

P-BEE: プレースホルダを用いた送信者情報抽出用 ビジネスメール生成フレームワークの初期検討

根本颯汰 大田尾匠 橋本航

Sansan 株式会社

{sota.nemoto, sho.otao, wataru.hashimoto}@sansan.com

概要

ビジネスメールから送信者の氏名や所属などの属性情報を抽出する送信者情報抽出タスクでは、既存の公開メールデータセットでは多様性や匿名化の問題により本タスクへの適用が困難である。本研究では、LLMを用いた日本語ビジネスメール生成フレームワーク Placeholder-based Business-Email Exemplar (P-BEE) を提案する。類似度や多様性の定量評価や送信者情報抽出タスクに対する質問応答モデルの性能の検証をした結果、P-BEEによる生成メールは社内で分析可能なメールデータと比べて少し乖離があり、更なる改善が必要であることを示した。

1 はじめに

電子メールは多くの人々が利用する代表的なコミュニケーションツールの一つである。特にビジネスシーンでは、業務連絡や合意形成等の重要なプロセスは依然としてメールを中心に行われている [1, 2, 3, 4]。NLP 技術の発展により、これらのプロセスの自動化を目的としたタスクがいくつか研究されている [5, 6, 7]。その中でも送信者情報抽出は送信者の属性 (氏名や所属等) を抽出するタスクであり、送受信履歴や企業間関係の解析などに貢献するため、その性能改善は重要な課題である [8, 9, 10]。

送信者情報抽出タスクにおいて、既存の公開されているメールデータセットでは2つの限界が存在する。1つ目は代表的なデータセットの多くが英語かつ旧来のものであり、日本語や現代のビジネスシーンの文脈に対応できていない点である。例えば、Enron Email Corpus [11] や Avocado Email Collection ¹⁾ は 2015 年以前に公開された英語のみのデータセットであり、現代の社会情勢などを網羅できていな

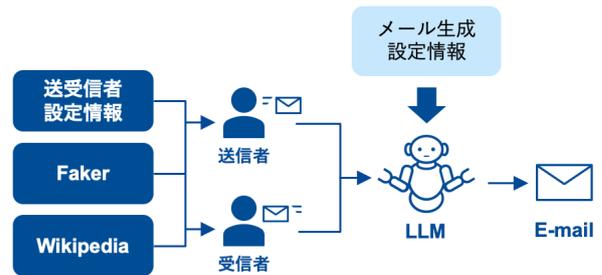


図1 P-BEEによる擬似メール生成の全体像。

い。2つ目はこれらのデータセットがプライバシー保護を目的とした過度な匿名化が施されている点である。実際に、属性情報が汎用的な文字列へ置換・削除されており、現実的な出現パターンが損なわれている²⁾。結果として、実用的な属性情報を保持しつつ現代のビジネスシーンに即した日本語のメールデータを十分に確保することは困難である。

本研究では、LLMを用いて擬似的な日本語ビジネスメールを生成するフレームワーク Placeholder-based Business-Email Exemplar (P-BEE) を提案する (Fig. 1)。LLMは近年のデータを含む膨大なコーパスを用いた事前学習をしており、現代の社会情勢やビジネス表現を高度に再現可能である。また、P-BEEはプロンプトのプレースホルダを介して任意の属性情報を動的に注入する機構を持つため、安全な属性情報を自在に組み込んだメール生成が可能となり、匿名化に伴う情報の欠落やプライバシー上の制約を回避できる。我々はP-BEEを用いて5,000件のメールデータを生成し、語彙や文脈の類似度・多様性の観点から品質を定量評価した。さらに、生成したメールデータと社内で分析可能なメールデータ (実メール) を使用して、抽出性能の比較を検証した。その結果、生成したメールでは実メールほどの抽出性能を引き出すことはできず、改善の余地があることが確認できた。

