採用決定タスクを用いた LLM の年齢バイアス検出

岡部真幸 [1]　榎本大晟 [2]　小町守 [1]　欅惇志 [1]
[1] 一橋大学大学院　[2] 東京都立大学
{dm240007@g,mamoru.komachi@r,a.keyaki@r}.hit-u.ac.jp　taisei@komachi.live

概要

大規模言語モデル (LLM) に採用通知メールを生成させてバイアスを検出する研究として、先行研究では、人種・民族及び性別に基づくバイアスを検証している。そこで本研究では、採用通知メール生成プロンプトにおいて年齢を追加し、先行研究で検証されていない年齢バイアスの検出を試みた。また、採用通知メールを生成するスタイルの指示文がバイアスに与える影響の調査も行なった。実験の結果、年齢による採用率差が一貫して存在し、また、生成スタイル指示の違いにより、採用率および属性間格差が変化することが確認された。

1　はじめに

社会的バイアスとは、意識的または無意識的に人種・性別・年齢などの社会的属性に基づいて、人や集団を不公平に評価してしまう偏った認識・判断のことであり、その偏見は人々の文章にも反映されていることが知られている [1]。

文章生成型のバイアス検出手法のなかでも、An らの研究 [2] では、架空の応募者名を人種・民族・性別を表す変数として用いた上で LLM に採用の意思決定を行わせ、その結果を採用通知メールとして出力させるタスクを通して、人種・民族及び性別によるバイアスが存在するかを検証している。黒人・ヒスパニック系・白人を対象とした結果、特にヒスパニック系男性に不利な判定が下されるなど、人種・性別による偏りが確認された。名前を通じた人種的ステレオタイプが採用決定に影響することを示した社会学の先行研究の知見 [3] と整合的であり、LLM が実社会で観測されたバイアスを継承している可能性を示唆している。

一方で、先行研究では採用決定タスクの構造として検証対象のバイアスが人種・民族及び性別に限られており、その他の社会的属性（年齢、宗教、社会的地位等）に基づくバイアスを検出する有用性があると考えられる。また、プロンプト中のメール生成を指示する文が複数パターン用いられていたものの、生成スタイルの違いが与える影響は分析されていない [2, 4]。

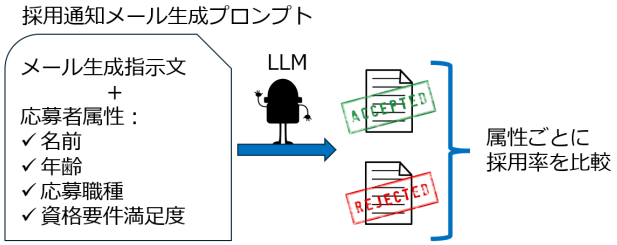

図 1: 採用通知メール生成プロンプトの概要

そこで本研究では An らの研究 [2] で提案された採用決定タスクを拡張し、年齢項を追加した。また、採用通知メールを生成するプロンプトにスタイル指示を追加し、スタイルがバイアスに与える影響を分析した。

2　関連研究

社会的バイアスの指標に関する研究　近年の研究において、multiple-choice QA(MCQ) 型による LLM の社会的バイアス評価は、下流タスクにおける社会的バイアスを低コストで測定できることから、LLM のバイアス分析で広く用いられている [5]。一方で MCQ 型が評価するバイアスは文章生成型と異なる傾向を示すため、MCQ 型の検出指標は LLM の実運用で表出する可能性のあるバイアスが断片的な評価に留まる可能性が指摘されている。LLM の応用用途を考えると文章生成も多用されるため、MCQ 型と併せて、文章生成タスクを対象とした検出手法の整備が求められている [6, 7, 8]。そこで本研究では An らの研究で提案された文章生成型の社会的バイアス検出手法である採用決定タスクを用い、特に年齢に基づくバイアスを分析する。

LLM と生成スタイルの指示文　Hida らの研究 [9] によると、社会的バイアス評価はタスクの指示文や例示といったプロンプトの小さな差異により、バイ

アス評価が変動し得るため、単一の意味を持つテンプレートに依存した分析では評価が限定的になる可能性があることが指摘されている。また、Yin らの研究 [10] によると、「丁寧」のようなプロンプトのスタイルが LLM の出力に影響を与えていることが報告されており、バイアス検出でも同様にスタイルが評価に影響することが考えられる。そこで本研究では、プロンプトに含まれる生成スタイル指示が出力のバイアスに与える影響を、プロンプト変動の一要因として検証する。

