

上場会社の有価証券報告書における 研究開発活動のトピックの長期変化分析

土井 惟成

株式会社日本取引所グループ 東京大学大学院 工学系研究科

n-doi@jpx.co.jp

概要

本研究では、上場会社の有価証券報告書における「研究開発活動」欄のテキストを用いて、上場会社の研究開発活動の内容の経時的な変化を定量的に把握することを目指す。具体的には、2013年度から2024年度までの年次別のコーパスを対象に、文書属性を取り込める Structural Topic Model を適用し、研究開発活動の主要トピックとその出現比率の推移を分析する。実験の結果、対象期間における、生成 AI 関連およびサステナビリティ関連のトピックの拡大と、業種間のトピックの推移の差異が確認された。

1 はじめに

企業の研究開発は、企業の将来の収益性を左右する重要な要素であり、研究開発費の総額や研究開発の内容等に関する情報は、企業の経営方針や将来の収益予測に関する重要な投資情報として位置づけられている [1]。日本では、法定開示書類である有価証券報告書を通じて、上場会社による研究開発活動について開示が行われ、投資家は EDINET を通じてそれらを閲覧することが可能である。しかしながら、有価証券報告書における研究開発活動の記述は、研究開発費を除くと定性的要素が大きく、大量の上場会社についてその内容を長期にわたり体系的に比較することは容易ではない。

一方で、開示資料の電子化と自然言語処理技術の進展を背景に、上場会社の開示資料におけるテキストを大規模に扱うアプローチが活用されている [2, 3, 4, 5]。そのなかでも、語彙が多様な領域のテキストでは、キーワード集計に代えて、文書集合から潜在テーマを抽出するトピックモデルが有用であると考えられる。ただし、代表的なトピックモデルの手法である Latent Dirichlet Allocation (LDA) [6] では、年次や業種といった文書属性を、モデル内で直

接扱えないため、属性別の比較は推定後の集計に依存しやすい。これに対して、Structural Topic Model (STM) [7, 8, 9] は、トピック比率を文書のメタデータの関数としてモデル化できるため、年次や業種を踏まえたトピックの変化を確認できる。

そこで、本研究では、上場会社の有価証券報告書における「研究開発活動」欄のテキストを対象に、STM によって、研究開発活動の主要トピックとその出現比率の推移を、2013年度から2024年度までの長期間にわたって分析する。これにより、研究開発のトピックの推移を、業種の差異を考慮しつつ捉えることを目指す。

本稿の構成は次のとおりである。第2章では関連研究を述べ、第3章では本研究で使用するデータセットをまとめる。第4章では分析手法について述べ、第5章で実験について述べた後、第6章で結論と今後の課題を述べる。

2 関連研究

2.1 開示資料における自然言語処理

開示資料の電子化により、自然言語処理の技術を応用することで、上場会社の開示資料を対象として計量的に扱うアプローチが発展している。米国の法定開示書類である Form 10-K に関しては、テキストの指標の変化と潜在トピックの推移を同時に捉え、開示資料のテキストが経済的に意味のあるシグナルを含むことを示した研究がある [3]。上場会社の開示資料におけるテキストの数量化の手法としては、可読性、トーン、センチメント等に基づく指標が広く用いられるが、汎用辞書の金融文脈での誤分類が問題となりうるため、金融テキストに特化した辞書整備が進められてきた [10, 11]。一方で、テキストの内容を評価する場合は、前述の指標や単語頻度の集計のみでは体系的に把握しにくい。この点で、文

書集合から潜在テーマを抽出して内容を把握できるトピックモデルが有用となる場面がある [4, 6].

