

2025年8月1日 全10頁

# 機械学習による有価証券報告書(2025年3月期)の人的資本開示の可視化

経営戦略と人事戦略の連動や、指標及び目標の設定に課題

金融調査部 主席研究員 太田珠美  
経済調査部 研究員 中田理恵  
経済調査部 主任研究員 新田堯之

## [要約]

- 2023年3月期から有価証券報告書の人的資本に関する情報開示が義務化された。具体的には、人材の多様性の確保を含む人材育成の方針や社内環境整備方針及び当該方針に関する指標の内容等について、「サステナビリティに関する考え方及び取組」の欄に「戦略」と「指標及び目標」として記載する必要がある。
- 上場企業が公表した2025年3月期の有価証券報告書に基づき、構造的トピックモデルを用いて人的資本に関する記載内容を分析したところ、従来と同様に「戦略」には「人材の育成」や「キャリア形成」が、「指標及び目標」には「人材の多様性」や「ジェンダー平等」が多く記載されていることが確認できた。「指標及び目標」は「戦略」に記載した方針に基づいて記述することとなっているが、重きがおかれているトピックは異なる状況にある。
- 金融庁は、有価証券報告書の「サステナビリティに関する考え方及び取組」に関して「企業の中長期的な持続可能性に関する事項について、経営方針・経営戦略等との整合性を意識して説明するもの」と位置付けている。今回、「経営方針、経営環境及び対処すべき課題等」や「事業等のリスク」の記述内容も分析したところ、それぞれに記載されている人的資本に関する記述と、「サステナビリティに関する考え方及び取組」の人的資本の記述で重きが置かれているトピックスに若干の差異が見られた。自社の経営戦略と人事戦略の紐づけ、その取り組み状況を把握するのに適切なKPI(重要業績評価指標)の設定が求められる。

## 1. 有価証券報告書における人的資本開示

2023年の「企業内容等の開示に関する内閣府令」等の改正により、2023年3月期の有価証券報告書から「サステナビリティに関する考え方及び取組」の記載欄が新設された<sup>1</sup>。当該記載欄にはサステナビリティに関する「ガバナンス」、「戦略」、「リスク管理」及び「指標及び目標」を記載することが求められている。個別テーマに関しては各社の判断に応じて記載することになっているが、人的資本については全社記載することが求められている。具体的には「人材育成方針」や「社内環境整備方針」を「戦略」に、当該方針に関する指標の内容や、当該指標による目標・実績を、「指標及び目標」に記載する。

本稿では「サステナビリティに関する考え方及び取組」の記述内容を、機械学習による自然言語処理の一手法である構造的トピックモデル（以下、トピックモデルとする。詳細は巻末補論を参照）を用いて視覚的に把握することを試みる。サステナビリティに関する記載内容をトピック（テーマ）別に分類し、どのようなトピックを重点的に記載しているか、またそのトピック内にてどのような単語が多用されているかを可視化する。その上で、「戦略」と「指標及び目標」の人的資本に関する記載内容の整合性を確認していく。

なお、企業によっては人的資本に関する情報を「サステナビリティに関する考え方及び取組」だけでなく「経営方針、経営環境及び対処すべき課題等」や「事業等のリスク」にも記載しているケースがある。今回の分析では「サステナビリティに関する考え方及び取組」の記載欄だけでなく、「経営方針、経営環境及び対処すべき課題等」や「事業等のリスク」の欄に関しても同様の分析を試みた。

## 2. 人的資本開示の可視化

### （1）データ、分析の手法

本稿の分析対象は、EDINET から取得した有価証券報告書に記載された文章および表のテキストである。具体的には、①経営方針、経営環境及び対処すべき課題等、②サステナビリティに関する考え方及び取組、③事業等のリスクの3項目を対象とする（図表1）。②はさらに人的資本に関する戦略と指標及び目標を分析対象としている。後述のトピックモデルの構築には、事業年度が2024年3月期から2025年3月期の有価証券報告書のうち、2025年6月30日までに提出されたものを使用した（データの処理等については巻末補論を参照）<sup>2</sup>。その上で、構築したモデルを2025年3月期のみ有価証券報告書に適用し、分析を実施した<sup>3</sup>。本稿の分析対象とな

<sup>1</sup> 制度概要は藤野大輝「開示府令の改正が公布・施行」大和総研レポート（2023年2月7日）を参照。

<sup>2</sup> 今回の分析はEDINETから取得した情報を基に行った。2023年1月の「企業内容等の開示に関する内閣府令」等の改正がEDINETに反映されたのは、2023年12月に公表された「[2024年版EDINETタクソノミ](#)」であり、その適用は2024年3月期の有価証券報告書からである。そのため、今回の分析では2024年3月期より前の有価証券報告書は取り扱わない。<sup>3</sup> なお、2024年3月期～2025年2月期の有価証券報告書のトピック分布についても分析を行ったものの、2025年3月期と概ね同様の分布であったため割愛した。