1) <https://catalog.ldc.upenn.edu/LDC2015T03>

2) <https://www.cs.cmu.edu/~enron/>

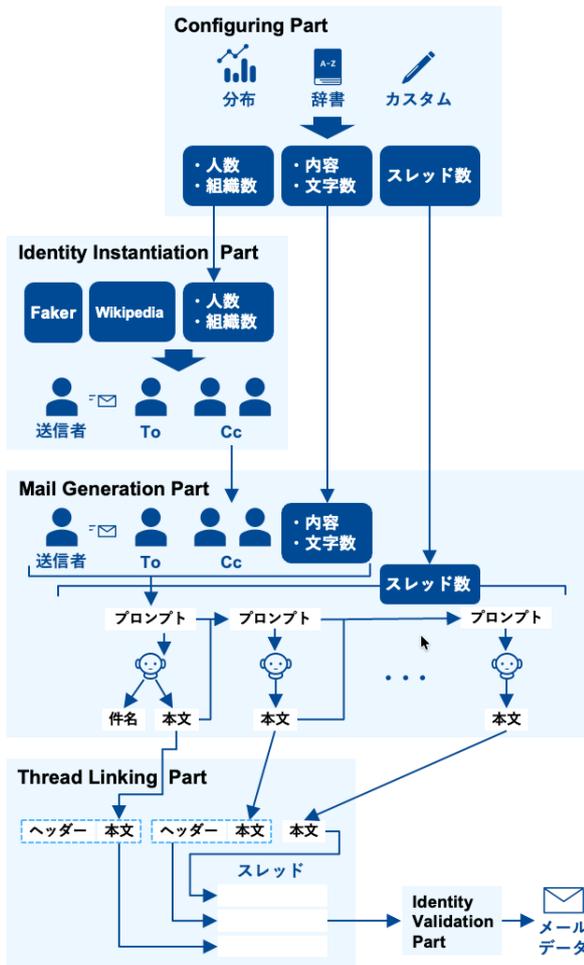


図2 P-BEEの5 Partによるプロンプト設計と生成フロー.

2 提案手法

我々は LLM を用いて擬似的に日本語ビジネスメールを生成するフレームワーク Placeholder-based Business-Email Exemplar (P-BEE) を提案する. P-BEE は 1) Configuring, 2) Identity Instantiation, 3) Mail Generation, 4) Thread Linking, 5) Identity Validation の 5 つの Part から構成される (Fig. 2).

Configuring Part ここでは柔軟なメール生成を可能にするために, 付録 A の表 4 の項目に対して送受信者やプロンプト, メールなどの生成時に必要なパラメータを設定する. 今回のメール生成において, パラメータの決定に必要な変数は社内でも分析可能なメールデータを参考にした.

Identity Instantiation Part ここでは Configuring Part で設定したパラメータを元に送受信者の属性情報を取得する (表 1). P-BEE のプレースホルダ形式によって任意の属性情報を注入できるが, 本研究で

表 1 Identity Instantiation Part で設定した属性情報の例.

項目	例
氏名	山田太郎
所属する企業名	株式会社佐藤情報
部署名	営業部
役職名	部長
会社の電話番号 (TEL)	03-1624-9184
会社の FAX 番号 (FAX)	03-1662-4307
個人の電話番号 (MOBILE)	090-0511-7042
会社の郵便番号	265-8376
会社の住所	京都府横浜市都筑区皇居外苑 31 丁目
メールアドレス	taro.yamada@satojoho.com
メールヘッダーの Display Name	山田太郎<taro.yamada@satojoho.com>

は属性情報を Faker と Wikipedia から取得して検証した. 詳細な生成方法は付録 A.1 に示す. また, 本手法で生成したメールの個人情報の取り扱いについては付録 A.2 で議論する.

Mail Generation Part Configuring Part のパラメータを注入したシステムプロンプト (付録 A.3) と Identity Instantiation Part で作成した送受信者情報を注入したユーザープロンプトから構成されるプロンプト p_1 を入力として時系列順にメールを生成する. 具体的に, p_1 を入力して最初のメールの件名 s と本文 c_1 を生成する. 次に, p_1 に生成した本文 c_1 を加えたプロンプト p_2 を入力して c_1 への返信本文 c_2 のみを生成する. 以降, スレッド数 N に応じて返信本文の生成を繰り返すことで全メール本文 $c = [c_1, c_2, \dots, c_N]$ を取得できる. この時, 各メールのプロンプト $p = [p_1, p_2, \dots, p_N]$ は送受信者の組み合わせ (From, To, Cc) をランダムに変更する.

Thread Linking Part 一般的にユーザーが取得できるメールはスレッド形式であり, 最新のメールが最上部に位置し, ヘッダーを挟んで時系列の逆順で連結される. 本 Part では, Mail Generation Part で生成した本文 c に対し, 最新のメール以外の各メールの冒頭にヘッダー $h = [h_1, h_2, \dots, h_{N-1}]$ を挿入し, 時系列の逆順に連結してスレッド $t = c_N + h_{N-1} + c_{N-1} + \dots + h_1 + c_1$ を構築する. この時, ヘッダーは事前に用意した 10 個のテンプレートからメールごとにランダムで選択し, Display Name などの属性情報を注入して作成した.

