

Supplementary file

Drivers of Public Engagement in Social Media Discourse on Self-Driving Vehicles: Psychological and Emotional Determinants

Contents

S1. Supplementary Materials for the Analyses	2
S2. Supplementary Materials for Validation	7
S2.1. Data validation and pre-processing	7
S2.2. Validation of Sentiment and Emotion Detection	8
S2.3. Validation of Topic Assignment.....	10
S2.4. Validation of Construal-Level Theory (CLT) Classification	13
S3. Supplementary Materials for Robustness Check	16
S3.1. Robustness Check using different engagement types as dependent variables.....	16
S3.2. Robustness Check using Predictive Models	25
S3.2.1. Results of robustness check with predictive modelling	25
S3.2.2. Reasoning for Choice of Predictive Models.....	31

List of Tables and Figures

Table S1 Usage of the variables.....	2
Table S2. Collection of the main Python libraries	3
Table S3. Keywords for off-/on-topic	3
Table S4. Variance Inflation Factor (VIF) scores of the independent variables.....	4
Table S5. Summary statistics of the compound sentiment and emotion scores.....	6
Table S6. Basic statistics for the posts' engagement metrics	7
Table S7 Annotators (L1-L4) agreement measures in the sample of 100 tweets.....	9
Table S8 Library accuracy (human annotators: 1-3) on 100 tweet sample.....	10
Table S9 Subset and Topic-level validation results.....	12
Table S10 Tweet classification across transport segments	13
Table S11 Annotators' agreement summary on CLT classification.....	15
Table S12 CLT distributions across subsets and topics.....	15
Table S13 Summary results of models per engagement type (Poisson and Negative Binomial Regression models).....	16
Table S14 Results for the NB Regression model—as reported in the main study—for Total Engagement Counts (sum of replies, retweets, and likes)	18
Table S15 Results for the NB Regression model for reply counts as dependant variable.	20
Table S16 Results for the NB Regression model for retweet counts as dependant variable.....	21
Table S17 Results for the NB Regression model for like counts as dependant variable	22
Table S18 Positive coefficient variables ranked by importance across engagement types and datasets.	22
Table S19 Negative coefficient variables ranked by importance across engagement types and datasets.	23
Table S20 Results of predictive model for overall sample.....	25

Table S21 Results of predictive model for Commercial AV sample	30
Table S22 Results of predictive model for AV Private Cars sample	31
Table S23 Results of predictive model for AV Public Transport sample	31
Figure S1 Most frequent words (nouns, verbs, adjectives, adverbs) by the final subsets and features ..	5
Figure S2 Feature importance for emotions and topics for XGBoost model in overall sample	26
Figure S3 Partial dependence of Happy emotion (overall sample).....	26
Figure S4 Partial Dependence of Angry emotion (overall sample)	27
Figure S5 Partial Dependence of Surprise emotion (overall sample).....	27
Figure S6 Partial Dependence of Sad emotion (overall sample)	27
Figure S7 Partial Dependence of Fear emotion (overall sample)	28
Figure S8 Partial Dependence of Safety Feature (overall sample)	28
Figure S9 Partial Dependence of Technology Feature (overall sample).....	28
Figure S10 Partial Dependence of Traffic Feature (overall sample).....	29
Figure S11 Partial Dependence of Employment Feature (overall sample).....	29
Figure S12 Partial Dependence of Events Feature (overall sample).....	29
Figure S13 Partial Dependence of Pricing Feature (overall sample).....	30
Figure S14 Partial Dependence of Infrastructure Feature (overall sample).....	30

S1. Supplementary Materials for the Analyses

The content of this supplementary file pertains to several sections in the main body of our study. To complement Section 4, Table S1 lists the variables and their utilisation; Table S2 summarises the various Python libraries used in our analyses; and Table S3 lists the off-topic and on-topic keywords. In relation to Section 4.3.1, basic statistics for the dataset and its subsets are presented in Tables S5 and S6, while Table S4 reports the Variance Inflation Factor (VIF) scores for the independent variables. Finally, Figure S1 illustrates the ten most frequent words (nouns, verbs, adjectives, and adverbs) and their importance across the three market-segment subsets and seven features.

Table S1 Usage of the variables

variables	Description	Notes on usage
tweet_url	url address of the tweet containing the unique tweet ID	For the timeline of tweets to count them
date	the date the tweet was posted (YYYY-MM-DD HH:MM:SS+00:00)	Base for custom made columns: 1) date_time (DD/MM/YYYY HH:MM) for hour_of_the_day column to be used in the NB regression model; 2) only_date (DD/MM/YYYY) to be used in the timeline as well to calculate the difference between scrape day and get the post_age for NB regression model.
content	the tweet itself max 140 character, urls only as .tco	Base for the clean_content column for all further analysis.
user_displayname	display name of the user	to get statistics about the users (e.g. unique users) and to drop duplicates from same user with same clean_content.
replyCount	the number of reply the tweet has (at the time of the scraping)	used for the NB regression model as dependent variable
retweetCount	the number of retweet the tweet has (at the time of the scraping)	used for the NB regression model as dependent variable

variables	Description	Notes on usage
likeCount	the number of like the tweet has (at the time of the scraping)	used for the NB regression model as dependent variable
lang	language of the tweet	only "en" English tweets were used

Table S2. Collection of the main Python libraries

Steps	Python package(s)	Rule(s)
Data gathering	Snsrape	"Autonomous" OR "Self-drive" OR "Driverless" lang=en; since:2022-03-01 until:2022-06-01
Data cleaning	pandas, re	- keep only English tweets, and - drop duplicates from same user. Cleaning the text: - remove url links, cashtags, mentions - remove special characters - remove leading and whitespaces - convert all characters to lower case
Lemmatization	re, spaCy, gensim	allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV'] column data_lemmatized contains all postags, while column nouns only the NOUNs.
Sentiment analysis	nltk (Vader Lexicon): SentimentIntensityAnalyzer	Use of compound scores only, range from -1 to +1. negative < -0.05; neutral = {-0.05, 0.05}; positive > 0.05
Emotion detection	text2emotion	Scores the text for 5 emotions: Happy, Angry, Surprise, Sad and Fear. The scores of the emotions are adding up either to 0 or 1 altogether. If no emotion is detected (all 0s), then is named in the analysis as Neutral.
Topic Modelling	sklearn, gensim, pyLDAvis, LatentDirichletAllocation, TruncatedSVD, CountVectorizer, TfidfVectorizer, and GridSearchCV	(1) Lemmatized tweets. (2) Best LDA model: 0.5 learning decay, 10 topics
Frequency and Bigrams	nltk, itertools, networkx	Top frequency (50 for initial off-topic detection, 30 in latter analyses), top 50 bigram.
Timeline	datetime	Sum of the daily counts.
Validation / CLT classification	Transformers: AutoTokenizer, AutoModelForCausalLM, pipeline	To use Mistral AI LLM, through Hugging face open-source transformers for various validation and classification purposes
Regression	statsmodels, sklearn	Negative Binomial Regression: default statmodels settings.
Visualisation	matplotlib, seaborn, plotly	

Table S3. Keywords for off-/on-topic

Keywords for off/on-topic (whole dataset)	
off-topic	'autonomous region', 'autonomous organization', 'autonomous republic', 'autonomous zone', 'tibet', 'crimea', 'autonomous weapons', 'autonomous territory', 'autonomous country', 'autonomous university'
on-topic	'self driving', 'selfdriving', 'driverless', 'autonomous vehicle', 'autonomous car', 'delivery', 'autonomous truck', 'driverless truck', 'robotruck', 'autonomous driving', 'tesla', 'semi truck', 'autonomousvehicle', 'autonomous drive', 'autonomous passenger', 'autonomous bus', 'autonomous material movement', 'last mile', 'autonomous transport', 'automotive is autonomous', 'self guided vehicle', 'electric vehicle',

'autonomous ship', 'autonomous port', 'autonomous boat', 'autonomous solution', 'autonomous weeding robot', 'farming', 'autonomous equipment', 'electric tractor', 'autonomous deliver', 'autonomous drone', 'delivery drone', 'flying car', 'flying vehicle', 'autonomous ride', 'autonomous valet', 'autonomous surface vessel', 'autonomous rail', 'autonomous naval', 'autonomous racing', 'autonomous unmanned', 'connected autonomous', 'connected and autonomous', 'autonomous plane', 'autonomous fleet', 'robotaxis', 'autopilot', 'autonomous shuttle', 'autonomous flight', 'autonomous aircraft', 'flight without pilot', 'automotive is autonomous', 'autonomous material movement', 'autonomous emergency breaking', 'autonomous grocery shuttle', 'autonomous on demand trip', 'futureofdelivery', 'autonomous mobile robot', 'public transportation', 'autonomous recharging', 'autonomous capabilities', 'autonomous mobility', 'autonomous and air vehicle', 'auto industry', 'car company', 'advanced vehicle', 'driver assistance technology', 'supply chain', 'truck', 'evtruck', 'volvo', 'renault', 'mercedes', 'toyota', 'net zero', 'decarbonisation', 'transport', 'fleet', 'ride sharing', 'carbon', 'co emission', 'electromobility', 'fuel efficiency', 'battery', 'batteries', 'hydrogen', 'charging', 'samsung', 'kwh', 'mwh', 'power capacity', 'cobalt', 'lithium', 'nickel', 'mining', 'electrode', 'cell production', 'battery pack', 'environmental footprint'

Table S4. Variance Inflation Factor (VIF) scores of the independent variables

	Variable	VIF
0	const	28.80
1	Happy	1.32
2	Angry	1.18
3	Surprise	1.73
4	Sad	1.53
5	Fear	2.17
6	CLT_low	2.24
7	CLT_high	2.45
8	sentiment_score_low	1.22
9	sentiment_score_high	1.33
10	Feature_1_Safety_v2_low	1.07
11	Feature_1_Safety_v2_high	1.10
12	Feature_2_Tecnology_v2_low	1.10
13	Feature_2_Tecnology_v2_high	1.13
14	Feature_3_Traffic_v2_low	1.01
15	Feature_3_Traffic_v2_high	1.01
16	Feature_4_Employment_v2_low	1.00
17	Feature_4_Employment_v2_high	1.00
18	Feature_5_Events_v2_low	1.00
19	Feature_5_Events_v2_high	1.01
20	Feature_6_Pricing_v2_low	1.00
21	Feature_6_Pricing_v2_high	1.00
22	Feature_7_Infrastructure_v2_low	1.00
23	Feature_7_Infrastructure_v2_high	1.01
24	post_age	1.01
25	word_count	1.39
26	has_link	1.70
27	has_hashtag	1.18
28	has_mention	1.49
29	has_cashtag	1.02
30	has_emoji	1.02
31	is_verified	1.08
32	user_friendsCount	1.04
33	user_followersCount	1.08
34	user_statusesCount	1.92
35	user_mediaCount	1.89

