Supplementary Materials

for

Corporate Sustainability Communication as 'Fake News': Firms' Greenwashing on Twitter

Divinus Oppong-Tawiah *(corresponding author)*
Schulich School of Business
York University. Toronto, Canada
divinus@schulich.yorku.ca

Jane Webster
Smith School of Business
Queen's University, Kingston, Canada

Contents

Sections S1-S4: Supplementary Materials for Research Question 1

S1. Greenwashing as Fake News

Table S1. Greenwashing as Different Types of Fake News

Fake News Concept	**Examples**	**Authenticity** *(Ease of Verification)*	**Intention**	**Information Source**	**Sample Papers**
Disinformation	Deceptive news	Non-factual *(Easy)*	Mislead	News / Message	Allcott and Gentzkow (2017)
	False News	Non-factual *(Easy)*	Mislead/ Mistake	News	Vosoughi et al. (2018)
	Hoax	Non-factual *(Easy)*	Mislead	Message / News	Wardle and Derekshan (2017)
	Fake reviews	Non-factual *(Moderate)*	Mislead	Message	Wu et al. (2020)
	Greenwashing	Non-factual *(Difficult)*	Mislead	Message	Marquis et al. (2016)
Misinformation	Conspiracy	Non-factual / factual *(Moderate)*	Mislead / Mistake	Message	Zannettou et al. (2019)
	Satire News	Non-factual *(Easy)*	Entertain	News	Tandoc et al. (2018)
	Pseudoscience	Non-factual *(Easy)*	Mislead / Mistake	Message	Rubin et al (2015)
	False connection	Non-factual *(Easy)*	Mislead / Mistake	News	Ireton and Posetti (2018)
	Greenwashing	Non-factual *(Moderate)*	Mistake	Message	Szabo & Webster (2021)
Malinformation	Cherry-picking	Factual *(Moderate)*	Mislead	Message / News	Asudeh et al. (2020)
	Clickbait	Factual *(Moderate)*	Mislead	News	Chen et al. (2015)
	Rumor	Non-factual / factual *(Difficult)*	Mislead / Mistake	Message	Zubiaga et al. (2018)
	Trolling	Non-factual / factual *(Difficult)*	Harass	Message	Wardle 2018 (2018)
	Greenwashing	Non-factual / factual *(Difficult)*	Mislead	Message	Bowen and Aragon-Correa (2014)

S2. Comparisons with Current Fake News Detection Methods

Current automatic fake news detection methods are broadly categorized as knowledge-, propagation-, source-, and linguistic style-based approaches (Zhou & Zafarani, 2020) – see Table S2. Knowledge-based detection methods attempt to address the authenticity problem by using fact extraction and knowledge graphs from ground truths to automate fact-checking (e.g., Nickel et al., 2016). However, this approach is known for problems including redundant information (Altowim et al., 2014), incorrect timing of (Hoffart et al., 2013) and conflicting (Kang & Deng, 2019) facts, unreliable sources (Ye & Skiena, 2019), and data-driven Machine Learning (ML) models with little interpretability (i.e. little theoretical explanations for why new information may or may not be authentic) (Zhou & Zafarani, 2020).

To address the intention problem, current approaches analyse the message's propagation patterns, source credibility, or writing style. Propagation-based methods focus on how fake news spreads in a user network by analysing, for example, news cascades (i.e. tree-like structures that directly capture the propagation of news articles on a social network (Vosoughi et al., 2018)). Source-based methods examine whether news sources create, publish, or share content from reliable writers and publishers (Sitaula et al., 2020), pass spam detection algorithms (e.g., see Spirin & Han, 2012), appear on reputation ranking systems such as MediaRank (Ye & Skiena, 2019), or exhibit certain user account characteristics in posts, friends, and behaviors (e.g., Ferrara et al., 2016; Shao et al., 2018). Propagation and source-based methods highlight several insights: for example, compared to true news, fake news spreads further, faster, and more widely (Vosoughi et al., 2018) and is spread by more users, attains stronger user-engagement, and circulates in denser networks (echo chambers) (Zhou & Zafarani, 2019). However, as these methods are mostly data-driven, there is little theoretical foundation for classifying news cascades and sources as genuine or fake.

Style-based linguistic detection methods use linguistic cues with ML models to classify new information as fake or genuine (Feng et al., 2012; Zhou et al., 2020). The selection of features is often driven by linguistic theories and shares many theoretical foundations with the linguistic deception detection literature but rarely considers empirical evidence related to specific linguistic cues. Yet deceptive writing styles constantly evolve (Castelo et al., 2019) and data-driven features that represent the style of fake news in one context may not be compatible or interpretable in another context (Vilone & Longo, 2020).

For our greenwashing detection method, we developed a deviation-based linguistic style approach that extends the style-based approaches in three specific ways. First, it avoids the need for ground truth data. Second, it uses a multi-label rather than a binary classification, thereby introducing a new method suited to detecting 'non-traditional' fake news such as greenwashing and addressing a major challenge faced by existing detection methods. Third, it selects theoretically and empirically established linguistic cues, making our approach is more "explainable" and replicable in different contexts. Thus, it addresses limited research on model interpretability by using related theories and domain knowledge to guide greenwashing detection.

