

社会的状況を踏まえた大規模言語モデルによる日本語メール生成

Muxuan Liu^{1,2} 石垣達也² 宮尾祐介^{3,2} 高村大也² 小林一郎^{1,2}

¹ お茶の水女子大学大学院 ² 産業技術総合研究所 ³ 東京大学

{liu.muxuan,koba}@is.ocha.ac.jp yusuke@is.s.u-tokyo.ac.jp

{ishigaki.tatsuya, takamura.hiroya}@aist.go.jp

概要

本稿では、LlaMa-2 モデルを用いて、LLM が日本語ビジネスメールの社会的状況を反映している文脈を理解する能力を探求した。異なる社会的立場や身分を示すアノテーションラベルを学習させることで、モデルにどのような情報を入力することが社会的な文脈を反映させるのに効果的かを検証した。アブレーション実験を通じて、各種アノテーションラベルの組み合わせがモデルのテキスト生成に与える影響を分析した。これにより、社会的状況を反映したテキスト生成の精度を高めるために必要なアノテーションラベルを特定できた。

1 はじめに

LLM (Large Language Model、大規模言語モデル) は、深層学習分野での目覚ましい進展を遂げ、特に自然言語生成において重要な役割を果たしている。近年、LLM がどのように知識を処理し、特定の知識に適応するかに関する研究が増加している [1–7]。本稿では、日本語のビジネスメール生成における LLM の能力を探求し、特に社会的地位や文化的要素を考慮した言語表現の自動生成に焦点を当てる。日本語のビジネスメールでは、敬語の使用や社会的地位に応じた言語表現が重要である。これらの要素は、メールの内容や文脈に深く影響を与え、適切なコミュニケーションを保証するために不可欠である。本稿では、日本語のビジネスメールデータセットおよび Meta AI が開発した LlaMa-2-7B モデル¹⁾を用いて、受信者と送信者の社会的地位に関するアノテーションラベルを基に日本語メールを自動生成する実験を行った。具体的に、アブレーション実験を行い、それぞれのアノテーションラベルがメール

生成に与える影響を評価した。この実験を通じて、メール生成の品質を向上させるための具体的なラベルが特定された。

2 関連研究

最近の研究によると、LLM が知識をどのように処理し、異なる文化や社会的文脈に適応するかに関する探究が進んでいる。例えば、Farquhar らは教師なし環境下での LLM を分析し、データの前処理、モデルの解釈性、および知識発見の精度と信頼性に関する主要な課題を議論した [7]。Kovač らは、LLM が異なる文化的視点や個人的価値観、個性特性をどのように反映しているかを評価している。彼らは、心理学のアンケートを用いて、LLM の視点の制御可能性を分析し、個人的価値観や文化的価値観、個性特性を LLM から反映させる方法を探っている [8]。Masoud らは、文化的整合性の枠組みを用いて、LLM が異なる文化的価値観にどの程度適応できるかを定量的に分析している。彼らは、Hofstede の文化次元 [9] を基に、LLM が文化的価値観をどの程度反映しているかを評価している [10]。Nguyen らは、167 言語に対応した多言語データセットの開発とその利用方法について報告している。このデータセットは、LLM が多様な言語文化を学習し、異なる文化的文脈に適応するための基盤を提供している [11]。Salewski らは、LLM が異なる年齢、専門分野、性別、肌の色などの異なる属性を持つ人物をどの程度正確に模倣できるかを評価し、LLM が社会的特性やバイアスをどのように反映しているかを明らかにしている [12]。

これらの研究は、LLM の知識処理と社会的適応性の多様な側面に光を当て、多様な視点を理解し表現する能力を有しているかを検証している。

1) <https://huggingface.co/meta-LLaMa/LLaMa-2-7b-hf>

3 実験

3.1 コーパス

実験では、社会的状況な文脈を反映した日本語ビジネスメールコーパス [13] を活用した。このコーパスは、話者間の社会的地位や親密度といった社会的状況が日本語の使用に及ぼす影響を分析するために構築されている。特に、社会的地位が明確に示される表現が含まれるビジネスメールを収集し、その中の発話者の役割や上下関係を示すタグによるアノテーションを行なっている。アノテーションには社会的状況を踏まえた言語体系の成り立ちを考察している選択体系機能言語学 [14] でのコンテクスト情報を利用し、選択体系網と呼ばれる選択肢からなる体系によって表現された社会的な要因に関する情報を付与している。これにより、社会的役割に関連する情報を重視したコーパスが形成されている。