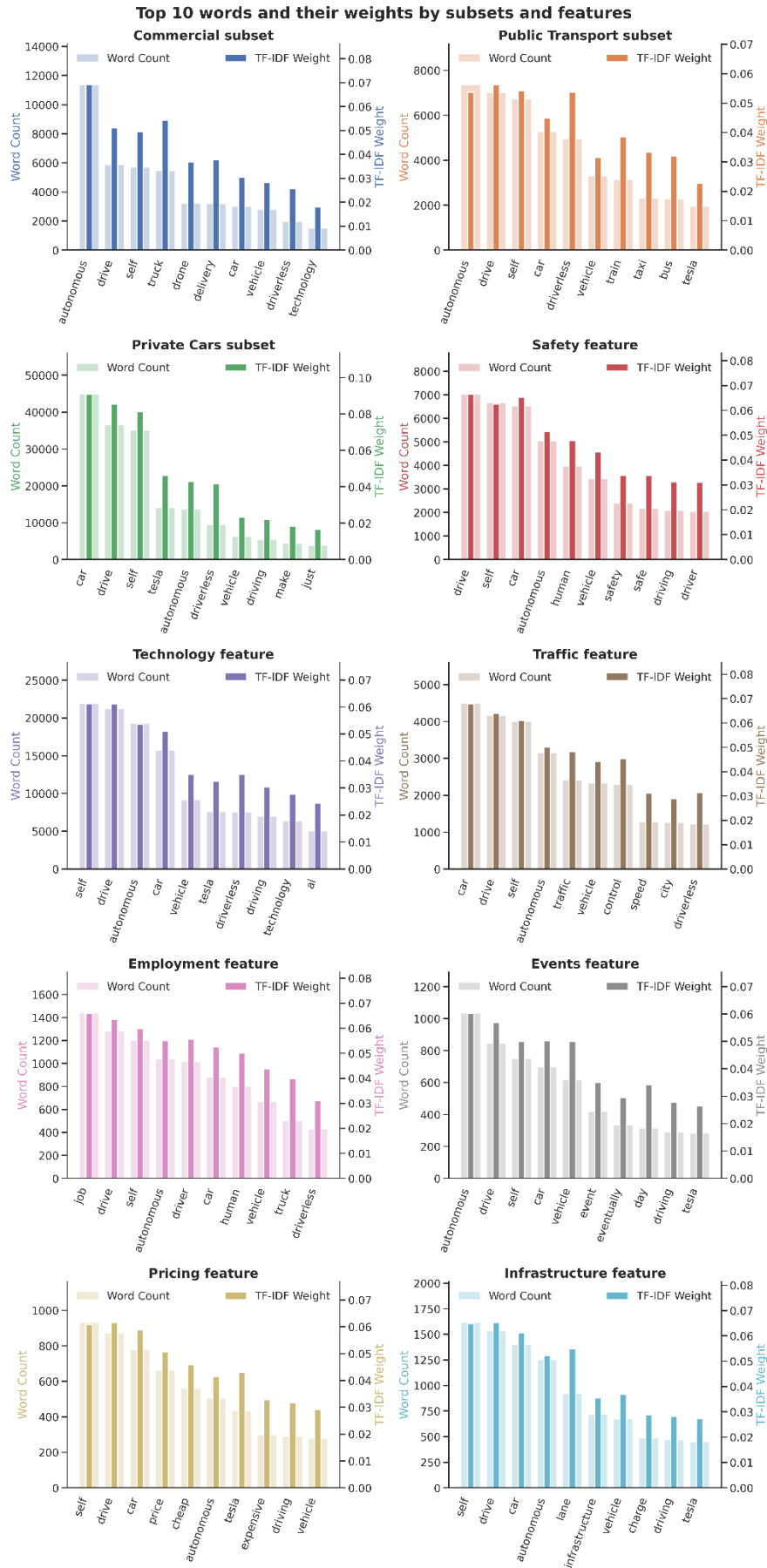


Figure S1 Most frequent words (nouns, verbs, adjectives, adverbs) by the final subsets and features

Table S5. Summary statistics of the compound sentiment and emotion scores

		Whole Dataset n= 102,174 unique user= 58,827 users with only 1 tweet= 46,057			Commercial Subset n=17,237 unique users= 11,871 users with only 1 tweet= 9,806			Public Transport Subset n=17,057 unique users= 11,970 users with only 1 tweet= 9,847			Private Cars subset n= 52,944 unique users= 35,054 users with only 1 tweet= 29,029		
	variable	Count	Mean	St. Dev	Count	Mean	St. Dev	Count	Mean	St. Dev	Count	Mean	St. Dev
Total	<i>Compound</i>												
	score	102,174	0.11	0.41	17,237	0.14	0.40	17,057	0.13	0.41	52,944	0.07	0.43
	Happy	102,174	0.05	0.14	17,237	0.06	0.15	17,057	0.06	0.15	52,944	0.05	0.12
	Angry	102,174	0.03	0.11	17,237	0.04	0.12	17,057	0.05	0.13	52,944	0.03	0.10
	Surprise	102,174	0.15	0.23	17,237	0.18	0.26	17,057	0.18	0.25	52,944	0.15	0.22
	Sad	102,174	0.11	0.19	17,237	0.11	0.19	17,057	0.11	0.19	52,944	0.11	0.18
Feature A) Safety	Fear	102,174	0.59	0.33	17,237	0.55	0.35	17,057	0.54	0.34	52,944	0.60	0.31
	<i>Compound</i>												
	score	11,827	0.12	0.54	1,685	0.19	0.53	1,723	0.15	0.55	6,727	0.05	0.55
	Happy	11,827	0.05	0.11	1,685	0.05	0.11	1,723	0.05	0.10	6,727	0.05	0.10
	Angry	11,827	0.03	0.09	1,685	0.03	0.09	1,723	0.04	0.10	6,727	0.03	0.09
	Surprise	11,827	0.15	0.19	1,685	0.20	0.24	1,723	0.16	0.18	6,727	0.16	0.19
Feature B) Technology	Sad	11,827	0.12	0.17	1,685	0.11	0.16	1,723	0.11	0.16	6,727	0.13	0.16
	Fear	11,827	0.63	0.27	1,685	0.60	0.28	1,723	0.60	0.27	6,727	0.63	0.26
	<i>Compound</i>												
	score	43,831	0.13	0.43	9,139	0.16	0.40	10,287	0.14	0.42	20,875	0.10	0.46
	Happy	43,831	0.06	0.14	9,139	0.06	0.15	10,287	0.06	0.15	20,875	0.05	0.12
	Angry	43,831	0.04	0.11	9,139	0.04	0.12	10,287	0.05	0.13	20,875	0.03	0.10
Feature C) Traffic	Surprise	43,831	0.16	0.22	9,139	0.17	0.24	10,287	0.17	0.24	20,875	0.16	0.21
	Sad	43,831	0.12	0.18	9,139	0.11	0.18	10,287	0.11	0.19	20,875	0.12	0.18
	Fear	43,831	0.59	0.31	9,139	0.58	0.33	10,287	0.55	0.33	20,875	0.60	0.29
	<i>Compound</i>												
	score	7,492	0.11	0.45	940	0.19	0.43	1,538	0.17	0.45	4,310	0.08	0.47
	Happy	7,492	0.05	0.12	940	0.06	0.13	1,538	0.06	0.12	4,310	0.06	0.12
Feature D) Employment	Angry	7,492	0.05	0.13	940	0.04	0.11	1,538	0.06	0.12	4,310	0.05	0.13
	Surprise	7,492	0.16	0.22	940	0.19	0.22	1,538	0.17	0.21	4,310	0.15	0.19
	Sad	7,492	0.12	0.18	940	0.11	0.16	1,538	0.12	0.18	4,310	0.13	0.18
	Fear	7,492	0.58	0.30	940	0.57	0.28	1,538	0.56	0.29	4,310	0.60	0.28
	<i>Compound</i>												
	score	2,445	0.13	0.46	738	0.12	0.46	563	0.13	0.47	1,177	0.13	0.46
Feature E) Events	Happy	2,445	0.05	0.11	738	0.04	0.10	563	0.07	0.12	1,177	0.05	0.12
	Angry	2,445	0.03	0.09	738	0.03	0.09	563	0.05	0.11	1,177	0.03	0.08
	Surprise	2,445	0.13	0.17	738	0.12	0.16	563	0.15	0.17	1,177	0.15	0.18
	Sad	2,445	0.12	0.17	738	0.11	0.16	563	0.12	0.17	1,177	0.12	0.18
	Fear	2,445	0.65	0.26	738	0.68	0.24	563	0.60	0.24	1,177	0.64	0.26
	<i>Compound</i>												
Feature F) Pricing	score	1,841	0.16	0.45	350	0.16	0.43	317	0.19	0.46	841	0.13	0.47
	Happy	1,841	0.11	0.22	350	0.10	0.19	317	0.09	0.17	841	0.10	0.21
	Angry	1,841	0.04	0.12	350	0.04	0.12	317	0.04	0.11	841	0.03	0.09
	Surprise	1,841	0.15	0.21	350	0.17	0.24	317	0.16	0.21	841	0.16	0.21
	Sad	1,841	0.11	0.17	350	0.12	0.18	317	0.11	0.15	841	0.12	0.17
	Fear	1,841	0.55	0.30	350	0.53	0.33	317	0.57	0.29	841	0.56	0.29
Feature G) Infrastructure	<i>Compound</i>												
	score	1,496	0.17	0.46	253	0.12	0.48	417	0.19	0.46	909	0.17	0.47
	Happy	1,496	0.06	0.11	253	0.06	0.12	417	0.06	0.13	909	0.05	0.10
	Angry	1,496	0.05	0.10	253	0.05	0.10	417	0.07	0.13	909	0.04	0.09
	Surprise	1,496	0.16	0.19	253	0.16	0.19	417	0.17	0.20	909	0.16	0.18
	Sad	1,496	0.14	0.17	253	0.16	0.17	417	0.14	0.16	909	0.15	0.16
Feature G) Infrastructure	Fear	1,496	0.58	0.26	253	0.55	0.26	417	0.55	0.27	909	0.60	0.25
	<i>Compound</i>												
	score	2,819	0.19	0.45	528	0.19	0.40	662	0.26	0.45	1,495	0.15	0.47
	Happy	2,819	0.09	0.15	528	0.09	0.14	662	0.12	0.17	1,495	0.08	0.13
	Angry	2,819	0.05	0.12	528	0.06	0.14	662	0.07	0.14	1,495	0.04	0.11
	Surprise	2,819	0.17	0.23	528	0.24	0.30	662	0.17	0.20	1,495	0.17	0.22
Feature G) Infrastructure	Sad	2,819	0.12	0.17	528	0.11	0.17	662	0.12	0.16	1,495	0.12	0.16
	Fear	2,819	0.55	0.29	528	0.46	0.32	662	0.52	0.26	1,495	0.58	0.27

Table S6. Basic statistics for the posts' engagement metrics

	Engagement attribute	Whole dataset			Commercial subset			Public transport subset			Private cars subset		
		Mean	St. Dev	Max	Mean	St. Dev	Max	Mean	St. Dev	Max	Mean	St. Dev	Max
Total	replyCount	0.87	26.04	5,984	0.54	4.18	331	0.73	9.68	1,041	1.02	32.82	5,984
	retweetCount	2.02	186.32	57,760	1.13	9.92	633	1.20	17.98	1,252	2.72	256.23	57,760
	likeCount	14.65	1,280.66	387,226	5.89	60.29	3,332	9.56	219.41	16,711	20.74	1,763.22	387,226
A)	replyCount	0.91	8.43	621	0.63	4.89	185	0.60	1.60	33	1.04	10.28	621
Safety	retweetCount	1.20	22.39	2,180	1.00	7.21	256	0.87	5.14	174	1.29	28.94	2,180
	likeCount	8.64	196.67	19,319	4.40	37.26	1,306	4.68	25.72	650	10.13	252.90	19,319
B)	replyCount	1.05	36.01	5,984	0.54	4.50	331	0.83	12.20	1,041	1.37	50.59	5,984
Tech-	retweetCount	1.66	46.94	7,275	1.23	10.64	633	1.39	20.95	1,252	1.75	42.68	4,505
nology	likeCount	12.08	516.92	97,245	6.12	65.47	3,332	10.97	244.36	16,711	15.09	706.83	97,245
C)	replyCount	1.87	72.04	5,984	1.01	6.73	185	0.78	2.62	52	2.49	92.87	5,984
Traffic	retweetCount	10.27	672.86	57,760	1.74	11.46	256	1.60	15.68	569	16.99	887.08	57,760
	likeCount	78.01	4,640.00	387,226	9.68	68.27	1,306	9.33	70.68	1,660	130.32	6,116.88	387,226
D)	replyCount	0.82	6.03	197	0.72	4.92	119	0.62	2.12	32	1.00	8.28	197
Employ-	retweetCount	0.71	9.37	445	0.50	1.95	34	0.62	2.98	36	0.83	13.22	445
ment	likeCount	4.91	57.74	2,656	3.35	19.05	371	3.57	17.25	239	6.32	80.48	2,656
E)	replyCount	0.62	2.40	46	0.35	1.06	12	0.55	1.58	19	0.83	2.70	46
Events	retweetCount	0.94	7.15	244	0.63	2.23	27	0.61	1.71	12	0.92	5.76	114
	likeCount	5.84	68.08	2,792	2.67	8.44	84	4.30	16.68	224	5.08	23.04	319
F)	replyCount	1.00	3.52	87	0.91	2.36	29	0.81	1.49	14	1.08	4.20	87
Pricing	retweetCount	0.77	10.65	390	0.41	2.51	30	0.58	4.83	92	0.96	13.53	390
	likeCount	9.15	170.47	6,522	3.40	17.55	218	4.89	26.38	368	12.47	218.23	6,522
G)	replyCount	1.02	13.05	668	0.70	4.07	84	0.78	2.90	59	1.36	17.80	668
Infra-	retweetCount	3.18	112.62	5,869	0.84	6.45	114	1.19	6.47	92	5.29	154.58	5,869
structure	likeCount	25.93	857.20	43,161	5.70	58.31	1,238	8.50	54.07	985	44.40	1,176.48	43,161

S2. Supplementary Materials for Validation

Three validation procedures were conducted to assess the reliability of the automated tools used in this study: (1) data validation and pre-processing, (2) sentiment and emotion detection, (3) topic assignment, and (4) construal-level (CLT) classification. Across these tasks, the Mistral LLM (mistralai/Mistral-7B-Instruct-v0.2) was evaluated as a potential annotator. Its performance varied substantially across tasks, and the final modelling choices reflect these empirical results.

