

# 大規模言語モデルを用いた従業員価値の測定

## Measuring Employee Value Using LLM

濱田祐馬

Yuma HAMADA

広島経済大学大学院

Hiroshima University of Economics Graduate School

**概要** 近年、企業は、環境（Environment）や社会（Social）、ガバナンス（Governance）面が重視される ESG 評価が行われるようになってきている。本研究では、代表的な ESG の評価指標を統合して作成された「価値モデル」を評価指標として採用し、企業が ESG 活動を報告するために開示しているサステナビリティ報告書から、大規模言語モデル（LLM：Large Language Model）を利用して ESG 評価を自動で行うことを目指す。

具体的には、サステナビリティ報告書の内容に対し、ESG 評価指標に関する質問を行い、大規模言語モデルに回答を生成させることで ESG 評価を行った。性能評価の実験の結果、0.78 と高い正答率を得た。

### 1 はじめに

近年、グローバルな環境・社会課題を背景に、企業は財務評価だけでなく、環境（Environment）や社会（Social）、ガバナンス（Governance）面が重視される ESG 評価が行われるようになってきている。しかしながら、現代の ESG 評価には、主要な問題点が 2 つある[1]。1 つ目は、複数の ESG 評価機関が、同一の企業に ESG 評価を行った場合に、ESG 評価の不一致が起こる点である。2 つ目は、ESG 評価を手で行うには多大なコストがかかる点である。

本研究では、これらの 2 つの問題を解決するため、多様な環境・社会・ガバナンス（ESG）のフレームワークを統合して開発された「価値モデル[2]」を採用し、企業のサステナビリティ報告書から、AI 技術を利用して ESG 評価を自動で行うことを目指す。

近年、AI 技術の一つとして、ChatGPT などに代表される大規模言語モデル（LLM：Large Language Model）が注目を集めている。本研究では、大規模言語モデルを企業の ESG 評価に活用するための手法を提案し、性能を評価することで大規模言語モデル活用のための基礎検討を行う。

具体的には、サステナビリティ報告書の内容に対し、ESG 評価指標に関する質問を行い、大規模言語モデルに回答を生成させることで ESG 評価を行う。

### 2 関連研究

近年では、ESG 分野に AI を用いる研究が進んでいる。

大規模言語モデルを用いて ESG 情報を抽出する研究には [3, 4] がある。指田ら [3] は、サステナビリティ報告書などの PDF ベースで作成されている報告書から、SDGs 関連の文章を抽出し、17 のゴール別に分類を行うモデルを作成している。小杉ら [4] は、ChatGPT [5] を用いて、ニュース記事の ESG トピックを抽出する手法を提案している。指田らや小杉らは、

大規模言語モデルを用いて ESG 情報の抽出を行っているが、本研究では大規模言語モデルを用いて ESG 評価を行う。

AI を活用して ESG 評価を行う研究には石野らの研究 [6] がある。石野ら [6] は、サステナビリティ報告書から ESG 評価を AI で推定するモデルを作成している。石野は、ニューラルネットワークを用いて ESG 評価を行っているが、本研究では大規模言語モデルを用いて ESG 評価を行う。

### 3 価値モデルの概要

本章では、本研究で ESG 評価指標として採用する価値モデルについて説明する。価値モデルは、同志社大学の社会価値研究センターが開発した、企業のサステナビリティ報告書などの報告書を使用して、ESG 評価を行うための指標である。価値モデルは、GRI や SASB、TCFD などの 40 以上のサステナビリティ開示基準、フレームワーク、およびモデルを統合し、1200 以上の個別インパクト測定を含む包括的なモデルとなっている。表 1 に価値モデルの構成の例を示す。

表 1 価値モデルの構成

ステークホルダ	テーマ	ゴール	サブゴール
Employee	E1: Diversity & Equity	E1-A: Full-time Employment	E1-A#1: Organization hires 85% or more of total workforces as full-time employees.
		E1-B: Ethnic Diversity	E1-B#1: Organization's workforce matches 100% with ethnicity of surrounding population.