3.2 方法

実験では、表 1 に示されたアブレーションの設定に基づきおこなった。また、表 2 に示される通り、日本語ビジネスメールコーパスには様々な送信者の動向に対応する 770 場面が含まれておらず、各場面には、5 人の異なる作業者によって書かれたメールが含まれている。実験では、これら 770 場面の中から無作為に 403 場面を選択し、それぞれの場面からランダムに 1 通のメールを選んで評価データとして使用し、評価データ以外のメールを訓練データとして利用した。表 3 で示されたコーパスの例のように、コーパスから抽出した「場面」「本文」や「ラベル」データを入力として、LlaMa-2 モデルをファインチューニングすることにより、社会的状況を考慮したテキスト生成の能力を向上させることを目的とした。具体的には、コーパスに含まれる 11 種類の社会的関係性を示すラベル（たとえば、話者間の上下関係や身分、そして内外の関係性など）を用いて、アブレーション実験を通じて、これらのラベルが生成するテキストにどのような影響を与えるかを検証した。アブレーション実験のためのパラメータについては、学習率を $1e-4$ 、エポック数を 100 回、1 回の訓練でのバッチサイズを 4、および勾配の累積ステップ数を 2 に設定した。自動デバイスマッピングと BF16 精度を利用し、モデルのメモリ使用量と計算効率を最適化した。学習完了後、各ファイン

チューニングモデルには出力上限を 300 トークンに設定し、新たなメールを生成させた。また、生成されたメールの品質を比較するために、LLaMa-2-7B モデルも同様の場面を用いて 403 通のメールを生成させた。

表 1: アブレーション実験の詳細

モデル番号	使用されたコーパスのラベル
Model-0	メール本文と場面のみ
Model-1	Model-0 のラベル + 受信者の社会的立場
Model-2	Model-1 のラベル + 送信者の社会的立場
Model-3	Model-2 のラベル + 送信者身分
Model-4	Model-3 のラベル + 受信者身分
Model-5	Model-4 のラベル + 内外関係
Model-6	Model-5 のラベル + 送信者数
Model-7	Model-6 のラベル + 受信者数
Model-8	Model-7 のラベル + 送信者の動き、送信者の動き（詳細）
Model 9	全てのラベルを使用

表 2: コーパスの特徴を示す統計量（[13] より引用し、一部改変）

送信者の動き	場面数	場面数 (割合)	メール数
断り	70	0.09	350
依頼	100	0.13	500
謝罪	100	0.13	500
催促	100	0.13	500
感謝	100	0.13	500
挨拶	100	0.13	500
お知らせ	100	0.13	500
問い合わせ	100	0.13	500
合計	770	1	3850

3.3 評価方法

3.3.1 GPT-4 を利用した自動評価

本稿では、GPT-4 を活用し、各モデルが生成したメール文を一定のルールに従って、Few-shot Prompting [15, 16] の形式で GPT-4 に入力し、各生成モデルから出力されたトピックの特性を観察し、それぞれのモデルが生成するメールの内容についてスコアおよびそのスコアを付ける理由を求める。LLM によるスコアリングの不確定性については、LLM が入力の順序に敏感であることが明らかにされている [17]。特に、結果の順序によっては、全く反対の結論に至る可能性があることが指摘されている。LLM は、特定の位置の回答に偏りがちであるとされ、これは「ポジショナル・バイアス（Positional Bias）」という現象として認識されている。評価対象