<sup>3</sup> なお、2024年3月期～2025年2月期の有価証券報告書のトピック分布についても分析を行ったものの、2025年3月期と概ね同様の分布であったため割愛した。

る有価証券報告書数は図表1の通りである。また、1社当たり各項目の平均記載字数は図表2の通りである。

図表1 本稿の分析対象とする有価証券報告書の記載欄（左）と有価証券報告書の数（右）

	プライム	スタンダード	グロース	総計
経営方針	2,777	2,540	762	6,079
サステナビリティに関する考え方	2,793	2,545	772	6,110
ガバナンス	2,647	2,450	745	5,842
戦略	2,448	2,410	745	5,603
戦略の人的資本に関する記述	2,659	2,474	757	5,890
リスク管理	2,615	2,420	734	5,769
指標及び目標	2,429	2,378	726	5,533
指標及び目標の人的資本に関する記述	2,647	2,440	733	5,820
事業等のリスク	2,793	2,548	772	6,113

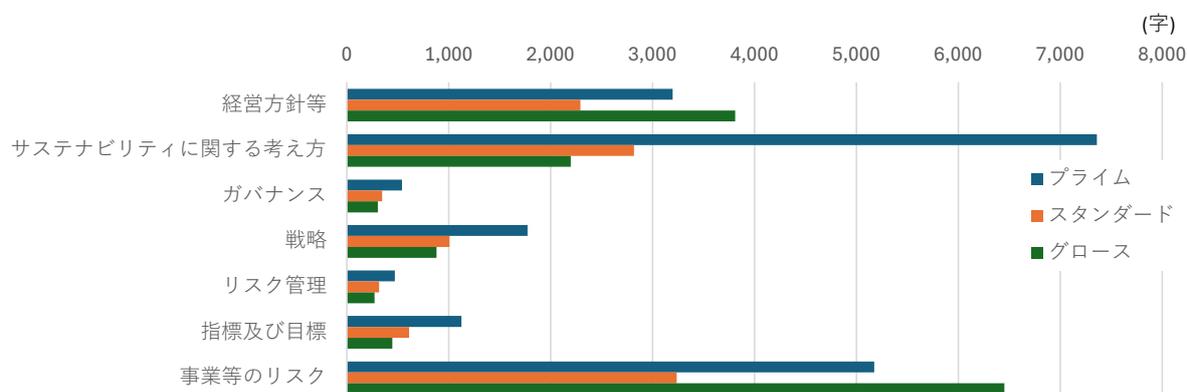
  

第一部【企業情報】 [略]						
第2【事業の状況】						
1【経営方針、経営環境及び対処すべき課題等】 当社グループの経営方針経営環境及び対処すべき課題等は………。						
2【サステナビリティに関する考え方及び取組】 当社グループのサステナビリティに関する考え方及び取組は、次のとおりである。なお、文中の将来に関する事項は、当連結会計年度末現在において当社グループが判断したものである。						
(1) ガバナンス 当社グループは、………。 なお、詳細は、「XXXX」に記載している。						
(2) 戦略 当社グループの………。 また、当社グループにおける、人材の多様性の確保を含む人材の育成に関する方針及び社内環境整備に関する方針は、………。						
(3) リスク管理 当社グループは、………。						
(4) 指標及び目標 当社グループの………。 また、当社グループでは、上記「(2) 戦略」において記載した人材の多様性の確保を含む人材の育成に関する方針及び社内環境整備に関する方針に関する指標として、次の指標を用いている。当該指標に関する目標及び実績は、次のとおりである。						
<table border="1"> <thead> <tr> <th>指標</th> <th>目標</th> <th>実績</th> </tr> </thead> <tbody> <tr> <td>-</td> <td>-</td> <td>-</td> </tr> </tbody> </table>	指標	目標	実績	-	-	-
指標	目標	実績				
-	-	-				
3【事業等のリスク】 当社グループの経営成績及び財政状態に重要な影響を及ぼす可能性のある主要なリスクは………。						