日本の有価証券報告書におけるテキストについても、テキストマイニングを活用した研究 [12, 13] や、テキストの数量化に関する研究 [14] が報告されている。有価証券報告書における特定のセクションを対象としたトピック分析に関する研究としては、サステナビリティ情報 [15], MD&A [16], 事業等のリスク [17, 18] を対象とした研究が挙げられる。その一方で、有価証券報告書における研究開発活動を対象としたテキストマイニングは限定的である。

2.2 研究開発活動の開示の計量化

米国では、研究開発に関する物語的な開示の量、詳細度、トーン等が、アナリスト行動等と関連することが示されている [19]. また、テキストを用いて企業のイノベーション活動を測定する試みも進み、Form 10-K の潜在トピック分布等からイノベーション度を推定する研究 [20] や、特許を持たない企業も含めてテキストからイノベーション活動を捉える研究 [21] が提案されている。

日本の有価証券報告書における「研究開発活動」欄を対象とした研究として、石光 [22] は、当該テキストの総単語数や語彙多様性を指標化し、開示内容の量的特性を検討している。しかし、研究開発活動のテキストを対象に、企業横断かつ長期で内容を定量化し、経年で変化を追跡する研究は限定的である。

2.3 トピックモデル

トピックモデルの代表例である LDA [6] は、文書集合から潜在トピックを抽出して内容構造を要約できる一方で、年次や業種といった文書属性をモデル内で直接扱うことは難しい。トピックモデルの手法としては、時系列変化を明示的に表現する拡張として Dynamic Topic Model (DTM) [23] が提案されており、さらに Correlated Topic Model (CTM) [24], Hierarchical Dirichlet Processes (HDP) [25], 教師ありトピックモデル (Labeled LDA) [26] 等、多様な派生手法が存在する。本研究では、年次と業種という複数の文書属性を同一枠組みで組み込み、トピック比率をメタデータの関数として推定できる点から、STM [7, 8, 9] を採用する。

3 データセット

本研究では、日本国内の証券取引所に上場する内国会社のうち、2013 年度から 2024 年度（具体的には、2013 年 12 月期から 2025 年 3 月期）を対象に、各年 6 月 30 日まで（2020 年度については 9 月 30 日まで¹⁾）に提出された最新の有価証券報告書を、EDINET API で収集した。なお、分析に当たり、2013 年 12 月期から 2014 年 3 月期までを 2013 年度、それ以降は N 年 4 月期から N+1 年 3 月期までを N 年度として分類した。例えば、2024 年度の有価証券報告書に関する集計は、2024 年 4 月期から 2025 年 3 月期までの有価証券報告書に基づくものである。収集した有価証券報告書から、「研究開発活動」欄のテキストを抽出した。このとき、研究開発活動を行っていない上場会社を除外する目的で、当該テキストが 100 文字以下の文書を分析対象から除外した²⁾。

この結果、研究開発活動欄を抽出できた文書は全期間で 35,066 件であり、そのうち 100 文字超（分析対象）は 23,671 件となった。また、それぞれの上場会社に対しては、東証 33 業種分類に基づく業種を付与した。データセットの記述統計を付録に示す。

4 分析手法

4.1 前処理とトークナイズ

上場会社の開示資料には、定型的な注意書きや一般的な説明表現が繰り返し含まれることがあるため、定型的な表現を予め除外する。また、前処理として、Unicode 正規化による表記ゆれの正規化を行う。形態素解析には SudachiPy [27, 28] を用い、分割モードを選択して技術用語や複合語が過度に分割されないようにする。なお、Sudachi の分割モードは Mode C を使用し、名詞のみを採用した。

4.2 Structural Topic Model

STM は、文書 d を K 個の潜在トピックの混合とみなし、文書ごとのトピック比率 $\theta_d = (\theta_{d1}, \dots, \theta_{dK})$ を推定する手法である。STM の特徴は、トピック比率 (prevalence) を文書属性の関数として同時にモデ

1) 2020 年度は、新型コロナウイルス感染症に関連し、有価証券報告書の提出期限が 9 月 30 日まで延長されたことを考慮した。

2) 研究開発活動について有価証券報告書に特段の記述を行わない上場会社は、「研究開発活動」の欄に「該当事項はありません」といった短文のテキストを記載することが多い。

