

SentiPromiseESG: Sentiment Analysis of Sustainability Promises Across Industries

Wei-Chen Huang¹
wesleyseawatch@gmail.com

Hsin-Ting Lu²
hsintinglubob@gmail.com

Wen-Ze Chen²
50712andy@gmail.com

Yu-Han Huang³
5monica.cupcake@gmail.com

Min-Yuh Day^{2*}
myday@gm.ntpu.edu.tw

Department of Statistics¹, Graduate Institute of Information Management²
, Department of Accounting³
National Taipei University, New Taipei City, Taiwan

Abstract

Sustainability reporting has become a primary channel for firms to communicate their environmental, social, and governance (ESG) performance, yet the voluntary and narrative nature of these disclosures raises persistent concerns about selective reporting and greenwashing. Despite the growing use of large language models (LLMs) and ESG-focused language models to process sustainability texts, there is still limited empirical evidence on how the tone of ESG disclosures relates to the presence, content, and verifiability of sustainability promises, particularly in non-English, industry-specific settings. This study aims to examine how sentiment is associated with sustainability promise behavior by developing a sentiment-aware, LLM-based framework for identifying and evaluating sustainability promises across industries. Using ESG reports from 29 large Taiwanese listed firms in the semiconductor, financial, and computer-peripherals sectors, we construct the SentiPromiseESG Dataset with 15,345 sentence-level spans annotated for promise status, evidence status, evidence quality, verification timeline, and ESG type via a two-stage Gemini 2.5 Pro pipeline, and derive weighted sentiment scores from Gemini 2.5 Pro and FinBERT-ESG that are incorporated into span-level logistic regression models. The empirical results show that more positive sentiment is consistently associated with a higher likelihood of making sustainability promises and with a stronger tendency for those commitments to fall within the social dimension rather than environmental or governance domains, whereas sentiment exhibits only weak and unstable links to whether supporting evidence is provided and how clearly that evidence is presented. These findings contribute a scalable LLM-based annotation and sentiment-analysis pipeline, introduce the SentiPromiseESG Dataset as a new resource for ESG and greenwashing research, and offer practical implications for investors, regulators, and practitioners seeking to use textual analytics and LLM-based tools to assess the credibility and narrative structure of corporate sustainability communication.

Keywords:

Sustainability Promises, Sentiment Analysis, ESG, Greenwashing, Large Language Model