



Affective, cognitive, and contextual cues in Reddit posts on artificial intelligence

Nina Savela¹ · Max Pellert^{2,3,4} · Rita Latikka^{5,6} · Jenna Bergdahl⁵ · David Garcia^{3,4,7} · Atte Oksanen⁵

Received: 19 March 2024 / Accepted: 28 September 2024
© The Author(s) 2024

Abstract

Artificially intelligent technologies have become a common topic in our everyday discussions where arguments about the subject can take different forms from cognitive reasoning to emotional expressions. Utilizing persuasion theories and research on the appeal of content characteristics as the theoretical approach to examine affective–cognitive language, we investigated social media posts on artificial intelligence (AI). We examined Reddit posts from 2005 to 2018 referring to AI ($N=455,634$) using automated content analysis tools. The results revealed that although both the tone positivity and affective–cognitive ratio were dependent on the specific context, the language in AI posts was more analytically than emotionally oriented in general. Other users were more likely to engage with Reddit posts on AI that were high in cognitive and analytic content compared to affective and emotional content. In addition to the practical contribution of public opinion on AI, the results contribute to the theoretical discussions on affective and cognitive language in social media discussions.

Keywords Artificial intelligence · Social media · Sentiment analysis · Contextual dictionaries · Affect · Cognition

✉ Nina Savela
nina.savela@lut.fi

¹ LUT University, Yliopistonkatu 34, Lappeenranta 53850, Finland

² University of Mannheim, Baden-Württemberg, Germany

³ Medical University of Vienna, Wien, Austria

⁴ Complexity Science Hub Vienna, Wien, Austria

⁵ Tampere University, Tampere, Finland

⁶ Demos Helsinki, Helsinki, Finland

⁷ University of Konstanz, Konstanz, Germany

Introduction

Artificially intelligent technologies have become a common topic in our everyday life discussions — even in life domains where artificial intelligence (AI) has been relatively excluded from the public eye until recent years, such as its use in art and urban environments. Novel technologies in new contexts induce public discussions and arguments that can take different forms from cognitive reasoning to emotional expressions. Research has demonstrated that the cognitive–emotional language ratio differs in public discussions depending on the topic [1], highlighting the importance of context-dependency. In social media persuasion research, both emotional cues and rational appeals in the original message have been found to influence the decisions of other users to engage with it [2–3]. Thus, the affective–cognitive orientation of messages in social media can be a crucial factor in how socially engaging the conversation starters are and how societally impactful the discussions around a certain topic will eventually be.

As a rapidly evolving field and timely discussion topic, defining AI and its components and boundaries is a complex task that is unlikely to arrive at a unified opinion any time soon [4]. A broad dictionary definition states that AI is a field involving computer systems that have the capacity to execute tasks traditionally associated with human intelligence [4]. In addition to the ununified professional definitions, AI is likely to be used and understood in versatile ways in public discussions. When investigating language in larger currents of public discussions where individual meanings cannot be inquired from the subjects or clarified by the researchers, we are approaching the target construct from the perspective of symbolic or sociocognitive representations. These representations are collectively shared meanings and ways to use a certain concept for the sake of communication [5–7].

We utilized emotional–analytic persuasion appeals [1–2, 8–9] as a theoretical framework to examine affective–cognitive content characteristics of the language used in social media posts on AI. As a social media data set, we used posts submitted to the Reddit platform during 2005–2018 that referred to AI. The automated content analyses were performed with Linguistic Inquiry and Word Counts (LIWC) and Valence Aware Dictionary and sEntiment Reasoner (VADER) sentiment analysis tools. First, we considered the context-dependency by comparing the tone of AI posts with different contextual cues. Second, we compared the affective and emotional language with cognitive and analytic language in the posts and examined how they were associated with other users' engagement with the original messages. Our study is the first to investigate social influence and persuasion of affective- versus cognitive-oriented social media posts on the topic of AI.

AI in various contexts and life domains

Because AI applications and perceptions of them are often domain-specific, it is advisable to consider context-dependency [10]. A report by a study panel consisting of experts on AI examined the past and future development of AI over 15 years within domains that have an effect on urban life [11]. The domains they considered the most salient perspectives to focus on consisted of, for example, transportation,

service, health care, education, public safety, and entertainment [11]. When anticipating the effect AI has on individuals and societies, creating markets (e.g., music, literature, and journalism) are highlighted in addition to service-providing markets (e.g., medicine, transportation, and education) [12].

The insights from researchers are in line with marketing analyses. The domains that have been estimated to have the most AI applications and rapid growth in usage included sectors such as advertising and media (including gaming and entertainment), finance, health care, transportation, and others (e.g., education, security, and surveillance) [13]. The marketing research for the World Economic Forum surveyed by Ipsos [14] addressed global AI opinions and expectations and stated that education, safety, transportation, and entertainment are among the top areas expected to change people's daily lives in the next 3–5 years due to the increased use of AI. Education (77%), entertainment (77%), transportation (74%), and safety (69%) were also among the areas where people expected the increased use of AI would improve their and their families' lives [14–15]. However, Eurobarometer conducted in 2019 showed that only half of Europeans perceived the best uses of AI to be in medicine, a third perceived its best use could be to alleviate traffic problems, and nearly one-fifth saw it as best used for societal safety and security [16].