Table S2. Comparisons of Automatic Deception Detection Methods

Method	Description	Automation Tools	Authenticity: Ground Truth	Intention: Fake News Classification	Explainability
Knowledge-based	Using knowledge graphs to automate fact-checking	Machine Learning, Deep Learning	Ground truth required: Manual labeling	Binary (True/False)	Data driven
Source-based	Ranking the credibility of information (news) sources	Feature engineering, Machine Learning	Ground truth required: Manual labeling	Binary (Credible /Not credible)	Data driven
Propagation-based	Graphing the spread of false information (news)	Graph optimization, Deep Learning	Ground truth required: Manual labeling	Binary (True /False)	Data driven
Style-based	Mining deceptive intentions from language writing styles	Feature engineering, Machine Learning, Deep Learning	Ground truth required: Manual labeling or manipulation	Binary (True /False)	Theory driven
Deviation-based linguistic style *(our method)*	Mining deceptive intentions from language writing styles	Feature Specification, Profile Deviation	No ground truth required	Multi-class (Deception patterns)	Theory and data driven

S3. Robustness of the Multi-Category Greenwashing Patterns Measure

We introduced the notion of "greenwashing patterns" to classify tweets into deciles (ten quantile splits) of increasing Euclidean distances from the ideal non-greenwashing profile. One concern may be that classification into multiple categories may be inconsistent with the more common binary categorization (e.g., greenwashed or not) typical in prior research. Therefore, in Table S3, we compare correlational results for several ways to classify high and low greenwashing, including two binary splits (Mean Split, Quantile Split) and our preferred decile split (Quantile Range) using both normalized and scaled distance measures. Our Quantile Range measure correlates highly with the raw Euclidean distance (Pearson = 0.75), Mean Split (Pearson = 0.80) and reasonably well with Quantile Split (Pearson = 0.68). Thus, our greenwashing patterns measure compares reasonably with binary categorization with the advantage of avoiding strong claims about whether an organizational tweet is greenwashed. By measuring degrees of variation, we acknowledge the practical difficulty of inferring deception with absolute certainty in organizational greenwashing.

Table S3. Correlations of Different Measures of Greenwashing Patterns

Item	Greenwashing measure*	1	2	3	4	5	6	7	8
1	Euclidean Distance	1.00							
2	Mean Split	0.71	1.00						
3	Quantile Split	0.77	0.73	1.00					
4	Quantile Range (our measure)	0.75	0.80	0.68	1.00				
5	Scaled Distance	1.00	0.71	0.78	0.76	1.00			
6	Scaled Mean Split	0.70	0.99	0.73	0.80	0.71	1.00		
7	Scaled Quantile Split	0.77	0.73	0.98	0.68	0.78	0.73	1.00	
8	Scaled Quantile Range	0.75	0.80	0.68	1.00	0.76	0.80	0.68	1.00

***Key**	***Definitions***
Euclidean Distance	*Greenwashing as raw Euclidean distances between normalized (z) scores*
Scaled Distance	*Greenwashing as raw Euclidean distances between unitary scaled [0,1] scores*
Mean Split	*Greenwashing as mean split of Euclidean distances*
Quantile Split	*Greenwashing as 20th quantile split of Euclidean distances*
Quantile Range	*Greenwashing as decile quantile range (10th, 20th, …90th) splits of Euclidean distances*

S4. Validation Approaches for Detecting Greenwashing

As described in the paper, our validation tests address the third issue in profile deviation, i.e., the predictive power of our method for detecting greenwashing patterns. The classical approach in profile deviation is to specify a baseline model to demonstrate that the predictive power of the Euclidean distance measure of deviation from the ideal profile is significantly better than a measure calculated as deviation from a random profile (Venkatraman, 1989). However, empirical implementation of this test tends to vary. For example, Hult et al. (2006) compared performance outcomes of deviation from the ideal profile with a non-ideal baseline of profile deviation from ''average performers'' (cases at the median on the performance scale), whereas Vorhies and Morgan (2005) calibrated a non-ideal baseline of profile deviation from a random selection of five firms with unknown performance scores. We adopt a different approach to be more conservative by using multiple methods (methodological, industry comparison, public perception, and expert opinions). Collectively, our multiple validation tests demonstrate that: a) profile deviation is replicable with an alternative (clustering) method, b) our greenwashing measure distinguishes potentially greener industries from potentially less-green industries, c) our measure relates as expected to public sentiment, and d) a firm's tweet that scores high on our greenwashing measure correlates highly with the likelihood that the tweet will be tagged as greenwashed. We discuss each test in detail next.

For methodological validation, we employ clustering (McQueen, 1967) as an unsupervised machine learning approach to test the predictive power of the linguistic cues. The rationale for this validity test is to check whether an alternate method will yield similar greenwashing scores as the profile deviation method. While both profile deviation and clustering are distance-based methods, the former calculates distances of composite linguistic dimension scores from a pre-defined fixed point, whereas the latter exploits distances between the linguistic dimension scores themselves. Hence, a clustering solution that is valid, stable, robust and yields cluster patterns consistent with our greenwashing patterns will provide external validation to the profile deviation model.

To begin, a clustering tendency test for the presence of meaningful clusters in the data yielded a Hopkins statistic of 0.056 which is far below the test value of 0.5. Thus, we reject the null hypothesis that the data is uniformly distributed and has no meaningful clusters (Hopkins & Skellam, 1954). Visual confirmation of clustering tendency is plotted in a dissimilarity matrix image in Figure S4.1, in which color level representing the dissimilarity between observations shows a clear cluster structure in the data. The next natural question is to identify the optimal number of clusters. We explored preliminary suggestions from seven hierarchical clustering algorithms and summarize results obtained with a universal cut-off height of 15 units on their dendrograms in Table S4.1 (e.g., see Figure S4.2 for the cluster dendrogram for Ward's linkage). The cluster quality measures—the cophenetic correlation of original and clustered distances between observations—are all above the minimum acceptable value of 0.75 (Kaufman & Rousseeuw, 2009; Mather, 1976). Ward Linkage yielded 11 clusters with the best-balanced average cluster size.