価値モデルは、ステークホルダ、テーマ、ゴール、サブゴールという構成である。価値モデルには、企業に関わる 7 件のステークホルダが設定されている。ステークホルダには、細分化したテーマが設定されている。7 件のステークホルダには合計 27 件のテーマが設定されている。さらに、テーマに関わる 81 のゴールが設定されている。ゴールには段階的な要素が含まれているため、いくつかのサブゴールに分割されている。このサブゴールが ESG 評価を行う際の指標となる。

価値モデルのサブゴールは、表 1 の下線部の「85%」のように、定量的な目標が設定されており、従来の ESG 評価指標よりも客観的に企業の取り組みの成果を評価できる。

さらに、価値モデルでは、サブゴールに対する情報開示の質についても評価する。情報開示の質は 0~3 点の整数で評価され、これを情報開示スコアと呼ぶ。情報開示スコアの基準を表 2 に示す。

表 2 情報開示スコアの基準

スコア	基準
0	サブゴールに関する企業の活動結果が <u>記述されていない</u> 。
1	サブゴールに関する企業の活動結果が <u>簡潔に記述されているか、定性情報のみ記述されている</u> 。
2	サブゴールに関する企業の活動結果の <u>定量情報が記述されているが、価値モデルで測定できない情報が記述されている</u> 。
3	サブゴールに関する企業の活動結果の <u>定量情報が記述されており、価値モデルで測定可能な情報が記述されている</u> 。

情報開示スコアの 3 点と 2 点は、価値モデルでの測定の可否で区別される。表 3 に、E1-A#1 のサブゴールである「Organization hires 85% or more of total workforces as full-time employees」に対して、情報開示スコアでそれぞれ 3 点と 2 点が付与される記載の例を示す。

表 3 E1-A#1 における情報開示スコアの付与例

3 点が付与される記載	2 点が付与される記載
<u>99%</u> of all employees are full-time employees	The number of full-time employees is <u>33,043</u>

表 3 に示すように、サステナビリティ報告書で、「99% of all employees are full-time employees」という記載があれば、情報開示スコアは 3 点が付与される。これは、常勤従業員の情報が下線部のように「99%」と割合で記載されており、価値モデルで評価することができるからである。また、サステナビリティ報告書で、「The number of full-time employees is 33,043」という記載があれば、情報開示スコアは 2 点が付与される。これは、常勤従業員の数が下線部のように「33,043」と定量情報で記載されているが、割合が不明なため価値モデルで評価できないためである。本研究では、情報開示スコアの 3 点と 2 点とそれ以外の評価を、大規模言語モデルを用いて行う。

#### 4 ChatPDF を用いた企業の ESG 評価手法

本章では、大規模言語モデルの一つである ChatPDF[7]を用いて企業の ESG 評価を行う手法について説明する。

ChatPDF は、OpenAI[8]が開発した大規模言語モデルをベースとした ChatGPT を活用したサービスであり、読み込んだ PDF ファイルに関する質問を投げかけると、それに応じた回答を生成する。ChatPDF は、その名の通り PDF ファイル

の分析や解析に特化したモデルであり、ChatGPT と比較して、PDF ファイルの読み込みの速さ、質問に対する回答性能が優れている。

本研究では、ChatPDF を用いて価値モデルをもとに ESG 評価を行う。価値モデルでは、サブゴールをもとに ESG 評価を行うため、サブゴールに関する質問を作成する。

本研究では、人手で正解データを作成するのが比較的容易と判断した、ステークホルダ「Employee」のテーマ「Diversity & Equity」と「Human rights」のサブゴール 15 件に関する情報開示スコアを評価するための質問を作成した。サブゴールに対し、それぞれ情報開示スコアの 3 点と 2 点を評価する質問を作成した。各サブゴールに対し、3 点の質問が当てはまらなかった場合、2 点の質問に推移するように設定した。