Perceptions of people in various occupational fields and life domains have been compared in research focusing on robots, which can offer tentative implications of what the results could be for the perhaps more complex and abstract concept of AI. People have been reported to be generally quite positive toward robots doing more cognitively oriented tasks and in historically more familiar domains (e.g., manufacturing and military) but are more hesitant and divided about deploying robots to fields that require artistic or social skills such as education and leisure [17–19]. Although general and some task-specific attitudes toward robots have become more negative based on comparisons of 2012, 2014, and 2017 Eurobarometer datasets, Europeans' attitudes toward travel in driverless cars (from 2014 to 2017) and especially toward using robots for medical operations (from 2012 to 2017) have become more positive [20].

Another issue to consider is the difference between explicit perceptions of novel technologies (e.g., in surveys) compared to more implicit implications of sentiments and other expressions found in textual data. Fast and Horvitz [21] studied text corpora of news over 30 years and from various industries, and they included education, transportation, health care, and entertainment in the category of “hopes for AI” and military in the category of “concerns for AI.” They concluded that discussions of AI in news media rapidly increased after 2009 and have been more optimistic than pessimistic in nature [21]. Within the timeframe of the study (1985–2015), hopes for and positive views of AI in fiction, health care, and education trended upwards, whereas similar positive trends were not found for entertainment in general, transportation, and military topics [21].

Another study on text corpora compared public and news sentiments on robots and reported that although topics such as military robots have been a major concern over the past decade, topics such as health care and especially educational robots trended in public discussions and news in 2018 [22]. Robots in entertainment were shown to be a more positively discussed topic than military ones [22]. In their review

of the literature, Javaheri et al. demonstrated that although the public's general view of AI is positive and less conservative than that of the experts, the attitudes vary between the application contexts. Even though some evidence for the positive connection between people's ratings and written sentiments has been reported [23], caution must be exercised when contrasting explicit opinions with implicit attitude results arrived at through linguistic analyses. The traditional approach of language analysis has established that words can convey the attention and focus of the messenger [24]. However, language is not an intra-psychological phenomenon free of social influence, but a means of communication and affected by its social context offline or online [25].

Affective and cognitive language in social media

Affect and cognition, along with the behavioral factor, are traditionally seen as the three components of the mind, as illustrated, for example, in the tripartite model of attitudes [26–27]. Cognition includes all forms of knowledge (e.g., language), awareness, memory, and cognitive processes [28]. Affect is sometimes used as a superordinate category or a collective term for emotion, feeling, and mood constructs [29]. Emotion is also considered to be a broad construct encompassing behavioral and physiological elements in addition to involving subjective mental experiences such as feelings [30]. The complex relationship between cognition and affective dimensions such as emotions is apparent in theoretical frameworks for both cognitive and affective constructs. Emotional components are integrated in theories of cognitively oriented constructs such as attitudes [26], and the cognitive aspects of emotions have been extensively studied in relation to emotion theories [31].

People use verbal (i.e., words that describe feelings) and nonverbal cues (e.g., facial expressions) to express emotions. In natural language, less than 5% of English words can be categorized as emotional [32], which highlights the multimodality of emotional expression through means other than only verbal cues. The linguistic patterns and frequency of using emotional and cognitive words can still reveal valuable insights, for example, about the development of people's psychological state [32]. The small proportion of emotion words demonstrates the weight of language in conveying nonemotional words and constructs, such as cognitive and analytic ones. Much research has been devoted to examining the interaction between language and cognition and its potential brain–mind mechanisms resulting in findings that language predefines several cognitive concepts [33–34].

Analytic thinking is often defined in contrast with intuitive gut feelings and has been found to have significant consequences for individual psychology [35]. Analytic language can arguably have a significant societal impact and has been found to predict academic achievement [36], to distinguish one writer from another [37], to identify communication style trends in politics [38], and to persuade others to change their mind [39]. Analytic thinking style has been studied to capture formal and logical and less personal and emotional language while distinguishing it from more general cognitive processing, which refers to meaning making and the expressions of cognitive activity [40]. Other studies have investigated, for instance, information processing and cognitive–affective styles associated with war attitudes [41], cognitive and

affective cues to detect nostalgia [42], and cognitive processes and emotions in the context of alternating cognitive and emotional fluctuations in public conversation [1].