Identity Validation Part 最後に Thread Linking Part で完成したスレッド t では LLM の hallucination の影響から, 送信者の属性情報のうち一部文字列が欠けたり, 全く含まれない可能性がある. そのため, Identity Validation Part では正規表現を用いて送受信者情報の修正をした.

表 2 P-BEE による生成メールに対する品質検証結果.

LLM	Similarity		Diversity		
	Jaccard ↑	BERTScore ↑	Distinct-1 ↑	Distinct-2 ↑	BERTScore ↓
GPT-4.1	0.1401	0.8149	0.9672	0.6932	0.8432
Qwen3 30B	0.1398	0.8153	0.9698	0.7277	0.8415
Gemma3 27B	0.1336	0.8163	0.9676	0.7442	0.8422
Sarashina2.2 3B	0.1637	0.8064	0.9781	0.8305	0.8349

3 実験

3.1 品質評価

P-BEE により生成された疑似メールデータの品質を定量的に評価する. 具体的には, 社内で分析可能なメールデータ (実メール) との類似度 (Similarity) および生成データ内の多様性 (Diversity) の観点から, 複数の LLM による生成品質の差異を検証する.

3.1.1 評価用データセットとモデル設定

本評価では, 実メール 1,000 件と, P-BEE が生成したメール 5,000 件から無作為に抽出した 1,000 件を比較対象とする. 実メールは社内で分析可能なメールデータより選定した. また, 生成モデルにはクローズモデルである GPT-4.1 および 3 つのオープンモデルである Qwen3 [12] 30B³⁾, Gemma3 [13] 27B⁴⁾, Sarashina2.2⁵⁾ 3B⁶⁾ の計 4 つの LLM を用いてモデル間のメールの品質差を検証する. 全モデルにおいて, 多様性と再現性のバランスを考慮し, パラメータは temperature を 1.0, top_p を 0.9 に固定した.

3.1.2 評価方法

生成メールの品質を, 実データへの忠実性とデータセット内のバリエーションの両面から評価するため, Similarity および Diversity の 2 項目を定義し, それぞれ 2 つの指標を用いて定量的に評価する.

Similarity 実メールと生成メールの類似度を検証し, P-BEE の実メールに対する再現性を測るために以下の 2 つの指標を用いる. 1 つ目は **Jaccard 係数** による語彙の共通性の評価である. これは実メール群と生成メール群から抽出されたユニークな単語 (1-gram) の集合を対象とし, 両集合間での語彙の一

致率を計算する. 2 つ目は **BERTScore** [14] を用いた意味的な類似度の評価である. これは事前学習済みモデル⁷⁾ の分散表現間のコサイン類似度に基づき, 文脈的な意味の類似性を評価する. 本研究では, 実メールと生成メールの全ペアにおける F1 スコアの平均値を用いる.

Diversity 生成データが特定のパターンに偏らず, 多様な表現を保持しているかを検証するため以下の 2 つの指標を用いる. 1 つ目は **Distinct-n** [15] による語彙の多様性評価である. 生成メールに含まれる全 n -gram ($n \in \{1, 2\}$) に対するユニークな n -gram 数の割合を算出する. 2 つ目は **BERTScore** [14] を用いた意味的な多様性の評価である. 本指標では, 評価用生成メールに対する全ペア ($1000C_2$) の BERTScore を計算し, その平均値を算出する. Distinct-n はスコアが高く, BERTScore のスコアは低いほど生成メールの語彙や文脈が重複せず, 広範な表現を獲得できていることを示す.

3.2 送信者情報抽出性能

P-BEE により生成されたメールに対して, 情報抽出タスクのデータとしての有効性を検証するため, 実メールに対する送信者情報の抽出実験を行った.

設定 抽出エンジンとして Gemma3 [13] の 12B Instruct⁸⁾ を使用し, パラメータは再現性を重視して temperature を 0.0, top_p を 0.5 とした. 抽出の際の Few-shot (3-shot) として与えるメールのソースを以下の 5 パターンで切り替え, それぞれの抽出性能を比較した: 1) 実メール, 2) GPT-4.1, 3) Qwen3 30B, 4) Gemma3 27B, 5) Sarashina2.2 3B. 実メールを用いた際の結果を本設定のベースラインとし, 各 LLM の生成メールがどの程度代替可能かを評価する. なお, 抽出用プロンプトの詳細は付録 A.4 に示す.

