

# ESG-QA: 日本語 ESG 文書を対象としたマルチモーダル質問応答ベンチマークの構築

佐多亮明<sup>1,2</sup> 赤部晃一<sup>1</sup> 神田峻介<sup>1</sup> 小田悠介<sup>1,3</sup>

<sup>1</sup> シェルパ・アンド・カンパニー株式会社 <sup>2</sup> 筑波大学大学院 システム情報工学研究群

<sup>3</sup> 国立情報学研究所 大規模言語モデル研究開発センター

{yoshiaki.sata,koichi.akabe,shunsuke.kanda,yusuke.oda}@cierpa.co.jp  
s2420612@u.tsukuba.ac.jp odashi@nii.ac.jp

## 概要

統合報告書などの日本語 ESG 文書は、非定型なレイアウトや図表を含むため、正確な文書理解や質問応答が困難である。こうした ESG 文書に対するモデルの実用的な能力を適正に評価するため、本研究では専門家が作成した質問、回答、及び根拠箇所の矩形を含む日本語 ESG 文書データセット「ESG-QA」を構築した。該当ページ検索、根拠箇所からの回答生成、根拠箇所の領域検出の3つのタスクを通じて、モデルの ESG 文書理解能力を多角的に評価するとともに、データセットが抱える課題を明らかにした。

## 1 はじめに

ESG とは、環境 (Environment)、社会 (Social)、ガバナンス (Governance) の頭文字を取った言葉であり、これらの要素を考慮した経営・事業活動を指す [1]。ESG 経営を重視する投資家などのステークホルダーは、企業が発行する ESG 関連文書を参照し、企業経営の ESG に関する透明性や有効性を評価している。

ESG 関連文書は、その形式により大きく2つに分類される。一つは、有価証券報告書やコーポレートガバナンス報告書<sup>1)</sup>のように、法律や証券取引所によって開示が義務付けられ、指定されたフォーマットで作成される文書である。もう一つは、統合報告書のように開示が任意であり、企業独自の構成、非定型なレイアウト、複雑な図表を含む文書である。本研究では、後者の統合報告書を分析対象とする。

非定形な ESG 文書の解析には、ESG ドメイン固有の事前知識を要する高度な解析技術が必要であり、解析技術の開発や解析能力の測定には ESG ドメイ

ンに特化した評価ベンチマークが必要である。

本研究では、日本語の ESG 文書に対する文書理解能力を測定するための評価ベンチマーク「ESG-QA」を構築する。<sup>2)</sup> 本ベンチマークでは、投資家や企業担当者が ESG の文脈で解決すべき問いに基づいて文書を探索・解析するという実務上のワークフローを想定しており、その過程で不可欠となる「該当ページ検索」「根拠箇所からの回答生成」「根拠箇所の領域検出」の3タスクを定義した。

専門領域のアノテーションにおいて、非専門家による作業の品質担保が困難であることが先行研究で指摘されている [2, 3]。そのため、本研究ではベンチマークの信頼性を担保するため、クラウドソーシング等の非専門家による作成ではなく、実務経験を持つ専門家による作成プロセスを採用した。作成プロセスを図1に示す。

データセット構築に加えて、最先端モデルを対象とした評価実験を行い、ESG-QA におけるベースライン性能を算出した。実験結果の検討を通じて、ESG ドメインにおけるモデルの現状の能力や、ESG-QA に残された課題を明らかにする。

## 2 関連研究

ESGenius[4] は、231 件の英語 ESG 関連文書を対象としたベンチマークである。大規模言語モデル (LLM) が文書内の記述に基づいて多肢選択式問題の候補を作成し、それらを専門家が検証することで、最終的に 1,136 件の質問応答ペアを構築している。しかし、文書がテキスト情報のみに基づく、質問が多肢選択式に限られるなどの制限が存在し、実務との関連性の観点で有効な問題設定を扱っているとは言い難い。