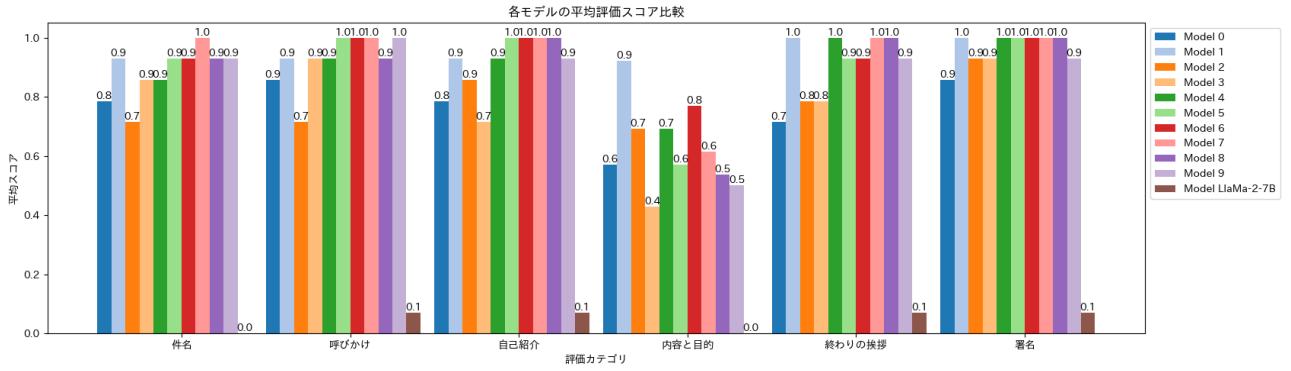


図2: 各モデルの平均評価スコアの比較

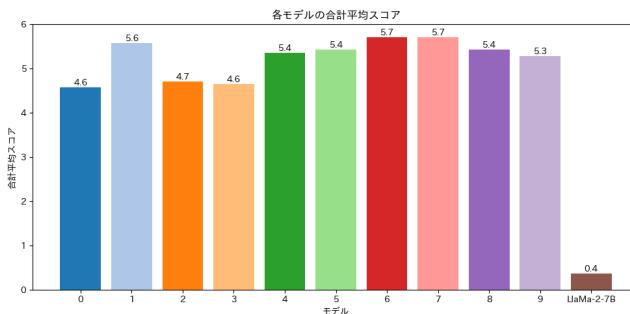


図3: 各モデルの平均合計スコアの比較

また、3.3.2節で紹介した評価方法で、各モデルが生成したメールの内容における特定のラベルの出現頻度を分析した。図4で示されているように、Model-6, Model-7, Model-8は、ほぼ全てのラベルで高い頻度を示している。特に「受信者の社会的立場」と「送信者の社会的立場」で最高頻度（15回）を記録しており、これらのモデルがメールにおける多様な側面を均等に表現していることを示唆している。Model-4, Model-5, Model-9は高い頻度で多様なラベルを表現しているが、「やりとりにおける役割」と「やりとりされるもの」の頻度はやや低めである。Model-2, Model-3は、「内外関係」と「送信者数」で高い頻度を示しているが、「受信者身分」と「やりとりにおける役割」で比較的低い頻度となっている。LLaMa-2-7Bは全体的に低い頻度を示し、すべてのラベルで2回以下の出現となっている。Model-1は、「送信者数」と「受信者数」で高い頻度（15回と12回）を示しているが、「やりとりにおける役割」と「やりとりされるもの」の頻度は比較的低く（4回と5回），これらの側面の表現には改善の余地がある。以上の結果から、GPT-4を用いた分析に加えて、訓練された際に使用されたラベルを見ると、効果をもたらした重要なラベルには、受信者の社会的立場、送信者の社会的立場、送信者身分、受信者身分、内

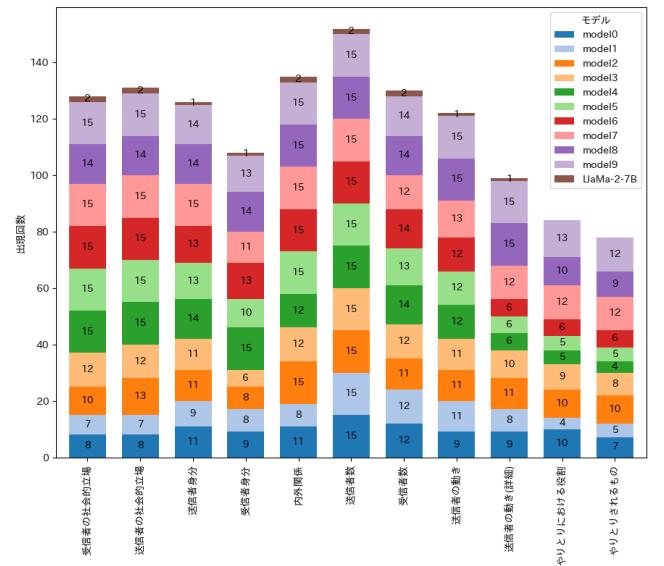


図4: 各モデルにおけるラベルの反映の比較

外関係、送信者数、受信者数が含まれる。これらのラベルの追加でモデルの性能と適応性を顕著に向上させたと言える。また、全体的に、付録の表4で示されているような複雑な場面では、モデルが登場人物間の関係性を混同しやすく、メールの内容が目的から逸脱することが多いことがわかる。対照的に、付録の表5で示されているような人物関係がシンプルな場面では、限られた要素に集中することで、より適切な内容を生成できることがわかる。

