



JAPAN EXCHANGE GROUP

JPX WORKING PAPER

Automated Determination of
TCFD Recommended Disclosures
through Zero-shot Text Classification
Using Large Language Models

Nobushige Doi, Yusuke Oda, Naho Nakakubo, Jun Sugimoto

March 4, 2024

Vol. 43

Remarks

This material was compiled based on the results of research and studies by directors, officers, and/or employees of Japan Exchange Group, Inc., its subsidiaries, and affiliates (hereafter collectively the “JPX group”) and external researchers with the intention of seeking comments from a wide range of persons from academia, research institutions, and market users. The views and opinions in this material are the writer’s own and do not constitute the official view of the JPX group. This material was prepared solely for the purpose of providing information and was not intended to solicit investment or recommend specific issues or securities companies. The JPX group shall not be responsible or liable for any damages or losses arising from the use of this material. This English translation is intended for reference purposes only. In cases where any differences occur between the English version and its Japanese original, the Japanese version shall prevail. This translation is subject to change without notice. The JPX group shall accept no responsibility or liability for damages or losses caused by any error, inaccuracy, misunderstanding, or changes with regard to this translation.

Automated Determination of TCFD Recommended Disclosures through Zero-shot Text Classification Using Large Language Models *

Nobushige Doi[†] Yusuke Oda[‡] Naho Nakakubo[§] Jun Sugimoto[¶]

2024/03/04

Abstract

To meet the increasing global demand for sustainability information disclosures, the Task Force on Climate-related Financial Disclosures (TCFD) has established 11 recommended disclosures related to climate change, which are intended as a standard for actual disclosures. Because disclosure information is described in various formats, investigating the fulfillment status of the recommended disclosure items requires analyzing many documents. Methods that mechanically determine the fulfillment status of the disclosure information would largely reduce the cost of this task.

The present paper reports a series of measures for investigating the fulfillment status of TCFD recommended disclosures. To automatically determine individual items, the disclosure information is broken down into more basic units (TCFD recommended disclosure criteria) and the performance of zero-shot text classification is validated under each criterion. Furthermore, the method was applied to the fulfillment status of TCFD recommended disclosures in the annual security reports of companies listed on the Tokyo Stock Exchange.

*In preparing this paper, valuable comments were received from the employees of Japan Exchange Group, Inc. and others. We would like to express my deep gratitude here.

[†]Chief Researcher, Corporate Strategy, Japan Exchange Group, Inc., and Doctoral Student, Department of Advanced Interdisciplinary Studies, Graduate School of Engineering, The University of Tokyo (n-doi [at] jpx.co.jp)

[‡]Affiliate Assistant Professor, Graduate School of Science and Technology, Nara Institute of Science and Technology, and Technical Director, AI Lab, Cierpa & Company, Inc. (yusuke.oda [at] naist.ac.jp)

[§]Chief ESG Innovation Officer, Cierpa & Company, Inc. (naho.nakakubo [at] cierpa.co.jp)

[¶]Chief Executive Officer, Cierpa & Company, Inc. (jun.sugimoto [at] cierpa.co.jp)

Contents

1	Introduction	3
1.1	Background of the study	3
1.2	Purpose of this study	3
1.3	Structure of this study	5
2	Related work	6
3	TCFD disclosures	7
4	Sustainability information disclosure dataset	9
4.1	Dataset creation process	9
4.2	Constraints of the dataset	10
4.3	Statistics of the dataset	11
4.4	Creation of evaluation datasets	11
5	Proposed method	12
5.1	TCFD recommended disclosure criteria	12
5.2	Approach to zero-shot text classification	13
6	Classification-accuracy evaluation experiment	15
6.1	Experimental environment	15
6.2	Experimental results and discussion	15
6.2.1	Overall trend assessments	15
6.2.2	Evaluation of core elements and criteria	16
6.2.3	Summary of the experimental results	17
7	Investigation of the fulfillment status of TCFD recommended disclosures	18
8	Conclusion	21
A	By-category results of the classification evaluation experiments	24
B	Relationship between TCFD recommended disclosures and their criteria	25
C	Example of a prompt input into the LLMs	26

1 Introduction

1.1 Background of the study

Issues related to medium to long-term sustainability, such as climate change (including Environmental, Social, and Governance (ESG) factors), have become important management issues as companies seek to increase their medium- to long-term corporate value. To facilitate discussions on which information must be disclosed for appropriate climate risk assessment and the financial evaluation of those disclosures by investors, the Financial Stability Board established the Task Force on Climate-related Financial Disclosures (TCFD) and published “Recommendations of the Task Force on Climate-related Financial Disclosures” [1] in June 2017. The Recommendations set out four core elements related to climate change: Governance, Strategy, Risk Management, and Metrics & Targets, along with 11 recommended disclosure items (hereafter referred to as TCFD recommended disclosures).

Information disclosures on risks and opportunities related to climate change, which are organized by listed companies and also in accordance with the TCFD recommendations, are called TCFD disclosures. Through TCFD disclosures, investors can make investment decisions based on appropriate information disclosures related to climate change. However, whether TCFD disclosures provide sufficient information to investors is uncertain. For example, Binger et al. [2] investigated the TCFD disclosures of 818 companies worldwide that supported TCFD from 2015 to 2020. They found that these companies primarily disclosed information on Governance and Risk Management, with less emphasis on Strategy and Metrics & Targets. Ding et al. [3] analyzed the impact of carbon emissions on the TCFD disclosures of listed companies in the USA, UK, Canada, and Australia from 2010 to 2018. They revealed that companies with higher carbon emissions more proactively disclose their Strategy, Risk Management, and Metrics & Targets. To resolve the contention on TCFD disclosure trends, the actual situation of TCFD disclosures must be investigated in detail. In addition, the existing investigations are usually conducted at the level of the core elements of TCFD recommendations rather than at the detailed level of TCFD recommended disclosures.

Investigating the fulfillment status of TCFD recommended disclosures requires the manual analysis of all documents of the listed companies, which is burdensome and time-consuming. This burden can be reduced by automated classification using machine learning. However, TCFD disclosures and other disclosures of sustainability information (hereafter called sustainability information disclosures) are described in various formats, including natural language, on company materials, and the data are usually unstructured. Therefore, reading comprehension is essential for interpreting whether these materials fulfill the TCFD recommended disclosures. These necessities increase the cost of building the training data for traditional machine learning-based methods.

This challenge can be overcome by zero-shot text classification, a natural-language-processing task that inputs a text and a list of candidate classes and outputs the class of the input text without the need for training data. Using the zero-shot text classification model developed by Davison [4], Auzepy et al. [5] investigated the fulfillment statuses of TCFD recommended disclosures in 3,335 disclosure documents submitted by banks worldwide that supported TCFD from 2010 to 2021. However, the Micro F1-score of this method (equivalent to the accuracy measure defined later in this paper) was only 0.6029, and the zero-shot text classification model in [4] is limited to English texts. Doi et al. [6] proposed large language models (LLMs) for zero-shot text classification of TCFD recommended disclosures in the annual securities reports of Japanese listed companies. However, the accuracy of their LLMs was not improved beyond 86.0%.

1.2 Purpose of this study

Similar to Doi et al. [6], we propose an LLM-based method that mechanically classifies the fulfillment statuses of TCFD recommended disclosures. The LLMs are used for zero-shot text classification of the sustainability information disclosures in the annual securities reports of companies listed on the Tokyo Stock Exchange. In addition, as the judgment criteria of the original 11 TCFD recommended disclosures are