1) <https://www.jpx.co.jp/equities/listing/cg/01.html>

2) <https://huggingface.co/datasets/cierpa/esg-qa-dataset>

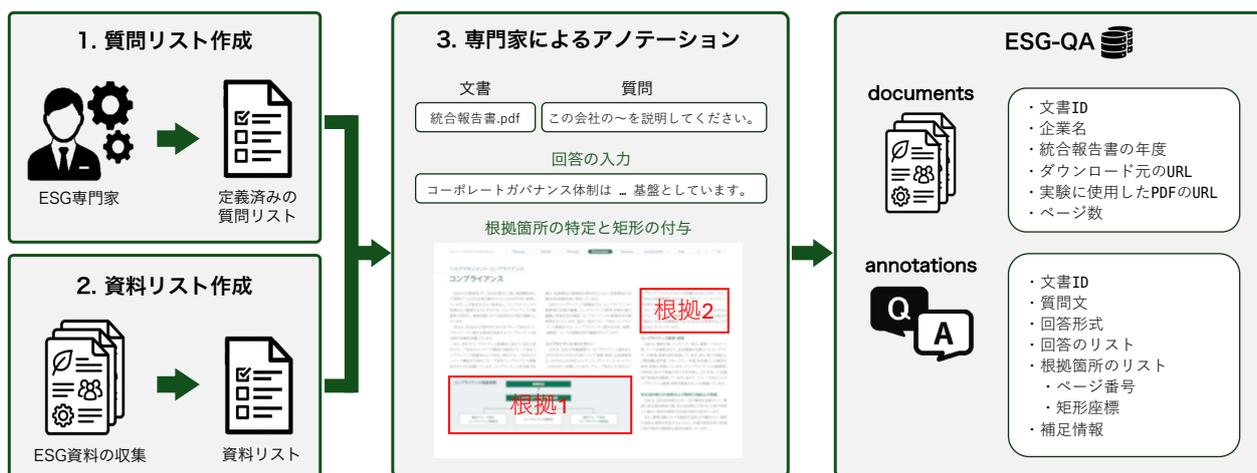


図1 ESG-QA 構築の流れ

MMESGBench[5] は、45 件の英語 ESG 関連文書を対象としたベンチマークである。マルチモーダル LLM に文書内のテキストや図表を入力し、それらに基づく質問応答ペアの候補を作成させるアプローチにより、933 件のデータセットを構築している。

ClimRetrieve[6] は、30 件の気候関連文書と事前に定義された 16 の質問を用い、専門家が YES/NO 形式で回答を作成した 743 件のペアからなる情報検索のみを対象としたベンチマークである。

JDocQA[7] は、5,508 件の官公庁公開 PDF 文書を対象とし、視覚情報とテキスト情報の双方を参照する必要がある 11,600 件の質問応答ペアで構成されている。日本語のドキュメント理解ベンチマークとして有用であるが、行政文書全般を扱っているため、ESG ドメイン特有の専門知識や文脈理解能力を測るには適していない。

これらのデータセットには、日本語の ESG 文書に対する質問応答の評価をする上で大きく 2 つの課題がある。1 つは、日本語特有の ESG 文脈を評価できるデータセットでない点、もう 1 つは、「文書起点」で質問を作成している点である。実務では「解決したい問い」が先に存在し、それに基づいて文書を探索するというプロセスが一般的である。そのため、文書から逆算して作られた質問は、文書に関する事前情報が一定割合で質問に織り込まれてしまうため、実際の利用シーンにおける需要から乖離する可能性がある。

これらの課題を解決するために、本研究では国内企業が公開する日本語の統合報告書を対象に、専門家が実務で直面する問いを先に作成する「質問起点」のアプローチを採用した。これにより、文書の記

述内容に制約されず、ユーザーの課題解決に即した、より実践的なベンチマークの構築が可能となる。

## 3 ESG-QA

### 3.1 データセットの構築

ESG-QA は、文書データである *documents* と、質問回答データである *annotations* の 2 つのサブセットから構成される。本節では、それぞれの構築手段について記載する。図 1 にその概要を示す。

**質問の作成** 金融・証券業者、投資家、企業のサステナビリティ担当者にとって有益と考えられる質問を、ESG の専門家 1 名によって 20 件作成した。質問には回答形式が割り当てられており、「はい/いいえ」「事実抽出」「数量」「自由記述」の 4 種類が存在する。

**ドキュメントの収集** *documents* の構築にあたり、本研究では日本国内の統合報告書を発行している上場企業の中から無作為に 28 社を抽出し、各社が公表している PDF 形式の報告書を収集した。

**アノテーション手法** 2 節に記載の通り、ESG-QA では質問起点の作成プロセスを採用した。具体的なアノテーション工程は以下の 3 ステップであり、2 名の ESG 専門家によって実施した。

1. 質問に対し、回答の根拠となる箇所を文書中から特定する。
2. Computer Vision Annotation Tool[8] を用い、特定した根拠箇所に矩形のアノテーションを付与する。根拠箇所が複数存在する場合は、すべての箇所にアノテーションを付与する。
3. アノテーションした根拠箇所に基づき、質問に対する回答を入力する。