S2.1. Data validation and pre-processing

We collected tweets posted between 1 March 2022 and 1 June 2022 using the public Python library SnsCrape on 4 June 2022. When we collected the data, only 3 months past data was allowed to be scrapped using the python library. As presented in Section 4.2, to remove off-topic tweets, we first extracted the 50 most frequent bigrams from the dataset. However, increasing this number did not improve filtering, as many irrelevant entries remained within the 171,391 tweets. To refine the dataset, we randomly sampled 1,000 tweets and manually classified them as on- or off-topic using extended domain knowledge and a keyword heuristic method (Huyen, 2022). The final set of applied keywords is available in Table S3. Although this reduced the dataset to 170,749 tweets, many unique off-topic tweets persisted. Instead of eliminating off-topic terms, we then extracted on-topic expressions from the manually labelled sample to improve accuracy. This two-step approach ultimately reduced the dataset from 198,597 to 102,174 relevant tweets (51.45%). Standard Natural Language Processing (NLP) approach for preprocessing was applied using Python's re library to remove URLs, mentions, special characters, and whitespaces, and to convert text to lowercase. Tokenisation was conducted using gensim (Rehurek and Sojka, 2010), while lemmatisation was performed using spaCy (spaCy, 2023). Topic-level sentiment was derived by parsing each tweet with spaCy, identifying predefined keywords, and extracting the corresponding dependency subtrees, on which VADER was applied to compute sentiment scores; when no keyword was present, the topic column was assigned a value of zero. The validation of the final 3 subsets and 7 features are detailed in Section S2.3 of this Supplementary File.

S2.2. Validation of Sentiment and Emotion Detection

This validation served two purposes. First, to assess how accurately state-of-the-art sentiment and emotion detection libraries classify SDV-related tweets. Second, to determine whether VADER and Text2Emotion or the NRC lexicon should be used in the study. To evaluate these tools, three independent human annotators (L1–L3) labelled a subset of 100 tweets.

As shown in Table 2 in the main study, agreement among human annotators was high:

- Sentiment polarity: $\kappa \approx 0.63$ – 0.66 ; $\alpha_{123} = 0.646$
- Text2Emotion categories: $\kappa \approx 0.70$ – 0.90 ; $\alpha_{123} = 0.736$ – 0.859

These values fall within the substantial–almost perfect range reported in the literature for short-text affect annotation (Kiritchenko et al., 2014). This confirms that the sentiment and emotion categories used by VADER and Text2Emotion are reliably identifiable by humans.

LLM Validation Attempt

A fourth annotator based on an instruction-tuned Mistral LLM was evaluated on the same 100 tweets, with the intention of using it to annotate the remaining 900 tweets. The system prompt used was:

```
SYSTEM_PROMPT = """
You are an expert annotator.
Your task is to identify which sentiments and emotions are
present in a tweet.
Rules:
1. You MUST output a single valid JSON dictionary.
2. The JSON MUST contain exactly the keys listed in the
"sentiments" list.
3. For "sentiment_cat", output ONLY one of the following
strings:
    - "positive"
    - "negative"
    - "neutral"
4. For all other sentiment and emotion keys, output ONLY 0
or 1.
5. A tweet may contain zero, one, or multiple emotions.
6. Do NOT output explanations, comments, or text before or
after the JSON.
7. Do NOT include trailing commas or additional fields.
8. Output ONLY the JSON object. Nothing else.
"""

sentiments = [
    "sentiment_cat", "Happy", "Angry", "Surprise", "Sad",
    "Fear", "fear_NRC", "anger_NRC", "trust_NRC",
    "surprise_NRC", "positive_NRC", "negative_NRC",
    "sadness_NRC", "disgust_NRC", "joy_NRC",
    "anticipation_NRC"
]

sentiment_descriptions = {
    "sentiment_cat": "Overall sentiment of the tweet:
positive, negative, or neutral.",
    "Happy": "Tweet contains happy, happiness, or joy-
related expressions.",
    "Angry": "Tweet contains anger-related expressions.",
```

```

    "Surprise": "Tweet contains surprise-related
expressions.",
    "Sad": "Tweet contains sadness-related expressions.",
    "Fear": "Tweet contains fear-related expressions.",
    "fear_NRC": "Tweet contains fear-related expressions.",
    "anger_NRC": "Tweet contains anger-related
expressions.",
    "trust_NRC": "Tweet contains trust-related
expressions.",
    "surprise_NRC": "Tweet contains surprise-related
expressions.",
    "positive_NRC": "Tweet contains positive-related
expressions.",
    "negative_NRC": "Tweet contains negative-related
expressions.",
    "sadness_NRC": "Tweet contains sadness-related
expressions.",
    "disgust_NRC": "Tweet contains disgust-related
expressions.",
    "joy_NRC": "Tweet contains joy-related expressions.",
    "anticipation_NRC": "Tweet contains anticipation-related
expressions."
}

```

Despite this structured prompt, the LLM performed extremely poorly on this task. Agreement with human annotators was very low (e.g., Krippendorff’s α_{1234} dropped to 0.238 for sentiment). The LLM frequently defaulted to neutral sentiment, under-detected positive emotions, and produced degenerate distributions for several categories. These behaviours are consistent with known limitations of LLMs in fine-grained affect detection (Herrera-Poyatos et al., 2025; Zhang et al., 2024).

Conclusion for Task 1

- VADER and Text2Emotion were retained.
- NRC was not used (lower human agreement; not aligned with modelling needs).
- LLM annotation was rejected due to poor reliability.

Table S7 Annotators (L1-L4) agreement measures in the sample of 100 tweets

category	Cohen 12	Cohen 13	Cohen 14	Cohen 23	Cohen 24	Cohen 34	Krippendorff alpha 123	Krippendorff alpha 1234
sentiment_cat	0.656	0.647	0.130	0.635	0.031	0.112	0.646	0.238
happy	0.765	0.748	-0.094	0.701	-0.152	0.030	0.736	0.287
angry	0.834	0.863	0.497	0.879	0.476	0.557	0.859	0.686
surprise	0.885	0.788	-0.182	0.900	-0.105	-0.030	0.858	0.249
sad	0.834	0.864	0.188	0.840	0.228	0.247	0.846	0.482
fear	0.836	0.811	0.099	0.773	0.135	0.110	0.806	0.312
NRC_fear	0.722	0.695	-0.105	0.600	0.100	0.015	0.673	0.343
NRC_anger	0.619	0.690	0.652	0.438	0.397	0.414	0.578	0.530
NRC_trust	0.598	0.610	0.167	0.474	0.328	0.122	0.557	0.359

NRC_surp	0.737	0.715	0.023	0.535	-0.053	0.033	0.663	0.334
NRC_pos	0.633	0.633	0.182	0.371	0.092	0.050	0.544	0.323
NRC_neg	0.521	0.606	0.386	0.384	0.371	0.253	0.502	0.352
NRC_sad	0.667	0.667	0.035	0.522	0.173	0.120	0.617	0.377
NRC_disg	0.484	0.490	0.285	0.199	0.075	0.048	0.377	0.260
NRC_joy	0.724	0.661	0.051	0.512	-0.115	0.049	0.630	0.321
NRC_anticip	0.590	0.558	0.122	0.465	0.118	0.093	0.533	0.305

Table S8 Library accuracy (human annotators: 1-3) on 100 tweet sample

Lexicon	Category	Accuracy (TP+TN/N)
VADER	Total (positive, neutral and negative)	85%
Text2Emotion	Happy	87%
	Angry	87%
	Surprise	86%
	Sad	82%
	Fear	79%
NRC	anger	83%
	anticipation	80%
	disgust	88%
	fear	66%
	joy	74%
	negative	78%
	positive	67%
	sadness	77%
	surprise	56%
	trust	80%

S2.3. Validation of Topic Assignment

Topic assignment in this study was performed using a keyword-based heuristic, as described in Methodology Section 4.3 and Section S2.1. Tweets were mapped to predefined thematic categories relevant to the SDV discourse (e.g., safety, technology, traffic, employment, events, pricing, infrastructure). These categories were operationalised through explicit lexical cues derived from domain knowledge and exploratory corpus analysis using LDA topic modelling and bigram extraction. Because the rules were deterministic and transparent, topic assignment did not require human annotation for initial labelling.

To validate the heuristic, the Mistral LLM (mistralai/Mistral-7B-Instruct-v0.2) was asked to independently assign topics to a stratified random sample of 996 tweets, ensuring that approximately 100 tweets from each of the ten topic categories were included. The LLM was prompted to identify which topics were present in each tweet using the following instruction:

```
SYSTEM_PROMPT = """
You are an expert annotator.
Your task is to identify which of the following topics
are present in a tweet.
A tweet may contain zero, one, or multiple topics.
```

For each topic, decide if it is present (1) or not present (0) in the tweet, based on the topic name and its description.

Return your answer strictly as a JSON dictionary with topic names as keys and 0 or 1 as values.

Do not explain your reasoning.

"""

```
topic_descriptions = {
    "Commercial_subset": "Tweets referring to commercial
or industrial autonomous vehicles: trucks, lorries,
HGVs, freight, delivery fleets, long-haul transport,
drones, cargo, logistics, or supply chain
applications.",
    "Public_Transport_subset": "Tweets discussing public
or shared autonomous mobility: buses, trains, robotaxis,
shuttles, transit systems, stations, smart city
mobility, or mass passenger transport.",
    "Private_Cars_subset": "Tweets referring to personal
or consumer autonomous vehicles: cars, SUVs, Tesla,
Polestar, private ownership, or the automotive
industry.",
    "safety_subset_v2": "Tweets discussing the safety of
self-driving vehicles, including accidents, collisions,
near-misses, human harm, pedestrian risk, system
failures, recalls, or general concerns about whether
autonomous vehicles are safe.",
    "technology_subset_v2": "Tweets referring to the
technical aspects of autonomous vehicles: AI, software,
automation, engineering, sensors, beta features,
autopilot systems, robotics, or technological
innovation.",
    "traffic_subset_v2": "Tweets discussing traffic
flow, congestion, speed, parking, city movement, or road
control in relation to autonomous vehicles.",
    "employment_subset_v2": "Tweets about jobs and
labour related to autonomous vehicles: driver shortages,
job losses, new careers, human vs automated work, or
employment impacts on drivers, engineers, or related
professions.",
    "events_subset_v2": "Tweets referring to events,
expos, shows, demonstrations, festivals, marketing
events, or public showcases involving autonomous
vehicles.",
    "pricing_subset_v2": "Tweets discussing cost,
pricing, affordability, value for money,
```

```

expensive/cheap, or financial considerations of
autonomous vehicles.",
    "infrastructure_subset_v2": "Tweets about charging
stations, lanes, road infrastructure, superchargers, or
physical systems required to support autonomous
vehicles."
}

```

Agreement between the LLM and the keyword-based heuristic was moderate to high, with accuracy ranging from 0.723 to 0.932 across categories (Table S9). Average agreement across the ten topics was 0.843. High agreement was observed for concrete, lexically well-defined topics, such as, Pricing (0.932), Events (0.903), Employment (0.891). Lower agreement occurred for semantically diffuse or context-dependent topics, such as, Technology (0.723) and Private Cars (0.776).

In these cases, keyword rules tended to capture literal lexical matches, whereas the LLM relied more heavily on contextual interpretation. Confusion matrices indicated that disagreements were primarily due to false negatives for broad topics and false positives for context-dependent topics.

Topic-Level Results

Table S9 Subset and Topic-level validation results

Topic	TP	FP	FN	TN	Accuracy
Commercial	72	19	181	724	0.799
Public Transport	86	3	172	735	0.824
Private Cars	100	9	224	663	0.776
Safety	133	72	95	696	0.832
Technology	197	94	182	523	0.723
Traffic	49	15	125	807	0.859
Employment	65	50	59	822	0.891
Events	31	15	82	868	0.903
Pricing	69	23	45	859	0.932
Infrastructure	43	22	88	843	0.890

The LLM demonstrated:

- strong sensitivity to domain-specific terminology
- consistent topic boundaries
- no systematic misclassification patterns
- high alignment with the intended semantic structure of the topic taxonomy

This stands in sharp contrast to the LLM's poor performance on sentiment and emotion detection, indicating that topic classification is a task LLMs handle reliably, particularly when categories are semantically distinct and lexically grounded.

The LLM validation confirmed that the keyword heuristic captured the intended semantic distinctions. Human annotators were therefore deemed unnecessary, as manual annotation would not meaningfully improve reliability and could introduce unnecessary subjectivity.

Conclusion for Task 2

- Keyword-based topic assignment was validated and retained.
- LLM validation confirmed high accuracy.
- Human annotation was not required.

Additional Robustness: Segment Overlap Analysis

Because cross-segment comparison is central to the manuscript’s contribution, we further examined the overlap between the three transport market segments (commercial, public, private). The segments were not defined as mutually exclusive, since tweets may reference multiple contexts simultaneously. Tweets containing keywords from more than one segment were included in all relevant categories, enabling segment-wise sentiment analysis even when users expressed contrasting views across contexts within a single post.