(注) 右図は2024年3月期～2025年2月期と2025年3月期の有価証券報告書を対象

(出所)「企業内容等の開示に関する内閣府令 第三号様式」および各社有価証券報告書より大和総研作成

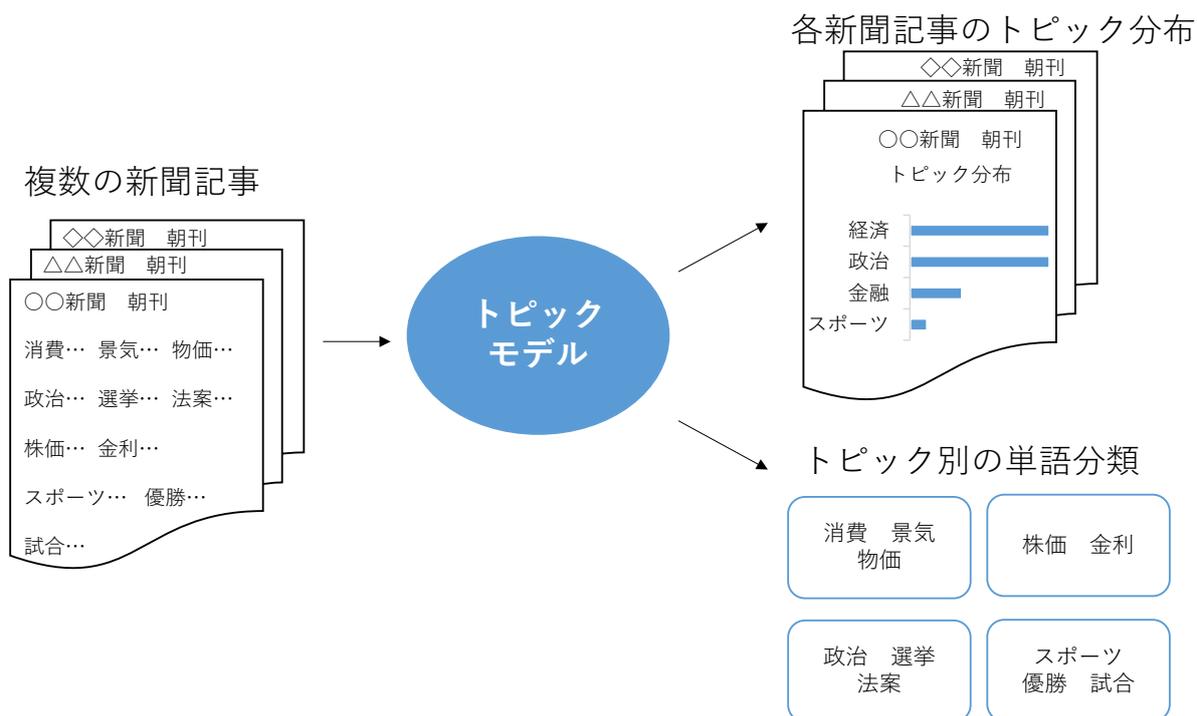
図表2 1社当たり各項目の記載字数（2025年3月期平均値）



(出所) 各社の有価証券報告書より大和総研作成

今回、分析手法として採用するトピックモデルとは、複数の文書のテキストを読み込み、各文書に含まれる単語の出現頻度を元に、どのようなトピック（テーマ）が潜在的に存在するかを推定する手法の一つである。例えば複数のトピックを持つ新聞記事のデータをモデルに読ませることで、各新聞記事におけるトピックの分布の推定やトピック別に単語を分類することができる（図表3）。

図表3 トピックモデル概念図



(出所) 大和総研作成

## (2) 推定結果

### ① トピッカー一覧

トピックの推定結果は図表4の通りである。図表4ではトピックと当該トピックに属する特徴的な単語を並べている。トピックごとの特徴的単語はモデルが算出したFREXという値に基づき上位10単語を抽出した。FREXはトピック内における単語の出現頻度の高さ与其他のトピックと比較した排他性(特定のトピックによく出現し、他のトピックでは出現しにくい)の両者を加味した値である。トピックの名称はトピックモデルが分類した単語に沿って筆者が決定している。

トピッカー一覧を見ると、人材育成、人材の多様性、ジェンダー平等、キャリア形成、健康と安全、人権、リスクマネジメント、ガバナンス体制等が抽出されている。そのほか、金融のように特定業種と関連の強いトピックもみられる。各トピックに属する単語グループの内容を見ると、人による直感的な分類と近い形で分類できているといえよう。