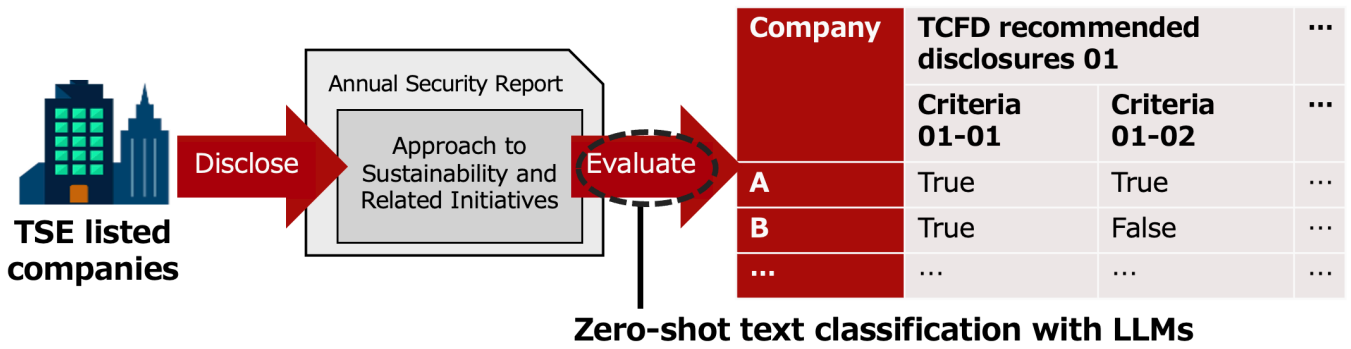


Figure 1: Overview of the proposed method

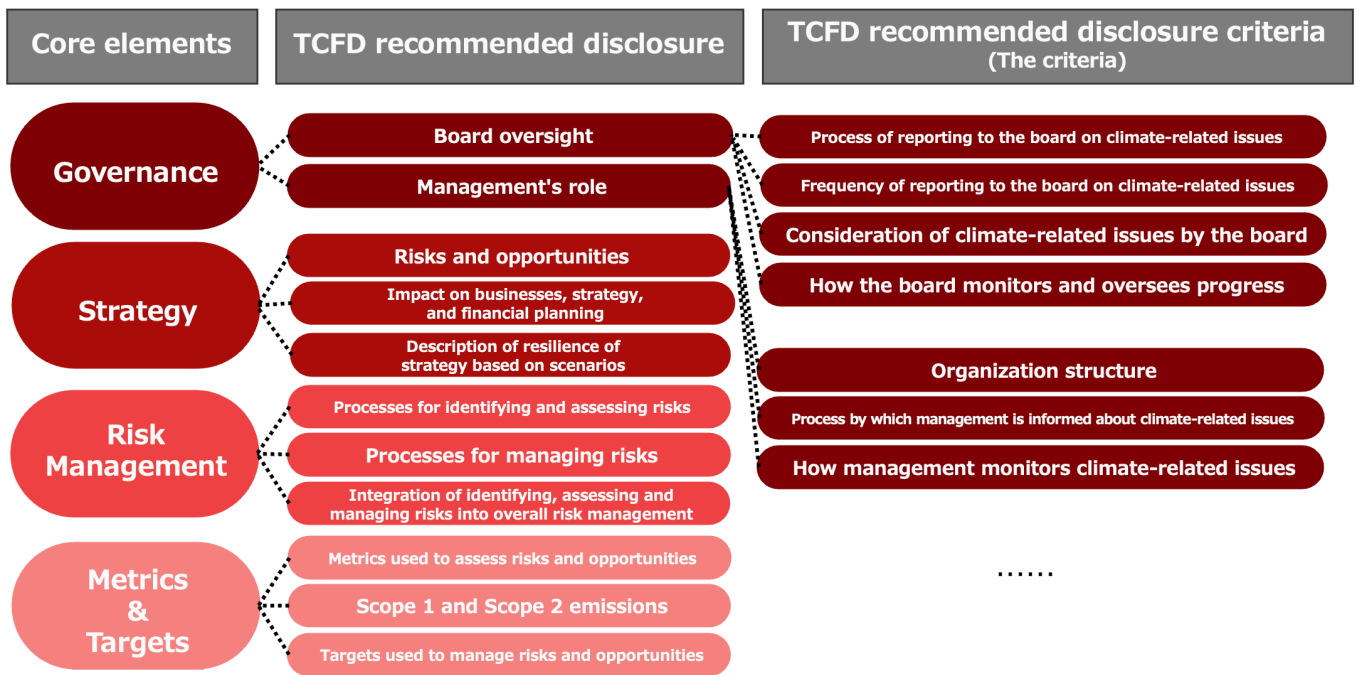


Figure 2: Abstract of the criteria

ambiguous, their applicability can be determined only by experts. To resolve this problem, we created 27 types of TCFD recommended disclosure criteria (hereafter simply called criteria) as standards for determining whether the disclosure content of each company aligns with each item. Under these criteria, both humans and machines can more easily evaluate the fulfillment status of TCFD recommended disclosures. After confirming the classification accuracy of the method, we investigate the fulfillment statuses of TCFD recommended disclosures in the companies listed on the Tokyo Stock Exchange. The proposed method is overviewed in Figure 1, and the criteria are organized as shown in Figure 2.

Our study improves that of Doi et al. [6] by refining the criteria and adopting the latest models. More specifically, we revise the contents of the criteria and employ current state-of-the-art LLMs to improve the classification accuracy from that of the previous method.

This study is divided into two phases: evaluating the proposed method and investigating the fulfillment statuses of TCFD recommended disclosures. The proposed method was evaluated as outlined in Figure 3.

This study makes several academic and practical contributions to the field. First, it proposes criteria for the systematic construction of standards, enabling mechanical evaluations of whether disclosures fulfill the TCFD recommended disclosures. Our concrete and practical indicators can improve the quality of disclosure information. Second, the study proposes an automatic evaluation method using zero-shot text classification with LLMs, which analyzes the statuses of TCFD disclosures and enables the mechanical

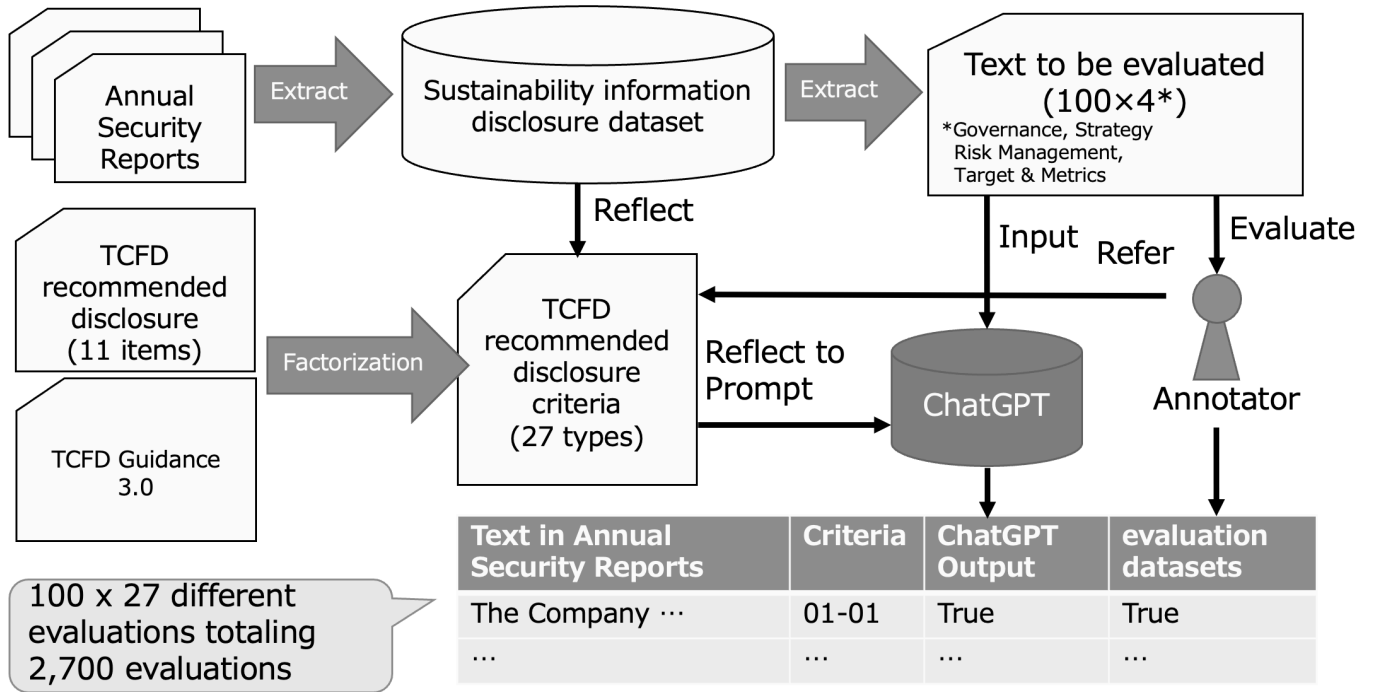


Figure 3: Flowchart of the evaluation experiment of the proposed method

evaluation of numerous text data. Such evaluations are not easily performed using traditional methods.

However, the present study has three main limitations. First, the proposed method is evaluated on a high-quality but small dataset (only 400 entries), which potentially limits the applicability of the experimental results. Second, the details of the LLMs used in the experiments are not clarified, which limits the reproducibility of the evaluation results. Third, only the texts of the annual security reports and the converted tables are analyzed, excluding images and other documents from the study.