ル化できる点にある。本研究では、年次と業種の影響を同時に捉えるため、次のとおり設定した。

$$\eta_d = \mathbf{X}_d \gamma + \varepsilon_d \quad (1)$$

$$\theta_d = \text{softmax}(\eta_d) \quad (2)$$

$$\theta_{dk} = \frac{\exp(\eta_{dk})}{\sum_{j=1}^K \exp(\eta_{dj})} \quad (3)$$

ここで、 \mathbf{X}_d には年次の非線形効果を表すスプライン基底 $s(\text{fiscal_year}_d)$ と業種ダミーを含め、推定時は $\text{prevalence} \sim s(\text{fiscal_year}) + \text{industry}$ として指定する。初期値は Spectral を用い、推定後の $\hat{\theta}_{d,k}$ を年次・業種別に集計して推移を分析する。

4.3 トピック数 K の選択

トピック数 K はモデルの表現力と解釈可能性に影響するため、複数の候補で推定し診断指標を比較した。本研究では、 $K \in \{20, 40, 60, 80, 100\}$ について、stm パッケージの searchK 関数を使って算出することができる、Held-out likelihood, Residuals, Semantic coherence, Exclusivity を確認した。具体的には、これらを $[0, 1]$ に正規化した上で、Held-out likelihood, Semantic coherence, Exclusivity を加点し Residuals を減点する合成スコアを定義し、最大となる K を採用した。この結果、合成スコアが最大となったトピック数として $K = 100$ を採用した。

4.4 年次推移・業種差シフトの指標

推定された文書トピック比率を $\theta_{d,k}$ (文書 d のトピック k 比率) とし、年度 t の年次平均を

$$\bar{\theta}_{t,k} = \frac{1}{N_t} \sum_{d:\text{fiscal_year}(d)=t} \theta_{d,k} \quad (4)$$

と定義した。業種 s に限定した年次平均は

$$\bar{\theta}_{t,s,k} = \frac{1}{N_{t,s}} \sum_{d:\text{fiscal_year}(d)=t, \text{industry}(d)=s} \theta_{d,k} \quad (5)$$

とした。ここで N_t は年度 t に属する対象文書数、 $N_{t,s}$ は年度 t かつ業種 s に属する対象文書数である。

5 実験

5.1 主要トピックの選定基準

本研究では、 $K = 100$ トピックの全体推定を行った上で、本文で詳細に議論する「主要トピック」を、恣意性を抑えるために、次の手続きで選定した。

まず、長期変化を捉える目的に照らし、期間の始点と終点における年次平均トピック比率の増加倍率

$$G_k = \bar{\theta}_{2024,k} / \bar{\theta}_{2013,k} \quad (6)$$

を算出し、 G_k が大きいトピックを「構造変化候補」として抽出した。

次に、抽出候補のうち、確率上位語および FREX³⁾ の上位語が、技術領域として解釈可能であること(定型句・一般語や企業固有名詞のみで支配されないこと)を確認し、ラベル付けを行った。なお、本研究で議論する主要トピックは、複数業種で観測されるものを優先して選定した。

5.2 主要トピックの同定とラベル付け

主要トピックの比率の推移を表 1 に示す。なお、主要トピックの代表語の内訳の抜粋は付録に示す。

Topic 33 は「生成 AI」「LLM」「大規模言語モデル」等が上位語に現れ、2020 年代に顕在化した生成 AI・大規模言語モデル関連の研究開発を反映するトピックとして解釈できる。一方で、FREX 上位語には「暗号」「テキスト」「量子」等、生成 AI に必ずしも限定されない語彙も含まれる。これに起因して、生成 AI の登場以前である 2010 年代から、Topic 33 が一定比率で観測されたものと推察される。