質問の構成は以下のように設計した。

- (1) 回答は必ず Yes・No から始めることという指示
- (2) 回答の根拠となる箇所（該当箇所）を含むという指示
- (3) サブゴールに関する質問
- (4) 指定の単語を同じ意味として扱うことという指示

「(1) 回答は必ず Yes・No から始めること」と、「(2) 回答の根拠となる箇所（該当箇所）を含めること」という指示は、全ての質問に共通の指示として付与した。実際に作成した指示を表 4 に示す。

表 4 全ての質問に共通して付与する指示

質問
Please be sure to answer Yes or No at the beginning.
Please be sure to include quantitative information and page numbers in your answer.

「(3) サブゴールに関する質問」について、「Organization hires 85% or more of total workforces as full-time employees」という E1-A#1 のサブゴールを用いて説明する。E1-A#1 の 3 点と 2 点の質問を表 5、表 6 に示す。

表 5 E1-A#1 の 3 点を判定する質問

質問
Does this report provide information about the percentage of full-time employees?

表 6 E1-A#1 の 2 点を判定する質問

質問
Does this report provide information on the number of full-time employees?

ChatPDF に質問を投げかける際には、先に 3 点を判定する質問から行う。もし、ChatPDF がこの質問に No と判断すれば、同じサブゴールの 2 点を判定する質問へ推移する。さらに、2 点を判定する質問でも、ChatPDF が No と判断すれば、情報開示スコアは 1 点か 0 点のいずれかに相当する。この推移から、情報開示スコアの 3 点と 2 点とそれ以外の評価を行うことができる。「(4) 指定の単語を同じ意味として

扱うこと」という指示について、(3)と同様に E1-A#1 のサブゴールを用いて説明する。E1-A#1 のサブゴールでは、常勤従業員を「full-time employees」という単語で示している。しかし、複数のサステナビリティ報告書を確認したところ、常勤従業員を示す単語は、「permanent, direct, regular employee, staff member, in-house employee」など異なる単語を用いている場合が多々あった。このような場合に、質問で「full-time employees」のみを使用すると、ChatPDF が正しく回答できないことが明らかになった。そのため、「permanent, direct, regular employee, staff member, in-house employee」という単語を「full-time employees」と同じ意味として扱うよう指示を与えた。作成した指示を表 7 に示す。

表 7 指定の単語を同じ意味として扱う指示

質問
Please note that the following words are also used interchangeably with full-time: permanent, direct, regular employee, staff member, in-house employee

## 5 実験

本章では、本研究の提案手法である ChatPDF を用いた ESG 評価手法の有効性を確認するために行った実験について報告する。

### 5.1 実験に使用したデータ

実験は、主要な ESG 評価機関の 1 つである S&P グローバルが実施した ESG の格付け結果[9]で、ESG スコアの高い 10 社の企業のサステナビリティ報告書を候補とした。そのうち 2 社のサステナビリティ報告書のファイルサイズが、ChatPDF の API からアップロードできる 32MB を超過していたため、本実験の対象からは除外した。実験に使用した 8 社の企業を以下に示す。

- LG H&H Co.,Ltd,
- Indra Sistemas, S,A,
- Coca-cola HBG AG
- Deutsche Telecom AG
- Reckitt Benckiser Group plc
- PTT Global Reckitt Chemical Public Company Limited
- Moncler S,p,A
- Tama-Rete Elettrica Nazionale Societa per azioni

### 5.2 評価の観点と評価尺度

本節では、実験の結果を評価する方法について説明する。本実験では、ChatPDF が正しく回答できたかどうかを、以下の 2 つの観点で評価した。2 つの観点を、図 1 の回答を例に説明する。

