのち、スチューデントのt検定、Welchのt検定、マン・ホイットニーのU検定のいずれかを用いた。第4列は有意確率、第5列は検定に対応する効果量(rank biserial rまたはCohen's d)を示す。効果量が $|r| \geq 0.3$ または $|d| \geq 0.5$ を満たす場合、意味のある差と解釈する。有意水準は0.01とするが、多重比較の影響を考慮してBonferroni補正を行い、補正後の有意水準0.001を用いた。

分析の結果、総文字数や句読点数、文数、1文あたりの文字数に有意差は認められなかった。ただし、漢字数やカタカナ数、総語彙数、異なり語彙数は

ChatGPT 生成の志望理由書が有意に多く、語彙の多様性も有意に高かった。一方で、ひらがな数は協力者作成の志望理由書が有意に多かった。また、これらの効果量は上記基準を満たしており、意味のある差であったと判断できる。以上より、ChatGPT 生成の志望理由書は漢字・カタカナの比率が高く、多様な語彙が使用される傾向が示唆された。たとえば、“在学中はアクセシビリティハッカソンを主催し、異文化チームを率いるリーダーシップを磨く”、“スペイン語と統計モデリングを修得し、映像制作技術によるリスクコミュニケーションも学ぶ”、“語彙選択と文化的背景知識が説得力を左右することを痛感した”等、難解な表現で高度な文体が多く見られた。一方で、協力者作成の志望理由書はひらがなが多く用いられ、たとえば“人と話すことがあまり好きではありませんでした。”、“オーストラリアの文化をもっと深く知りたいと思うようになりました”、“自分が当たり前と感じていたことでも他の国では当たり前ではないということが分かりました。”等、ChatGPT 生成の志望理由書と比較すると、読みやすい表現で平易な文体が多く見られた。なお、協力者の 1 文あたりの文字数は 72.34 (71.38) で標準偏差が大きくなったが、これは 1 件の志望理由書で句点が未使用であったことが原因と考えられる。該当文書を除いた場合の平均値 (標準偏差) は 55.59 (10.40) となり、ChatGPT による結果と近い値になった。

3.2 志望理由書の意味的分析

各志望理由書の意味的分析を行うために、特徴表現の可視化を行った。ここでいう特徴表現とは、文書の語彙的・意味的な情報を反映した数値ベクトルであり、文書間の意味的な類似度を計算する際に用いられるものである。まず前処理として、形態素解析器の MeCab (NEologd 辞書) を用いて、名詞・動詞・形容詞の抽出、非自立・接尾に分類される語やストップワード (こと、よう、する等) の除外をした。そして、志望理由書ごとに抽出した語を元の語順で連結したうえで、Bidirectional Encoder Representations from Transformers (BERT) に入力する処理を、書類の件数に応じて実施した。本研究では、Transformer ブロック数 12 層、隠れ層の次元数 768、自己注意ヘッド数 12、活性化関数に GELU を採用した事前学習済み日本語 BERT モデル (Tohoku NLP, 2020) を使用した。文書の特徴表現としては、Classification (CLS) トークンの 768 次元の出力ベクトルを使用しており、計 42 件の結果が得られた。なお、BERT は文脈を考

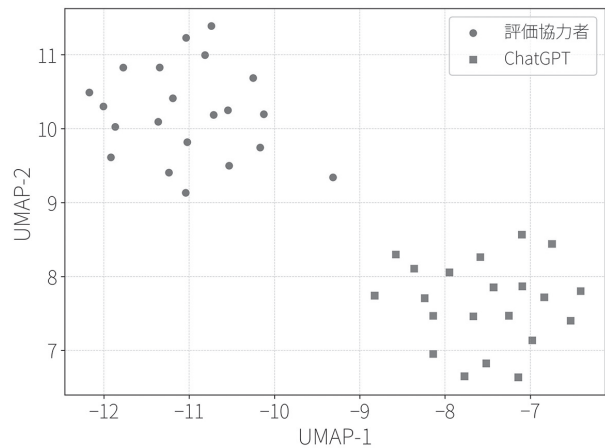


図 3 各志望理由書の特徴表現

慮して特徴表現を生成できるため、必ずしも助詞等を除外する必要はない。しかし、本研究では志望理由書間の意味的距離や分布傾向の比較に重点を置いている。文脈情報を詳細に保持するよりも、各文書に含まれる名詞や動詞等の内容語に基づく特徴を明確化するために、意味的寄与の小さい語を除外した。