評価方法 評価対象として, 実メール 1,000 件から送信者情報 (氏名, 会社名, 部署, 役職) の抽出を

3) Qwen/Qwen3-30B-A3B-Instruct-2507

4) google/gemma-3-27b-it

5) <https://www.sbintuitions.co.jp/blog/entry/2025/03/07/093143>

6) sbintuitions/sarashina2.2-3b-instruct-v0.1

7) FacebookAI/xlm-roberta-large

8) google/gemma-3-12b-it

表 3 Gemma3 (12B Instruct) による質問応答モデルの送信者情報抽出性能結果.

Few-shot	氏名			会社名			部署			役職		
	Prec.	Rec.	F1									
実メール	0.660	0.651	0.655	0.641	0.638	0.639	0.372	0.467	0.414	0.068	0.478	0.118
GPT-4.1	0.531	0.525	0.528	0.493	0.491	0.492	0.234	0.298	0.262	0.118	0.574	0.196
Qwen3 30B	0.478	0.478	0.478	0.583	0.574	0.579	0.232	0.289	0.258	0.084	0.487	0.143
Gemma3 27B	0.476	0.469	0.473	0.623	0.619	0.621	0.260	0.325	0.289	0.108	0.617	0.183
sarashina2.2 3B	0.601	0.595	0.598	0.594	0.593	0.593	0.238	0.326	0.276	0.107	0.635	0.183

試みた. 各抽出ターゲットに対し, スペースおよび改行文字を無視した完全一致により判定を行った. モデルが情報を出力した数を N_{output} , 正解データが存在する数を N_{exist} , 正解が存在する場合に正解とモデルの出力が一致した数を N_{tp} と定義し, 以下の通り Prec, Rec, F1 スコアを算出した.

$$\text{Prec} = \frac{N_{tp}}{N_{output}}, \text{Rec} = \frac{N_{tp}}{N_{exist}}, \text{F1} = \frac{2 \times \text{Prec} \times \text{Rec}}{\text{Prec} + \text{Rec}}$$

4 結果

4.1 品質評価

各 LLM で生成したメールの品質評価の結果を表 2 に示す. Similarity において, Jaccard は Sarashina2.2 が最も高く, BERTScore は Qwen3 [12] が最も高かったが, 大差はなかった. これより, 実メールと語彙は異なるもののビジネス的な文脈は再現できていることがわかった. また Diversity においては, すべての指標で Sarashina2.2 が最も優れた結果となった.

4.2 送信者情報抽出タスク

実メールと各 LLM で生成したメールごとの送信者情報抽出性能の結果を表 3 に示す. 氏名・会社名・部署名では実メールを使用すると最も性能が高く, 生成メールに比べて 5~15% も高い結果となった. 一方で, 役職名では実メールを使用すると最も性能が低く, GPT-4.1 が最も高く, 生成メールの中で一番低い Qwen3 [12] でも 2.5% の差が生じた.

5 Discussion

表 2 の結果より, Similarity と Diversity の双方で Sarashina が優れた値を示した. そこで Jaccard や Distinct の算出対象 (名詞, 動詞, 形容詞, 副詞, 連体詞, 感動詞) のユニークな単語数を確認したところ, 実メールの 44,868 語に対し, GPT-4.1 や Gemma3 [13],

Qwen3 [12] は約 15,000 ~ が 16,000 語であったが, Sarashina2.2 は 25,812 語と顕著に多かった. しかし, プロンプトで指定した文字数以下で生成できた割合は他モデルが 9 割を超える一方で, Sarashina は約 1 割に留まった. これより, P-BEE の複雑なプロンプトの理解は 3B のモデルでは困難であり, 指示追従性と実メールに近い語彙を実現するためには語彙とモデルサイズが伴ったモデルが必要である.

また, 表 2 ではモデル間に品質指標に差があるものの, 表 3 では抽出性能に同様の傾向が見られなかった. これは生成メールの語彙や文脈の多様性が抽出性能に与える影響が限定的であることを示唆する. 一方で, 実メールに比べ抽出性能が比較的高かった Gemma3 [13] の生成メールでもユニークな送信者の会社名数が約 200 件不足していた. これは Faker の生成パターンが「法人格+苗字+任意のドメイン (ex. 建設, 情報)」に固定されており, 多様性に欠けることが要因である. また実メールでは 1 人に対する複数部署の所属や日英や省略形などの役職の表記揺れなどの多様性が一定数存在した. 以上より, 非匿名性を維持しつつ属性情報のパターンをさらに増強する必要がある.