結果として、質問に対する回答、及びその根拠となるページと矩形領域が得られた。

### 3.2 統計情報

*documents* は 28 件の ESG 文書の PDF からなる。ページ総数は 1,522, 1 つの文書あたりの平均ページ数は 54.36, 標準偏差は 25.83 である。

*annotations* は、総計 560 件の質問回答事例からなる。これは、作成した全 20 質問と収集した全 28 文書の全対についてアノテーションを実施した結果である。この中には文書内の情報からは回答不能な質問も含まれており、実際に回答が存在する事例は 279 件である。

### 3.3 タスク設定

ESG-QA は、実用上の重要性を考慮した 3 つのタスクにより構成される。

**該当ページ検索** 入力として質問文と文書全体を受け取り、回答の根拠が含まれるページ番号を出力するタスクである。本タスクは、多ページにわたる文書から、回答に関連するページ候補を効率的に絞り込む能力を評価することを目的とする。評価指標には、Recall@k, MAP@k, nDCG@k を使用する [9]。

**根拠箇所からの回答生成** 入力として質問文と、正解の根拠箇所が含まれる文書ページを受け取り、回答文を出力するタスクである。文書ページの入力形式としては、テキスト情報のみを用いる場合と、画像情報のみを用いる場合の 2 通りを設定する。本タスクは、与えられた情報源から ESG ドメイン特有の語彙や概念を正確に捉え、適切な回答を生成する能力を評価することを目的とする。評価指標には、正解テキストに対する生成テキストの Character F1[10], Exact Match[10], BLEU[11], および LLM-as-a-Judge[12] による正誤判定を使用する。

**根拠箇所の領域検出** 入力として質問文と、正解の根拠箇所が含まれる文書ページの画像を受け取り、根拠箇所に該当する矩形領域リストを出力するタスクである。本タスクは、回答の根拠となる図表やテキストブロックを文書画像から正確に抽出する能力を評価することを目的とする。評価指標には、正解領域  $A_{ans}$  と予測領域  $A_{pred}$  の一致度を測る Intersection over Union (IoU) を使用する。IoU は以下の式で定義される。

$$\text{IoU} = \frac{|A_{ans} \cap A_{pred}|}{|A_{ans} \cup A_{pred}|}$$

## 4 評価実験

### 4.1 実験設定

ESG-QA のベースライン性能を確立するため、複数のモデルについて実験的評価を実施した。

**該当ページ検索** 本実験では、埋め込みモデルを用いて PDF 文書の各ページと質問クエリをベクトル化し、コサイン類似度に基づいて根拠となるページの候補を順に取得した。商用モデルからは text-multilingual-embedding-002<sup>3)</sup>, オープンソースモデルからは sarashina-embedding-v2-1b<sup>4)</sup>, colqwen2-v1.0<sup>5)</sup> [13] を採用した。このうち、マルチモーダルモデルである colqwen2-v1.0 については、文書ページの画像を入力し検索を行う。一方、text-multilingual-embedding-002 と sarashina-embedding-v2-1b は、Google Cloud Vision API<sup>6)</sup> でページ画像を OCR した結果を入力とする。指標の算出には ranx[14] を用いた。

**根拠箇所からの回答生成** 商用モデルとして、gemini-2.5-flash<sup>7)</sup> [15], gpt-5-mini<sup>8)</sup> を用いた。オープンソースモデルからは、gpt-oss-120b<sup>9)</sup> [16], llm-jp-3.1-8x13b-instruct4<sup>10)</sup> を採用した。マルチモーダル入力に対応した商用モデルについては、入力形式としてテキスト情報のみを与える場合と、画像情報を与える場合の双方を比較する。入力テキストがモデルのコンテキストサイズを超過する場合は、未回答とする。LLM-as-a-Judge の評価者モデルには gemini-2.5-pro<sup>7)</sup> を使用した。