次に、得られたベクトル群を Uniform Manifold Approximation and Projection (UMAP) で 2 次元に圧縮し、Matplotlib で散布図を作成した。なお、これらの処理は Google Colaboratory 上で Python を用いて行った。作成した散布図を図 3 に示す。丸が協力者、四角が ChatGPT による志望理由書の特徴表現で、それぞれ 21 個ある。この散布図では、近接する点ほど BERT の意味空間上で類似していることを示す。可視化の結果、特徴表現が協力者と ChatGPT で分離しており、2 つのクラスが形成されていた。形態素解析で助詞等を除外しているため、両志望理由書における文意に寄与する語の構成や意味的傾向が、大きく異なる可能性が示唆される。

どのような語が用いられているかを調査するために、各志望理由書の上位 15 件の頻出語と正規化頻度を調査した。正規化頻度は 0 ~ 1 の範囲をとる。表 3 に調査結果を示す。両文書において頻度に大差はないが、協力者作成の志望理由書には文化・留学・英語・コミュニケーション・海外等の国際体験に関する語や、私・考える・感じる・思う等の自己言及を表す語が多く出現し、実体験重視の志望傾向が強いことが考えられる。一方で、ChatGPT 生成の志望理由書には多言語・国際・通訳・インターン等の語学技術を示す語や、卒論・検証・教授・ゼミ・開発等の研究活動に関する語が多くを占めており、研究開発重視の志望傾向が強いことが考えられる。実際、各クラスターの重心に最も

近い代表文 5 件を調査した結果、協力者の代表文には“～した経験から”や“～に魅力を感じた”、“将来は～したい”等が用いられており、留学や国際交流等、過去の体験を起点に志望動機を語る記述が中心だった。一方で、ChatGPT の代表文は“研究に参画”や“プログラム開発”、“学習効果の統計的可視化”、“言説分析”等、今後取り組む具体的な研究手法や開発計画が詳細に述べられており、専門用語も多用されていた。したがって、前述した実体験重視の志望傾向と研究開発重視の志望傾向という特徴は代表文にも現れており、両志望理由書の記述傾向が異なることが確認できた。なお、本研究では各文書の特徴に明確な差異が見られたが、ChatGPT が生成する志望理由書はプロンプトの内容に大きく左右される。そのため、“高校 3 年生が書くような”等の条件を加えることで、生成される志望理由書の内容が大きく変化する可能性がある

表 3 上位頻出語

協力者		ChatGPT	
語	頻度	語	頻度
私	0.03	学ぶ	0.01
考える	0.02	多言語	0.01
学部	0.02	学部	0.01
文化	0.02	卒論	0.01
感じる	0.01	言語	0.01
経験	0.01	将来	0.01
留学	0.01	国際	0.01
学ぶ	0.01	修得	0.01
言語	0.01	考える	0.01
理解	0.01	検証	0.01
志望	0.01	教授	0.01
思う	0.01	ゼミ	0.01
英語	0.01	開発	0.01
コミュニケーション	0.01	通訳	0.01
海外	0.01	インターン	0.01