6 おわりに

本研究では LLM を用いて擬似的な日本語ビジネスメールを生成するフレームワーク P-BEE を提案した. P-BEE は, プレースホルダを介して属性情報を動的に注入する機構を持ち, 柔軟なメール生成を可能にする. 定量評価の結果, 単語単位での重なりは少ないものの, 文脈レベルでは実メールと高い類似性を有することを確認した. しかし, 抽出タスクの例としての利用にはまだ実メールとの性能乖離が残る. 今後は, この乖離を捉える評価指標の策定に加え, モデルの選定や属性情報のパターンの多様性を図ることで, 更なる性能改善が期待できる.

参考文献

- [1] Marina Jovic and Salaheddine Mnasri. Evaluating ai-generated emails: A comparative efficiency analysis. **World Journal of English Language**, Vol. 14, No. 2, 2024.
- [2] Aynura Valiyeva and Basil John Thomas. Successful organizational business communication and its impact on business performance: An intra-and inter-organizational perspective. **Journal of Accounting, Business and Finance Research**, Vol. 15, No. 2, pp. 83–91, 2022.
- [3] Celia T Romm and Nava Pliskin. Toward a virtual politicking model. **Communications of the ACM**, Vol. 40, No. 11, pp. 95–100, 1997.
- [4] Salaheddine Mnasri and Stavros Papakonstantinidis. The teaching-learning dynamic from behaviourism to social constructionism: a communication-centred narrative. **International Journal of Learning and Change**, Vol. 15, No. 3, pp. 312–327, 2023.
- [5] Sudipto Mukherjee, Subhabrata Mukherjee, Marcello Hasegawa, Ahmed Hassan Awadallah, and Ryen White. Smart to-do: Automatic generation of to-do items from emails. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 8680–8689, Online, July 2020. Association for Computational Linguistics.
- [6] Kexun Zhang, Jiaao Chen, and Diyi Yang. Focus on the action: Learning to highlight and summarize jointly for email to-do items summarization. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, **Findings of the Association for Computational Linguistics: ACL 2022**, pp. 4095–4106, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [7] Saurabh Srivastava, Gaurav Singh Tomar, Shou Matsumoto, Ali Raz, Paulo Costa, Joshua Poore, and Ziyu Yao. Mailex: Email event and argument extraction. In **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 12964–12987, 2023.
- [8] Vitor R Carvalho and William W Cohen. Learning to extract signature and reply lines from email. In **Proceedings of the Conference on Email and Anti-Spam**, Vol. 2004, 2004.
- [9] Einat Minkov, Richard C Wang, and William Cohen. Extracting personal names from email: Applying named entity recognition to informal text. In **Proceedings of human language technology conference and conference on empirical methods in natural language processing**, pp. 443–450, 2005.
- [10] Jelena Graovac, Ivana Tomašević, and Gordana Pavlović-Lažetić. Meet your email sender-hybrid approach to email signature extraction. In **Asian Conference on Intelligent Information and Database Systems**, pp. 547–558. Springer, 2022.
- [11] Bryan Klimt and Yiming Yang. The enron corpus: A new dataset for email classification research. In **European conference on machine learning**, pp. 217–226. Springer, 2004.
- [12] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. **arXiv preprint arXiv:2505.09388**, 2025.
- [13] Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivièrè, et al. Gemma 3 technical report. **arXiv preprint arXiv:2503.19786**, 2025.
- [14] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. **arXiv preprint arXiv:1904.09675**, 2019.
- [15] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In Kevin Knight, Ani Nenkova, and Owen Rambow, editors, **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational**

Linguistics: Human Language Technologies, pp. 110–119, San Diego, California, June 2016. Association for Computational Linguistics.

A 付録

表4 Configuring Part で設定するパラメータ一覧.