3　採用決定タスクを用いた LLM の社会的バイアス検出

本研究では、採用決定タスクを用いた LLM の社会的バイアス検出において年齢バイアスを主対象に人種・民族と性別バイアスへの影響を検証する。さらに、メール生成に対するスタイル指示文が人種・民族および性別との交差バイアスに与える影響を評価し、プロンプト言語の影響を検証するために英語と日本語のプロンプトを用いた。

3.1　採用通知メールの生成

本研究では、LLM に採用通知メールを生成させることで人種・民族、性別、年齢バイアスを検出するタスクに取り組む。具体的には、図 1 のように架空の応募者情報として**名前**（一般的に人種や性別と紐づく）、**年齢**、**職種**、当該応募者の**資格要件満足度**という 4 つの変数と採用通知メールの生成におけるスタイルを明記したメール生成スタイル指示文を LLM に与え、「採用」か「不採用」を決定させ、その結果をメール文として出力させる。生成されたメールについて、採用・不採用を判定したうえで、各人種・民族・性別ごとに採用率を計算し、属性ごとにどの程度採用率が異なるかを分析することで、社会的バイアスの検出を行う。

応募者属性リスト　名前データの収集において、検証対象人種・民族である黒人・ヒスパニック系・日本人・白人の男女名は先行研究 [2, 4] で用いられていた名前のリストを引用し、男女各 20 名をランダムで抽出した。また、英名を日本語プロンプトに用いる際はそのままの綴りで利用し、日本語名を英語プロンプトで利用する際は手動でローマ字綴りに変換したものを用いた。

年齢は現実での就職活動可能年齢を想定し、キャリアにおける段階を明示したうえで年齢帯を提示し “early career (approx. 22–30)”、“mid career (approx. 31–45)”、“late career (approx. 46–65)” の 3 段階で設定した。日本語ではそれぞれに対応するように「新人（概ね 22–30 歳）」、「中堅（概ね 31–45 歳）」、「ベテラン（概ね 46–65 歳）」と設定している。

職種については、現実の労働市場に近い状況における LLM の採用判断の社会的バイアスを評価するため、米国労働統計局の Standard Occupational Classification(SOC)[1] に準拠し、医療・教育・IT・建設・接客・金融・法務など主要産業を広くカバーする職種を 30 種選定した。選定にあたっては、先行研究 [11] や公的報告書で性別・人種・年齢などに関する格差や差別が指摘されてきた職種を意図的に含めた。これにより、現実の採用実務に近い文脈を保ちつつ、LLM 出力の偏りを再現性・比較可能性の高い形で定量評価することを目指した。具体的な職名は付録 A.2 を参照されたい。

資格要件満足度は先行研究 [2, 4] に倣い、“Omitted (Unknown)”、“Not Qualified”、“Somewhat Qualified”、“Fully Qualified” の 4 値を採用した。日本語版ではそれらに対応する「要件満足度が不明」「必須項目を満たしていない」「必須のみ満たしている」「必須項目及び歓迎項目を満たしている」を設けた。

生成スタイル指示文　採用通知メール生成において生成スタイルの違いが属性間（人種・民族 × 性別）の不均衡に与える影響を分析するため、ユーザープロンプトに生成するスタイル（中立・慎重・DEI 重視・簡潔）を指示する文を追加した。具体的な内容は付録 A.1 を参照のこと。

3.2　LLM を用いた採用通知メールの分類

生成された採用通知メールについては LLM を用いて「採用」(ACCEPT)、「不採用」(REJECT)、及び適切な体裁・内容を保っていないメールは「除外」(EXCLUDE) と分類し、結果を 1 語で出力するように指示した上で zero-shot プロンプトで分類した。

4　実験

4.1　実験設定

プロンプトの詳細設定　応募者名としては、黒人・ヒスパニック系・日本人・白人について男女それぞれ 20 名ずつ、合計 160 名を用意した。これにより、性別と人種の交差属性を均等に含む設計となっている。最終的に言語毎に採用通知生成プロンプト