点に留意が必要である。

3.3 志望理由書の採点

3 人の評価者による志望理由書の採点結果を表 4 に示す。項目は、ループリック評価表の 4 つの評価項目に総合点を加えた 5 項目で、総合点は 4 つの評価項目（各 3 点）の総和（12 点満点）を示し、第 2 列と第 3 列には各項目の平均値（標準偏差）になる。検定には Shapiro-Wilk 検定で正規性を確認したのち、Wilcoxon 符号付き順位検定または対応のある t 検定を実施した。第 4 列は有意確率、第 5 列は対応する効果量（rank biserial r または Cohen's dz）を示す。効果量が $|r| \geq 0.3$ または $|dz| \geq 0.5$ を満たす場合、意味のある差と解釈する。有意水準は 0.01 とするが、多重比較の影響を考慮して Bonferroni 補正を行い、補正後の有意水準 0.002 を用いた。

結果として、多文化の好奇心を除くすべての項目で、ChatGPT が生成した志望理由書の得点が有意に高かった。これらの効果量は上記基準を満たしており、意味のある差であったと判断できる。総合点の平均値（標準偏差）は、協力者作成が 6.65(2.33) で、ChatGPT 生成が 10.52(1.72) であり、後者が有意に高くまた効果量が大きい。そのため、文書間の得点には顕著な差が確認された。そのため、本研究では協力者より ChatGPT 生成の志望理由書の方が高く評価される結果になった。

ループリック評価表の項目別では、多文化の好奇心の得点差は小さかった。これは協力者作成の志望理由書は実体験重視の志望傾向が強く、3.2 節で述べた通り文化・留学・英語等の多・異文化への関心を示す語が頻出していたことが、影響した可能性がある。実際、文中には“文化や価値観の違いを学ぶ”、“国際社会に貢献できる人材”、“異文化間の相互理解を促進”等の表現が記述されていた。この項目を除けば、ChatGPT 生成の志望理由書の得点が特に高く、研究開発重視の志望傾向が強いことで、とりわけ探究心と研究意欲に関して、より高い評価を獲得した可能性が考えられる。

表 4 評価者による採点結果

	協力者	ChatGPT	p 値	効果量
多文化への好奇心	2.03(0.76)	2.33(0.80)	0.02	-0.73(r)
実践的な語学力への関心	1.71(0.82)	2.38(0.79)	1.61×10^{-6} ***	-0.82(r)
国際的な視野	1.52(1.33)	2.86(0.43)	1.35×10^{-10} ***	-0.82(r)
探究心と研究意欲	1.38(0.58)	2.95(0.21)	2.92×10^{-12} ***	-0.87(r)
総合点	6.65(2.33)	10.52(1.72)	2.51×10^{-20} ***	-1.72(dz)

*** $p < .002$, $|r| \geq 0.3$ または $|dz| \geq 0.5$ が意味のある差とする。

各文書の作成者は伏せているが、採点後の評価者のコメントとして、協力者作成の志望理由書に対しては“私で始まり、志願がほぼ必ず出てくる。”や“将来は国際社会で活躍できる人になりたい等、私視点で漠然とした未来を語る文が多い。”、“使用される単語や文章の表現が平易。”といった意見があった。一方、ChatGPT生成の志望理由書には“難解な単語の羅列で読みにくい。字数内に目一杯詰め込んでいる。”や“内容が多岐にわたるため、ものすごく有望に見えるが、結局何がしたいのかわからない文書がある。”、“高校生としては現実離れした設定で、その完成度の高さから入学の必然性を感じないものもある”といった意見があった。各志望理由書の文体や志望傾向が異なることで、このようなコメントが得られた可能性が考えられる。

以上より、本研究では協力者よりChatGPTが作成した志望理由書が、高く評価される結果となった。生成AIは誰でも利用できるため、大学側が使用を制限したとしても、今後より多くの受験生が活用することが十分に予想される。生成AIの利用を検出するツールが存在するが、その精度は入試現場で求められる水準には達しておらず、現場への導入には課題が残る。さらに、本研究では生成AIのモデルにChatGPT o3を用いて志望理由書を作成したが、技術の進化に伴いより高性能なモデルが登場することは確実である。出願書類を点数化する評価方法では公平性の担保が難しくなるため、たとえば志望理由書は面接の参考資料にする等、生成AIの使用を前提とした入試設計が求められる。

