

# Inclusion in CSR Reports: The Lens from a Data-driven Machine Learning Model

**Lu Lu, Jinghang Gu, Chu-Ren Huang**

Hong Kong Polytechnic University  
lulubfsu@gmail.com, {jinghang.gu, churen.huang}@polyu.edu.hk

## Abstract

Inclusion, as one of the foundations in the diversity, equity, and inclusion initiative, concerns the degree of being treated as an ingroup member in a workplace. Despite of its importance in a corporate’s ecosystem, the inclusion strategies and its performance are not adequately addressed in corporate social responsibility (CSR) and CSR reporting. This study proposes a machine learning and big data-based model to examine inclusion through the use of stereotype content in actual language use. The distribution of the stereotype content in general corpora of a given society is utilized as a baseline, with which texts from a corporate under discussion are compared. This study not only propose a model to identify and classify inclusion in language use, but also provides insights to measure and track progress by including inclusion in CSR reports and to build an inclusive corporate team.

**Keywords:** corporate social responsibility (CSR) reports, inclusion, word embeddings, baseline, big data

## 1. Introduction

Diversity, equity and inclusion (DE&I) are essential cornerstones of Corporate Social Responsibility (CSR) (Fenwick and Bierema, 2008; Grosser and Moon, 2005). The CSR report, as one of the most important outlets to communicate a corporate’s social engagement and impact, provides insights to a corporate’s social norms. However, DE&I, especially the inclusion, lacks adequate attention in CSR reports (Hunt et al., 2020, among many others). As inclusion focuses on treating a person as ‘we’ (ingroup) rather than ‘them’ (outgroup), and is mediated by the use of stereotyping languages, we are interested to examine inclusion through Warmth and Competence, the two universal dimensions in stereotype content (Fiske et al., 2002). This niche in CSR reporting motivates us to employ machine learning techniques and big data samples to study stereotype content in actual language use. Word embedding, as a powerful technique in studying the semantic association between words, are employed to identify the stereotype content and analyze its distribution in a general corpus and use the results as a baseline. Furthermore, the collection of internal and external documents and texts in a corporate can be utilized to measure and compare with the baseline, hence developing a better understanding of the progress of a corporate’s efforts in promoting inclusion. This study aims to establish and demonstrate a model that can be applied to CSRs in the future in order to unveil the corporate’s stand on inclusion and their DE&I efforts at large. Measuring a corporate’s use of stereotype content against the society’s norm of stereotypes in general corpora, this study gives CSR reports a clear baseline to do better than.

The rest of the paper is organized as follows. Section 2 reviews the previous studies regarding inclusion,

stereotypes, CSR reports and machine learning approaches. In Section 3, a word embedding based model to study stereotype content in general corpora (i.e. baseline) and the corporate corpus is proposed and referred to as the Word Embedding Inclusion Model (WEIM). We then, in Section 4, report a research plan to examine the distribution of Warmth and Competence of different temporal points in the US society from American English corpora and use the findings as a baseline, which can be compared with corporate datasets. The paper is concluded in Section 5.

## 2. Literature review

### 2.1 Diversity, Equity and Inclusion in Businesses

Social responsibility, also known as corporate social responsibility, refers to the ways that a corporate positions itself to make positive influence on the community and society at large (Fenwick and Bierema, 2008). CSR initiatives require a corporate not only to focus on making money and profitable gains, but also take a longer and strategic view of their ‘impact economically, socially, environmentally, and in terms of human rights’, on a wide range of stakeholders (CIPD, 2003). CSR is an important approach to align business strategies with the value and well-being of the society, thus strengthening the connection between employers and employees. CSR initiatives have been incorporated into the branding of the corporate (Hon and Gamor, 2021).

While CSR relating to personal well-being starts from inside the corporate, it is communicated externally through outlets, such as corporate websites, blogs, and public reports showing their businesses’ cultures and priorities. One of the most important public-facing reporting outlets is the CSR reports, providing insights

into an organization's workplace norms, hiring practices, and overarching aspects of organizational culture (De Stefano et al., 2018).