項目	値の決め方
メールのスレッド数	$\mu = 6.0, \sigma = 3.9$ の対数正規分布を $3 \leq x \leq 10$ で離散化†
メールのトピック	16 個のビジネストピックから選択 (ex. 商談, 発注)
生成するメール本文の文字数	70 文字: 50%, 600 文字: 50%
宛名, 名乗り, 署名のパターン	属性情報全て, 会社名+氏名, 氏名のみ, なし, からランダム
引用符 (>) をつける割合	あり: 50%, なし: 50%
登場する受信者数	$\mu = 5.5, \sigma = 2.7$ の対数正規分布を $1 \leq x \leq 10$ で離散化†
To と Cc の割合	to: 20%, cc: 80%
受信者が属する企業のユニーク数	$\mu = 2.8, \sigma = 0.7$ の対数正規分布を $1 \leq x \leq 4$ で離散化†
法人格のパターン	*株式会社: 70%, "(株)": 5%, その他: 25%
部署・役職を記載する割合	各 50%

† 実メールに近いロングテールな分布を再現するために対数正規分布を採用した.

A.1 属性情報の作成方法の詳細

Faker とは Python でダミーデータを簡単に生成できるパッケージで, 氏名, 企業名, 住所の 3 項目を生成した. Faker からは提供していない部署, 役職の 2 項目に関しては, Wikipedia から部署, 役職の辞書を作成し, その辞書からランダムに採用した. メールアドレスに関して, @ の前の部分は Faker から生成した氏名を pykakasi⁹⁾ でローマ字に変換し, "{first name}.{last name}" とした. また, ドメイン部分は example の部分を Faker から取得できる会社名を pykakasi によってローマ字に変換して置換した. さらに, TEL, FAX, MOBILE の 3 項目に関して, 日本の電話番号のルールから割り当てられない数字のパターンを考慮し, 桁数を揃えてランダムに生成した. 具体的には, 固定電話においては 1 から始まる市内局番は割り当てられていないため, 03-1xxx-xxxx を TEL, FAX で使用し, 携帯電話番号においては 090 の後に 0 から始まる番号 (090-0xxx-xxxx) は割り当てられていないことから MOBILE で使用した¹⁰⁾.

A.2 倫理的配慮事項

本研究のデータ生成において, 他者の著作権およびプライバシーを侵害しないよう以下の措置を講じた. まず, 注入する属性情報は CC BY-SA 4.0 ライセンス下の Wikipedia¹¹⁾ および MIT ライセンス下の Faker¹²⁾ から取得しており, 公開・再利用が適法に許可された情報のみを用いている. また, 生成されたメールに含まれる氏名・会社名・住所はすべて擬似的に生成された架空の情報であり, メールアドレスも先の氏名や会社名をもとに容易かつ疑似的に生成できる架空の情報である. これにより, 本手法は言語処理学会の生成 AI 利用指針および本大会の投稿規定に準拠している.

A.3 P-BEE のシステムプロンプト

あなたは会社員、{sender_address} さんです。
以下の条件をもとに、{receivers_address} 宛に送信する {content} に
関するビジネスメールを {num_words} 文字で作成してください。

9) <https://github.com/miurahr/pykakasi>

10) https://www.soumu.go.jp/main_sosiki/joho_tsusin/top/tel_number/number_shitei.html

11) https://en.wikipedia.org/wiki/Wikipedia:Reusing_Wikipedia_content

12) <https://github.com/joke2k/faker/blob/master/LICENSE.txt>

出力ルール

- 件名 (Subject) を 1 行目に生成してください。
- 形式: 「件名: ...」
- 内容は本文に沿った自然な日本語タイトルにしてください。
- その後に本文を出力してください。- 出力はプレーンテキストのみ。出力はプレーンテキストのみ (引用符や余計な説明を含めない)。

文体と内容

- 実際の日本企業でのやり取りのように, ややラフで自然な文体にしてください。
- 過剰にかしこまった文章ではなく, 自然な敬語・丁寧語を使ってください。
- 架空のプロダクト名やサービス, URL, 添付ファイル (PDF・画像など) に触れるなど, 具体的な業務メールとしてリアリティを持たせてください。
- 適度に改行を入れ, 人間らしいメールのリズムで書いてください。
- 「〇〇様, お世話になっております。××株式会社のです。」などの挨拶や受信者への宛名や送信者自身の名乗りを記述してください。

出力形式

件名: ...
本文...

A.4 質問応答モデルのプロンプト

送信者のメールヘッダとメール本文を使って, メール送信者の氏名を応答に出力してください。
以下の制約に従って出力してください。
- 送信者のメールヘッダもしくはメール本文に存在する文字列のみ出力してください。
- 送信者のメールヘッダもしくはメール本文にメール送信者の氏名が存在しない場合は, 空白を出力してください。
- 他人の情報は含めず, 送信者の情報のみを出力してください。
- 送信者の組織名や部署や役職は出力せず, 送信者の氏名だけを出力してください。