3.4 本研究の限界と今後の課題

協力者とChatGPTが作成した志望理由書を分析したが、協力者は大学生であり、受験生が実際に出願した志望理由書との比較はできていない。また、生成AIで使用したプロンプトやルーブリック評価表の一般性は議論できておらず、独自設計した語学系学部のみを対象に調査したため、分析結果の解釈には注意が必要である。さらに、本研究では、被験者に提示した情報を学部概要やアドミッション・ポリシー等に限定していたため、大学の雰囲気や教育・研究環境等を踏まえた具体的な志望理由を検討することが難しい状況であった。そのため、一般的な回答や模範的な表現が得意な生成AIに、有利な条件となっていた可能性が考えられる。今後は、紹介動画や模擬面談、さらには実在する学部での実験等、より現実的な出願環境を再現した比較分析を実施する必要がある。

本研究では、プロンプトに具体的な指示や役割の付与等の手法を用いて、ChatGPTで志望理由書を作成した。しかし、プロンプトの手法は他にもあり、たとえば複数の具体例を提示することで精度を高めるFew-Shot Promptingがある。この手法を用いて、高く評価された過去の志望理由書をプロンプトに含めることで、より高得点が期待できる文書を生成できる可能性がある。また、ChatGPT以外にもGeminiやDeepSeek等、他の生成AIを用いることも可能である。さらには、生成AIに特定の知識を学習させるファインチューニングや、外部知識を活用するRetrieval-Augmented Generation (Lewis et al., 2020)等の手法もある。志望理由書の作成方法は多岐に渡り、手法により文書の質は大きく異なる可能性がある。そのため、他手法で作成した志望理由書との比較評価の実施が、今後の課題の一つにあげられる。

本研究ではChatGPTで作成した志望理由書をそのまま用いたが、実際の利用場面では人手による加筆・修正が加わる可能性が高い。たとえば、ChatGPT生成の志望理由書には“入学後はロンドンでインターシップを行い、アラビア語を修得したい”等、文脈上やや不自然で補足説明が必要と考えられる箇所がある。また、プロンプトに依存するが、生成AIの出力は画一的な内容になる傾向がある。このような問題に対して、人間の介入により自然で魅力的な表現に修正できるため、より質の高い文書を作成できる可能性がある。生成AIが出願書類の作成に与える影響を把握する上で、ChatGPT等の出力に人手が介入することで、どのような文書が作成されるかの調査も行う必要がある。

3.1節と3.2節の比較分析では、志望理由書の形式的かつ意味的分析を行った。しかし、前者では文字種の出現頻度等の表層的な特徴に限定しており、たとえば係り受け関係といった構造的な特徴は対象としていない。また、後者では各志望理由書の特徴表現に着目したが、文書間の違いをより深く調査するために、クラスタリングや内容分類、トピック分析等、様々な手法が存在する。今後はこのような手法を活用して、文書構造や内容の差異をより詳細に明らかにしたい。

4 おわりに

本研究では、入学者選抜における志望理由書が生成AIによって作成される可能性を懸念して、独自に語学系学部を設計し、生成AIがどの程度の文書を生成できるかを調査した。分析の結果、人間作成の志望理由書は体験重視の志望傾向が強く、ひらがなを多用し

た平易な文体が多かった。一方、ChatGPT 生成の志望理由書は研究開発重視の志望傾向が中心で、漢字・カタカナの比率が高く、多様な語彙を用いたより高度な文体で構成されていた。また、評価者が採点した結果、人間より ChatGPT 生成の志望理由書が有意に高い評価を得た。