1.3 Structure of this study

The remainder of this study is structured as follows. Section 2 overviews the relevant existing research and clarifies the present research. Section 3 explains the abstraction of the TCFD disclosures and Section 4 details the sustainability information disclosure dataset used in this study. Section 5 introduces the proposed approach based on zero-shot text classification, which automatically determines TCFD recommended disclosures, and introduces the TCFD recommended disclosure criteria. Section 6 reports the environment, methods, and results of the verification experiment of the proposed method and discusses the performance results. In Section 7, the fulfillment status of TCFD recommended disclosures in the annual security reports of companies listed on the Tokyo Stock Exchange are evaluated using the proposed method. Section 8 discusses the contributions of the proposed method and suggests future research directions.

2 Related work

This section reviews the existing research on TCFD disclosures and the application of zero-shot text classification.

We first discuss the studies on TCFD recommended disclosures. Bingler et al. [2] used a pretrained Bidirectional Encoder Representations from Transformers (BERT) [7] model called ClimateBERT [8], which classifies English input texts into Governance, Strategy, Risk Management, or Metrics & Targets. They reported that TCFD recommendations have not significantly influenced the disclosures of TCFD-supporting companies worldwide. The TCFD then developed a classification model that determines whether company disclosures actually align with the 11 items in the TCFD recommended disclosures. This classification model, included in the 2023 Status Report [9] of the TCFD, creates training data and fine-tunes a language model called the Robustly optimized BERT pretraining approach (RoBERTa) [10]. The model of Friederich et al. [11] classifies climate-related risks in the annual reports of European companies into five types using RoBERTa and other machine learning methods [10]. Moreno and Caminero [12] classified TCFD recommended disclosures in text with named entity recognition, targeting the corporate reports of Spanish financial institutions. Unlike the previous studies, the current study adopts Japanese as the input language and identifies 27 multiclass types through zero-shot text classification.

Zero-shot learning with LLMs largely reduces the dataset construction and labeling costs. Specifically, it analyzes and classifies unknown texts without requiring prior training for a specific task. Kuzman et al. [13] reported that OpenAI's ChatGPT achieves higher accuracy than finely tuned models in zero-shot text classification of nine types of sentences, such as news and promotions, in English and Slovenian datasets. Doi et al. [14] tested the performances of multiple models in classifying ten topics related to key audit matters disclosed in the audit reports of Japanese listed companies. They reported that the ChatGPT model GPT-4 [15] outperformed the other models. Based on these findings, the present study demonstrates the utility of LLMs in zero-shot text classification.

3 TCFD disclosures

The importance of disclosing climate-change information is recognized worldwide because such disclosures exert extensive and urgent impacts. To mitigate the climate-change-induced destabilization of financial markets, the Financial Stability Board launched the TCFD in 2015 and published the TCFD recommendations in June 2017, which focus on climate change within sustainability information disclosures.

The TCFD recommendations provide a framework on which climate-change-related risks and opportunities of corporate finances can be reported to investors and other stakeholders [16]. Based on the “Final Report: Recommendations of the Task Force on Climate-related Financial Disclosures” [17], four core elements have been established for disclosing risks and opportunities related to climate change:

Governance: The governance of an organization regarding climate-related risks and opportunities

Strategy: The actual and potential impacts of climate-related risks and opportunities on the business, strategy, and financial planning of an organization

Risk Management: Processes by which an organization identifies, assesses, and manages climate-related risks

Metrics & Targets: The metrics and targets used for assessing and managing relevant climate-related risks and opportunities

Based on these core elements, we can understand and assess how a company perceives and evaluates its climate change-related risks and opportunities and how these risks and opportunities will financially affect the company. Thus, TCFD disclosures enable investors and other stakeholders to comprehend the potential impacts of climate change on companies and the management strategies and methods of these companies, facilitating informed decision-making.

Japan’s Corporate Governance Code [18] is revised in 2021, encouraging listed companies to actively and proactively address their sustainability-related issues under a comply-or-explain framework. In particular, companies listed on the Prime Market must collect and analyze necessary data on the impact of climate-related risks and revenue opportunities on their business activities and earnings. They must also improve the quality and quantity of their disclosures based on the TCFD or an equivalent framework.

Furthermore, since the “Cabinet Office Ordinance on Disclosure of Corporate Affairs” was amended in January 2023, disclosures related to sustainability initiatives must be included in the “Approach to Sustainability and Related Initiatives” section of the annual securities reports for the fiscal year ending in March [19]. The four core elements in the disclosure structure are those of the TCFD: Governance, Risk Management, Strategy, and Metrics & Targets. To mitigate the disclosure burden on companies, all companies are expected to disclose their Governance and Risk Management but the disclosure of Strategy and Metrics & Targets is optional and depends on the assessment of its importance by companies [20]. As a reference, Figure 4 displays part of the sustainability disclosure in the annual securities report of the Japan Exchange Group, Inc. for the fiscal year ending March 2022 [21, 22]. The text is displayed in both Japanese and English.

2 【サステナビリティに関する考え方及び取組】

(1) 考え方・体制

当社グループは、企業理念で掲げる「市場の持続的な発展を図り、豊かな社会の実現に貢献」に向け、我々を取り巻く環境や社会課題、それらとの関係に目を向け、企業価値の向上につながる取組を進めることが重要な経営課題の一つであると認識し、経営方針を定め、経営計画等を策定しています（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参照）。

公正性・信頼性を備えた利便性・効率性及び透明性が高い市場と魅力的なサービスを提供するという当社グループのビジネスモデルを踏まえると、市場メカニズムを活用した取組を進めていくことが肝要と考え、長期ビジョンのもと、中期経営計画2024では3つのFocusの一つに「社会と経済をつなぐサステナビリティの推進」を掲げ、サステナビリティ関連情報の発信に係る機能強化や、関連指数算出・商品の上場、排出量市場創設の推進等に取り組んでいます。

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については、当社ウェブサイトもご参照ください。

<https://www.jpx.co.jp/corporate/sustainability/index.html>

2 Approach to Sustainability and Related Initiatives

(1) Approach and Organization

In order to “contribute to the realization of an affluent society by promoting sustainable development of the market” as set out in its Corporate Philosophy, the Group works on initiatives to improve its corporate value with attention to the surrounding environment and societal issues and the Group’s relationship to them. The Group considers this an important management issue and formulates its management policies and plans accordingly (see “1 Management Policy, Business Environment, and Issues to be Addressed” in “II Overview of Business”).

Given its business model, which is to provide attractive services along with a highly convenient, efficient, and transparent market that ensures fairness and reliability, the Group considers it essential to work on initiatives that utilize market mechanisms. Accordingly, in line with its long-term vision, the Group set out “promoting sustainability that connects society and economy” as one of the three Focus areas in its Medium-Term Management Plan 2024 and is working on related initiatives such as enhancing dissemination functions for sustainability-related information, calculating sustainability-related indices and listing-related products, and advancing the creation of an emissions trading market.

For more information on the Group’s approach to sustainability and related initiatives, please see the below webpage.

<https://www.jpx.co.jp/english/corporate/sustainability/index.html>

Figure 4: Part of the sustainability disclosures of the Japan Exchange Group, Inc. for the fiscal year ending March 31, 2022 (top: Japanese, bottom: English)

4 Sustainability information disclosure dataset

Sustainability information disclosures are found in annual securities reports, integrated reports, annual reports, ESG/CSR/environmental/sustainability reports, and TCFD reports. Media other than annual security reports, commonly published in PDF, are disclosed voluntarily by listed companies. To ensure a consistent analysis, the present study is limited to annual security reports published in HTML format.

The sustainability information disclosure dataset (hereafter simply called the dataset) aggregates the text related to sustainability information disclosures from the annual securities reports of companies listed on the Tokyo Stock Exchange as of October 31, 2023. The aim is to investigate the fulfillment status of TCFD recommended disclosures. The classification accuracy of the proposed method was verified on subsets of this dataset (100-text evaluation datasets for each of the four core elements).

The following subsections describe the dataset creation process, statistical measures, and evaluation datasets in the present study.

4.1 Dataset creation process

The dataset creation process includes the following steps: (1) collecting the annual securities reports, (2) extracting the text fields related to sustainability, (3) extracting text according to XBRL tags, (4) extracting the XBRL tags related to TCFD disclosures, and (5) normalizing the text. These five steps are detailed below.

(1) Collection of annual securities reports: Annual securities reports submitted from April 1, 2023, to October 31, 2023, were collected using the Electronic Disclosure for Investors' NETwork (EDINET) application programming interface (API) and filtered under the following conditions:

- The fiscal year ended on March 31, 2023, or later.
- The issuer is a domestic stock-listed company on the Tokyo Stock Exchange as of October 31, 2023.