**根拠箇所の領域検出** 本実験では、レイアウト解析モデルが PDF 画像からテキストや図表を矩形として抽出し、マルチモーダル LLM が各矩形を根拠箇所として適切かを二値分類する、2 段階のパイプラインからなるアプローチを採用した。レイアウト解析モデルには docling-layout-heron<sup>11)</sup> [17], 判定を行うマ

3) <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/model-reference/text-embeddings-api?hl=ja>

4) <https://huggingface.co/sbintuitions/sarashina-embedding-v2-1b>

5) <https://huggingface.co/vidore/colqwen2-v1.0>

6) <https://cloud.google.com/vision>

7) <https://ai.google.dev/gemini-api/docs/models?hl=ja>

8) <https://platform.openai.com/docs/models/gpt-5-mini>

9) <https://platform.openai.com/docs/models/gpt-oss-120b>

10) <https://huggingface.co/llm-jp/llm-jp-3.1-8x13b-instruct4>

11) <https://huggingface.co/docling-project/docling-layout-heron>

表 1 該当ページ検索タスクの評価結果

モデル	Top-K	Recall	MAP	nDCG
text-multilingual-embedding	5	0.803	0.639	0.691
	10	<b>0.922</b>	0.660	0.733
sarashina-embedding-v2-1b	5	<b>0.864</b>	<b>0.670</b>	<b>0.729</b>
	10	<b>0.953</b>	<b>0.682</b>	<b>0.751</b>
colqwen2-v1.0	5	0.443	0.285	0.332
	10	0.572	0.302	0.375

マルチモーダル LLM には gemini-2.5-flash を採用した。

## 4.2 実験結果

該当ページ検索タスクの実験結果を表 1 に示す。日本語特化の埋め込みモデルである sarashina-embedding-v2-1b が、ほぼ全ての指標において他モデルを上回る最高性能を達成した。対照的に、マルチモーダル検索を行った colqwen2-v1.0 は、テキストベースのモデルと比較して大きく劣る結果となった。

根拠箇所からの回答生成タスクの実験結果を表 2 に示す。商用モデルはオープンソースモデルと比較して全体的に高い性能を示した。指標別では、gemini-2.5-flash が Character F1, Exact Match, BLEU で最高値を記録した一方、gpt-5-mini は LLM-as-a-Judge において最も高い評価を得た。なお、入力形式の違いによるスコアへの影響は限定的であった。

根拠箇所の領域検出タスクの IoU スコアは **0.297** となった。

## 4.3 データセットの妥当性評価

本節では、評価の妥当性を検証することを目的として、4.2 節において最良の性能を示したモデルの誤答事例に着目し、ESG 専門家による再評価を行った結果を述べる。

該当ページ検索タスクについて、Recall@10 が 0 であった 13 事例における予測ページ総数 130 件を再評価した。専門家による精査の結果、誤答とされた 130 ページのうち約 26.2%にあたる 34 件については、実際には回答の根拠として妥当な記述を含んでいることが確認された。

根拠箇所からの回答生成タスクについても、gpt-5-mini において LLM-as-a-Judge により不正解と判定された 80 事例から 20 事例を無作為抽出し、再評価を行った。その結果、半数にあたる 10 件 (50.0%) は、実際には正しい回答であった。その要因を調査し

表 2 根拠箇所からの回答生成タスク評価結果

モデル	入力形式	Char F1	EM	BLEU	Judge
gemini-2.5-flash	テキスト	<b>0.490</b>	<b>0.327</b>	<b>0.345</b>	0.631
	画像	0.454	0.289	0.325	0.649
gpt-5-mini	テキスト	0.444	0.282	0.327	<b>0.713</b>
	画像	0.451	0.297	0.328	0.634
gpt-oss-120b	テキスト	0.448	0.280	0.315	0.609
llm-jp-3.1-8x13b	テキスト	0.386	0.225	0.269	0.409

た結果、9 件は正解データの不完全さに起因し、予測した回答が正しかったにも関わらず、正解データとの乖離により不正解と判定されたものであった。正解の回答は一意では無く、参照箇所や解釈によって異なる回答もあり得るという制限が明らかとなった。残りの 1 件は、LLM-as-a-Judge 自体が誤っており、正解データと予測した回答の意味が一致しているにも関わらず不正解と判定されていた。

同様に根拠箇所の領域検出タスクについても、IoU が 0 となった 197 事例から 20 事例を無作為に抽出し、モデルが根拠箇所として予測した計 221 件の矩形について、それが実際に正解であるかどうかの再評価を行った。その結果、約 14.0%にあたる 31 個の領域については、実際には根拠箇所として正解であることが判明した。