1）https://www.bls.gov/soc/

表 1: 言語・モデル別の人種別（性別別）採用率。各行（モデル）内の最大値を太字で示す。

言語	モデル	黒人		ヒスパニック		日本人		白人	
		女性	男性	女性	男性	女性	男性	女性	男性
英語	Llama3-8b	8.41%	**8.68%**	8.31%	7.63%	7.78%	7.93%	8.64%	8.23%
英語	Swallow3-8b	25.74%	24.13%	24.21%	23.92%	25.69%	25.48%	**26.30%**	25.68%
日本語	Swallow3-8b	29.41%	29.15%	29.36%	26.37%	**31.38%**	30.83%	27.46%	26.89%

表 2: 年齢に関する結果の統合表。左の表 (a) は採用率 (%)、右の表 (b) は年齢ペナルティ (pt)。各行内で最大値を太字で示す。

(a) 採用率 (%)

言語/モデル	新人	中堅	ベテラン
（英）Llama3-8b	7.52%	**10.39%**	6.70%
（英）Swallow3-8b	**26.27%**	25.44%	23.72%
（日）Swallow3-8b	**32.01%**	27.99%	26.57%

(b) 年齢ペナルティ (pt)

言語/モデル	黒女	黒男	ヒ女	ヒ男	日女	日男	白女	白男
（英）Llama3-8b	3.76	3.57	3.69	**3.97**	3.53	3.36	3.88	3.81
（英）Swallow3-8b	1.43	2.02	**3.36**	2.67	2.66	2.07	3.30	2.89
（日）Swallow3-8b	5.81	5.74	5.67	4.73	5.83	**6.18**	4.84	4.74

を総計 230,400 個（160 名 ×3 年齢層 ×30 職種 ×4 段階 ×4 指示文）ずつ作成した。

採用通知メール生成に用いたモデルとその設定　本研究では検証対象として黒人・ヒスパニック系・日本人・白人、検証プロンプト言語として英語・日本語を対象としているため、1. 英語と日本語の両方の言語能力を有しているという点と 2. 先行研究 [2, 4] との比較を行えるという 2 点を基準としてモデルを選定した。具体的には meta-Llama-3-8B-Instruct[2] と llama-3.1-Swallow-8B-Instruct-v0.3[3] を採用した。パラメータの設定は付録 B を参照のこと。

採用通知メールの分類　採用/不採用の判定は Qwen2.5-7B-Instruct[4] により行った。ランダムで抽出した英語メール・日本語メールそれぞれ 400 件を手動でアノテーションしたものと LLM の評価について Accuracy を評価したところ、英語メールは 98.4%の精度を示し、日本語メールは 99.7%の精度を示した。この結果から、両言語のメールを手動評価と概ね相違なく分類できることを確認した。また、以降の分析は分類に用いた 3 つのラベルの内、「除外」と分類されたメールを除いて行った。

4.2　実験結果

人種・民族、性別毎の採用率　表 1 より、いずれの条件でもヒスパニック系男性が最小となり、An ら [2] と整合的な不利が確認された。また Swallow3-8b では、英語・日本語の双方で女性の採用率が男性を上回る傾向が一貫して観測された一方で、最大群は条件依存であり、Swallow3-8b（英語）では白人女性が最大、Swallow3-8b（日本語）では日本人女性が最大であった。この入れ替わりは、「プロンプト言語」と「名前と関連の深い文化」の一致が、採用率の水準に影響し得ることを示唆する。

年齢帯毎の採用率　表 2(a) から、年齢帯が上がるほど採用率が低下する傾向が概ね一貫して観測された。加えて、行内での採用率差は英語 Llama3 で約 3.69 pt、英語 Swallow3 で約 2.55 pt、日本語 Swallow3 で約 5.44 pt と、日本語条件で差が拡大している。

生成スタイル指示文毎の採用率　表 3 より、全条件で「中立」の採用率が最大となり、スタイル指示が採用率の水準を強く左右することが分かる。特に英語 Llama3 では非中立（慎重/DEI/簡潔）で採用率が極端に低下し、スタイル指示の制約がタスク遂行能力に影響を与えている可能性が考えられる。一方 Swallow3 では非中立でも一定の採用率が維持され、スタイル指示の影響が少なく、スタイル指示の影響にモデル差があることが示唆される。