Topic 62 は「機械学習」「ディープラーニング」「画像」「認識」等を含み、生成 AI の登場以前に特に普及した AI 技術を表すトピックと考えられる。

Topic 37 は「クラウド」「IoT」「プラットフォーム」「セキュリティ」等を含み、DX や IT 基盤整備に関するトピックとして位置付けられる。

Topic 52 は「環境」「カーボンニュートラル」「CO2」等が上位語に現れ、サステナビリティを中心としていると考えられる。

5.3 年次推移の結果

図 1 は、主要トピックの年次平均トピック比率の推移である。網掛けは、年度 t に属する文書 d のトピック k の推定比率 $\theta_{d,k}$ を年度で集計した平均 $\bar{\theta}_{t,k}$ に対する 95% 信頼区間であり、それぞれの標本標準偏差 $s_{t,k}$ と標本数 N_t から標準誤差 $\text{SE}_{t,k} = s_{t,k} / \sqrt{N_t}$ を求め、 $\bar{\theta}_{t,k} \pm t_{0.975, N_t-1} \text{SE}_{t,k}$ (t 分布, 自由度 $N_t - 1$) として算出した。なお、この信頼区間には、STM 推

3) STM では、トピックごとの特徴的単語として、各トピック中の単語ごとに FREX という値が算出される。FREX は、トピック内における単語の出現頻度と他のトピックと比較した排他性を踏まえて計算される。

表1 主要トピックのトピック比率の推移 (2013年度→2024年度)

| Topic | ラベル | 2013年度 (%) | 2024年度 (%) | 増加倍率 G_k | 平均 (%) |
|-------|----------|------------|------------|------------|--------|
| 33 | 生成 AI 関係 | 0.17 | 1.43 | 8.3 | 0.56 |
| 62 | AI/機械学習 | 0.46 | 2.50 | 5.5 | 1.74 |
| 37 | DX/クラウド | 1.31 | 2.21 | 1.7 | 1.86 |
| 52 | サステナビリティ | 0.30 | 3.34 | 11.2 | 1.40 |

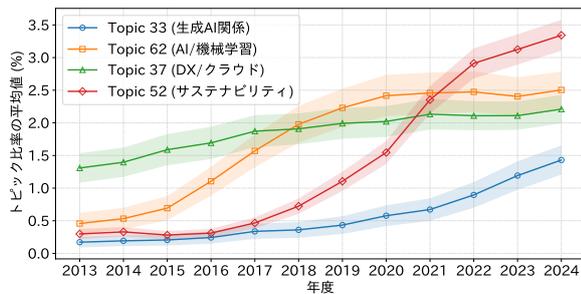


図1 主要トピックの年次平均トピック比率の推移

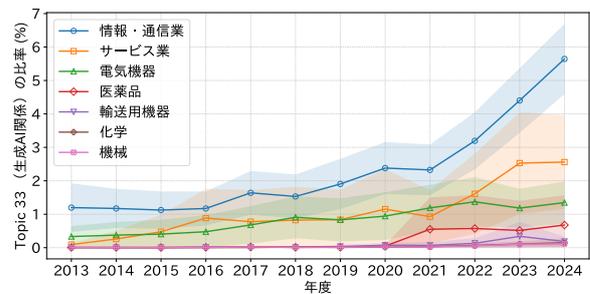


図2 Topic 33 (生成 AI 関係) の業種別の推移

定自体の不確実性は含まない。

Topic 33 (生成 AI 関係) は、2013 年度の 0.17% から 2024 年度の 1.43% へと増加しており、期間全体で約 8.3 倍であった。ただし、本トピックは「AI」「暗号」「テキスト」等の隣接領域の語彙も含んでいるため、2010 年代から観測される一定の出現は、生成 AI に特有なものではない、AI 技術一般や、暗号、自然言語処理といった技術への言及を反映している可能性がある。その上で、2022 年度以降の伸びの大きさは、生成 AI により強く結びつく語彙 (例として、生成 AI, LLM, 大規模言語モデル) が研究開発活動の記述において普及したことを示唆する。