(2) Extraction of text fields related to sustainability: The text under "Approach to Sustainability and Related Initiatives" in the annual securities reports is described within the XBRL tag "DisclosureOfSustainabilityRelatedFinancialInformationTextBlock." The text within this XBRL tag is extracted.

(3) Extraction of text according to the XBRL tags: The element names of the XBRL tags are extracted along with the associated text. The text within "Approach to Sustainability and Related Initiatives" forms a multilayered structure of various XBRL tags, including those uniquely defined by each company. Therefore, the extraction of XBRL tags is limited to two levels at most, treating the text within more deeply defined XBRL tags as regular text. Texts enclosed in p tags, h3 tags, or h4 tags are treated without conversion, whereas the text in tables is converted to double-listed strings.

(4) Extraction of XBRL tags related to TCFD disclosures: Texts related to TCFD disclosures are extracted from "Approach to Sustainability and Related Initiatives," which also includes texts unrelated to climate issues, such as human capital. To this end, among the texts associated with the extracted XBRL tag element names (first and second levels), we extract the texts containing any of the terms "Governance," "Strategy," "RiskManagement," or "MetricsAndTargets" in either the first or second level of the XBRL tag element names.

(5) Text normalization: The text is normalized through the following steps:

- NFKC normalization of characters, excluding ellipse and circle numbers
- Replacement of characters resembling hyphens with hyphens

Table 1: Example dataset (some text excerpts from the Japan Exchange Group, Inc.)

Code	1stXBRL	2ndXBRL	Text (Reference Translation)
86970	Governance TextBlock		Governance In line with the above approach, the Group has established a Sustainability Committee with the Group CEO as Chair and Group COO as Vice-Chair. The Committee has analyzed how related issues affect business operations and is proceeding to address these. The Group has also established a system where overall approaches and important matters relating to these issues are reported to the board of directors where necessary, ensuring appropriate oversight (see “4 Corporate Governance” in “TV Information about the Reporting Company”). (Omitted)
86970	Risk Management TextBlock		Risk Management JPX Group has established a Risk Policy Committee chaired by an outside director and a Risk Management Committee chaired by the Group CEO in order to address the various risks faced by the Group. In line with the Group’s Risk Management Policy, these committees are responsible for preventing risk materialization through identification of risks and development and management of countermeasures, as well as preparing processes for a swift and appropriate response in the case risks do or are likely to materialize. (Omitted)
86970	Strategy TextBlock		Strategy JPX Group has considered the possible risks and opportunities brought on by climate change and their effects on operations, strategy, and financial planning. It is accordingly implementing measures to reduce risks and increase corporate value and has summarized these initiatives as its Green Strategy in the Medium-Term Management Plan 2024. (Omitted)
86970	Strategy TextBlock	Reference ToOther Information Strategy	For details on the Group’s scenario analysis, please see the below page. https://www.jpx.co.jp/english/corporate/sustainability/jpx-esg/environment/01.html

- Replacement of consecutive whitespace characters with a single space

These processes refined and selected the texts related to TCFD disclosures from the annual security reports of companies listed on the Tokyo Stock. The resulting dataset contained the texts on Governance, Strategy, Risk Management, and Metrics & Targets of each listed company, allowing analyses of the fulfillment statuses of the TCFD disclosures.

Table 1 is an example dataset created through the above procedures.

4.2 Constraints of the dataset

The dataset is limited by the following three constraints.

- (1) Restricted period of annual securities report submissions:** The dataset is based on annual securities reports submitted to EDINET from April 1, 2023, to October 31, 2023. Therefore, it covers only part of the 2023 fiscal year (ending March 2024) and excludes reports submitted after this period, for example, the reports of companies with a fiscal year ending in August 2023 or later. This constraint limits the comprehensiveness of the data, preventing a complete picture of companies listed on the Tokyo Stock Exchange.

Table 2: Statistics of the dataset arranged by XBRL tags

Strings contained in the XBRL tags	Number of Texts	Number of Companies	Number of Characters	Average Number of Characters per Company
Governance	2,721	2,178	1,071,287	491.9
Strategy	4,551	2,164	4,291,547	1,983.2
RiskManagement	2,667	2,145	987,729	460.5
MetricsAndTargets	4,233	2,118	2,475,695	1,168.9
Whole	13,530	2,198	8,089,292	3,680.3

- (2) **Use of text information only:** The dataset is limited to text information and tables (which are converted to double lists and used as text information) and excludes visual details such as images. Visual information, such as numerical data and graphs depicting future scenario analyses, is also crucial in TCFD disclosures. The inability to capture visual elements in the dataset limits the analysis.
- (3) **Non-consideration of external materials:** Annual securities reports may include limited TCFD disclosures by some listed companies and instead provide references to external materials, including sustainability reports. The dataset extracts text only from annual security reports and excludes text from external materials. Consequently, the present analysis is limited to the information described in the annual securities reports, preventing a comprehensive capture of the full scope of sustainability initiatives by the listed companies.

4.3 Statistics of the dataset

The dataset was extracted from the annual securities reports of Japanese companies listed on the Tokyo Stock Exchange as of October 31, 2023, and submitted from April 1, 2023, to October 31, 2023. From these annual security reports, we extracted the text related to sustainability information, finally obtaining the data of 2,198 companies totaling 13,530 entries with a character count of 8,089,292.

Table 2 shows the number of texts and characters contained in the XBRL tags. Note that some XBRL tag element names, such as “MetricsAndTargetsDataGovernanceTextBlock,” contain the names of multiple core elements. Therefore, the summed number of texts (characters) in each core element does not always match the total number of texts (characters).

4.4 Creation of evaluation datasets

The performance of the model, which is designed to automatically assess the fulfillment status of TCFD recommended disclosures, was tested on evaluation datasets of texts extracted from the dataset. The evaluation datasets contained 100 randomly selected texts of the four core elements, with each sample manually labeled by the authors. The criteria of these labels are mentioned later.

The evaluation datasets were created as follows:

1. Randomly extract 100 texts corresponding to Governance, Strategy, Risk Management, and Metrics & Targets from the dataset (400 texts in total).
2. Manually label each extracted text with the fulfillment status (binary classification: fulfill or does not fulfill) under Governance, Strategy, Risk Management, or Metrics & Targets.

On these evaluation datasets, we can test the performance of the proposed automatic evaluation model and accurately capture the fulfillment statuses of TCFD recommended disclosures in annual securities reports.

The fulfillment status of each criterion in the evaluation datasets is shown in Appendix A.

5 Proposed method

In this study, the fulfillment statuses of TCFD recommended disclosures are mechanically classified using LLMs for zero-shot text classification. The LLMs are pre-trained, eliminating the need for additional training data for the specific classification task. Moreover, the judgment criteria of the original 11 TCFD recommended disclosures are potentially ambiguous to both humans and machines. To evaluate the alignment between a company’s disclosures and each item, the present study develops 27 criteria based on the 11 TCFD recommended disclosures.

In the following subsections, we detail the criteria and discuss the proposed method based on zero-shot text classification.

5.1 TCFD recommended disclosure criteria

<p>ID: TCFD-G-01-01</p> <p>Title: Process of reporting to the board on climate-related issues</p> <p>Definition: Does the text describe the processes by which the board and/or board committee are informed about climate-related issues?</p> <p>Remarks: <ul style="list-style-type: none">• Examples of board committees could include an audit committee, risk committee, etc.• Climate-related issues refer to climate-related risks and opportunities.• Sustainability efforts are considered to be related to climate-related issues.• Content related to future plans will not be considered.</p>
--

Figure 5: Sample criterion

The 11 items of the TCFD recommended disclosures were divided into 27 criteria, providing specific evaluation standards for each item and enabling detailed analysis and precise assessments of the disclosures. Although a multitiered scale indicating the compliance degree of each criterion would enable a more specific analysis, a binary assessment (whether the criterion is fulfilled or not) was chosen to simplify the classification.

Each criterion consists of an ID, a title, a related TCFD recommended disclosure, a definition of the criterion, and notes. Each criterion, which is uniquely determined by its ID and title, is linked to the TCFD recommended disclosures. Therefore, the item associated with each criterion is clearly stated within that criterion. The core of a criterion is its definition, which was maintained as concisely as possible to eliminate ambiguity. Notes include definitions of terms or exceptional conditions. When improving the accuracy of the proposed method, the content of each criterion was revised with minimal adjustments to the definition of the criteria, and the classification results were corrected in the notes. Figure 5 is an image of a criterion.