DE&I initiatives, aiming to create a welcoming environment for less privileged identities, are an essential aspect in employers' CSR strategies, due to social pressure, increased diversity in clients, and public policies (Moore et al., 2017). Grosser and Moon (2005), for example, report the criteria and benefits of including gender equality into CSR reports. The 2020 analyst report by McKinsey & Company shows that the gap in ethnic diversity is larger than gender diversity between the top-quartile and the fourth quartile corporates, and this trend is likely to continue (Hunt et al., 2020). The lack of ethnic minority diversity is even more evident in the UK and US. The representation of ethnic minorities in the executive team in the UK and US is 13 percent in 2019, which only increased from 7 percent in 2014, whereas the global dataset shows that 14 percent of ethnic minorities are represented in the executive team, increased from 12 percent in 2017. Even though DE&I is the crucial aspect of marketing and talent acquisition, inclusion, which is the 'degree to which an employee is accepted and treated as an insider by others in a work system' (Pelled et al., 1999), is not yet prioritized in corporates' CSR reporting and strategies. In the tourism workforce, for example, Hon and Gamor (2021) have advocated for the inclusion of minority groups as CSR strategies and corporate images. Furthermore, in the advancement of DE&I culture in industrial settings, employees' negative sentiments towards inclusion in their workplace experience is markedly worse than the ones towards diversity in McKinsey's diversity report in 2020 (Hunt et al., 2020). This challenge is still visible for relatively diverse corporates. Among many aspects of inclusion, freedom from bias and discrimination is one of the important factors.

As businesses face increased demands for inclusion, it is worth continued research to help corporates identify their advantages and problems compared with the norm in the society, and thus, based on that, corporates can further enhance inclusive practices, organizational cultures, and policies.

## 2.2 Inclusion, Stereotypes, and Language Use

Inclusion can be affected by negative attitudes and stereotypes (Sanders and Sullivan, 2010; Krischler et al., 2018). Being both positive and negative, stereotypes in both polarities can be found in a given social group. In Stereotype Content Model (SCM), Warmth and Competence are two universal dimensions to evaluate stereotype content (Fiske et al., 2002; Fiske, 2018). Warmth (trustworthiness, sociability) can be depicted as 'good-hearted' and 'benevolent', and features such as 'competent', 'intelligent' are used to describe Competence (capable, agentic). The degree that Warmth and Competence are

ascribed to high and low levels reflects how 'we' believe and evaluate 'others'. The arrays of Warmth and Competence are further classified based on their high and low levels: High Warmth (HW), Low Warmth (LW), High Competence (HC), and Low Competence (LC). The combined interpretation of the stereotype content and its levels define different social groups: for instance, elderly people are regarded as HW-LC, Whites are HW-HC, the rich are commonly believed as LW-HC, and Blacks as LH-LC (Fiske et al., 2002, Durante et al., 2017a, 2017b). According to the SCM, a society's default group or ingroup is believed to be 'us' that are high on both Warmth and Competence, whereas the group of 'them', depicting a stereotype of exclusion, is low on both dimensions. The rest of the combinations are ambivalent, meaning that they are high on one dimension only, such as being high in Warmth but low in Competence (Durante et al., 2017b).

There have been several attempts to apply the SCM to investigate the actual use of languages. In Dupree and Fiske's (2019) study, they apply the SCM to analyze the past campaign speeches of the White Republican and Democratic presidential candidates and compared their speeches with different target audiences. The findings exhibit that, when addressing audiences who are mostly minority groups, Democrats use more Warmth than Competence. As the use of stereotype is reflective of social inclusion, analyzing the distribution along the Warmth and Competence dimensions in actual language use can reveal inclusion towards a group in a natural way.

## 2.3 Word Embedding in Stereotypes

The study of stereotypes has been broadly explored with human subject research (Katz and Braly, 1933; Fiske et al., 2002) and text-based analysis (Henley, 1989). Recent development in machine learning offers great promise and valuable insights to understand stereotypes. Word embeddings are an unsupervised neural network-based technique to capture semantic associations of words with relationships between vectors. Word2vec (Mikolov et al., 2013a, 2013b), as one of the most popular techniques in word embeddings, takes a large amount of textual data as input and represents a word as a list of low-dimension vectors. The cosine similarity function between the vectors indicates the degree of semantic similarities between the words. For example, a higher cosine similarity score can be found between words 'man' and 'woman' than the pair of 'man' and 'pen'. The vector representation can be obtained by the models of Skip-gram and Common Bag of Words. This project will choose the Skip-gram model as we are more interested in predicting a word within a certain range before and after the target word in the same sentence (a.k.a. window size).

There have been some studies using word embedding techniques to study stereotype languages. Garg et al.

(2018) have proved that word embeddings are robust in extracting and analyzing ethnic and gender stereotypes over 100 years. Since Garg et al.'s (2018) longitudinal survey, word embeddings have been widely applied as a method of extracting features out of texts and using those features as an input to machine learning model to shed light on stereotype expressions and the attitudes towards them (Charlesworth et al., 2021; Kroon et al., 2021). However, there has been less work on applying machine learning techniques to examine the SCM and analyze how the properties of stereotype content are manifested in actual languages.