一方で、Topic 62 (AI/機械学習) は、2010 年代後半に増加した後、2020 年以降はおおむね 2.4–2.5% で推移しており、従来型の AI の応用が「成熟した基盤技術」として定着した可能性がある。Topic 37 (DX/クラウド) は緩やかな増加に留まり、Topic 52 (サステナビリティ) は 2018 年度以降に急増し、2024 年度には 3.34% に達する。

以上を踏まえると、本データセットでは、生成 AI とサステナビリティの二つが、2020 年代における構造変化として観測される。

5.4 業種別の差異

図 2 は、Topic 33 (生成 AI 関係) の業種別推移の抜粋である。網掛けは、図 1 と同様の手法で算出した、95% 信頼区間である。

情報・通信業では Topic 33 が一貫して高く、2024

年度には 5.64% に達する一方、機械・化学では 2024 年度でも 0.1% 台に留まる。電気機器やサービス業は情報・通信業ほどではないが上昇が明瞭であり、医薬品も 2021 年度以降に増加が観測される。これらは、生成 AI が様々な企業で活用されているが、その傾向は業種により異なることを示す。

なお、情報・通信業で Topic 33 が早期から一定水準で観測される点は、生成 AI の登場以前から暗号関係やテキスト解析等の隣接領域の研究開発が相対的に活発であった可能性とも整合的である。2022 年度以降の上昇局面では、生成 AI 関係の語彙の増加がこれに上乗せされた可能性がある。

6 おわりに

本研究は、有価証券報告書の「研究開発活動」テキストに STM を適用し、年次と業種別のトピック推移を推定した。2013 年度から 2024 年度の分析により、生成 AI 関係のトピックの急速な拡大と、サステナビリティのトピックの継続的な上昇という、2020 年代に顕在化した二つの構造変化が確認された。また、生成 AI 関係は情報・通信業で高水準で推移する一方、機械・化学では相対的に低位に留まる等、生成 AI の普及が業種で異なることを示した。

今後の課題として、企業規模、研究開発費、財務指標等の追加共変量を導入した上での要因分解や、他のセクションのテキストと横断した分析が挙げられる。これらを通じて、研究開発開示の変化をより具体的に捉えることを目指す。