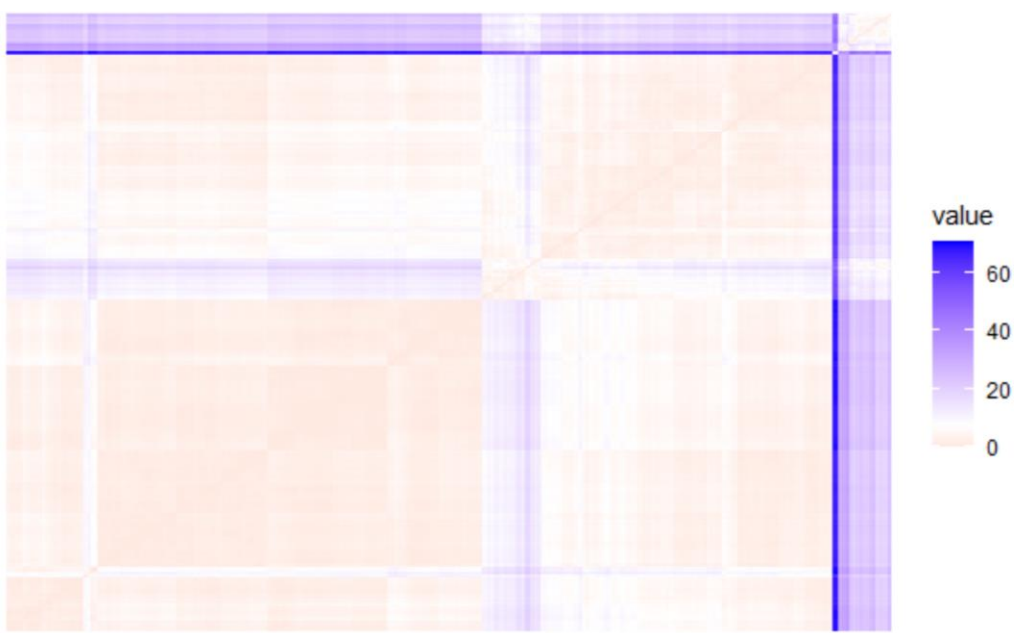

Figure S4.1. Dissimilarity Matrix Image of Cluster Tendency*

**The color level is proportional to the value of the dissimilarity between observations: pure red if dist(x_i, x_j) = 0 and pure blue if dist(x_i, x_j) = 1. Objects belonging to the same cluster are displayed in consecutive order. The dissimilarity matrix image confirms the presence of a cluster structure in the data.*

Table S4.1. Hierarchical Clustering Solutions with Dendrogram Cut-off at h=15

Item	Linkage Method	No. of Clusters	Average Cluster Size	Cophenetic Correlation
1	Single	3	27409	0.8857
2	Complete	6	13706	0.8590
3	Average	4	20557	0.9483
4	Weighted	4	20557	0.8703
5	Centroid	3	27409	0.9490
6	Median	4	20557	0.8410
7	Ward	11	7475	0.7551

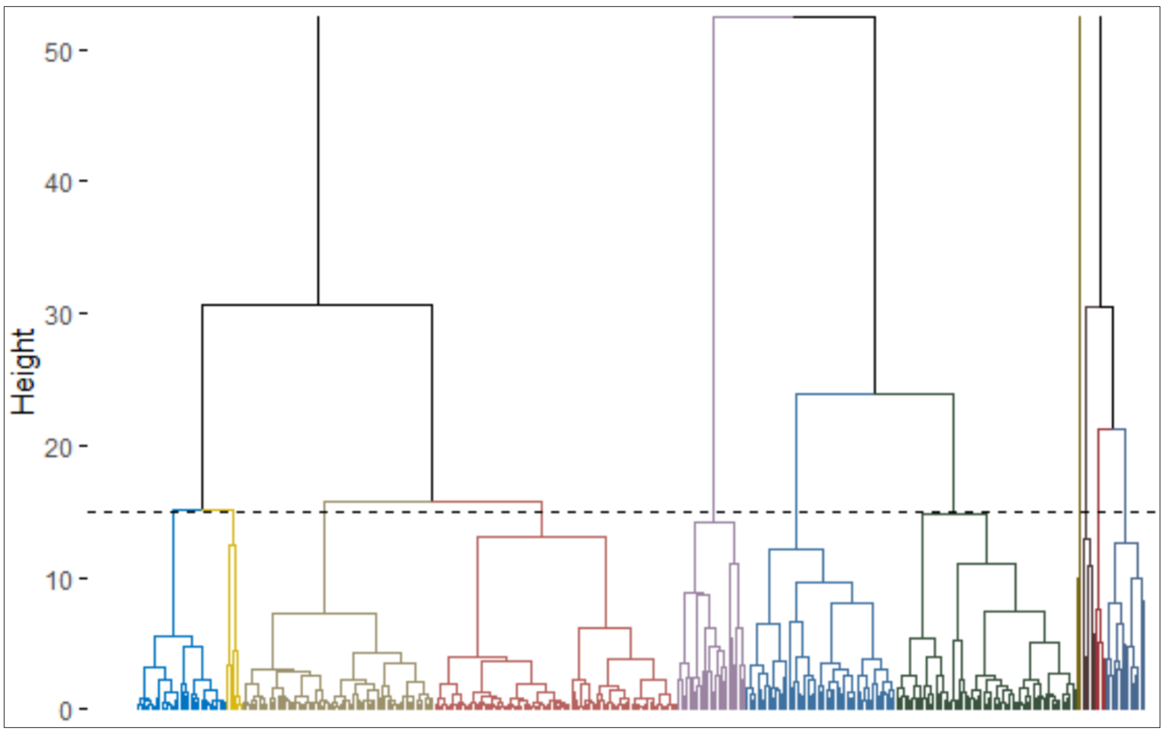

Figure S4.2. Hierarchical Clustering Dendrogram (Ward Linkage) *
**Zoomed in for clearer visualization (truncated at height=50)*

However, hierarchical clustering solutions are notoriously rigid (i.e., once formed, linkages cannot be broken). Hence, we use K-means clustering to explore the stability of the clustering solution. With a potential solution at K=11 known from hierarchical clustering; we test for the optimal K from 1 to 15 clusters. We explored the elbow method (Thorndike, 1953), average silhouette method (Kaufman & Rousseeuw, 2009) and the gap statistic method (Tibshirani et al., 2001). The well-known elbow method suggested 5 clusters, the silhouette method suggested 9 clusters, while the gap statistic method suggests either 11 or 12 clusters. Given that the Gap Statistic method is a much more efficient improvement over both the elbow and silhouette methods, we proceeded to examine the k=11 versus k=12. We used the NbClust functions in R version 4.0.2 (Charrad et al., 2014; Kassambara, 2017) that examine 30 indices to choose the best cluster via majority voting. The results suggested an optimal number of 12 clusters, a solution that is close to the ten greenwashing patterns from profile deviation.