図表4 トピック及びトピックに属する単語一覧

トピック	トピックに属する単語 (FREX 上位 10 単語)
人材育成	人財、育成方針、グローバル、育成、財、変革、海外、プログラム、リーダー、自律
人材育成2	人材、人材育成、能力、整備、国籍、性別、年齢、社内、確保、方針
人材の多様性	外国人、中途採用、女性、登用、女性社員、管理職、職、比率、外国籍、課長
ジェンダー平等	女性労働、労働者、差異、賃金、男女、育児休業、労働、事業年度、男性、号
キャリア形成	休暇、短時間、階層、新入社員、自己啓発、定年、資格、勤務、通信教育、新任
従業員エンゲージメント	報酬、共感、挑戦、人材戦略、人事、well-being、パーパス、組織、バリュー、ポジション
健康と安全	健康経営、優良、法人、株、認定、増進、健康、連続、心身、宣言
柔軟な働き方	ミッション、エンジニア、リモートワーク、オフィス、出社、マネージャー、スタッフ、コンテンツ、学生、書籍
気候変動	SCOPE、t-CO <sub>2</sub> 、排出、温室効果ガス、量、GHG、算定、スコープ、CO <sub>2</sub> 、カテゴリ
気候変動2	炭素税、シナリオ、気温、コスト、上昇、激甚、小、価格、政策、評判
サーキュラーエコノミー (循環経済)	プラスチック、リサイクル、容器、原料、食品、包装、廃棄、再利用、回収、食品ロス
人権	人権、人権尊重、due diligence、step、デューデリジェンス、web サイト、救済、バリュー・チェーン、ガイドライン、生物多様性
リスクマネジメント	リスク管理、リスクマネジメント、識別、リスク、規程、文中、内部統制、判断、監視、策
ガバナンス体制	審議、委員長、取締役会、執行、委員、監督、取締役、委員会、代表取締役社長、会議
金融	当行、投融資、融資、与信、ファイナンス、セクター、行員、銀行、職員、金融
情報セキュリティ	セキュリティ、医療、知的財産、医薬品、データセンター、患者、情報セキュリティ、ソリューション、顧客、テクノロジー
持続可能性	地球環境、遵守、豊か、発展、公正、経営理念、安全、企業活動、貢献、社会
その他	スコア、回答、単体、件数、KPI、度数、延べ、受講、前年度、pt

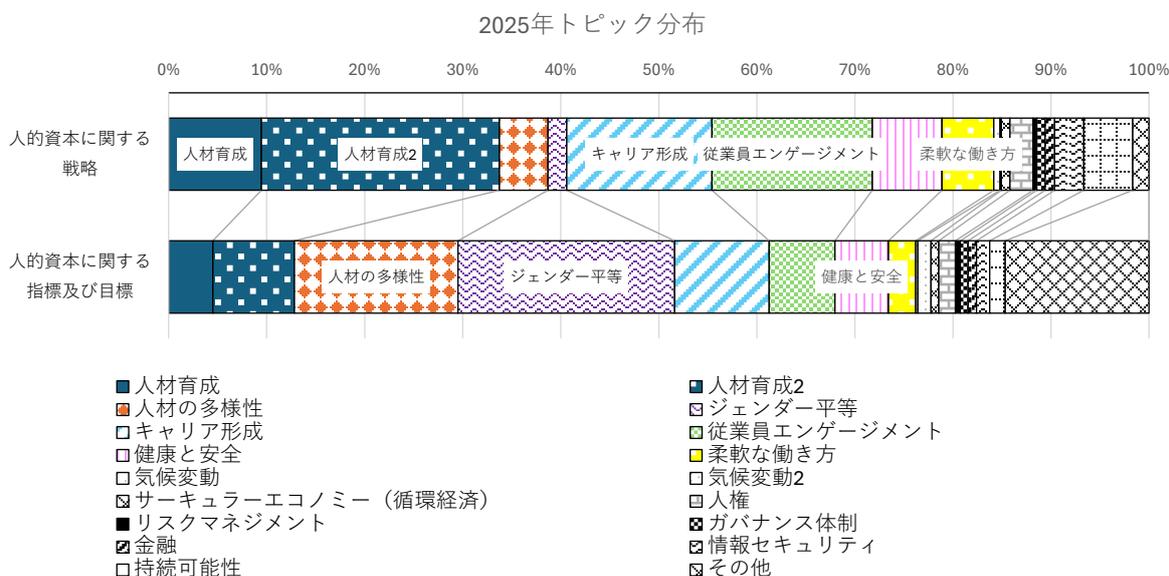
(出所) 各社有価証券報告書より大和総研作成

## ②項目別トピック分布 (人的資本の「戦略」と「指標及び目標」)

図表5が2025年3月期の有価証券報告書の人的資本に関する「戦略」、「指標及び目標」のトピック分布(トピックごとの記述量の割合)を示したものである(先に述べた通りモデルによる推定値であるため、実際の記述の分布と正確に一致しているとは限らない点には留意されたい)。「戦略」に関しては人材育成とキャリア形成、従業員エンゲージメント関連のトピックが多く記載されており、多くの企業において人材育成に関する施策に力を入れている状況が確認された。