参考文献

- [1] 企業会計審議会. 研究開発費等に係る会計基準の設定に関する意見書. https://www.fsa.go.jp/p_mof/singikai/kaikai/tosin/1a909e1.htm (accessed on 2026-01-09), 1998.
- [2] Feng Li. The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. **Journal of Accounting Research**, Vol. 48, No. 5, pp. 1049–1102, 2010.
- [3] Travis Dyer, Mark Lang, and Lorien Stice-Lawrence. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. **Journal of Accounting and Economics**, Vol. 64, No. 2, pp. 221–245, 2017.
- [4] Justin Grimmer and Brandon M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. **Political Analysis**, Vol. 21, No. 3, pp. 267–297, 2013.
- [5] Tim Loughran and Bill McDonald. Textual analysis in accounting and finance: A survey. **Journal of Accounting Research**, Vol. 54, No. 4, pp. 1187–1230, 2016.
- [6] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, Vol. 3, pp. 993–1022, 2003.
- [7] Margaret E. Roberts, Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. Structural topic models for open-ended survey responses. **American Journal of Political Science**, Vol. 58, No. 4, pp. 1064–1082, 2014.
- [8] Margaret E. Roberts, Brandon M. Stewart, and Edoardo M. Airoldi. A model of text for experimentation in the social sciences. **Journal of the American Statistical Association**, Vol. 111, No. 515, pp. 988–1003, 2016.
- [9] Margaret E. Roberts, Brandon M. Stewart, and Dustin Tingley. stm: An R package for structural topic models. **Journal of Statistical Software**, Vol. 91, No. 2, p. 1–40, 2019.
- [10] Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. **The Journal of Finance**, Vol. 66, No. 1, pp. 35–65, 2011.
- [11] Tim Loughran and Bill McDonald. The use of word lists in textual analysis. **Journal of Behavioral Finance**, Vol. 16, No. 1, pp. 1–11, 2015.
- [12] 近藤隆史, 石光裕. 有価証券報告書へのテキストマイニングの適用に関する文献レビュー. 京都マネジメント・レビュー, Vol. 42, pp. 93–118, 2023.
- [13] Kenichi Yazawa, Takeaki Ito, and Hyonok Kim. テキストマイニングを用いた MD&A, リスク, ガバナンス情報の分析. 青山経営論集= Aoyama journal of business/青山学院大学経営学会 編, Vol. 56, No. 1, pp. 59–84, 2021.
- [14] 加藤大輔, 五島圭一. 有価証券報告書のテキスト分析: 経営者による将来見通しの開示と将来業績. 金融研究, Vol. 40, No. 3, pp. 45–75, 2021.
- [15] 中田理恵, 鈴木裕. 機械学習を用いた有価証券報告書のサステナビリティ情報の可視化 構造的トピックモデル (Structural Topic Model) による記述情報把握. https://www.dir.co.jp/report/research/capital-mkt/esg/20231006_023747.pdf (accessed on 2026-01-09), 2023.
- [16] 中野貴之, 五十嵐未来, 湯浅大地. MD&A 情報における「トピック」の分析. 証券アナリストジャーナル, Vol. 60, No. 10, pp. 27–35, 2022.
- [17] 土屋和之. 事業等のリスクの分析: 記載内容のトピックにもとづくアプローチ. 千葉商大論叢, Vol. 57, No. 3, pp. 185–197, 2020.
- [18] 土屋和之. 事業等のリスクの重要度の分析. 千葉商大論叢, Vol. 58, No. 3, pp. 151–166, 2021.
- [19] Kenneth J. Merkley. Narrative disclosure and earnings performance: Evidence from R&D disclosures. **The Accounting Review**, Vol. 89, No. 2, pp. 725–757, 2014.
- [20] Essi Vesisalmi, Mikko Ranta, Mika Ylinen, and Marko Järvenpää. Using machine learning and 10-K filings to measure innovation. **Accounting & Finance**, Vol. 64, No. 4, pp. 3211–3239, 2024.
- [21] Gustaf Bellstam, Sanjai Bhagat, and J. Anthony Cookson. A text-based analysis of corporate innovation. **Management Science**, Vol. 67, No. 7, pp. 4004–4031, 2021.
- [22] 石光裕. 有価証券報告書における研究開発活動の記載内容の分析. PhD thesis, KYOTO SANGYO UNIVERSITY, 2018.
- [23] David M. Blei and John D. Lafferty. Dynamic topic models. In **Proceedings of the 23rd International Conference on Machine Learning (ICML '06)**, pp. 113–120, 2006.
- [24] David M. Blei and John D. Lafferty. A correlated topic model of science. **The Annals of Applied Statistics**, Vol. 1, No. 1, pp. 17–35, 2007.
- [25] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. **Journal of the American Statistical Association**, Vol. 101, No. 476, pp. 1566–1581, 2006.
- [26] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D. Manning. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In Philipp Koehn and Rada Mihalcea, editors, **Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing**, pp. 248–256, Singapore, August 2009. Association for Computational Linguistics.
- [27] Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, Yoshitaka Uchida, and Yuji Matsumoto. Sudachi: a Japanese tokenizer for business. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, H el ene Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**, Miyazaki, Japan, 2018. European Language Resources Association (ELRA).
- [28] Works Applications. SudachiPy: Python version of Sudachi, the Japanese Morphological Analyzer. <https://pypi.org/project/SudachiPy/>, 2025. Version 0.6.10 (released 2025-01-10).