Table S10 Tweet classification across transport segments

Segment	Total Tweets	Overlap with Commercial	Overlap with Public Transport	Overlap with Private Cars	Overlap with All Three	Exclusive to Segment	% Exclusive
Commercial	17,237	–	2,988 (17.3%)	4,784 (27.8%)	886 (5.1%)	10,351	60.0%
Public Transport	17,057	2,988 (17.5%)	–	6,529 (38.3%)	886 (5.2%)	8,426	49.4%
Private Cars	52,944	4,784 (9.0%)	6,529 (12.3%)	–	886 (1.7%)	42,517	80.3%
Unclassified	28,351	–	–	–	–	–	–
Total Dataset	102,174						

Table S10 reports the distribution of overlaps, exclusive tweets, and unclassified tweets. For example, 17.3% of commercial tweets also referenced public transport, 27.8% referenced private cars, and 5.1% referenced all three segments, leaving 60% of commercial tweets exclusive to that segment. Similar proportions were observed for public transport (49% exclusive) and private cars (80% exclusive). In addition, 28,351 tweets (27.7% of the dataset) did not match any segment keywords; these were retained for overall sentiment and emotion analysis but excluded from segment-wise comparisons.

This overlap analysis demonstrates that while categories are not mutually exclusive, the majority of tweets in each segment are exclusive, and results are robust to stricter definitions.

S2.4. Validation of Construal-Level Theory (CLT) Classification

Tweets were classified into low-CLT, high-CLT, or none using the Mistral LLM (mistralai/Mistral-7B-Instruct-v0.2) in Colab Pro with a G4 GPU. Classification was performed in batches of eight tweets per request, requiring approximately 12 minutes to process 10,000 tweets. Unlike sentiment and emotion detection, CLT classification is a conceptual, interpretive task involving psychological abstraction levels — a domain where LLMs typically perform well due to their ability to reason over linguistic structure and contextual cues.

The LLM was guided with the following structured system prompt:

```
CLT_SYSTEM_PROMPT = """
You are an expert in Construal Level Theory (CLT).
```

Your task is to classify a tweet as LOW-level construal, HIGH-level construal, or NONE.

Operational definitions:

- LOW-level construal = concrete, specific, detailed, action-focused, describing immediate events, near in time/space/social distance.
- HIGH-level construal = abstract, general, conceptual, value-focused, long-term, hypothetical, or distant in time/space/social distance.
- NONE = the tweet does not contain enough meaningful content to determine a construal level.

Decision rules:

1. If the tweet is ambiguous between NONE and LOW/HIGH → choose LOW or HIGH (do NOT output NONE).
2. If the tweet is ambiguous between LOW and HIGH → choose LOW.
3. If the tweet mixes abstract and concrete elements → choose LOW.
4. If the tweet is vague but mentions future implications, values, or societal impact → choose HIGH.
5. Output MUST be exactly one of: "low", "high", "none".
6. NEVER output explanations, parentheses, qualifiers, or multiple labels.

Examples:

LOW examples:

- "The self-driving car hit a lamp post on my street today."
- "I tried the new robotaxi yesterday and the seatbelt was broken."
- "Traffic is crazy right now because an AV blocked the junction."

HIGH examples:

- "Autonomous vehicles will reshape society in the next decade."
- "We need to think about the ethical implications of AI-driven transport."
- "Driverless technology represents a major shift in human mobility."

NONE examples:

- "Check this out!"
- "Lol what is this"
- "Buy now!"

Return ONLY a JSON dictionary:

```

{"CLT": "low"} or {"CLT": "high"} or {"CLT": "none"}.
Do not explain your reasoning.
"""

full_prompt = f"""
<s>[INST] <<SYS>>
{CLT_SYSTEM_PROMPT}
<</SYS>>

Tweet: "{tweet}"
Classify the construal level.
[/INST]
"""

```

LLM Performance

The LLM produced coherent and consistent CLT assignments across the dataset. To validate these outputs, three domain experts independently reviewed a stratified sample of 99 tweets (33 per CLT category). Experts confirmed the validity of the LLM’s low/high CLT distinctions, observed no systematic misclassification patterns, and judged that the model reliably captured abstract versus concrete framing in line with CLT theory. Agreement statistics are shown in Table S11.

Table S11 Annotators' agreement summary on CLT classification

statistic	CLT_level agreement
cohen_L1_L2	0.86
cohen_L1_L3	0.89
cohen_L2_L3	0.88
krippendorff_alpha_humans	0.88
cohen_L1_L4	0.94
cohen_L2_L4	0.89
cohen_L3_L4	0.92
krippendorff_alpha_all	0.9

As observed, agreement between the experts and the LLM was high. This validation demonstrates that the LLM was able to generalise the conceptual distinctions underlying CLT and apply them consistently to real-world SDV discourse.

CLT Distribution Across Topics

Table S12 summarises the proportion of low, high, and none categories across the subsets and topics. As observed, all subsets and topics dominated by high-level construal discussions (~60–73%), followed by low-level (~19–31%), and the smallest proportion in each were none (~5–14%).

Table S12 CLT distributions across subsets and topics

subset or topic feature	Dominant CLT	Low proportion	High proportion	None proportion
Commercial subset	high	0.26	0.66	0.08
Public Transport subset	high	0.3	0.61	0.09
Private Cars subset	high	0.32	0.54	0.14
Safety	high	0.21	0.73	0.06
Technology	high	0.27	0.62	0.11

Traffic	high	0.28	0.67	0.05
Employment	high	0.21	0.7	0.09
Events	high	0.19	0.68	0.13
Pricing	high	0.27	0.67	0.06
Infrastructure	high	0.28	0.66	0.06

This indicates that discourse around SDVs overwhelmingly adopts a high-level construal: SDVs are framed as futuristic, abstract, societal, long-term, and conceptual. This finding reinforced the need for our modelling framework to incorporate how engagement differs between low- and high-level CLT tweets.

Conclusion for Task 3

- LLM-based CLT classification was validated and retained.
- Human experts confirmed the LLM’s accuracy.
- Regression models were designed to directly measure engagement effects in low- and high-CLT contexts.

S3. Supplementary Materials for Robustness Check

S3.1. Robustness Check using different engagement types as dependent variables

To assess the robustness of our results, we re-estimated all models using each engagement type—like_count, retweet_count, and reply_count—as the dependent variable. Across all specifications, the Negative Binomial model consistently outperformed alternative count models based on AIC scores (Table S13).

The results for total engagement (sum of likes, retweets, and replies), which are presented in the main manuscript, are reproduced in Table S14. These are then compared with the individual engagement models in Tables S15–S17.

Finally, Tables S18 and S19 summarise the positive and negative effects, respectively, ranked by effect strength across all model specifications.

Across the three engagement types, the direction and significance of the key predictors remained broadly consistent, indicating that the observed relationships are not driven by any single engagement metric. However, the magnitude of effects varied in theoretically meaningful ways: replies tended to be more sensitive to negative and high-arousal emotions, retweets were more strongly associated with abstract (high-CLT) framing and informational topics, and likes reflected low-effort affective approval. These patterns align with prior research on differentiated engagement behaviours and further support the robustness of our findings.

Table S13 Summary results of models per engagement type (Poisson and Negative Binomial Regression models)

Dataset	Model #	Model	Dependent Variable	Num Observations	AIC	LR Statistic	LR df	LR p-value
Whole	1.0	Poisson	replyCount	102,174	481,577	212548.27	35	0.000
Whole	1.0	Negative Binomial	replyCount	102,174	210,374	54121.46	35	0.000
Commercial	1.1	Poisson	replyCount	17,237	46,863	8847.31	35	0.000
Commercial	1.1	Negative Binomial	replyCount	17,237	29,338	5257.57	35	0.000
Public Transport	1.2	Poisson	replyCount	17,057	52,072	24204.65	35	0.000

Public Transport	1.2	Negative Binomial	replyCount	17,057	32,634	7566.88	35	0.000
Private Cars	1.3	Poisson	replyCount	52,944	274,090	154204.83	35	0.000
Private Cars	1.3	Negative Binomial	replyCount	52,944	115,772	32352.11	35	0.000
Whole	2.0	Poisson	retweetCount	102,174	2,019,591	467250.20	35	0.000
Whole	2.0	Negative Binomial	retweetCount	102,174	308,977	82642.61	35	0.000
Commercial	2.1	Poisson	retweetCount	17,237	100,887	25980.65	35	0.000
Commercial	2.1	Negative Binomial	retweetCount	17,237	40,807	9990.81	35	0.000
Public Transport	2.2	Poisson	retweetCount	17,057	116,721	39064.66	35	0.000
Public Transport	2.2	Negative Binomial	retweetCount	17,057	39,390	12365.77	35	0.000
Private Cars	2.3	Poisson	retweetCount	52,944	1,466,994	463552.29	35	0.000
Private Cars	2.3	Negative Binomial	retweetCount	52,944	172,622	56625.81	35	0.000
Whole	3.0	Poisson	likeCount	102,174	13,340,780	4763426.16	35	0.000
Whole	3.0	Negative Binomial	likeCount	102,174	609,277	150516.43	35	0.000
Commercial	3.1	Poisson	likeCount	17,237	540,233	124142.22	35	0.000
Commercial	3.1	Negative Binomial	likeCount	17,237	84,372	14052.55	35	0.000
Public Transport	3.2	Poisson	likeCount	17,057	1,016,336	392913.30	35	0.000
Public Transport	3.2	Negative Binomial	likeCount	17,057	96,558	16380.33	35	0.000
Private Cars	3.3	Poisson	likeCount	52,944	10,050,910	4275632.04	35	0.000
Private Cars	3.3	Negative Binomial	likeCount	52,944	339,342	90171.93	35	0.000
Whole	4.0	Poisson	total_eng_count	102,174	15,513,786	5350814.66	35	0.000
Whole	4.0	Negative Binomial	total_eng_count	102,174	648,184	147290.55	35	0.000
Commercial	4.1	Poisson	total_eng_count	17,237	644,220	150102.48	35	0.000
Commercial	4.1	Negative Binomial	total_eng_count	17,237	92,485	13971.35	35	0.000
Public Transport	4.2	Poisson	total_eng_count	17,057	1,134,247	443964.51	35	0.000
Public Transport	4.2	Negative Binomial	total_eng_count	17,057	104,313	14605.81	35	0.000
Private Cars	4.3	Poisson	total_eng_count	52,944	11,634,284	4800804.38	35	0.000
Private Cars	4.3	Negative Binomial	total_eng_count	52,944	738,696	-292027.55	35	1
Private Cars	4.3r	Negative Binomial	total_eng_count	52,944	354,667	91999.83	34	0.000

Table S14 Results for the NB Regression model—as reported in the main study—for Total Engagement Counts (sum of replies, retweets, and likes)