一方で、「指標及び目標」に関してはジェンダー平等や人材の多様性が上位に挙がっている。本来「指標及び目標」は「戦略」に関する内容を記載すべきである。「戦略」において人材の多様性やジェンダー平等に関するトピックも一定程度記載されているが、「戦略」においてウエイトが高い人材育成やキャリア形成、従業員エンゲージメントに関する「指標及び目標」の開示には課題がある状況といえそうだ。

図表5 「サステナビリティに関する考え方及び取組」の人的資本に関する「戦略」、「指標及び目標」に関するトピック分布推定結果



(注) 分析対象は2025年3月期の有価証券報告書、サンプルサイズは1050社。

(出所) 各社有価証券報告書を基に大和総研作成

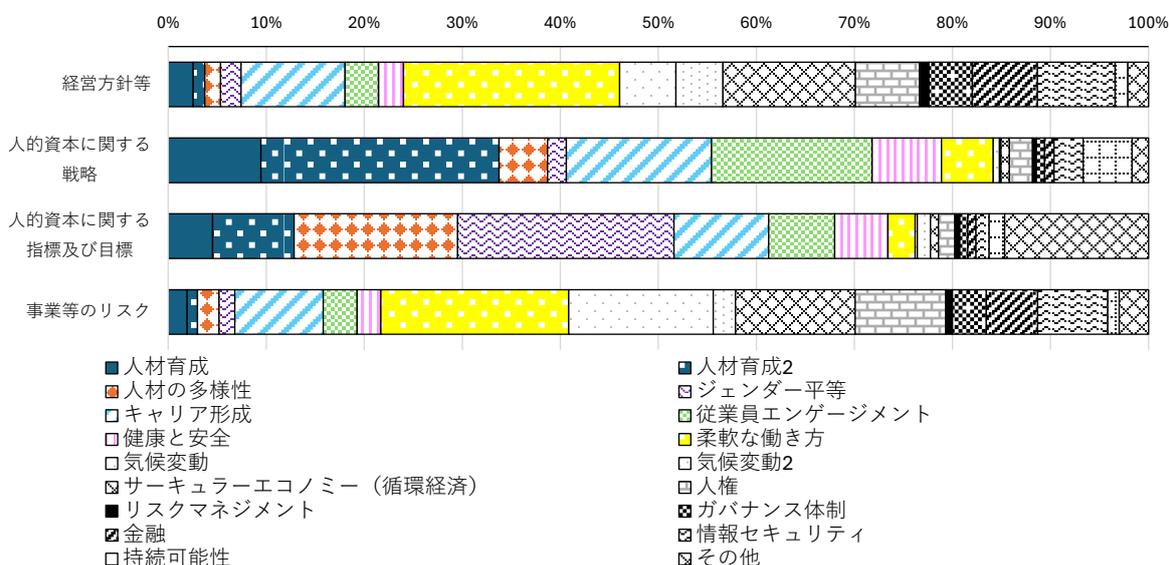
### ③項目別トピック分布（「経営方針、経営環境及び対処すべき課題等」や「事業等のリスク」との比較）

金融庁は2023年の開示府令改正と併せて公表した「記述情報の開示に関する原則（別添）—サステナビリティ情報の開示について—」において、有価証券報告書の「サステナビリティに関する考え方及び取組」に関して「企業の中長期的な持続可能性に関する事項について、経営方針・経営戦略等との整合性を意識して説明するもの」と位置付けている。ここでは、人的資本の「戦略」や「指標及び目標」が、経営方針・経営戦略等との整合性を意識した開示になっているのか、トピック分布を確認していきたい<sup>4</sup>。

分析結果を確認すると「経営方針、経営環境及び対処すべき課題等（以下、経営方針等）」では、柔軟な働き方、サーキュラーエコノミー（循環経済）、キャリア形成、気候変動等が、「事業等のリスク」では、柔軟な働き方、気候変動、サーキュラーエコノミー、キャリア形成、人権等のトピックへの言及が多いことが確認できた（図表6）。人的資本に関連する内容に注目すると、柔軟な働き方、キャリア形成に重きが置かれている様子が窺える。「サステナビリティに関する考え方及び取組」の人的資本に関する記述と比較すると、キャリア形成というトピックが上位である点は整合的といえる。一方で、「サステナビリティに関する考え方及び取組」で上位にあった人材育成や人材の多様性、ジェンダー平等といったトピックは、経営方針等や事業等のリスクではあまり重きを置かれていない可能性が示唆される。

<sup>4</sup> なお、18のトピックに言及している箇所の推計であり、全体の記載割合を示すものではない点に留意されたい。「経営方針、経営環境及び対処すべき課題等」と「事業等のリスク」に関しては、サステナビリティに直接的な記載がないケースも多い。