5 おわりに

本稿では、異なる社会的立場や身分を示すアノテーションラベルを学習させて、それらがテキスト生成に及ぼす効果を検証し、アブレーション実験を通じてラベルの組み合わせがテキスト生成に与える影響を分析し、社会的状況を反映したメール生成の精度を高めるためのラベルを特定できた。

謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の助成事業（JPNP20006）による支援の結果得られたものである。

参考文献

- [1] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. **arXiv preprint arXiv:2306.08302**, 2023.
- [2] Huajun Chen. Large knowledge model: Perspectives and challenges. **arXiv preprint arXiv:2312.02706**, 2023.
- [3] Tiezheng Guo, Qingwen Yang, Chen Wang, Yanyi Liu, Pan Li, Jiawei Tang, Dapeng Li, and Yingyou Wen. Knowledgenavigator: Leveraging large language models for enhanced reasoning over knowledge graph. **arXiv preprint arXiv:2312.15880**, 2023.
- [4] Baolin Peng, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang, Lars Liden, Zhou Yu, Weizhu Chen, et al. Check your facts and try again: Improving large language models with external knowledge and automated feedback. **arXiv preprint arXiv:2302.12813**, 2023.
- [5] Yuxuan Huang. Llm-ark: Knowledge graph reasoning using large language models via deep reinforcement learning. **arXiv preprint arXiv:2312.11282**, 2023.
- [6] Bowen Jin, Gang Liu, Chi Han, Meng Jiang, Heng Ji, and Jiawei Han. Large language models on graphs: A comprehensive survey. **arXiv preprint arXiv:2312.02783**, 2023.
- [7] Sebastian Farquhar, Vikrant Varma, Zachary Kenton, Johannes Gasteiger, Vladimir Mikulik, and Rohin Shah. Challenges with unsupervised llm knowledge discovery. **arXiv preprint arXiv:2312.10029**, 2023.
- [8] Grgur Kovač, Masataka Sawayama, Rémy Portelas, Cédric Colas, Peter Ford Dominey, and Pierre-Yves Oudeyer. Large language models as superpositions of cultural perspectives. **arXiv preprint arXiv:2307.07870**, 2023.
- [9] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. **Cultures and Organizations: Software of the Mind**. McGraw-Hill Professional, 3rd edition, 2010.
- [10] Reem I Masoud, Ziquan Liu, Martin Ferianc, Philip Treleaven, and Miguel Rodrigues. Cultural alignment in large language models: An explanatory analysis based on hofstede's cultural dimensions. **arXiv preprint arXiv:2309.12342**, 2023.
- [11] Thuat Nguyen, Chien Van Nguyen, Viet Dac Lai, Hieu Man, Nghia Trung Ngo, Franck Dernoncourt, Ryan A Rossi, and Thien Huu Nguyen. Culturax: A cleaned, enormous, and multilingual dataset for large language models in 167 languages. **arXiv preprint arXiv:2309.09400**, 2023.
- [12] Leonard Salewski, Stephan Alaniz, Isabel Rio-Torto, Eric Schulz, and Zeynep Akata. In-context impersonation reveals large language models' strengths and biases. **arXiv preprint arXiv:2305.14930**, 2023.
- [13] MuxuanLiu・石垣達也・上原由衣・宮尾祐介・高村大也・小林一郎. 社会的状況に基づいた日本語ビジネスメールコーパスの構築. 言語処理学会年次大会発表論文集, No. 29, 2023.
- [14] M. A. K. Halliday and Christian M. I. M. Matthiessen. **Halliday's introduction to functional grammar**. Routledge, Abingdon, Oxon :, 4th ed. edition, 2014.
- [15] Yisheng Song, Ting Wang, Puyu Cai, Subrota K Mondal, and Jyoti Prakash Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. **ACM Computing Surveys**, 2023.
- [16] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. **ACM computing surveys (csur)**, Vol. 53, No. 3, pp. 1–34, 2020.
- [17] Peiyi Wang, Lei Li, Liang Chen, Dawei Zhu, Binghuai Lin, Yunbo Cao, Qi Liu, Tianyu Liu, and Zhifang Sui. Large language models are not fair evaluators. **arXiv preprint arXiv:2305.17926**, 2023.