Next, we validate the cluster solution with three internal measures: cluster cohesion, separation, and connectedness. Dunn Index (0.22) and average silhouette width (0.29) all point to acceptable cluster compactness or cluster cohesion. Within and between cluster sums of squares further showed good cluster separation. Cluster connectedness measures the extent to which items are placed in the same cluster as their nearest neighbors in the data space. Silhouette plot and further analysis revealed that only 6% of observations were not placed with their nearest neighbors (i.e., not in the right cluster). Given these reasonably stable cluster metrics, we finally tested the extent to which cluster assignments correlated with profile deviation assignments of observations. For each cluster, we use the resulting mean cluster scores for each cue (see Table S4.2 and Figure S4.3) to compute Euclidean distances from our theoretically ideal profile. Table S4.3 shows that the raw Euclidean distance scores and quantile range measures for profile deviation patterns and cluster memberships correlated highly at Pearson = 0.98 and Pearson = 0.88, respectively. This methodological cross-validation presents strong evidence that our greenwashing patterns are stable. For comparison to Table 2 in the main paper, Table S4.4 includes the average greenwashing scores per cluster.

Table S4.2. Cluster Means on Linguistic Cues

Cluster No.	**Linguistic Cues**						
	Quantity	*Specificity*	*Complexity*	*Diversity*	*Hedging*	*Affect*	*Vividness*
1	13.99	40.41	37.09	12.95	-17.84	-14.48	0.30
2	5.24	16.37	16.62	4.85	-10.31	-5.44	1.15
3	8.27	10.50	8.72	8.97	-3.88	-2.28	1.30
4	3.61	9.60	11.61	3.54	-5.06	-5.02	0.04
5	2.20	4.75	4.69	1.92	-4.25	-1.92	0.76
6	0.83	1.77	1.81	0.77	-1.37	-0.78	0.29
7	-0.11	0.03	0.05	-0.06	0.43	0.14	-1.49
8	0.09	-0.11	-0.34	0.05	-0.44	0.06	1.23
9	-0.65	-1.62	-1.67	-0.60	1.10	0.63	0.27
10	-1.19	-2.79	-2.80	-1.16	1.69	1.09	0.85
11	-1.16	-2.63	-2.66	-1.10	1.71	1.07	-2.08
12	-1.17	-2.61	-2.78	-1.14	1.13	1.14	-6.54

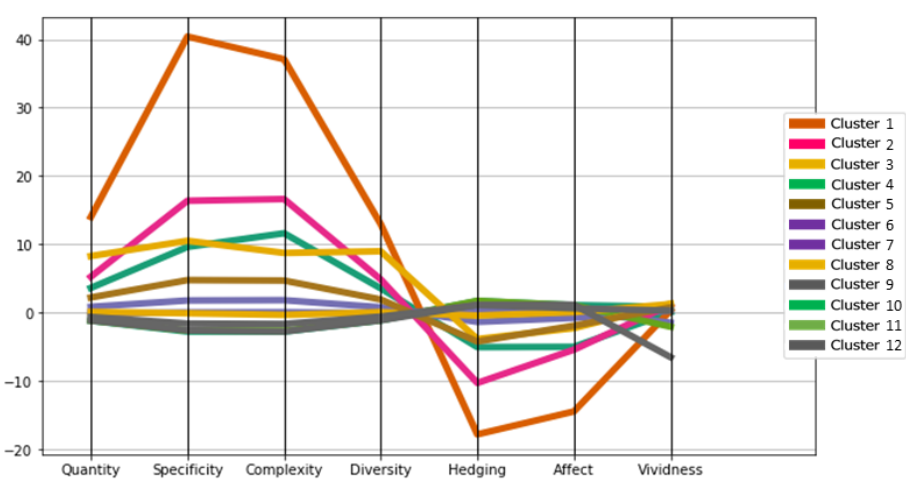

Figure S4.3. K-Means Cluster Profiles Plot

Table S4.3. Correlations between Greenwashing Patterns by Profile Deviation (PD) and Clustering

Item	Greenwashing measure	1	2	3	4
1	PD Euclidean Distance	1.00			
2	PD Quantile Range	0.75	1.00		
3	Cluster Euclidean Distance	0.98	0.73	1.00	
4	Cluster Quantile Range	0.89	0.88	0.91	1.00

Table S4.4. Average Greenwashing Score per Cluster

Cluster No.	No. of Observations	Average Greenwashing Score
1	1817	73.25
2	10134	76.66
3	4381	78.93
4	1044	80.64
5	15867	83.34
6	1145	84.65
7	5433	85.55
8	11316	86.57
9	10112	87.56
10	9339	88.48
11	6926	89.84
12	4712	90.65

The second validation approach assessed the accuracy of our method by distinguishing greenwashing by our firms in the oil/gas and auto industries from firms in the environmental management industry. The rationale is that our firms face growing institutional pressures to adopt more green practices and have higher motivation to communicate green actions through greenwashing. In contrast, firms in the environmental management industry are perceived as environmentally proactive (e.g., Henriques & Sadorsky, 1999) and have fewer incentives to greenwash their communications. We sampled all firms classified under the environmental industry in the Forbes Global 2000 classification and selected those with tweets spanning the same period of observation. Next, we measured the level of greenwashing in their tweets with the same profile deviation method used for our focal firms. We found, as expected, that green tweets from the environmental firms fell in the lowest decile on the quantile range splits, indicating the closest distance to the ideal non-greenwashing (truthful) profile[1].