When creating the criteria, we referred to “Guidance on Climate-related Financial Disclosures 3.0” (hereafter referred to as TCFD Guidance 3.0) [23, 24]. Each criterion reflects the guidance content of a TCFD recommended disclosure. In other words, the criteria were designed to factorize the TCFD recommended disclosures and TCFD Guidance 3.0 and to ensure that compliance with one criterion was independent of other outcomes; that is, the judgment criteria are independent. The example in Figure 6

Guidance on Climate-related Financial Disclosures 3.0 (TCFD Guidance 3.0)

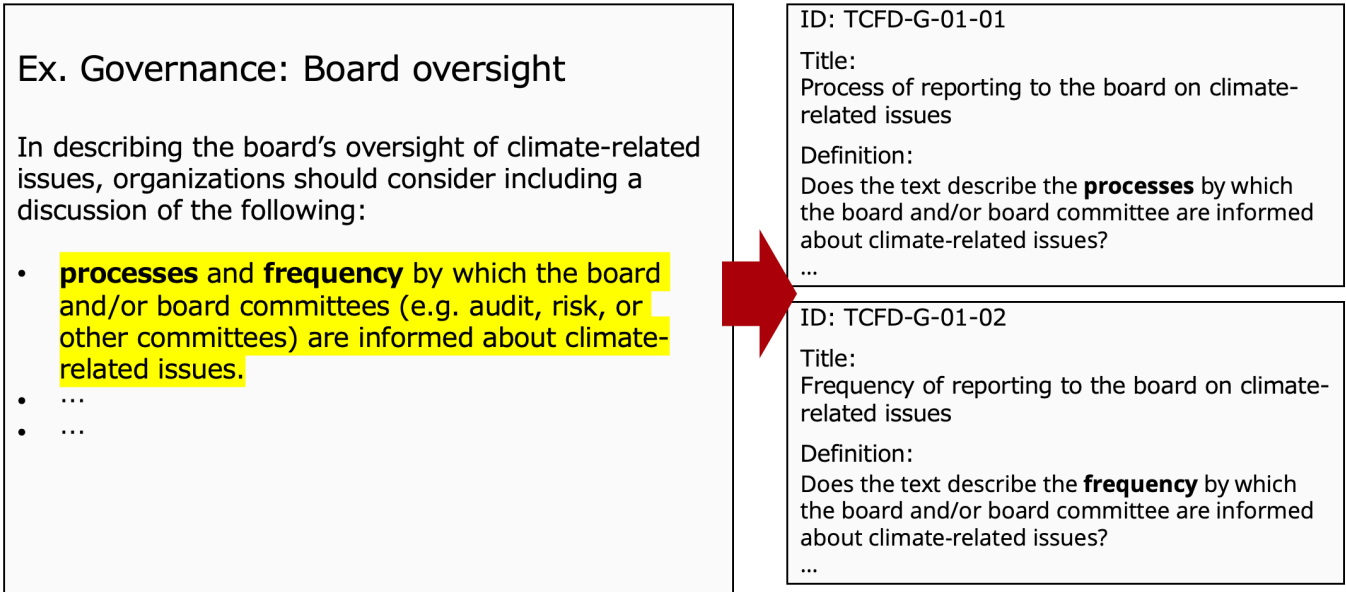


Figure 6: Example showing the design of the criteria

shows the extraction of the content disclosed by an organization as an independent criterion from the TCFD recommended disclosure “Board oversight.”

Modifications to the criteria reflect (to some extent) the actual content of disclosures by Japanese listed companies at the current time in the dataset. For example, the terms used in actual disclosures are provided as examples of ambiguous terms in the TCFD recommended disclosures. Instead of targeting all texts in the dataset, representative samples were extracted for criteria-tuning. Specifically, texts were compressed into 200 dimensions using Latent Semantic Analysis [25], and 30 texts representing each of the four core elements were extracted using K-means.

The relationship between the TCFD recommended disclosures, and the criteria is documented in Appendix B.

The criteria are available at the following GitHub repository link:

https://github.com/cierpa/tcfid_criteria

5.2 Approach to zero-shot text classification

The text from the dataset or evaluation datasets, along with the content of the criteria, are input as prompts to the LLMs to determine whether the text fulfills the criteria. The prompts for each criterion corresponding to each core element are executed by referencing the element name of the XBRL tag corresponding to the text. For each text, prompts are executed for seven Governance criteria, ten Strategy criteria, four Risk Management criteria, and six Metrics & Targets criteria. When the XBRL tag element name of a text includes multiple core elements, such as “MetricsAndTargetsDataGovernanceTextBlock,” prompts are executed for the criteria corresponding to each core element.

Moreover, the prompt input to the LLMs explicitly states that the output be True if the criterion is fulfilled and False if the criterion is not met.

In general, the accuracy of text classification by LLMs largely depends on the prompt design. In this study, the prompts are the TCFD recommended disclosure criteria, and supplementary information is entered as notes.

Figure 7 is an overview of zero-shot text classification in the proposed method, and Appendix C provides examples of prompts.

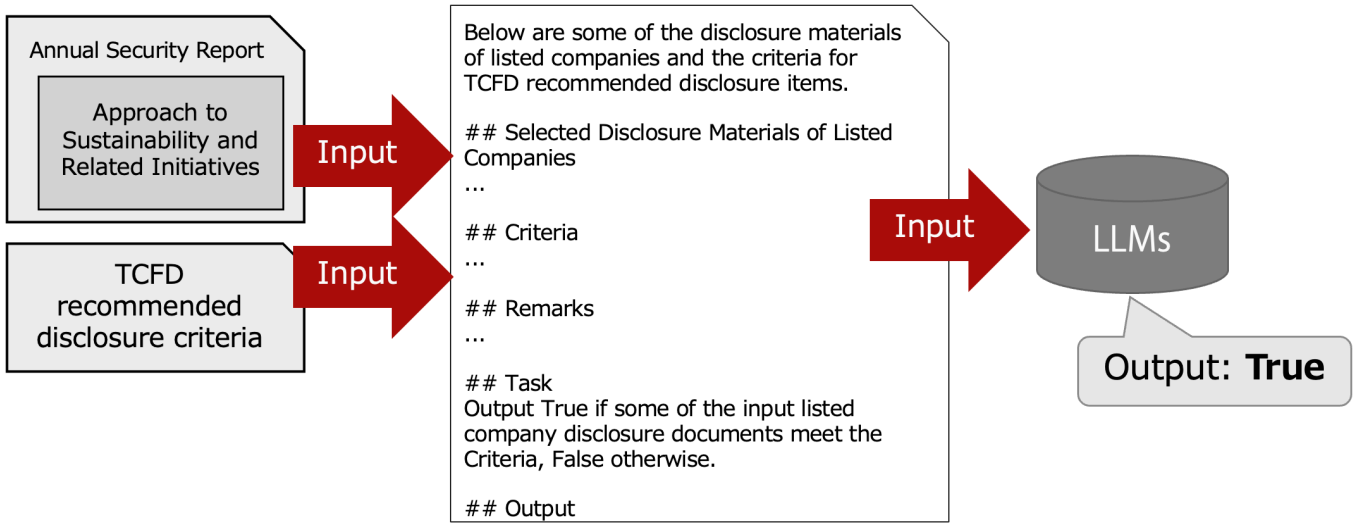


Figure 7: Overview of zero-shot text classification using the proposed method

6 Classification-accuracy evaluation experiment

The effectiveness of the proposed method was evaluated through a classification accuracy test on four evaluation datasets. Each evaluation dataset contains 100 texts related to a core element (Governance, Strategy, Risk Management, or Metrics & Targets).

Below, we describe the experimental environment and discuss the experimental results and considerations.

6.1 Experimental environment

Experiments were performed using ChatGPT (OpenAI) models¹ (GPT-3.5’s gpt-3.5-turbo-0125 and GPT-4’s gpt-4-0125-preview), which execute prompts through the ChatGP API².

The evaluation datasets comprise 400 texts in total: 100 texts corresponding to each of the four core elements. For each text, prompts were executed for seven Governance criteria, 10 Strategy criteria, four Risk Management criteria, and six Metrics & Targets criteria. This process obtains a True (fulfilled) or False (not fulfilled) judgment of each criterion within each text.

The criteria and core elements were evaluated in terms of accuracy, precision, recall, and F1-score. The accuracy defines the ratio of correctly classified evaluations among all evaluations. Precision is the ratio of predicted True data that are actually True, Recall is the ratio of actual True data that are predicted as True, and the F1-score (the harmonic mean of Precision and Recall) evaluates the overall accuracy of the proposed method. The accuracy, precision, recall, and F1 score are calculated using Eqs. (1), (2), (3), and (4), respectively, where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

6.2 Experimental results and discussion

After presenting the overall trends in the results and considerations, we analyze and discuss the trends in the individual core elements and several specific criteria.