これらの分析から、本データセットには正解ラベルの欠落が存在することが明らかとなった。これは、モデルが適切な根拠を導出しているにもかかわらず、評価上は誤答として扱われていることを意味しており、現状の自動評価スコアはモデルの真の能力を過小評価している可能性がある。

## 5 おわりに

本研究では、専門家の知見を取り入れた日本語 ESG 文書ベンチマーク「ESG-QA」を構築し、複数のモデルを用いてベースライン性能を導出した。専門家による妥当性評価の結果、モデルが正当な出力を行っているにもかかわらず、データセット側のアンノテーション漏れにより誤答と判定される事例が確認された。この正解ラベルの網羅性の欠如は本データセットにおける現時点での限界であり、自動評価結果の解釈において留意すべき点である。加えて、データセットが小規模であることや、専門家による主観で質問回答が作成されていることも課題である。今後は、これらの課題を克服するために、データセットの品質向上が求められる。

## 参考文献

- [1] 内閣府. 令和2年度障害者差別の解消の推進に関する国内外の取組状況調査報告書第2章 国外調査 2.2.1 esg とは何か. [https://www8.cao.go.jp/shougai/suishin/tyosa/r02kokusai/h2\\_02\\_01.html](https://www8.cao.go.jp/shougai/suishin/tyosa/r02kokusai/h2_02_01.html), 2020. (閲覧日: 2025-11-28).
- [2] Ting Wai Terence Au, Vasileios Lampos, and Ingemar Cox. E-NER — an annotated named entity recognition corpus of legal text. In Nikolaos Aletras, Ilias Chalkidis, Leslie Barrett, Cătălina Goanță, and Daniel Preoțiuc-Pietro, editors, **Proceedings of the Natural Legal Language Processing Workshop 2022**, pp. 246–255, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.
- [3] Malte Michel Multusch, Lasse Hansen, Mattias Paul Heinrich, Lennart Berkel, Axel Saalbach, Heinrich Schulz, Franz Wegner, Joerg Barkhausen, and Malte Maria Sieren. Impact of radiologist experience on ai annotation quality in chest radiographs: A comparative analysis. **Diagnostics**, Vol. 15, No. 6, 2025.
- [4] Chaoyue He, Xin Zhou, Yi Wu, Xinjia Yu, Yan Zhang, Lei Zhang, Di Wang, Shengfei Lyu, Hong Xu, Wang Xiaoqiao, Wei Liu, and Chunyan Miao. ESGenius: Benchmarking LLMs on environmental, social, and governance (ESG) and sustainability knowledge. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, **Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing**, pp. 14623–14664, Suzhou, China, November 2025. Association for Computational Linguistics.
- [5] Lei Zhang, Xin Zhou, Chaoyue He, Di Wang, Yi Wu, Hong Xu, Wei Liu, and Chunyan Miao. Mmesgbench: Pioneering multimodal understanding and complex reasoning benchmark for esg tasks. In **Proceedings of the 33rd ACM International Conference on Multimedia**, MM '25, p. 12829–12836, New York, NY, USA, 2025. Association for Computing Machinery.
- [6] Tobias Schimanski, Jingwei Ni, Roberto Spacey Martín, Nicola Ranger, and Markus Leippold. ClimRetrieve: A benchmarking dataset for information retrieval from corporate climate disclosures. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 17509–17524, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [7] Eri Onami, Shuhei Kurita, Taiki Miyanishi, and Taro Watanabe. JDocQA: Japanese document question answering dataset for generative language models. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 9503–9514, Torino, Italy, May 2024. ELRA and ICCL.
- [8] CVAT.ai. Computer vision annotation tool (CVAT). <https://github.com/cvat-ai/cvat>, 2024. Accessed: 2025-11-28.
- [9] 酒井哲也. 情報アクセス評価方法論: 検索エンジンの進歩のために. コロナ社, 2015.
- [10] Kentaro Kurihara, Daisuke Kawahara, and Tomohide Shibata. JGLUE: Japanese general language understanding evaluation. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 2957–2966, Marseille, France, June 2022. European Language Resources Association.
- [11] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin, editors, **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [12] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging llm-as-a-judge with mt-bench and chatbot arena. In **Proceedings of the 37th International Conference on Neural Information Processing Systems**, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.
- [13] Manuel Faysse, Hugues Sibille, Tony Wu, Bilel Omrani, Gautier Viaud, CELINE HUDELOT, and Pierre Colombo. Colpali: Efficient document retrieval with vision language models. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, **International Conference on Representation Learning**, Vol. 2025, pp. 61424–61449, 2025.
- [14] Elias Bassani. ranx: A blazing-fast python library for ranking evaluation and comparison. In **ECIR (2)**, Vol. 13186 of **Lecture Notes in Computer Science**, pp. 259–264. Springer, 2022.
- [15] Gemini Team and Google. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities.
- [16] OpenAI. gpt-oss-120b gpt-oss-20b model card, 2025.
- [17] Nikolaos Livathinos, Christoph Auer, Maksym Lysak, Ahmed Nassar, Michele Dolfi, Panos Vagenas, Cesar Berrospi Ramis, Matteo Omenetti, Kasper Dinkla, Yusik Kim, Shubham Gupta, Rafael Teixeira de Lima, Valery Weber, Lucas Morin, Ingmar Meijer, Viktor Kuropiatnyk, and Peter W. J. Staar. Docling: An efficient open-source toolkit for ai-driven document conversion, 2025.