target variable: total_eng_count (replyCount + retweetCount + likeCount)										
*** p < 0.001, ** p < 0.01, * p < 0.05										
	Whole Dataset (4.0)		Commercial (4.1)		Public Transport (4.2)		Private Cars (4.3) Non-significant		Private Cars (4.3r) Robustness Check	
	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.
(intercept)	1.41***	0.019	1.03***	0.050	0.55***	0.051	2.24***	0.025	0.74***	0.027
CLT low	-0.16***	0.011	0.37***	0.036	0.24***	0.035	0.10***	0.014	0.00	0.016
CLT high	0.14***	0.011	0.12***	0.034	1.26***	0.033	0.98***	0.014	0.38***	0.015
Sentiment score x CLT low	-0.19***	0.018	-0.24***	0.052	0.64***	0.046	0.53***	0.022	-0.34***	0.023
Sentiment score x CLT high	0.17***	0.011	0.30***	0.028	0.37***	0.027	-0.97***	0.014	0.20***	0.015
Happy emotion detection score	1.79***	0.027	-0.66***	0.066	-0.34***	0.067	13.83***	0.041	3.66***	0.043
Angry emotion detection score	0.80***	0.032	-0.40***	0.076	-0.36***	0.075	6.18***	0.046	1.99***	0.048
Surprise emotion detection score	-0.87***	0.020	-0.96***	0.044	-0.58***	0.047	-2.69***	0.028	-0.90***	0.032
Sad emotion detection score	-0.73***	0.023	-0.86***	0.055	-0.35***	0.057	-1.64***	0.031	-0.39***	0.035
Fear emotion detection score	0.18***	0.016	-0.91***	0.036	0.73***	0.038	2.31***	0.021	0.65***	0.025
Safety sentiment x CLT low	0.05	0.061	0.33	0.176	-0.21	0.137	1.15***	0.074	0.73***	0.079
Safety sentiment x CLT high	0.10**	0.033	-0.71***	0.093	-0.33***	0.084	-2.65***	0.042	0.33***	0.044
Technology sentiment x CLT low	0.14**	0.053	0.90***	0.154	0.45***	0.124	-0.88***	0.063	0.40***	0.068
Technology sentiment x CLT high	0.23***	0.029	-0.77***	0.069	-0.81***	0.065	4.57***	0.037	0.62***	0.039
Traffic sentiment x CLT low	-0.93***	0.182	0.42	0.575	1.73**	0.526	0.41	0.213	-1.15***	0.229
Traffic sentiment x CLT high	0.23*	0.099	-0.30	0.273	1.71***	0.218	0.24	0.133	0.40**	0.138
Employment sentiment x CLT low	-0.38	0.293	1.54**	0.578	0.45	0.824	-2.73***	0.398	-1.57***	0.468
Employment sentiment x CLT high	0.09	0.173	0.71*	0.299	1.11**	0.337	-2.39***	0.239	-1.10***	0.248
Events sentiment x CLT low	-2.27***	0.296	-2.97***	0.602	-4.40***	0.976	-6.31***	0.394	-1.82***	0.419
Events sentiment x CLT high	-0.52***	0.138	-1.66***	0.355	-0.58	0.317	4.18***	0.209	0.07	0.215
Pricing sentiment x CLT low	0.26	0.304	-6.97**	2.136	1.15	0.852	-3.63***	0.347	0.01	0.361
Pricing sentiment x CLT high	-1.77***	0.203	-0.91*	0.436	0.32	0.378	-3.33***	0.257	-1.51***	0.266
Infrastructure sentiment x CLT low	-0.20	0.275	-2.36**	0.802	2.42***	0.650	0.65*	0.324	1.03**	0.337
Infrastructure sentiment x CLT high	-1.52***	0.134	3.18***	0.310	-0.60*	0.238	-3.01***	0.184	-0.72***	0.191
Post age	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
Word count	0.03***	0.000	0.03***	0.001	0.00	0.001	0.00***	0.000	0.02***	0.000
Tweet has link(s)	-0.07***	0.009	0.06**	0.023	-0.24***	0.023	2.32***	0.012	0.08***	0.011
Tweet has hashtag(s)	-0.84***	0.009	-0.27***	0.021	-0.54***	0.022	-2.48***	0.012	-1.08***	0.013
Tweet has mention(s)	-0.44***	0.008	0.20***	0.020	0.19***	0.021	0.98***	0.011		

<u>target variable: total_eng_count (replyCount + retweetCount + likeCount)</u> *** p < 0.001, ** p < 0.01, * p < 0.05										
	Whole Dataset (4.0)		Commercial (4.1)		Public Transport (4.2)		Private Cars (4.3) Non-significant		Private Cars (4.3r) Robustness Check	
	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.
Tweet has cashtag(s)	0.48***	0.016	0.74***	0.036	0.96***	0.041	4.23***	0.025	0.75***	0.026
Tweet has emoji(s)	1.17***	0.009	0.21***	0.023	1.50***	0.021	4.78***	0.011	1.38***	0.012
User is verified	1.98***	0.015	2.19***	0.035	2.11***	0.034	3.06***	0.022	1.90***	0.022
User friends Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User followers Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User statuses Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User media Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
AIC	648,184		92,485		104,314		738,696		354,667	

Table S15 Results for the NB Regression model for reply counts as dependant variable.

target variable: replyCount								
*** p < 0.001, ** p < 0.01, * p < 0.05								
	Whole Dataset (1.0)		Commercial (1.1)		Public Transport (1.2)		Private Cars (1.3)	
	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.
(intercept)	-0.95***	0.031	-0.98***	0.088	-1.37***	0.091	-1.27***	0.050
CLT low	-0.08***	0.019	0.25***	0.066	0.07	0.063	-0.01	0.025
CLT high	0.12***	0.018	0.25***	0.062	0.31***	0.060	0.10***	0.024
Sentiment score x CLT low	-0.06*	0.028	-0.28***	0.083	0.03	0.072	-0.01	0.034
Sentiment score x CLT high	-0.17***	0.016	-0.18***	0.043	0.01	0.040	-0.12***	0.021
Happy emotion detection score	0.25***	0.045	-0.39***	0.114	-0.06	0.116	0.88***	0.069
Angry emotion detection score	0.14**	0.053	-0.42**	0.133	-0.03	0.127	0.94***	0.076
Surprise emotion detection score	-0.37***	0.034	-0.78***	0.081	0.09	0.083	-0.04	0.054
Sad emotion detection score	-0.32***	0.038	-0.66***	0.095	0.20*	0.096	0.03	0.058
Fear emotion detection score	0.03	0.026	-0.59***	0.064	-0.03	0.071	0.46***	0.044
Safety sentiment x CLT low	0.10	0.093	0.48	0.279	0.13	0.220	0.29*	0.117
Safety sentiment x CLT high	0.14**	0.047	-0.30*	0.144	0.14	0.127	0.09	0.062
Technology sentiment x CLT low	0.20*	0.081	0.52*	0.236	0.94***	0.185	0.38***	0.101
Technology sentiment x CLT high	-0.22***	0.043	-0.26*	0.108	-0.34***	0.097	-0.36***	0.056
Traffic sentiment x CLT low	0.08	0.275	0.82	0.886	-0.36	0.807	-0.40	0.336
Traffic sentiment x CLT high	-0.07	0.147	0.38	0.412	1.01**	0.311	0.02	0.195
Employment sentiment x CLT low	-0.05	0.458	1.20	0.941	0.42	1.227	-1.51	0.773
Employment sentiment x CLT high	0.56*	0.242	1.00*	0.446	0.53	0.487	0.52	0.340
Events sentiment x CLT low	-1.65***	0.461	-1.85	0.980	-3.19*	1.291	-1.25	0.657
Events sentiment x CLT high	-0.20	0.205	-1.06	0.578	-0.39	0.534	0.14	0.297
Pricing sentiment x CLT low	1.51***	0.401	-3.07	2.858	-0.32	1.344	1.39**	0.451
Pricing sentiment x CLT high	-0.31	0.278	0.38	0.579	0.33	0.530	-0.84	0.364
Infrastructure sentiment x CLT low	0.17	0.400	-3.44*	1.561	0.85	0.875	0.29	0.477
Infrastructure sentiment x CLT high	-0.17	0.192	1.91***	0.392	-0.58	0.381	0.48	0.263
Post age	0.00***	0.000	0.00***	0.000	0.00	0.000	0.00*	0.000
Word count	0.02***	0.000	0.03***	0.001	0.03***	0.001	0.02***	0.001
Tweet has link(s)	-0.36***	0.014	-0.78***	0.037	-0.64***	0.036	-0.36***	0.019
Tweet has hashtag(s)	-0.88***	0.017	-0.45***	0.040	-0.96***	0.043	-0.94***	0.024
Tweet has mention(s)	-0.02	0.013	0.02	0.033	-0.15***	0.033	-0.24***	0.017
Tweet has cashtag(s)	0.61***	0.023	0.55***	0.057	0.77***	0.057	0.86***	0.033
Tweet has emoji(s)	0.21***	0.013	0.30***	0.036	0.15***	0.033	0.27***	0.017
User is verified	2.22***	0.018	2.03***	0.045	1.83***	0.044	2.39***	0.026
User friends Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User followers Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User statuses Count	0.00***	0.000	0.00*	0.000	0.00***	0.000	0.00***	0.000
User media Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
AIC	210,374		29,338		32,635		115,772	

Table S16 Results for the NB Regression model for retweet counts as dependant variable

target variable: retweetCount								
*** p < 0.001, ** p < 0.01, * p < 0.05								
	Whole Dataset (2.0)		Commercial (2.1)		Public Transport (2.2)		Private Cars (2.3)	
	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.
(intercept)	-0.98***	0.025	-1.89***	0.073	-2.56***	0.084	0.09*	0.038
CLT low	-0.38***	0.016	0.42***	0.054	0.17**	0.060	0.02	0.021
CLT high	-0.25***	0.015	0.26***	0.052	0.91***	0.056	0.79***	0.020
Sentiment score x CLT low	-0.37***	0.025	-0.43***	0.074	0.57***	0.078	-0.52***	0.032
Sentiment score x CLT high	0.00	0.014	0.27***	0.040	0.34***	0.039	-0.08***	0.020
Happy emotion detection score	1.71***	0.032	-0.52***	0.091	-0.52***	0.103	4.56***	0.051
Angry emotion detection score	1.25***	0.042	-0.27**	0.102	-0.67***	0.122	1.48***	0.057
Surprise emotion detection score	-0.70***	0.028	-0.70***	0.060	-0.54***	0.072	-1.43***	0.043
Sad emotion detection score	-1.01***	0.033	-0.67***	0.076	-0.47***	0.090	-1.32***	0.048
Fear emotion detection score	-0.06**	0.021	-0.80***	0.049	0.15**	0.059	0.43***	0.032
Safety sentiment x CLT low	0.03	0.087	0.40	0.235	-1.01***	0.217	0.80***	0.109
Safety sentiment x CLT high	-0.05	0.041	-0.77***	0.130	-0.32**	0.116	0.28***	0.057
Technology sentiment x CLT low	-0.24**	0.075	-0.16	0.213	0.78***	0.203	-0.12	0.089
Technology sentiment x CLT high	0.31***	0.036	-0.48***	0.096	-0.76***	0.092	0.93***	0.050
Traffic sentiment x CLT low	0.87***	0.255	0.45	0.856	0.57	0.962	-0.34	0.297
Traffic sentiment x CLT high	1.57***	0.123	-0.31	0.391	1.96***	0.284	1.24***	0.175
Employment sentiment x CLT low	-0.18	0.409	0.99	0.837	-0.11	1.535	-2.28**	0.761
Employment sentiment x CLT high	-0.15	0.211	0.29	0.410	-0.40	0.467	-1.73***	0.329
Events sentiment x CLT low	-1.71***	0.423	-1.64*	0.797	-5.87***	1.296	-3.40***	0.599
Events sentiment x CLT high	-0.72***	0.167	-1.81***	0.494	-0.72	0.441	0.05	0.276
Pricing sentiment x CLT low	-0.54	0.469	-3.68	2.690	-5.35*	2.640	-0.61	0.506
Pricing sentiment x CLT high	-1.46***	0.261	-0.86	0.607	0.47	0.571	-1.21***	0.358
Infrastructure sentiment x CLT low	-0.23	0.386	-1.57	1.214	2.42**	0.896	-0.09	0.448
Infrastructure sentiment x CLT high	-1.35***	0.178	3.49***	0.355	-0.79*	0.345	-0.93***	0.262
Post age	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
Word count	0.04***	0.000	0.03***	0.001	0.03***	0.001	0.00	0.001
Tweet has link(s)	0.80***	0.012	0.99***	0.035	1.03***	0.034	-0.45***	0.016
Tweet has hashtag(s)	-0.23***	0.012	0.21***	0.028	-0.12***	0.030	-0.48***	0.017
Tweet has mention(s)	-0.38***	0.011	0.11***	0.028	-0.14***	0.030	-1.83***	0.016
Tweet has cashtag(s)	0.52***	0.020	0.28***	0.050	0.89***	0.050	0.77***	0.030
Tweet has emoji(s)	1.00***	0.010	0.43***	0.031	0.48***	0.030	1.92***	0.015
User is verified	2.40***	0.016	2.03***	0.039	2.07***	0.038	2.17***	0.024
User friends Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User followers Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User statuses Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User media Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
AIC	308,977		40,808		39,391		172,622	