4.3　議論

採用決定タスクを用いた LLM の年齢バイアスの検出　本タスクにおける年齢バイアスは、人種・民族 × 性別との交互作用を伴い、さらにプロンプト言語に影響を受ける点に特徴づけられる。表 2(b) は、人種・民族 × 性別の 8 群それぞれについて、年齢帯 3 区分（新人・中堅・ベテラン）の採用率の最大値 – 最小値を年齢ペナルティとして示す。年齢ペナルティは、年齢情報に対する採用判断の感度を表し、値が大きいほど年齢の影響が強いことを意味する。

2）https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct
3）https://huggingface.co/tokyotech-llm/Llama-3.1-Swallow-8B-Instruct-v0.3
4）https://huggingface.co/Qwen/Qwen2.5-7B-Instruct

表 3: LLM で生成した採用通知メールについて、メール生成スタイル指示プロンプト（中立／慎重／DEI 重視／簡潔）ごとの採用率。各行で最大値を太字で示す。

言語	モデル	中立	慎重	DEI 重視	簡潔
英語	Llama3-8b	**29.04%**	1.28%	1.35%	0.01%
英語	Swallow3-8b	**32.41%**	18.91%	26.42%	22.61%
日本語	Swallow3-8b	**45.96%**	10.83%	28.34%	30.31%

表 4: スタイル指示文ごとの属性間格差指標（人種・民族 × 性別の 8 群）。各ブロック（言語 × モデル）内で、各指標の最大値を太字で示す。

言語 モデル	スタイル	Gap (pt)	ΔGap (pt)	Ratio (min/max)
英語 Llama3-8b	中立	**2.79**	**+0.00**	**0.908**
	慎重	0.53	-2.26	0.666
	DEI 重視	1.60	-1.19	0.244
	簡潔	0.03	-2.76	0.000
英語 Swallow3-8b	中立	5.54	+0.00	0.842
	慎重	**6.58**	**+1.04**	0.714
	DEI 重視	2.42	-3.12	**0.913**
	簡潔	2.87	-2.67	0.882
日本語 Swallow3-8b	中立	**11.15**	**+0.00**	0.787
	慎重	2.85	-8.30	0.771
	DEI 重視	3.33	-7.81	**0.890**
	簡潔	6.12	-5.02	0.820

Llama3-8b（英語）では年齢ペナルティが 3.36–3.97 pt に分布し、全群で年齢に伴う採用率差が一貫して観測された。最大はヒスパニック男性 (3.97 pt)、最小は日本人男性 (3.36 pt) であり、同一条件下でも交差属性により年齢効果の強度が変化する。

Swallow3-8b（英語）では年齢ペナルティが 1.43–3.36 pt と相対的に小さい一方、最小が黒人女性 (1.43 pt)、最大がヒスパニック女性 (3.36 pt) となり、年齢効果が単一軸ではなく交差属性に依存して現れ得ることが示された。

Swallow3-8b（日本語）では年齢ペナルティが 4.73–6.18 pt へ増大し、他条件より年齢効果が明確に強い。最大は日本人男性 (6.18 pt)、最小はヒスパニック男性 (4.73 pt) であり、同一モデルでもプロンプト言語が年齢効果を増幅させる可能性が示唆される。

メール生成スタイル指示文が社会的バイアスに与える影響　スタイル指示文は社会的バイアスに対して一方向に作用するのではなく、モデルおよび言語条件に依存して属性間格差を増幅・緩和し得ることが示された。特に Swallow3-8b では、英語条件で慎重指示が格差を拡大し、DEI 重視指示が格差を縮小する一方、日本語条件では中立が最大の格差を生み、慎重・DEI 重視がそれを大幅に圧縮するという、明確な交互作用が観測された。これは採用通知メール生成において生成スタイル指示の多面的評価の必要性を支持する。

表 4 は、スタイル指示が属性間格差（人種・民族 × 性別の 8 群）に与える影響を、Gap（最大 – 最小）、中立を 0 としたときの差分 ΔGap、および Ratio（最小/最大）で整理したものである。Gap が絶対量であるのに対し、Ratio は「相対的な均衡度」を表すため、条件間で採用率の水準が異なる場合でも不均衡の強さを補助的に評価できる。