## A 使用したプロンプト

本研究で使用したプロンプトを以下に示す。なお、文脈や質問等が入力される箇所は紙面の都合上省略した。

### 根拠箇所からの回答生成

文脈の内容を元にして以下の質問に答えてください。

以下のタイプに応じて回答を変えてください。

- YES\_NO: YES or NO で回答

- FREE\_TEXT: 自由記述で回答

- EXTRACTION: 文脈中の該当する文字列を抽出して回答

- NUMBER: 数字で回答

EXTRACTION のみ複数個の回答生成が可能です。他のタイプの場合は、必ず一つだけ出力してください。

また、与えられた文脈からは回答できない場合は、answers に ["-"] を出力してください。

### 根拠箇所の領域検出

添付した画像に、質問に対する回答としての根拠が記載されているか否かを判断してください。

根拠はテキストだけでなく、あらゆる情報源が該当します。

根拠が記載されていると判断したら YES、正しくないと判断したら NO を与えなさい。

### 回答生成における LLM-as-a-Judge

与えた質問と正解のペアを元に、予測回答が正しいか否かを判断してください。

予測回答が正しいと判断したら 1、正しくないと判断したら 0 のスコアを与えなさい。

正解タイプに応じて、以下のように判断基準を変えてください。

- YES\_NO: 正解と YES or NO が一致しているかどうか

- FREE\_TEXT, EXTRACTION: 正解に含まれる要素が全て含まれているか

- NUMBER: 正解と数字が一致しているかどうか

正解・予測回答の中身が "-" である場合、ドキュメントの情報からは質問に答えられないという意味です。

に示す。

### 環境

- 環境方針を定めていますか。
- スcope 1, 2, 3 の各排出量について、温室効果ガス削減のための施策を列挙してください。
- 水関連のリスクを評価しているか。
- 公害管理に関する証明書を取得しているか。
- プラスチック廃棄物を削減するための会社の活動をリストアップしてください。
- 生物多様性に関する方針を持っていますか。

### 社会

- 人権方針を定めていますか。
- サプライチェーンの人権デューデリジェンスを行っていますか。
- 会社の責任ある調達方針にはどのようなトピックが含まれていますか。
- 安全衛生に関する研修を実施していますか。
- サプライチェーンにおける児童労働のリスクを認識していますか。
- その会社の休業災害度数率 (LTIFR) はどうですか。
- 社会的持続可能性の文脈で、この企業は何に重点を置いていますか。
- 社会的持続可能性の観点から、同社は優良企業とみなされていますか。
- 人的資本にどのように投資していますか。

### ガバナンス

- 取締役会の独立性について説明してください。
- 直近 1 年間に、ガバナンスを改善するために新たに設定した制度や方針はありますか。
- コンプライアンスに関する報告件数はいくつですか。
- 汚職・贈収賄にどのように対処していますか。
- この会社のガバナンス・システムを説明してください。

## B 質問項目一覧

ESG-QA で使用された 20 件の質問項目を、環境・社会・ガバナンスの 3 つのカテゴリに分類して以下