6.2.1 Overall trend assessments

Table 3 presents the accuracies, precisions, recalls, and F1-scores of GPT-3.5 and GPT-4 on the 400 texts on the evaluation datasets.

From Table 3, we observe that GPT-4 achieves a slightly higher classification accuracy than GPT-3.5 in zero-shot text classification. The improvements in both accuracy and F1-score suggest that GPT-4 is more suited for automatically determining TCFD recommended disclosures than GPT-3.5, likely because GPT-4 possesses an inferencing capability that improves the interpretability of complex natural language contexts in the sustainability disclosures. Therefore, GPT-4 can infer accurate classification judgments.

¹<https://platform.openai.com/docs/models>

²<https://platform.openai.com/docs/api-reference>

Table 3: Comparison of classification accuracies of GPT-3.5 and GPT-4

LLMs	Accuracy	Precision	Recall	F1-Score
GPT-3.5	59.3%	36.3%	95.8%	52.7%
GPT-4	92.8%	77.2%	94.4%	84.9%

Table 4: Classification results of GTP-4 on the core elements of the dataset

Core elements	Accuracy	Precision	Recall	F1-Score
Governance	89.6%	86.9%	97.1%	91.6%
Strategy	95.3%	68.0%	94.1%	77.9%
Risk Management	87.3%	72.7%	92.7%	81.4%
Metrics & Targets	96.0%	84.2%	93.0%	88.1%
Whole	92.8%	77.2%	94.4%	84.9%

Compared with the experimental results of Doi et al. [6], the proposed approach improved the accuracy and F1-score of GPT-4 from 86.0% to 92.8% and from 72.9% to 84.9%, respectively. These results indicate that refining the criteria and applying the latest model improves the classification accuracy from that of prior research.

Both GPT-3.5 and GPT-4 achieved a high recall score, indicating that the LLMs captured the content related to TCFD recommended disclosures. Conversely, the relatively low precision scores suggest difficulties in classifying unrelated content. The low precision is possibly sourced from the design of the prompts or context misinterpretation by the LLMs and could be improved by a more detailed prompt design or additional tuning that enhances the LLMs’ understanding of context.

6.2.2 Evaluation of core elements and criteria

This subsection evaluates the classification outcomes of GPT-4, which outperformed GPT-3.5 in the previous subsection. The classification results of GPT-3.5 and GPT-4 on all criteria are documented in Appendix A.

Table 4 presents the classification results on each core element using the proposed method with GPT-4.

Overall, GPT-4 with the proposed method achieved very high classification accuracies across all core elements.

The higher F1-scores of Governance and Metrics & Targets than those of the other elements suggest that the proposed method is particularly suitable for classifying these categories.

Meanwhile, high recall scores were obtained for all core elements, indicating that the proposed method can recognize texts related to TCFD recommended disclosures. The recall results were considerably improved from those of Doi et al. [6], who reported low recalls on some Governance criteria. It was inferred that our approach clarifies ambiguous judgment criteria by refining the criteria and enhancing the interpretive abilities using the latest model.

Finally, the precision scores were lower for Strategy and Risk Management than for the other core elements, leading to lower F1-scores for these categories. This finding indicates that unrelated content can be misclassified under the criteria in these core elements. Examining the trends of specific criteria, the precision scores were lowered for Strategy03-02 (precision = 55.6%), Strategy03-03 (precision = 52.9%), Strategy04-01 (Precision = 54.2%), Risk Management08-01 (Precision = 61.5%) and Risk Management08-02 (Precision = 66.7%).

These results suggest that the prompt design should be improved or the classification process optimized to enhance the classification accuracy of specific criteria. Specifically, for the low precision criteria, texts of low relevance might be incorrectly classified as fulfilling the criteria in many cases; thus, incorporating more specific contexts or terms into the prompt could improve the classification accuracy.

6.2.3 Summary of the experimental results

Overall, the proposed method using GPT-4 achieved high classification accuracy and high precision and recall for the core elements of Governance and Metrics & Targets, indicating that the proposed method is especially suited for classifying these categories. However, judging from the low precision for the core elements of Strategy and Risk Management, the design of the prompts should be improved, and the classification process should be optimized.

The results should be generalized with caution because the evaluation datasets are relatively small (only 400 cases in total). Moreover, the inner workings of the LLMs are opaque and should be clarified to enhance the validity of the evaluation results. Furthermore, the evolution of LLMs has provided new insights into model selection and prompt design methods, highlighting the importance of regular model updates and method reevaluations.

7 Investigation of the fulfillment status of TCFD recommended disclosures

Using the proposed method, this section investigates the fulfillment status of TCFD recommended disclosures in the sustainability information in the annual securities reports of companies listed on the Tokyo Stock Exchange. The proposed method was applied to the previously created dataset to mechanically determine which TCFD recommended disclosures were addressed by each listed company. As mentioned in subsection 4.2, the dataset is limited to text information in the annual security reports and excludes information related to images or external materials.

Among 13,530 texts from all 2,198 companies, texts containing the strings “Governance,” “Strategy,” “RiskManagement,” and “MetricsAndTargets” were assigned to Governance, Strategy, Risk Management, and Metrics & Targets, respectively, and were classified based on these criteria. Specifically, if the text of a listed company fulfilled a criterion, that company was considered to meet that criterion, aggregating the classification results at the company level.

Table 5 compares the classification results of the proposed method on the dataset and the correct-answer ratios of the evaluation datasets during the classification evaluation experiment. The evaluation datasets consist of 100 pieces randomly extracted from their corresponding datasets. If the accuracy of the proposed method is sufficiently high, the classification results of the method on the dataset should largely match the correct answers in the evaluation datasets. Hence, we subtracted the percentage of True results (fulfilling each criterion) of the proposed method from the percentage of True results on the evaluation datasets during the classification evaluation experiment. As shown in Table 5, this difference was small for most criteria but deviated for the low precision criteria during the classification evaluation experiment, including 04–01. When the precision was low in the classification evaluation experiment, the proposed method could overestimate the true fulfillment rate in the actual distribution. Therefore, to more accurately grasp the fulfillment status of each criterion across the entire dataset, we must improve the precision scores of these criteria.

Table 6 presents the aggregated results of the number of companies fulfilling each TCFD recommended disclosure obtained from the classification results of the proposed method on the dataset. The results allow us to observe an overall trend in fulfilling TCFD recommended disclosures in the annual securities reports of companies listed on the Tokyo Stock Exchange. Notably, disclosures related to Governance and Risk Management are reasonably well fulfilled, although disclosures related to Strategy and Metrics & Targets remain unfulfilled by many companies. This trend aligns with the Cabinet Office Ordinance on Corporate Disclosure, which mandates the disclosure of Governance and Risk Management by all listed companies from the fiscal year ending March 2023 onward but allows optional disclosures on Strategy and Metrics & Targets [20]. As Strategy and Risk Management obtained lower precision scores than the other categories in the classification evaluation experiment, the method might over-represent the fulfillment numbers in these categories from those in the true distribution.

Furthermore, many specific criteria reveal disclosure deficiencies in certain items, suggesting that investors and stakeholders cannot adequately assess information related to climate change risks and opportunities. For example, although 77.9% of companies fulfill the “Process of reporting to the board on climate-related issues,”(01-01) only 33.4% mention the “Frequency of reporting to the board on climate-related issues.”(01-02) This result indicates that although companies provide TCFD recommended disclosures, they might not sufficiently provide more detailed information, such as the reporting frequency.

The Japan Exchange Group Inc. aims to understand the extent to which the climate-related information disclosures of Japanese companies comply with the TCFD recommendations and to provide assistance for listed companies disclosing their climate-related information. To this end, it published its “Survey on the Reality of Information Disclosure in Compliance with TCFD Recommendations” [26] in November 2021, which targeted 259 TCFD-endorsing listed companies (as of the end of March 2021), and the “Survey on the Reality of Information Disclosure in Compliance with TCFD Recommendations (Fiscal Year 2022)” [27] in January 2023, which targeted constituents of the JPX-Nikkei Index 400. Based on the classification results, a more detailed survey will be conducted in fiscal year 2023, and the results will be published.