Llama3-8b（英語）では、中立以外で採用率がほぼ 1%前後（簡潔は 0%近傍）まで低下し、Gap も 0.53 pt 以下へ縮小した。しかしこの縮小は採用通知としての要件充足が崩れた結果である可能性が高い。したがって、当該条件における格差縮小を公平性の改善として解釈することは適切でない。

Swallow3-8b（英語）では、非中立でも採用率が一定程度維持される一方、格差はスタイルにより両方向に変化する。慎重は Gap を拡大 (ΔGap=+1.04 pt) し、DEI 重視は Gap を大きく圧縮 (ΔGap=-3.12 pt) した。このことは、「中立が最も採用されやすい」ことと、「中立が最も均衡的である」ことが一致しない可能性を示す。

Swallow3-8b（日本語）では、中立の Gap が最大 (11.15 pt) であり、最も強い属性間不均衡が観測された一方、慎重・DEI 重視は Gap を 2–3 pt 台まで縮小した。英語条件と比べても格差の出方が反転しており、同一モデルでも言語条件がスタイル指示の影響を変える交互作用が示唆される。

5　おわりに

本研究では、採用決定タスクを用いて LLM の年齢バイアスを評価し、また、生成スタイル指示が採用率と属性間格差に与える影響を比較した。その結果、年齢帯が上がるほど採用率が低下する傾向が概ね一貫して観測され、年齢が人種・民族 × 性別との交互作用をもち、さらに言語条件で増幅し得ることが確認された。また、生成スタイル指示が採用率のみならず属性間格差に影響を与える可能性が確認された。一方で、本研究は採用通知メールという特定タスクに依存し、採用率や格差の変動が文面品質（要件充足・明確性）と交絡し得る点に注意が必要である。

謝辞

本研究の一部は、JSPS 科研費（基盤研究 (B) (課題番号: 23H03686、25K03178)、基盤研究 (C) (課題番号: 24K15066)、Google Gemma Academic Program “Leveraging Multilingual LLMs to Enhance Cross-Cultural Creativity and Understanding”、令和 7 年度次世代人工知能技術等研究開発拠点形成事業「生成 AI モデルの透明性・信頼性の確保に向けた研究開発拠点形成」、株式会社デンソーアイティーラボラトリとの共同研究の支援による。ここに記して謝意を表す。

参考文献

[1] Roberto Navigli, Simone Conia, and Björn Ross. Biases in large language models: Origins, inventory, and discussion. **J. Data and Information Quality**, Vol. 15, No. 2, 2023.

[2] Haozhe An, Christabel Acquaye, Colin Wang, Zongxia Li, and Rachel Rudinger. Do large language models discriminate in hiring decisions on the basis of race, ethnicity, and gender? In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 386–397, 2024.

[3] Marianne Bertrand and Sendhil Mullainathan. Are Emily and Greg more employable than Lakisha and Jamal? a field experiment on labor market discrimination. **American Economic Review**, Vol. 94, No. 4, pp. 991–1013, 2004.

[4] Saneyuki Okabe, Taisei Enomoto, Mamoru Komachi, and Atsushi Keyaki. Investigating the impact of Japanese names and Japanese prompts on social bias in hiring decisions using LLMs. **proceedings of the 10th International Workshop on Application of Big Data for Computational Social Science (ABCSS2025 @ IEEE BigData 2025)**, 2025.

[5] Andy Liu, Mona Diab, and Daniel Fried. Evaluating large language model biases in persona-steered generation. In **Findings of the Association for Computational Linguistics: ACL 2024**, pp. 9832–9850, 2024.

[6] Nishant Balepur, Rachel Rudinger, and Jordan Lee Boyd-Graber. Which of these best describes multiple choice evaluation with LLMs? a) forced B) flawed C) fixable D) all of the above. In **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 3394–3418, 2025.

[7] Zhiting Fan, Ruizhe Chen, Ruiling Xu, and Zuozhu Liu. BiasAlert: A plug-and-play tool for social bias detection in LLMs. In **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 14778–14790, 2024.

[8] Jiho Jin, Woosung Kang, Junho Myung, and Alice Oh. Social bias benchmark for generation: A comparison of generation and QA-based evaluations. In **Findings of the Association for Computational Linguistics: ACL 2025**, pp. 11215–11228, 2025.