For our third validation test, we examined the extent to which our greenwashing measure correlates with public perceptions of firms' green tweets. Prior studies have examined various metadata

[1] To avoid any ramifications of publicly disclosing the greenwashing scores of firms, we only report average greenwashing scores for deciles and clusters in the paper. Interested readers can contact the authors to verify individual greenwashing scores for firms.

about social media posts, such as likes, favorites, retweets, and votes to study consumer engagement in greenwashing (Topal et al., 2019), fake online reviews (Zhang et al., 2016), social media virality (Han et al., 2020), information credibility (Castillo et al., 2011), and mobilization of public sentiment for political activism (Theocharis et al., 2015). Based on these studies and given the general climate consciousness or virtue signaling in social media, we expect Twitter users to be more likely to endorse green tweets perceived as genuine by retweeting, mentioning, or liking them. Although general Twitter users will not be adept at spotting greenwashing, their endorsement (or otherwise) of firm's green tweets serves a proxy for public sentiments about the tweet. Thus, perceived greenwashed tweets should be negatively related to retweets, favorites, and mentions on Twitter. We aggregated the number of retweets and favorites for all tweets and correlated them with our greenwashing measure. We found negative spearman correlations, significant for retweets (ρ = -0.13, $p < 0.01$) but not for favorites (ρ = -0.06, $p = 0.26$), with the retweet effect showing that higher greenwashing generally had fewer endorsements. Although the correlation size is small, the negative direction adds credibility to our greenwashing measure as potentially reflecting public sentiments on tweets. Further, this weak correlation is consistent with prior findings that humans are not apt at detecting deception (Crilly et al., 2016; George et al., 2013; L. Zhou et al., 2004). One potential limitation to using retweet as a validation measure is the possibility that users may retweet to ridicule or criticize the content, rather than the common practice of endorsing it. Future research could analyze sentiments in user comments accompanying retweets to rule out such cases.

Our final validation test examined the extent to which outsiders would agree with our greenwashing score for a firm's green tweet. The rationale for this test is that if a firm's green communication is deceptive and our greenwashing measure adequately labels it so, then those who specialize in calling out corporate greenwashing should come to a similar conclusion. Thus, we can expect a moderate to high correlation between our greenwashing measure and the likelihood of it being tagged as deceptive. Data for this test were scraped from all tweets labeled with the hashtags #greenwashing or #greenwash on Twitter from 2009 to 2019, overlapping the period of the study. These two hashtags are the established monikers widely used by environmental activists to shine a spotlight on both real and perceived instances of greenwashing by firms. Together, they returned over 26,000 posts that were tagged as greenwashing (across all topic areas, not just for our two industries). To focus on our study firms, we extracted mentions that tagged these firms up to one week following the firm posting of a green tweet. We created a binary variable named *Tagged* which scored firms 1 if their tweets were labeled as greenwashed and 0 otherwise. We believe that this tagging is appropriate for our validation because it includes posts by many well-known environmental activists, such as Extinction Rebellion and Greenpeace, who are more likely than the general public to spot instances of greenwashing on Twitter.

We analyzed whether the linguistic cues in Table 1 of the main paper would accurately predict the likelihood that a tweet with a high greenwashing score (i.e., high profile deviation) would be labeled with greenwashing hashtags in the following week. Using supervised machine learning techniques, we built and trained five competing classifiers including logistic regression, random forest, support vector machine (SVM), extreme gradient boosting (XGB), and neural net classifiers. The predictor variables include all eight linguistic cues and the outcome variable is *Tagged* (with two unbalanced classes: 1 = 12,665, 0 = 69,562). To improve classification accuracy, we used a balanced subsample of 25,000 tweets and a train-test split ratio of 70:30 for each classifier. We evaluated the model validation accuracy and F1 score (ability to identify more true than false positives) to select the best model. Finally, we calculated the correlation between the predicted propensity scores and our profile deviation scores.

Model evaluation results are included in Table S4.5 below. While all models had high accuracy scores, the Logistic regression classifier yielded the best model, with the highest F1 score indicating a better ability to accurately identify instances of tagged greenwashing than any other model. In the final step, the correlation between the propensity scores from the Logistic regression model and the profile deviation scores was 0.72 and significant ($p = 0.00$). Thus, our greenwashing measure correlates highly with the likelihood that the firm's communication will be tagged as greenwashed soon after the firm posts a green tweet. This result provides additional evidence that our profile deviation method represents a valid measure of greenwashing in firms' tweets.

In conclusion, while no single validation method on its own is sufficient, the collective results from four independent validation methods (i.e., methodological, industry, public perception, and expert opinion) provide strong evidence for the validity of our greenwashing measure.

Table S4.5. Classification model evaluation and selection

No.	Model	Training Accuracy	Validation Accuracy	F1-Score
1.	Logistic Regression	91.85%	91.33%	95.45%
2.	Random Forest	98.63%	79.70%	84.70%
3.	XGB	96.73%	85.55%	88.98%
4.	SVM	91.44%	76.34%	81.63%
5.	Neural Nets	93.56%	86.29%	82.71%

Section S5-S7: Supplementary Materials for Research Question 2

S5. Variables and Sources used in Regressions

Table S5. Variable Definitions

Variable	Definition	Data Source
Greenwashing	*Euclidean distance from ideal non-deceptive profile on linguistic cues of a firm's green tweet.*	Twitter
ESG Controversies	*Number of environmental, social and governance incidents listed in a firm's controversy report.*	Sustainalytics
Share Price	*The daily average stock price of a firm.*	Bloomberg
Gross Income	*A firm's net sales revenues minus cost of goods sold.*	Bloomberg
Return on Assets	*A firm's ratio of net income per total assets.*	Bloomberg
Operating Income	*A firm's earnings before interest, taxes, depreciation and amortization, adjusted per year.*	Bloomberg
Revenue	*A firm's income generated from normal business operations, adjusted per year*	Bloomberg
Profit	*A firm's total revenue minus total expenses, adjusted per year.*	Bloomberg