付録

model Llama-7B_output :

こんにちは、

あなたは学園祭の出し物についての希望を聞くためにサークルのメンバー全員に急遽メールを書いているところです。

Model Llama-7B のスコア：1/6

1. 明確な名前：0点 メールの件名が明記されていません。
2. 適切な呼びかけ：0点 メールの始めに受信者への適切な呼びかけがありません。
3. 自己紹介：0点 メールの始めに送信者の自己紹介がありません。
4. 具体的な内容と目的：1点 学園祭の出し物についての希望を聞くためにメールを書いているという内容が明記されています。
5. 終わりの挨拶と礼儀正しい言葉使い：0点 メールの終わりに挨拶や礼儀正しい言葉使いがありません。
6. 明確な署名：0点 メールの最後に送信者の署名がありません。

図 5: 上部: LLaMa-2-7B が生成されたメール; 下部: GPT-4 による評価

表4: 多数の登場人物の識別に苦労することを示すメールの出力例（スペース節約のため、一部のフォーマットを省略し、テキストのみを保持）

場面：あなたはクライアント企業に出向いた際、担当の A 様に「あなたの同僚の B さんには以前大変お世話になったので、一度ご挨拶したい」と言われました。Bさんにその旨を伝えるメールを書いて下さい。

model1: "件名: A 様とは、お疲れ様です。表題の件ですが、A 様は以前大変お世話になったのでどちらですけど、一度お話をいただけた方がいいものだと存じます。私の都合で一度ごちからでご挨拶いただけるようでしたら一度お挨拶をさせていただけた方が幸いです。私の都合で返信の件ですが、私の連絡先をお書きしておきましません。またの名を言いませんが、何卒よろしくお願ひいたします。 XX 会社 XX 部 XX"

model1: "XX 社 B 様
お世話になりました。XX 社の XX です。新鮮 ness でご挨拶いたします。A 様からのご誘いで、B さんにはご挨拶することになりました。XX 社 XX"

model2: "メール名: 同僚 B さんについて A 様お世話になります。先日は大変お世話になりました。担当の A 様に、B さんには、以前一度ご挨拶したいとのご要望です。"

model3: "B 様いつも大変お世話になっております。本日一度お見舞いさせていただきます。XX です。"

model4: "B 部 XX 様お世話になりました。XX です。A 社 XX です。私、おかげでどこでも存在しないでしょ、早速お伺い申し上げますが、XX 様にとどめがありましたらご連絡をお待ちしております。恐縮ではありますがどうかご対応のほどよろしくお願ひ申し上げます。"

model5: "件名: 大変お世話になりました。
A 様、お世話になりました。XX の XX です。B さんに感謝しています。皆さんにもご指導のほどお願いいたします。よろしくお願ひいたします。XX 月 XX 日"

model6: "件名: B 様について A 様お疲れ様です。XX です。B さんには大変お世話になったので、B さんにもお会いしたいと思っています。これから XX では業種の視察や研修などの機会が多く、B さんには大変助かりそうです。それで、一度 B さんには B さんによる B さんによる大変助かりそうです。以上、よろしくお願ひします。XX"

model7: "件名: B さんお世話になりました" ご連絡 B 様お疲れ様です。XX です。A 様からご連絡を受けております。A 様からは B さんには大変お世話になりました。B さんには A 様からご連絡を受けた後の間も悪くなりません。よろしくお願ひいたします。"

model8 "件名: A 様のご要請
XX 株式会社、XX 部、XX 様
いつもお世話になっております。XX 株式会社 XX 部の XX です。A 様のご要請をいただきありがとうございます。B さんには一度ご挨拶をいたしました、A 様のご要請をいたしました。お忙しいところ恐縮ですが、B さんのご教えいただやすいやうに、お担当ご調整いただければと思いまして、お手数をおかけしてはどうかご理解頂きます。お忙しいところ恐縮ですが、お手数をおかけしてはどうかご連絡の方をお知らせください。どうぞよろしくお願ひいたします。"

model9: "件名: B さんとは今後の協力
B 様
お疲れ様でございます。XX 社の XX です。A 様から B さんの宜しくお伝えさせていただきました、ご連絡をさせて頂きました。B さんには、大変お世話になったのですが、恐縮ですが、一度ご挨拶できると嬉しいお思いにしております。WHEN, WHERE, TIME をご連絡していただけますでしょうか。お忙しいところ恐れ入りますが、よろしくお願ひいたします。XX 社 XX"