Given the fact that little research exists about how inclusion is addressed in CSR reports with machine learning tools, this study attempts to address this niche by using word embedding techniques to analyze stereotype content. Specifically, the distribution of stereotype content in a general corpus will be employed as a baseline, with which the one in corporate corpora, consisting of published resources of a corporate, will be compared and reported in CSR reports. In the rest of the study, we will detail the model based on word embedding techniques on stereotype content with a preliminary case study.

### 3. Methodology

In this section, the Word Embedding-based Inclusion Model (WEIM) in CSR reports is proposed to address the niche on inclusion as reflected in the language use of stereotype content. Figure 1 illustrates the architecture of the proposed method, which is a machine learning framework that classifies stereotypes into Warmth and Competence, and identifies the keywords associated with the two categories based on deep semantic representation.

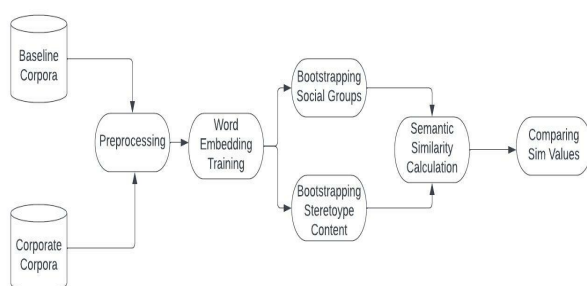


Figure 1: The pipeline of the Word Embedding-based Inclusion Model (WEIM).

In this model, the baseline corpus and the corporate corpus, respectively, go through the pipeline in the WEIM. The unstructured raw texts in each corpus, after preprocessing, was trained on word embeddings, which converts each word into a vector for calculation. In the following bootstrapping module, keywords about particular social groups (e.g. ethnic groups, gender groups) and stereotype content of Warmth and

Competence were automatically extracted from each corpus. For example, in a case study on ethnic groups in the USA, 'Asians', 'Blacks', 'Hispanics', and 'Whites' can be used as seed words to extract the 150 most similar words to those seed words for each ethnic group. In a similar way, seed words related to the positive and negative Warmth and Competence were chosen to automatically extract the 150 most similar words based on the Euclidian distance. After manual checking, the remaining words are the valid keywords used in the rest of the analysis. In the next module on the semantic similarity calculation, the keywords in each social group and in each positive/negative category of the two stereotype content dimensions were iterated and paired to calculate the average cosine similarity score of that particular social group in terms of its stereotype content. Specifically, for each corpus, we calculate the cosine similarity score of the keywords pairing between each social group and each component in the stereotype content. We have until now collected the average sim values of positive and negative Competence and positive/negative Warmth that pair with a given social group, in terms of the corporate corpus data and the baseline corpus data. Finally, we can compare the two sets of data and examine the performance of the language of inclusion in corporate data vis-à-vis the baseline general corpus data.

### 4. A Proposal for a Study on Ethnic Inclusion

In this section, we will use a set of American English corpora in general to illustrate how we can incorporate the data generated from this baseline corpus to better understand the use of stereotype in corporates. The baseline corpora for this case study are the Brown Family Corpora, consisting of a) the original Brown corpus (Francis and Kucera, 1979), b) the Freiburg update of the Brown corpus (Frown; Hundt et al., 1999), and c) the recent update of the Brown corpus around the year 2009 (Crown; Xu and Liang, 2009). These three corpora of American English follow the same sampling pattern in the Brown corpus. Each of the three corpus contains 500 documents with approximately 2,000 words on average, consisting of textual collections published in the years 1961, 1991, and 2009 ( $\pm 1$  year), respectively. The three corpora cover four broad text types: press, general prose, learned writing, and fiction, which is meant to present American language use in general. In total, the baseline corpora have approximately three million words and contain three temporal points in the 1960s, 1990s, to approximately the 2010s. Additionally, the corporate corpus can be composed of internal and external documents and texts published by a given corporate, such as news reports about this corporate, past public-facing reports (e.g., CSR reports), corporate websites, blogs, and transcripts of recorded meetings in internal and external channels (with prior ethical approval). Both baseline corpora and the

corporate corpora will undergo the same pipelines as detailed in the rest of the section, including but not limited to following the same preprocessing methods and using the same seed words to extract social group and stereotype content wordlists. In what follows, I will propose a preliminary study to identify and extract information in the baseline corpus, and build on the data reported in Lu et al. (to be submitted) to examine the baseline vis-à-vis the corporate corpus.