各志望理由書の文体や志望傾向には顕著な差異が認められ、ChatGPT が生成したと容易に判定できる文書も含まれていた。しかし、本研究で用いたルーブリック評価表では、そのような文書であっても高得点を獲得する傾向が確認された。さらに、プロンプトを調整すれば、人間が書いたような文体を表現できるため、文書の作成主体を判別することは一層困難になる。また、ChatGPT の出力に人手で修正を加えることで、得点がさらに向上する可能性もある。従って、志望理由書の点数化には慎重な判断が求められ、たとえば面接時の参考資料とする、当日にテーマを与えてその場で作成させる、活動報告書のように根拠資料がない書類は点数化を避ける、等の方針を検討する必要がある。

本研究では、出願書類を生成 AI の作成対象としたが、提出された書類を生成 AI で採点することも、技術的に可能である。ただし、生成 AI に評価を一任する場合、プライバシー侵害や情報漏洩、評価の公平性・透明性といった課題が懸念される。そのため、採点結果はあくまで参考情報として最終判断は人間が行う等、運用面の検討も進める必要がある。今後は、評価作業の支援を目指して、技術的課題と運用上の課題を整理したうえで、その実現可能性を検討したい。

謝辞

本研究はJSPS 科研費 24K16757 の助成を受けたものです。

参考文献

- Ali, S. R., Dobbs, T. D., Hutchings, H. A. and Whitaker, I. S. (2023). "Using ChatGPT to write patient clinic letters," *The Lancet Digital Health*, 5(4), e179-e181.
- Biswas, S. (2023). "ChatGPT and the Future of Medical Writing," *Radiology*, 307(2), 1-3.
- 井上敏憲, 中村裕行, 前村哲史, 植野美彦, 立岡裕士, 岡本崇宅, 大塚智子 (2017). 「四国地区国立 5 大学共通のインターネット出願と多面的・総合的評価への取り組み」『大学入試研究ジャーナル』 **27**, 91-96.
- 加美山若奈・倉元直樹 (2025). 「志願者本人記載書類作成に関する文章生成 AI の影響懸念—高校教員への質問紙調査から—」『大学入試研究ジャーナル』 **35**,

177-184.

- 神戸市外国語大学 (2023 年 4 月 7 日). 「学生の皆さんへ: 生成系 AI の使用についての注意 (第 1 報)」神戸市外国語大学 <https://www.kobe-cufs.ac.jp/news/2023/22365.html> (2025 年 6 月 26 日)
- Lewis, P. et al (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *NeurIPS 2020*, 33, 9459-9474.
- 文部科学省 (2023 年 7 月 13 日). 「大学・高専における生成 AI の教学面の取扱いについて【概要】」文部科学省 <https://www.mext.go.jp/kaigisiryoy/content/000245316.pdf> (2025 年 6 月 18 日)
- 関 陽介 (2025). 「用例ベースと生成 AI を併用したハイブリッド対話システム」『情報処理学会論文誌 デジタルプラクティス』 **6**(1), 1-11.
- Surameery, N. M. S. and Shakor, M. Y. (2022). "Use ChatGPT to Solve Programming Bugs," *International Journal of Information technology and Computer Engineering*, 3(1), 17-22.
- Tohoku NLP (2020 年 8 月 31 日). 「bert-base-japanese-v2」Hugging Face Hub <https://huggingface.co/tohoku-nlp/bert-base-japanese-v2> (2025 年 10 月 22 日)
- 東京学芸大学 (2024 年 5 月 13 日). 「出願書類作成における生成 AI の利用について」東京学芸大学入試課 https://www.u-gakugei.ac.jp/nyushi/upload/20240513_seiseiai.pdf (2025 年 6 月 18 日)
- 山梨大学 (2024 年 7 月 24 日). 「入学者選抜における生成 AI の取り扱いについて」山梨大学 <https://www.yamanashi.ac.jp/examination/49103> (2025 年 6 月 18 日)
- 横浜国立大学 (2024 年 3 月 22 日). 「生成 AI の活用に関するグッドプラクティス集」横浜国立大学 https://www.yec.ynu.ac.jp/about/pdf/ai_good_practice.pdf (2025 年 6 月 18 日)