Table 5: Comparison of (1) classification results of the proposed method on the dataset and (2) percentage of correct answers in the evaluation datasets during the classification evaluation experiment

ID	Title	(1) Proposed method	(2) Evaluation datasets	(1)-(2)
01-01	Process of reporting to the board on climate-related issues	71.3%	69.0%	2.3%
01-02	Frequency of reporting to the board on climate-related issues	28.7%	38.0%	-9.3%
01-03	Consideration of climate-related issues by the board	69.6%	70.0%	-0.4%
01-04	How the board monitors and oversees progress	47.5%	47.0%	0.5%
02-01	Organizational structure	73.9%	66.0%	7.9%
02-02	Process by which management is informed about climate-related issues	77.8%	71.0%	6.8%
02-03	How management monitors climate-related issues	66.5%	66.0%	0.5%
03-01	Time horizon(s) for the consideration of climate-related risks and opportunities	15.9%	14.0%	1.9%
03-02	Climate-related issues in each specified time horizon	10.8%	5.0%	5.8%
03-03	Process for determining risks and opportunities with a financial impact	18.4%	9.0%	9.4%
04-01	Impact on businesses or strategy	24.6%	13.0%	11.6%
04-02	Impact on financial planning	7.9%	5.0%	2.9%
05-01	2 ° C or lower scenario	11.0%	11.0%	0.0%
05-02	Climate-related scenarios and associated time horizon(s)	11.9%	7.0%	4.9%
05-03	Impact of climate-related scenarios on the strategy	13.5%	8.0%	5.5%
05-04	Strategic action taken in relation to climate-related scenarios	12.2%	7.0%	5.2%
05-05	Impact of climate-related scenarios on financial planning	11.8%	8.0%	3.8%
06-01	Process for identifying and assessing risks	60.1%	50.0%	10.1%
07-01	Process for managing risks	73.1%	62.0%	11.1%
08-01	Integration of the identification and assessment process into the overall risk management	27.6%	18.0%	9.6%
08-02	Integration of the management process into overall risk management	28.0%	21.0%	7.0%
09-01	Evaluation metrics	28.0%	25.0%	3.0%
09-02	Historical performance based on evaluation metrics	9.2%	6.0%	3.2%
10-01	Greenhouse gas emissions	20.8%	20.0%	0.8%
10-02	Historical greenhouse gas emissions	6.7%	6.0%	0.7%
11-01	Climate-related targets	26.5%	24.0%	2.5%
11-02	Time horizon(s) for climate-related targets	26.0%	23.0%	3.0%
	Average	32.6%	28.5%	4.1%

Table 6: Number of TCFD recommended disclosures fulfilled by each listed company, based on the dataset classification results of the proposed method

ID	Title	Number of Companies Fulfilling the Disclosure	Percentage of Total Listed Companies (2,198 companies)	Fulfillment Rate of the TCFD recommended Disclosure (Ave.)
01-01	Process of reporting to the board on climate-related issues	1713	77.9%	60.2%
01-02	Frequency of reporting to the board on climate-related issues	735	33.4%	
01-03	Consideration of climate-related issues by the board	1673	76.1%	
01-04	How the board monitors and oversees progress	1175	53.5%	
02-01	Organizational structure	1702	77.4%	78.3%
02-02	Process by which management is informed about climate-related issues	1851	84.2%	
02-03	How management monitors climate-related issues	1607	73.1%	
03-01	Time horizon(s) for the consideration of climate-related risks and opportunities	677	30.8%	29.7%
03-02	Climate-related issues in each specified time horizon	487	22.2%	
03-03	Process for determining risks and opportunities with a financial impact	796	36.2%	
04-01	Impact on businesses or strategy	1024	46.6%	31.4%
04-02	Impact on financial planning	358	16.3%	
05-01	2 ° C or lower scenario	495	22.5%	24.6%
05-02	Climate-related scenarios and associated time horizon(s)	534	24.3%	
05-03	Impact of climate-related scenarios on the strategy	604	27.5%	
05-04	Strategic action taken in relation to climate-related scenarios	542	24.7%	
05-05	Impact of climate-related scenarios on financial planning	528	24.0%	
06-01	Process for identifying and assessing risks	1459	66.4%	66.4%
07-01	Process for managing risks	1713	77.9%	77.9%
08-01	Integration of the identification and assessment process into the overall risk management	707	32.2%	32.3%
08-02	Integration of the management process into overall risk management	715	32.5%	
09-01	Evaluation metrics	1076	49.0%	32.9%
09-02	Historical performance based on evaluation metrics	372	16.9%	
10-01	Greenhouse gas emissions	822	37.4%	25.0%
10-02	Historical greenhouse gas emissions	276	12.6%	
11-01	Climate-related targets	1008	45.9%	45.2%
11-02	Time horizon(s) for climate-related targets	979	44.5%	

8 Conclusion

We proposed and evaluated an LLM-based method that classifies TCFD recommended disclosures using zero-shot text classification. The proposed method mechanically evaluates the fulfillment statuses of TCFD recommended disclosures among the vast amount of text data contained in annual securities reports. Based on these results, we could understand the fulfillment statuses of TCFD recommended disclosures in the annual securities reports of companies listed on the Tokyo Stock Exchange. Furthermore, the proposed method classified the fulfillment statuses of specific categories (especially Governance and Metrics & Targets) with high accuracy, indicating that it can efficiently evaluate the fulfillment statuses of TCFD recommended disclosures. This approach can provide valuable information for efficient and accurate analyses of sustainability information disclosure, aiding companies in disclosing their climate-related information and investors in their assessment and decision-making processes.

Possible future research avenues are continuous surveys of fulfillment status, improvements in classification methods (including their criteria), and expansion to disclosure media other than annual securities reports. In addition, the proposed method can automate analyses of sustainability information disclosures, providing a foundation for further applications. For instance, the method is applicable to disclosures related to other ESG elements and to international disclosure standards.

References

- [1] TCFD (Task Force Clim. Relat. Financ. Discl.). Recommendations of the task force on climate-related financial disclosures. 2017.
- [2] Julia Anna Bingler, Mathias Kraus, Markus Leippold, and Nicolas Webersinke. Cheap talk and cherry-picking: What climatebert has to say on corporate climate risk disclosures. *Finance Research Letters*, 47:102776, 2022. ISSN 1544-6123. doi: <https://doi.org/10.1016/j.frl.2022.102776>. URL <https://www.sciencedirect.com/science/article/pii/S1544612322000897>.
- [3] Dong Ding, Bin Liu, and Millicent Chang. Carbon emissions and tcfid aligned climate-related information disclosures. *Journal of Business Ethics*, 182(4):967–1001, 2023.
- [4] Joe Davison. Zero-shot learning in modern nlp, 2020. <https://joeddav.github.io/blog/2020/05/29/ZSL.html> (accessed on 2024-03-01) .
- [5] Alix Auzepy, Elena Tönjes, David Lenz, and Christoph Funk. Evaluating tcfid reporting: A new application of zero-shot analysis to climate-related financial disclosures, 2023.
- [6] Nobushige Doi, Yusuke Oda, Naho Nakakubo, and Jun Sugimoto. Automated Determination of TCFD recommended disclosures through zero-shot Text Classification (in Japanese). pages 1279–1284, 3 2024.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [8] Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, and Markus Leippold. Climatebert: A pre-trained language model for climate-related text, 2022.
- [9] Task Force on Climate-related Financial Disclosures. Task force on climate-related financial disclosures 2023 status report, 2023. <https://www.fsb.org/wp-content/uploads/P121023-2.pdf> (accessed on 2024-03-01) .
- [10] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.
- [11] David Friederich, Lynn H. Kaack, Alexandra Luccioni, and Bjarne Steffen. Automated identification of climate risk disclosures in annual corporate reports, 2021.
- [12] Angel-Ivan Moreno and Teresa Caminero. Application of text mining to the analysis of climate-related disclosures. *International Review of Financial Analysis*, 83:102307, 2022. ISSN 1057-5219. doi: <https://doi.org/10.1016/j.irfa.2022.102307>. URL <https://www.sciencedirect.com/science/article/pii/S1057521922002617>.
- [13] Taja Kuzman, Igor Mozetič, and Nikola Ljubešić. Chatgpt: Beginning of an end of manual linguistic data annotation? use case of automatic genre identification, 2023.
- [14] N. Doi, Y. Nobuta, and T. Mizuno. Topic classification of key audit matters in japanese audit reports by zero-shot text classification. In *2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 540–545, Los Alamitos, CA, USA, jul 2023. IEEE Computer Society. doi: 10.1109/IIAI-AAI59060.2023.00108. URL <https://doi.ieeeecomputersociety.org/10.1109/IIAI-AAI59060.2023.00108>.