Table S17 Results for the NB Regression model for like counts as dependant variable

target variable: likeCount *** p < 0.001, ** p < 0.01, * p < 0.05								
	Whole Dataset (3.0)		Commercial (3.1)		Public Transport (3.2)		Private Cars (3.3)	
	coef.	std. err.	coef.	std. err.	coef.	std. err.	coef.	std. err.
(intercept)	1.09***	0.019	0.85***	0.051	-0.04	0.054	1.04***	0.028
CLT low	-0.17***	0.012	0.39***	0.037	0.11**	0.037	-0.17***	0.015
CLT high	0.05***	0.011	0.09*	0.035	0.63***	0.036	0.16***	0.015
Sentiment score x CLT low	-0.18***	0.018	-0.23***	0.053	0.67***	0.050	-0.24***	0.023
Sentiment score x CLT high	0.20***	0.012	0.35***	0.029	0.81***	0.028	-0.22***	0.015
Happy emotion detection score	1.59***	0.028	-0.67***	0.068	-0.50***	0.070	4.70***	0.042
Angry emotion detection score	1.03***	0.033	-0.36***	0.078	-0.44***	0.079	2.31***	0.047
Surprise emotion detection score	-0.83***	0.021	-0.98***	0.046	-0.56***	0.050	-1.15***	0.031
Sad emotion detection score	-0.70***	0.024	-0.85***	0.056	-0.34***	0.060	-0.40***	0.034
Fear emotion detection score	0.28***	0.016	-0.94***	0.037	0.00	0.041	0.53***	0.024
Safety sentiment x CLT low	0.03	0.062	0.30	0.180	-0.60***	0.149	1.04***	0.079
Safety sentiment x CLT high	-0.03	0.033	-0.77***	0.096	-0.96***	0.086	-0.64***	0.044
Technology sentiment x CLT low	0.15**	0.054	0.99***	0.157	0.45***	0.135	0.19**	0.067
Technology sentiment x CLT high	0.24***	0.029	-0.91***	0.071	-1.53***	0.066	2.23***	0.039
Traffic sentiment x CLT low	-1.15***	0.188	0.20	0.585	1.42*	0.569	0.20	0.226
Traffic sentiment x CLT high	0.24*	0.101	-0.48	0.281	0.75***	0.224	0.22	0.137
Employment sentiment x CLT low	-0.26	0.298	1.30*	0.588	-0.29	0.899	-1.50***	0.447
Employment sentiment x CLT high	-0.13	0.175	0.65*	0.306	1.07**	0.342	-1.82***	0.249
Events sentiment x CLT low	-2.25***	0.308	-3.28***	0.639	-5.85***	1.024	-2.47***	0.420
Events sentiment x CLT high	-0.43**	0.141	-1.65***	0.367	-0.84**	0.325	0.71***	0.213
Pricing sentiment x CLT low	0.12	0.311	-6.35**	2.128	-0.22	0.880	-1.28***	0.367
Pricing sentiment x CLT high	-1.72***	0.205	-1.02*	0.444	-0.74	0.379	-1.42***	0.265
Infrastructure sentiment x CLT low	-0.16	0.281	-2.16**	0.812	1.74*	0.683	1.22***	0.336
Infrastructure sentiment x CLT high	-1.54***	0.137	3.36***	0.313	-1.10***	0.241	-1.18***	0.194
Post age	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
Word count	0.03***	0.000	0.03***	0.001	0.05***	0.001	0.03***	0.000
Tweet has link(s)	0.08***	0.009	0.02	0.023	0.40***	0.023	0.18***	0.012
Tweet has hashtag(s)	-0.98***	0.010	-0.36***	0.022	-0.89***	0.023	-1.24***	0.013
Tweet has mention(s)	-0.36***	0.008	0.17***	0.021	-0.54***	0.021	-0.62***	0.012
Tweet has cashtag(s)	0.53***	0.017	0.84***	0.036	0.67***	0.041	1.12***	0.025
Tweet has emoji(s)	1.01***	0.009	0.18***	0.024	0.65***	0.021	1.82***	0.012
User is verified	1.92***	0.015	2.22***	0.035	2.07***	0.035	1.64***	0.022
User friends Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User followers Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
User statuses Count	0.00***	0.000	0.00***	0.000	0.00*	0.000	0.00***	0.000
User media Count	0.00***	0.000	0.00***	0.000	0.00***	0.000	0.00***	0.000
AIC	609,278		84,373		96,559		339,343	

Table S18 Positive coefficient variables ranked by importance across engagement types and datasets.

Total engagement (positive coefficients by strength)			
Whole (4.0)	Comm. (4.1)	Public T. (4.2)	Private (where 4.3 & 4.3r agree)
User is verified	Infrastructure sentiment x CLThigh	Infrastructure sentiment x CLTlow	Happy emotion detection score
Happy emotion detection score	User is verified	User is verified	Angry emotion detection score
Tweet has emoji(s)	Employment sentiment x CLTlow	Traffic sentiment x CLTlow	User is verified
Angry emotion detection score	Technology sentiment x CLTlow	Traffic sentiment x CLThigh	Tweet has emoji(s)
Tweet has cashtag(s)	Tweet has cashtag(s)	Tweet has emoji(s)	Infrastructure sentiment x CLTlow

Technology sentiment x CLThigh	Employment sentiment x CLThigh	CLThigh	Tweet has cashtag(s)
Traffic sentiment x CLThigh	CLTlow	Employment sentiment x CLThigh	Safety sentiment x CLTlow
Fear emotion detection score	Sentiment score x CLThigh	Tweet has cashtag(s)	Fear emotion detection score
Sentiment score x CLThigh	Tweet has emoji(s)	Fear emotion detection score	Technology sentiment x CLThigh
CLThigh	Tweet has mention(s)	Sentiment score x CLTlow	CLThigh
Technology sentiment x CLTlow	CLThigh	Technology sentiment x CLTlow	Tweet has link(s)
Safety sentiment x CLThigh		Sentiment score x CLThigh	
		CLTlow	
		Tweet has mention(s)	
replyCount (positive coefficients by strength)			
Whole (1.0)	Comm. (1.1)	Public T. (1.2)	Private (1.3)
User is verified	User is verified	User is verified	User is verified
Pricing sentiment x CLTlow	Infrastructure sentiment x CLThigh	Traffic sentiment x CLThigh	Pricing sentiment x CLTlow
Tweet has cashtag(s)	Employment sentiment x CLThigh	Technology sentiment x CLTlow	Angry emotion detection score
Employment sentiment x CLThigh	Tweet has cashtag(s)	Tweet has cashtag(s)	Happy emotion detection score
Happy emotion detection score	Technology sentiment x CLTlow	CLThigh	Tweet has cashtag(s)
Tweet has emoji(s)	Tweet has emoji(s)	Sad emotion detection score	Fear emotion detection score
Technology sentiment x CLTlow	CLTlow	Tweet has emoji(s)	Technology sentiment x CLTlow
Safety sentiment x CLThigh	CLThigh		Safety sentiment x CLTlow
Angry emotion detection score			Tweet has emoji(s)
CLThigh			CLThigh
retweetCount (positive coefficients by strength)			
Whole (2.0)	Comm. (2.1)	Public T. (2.2)	Private (2.3)
User is verified	Infrastructure sentiment x CLThigh	Infrastructure sentiment x CLTlow	Happy emotion detection score
Happy emotion detection score	User is verified	User is verified	User is verified
Traffic sentiment x CLThigh	Tweet has link(s)	Traffic sentiment x CLThigh	Tweet has emoji(s)
Angry emotion detection score	Tweet has emoji(s)	Tweet has link(s)	Angry emotion detection score
Tweet has emoji(s)	CLTlow	CLThigh	Traffic sentiment x CLThigh
Traffic sentiment x CLTlow	Tweet has cashtag(s)	Tweet has cashtag(s)	Technology sentiment x CLThigh
Tweet has link(s)	Sentiment score x CLThigh	Technology sentiment x CLTlow	Safety sentiment x CLTlow
Tweet has cashtag(s)	CLThigh	Sentiment score x CLTlow	CLThigh
Technology sentiment x CLThigh	Tweet has hashtag(s)	Tweet has emoji(s)	Tweet has cashtag(s)
	Tweet has mention(s)	Sentiment score x CLThigh	Fear emotion detection score
		CLTlow	Safety sentiment x CLThigh
		Fear emotion detection score	
likeCount (positive coefficients by strength)			
Whole (3.0)	Comm. (3.1)	Public T. (3.2)	Private (3.3)
User is verified	Infrastructure sentiment x CLThigh	User is verified	Happy emotion detection score
Happy emotion detection score	User is verified	Infrastructure sentiment x CLTlow	Angry emotion detection score
Angry emotion detection score	Employment sentiment x CLTlow	Traffic sentiment x CLTlow	Technology sentiment x CLThigh
Tweet has emoji(s)	Technology sentiment x CLTlow	Employment sentiment x CLThigh	Tweet has emoji(s)
Tweet has cashtag(s)	Tweet has cashtag(s)	Sentiment score x CLThigh	User is verified
Fear emotion detection score	Employment sentiment x CLThigh	Traffic sentiment x CLThigh	Infrastructure sentiment x CLTlow
Technology sentiment x CLThigh	CLTlow	Sentiment score x CLTlow	Tweet has cashtag(s)
Traffic sentiment x CLThigh	Sentiment score x CLThigh	Tweet has cashtag(s)	Safety sentiment x CLTlow
Sentiment score x CLThigh	Tweet has emoji(s)	Tweet has mention(s)	Events sentiment x CLThigh
Technology sentiment x CLTlow	Tweet has mention(s)	CLThigh	Fear emotion detection score
CLThigh	CLThigh	Technology sentiment x CLTlow	Technology sentiment x CLTlow
		Tweet has link(s)	Tweet has link(s)
		CLTlow	CLThigh

Table S19 Negative coefficient variables ranked by importance across engagement types and datasets.

Total engagement (negative coefficients by strength)			
Whole (4.0)	Comm. (4.1)	Public T. (4.2)	Private (where 4.3 & 4.3r agree)
Events sentiment x CLTlow	Pricing sentiment x CLTlow	Events sentiment x CLTlow	Events sentiment x CLTlow
Pricing sentiment x CLThigh	Events sentiment x CLTlow	Technology sentiment x CLThigh	Employment sentiment x CLTlow
Infrastructure sentiment x CLThigh	Infrastructure sentiment x CLTlow	Infrastructure sentiment x CLThigh	Pricing sentiment x CLThigh
Traffic sentiment x CLTlow	Surprise emotion detection score	Surprise emotion detection score	Employment sentiment x CLThigh
Surprise emotion detection score	Fear emotion detection score	Tweet has hashtag(s)	Tweet has hashtag(s)

Tweet has hashtag(s)	Pricing sentiment x CLThigh	Angry emotion detection score	Surprise emotion detection score
Sad emotion detection score	Sad emotion detection score	Sad emotion detection score	Infrastructure sentiment x CLThigh
Events sentiment x CLThigh	Technology sentiment x CLThigh	Happy emotion detection score	Sad emotion detection score
Tweet has mention(s)	Safety sentiment x CLThigh	Safety sentiment x CLThigh	
Sentiment score x CLTlow	Happy emotion detection score	Tweet has link(s)	
CLTlow	Angry emotion detection score		
	Tweet has hashtag(s)		
	Sentiment score x CLTlow		
replyCount (negative coefficients by strength)			
Whole (1.0)	Comm. (1.1)	Public T. (1.2)	Private (1.3)
Events sentiment x CLTlow	Infrastructure sentiment x CLTlow	Events sentiment x CLTlow	Tweet has hashtag(s)
Tweet has hashtag(s)	Surprise emotion detection score	Tweet has hashtag(s)	Technology sentiment x CLThigh
Surprise emotion detection score	Tweet has link(s)	Tweet has link(s)	Tweet has link(s)
Tweet has link(s)	Sad emotion detection score	Technology sentiment x CLThigh	Tweet has mention(s)
Sad emotion detection score	Fear emotion detection score	Tweet has mention(s)	Sentiment score x CLThigh
Technology sentiment x CLThigh	Tweet has hashtag(s)		
Sentiment score x CLThigh	Angry emotion detection score		
CLTlow	Happy emotion detection score		
Sentiment score x CLTlow	Safety sentiment x CLThigh		
	Sentiment score x CLTlow		
	Technology sentiment x CLThigh		
	Sentiment score x CLThigh		
retweetCount (negative coefficients by strength)			
Whole (2.0)	Comm. (2.1)	Public T. (2.2)	Private (2.3)
Events sentiment x CLTlow	Events sentiment x CLThigh	Events sentiment x CLTlow	Events sentiment x CLTlow
Pricing sentiment x CLThigh	Events sentiment x CLTlow	Pricing sentiment x CLTlow	Employment sentiment x CLTlow
Infrastructure sentiment x CLThigh	Fear emotion detection score	Safety sentiment x CLTlow	Tweet has mention(s)
Sad emotion detection score	Safety sentiment x CLThigh	Infrastructure sentiment x CLThigh	Employment sentiment x CLThigh
Events sentiment x CLThigh	Surprise emotion detection score	Technology sentiment x CLThigh	Surprise emotion detection score
Surprise emotion detection score	Sad emotion detection score	Angry emotion detection score	Sad emotion detection score
CLTlow	Happy emotion detection score	Surprise emotion detection score	Pricing sentiment x CLThigh
			Infrastructure sentiment x CLThigh
Tweet has mention(s)	Technology sentiment x CLThigh	Happy emotion detection score	high
Sentiment score x CLTlow	Sentiment score x CLTlow	Sad emotion detection score	Sentiment score x CLTlow
CLThigh	Angry emotion detection score	Safety sentiment x CLThigh	Tweet has hashtag(s)
Technology sentiment x CLTlow		Tweet has mention(s)	Tweet has link(s)
Tweet has hashtag(s)		Tweet has hashtag(s)	Sentiment score x CLThigh
Fear emotion detection score			
likeCount (negative coefficients by strength)			
Whole (3.0)	Comm. (3.1)	Public T. (3.2)	Private (3.3)
Events sentiment x CLTlow	Pricing sentiment x CLTlow	Events sentiment x CLTlow	Events sentiment x CLTlow
Pricing sentiment x CLThigh	Events sentiment x CLTlow	Technology sentiment x CLThigh	Employment sentiment x CLThigh
Infrastructure sentiment x CLThigh	Infrastructure sentiment x CLTlow	Infrastructure sentiment x CLThigh	Employment sentiment x CLTlow
Traffic sentiment x CLTlow	Events sentiment x CLThigh	Safety sentiment x CLThigh	Pricing sentiment x CLThigh
Tweet has hashtag(s)	Pricing sentiment x CLThigh	Tweet has hashtag(s)	Pricing sentiment x CLTlow
Surprise emotion detection score	Surprise emotion detection score	Events sentiment x CLThigh	Tweet has hashtag(s)
			Infrastructure sentiment x CLThigh
Sad emotion detection score	Fear emotion detection score	Safety sentiment x CLTlow	high
Events sentiment x CLThigh	Technology sentiment x CLThigh	Surprise emotion detection score	Surprise emotion detection score
Tweet has mention(s)	Sad emotion detection score	Tweet has mention(s)	Safety sentiment x CLThigh
Sentiment score x CLTlow	Safety sentiment x CLThigh	Happy emotion detection score	Tweet has mention(s)
CLTlow	Happy emotion detection score	Angry emotion detection score	Sad emotion detection score
	Angry emotion detection score	Sad emotion detection score	Sentiment score x CLTlow
	Tweet has hashtag(s)		Sentiment score x CLThigh
	Sentiment score x CLTlow		CLTlow