S6. Descriptive Statistics and Correlations

Table S6.1 Descriptive Statistics

Variable	N	Mean	Std. Dev.	Min	Max
log Share Price	37535	3.45	1.04	0.32	5.94
log Greenwashing	37580	4.45	0.03	4.27	4.61
log ESG Controversies	22748	1.51	0.98	0.00	3.78
Industry (0=Auto, 1=Oil)	34532	0.44	0.50	0.00	1.00
Region (0=N.Am, 1=Global)	34532	0.72	0.45	0.00	1.00
Size (0=Bottom, 1=Top)	34532	0.68	0.47	0.00	1.00
log Gross Income	34352	2.40	1.30	-0.14	4.50
log Return on Assets	29592	1.62	0.61	-0.47	3.27
log Operating Income	33231	2.58	0.67	0.00	4.28
log Profit	34352	5.63	4.05	0.00	11.19
log Revenue	34532	8.99	3.34	0.00	13.05

Table S6.2 Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) log Share Price	1.00										
(2) log Greenwashing	-0.09	1.00									
(3) log ESG Controversies	0.17	0.11	1.00								
(4) Industry	0.10	-0.07	0.17	1.00							
(5) Region	-0.25	0.07	0.05	-0.33	1.00						
(6) Size	0.40	-0.06	0.36	0.21	-0.16	1.00					
(7) Gross Income	0.20	0.01	-0.08	-0.35	-0.02	0.05	1.00				
(8) log Return on Assets	-0.13	-0.01	-0.21	-0.11	-0.06	-0.18	0.18	1.00			
(9) log Operating Income	0.05	0.04	0.03	0.37	-0.07	0.04	-0.02	0.19	1.00		
(10) log Profit	0.38	-0.03	0.08	0.25	-0.48	0.25	0.49	-0.19	-0.01	1.00	
(11) log Revenue	0.17	-0.01	0.19	0.63	-0.40	0.24	-0.37	-0.35	0.22	0.57	1.00

S7. Robustness Tests

We include several robustness tests using alternative specifications with: our greenwashing measured from clustering (Table S7.1); week-level random effects models for market outcome (Table S7.2); and individual effects of greenwashing dimensions (Table S7.3). Overall, these results support the robustness of our main results to alternate specifications.

Table S7.1 Greenwashing by Clustering: Daily Financial Market Performance

	Dependent Variable: ***Share Price***			
Variables	Model (1)	Model (2)	Model (3)	Model (4)
Greenwashing (GW)		-0.65**		-0.93**
		(0.05)		(0.10)
ESG Controversies (ESGC)			-0.04**	-0.65*
			(0.00)	(0.26)
GW x ESGC				0.14*
				(0.06)
Industry (0=Auto, 1=Oil)	-0.44	-0.48	-0.76†	-0.82*
	(0.37)	(0.37)	(0.41)	(0.41)
Region (0=NA, 1=Global)	-0.66	-0.66	-0.89†	-0.84†
	(0.41)	(0.38)	(0.47)	(0.48)
Size (0=B20; 1=T20)	0.90**	0.88*	0.65	0.59
	(0.34)	(0.35)	(0.40)	(0.40)
Gross Income	-0.21**	-0.22**	-0.04	-0.06**
	(0.02)	(0.02)	(0.03)	(0.03)
Return on Assets	-0.01	-0.01	0.19**	0.19**
	(0.01)	(0.00)	(0.01)	(0.01)
Operating Income	0.45**	0.46**	0.27**	0.29**
	(0.01)	(0.01)	(0.01)	(0.01)
Profit	0.07**	0.07**	-0.05**	-0.04**
	(0.01)	(0.01)	(0.01)	(0.01)
Revenue	0.04**	0.05**	0.13**	0.14**
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.81**	4.62**	2.19**	6.17**
	(0.47)	(0.53)	(0.55)	(0.72)
Observations	29,477	29,477	17,435	17,435
Number of Firms	50	50	42	42

Random Effects regression estimates with standard errors in parentheses. Firm-day panel. *Region NA:* North America. Greenwashing measured as Cluster Score (see validation with Clustering Method in Section 4). *Size T, B*: Top & Bottom 20 rank by market cap (Forbes Global 2000). ** $p<0.01$, * $p<0.05$, † $p<0.10$.

Table S7.2 Greenwashing by Profile Deviation: Weekly Financial Market Performance

Variables	Model (1)	Model (2)	Model (3)	Model (4)
	Dependent Variable: ***Share Price***			
Greenwashing (GW)		-0.46**		-0.79**
		(0.10)		(0.18)
ESG Controversies (ESGC)			-0.06**	-0.87†
			(0.00)	(0.45)
GW x ESGC				0.18†
				(0.10)
Industry (0=Auto, 1=Oil)	-0.40	-0.43	-0.92*	-0.97*
	(0.37)	(0.37)	(0.40)	(0.41)
Region (0=NA, 1=Global)	-0.60	-0.60	-0.61	-0.57
	(0.41)	(0.42)	(0.47)	(0.48)
Size (0=B20; 1=T20)	0.83*	0.82*	0.44	0.37
	(0.35)	(0.35)	(0.39)	(0.36)
Gross Income	-0.26**	-0.26**	-0.09*	-0.12**
	(0.02)	(0.02)	(0.04)	(0.04)
Return on Assets	0.12**	0.12**	0.10**	0.10**
	(0.01)	(0.00)	(0.01)	(0.01)
Operating Income	0.33**	0.33**	0.30**	0.31**
	(0.01)	(0.01)	(0.02)	(0.02)
Profit	0.09**	0.09**	0.00	0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Revenue	0.03**	0.04**	0.19**	0.20**
	(0.01)	(0.01)	(0.02)	(0.02)
Constant	2.10**	4.09**	1.55**	4.93**
	(0.48)	(0.64)	(0.56)	(1.00)
Observations	10,484	10,484	6,969	6,969
Number of Firms	50	50	42	42

Random Effects regression estimates with standard errors in parentheses. Firm-week panel. Greenwashing measured as Profile Deviation Score. *Region NA:* North America. *Size T, B*: Top & Bottom 20 rank by market cap (Forbes Global 2000). ** p<0.01, * p<0.05, † p<0.10.