図表6 経営方針等、事業等のリスクと、人的資本に関する「戦略」、「指標及び目標」に関するトピック分布推定結果（人的資本に関する「戦略」、「指標及び目標」は図表5と同じ）



（注）分析対象は2025年3月期の有価証券報告書、サンプルサイズは1050社。

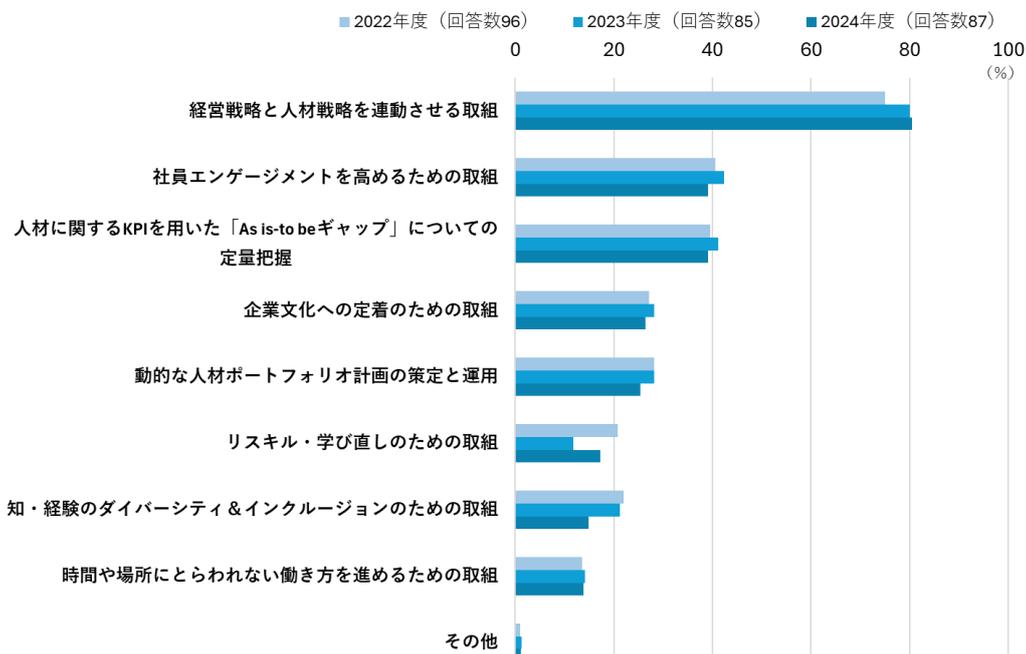
（出所）各社有価証券報告書を基に大和総研作成

### 3. 投資家の期待と現状の開示のギャップ

一般社団法人生命保険協会の「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2024年度版）投資家向けアンケート」によれば「人的資本経営のため、企業が取り組むべき主要テーマをお答えください」という設問に対し、回答者（機関投資家）の8割以上が「経営戦略と人材戦略を連動する取り組み」を選択している（図表7）。また、「経営戦略と人材戦略を連動させるため、企業は現時点で特にどのようなことに取り組むべきとお考えですか」という設問に対しては7割以上の回答者が「人材面における全社的課題の抽出」を挙げており、6割以上の回答者が「人材に関するKPIの設定」を挙げている（図表8）。

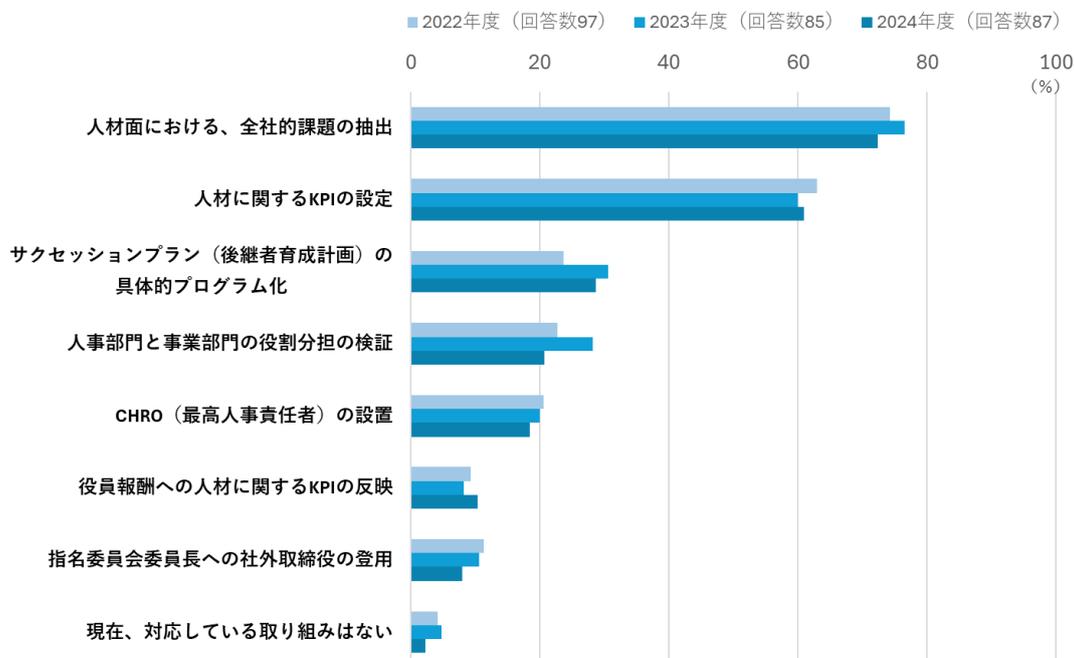
アンケート結果からは、機関投資家が投資先企業に対し、人的資本に関する機会とリスクを改めて精査し、経営戦略と人事戦略を連動させること、そしてその取り組み状況の把握に適切なKPIを設定することを求めている状況が窺える。本稿の分析では、上場企業全体として、有価証券報告書の人的資本開示において①経営方針等と事業等のリスクとの間で若干の相違があること、②（人的資本に関する）「戦略」と「指標及び目標」で重きが置かれている内容が異なることが確認された。自社にとって有価証券報告書に記載すべき重要な情報は何かを精査し、各記載項目に一貫性のある開示が期待される。