S3.2. Robustness Check using Predictive Models

To assess the robustness of our results, we conducted predictive modelling using the dataset and evaluated the performance of various models. Specifically, we employed Poisson regression, XGBoost, Random Forest, and Artificial Neural Networks (ANN). The rationale behind selecting these models is detailed in Subsection S3.2.2. Model training was performed in R, utilising the base package along with the following libraries: xgboost, ranger, and neuralnet. The dataset was randomly split into a 70:30 ratio for training and testing, respectively. To ensure replicability, we set a random seed (`set.seed = 123`). Each model was trained using the same set of variables and covariates as in the primary explanatory model. This procedure was applied to both the full sample and the three predefined subsets of the dataset. For each of these four cases (overall sample and three subgroups), we identified the best-performing model. When the top-performing model was a machine learning algorithm, we further analysed the relative importance of the independent variables and generated Partial Dependence Plots (PDPs). PDPs are interpretability tools commonly used in machine learning to visualise the marginal effect of a given feature (or pair of features) on the predicted outcome, averaging over the influence of all other features. They help answer the question: What is the impact on the predicted outcome when we change a single feature, keeping all others constant?

S3.2.1. Results of robustness check with predictive modelling

As shown in Table S20, XGBoost had the lowest prediction error among the four models tested. The relative importance curve (Figure S2) based on XGBoost indicates that happy is the most predictive emotion, and traffic sentiment stands out among topics. Partial dependence plots (PDPs) for key emotions and topics (Figures S3-S14) reveal that engagement decreases with higher levels of happy and fear, while anger shows a strong positive relationship, and sadness a weaker one. Surprise has a mixed, non-linear effect.

Among topics, safety shows an inverted U-shaped relationship, with excessive positivity reducing engagement. Technology sentiment is positively associated with engagement, while traffic sentiment displays a U-shaped pattern. Extremely positive sentiment on employment and infrastructure reduces engagement by approximately 150 and 200 units, respectively, while high feature sentiment increases it by about 45 units. Pricing sentiment shows no significant effect, with one potential outlier spike.

PDPs were not generated for the category-specific samples, as Poisson regression performed best in those cases (see Tables S21–S23), making the PDPs redundant with previously discussed regression results.

Table S20 Results of predictive model for overall sample

	Model Name	Out-sample RMSE	Out-sample MAE
1	Poisson Regression	700.285	30.134
2	Random Forest	727.621	29.723
3	XGBoost	713.648	28.324
4	Artificial Neural Network (ANN)	704.847	64.074

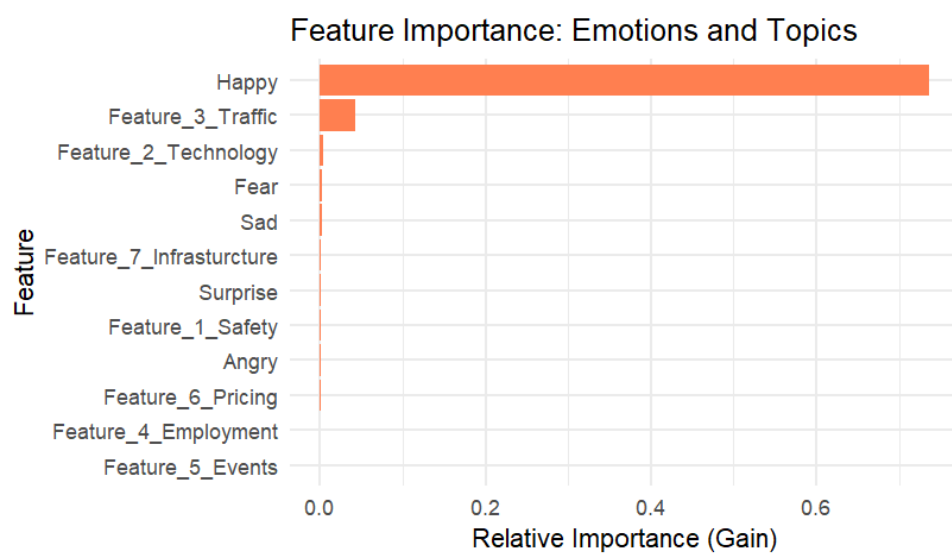


Figure S2 Feature importance for emotions and topics for XGBoost model in overall sample

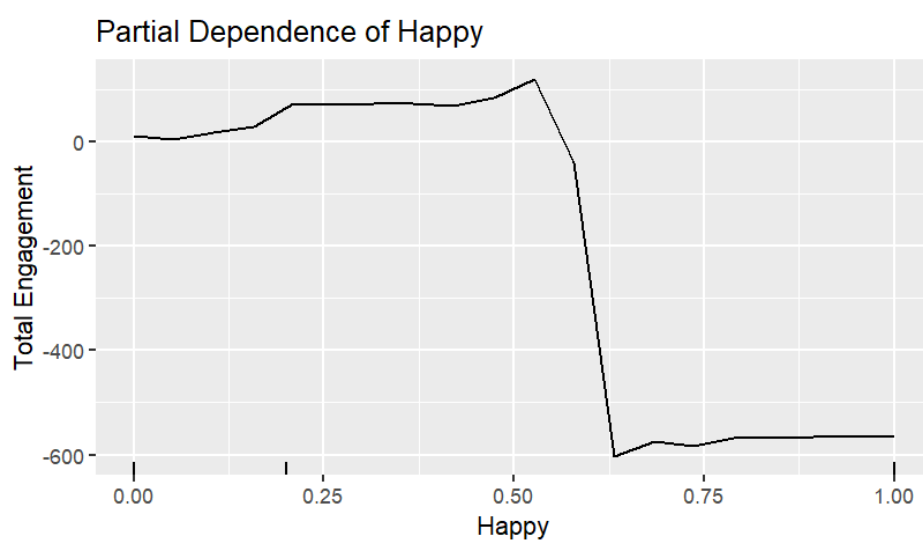


Figure S3 Partial dependence of Happy emotion (overall sample)

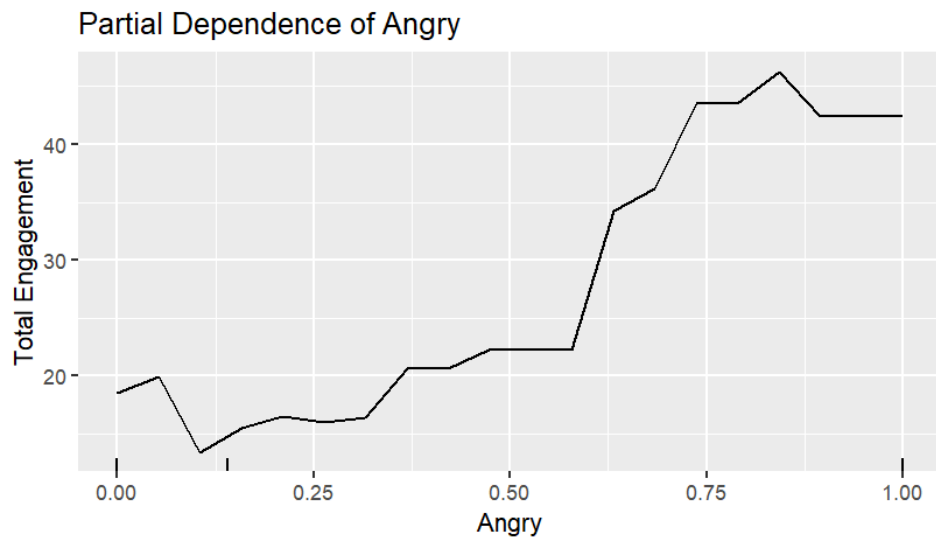


Figure S4 Partial Dependence of Angry emotion (overall sample)

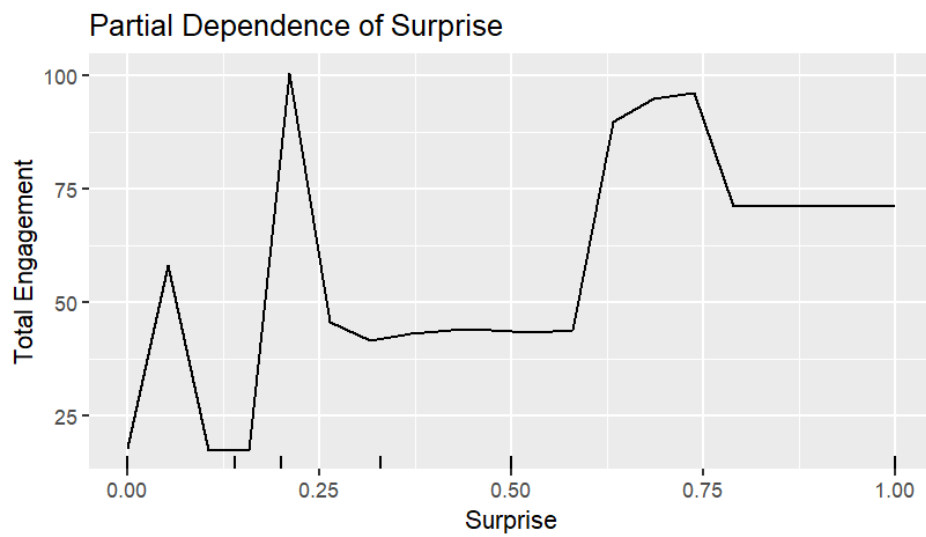


Figure S5 Partial Dependence of Surprise emotion (overall sample)

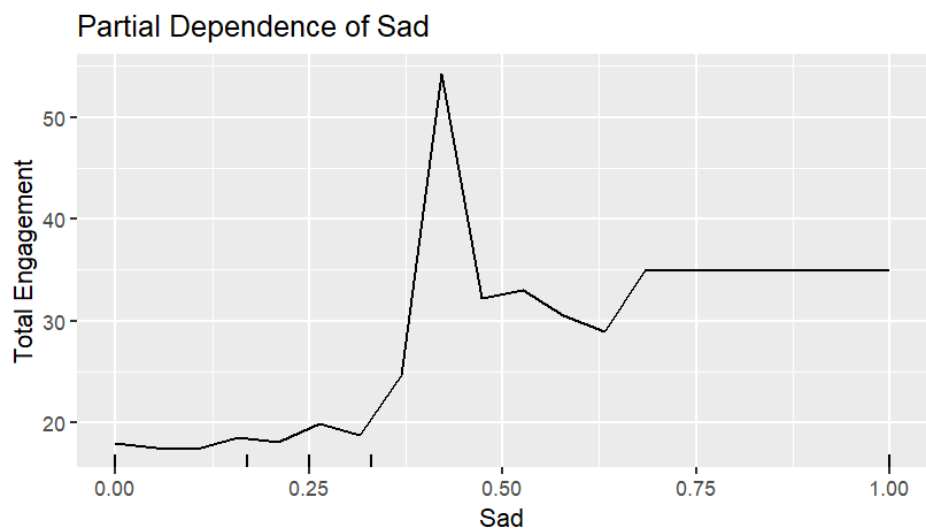


Figure S6 Partial Dependence of Sad emotion (overall sample)