- Yes・No が正しいか
- 回答の根拠となる箇所（該当箇所）が正しいか

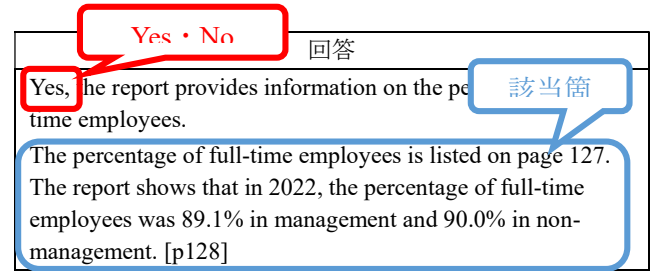


図 1 評価の観点

1 つ目の観点は、図 1 の赤枠で囲まれた回答の冒頭の Yes・No が正しいかどうかで判断する。図 1 では、ChatPDF は、Yes と回答している。この例に用いられたサステナビリティ報告書を人手で確認したところ、Yes という回答は正しかった。したがって、この場合は、ChatPDF の Yes・No は正しいと評価される。

2 つ目の観点は、図 1 の青枠の部分である ChatPDF の回答の根拠となる箇所（該当箇所）が正しく回答できているかどうかで判断する。人手で該当箇所を確認したところ、ページ数は正しいが、数値は誤りであった。したがって、この場合は、ChatPDF の回答の根拠となる箇所は誤りと評価される。

本実験では、上記の 2 つの観点から、以下の式で示す正答率と該当箇所正答率を求め、評価尺度として使用した。

$$\text{正答率} = \frac{\text{ChatPDF が Yes・No を正しく回答できたサブゴール数}}{\text{全てのサブゴール数}}$$

$$\text{該当箇所正答率} = \frac{\text{ChatPDF が該当箇所を正しく回答できたサブゴール数}}{\text{ChatPDF が Yes・No を正しく回答できたサブゴール数}}$$

### 5.3 実験結果と考察

実験の結果を表 8 に示す。表 8 より、正答率は 0.78 と高い数値であったが、該当箇所を含めると正答率が低下した。

表 8 実験の結果

正答率	該当箇所正答率
0.78	0.68

さらなる性能の向上を目指して、ChatPDF の回答が誤っていた事例について誤り分析を行う。

以下では、ChatPDF による回答の正答率低下の原因を検証するため、ChatPDF と人手（正解）で回答が異なった 26 件について原因を考察する。

ChatPDF が、「No」と回答すべきところを誤って「Yes」と回答した 24 件について分析したところ、以下のような 2 つのケースがあった。

- (1) PDF 内に測定指標の数値がない場合に測定指標と異なる数値を抽出してしまうケース (21 件)

ChatPDF に質問した測定指標の数値がサステナビリティ報告書内に記載がないにも関わらず、ChatPDF が測定指標とは異なる数値を該当箇所として抽出し、「Yes」と誤

って回答してしまうケースがあった。ChatPDF に測定指標として国別の上級管理職の割合について質問した場合を例に説明する。この質問の場合、サステナビリティ報告書に国別の上級管理職の割合の記載はないとき、ChatPDF は「No」と回答するのが正解である。しかし、ChatPDF は、女性中間管理職の割合を該当箇所として抽出して、「Yes」と誤った回答をしていた。

(2) PDF 内に測定指標の数値がない場合に外部の情報を参照するケース (3件)

ChatPDF に質問した測定指標の数値がサステナビリティ報告書内に記載がないにも関わらず、ChatPDF が外部にある情報を根拠に「Yes」と回答するケースがあった。サステナビリティ報告書内に測定指標の数値の記載はなく、測定指標の数値の参照先として外部サイトの URL の記載がある場合、ChatPDF は「please refer to the original source.」と外部サイトの参照を促し、「Yes」と誤った回答をしていた。

一方、ChatPDF が、「Yes」と回答すべきところを誤って「No」と回答した2件について分析をしたところ、以下のようなケースがあった。

- 文や表の構造が複雑で測定指標の数値の抽出が困難なケース (2件)

文や表の構造が複雑な場合、ChatPDF が測定指標の数値を適切に抽出できず、「No」と誤って回答してしまうケースがあった。例えば、文の構造が複雑な事例や、表の構造が複雑な事例である。