図表7 人的資本経営のため企業が取り組むべき主要テーマ（複数回答）



(出所) 生命保険協会「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2022年度版）投資家様向けアンケート」（2023年4月）、「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2023年度版）投資家様向けアンケート」（2024年4月）、「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2024年度版）投資家向けアンケート」（2025年4月）より大和総研作成

図表8 経営戦略と人材戦略を連動させるために取り組むべきこと（複数回答）



(出所) 生命保険協会「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2022年度版）投資家様向けアンケート」（2023年4月）、「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2023年度版）投資家様向けアンケート」（2024年4月）、「企業価値向上に向けた取り組みに関するアンケート集計結果一覧（2024年度版）投資家向けアンケート」（2025年4月）より大和総研作成

## <補論>

### 1. トピックモデルについて

トピックモデルとは複数の文書のテキストを読み込み、各文書に含まれる単語の出現頻度を基に、どのようなトピック（テーマ）が潜在的に存在するかを推定する手法の一つである。例えば複数のトピックを持つ新聞記事のデータをモデルに読ませることで、各新聞記事におけるトピックの分布の推定やトピック別に単語を分類することができる（本文の前掲図表3を参照）。

トピックモデルの代表的手法である「潜在的ディリクレ配分法（LDA）」では、各文書に含まれる単語の出現頻度及び事前に指定したトピック数を基に、各文書を構成するトピックの分布及び各トピックに属する単語の分布を推定する。本稿の分析ではLDAを改良した「構造的トピックモデル（Structural Topic Model）」を用いた。LDAはトピックの推定においてテキスト内の単語出現頻度のみを参照するが、構造的トピックモデルはテキストに加えて、その他の特徴量を元に推定する。このため、より詳細なトピック分類・分析が可能となる。

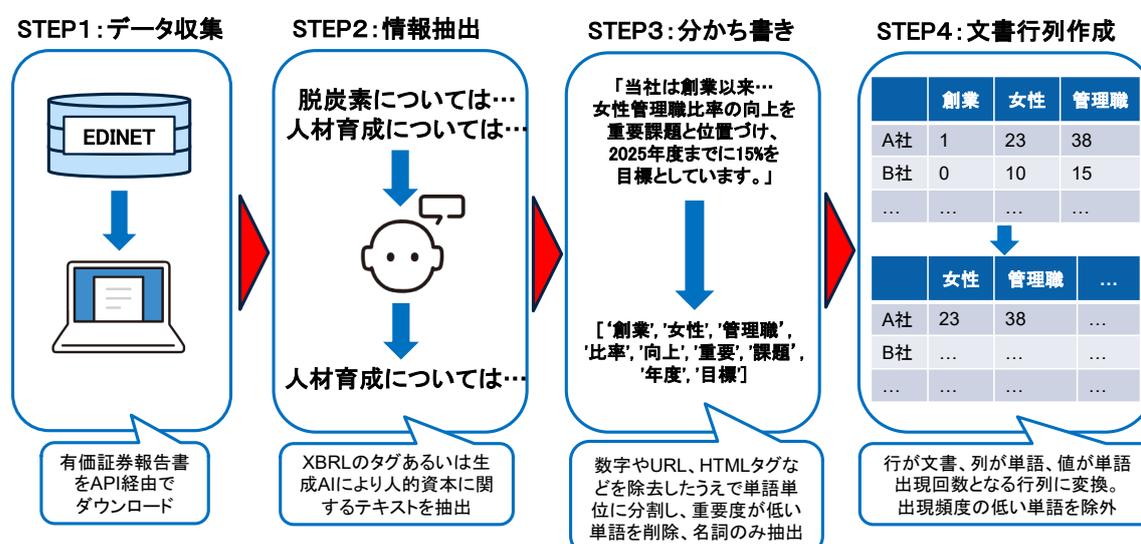
例えば新聞記事のテキストから複数のトピックを推定する際に、その記事を掲載した新聞社はどこであるかといった情報を加味することで、新聞社ごとによく取り上げるトピックの違いに配慮した推定や、同一のトピックでも新聞社により異なる単語を使う傾向にある場合でも一つのトピックとして認識するといったこと等が可能となる。