- [15] OpenAI. GPT-4 Technical Report, 2023. <https://cdn.openai.com/papers/gpt-4.pdf> (accessed on 2024-03-01) .
- [16] Financial Services Agency. Briefing material for the first meeting of the conference of relevant ministries and agencies on sustainability disclosure (in japanese), 2022. <https://www.cas.go.jp/jp/seisaku/sustainability/dai1/siryous2.pdf> (accessed on 2024-03-01) .
- [17] TCFD Consortium. Implementing the Recommendations of the Task Force on Climate-related Financial Disclosures (Japanese), 2021. https://tcfd-consortium.jp/pdf/about/2021_TCFD_Implementing_Guidance_2110_jp.pdf (accessed on 2024-03-01) .
- [18] Inc. Tokyo Stock Exchange. Japan's corporate governance code, 2021. <https://www.jpx.co.jp/english/equities/listing/cg/tvdivq0000008jdy-att/20210611.pdf> (accessed on 2024-03-01) .
- [19] Financial Services Agency. Revision of sustainability information, including establishment of a new section for sustainability information (explanatory material) (in japanese), 2023. <https://www.fsa.go.jp/policy/kaiji/sustainability01.pdf> (accessed on 2024-03-01) .
- [20] Financial Services Agency. Report of the working group on disclosure, financial system council (in japanese), 2022. https://www.fsa.go.jp/singi/singi_kinyu/tosin/20220613/01.pdf (accessed on 2024-03-01) .
- [21] Inc. Japan Exchange Group. Annual Securities Report FY2022 (in Japanese), 2023. <https://www.jpx.co.jp/corporate/investor-relations/ir-library/securities-reports/tvdivq0000008r0t-att/view22-0.pdf> (accessed on 2024-03-01) .
- [22] Inc. Japan Exchange Group. Annual Securities Report FY2022, 2023. https://www.jpx.co.jp/english/corporate/investor-relations/ir-library/securities-reports/tcgh5100000049ke-att/Annual_Securities_Report_fy2022.pdf (accessed on 2024-03-01) .
- [23] TCFD Consortium. TCFD Guidance 3.0 (in Japanese), . https://tcfd-consortium.jp/pdf/news/22100501/TCFD_Guidance_3.0_J.pdf.
- [24] TCFD Consortium. TCFD Guidance 3.0, . https://tcfd-consortium.jp/pdf/en/news/22100501/TCFD_Guidance_3.0_e.pdf.
- [25] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6): 391–407, 1990.
- [26] Inc. Japan Exchange Group. Survey of TCFD Disclosure in Japan , 2021. <https://www.jpx.co.jp/english/corporate/news/news-releases/0090/b5b4pj000004jllw-att/TCFDsurveyEN.pdf> (accessed on 2024-03-01) .
- [27] Inc. Japan Exchange Group. Survey of TCFD Disclosure in Japan (FY2022) , 2023. <https://www.jpx.co.jp/english/corporate/news/news-releases/0090/tcgh510000005j2x-att/TCFDsurveyen.pdf> (accessed on 2024-03-01) .

A By-category results of the classification evaluation experiments

Table 7: Results of the classification evaluation experiment

ID	Num. of fulfilled	GPT-3.5				GPT-4			
		Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
01-01	69	71.0%	70.4%	100.0%	82.6%	92.0%	89.6%	100.0%	94.5%
01-02	38	40.0%	38.8%	100.0%	55.9%	94.0%	92.1%	92.1%	92.1%
01-03	70	72.0%	71.9%	98.6%	83.1%	94.0%	92.1%	100.0%	95.9%
01-04	47	53.0%	50.0%	100.0%	66.7%	83.0%	75.9%	93.6%	83.8%
02-01	66	68.0%	67.7%	98.5%	80.2%	84.0%	82.9%	95.5%	88.7%
02-02	71	71.0%	71.0%	100.0%	83.0%	89.0%	86.6%	100.0%	92.8%
02-03	66	68.0%	67.7%	98.5%	80.2%	91.0%	89.0%	98.5%	93.5%
03-01	14	44.0%	19.1%	92.9%	31.7%	95.0%	73.7%	100.0%	84.8%
03-02	5	55.0%	8.3%	80.0%	15.1%	96.0%	55.6%	100.0%	71.4%
03-03	9	38.0%	12.7%	100.0%	22.5%	92.0%	52.9%	100.0%	69.2%
04-01	13	31.0%	15.0%	92.3%	25.8%	89.0%	54.2%	100.0%	70.3%
04-02	5	55.0%	8.3%	80.0%	15.1%	98.0%	80.0%	80.0%	80.0%
05-01	11	74.0%	29.7%	100.0%	45.8%	100.0%	100.0%	100.0%	100.0%
05-02	7	83.0%	29.2%	100.0%	45.2%	96.0%	63.6%	100.0%	77.8%
05-03	8	68.0%	20.0%	100.0%	33.3%	96.0%	66.7%	100.0%	80.0%
05-04	7	62.0%	14.0%	85.7%	24.0%	96.0%	66.7%	85.7%	75.0%
05-05	8	63.0%	17.8%	100.0%	30.2%	95.0%	66.7%	75.0%	70.6%
06-01	50	51.0%	50.5%	100.0%	67.1%	87.0%	80.3%	98.0%	88.3%
07-01	62	62.0%	62.0%	100.0%	76.5%	86.0%	82.4%	98.4%	89.7%
08-01	18	24.0%	18.5%	94.4%	30.9%	88.0%	61.5%	88.9%	72.7%
08-02	21	37.0%	25.0%	100.0%	40.0%	88.0%	66.7%	85.7%	75.0%
09-01	25	66.0%	42.1%	96.0%	58.5%	94.0%	85.2%	92.0%	88.5%
09-02	6	48.0%	10.3%	100.0%	18.8%	98.0%	75.0%	100.0%	85.7%
10-01	20	79.0%	48.8%	100.0%	65.6%	98.0%	95.0%	95.0%	95.0%
10-02	6	72.0%	15.6%	83.3%	26.3%	98.0%	83.3%	83.3%	83.3%
11-01	24	79.0%	53.7%	91.7%	67.7%	93.0%	84.0%	87.5%	85.7%
11-02	23	68.0%	42.6%	95.8%	59.0%	95.0%	82.8%	100.0%	90.6%
全体		59.3%	36.3%	95.8%	52.7%	92.8%	77.2%	94.4%	84.9%

B Relationship between TCFD recommended disclosures and their criteria

Table 8: Relationship between TCFD recommended disclosures and their criteria

#	Core elements	Title of disclosures	Criteria
01	Governance	Board oversight	01-01: Process of reporting to the board on climate-related issues 01-02: Frequency of reporting to the board on climate-related issues 01-03: Consideration of climate-related issues by the board 01-04: How the board monitors and oversees progress
02	Governance	Management's role	02-01: Organizational structure 02-02: Process by which management is informed about climate-related issues 02-03: How management monitors climate-related issues
03	Strategy	Risks and opportunities	03-01: Time horizon(s) for the consideration of climate-related risks and opportunities 03-02: Climate-related issues in each specified time horizon 03-03: Process for determining risks and opportunities with a financial impact
04	Strategy	Impact on businesses, strategy, and financial planning	04-01: Impact on businesses or strategy 04-02: Impact on financial planning
05	Strategy	Description of resilience of strategy based on scenarios	05-01: 2 ° C or lower scenario 05-02: Climate-related scenarios and associated time horizon(s) 05-03: Impact of climate-related scenarios on the strategy 05-04: Strategic action taken in relation to climate-related scenarios 05-05: Impact of climate-related scenarios on financial planning
06	Risk Management	Processes for identifying and assessing risks	06-01: Process for identifying and assessing risks
07	Risk Management	Processes for managing risks	07-01: Process for managing risks
08	Risk Management	Integration into overall risk management	08-01: Integration of the identification and assessment process into the overall risk management 08-02: Integration of the management process into overall risk management
09	Metrics & Targets	Metrics used to assess risks and opportunities	09-01: Evaluation metrics 09-02: Historical performance based on evaluation metrics
10	Metrics & Targets	Greenhouse gas emissions of Scope 1 and 2	10-01: Greenhouse gas emissions 10-02: Historical greenhouse gas emissions
11	Metrics & Targets	Targets used to manage risks and opportunities and performance against targets	11-01: Climate-related targets 11-02: Time horizon(s) for climate-related targets

C Example of a prompt input into the LLMs

Example of a prompt (Reference Translation)

Below are some disclosure materials of listed companies and the criteria for TCFD recommended disclosures.

Selected Disclosure Materials of Listed Companies

...

Criteria

Does the text describe the processes by which the board and/or the board committee are informed about climate-related issues?

Remarks

Examples of board committees could include an audit committee, risk committee, etc.

Climate-related issues refer to climate-related risks and opportunities.

Sustainability efforts are considered to be related to climate-related issues.

Content related to future plans will not be considered.

Task

Output True if some input listed company disclosure documents meet the Criteria, False otherwise.

Output