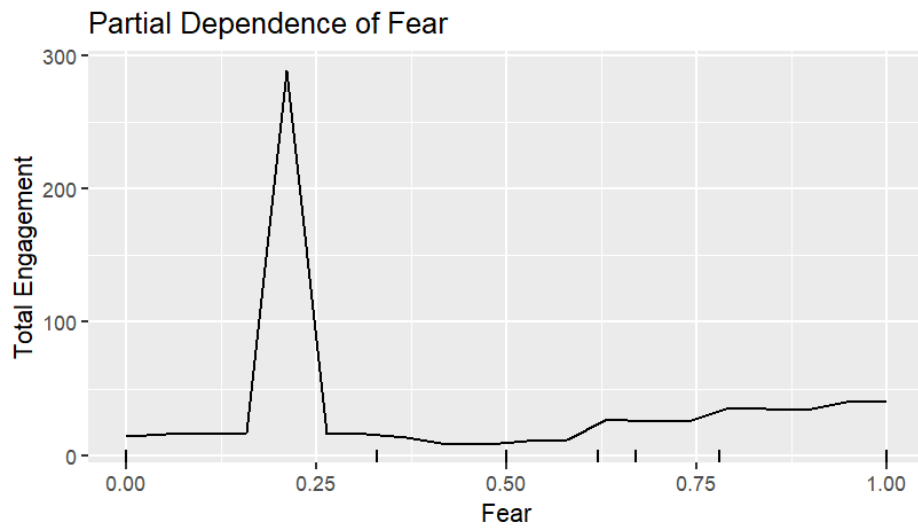


Figure S7 Partial Dependence of Fear emotion (overall sample)



Figure S8 Partial Dependence of Safety Feature (overall sample)

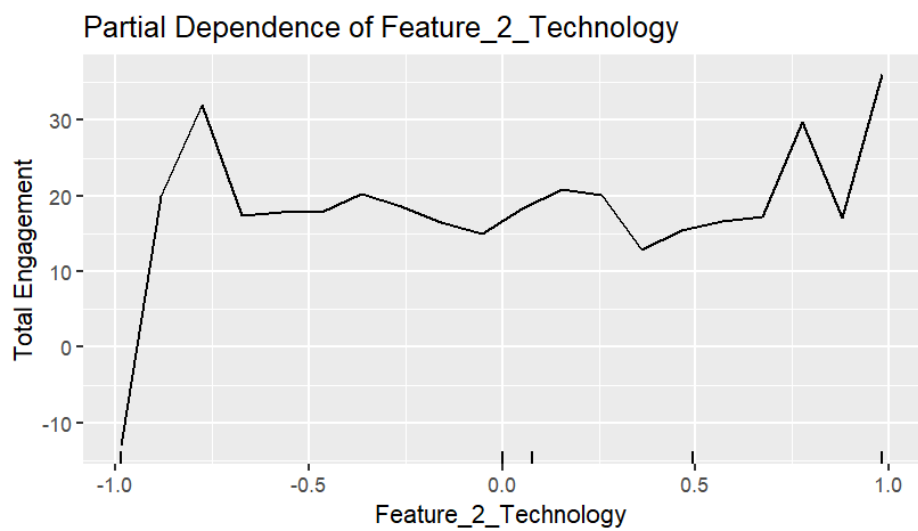


Figure S9 Partial Dependence of Technology Feature (overall sample)

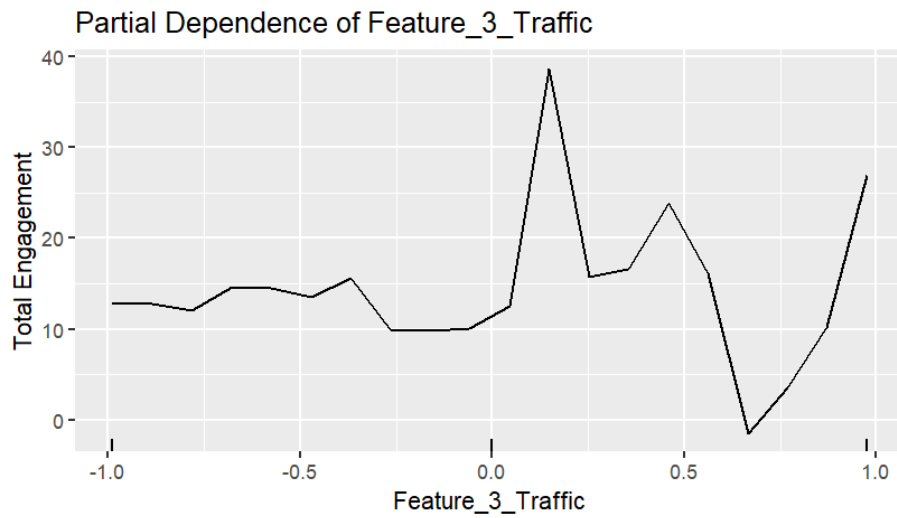


Figure S10 Partial Dependence of Traffic Feature (overall sample)



Figure S11 Partial Dependence of Employment Feature (overall sample)

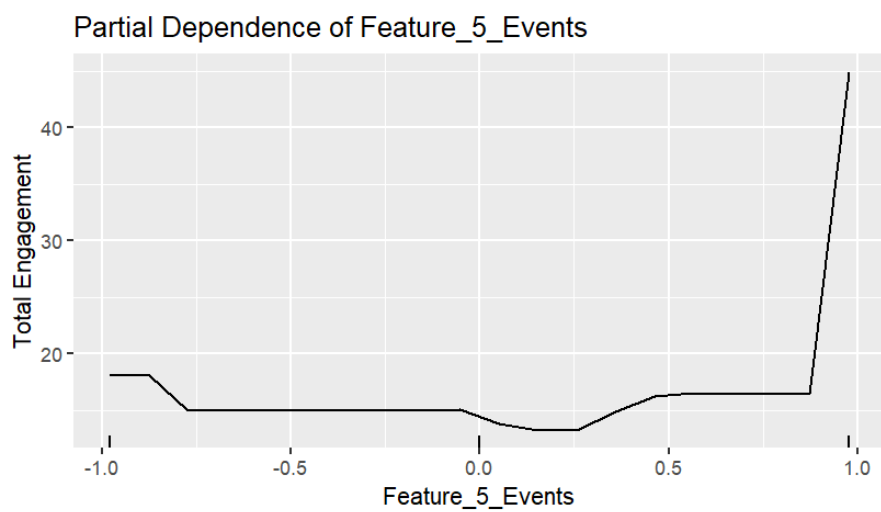


Figure S12 Partial Dependence of Events Feature (overall sample)



Figure S13 Partial Dependence of Pricing Feature (overall sample)

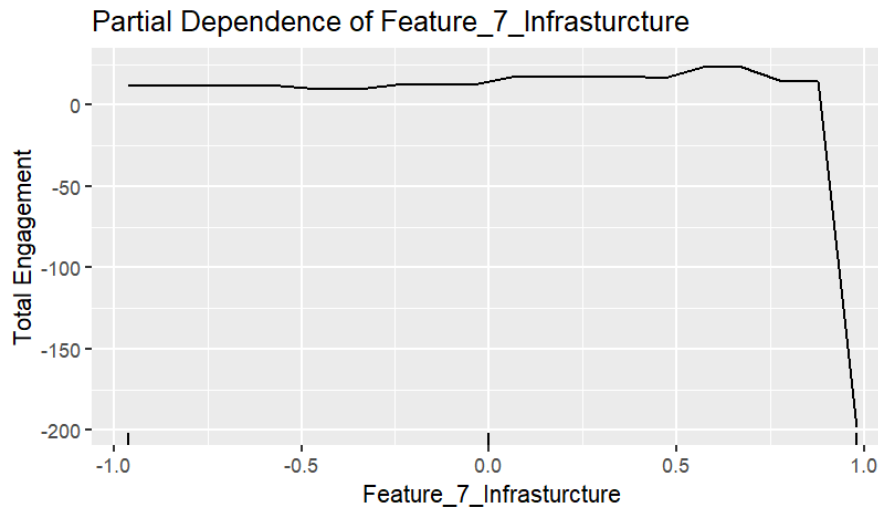


Figure S14 Partial Dependence of Infrastructure Feature (overall sample)

Table S21 Results of predictive model for Commercial AV sample

	Model Name	Out-sample RMSE	Out-sample MAE
1	Poisson Regression	77.295	12.559
2	Random Forest	78.675	13.373
3	XGBoost	87.609	15.273
4	ANN	77.922	14.042

Table S22 Results of predictive model for AV Private Cars sample

	Model Name	Out-sample RMSE	Out-sample MAE
1	Poisson Regression	462.628	39.468
2	Random Forest	737.205	46.271
3	XGBoost	7016.076	143.137
4	ANN	508.076	126.621

Table S23 Results of predictive model for AV Public Transport sample

	Model Name	Out-sample RMSE	Out-sample MAE
1	Poisson Regression	266.837	18.229
2	Random Forest	269.804	19.337
3	XGBoost	286.746	21.526
4	ANN	269.564	32.996

S3.2.2. Reasoning for Choice of Predictive Models

To comprehensively evaluate the predictive performance of various modelling approaches for public engagement with self-driving vehicle content, we employed four distinct algorithms: Poisson regression, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN). These models were purposefully chosen to represent a balance between interpretability, robustness, and predictive power, each offering unique advantages aligned with the structure and nature of our data.

Poisson regression was selected due to the count nature of the dependent variable—total engagement. It serves as a statistical baseline and provides interpretable coefficients, particularly helpful for understanding directional effects and testing theoretical frameworks such as Construal Level Theory (CLT). However, Poisson models assume equidispersion (equal mean and variance) and linear relationships, which may limit their predictive power in complex, non-linear settings (Xia, 2023).

Random Forest was chosen for its ability to handle non-linearity and variable interactions without strong parametric assumptions. It is robust to overfitting and naturally performs feature selection, making it suitable for moderately high-dimensional data with potential multicollinearity among sentiment variables (Fife and D’Onofrio, 2023).

XGBoost, a boosting-based ensemble method, was included for its state-of-the-art performance in structured tabular data. It iteratively reduces bias and variance, often

outperforming traditional machine learning models. Its built-in feature importance metrics and compatibility with interpretation tools such as partial dependence plots made it ideal for uncovering complex relationships in the data (Antar and Tayachi, 2025).

Artificial Neural Networks (ANN) were incorporated to assess the capacity of deep learning to model complex, non-linear functions, particularly in the presence of emotional and semantic content features. While less interpretable, ANN offers high flexibility and has shown success in many natural language and sentiment-based prediction tasks (Montesinos López et al., 2022).

Together, these models allowed us to assess trade-offs between interpretability and accuracy, providing a well-rounded evaluation of engagement prediction mechanisms.

References

- Antar, M., Tayachi, T., 2025. Partial dependence analysis of financial ratios in predicting company defaults: random forest vs XGBoost models. *Digit Finance*.
<https://doi.org/10.1007/s42521-025-00135-6>
- Fife, D.A., D’Onofrio, J., 2023. Common, uncommon, and novel applications of random forest in psychological research. *Behav Res* 55, 2447–2466.
<https://doi.org/10.3758/s13428-022-01901-9>
- Herrera-Poyatos, D., Peláez-González, C., Zuheros, C., Herrera-Poyatos, A., Tejedor, V., Herrera, F., Montes, R., 2025. An overview of model uncertainty and variability in LLM-based sentiment analysis: challenges, mitigation strategies, and the role of explainability. *Front. Artif. Intell.* 8. <https://doi.org/10.3389/frai.2025.1609097>
- Huyen, C., 2022. *Designing Machine Learning Systems, An Iterative Process for Production-Ready Applications.*, 1st edition. ed. O’Reilly Media,.
- Kiritchenko, S., Zhu, X., Mohammad, S.M., 2014. Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research* 50, 723–762.
<https://doi.org/10.1613/jair.4272>
- Montesinos López, O.A., Montesinos López, A., Crossa, Jose, 2022. Fundamentals of Artificial Neural Networks and Deep Learning, in: Montesinos López, O.A., Montesinos López, A., Crossa, José (Eds.), *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer International Publishing, Cham, pp. 379–425. https://doi.org/10.1007/978-3-030-89010-0_10
- Rehurek, R., Sojka, P., 2010. *gensim: Python framework for fast Vector Space Modelling*.
- spaCy, 2023. *spaCy · Industrial-strength Natural Language Processing in Python [WWW Document]*. URL <https://spacy.io/> (accessed 8.21.23).
- Xia, F., 2023. Why to use Poisson regression for count data analysis in consumer behavior research. *J Market Anal* 11, 379–384. <https://doi.org/10.1057/s41270-022-00166-7>
- Zhang, W., Deng, Y., Liu, B., Pan, S., Bing, L., 2024. Sentiment Analysis in the Era of Large Language Models: A Reality Check, in: Duh, K., Gomez, H., Bethard, S. (Eds.), *Findings of the Association for Computational Linguistics: NAACL 2024*. Presented at the Findings 2024, Association for Computational Linguistics, Mexico City, Mexico, pp. 3881–3906. <https://doi.org/10.18653/v1/2024.findings-naacl.246>