Table S7.3 Individual Effects of Greenwashing Dimensions ***(continues next page)***

	Dependent Variable = ***Share Price***														
Variables	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)	Model (9)	Model (10)	Model (11)	Model (12)	Model (13)	Model (14)	Model (15)
ESG Controversy (ESGC)	0.50**								1.00**	-0.25**	0.61**	2.39**	1.90**	-0.47*	-2.23**
	(0.02)								(0.04)	(0.08)	(0.05)	(0.19)	(0.78)	(0.18)	(0.28)
Quantity		-13.37**							-6.16**						
		(0.90)							(1.12)						
Specificity			-2.47**							-2.88**					
			(0.32)							(0.39)					
Complexity				-2.80**							1.21**				
				(0.32)							(0.42)				
Diversity					-3.42**							1.71			
					(1.26)							(1.54)			
Hedging						6.75**							7.12**		
						(1.12)							(1.50)		
Affect							2.34**							-1.03*	
							(0.36)							(0.44)	
Vividness								13.34**							-1.57
								(1.22)							(1.39)
Quantity x ESGC									-1.57**						
									(0.12)						
Specificity x ESGC										0.50**					
										(0.06)					
Complexity x ESGC											-0.12**				
											(0.06)				
Diversity x ESGC												-1.70**			
												(0.17)			
Hedging x ESGC													-0.36*		
													(0.20)		
Affect x ESGC														0.37**	

	Dependent Variable = *Share Price*														
	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
														(0.07)	
Vividness x ESGC															2.05**
															(0.21)
Industry (0=Auto,					-				-						
1=Oil)	-22.30†	-22.09†	-21.77†	-21.45†	21.97†	-21.87†	-21.96†	-20.48	22.44†	-22.22†	-22.43†	-22.26†	-22.18†	-22.15†	-21.38†
	(12.54)	(12.84)	(12.91)	(12.91)	(12.80)	(12.91)	(12.87)	(12.93)	(12.23)	(12.84)	(12.58)	(12.61)	(12.80)	(12.35)	(12.07)
Region (0=NA,	-	-	-	-	-	-	-		-	-	-	-	-	-	-
1=Global)	37.79**	37.63**	37.86**	37.71**	37.91**	37.79**	38.06**	-36.85*	37.37**	37.67**	37.85**	37.61**	37.67**	37.76**	37.45**
	(14.03)	(14.36)	(14.44)	(14.44)	(14.32)	(14.44)	(14.40)	(14.46)	(13.67)	(14.36)	(14.07)	(14.10)	(14.31)	(13.81)	(13.50)
Size (0=B20, 1=T20)	14.11	16.08	16.13	16.35	16.14	16.00	16.23	15.37	14.66	13.50	14.25	14.05	13.95	14.57	14.95
	(11.99)	(12.27)	(12.34)	(12.34)	(12.24)	(12.34)	(12.31)	(12.36)	(11.69)	(12.27)	(12.03)	(12.05)	(12.23)	(11.81)	(11.54)
Gross Income	-0.23**	-0.22**	-0.26**	-0.27**	-0.26**	-0.26**	-0.26**	-0.28**	-0.17**	-0.23**	-0.22**	-0.22**	-0.23**	-0.23**	-0.24**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Return on Assets	1.85**	1.89**	1.90**	1.90**	1.91**	1.91**	1.90**	1.92**	1.79**	1.83**	1.85**	1.83**	1.85**	1.85**	1.85**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Operating Income	0.00**	0.00**	0.00*	0.00*	0.00*	0.00*	0.00*	0.00†	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Profit	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Revenue	0.00**	0.00†	0.00	0.01	0.02	0.03	0.04	0.05	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00†
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	70.28**	75.35**	75.74**	74.50**	75.36**	45.06**	65.99**	54.09**	70.99**	74.92**	68.97**	67.97**	42.19*	72.74**	71.93**
	(15.67)	(16.04)	(16.14)	(16.13)	(16.05)	(16.73)	(16.11)	(16.24)	(15.28)	(16.05)	(15.73)	(15.84)	(16.85)	(15.47)	(15.19)
Observations	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532	34,532
Number of Firms	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52

Random Effects regression estimates with standard errors in parentheses. Firm-day panel. *Quantity, Specificity, Complexity, Hedging, Diversity, Affect,* and *Vividness* are the dimensions of greenwashing as profile deviation. *Region NA:* North America. *Size T, B*: Top & Bottom 20 rank by market cap (Forbes Global 2000). ** p<0.01, * p<0.05, † p<0.10.

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Altowim, Y., Kalashnikov, D. V., & Mehrotra, S. (2014). Progressive approach to relational entity resolution. *Proceedings of the VLDB Endowment*, *7*(11), 999–1010. https://doi.org/10.14778/2732967.2732975

Asudeh, A., Jagadish, H. V., Wu, Y. (Will), & Yu, C. (2020). On detecting cherry-picked trendlines. *Proceedings of the VLDB Endowment*, *13*(6), 939–952. https://doi.org/10.14778/3380750.3380762

Bowen, F., & Aragon-Correa, J. A. (2014). Greenwashing in corporate environmentalism research and practice: The importance of what we say and do. *Organization & Environment*, *27*(2), 107–112. https://doi.org/10.1177/1086026614537078

Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., & Freire, J. (2019). A topic-agnostic approach for identifying fake news pages. *Companion Proceedings of The 2019 World Wide Web Conference*, 975–980. https://doi.org/10.1145/3308560.3316739

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 675–684. https://doi.org/10.1145/1963405.1963500

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(1), 1–36. https://doi.org/10.18637/jss.v061.i06

Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as "false news." *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, 15–19. https://doi.org/10.1145/2823465.2823467

Crilly, D., Hansen, M., & Zollo, M. (2016). The grammar of decoupling: A cognitive-linguistic perspective on firms' sustainability claims and stakeholders' interpretation. *Academy of Management Journal*, *59*(2), 705–729. https://doi.org/10.5465/amj.2015.0171

Feng, S., Banerjee, R., & Choi, Y. (2012). Syntactic stylometry for deception detection. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 171–175. https://www.aclweb.org/anthology/P12-2034

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

George, J. F., Carlson, J. R., & Valacich, J. S. (2013). Media selection as a strategic component of communication. *MIS Quarterly*, *37*(4), 1233–1251. https://doi.org/10.25300/MISQ/2013/37.4.11

Han, Y., Lappas, T., & Sabnis, G. (2020). The importance of interactions between content characteristics and creator characteristics for studying virality in social media. *Information Systems Research*, isre.2019.0903. https://doi.org/10.1287/isre.2019.0903

Henriques, I., & Sadorsky, P. (1999). The relationship between environmental commitment and managerial perceptions of stakeholder importance. *Academy of Management Journal*, *42*(1), 87–99. https://doi.org/10.2307/256876

Hoffart, J., Suchanek, F. M., Berberich, K., & Weikum, G. (2013). YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, *194*, 28–61. https://doi.org/10.1016/j.artint.2012.06.001

Hopkins, B., & Skellam, J. G. (1954). A new method for determining the type of distribution of plant individuals. *Annals of Botany*, *18*(2), 213–227. https://doi.org/10.1093/oxfordjournals.aob.a083391

Hult, G. T. M., Ketchen, D. J., Cavusgil, S. T., & Calantone, R. J. (2006). Knowledge as a strategic resource in supply chains. *Journal of Operations Management*, *24*(5), 458–475. https://doi.org/10.1016/j.jom.2005.11.009

Ireton, C., Posetti, J., & UNESCO. (2018). *Journalism, "Fake News" and Disinformation: Handbook for Journalism Education and Training*. http://unesdoc.unesco.org/images/0026/002655/265552E.pdf

Kang, B., & Deng, Y. (2019). The maximum Deng entropy. *IEEE Access*, *7*, 120758–120765. https://doi.org/10.1109/ACCESS.2019.2937679

Kassambara, M. A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning* (1st edition). STHDA.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc.

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, *27*(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

Mather, P. M. (1976). *Computational Methods of Multivariate Analysis in Physical Geography*. Wiley.

McQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. L. Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification* (pp. 281–297). University of California Press.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, *104*(1), 11–33. https://doi.org/10.1109/JPROC.2015.2483592

Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–4. https://doi.org/10.1002/pra2.2015.145052010083

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 4787. https://doi.org/10.1038/s41467-018-06930-7

Sitaula, N., Mohan, C. K., Grygiel, J., Zhou, X., & Zafarani, R. (2020). Credibility-based fake news detection. In K. Shu, S. Wang, D. Lee, & H. Liu (Eds.), *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* (pp. 163–182). Springer International Publishing. https://doi.org/10.1007/978-3-030-42699-6_9

Spirin, N., & Han, J. (2012). Survey on web spam detection: Principles and algorithms. *ACM SIGKDD Explorations Newsletter*, *13*(2), 50–64. https://doi.org/10.1145/2207243.2207252

Szabo, S., & Webster, J. (2021). Perceived greenwashing: The effects of green marketing on environmental and product perceptions. *Journal of Business Ethics*, 719–739. https://doi.org/10.1007/s10551-020-04461-0

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. *Digital Journalism*, *6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Theocharis, Y., Lowe, W., van Deth, J. W., & García-Albacete, G. (2015). Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, *18*(2), 202–220. https://doi.org/10.1080/1369118X.2014.948035

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Topal, İ., Nart, S., Akar, C., & Erkollar, A. (2019). The effect of greenwashing on online consumer engagement: A comparative study in France, Germany, Turkey, and the United Kingdom. *Business Strategy and the Environment*, *0*(0). https://doi.org/10.1002/bse.2380

Venkatraman, N. (1989). The concept of fit in strategy research: Toward verbal and statistical correspondence. *Academy of Management Review*, *14*(3), 423–444. https://doi.org/10.5465/amr.1989.4279078

Vilone, G., & Longo, L. (2020). Explainable artificial intelligence: A systematic review. *ArXiv:2006.00093 [Cs]*. http://arxiv.org/abs/2006.00093

Vorhies, D. W., & Morgan, N. A. (2005). Benchmarking marketing capabilities for sustainable competitive advantage. *Journal of Marketing*, *69*(1), 80–94. https://doi.org/10.1509/jmkg.69.1.80.55505

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wardle, C. (2018). The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*, *6*(8), 951–963. https://doi.org/10.1080/21670811.2018.1502047

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* (No. 162317GBR). Council of Europe. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html (accessed on 10 May 2020)

Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, *132*, 113280. https://doi.org/10.1016/j.dss.2020.113280

Ye, J., & Skiena, S. (2019). Mediarank: Computational ranking of online news sources. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2469–2477. https://doi.org/10.1145/3292500.3330709

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality*, *11*(3), 10:1-10:37. https://doi.org/10.1145/3309699

Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, *33*(2), 456–481. https://doi.org/10.1080/07421222.2016.1205907

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation*, *13*(1), 81–106. https://doi.org/10.1023/B:GRUP.0000011944.62889.6f

Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2020). Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice*, *1*(2), 12:1-12:25. https://doi.org/10.1145/3377478

Zhou, X., & Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. *ACM SIGKDD Explorations Newsletter*, *21*(2), 48–60. https://doi.org/10.1145/3373464.3373473

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, *53*(5), 1–40. https://doi.org/10.1145/3395046

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys*, *51*(2), 32:1-32:36. https://doi.org/10.1145/3161603