In this preliminary study, following the WEIM model, we will use Python to preprocess raw data from corpora, such as turning all letters to lower cases, removing non-alphanumeric characters, before training word embeddings. Gensim's word2vec skip-gram model (Mikolov et al., 2013a) will be used and each word will be returned with 300 dimensions in our training corpus. Each corpus will be trained individually to examine the over-time variation of the three temporal points. Finally, some simple analogy tests (e.g. man is to king, as woman is to \_\_\_) will be performed to warrant the quality of the embedding models.

In the next module of bootstrapping, we consider stereotypes in the four ethnic groups in the US, namely Whites, Blacks, Hispanics, and Asians (e.g., Durante et al., 2017a). After word embedding training, seed words of 'Whites', 'Blacks', 'Hispanics', and 'Asians' can be used to bootstrap ethnic groups in the corpus. According to their cosine similarity scores, the first 150 most similar words to those seed words can be automatically extracted as the wordlist for ethnic groups. As for high/low levels of Warmth and Competence, seed words of 'warm' and 'warmth' (+W), 'unkind' and 'unfriendly' (-W), 'competence' and 'competent' (+W), and 'incompetence' and 'incompetent' (-C) will be used as seed words to extract, for example, the top 100 words that are similar to those seed words. Manual checking will be performed to 1) keep positive words in the positive groups and negative words in the negative groups; 2) remove irrelevant words generated from the wordlists. The wordlists of stereotype content will be paired with ethnic groups (e.g. (Japanese, kind) and (Chinese, friendly)) to compute the cosine similarity scores.

For this preliminary proposal, a sub-corpus with corporate texts from the baseline will be extracted and used as the corporate corpus. Based on the metadata of Brown, Frown, and Crown corpora, texts, such as news reports, about corporates or industries will be extracted to build the corporate corpus. This corpus of raw texts will follow the same pipelines as the baseline corpus to get word embedding score of each word and the similarity scores of the pair of the ethnic group and stereotype content.

Lu et al. (manuscript) used the same three baseline corpora as this current study, and also followed the similar approach as detailed above. In their research, data show that Asians are always ascribed to LW and

HC and Blacks to HW and LC, the stereotype of which are supported by many social psychological studies (e.g., Swencionis et al., 2017; Froehlich and Schulte 2019). Even though questionnaire-based studies (e.g. Fiske et al. 2002) show that high value of Competence and Warmth shows inclusion, Lu et al. (manuscript) argue that HW and HC may not necessarily be the indicators of inclusion in actual language use. In their data, they found that Asians do not use HW to represent their inclusion. Instead, the consistent pattern of LW and HC in Asians vis-à-vis the equally consistent pattern of HW and LC in Blacks, and the tendency that Asians are inclined to be grouped together with Hispanics and Whites imply that the (dis)association with a particular ethnic group is a special way to represent inclusion in the actual language use. The other finding that is worth our attention is that, while Blacks are usually assigned with high warmth category, this is not true in the Brown corpus, where Whites are assigned with high warmth. On the other hand, this unusual pattern may align with the white supremacy view in the 1960s when the Brown corpus were constructed.

Building upon their findings regarding the distribution of Warmth and Competence in the baseline corpus, we can compare the results generated from the corporate corpus. The practical application for this comparison in a CSR report can be captured threefold. Firstly, the baseline corpus presents the norm of stereotype use in a given society (USA in this case study) at different temporal points. For example, it is likely to see a surge of content in describing Whites are warmth in the 1960s' corporate data in the USA, whereas Blacks are increasingly perceived as being warm since 1990s. Secondly, the WEIM applies the same seed words to bootstrap and the same methodology to calculate, and thus compare the stereotype language use in corporate data versus the general social trend. For example, the high competence score of Asians does not necessarily imply that a corporate is inclusive in this aspect, because the high competence scores in Asians can be the baseline of a society in general. Thirdly, the WEIM encourages a balanced view of high and low stereotype content towards any given group of people, thus promoting inclusion in workplace and society.

## 5. Conclusion

This paper proposed the model of WEIM, a word embedding based approach to use general corpora as a baseline to better understand the stereotype content in corporate language dataset, thus promoting the inclusion in CSR reports, which rarely use machine learning techniques and big data samples. We then propose a preliminary case study to figure out the inclusion of ethnic minorities in the actual language use of American English in general corpus vis-à-vis the sub-corpus of corporate texts in three different temporal points. The results from general corpus data will be treated as a baseline to help corporates further

measure and compare the distribution of stereotype content in corporate datasets, and, eventually, promote the incorporation of inclusion in CSR reports.