文の構造が複雑な事例では、国籍別の管理職の割合が記載されている箇所が挙げられる。Reckitt Benckiser Group plc のサステナビリティ報告書には、「国名：国籍別の従業員の割合 (国籍別の管理職の割合)」という構造で記載されている。このような文から国籍別の管理職の割合を抽出したい場合、数値の前にある文を理解する必要があるが、ChatPDF は対応できていなかった。

表の構造が複雑な例では、人権トレーニングを受けた従業員の割合が記載されている箇所が挙げられる。図2に人権トレーニングを受けた従業員の割合が記載されている表を示す。

Average hours of training	GROUP (*)			TERNA		
	2022	2021	2020 (*)	2022	2021	2020
Average hours of training						
- per employee **	36	31	30	40	35	34
By category **						
- senior managers	32	12	17	36	13	19
- middle managers	35	25	25	37	26	25
- office staff	33	28	29	36	30	31
- blue-collar workers	42	43	36	52	54	47
By gender **						
- men	37	33	31	42	37	34
- women	27	18	22	30	19	24
Hours of training in human rights	496	1,584	996	466	1,584	996
Participants in the course on human rights (%) **	20.1	34.5	5.5	24.0	39.8	6.4
Percentage of employees	92	99	96	100	100	98
Hours provided						
Total	184,016	148,698	135,417	174,708	145,528	132,487

図2 人権トレーニングを受けた従業員の割合をChatPDF が抽出できなかった事例

図2の表では、左側の列に項目があり、各項目に対応した数値が同じ行に記載されている。また、項目の一部には単

位(%)が記載されている。この表から人権トレーニングを受けた従業員の割合を抽出するためには、表の行と列の対応を解析する必要があるが、ChatPDF は対応できていなかった。

## 6 おわりに

本研究では、ESG 評価指標に価値モデルを採用し、大規模言語モデルを企業の ESG 評価に活用するための手法を提案した。提案手法の有効性を確認するための実験を行った結果、正答率は 0.78 という高い性能を示しており、大規模言語モデルを活用した ESG 評価で一定の成果を得た。

今後の課題としては、ChatPDF の回答精度を上げるため、質問を洗練させる必要があることが挙げられる。

また、本研究では、ステークホルダ「Employee」の一部のテーマについて実験を行ったが、今後は他のテーマについても実験を行いたい。

## 7 参考文献

- [1] 北川 哲雄: サステナビリティ情報開示ハンドブック, 日本経済新聞, (2023)
- [2] 須貝フィリップ, 小泉レン, リナン ニコラス, パッタナプラユンウォン サタナン ペットハーン ジャッカラン: 責任あるビジネスのための価値モデル, 同志社大学社会価値研究センターWhite paper, (2022)
- [3] 指田 昌樹, 和泉 潔, 坂地 泰紀: BERT および ChatGPT を用いたサステナビリティレポートからの SDGs 関連文抽出, 人工知能学会第二種研究会 第 31 回金融情報学研究会, (2023)
- [4] OpenAI: ChatGPT, <https://openai.com/chatgpt> (参照 2023-10-26) .
- [5] 小杉 樹来, 小澤 誠一, 廣瀬 勇秀, 池田 佳弘, 中川 憲保, 飯塚 正昭, 西田 大輔: ChatGPT を用いたニュース記事の ESG トピック分析, 人工知能学会第二種研究会 第 31 回金融情報学研究会, (2023)
- [6] 石野 亜耶, 中尾 悠利子, 國部 克彦, 中久 保菜穂: AI による ESG 評価の推定モデルの構築, 同文館, (2023)
- [7] OpenAI: ChatPDF, <https://www.chatpdf.com/> (参照 2023-10-26) .
- [8] OpenAI: OpenAI, <https://openai.com> (参照 2023-10-26)
- [9] S&PGlobal: TheSustainabilityYearbook-2023Rankings, <https://www.spglobal.com/esg/csa/yearbook/2023/ranking/> (参照 2023-10-26)