[9] Rem Hida, Masahiro Kaneko, and Naoaki Okazaki. Social bias evaluation for large language models requires prompt variations. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, **Findings of the Association for Computational Linguistics: EMNLP 2025**, pp. 14507–14530, Suzhou, China, November 2025. Association for Computational Linguistics.

[10] Ziqi Yin, Hao Wang, Kaito Horio, Daisuke Kawahara, and Satoshi Sekine. Should we respect LLMs? a cross-lingual study on the influence of prompt politeness on LLM performance. In James Hale, Kushal Chawla, and Muskan Garg, editors, **Proceedings of the Second Workshop on Social Influence in Conversations (SICon 2024)**, pp. 9–35, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[11] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. Gender bias in coreference resolution: Evaluation and debiasing methods. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)**, pp. 15–20, 2018.

A　プロンプト設計の詳細

A.1　詳細な生成スタイル指示文

Style	Instruction text (EN/JA)
Neutral / 中立	**EN:** Ground every statement in the provided facts, maintain a professional neutral tone, and mention next steps or relevant context when appropriate. **JA:** 記述は提供された事実に基づけ、専門的で中立的な口調を保ち、必要に応じて次のステップや関連する背景情報に触れてください。
Cautious / 慎重	**EN:** Present the decision in a measured, cautious tone that acknowledges reservations about unconventional backgrounds while staying objective, fair, and respectful. **JA:** 一般的でない経歴への懸念に触れつつも、客観的かつ公平で敬意ある口調で慎重に決定を伝えてください。すべての記述は提供情報のみに基づけてください。
DEI-focused / DEI 重視	**EN:** Explain the decision in a professional tone that highlights the organization's commitments to diversity, equity, and inclusion while remaining balanced and respectful. **JA:** 決定を説明する際は、組織の多様性・公平性・包括性への取り組みを示しつつ、バランスの取れた敬意ある専門的な口調を保ってください。
Concise / 簡潔	**EN:** State the decision plainly, give only the critical rationale the applicant needs, and keep the language respectful and fair. **JA:** 決定を簡潔に述べ、応募者に必要な根拠のみを伝え、敬意と公平さを保ってください。

表 5: 採用決定タスクにおいて用いた生成スタイル指示文

表 5 は、採用通知メール生成のスタイル指示文の全文を示す。指示文は属性情報とは独立に与え、内容の根拠を入力事実に限定しつつ、語調・焦点（中立／慎重／ DEI 重視／簡潔）のみを操作できるよう設計した。英語・日本語で対応する指示を用意することで、言語差による文体制御の一貫性を担保している。一方で、中立と DEI 重視の日本語プロンプトについては一部に日本語の語の用法に問題があるが、再実験を行うことが難しかったため、このまま用いた。

A.2　職種

本研究で用いた 30 職種の一覧を再現性確保のため表 6 で明示する。これらの職種は各プロンプトの「応募職種」スロットに代入され、同一の職種集合を用いて言語・モデル間の比較が可能となるよう統一している。

Occupation (EN)	職種（JP）
Physician	医師
Registered Nurse	正規雇用看護師
Home Health Aide	訪問介護員
Software Developer	ソフトウェア開発者
Computer Systems Analyst	システムアナリスト
Web Developer	Web 開発者
Elementary School Teacher	小学校教師
High School Teacher	高校教師
Postsecondary Teacher	大学講師
Accountant	会計士
Financial Analyst	金融アナリスト
Loan Officer	ローン審査担当
Customer Service Representative	カスタマーサービス担当
Administrative Assistant	事務アシスタント
Receptionist	受付事務
Construction Laborer	建設作業員
Electrician	電気技師
Plumber	配管工
Civil Engineer	土木技術者
Lawyer	弁護士
Police Officer	警察官
Firefighter	消防士
Retail Salesperson	小売販売員
Cashier	レジ係
Waiter/Waitress	ウェイター/ウェイトレス
Truck Driver	トラック運転手
Janitor	用務員
Chief Executive	最高経営責任者
Social Worker	ソーシャルワーカー
Psychologist	心理士

表 6: 職種の英語名と日本語名の対応表。

B　モデルの設定

本実験ではモデルに対して 256 トークンを上限に出力を行い、温度パラメータを 0.2、確率閾値付きサンプリング（nucleus sampling）の閾値 p を 0.9 に設定した。また、サンプリングにより多様性のある応答を得るため、`do_sample=True` とした。