## References

- Charlesworth T. E. S., Yang V., Mann T. C., Kurdi B., & Banaji M. R. (2021). Gender Stereotypes in Natural Language: Word Embeddings Show Robust Consistency Across Child and Adult Language Corpora of More Than 65 Million Words. *Psychological Science*, 32(2):218–240.
- CIPD (2003), *Corporate Social Responsibility and HR's Role*. Chartered Institute of Personnel and Development, London.
- De Stefano, F., Bagdadli, S., & Camuffo, A. (2018). The HR role in corporate social responsibility and sustainability: A boundary-shifting literature review. *Human Resource Management*, 57(2), 549–566.
- Dupree, C. H., & Fiske, S. T. (2019). Self-presentation in interracial settings: The competence downshift by White liberals. *Journal of Personality and Social Psychology*, 117(3), 579–604.
- Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., ... & Teymoori, A. (2017a). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences*, 114(4), 669–674.
- Durante, F., Tablante, C. B., & Fiske, S. T. (2017b). Poor but warm, rich but cold (and competent): Social classes in the stereotype content model. *Journal of Social Issues*, 73(1), 138–157.
- Finwick, T., & Bierema, L. (2008). Corporate social responsibility: issues for human resource development professionals. *International Journal of Training and Development*, 12(1), 24–35.
- Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current directions in psychological science*, 27(2), 67–73.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, 82(6), 878–902.
- Francis, W. N., & Kucera, H. (1979). Brown corpus manual. *Letters to the Editor*, 5(2), 7.
- Froehlich, L., & Schulte, I. (2019). Warmth and competence stereotypes about immigrant groups in Germany. *PLoS ONE*, 14(9), e0223103.
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644.
- Grosser, K., & Moon, J. (2005). Gender mainstreaming and corporate social responsibility: Reporting workplace issues. *Journal of business ethics*, 62(4), 327–340.
- Henley N. M. (1989). Molehill or mountain? What we know and don't know about sex bias in language. In Crawford M. and Gentry M. (eds.) *Gender and Thought: Psychological Perspectives*. New York: Springer. pp 59–78.
- Hon, A. H., & Gamor, E. (2021). The inclusion of minority groups in tourism workforce: Proposition of an impression management framework through the lens of corporate social responsibility. *International Journal of Tourism Research*.
- Hope Pelled, L., Ledford, Jr, G. E., & Albers Mohrman, S. (1999). Demographic dissimilarity and workplace inclusion. *Journal of Management studies*, 36(7), 1013–1031.
- Hundt, M., A. Sand, and P. Skandera. (1999). *Manual of Information to accompany The Freiburg – Brown Corpus of American English ('Frown')*. Freiburg: Department of English. Albert-Ludwigs-Universität Freiburg.  
<http://khnt.aksis.uib.no/icame/manuals/frown/INDEX.HTM>.
- Hunt, V., Prince, S., Dixon-Fyle, S., & Dolan, K. (2020). *Diversity wins: How inclusion matters*. McKinsey & Company.
- Katz D., Braly K. (1933). Racial stereotypes of one hundred college students. *J Abnorm Soc Psychol* 28:280–290.
- Krischler, M., Pit-ten Cate, I. M., & Krolak-Schwerdt, S. (2018). Mixed stereotype content and attitudes toward students with special educational needs and their inclusion in regular schools in Luxembourg. *Research in Developmental Disabilities*, 75, 59–67.
- Kroon, A. C., Trilling, D., & Raats, T. (2021). Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage. *Journalism & Mass Communication Quarterly*, 98(2), 451–477.
- Lu, L., Gu, J., & Huang, C.-R. (manuscript). Warmth and Competence of ethnic stereotypes in American English: A data-driven machine learning approach.
- Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013a). Efficient estimation of word representations in vector space. In ICLR., *Proceeding of the International Conference on Learning Representations Workshop Track* (pp. 1301–3781).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

- Moore, K., McDonald, P., & Bartlett, J. (2017). The social legitimacy of disability inclusive human resource practices: the case of a large retail organisation. *Human Resource Management Journal*, 27(4), 514-529.
- Sanders, M. S., & Sullivan, J. M. (2010). Category inclusion and exclusion in perceptions of African Americans: Using the stereotype content model to examine perceptions of groups and individuals. *Race, Gender & Class*, 201-222.
- Schnabel, T., Labutov, I., Mimno, D., & Joachims, T. (2015, September). Evaluation methods for unsupervised word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 298–307).
- Swencionis, J. K., & Fiske, S. T. (2016). Promote up, ingratiate down: Status comparisons drive warmth-competence tradeoffs in impression management. *Journal of Experimental Social Psychology*, 64, 27–34.
- Xu, Jiajin & Maocheng Liang. (2013). A tale of two C's: Comparing English varieties with Crown and CLOB. *ICAME Journal*37: 175-183.