本稿の分析ではサステナビリティ関連、特に人的資本に関するトピックを重点的に割り出すことを企図し、有価証券報告書のテキストの記載欄を特徴量としている。

### 2. 今回使用したデータの処理について

データは以下の説明通りに取得および処理した（図表9）。

図表9：本稿で使用したデータの取得および処理の概要



（出所）大和総研作成（イラストはソコスト（<https://soco-st.com/>））

まず、EDINET の API 経由で、有価証券報告書に記載された文章および表のテキストをプログラミングにより自動で取得した。EDINET から収集したデータは、以下の前処理を行った。同一企業による同一会計期間の有価証券報告書が複数存在する場合（訂正報告書など）は、提出日が最も新しいものに限定し、旧版は分析対象から除外した。

人的資本に関する記述の抽出は、網羅性を確保するため二段階のアプローチを採用した。まず、EDINET の専用 XBRL タグを用いて抽出を試みた。この方法で情報が取得できない場合に限り、「サステナビリティに関する考え方及び取組」の項目から、生成 AI<sup>5</sup>を用いて補完的に抽出を行った。この補完的処理では、人的資本以外の情報（環境問題など）を除外し、関連記述のみを正確に抽出することを目的とした。生成 AI にはプロンプトで「企業の財務・非財務情報を分析する高度な専門知識を持つ AI アナリスト」という役割を与え、「人的資本」「人材育成」「エンゲージメント」「健康経営」といったキーワードリストを基に、該当文章を原文のまま過不足なく抜き出すよう厳密に指示した。生成 AI による情報の要約や改変を防ぐため、該当箇所がなければ「該当なし」と返すよう設定した。この抽出処理は、抽出元のテキストは存在するものの、まだ抽出結果が記録されていない有価証券報告書のみを対象として実行した。

さらに、トピックモデルの分析精度を高めるため、テキストデータの前処理を行った。処理内容は次の通りである。まず、HTML タグや URL、文字化けといった不要な情報を削除した上で、文章を単語単位に分割（分かち書き・形態素解析）した。分割により単語単位に分かれた文書データに対して、分析上意味を持たない単語（例：「当社」「下記」など）の削除、表記揺れの統一（例：「DX」を「デジタルトランスフォーメーション」に統一）を行った。なお、単語に分割する際の形態素解析エンジンには MeCab、辞書としては固有名詞に強い mecab-ipadic-Neologd を使用した。

構造的トピックモデルの学習用テキストデータには「サステナビリティに関する考え方及び取組」欄のみを利用している。これにより本稿の分析の主眼である人的資本に関するトピックをより詳細に分割させることを企図した処理である。モデルの精度を高めるため、学習用テキストデータ内の出現頻度が 50 文書以下の単語は除去し、単語数が 200 以下の文書は学習対象外としている。トピック判定の特徴量として文書の記載欄を加えている。

また、トピックモデルではトピック分割にあたってトピック数をあらかじめ指定する必要がある。5~40 のトピック数を検討し 18 に決定した。

なお、構造的トピックモデルは学習用テキストデータにない語彙についてトピック判定ができないため、「経営方針、経営環境及び対処すべき課題」欄と「事業等のリスク」欄のトピック判定はサステナビリティに関連した記述に絞ってトピック分布を算出している。

これら一連の処理を実施した後のサンプルサイズは図表 5 及び図表 6 の注釈に記載している。

<sup>5</sup> 具体的には、Google Cloud VertexAI 経由で Gemini2.5flash を使用した。出力を安定させるためにパラメータの temperature（ゼロに近いほど予測可能で正確な出力に近づき、1 に近いほどより多様で想像的な出力に近づく）をゼロに設定、および Seed 値（乱数発生の初期設定）は 42 に固定した。