A データセットの記述統計

表 2 は本研究で使用したデータセットの基本統計量を示す。

表 2 本研究で使用したデータセットの基本統計量 (p はパーセンタイルを示す)

| 年度 | 会社数 | | | テキスト長 (100 文字超のみ) | | | | |
|------|-------|--------|---------------|-------------------|--------|-------|-------|--------|
| | 上場会社数 | テキストあり | 100 文字超テキストあり | 平均 | 標準偏差 | 10p | 中央値 | 90p |
| 2013 | 2997 | 1612 | 1577 | 1337.3 | 1245.7 | 268.8 | 982 | 2757.0 |
| 2014 | 3117 | 1866 | 1822 | 1282.6 | 1225.4 | 246.1 | 961 | 2699.5 |
| 2015 | 3219 | 1897 | 1852 | 1281.9 | 1253.8 | 243.1 | 934 | 2693.9 |
| 2016 | 3300 | 1922 | 1874 | 1295.3 | 1316.7 | 247.0 | 924 | 2708.5 |
| 2017 | 3591 | 2067 | 2010 | 1289.7 | 1296.4 | 237.8 | 922 | 2697.0 |
| 2018 | 3699 | 2976 | 2049 | 1323.8 | 1599.9 | 236.4 | 922 | 2826.0 |
| 2019 | 3749 | 3734 | 2068 | 1353.0 | 1775.3 | 238.0 | 933 | 2823.3 |
| 2020 | 3638 | 3626 | 1979 | 1373.1 | 1958.4 | 229.0 | 930 | 2796.2 |
| 2021 | 3832 | 3811 | 2099 | 1409.9 | 2038.9 | 228.8 | 913 | 2911.8 |
| 2022 | 3866 | 3844 | 2107 | 1416.6 | 1947.2 | 221.6 | 911 | 3041.6 |
| 2023 | 3882 | 3862 | 2114 | 1449.2 | 2044.1 | 222.0 | 885.5 | 3175.5 |
| 2024 | 3893 | 3849 | 2120 | 1485.5 | 2017.4 | 218.0 | 902 | 3264.5 |

B 主要トピックの代表語

表 3 は主要トピックの代表語 (確率上位語および FREX 上位語) を示す。

表 3 主要トピックの代表語 (確率上位語および FREX 上位語)

| Topic | 解釈ラベル | 確率上位の 12 語 | FREX 上位の 12 語 |
|-------|----------|--|--|
| 33 | 生成 AI 関係 | データ, AI, 生成 AI, 活用, 予測, 量子, モデル, 学習, 開始, 分析, 提供, 実証 | 生成 AI, 量子, NEC, LLM, RS, 暗号, メタバース, テキスト, 仮想空間, 大規模言語モデル, アナリティクス, 行動 |
| 62 | AI/機械学習 | 技術, AI, 画像, 活用, 認識, IoT, 解析, データ, 処理, 自動, 情報, システム | AI, 人工知能, 機械学習, 認識, 画像, ディープラーニング, 深層学習, IoT, アルゴリズム, ビッグデータ, 地図, 自然言語処理 |
| 37 | DX/クラウド | ソリューション, システム, サービス, ビジネス, 技術, 提供, IoT, 活用, 情報, 分野, IT, クラウド | ソリューション, ビジネス, IT, クラウド, IoT, ICT, サービス, クラウドサービス, プラットフォーム, セキュリティー, 金融, ネットワーク |
| 52 | サステナビリティ | 社会, 貢献, 環境, 技術, 活用, 削減, 課題, カーボンニュートラル, 提供, CO2, 持続, エネルギー | カーボンニュートラル, 社会, 脱炭素, 持続, モビリティ, 排出量, 貢献, DX, CO2, 共創, 解決, サステナブル |