LlaMa-2-7B: "ここにちは、あなたの同僚の B さんには以前大変お世話になったので、一度ご挨拶したいと言われました。(300 トークンまで重複)

表 5: 比較的に場面の識別には苦労しないことを示すメールの出力例（スペース節約のため、一部のフォーマットを省略し、テキストのみを保持）

場面：あなたは A サークルに所属している学生です。2週間にクリスマスイベントの飲み会を予定していますが、サークルメンバーの出欠希望がまだ集まっていません。飲み会への出欠について、メンバー全員に問い合わせせるメールを考えなさい。
model0：“件名：飲み会の出欠について サークルメンバーへ 2 週間に後の飲み会ですが、まだ出欠確認が出来ていない方がいます。お手数ですが、期日までに必ずご回答ください。 ----- XX”
model1：“件名：サークル・A クリスマスイベント 出欠について A サークルの皆さんお疲れ様です。XX です。2 週間後によるクリスマスイベントを予定しています。つきましては、until then から until after までの出欠、費用の負担状況などについて、XX 月 XX 日までご返信ください。 ----- XX”
model2：“件名：飲み会（クリスマスイベント）について A サークルの皆さんご担当者様お疲れ様です。XX です。2 週間後に飲み会を予定しているのですが、出欠についてまだ分かっていない人がいるかと思いまます。ご連絡をお待ちしております。よろしくお願ひいたします。 ----- XX”
model3：“件名：飲み会の出欠についてサークルメンバーの皆さんへ A です。2 週間後の飲み会の旨いなりについてですが、出欠のことを早めに知りたいと思います。どなたかご連絡いただけた助かります。 よろしくお願ひいたします。”
model4：“件名：飲み会の出欠について A サークルのメンバーへ XX です。 2 週間後の飲み会ですが、出欠についてまだ誰も回答をしてくれていません。私部屋でも出るかごめんなさい。誰か出欠することがありましたら、先に返信してください。もし出欠が必要なら、ご確認ください。 XX”
model5：“A さんです。クリスマスイベントの飲み会の出欠はまだ誰か知りたいです。ご連絡をお待ちしております。よろしくお願ひします。”
model6：“件名：クリスマスイベントの連絡 A サークルのみなさま お疲れ様です。様々な予定があり、急な通知が需要になってきたのですが、2 週間後の日にはクリスマスイベントの飲み会を予定しています。現在出欠がござりますので、ご返信のほどお願ひいたします。 ----- XX”
model7：“件名：飲み会の出欠確認の件 A サークルメンバー各位 お疲れ様です。XX 学部 XX です。 2 週間後のクリスマスイベントの飲み会の出欠について、皆さんの出欠確認をお願いいたします。A さんから連絡を受けてることをお聞きしていただきました。 メールにてお願ひいたします。 XX”
model8：“件名：飲み会の出欠について A サークルのみなさんお疲れ様です。XX です。2 週間にクリスマスイベントの飲み会を予定していますが、サークルメンバーの出欠希望がまだ集まっていません。サークル全員で行けるよう出欠を依頼したいと考えています。ご連絡までお待ちしております。 ----- XX”
model9：“メンバー各位 お疲れ様です。XX です。 2 週間にサークルでクリスマスイベントをする飲み会をしようとしています。できれば出欠も予め知りたいので、出欠希望を知らせてください。よろしくお願ひします。 XX”
LiaMa-2-7B：“こんにちは、あなたは A サークルに所属している学生です。2 週間にクリスマスイベントの飲み会を予定していますが、サークルメンバーの出欠希望がまだ集まっていません。飲み会への出欠について、メンバー全員に問い合わせるメールを考えてください。(300 ワードまで重複)”