Studies on cognitive processes and analytic thinking cues in public discussions on AI or other emergent technologies are insufficient. However, research on affective dimensions such as emotions and sentiments have utilized sentiment analysis tools to study, for example, written reactions toward working with robots [43] and found AI social media comments on Reddit to be more often negative and more polarized than robot and bot comments [44]. Comparisons of cognitive and emotional language on other topics demonstrate that the ratio between cognitive and emotional words varies between different discussion topics [1], highlighting the context-dependency.

After considering analytic and emotional wording styles in written language, we are left with the question: What does it matter in terms of a specific topic such as AI discussions? Does emotional versus analytic language in social media writings have different outcomes, for example, in terms of social influence or societal ramifications? Examining the social consequences of cognitive or affective language use addresses Boyd and Schwartz's [24] call for extending the descriptive analyses of linguistic categories to explanatory investigations of social psychological phenomena. In persuasion research, researchers have studied the attributes of messages that drive people to change their attitude or behave a certain way [45–49]. In the context of social media, this could be specific ways of engagement with a piece of content, such as reposting, rating, liking, or other types of reactions enabled by the platform [50]. Researchers have demonstrated that in addition to the diversely measured content attributes of the message, emotions play an important role in deciding to engage [3, 51–56]. Dolan et al. [2] found that rational appeals in social media predict higher passive and active engagement, whereas emotional appeals predict only higher passive engagement.

Persuasion research provides some possible theoretical explanations for the different appeal of emotional and rational content characteristics. Influential theories of dual processing such as a heuristic–systematic model of information processing [8] and an elaboration likelihood model [9] typically include a more explicit, rational, analytic, and intentional reflection on the content of the message, in contrast to the other more unconscious, intuitive, and superficial component of thinking which is more prone to emotional appeals [57–58]. The duality of persuasion can be traced back to Aristotle's idea of separating passionate or emotional persuasion and rational persuasion [59]. Dual processing theories address also the content characteristics of the message and have been utilized to study the persuasive power of those cues [60–61], but these theories focus on the information processing styles rather than content cues solely.

Our study focuses on the emotional–analytic appeals of content characteristics addressed in persuasion theories. Some studies have deployed persuasion theories to investigate perceptions of AI in algorithmic journalism [62–63] or investigated attitude and trust implications of robot information in fiction versus fact-based media [64]. However, no studies have focused on the social influence and persuasion of cognitively or analytically versus affectively or emotionally oriented social media messages on AI.

Research questions

The theoretical framework of our research consists of studies and theories on affective and cognitive components in written language, persuasion and the appeal of content characteristics, and context-dependency. To consider the context-dependency [1, 10], we investigated the affective and cognitive language in social media posts on AI in relation to different contexts. Based on the relevant topics in the literature on AI in various contexts and life domains [11–12, 16], we chose eight contextual cues for our analyses: urban, art, hospital, school, military, gaming, music, and movies/books. Based on research on language in general [32], we expected to find that posts on AI would more often consist of cognitive and analytic than affective and emotional content. Based on research on persuasion in social media [2], we expected cognitive and analytic orientation of a message to facilitate more social interaction with the message by other users, but affective and emotional orientation to facilitate only more passive (voting) rather than active engagement (follow-up comments). Our research questions (RQs) were as follow:

- RQ1: How does the tone of the Reddit posts on AI associate with different contextual cues (urban, art, hospital, school, military, gaming, music, and movies/books)?
- RQ2: Do Reddit posts on AI consist of more cognitive and analytic than affective and emotional content?
- RQ3: Are cognitively and analytically oriented Reddit posts on AI more likely to receive a reaction (voting and follow-up comments) from other users than affective and emotional posts are?

Method

Data collection and processing

To answer our RQs, we utilized pushshift.io to identify Reddit posts called *submissions* referring to AI [65] and collected 471,525 Reddit submissions from 2005 to 2018. We prepared the submission corpus for statistical analyses by first excluding duplications, ergo, identical submissions by the same author to the same subreddit ($n=11,009$). We then excluded submissions with irrelevant search hits for the word *AIN'T* ($n=4,825$), submissions that had only the word “AI” in the title and no other words in the submission ($n=56$), and one observation that had the value -1 for number of comments. After these exclusions, we had our final text corpus for analysis ($N=455,634$). The posts were submitted to 23,075 subreddits by 223,600 authors.

We chose Reddit as the social media data source for its expansive datasets with thematically structured community discussions on various topics [50, 66]. Based on 2019 and 2021 surveys of U.S. adults, Reddit was among the top 10 most popular social media platforms [67]. Reddit has been the data source for diverse research designs and topics, especially in computer sciences, medical sciences, and sociology,

and had been increasingly utilized in 2010–2020 [66]. It has been previously utilized to study user engagement tendencies [e.g., 68–71]. Pushshift is a platform for Reddit data retrieval for researchers that includes historical archives and continuously updated data [65].

Instead of an individual person, the target unit of this investigation is an AI-related submission posted on Reddit during the time frame (2005–2018). Because the collected data includes all the target units of investigation, we used a total population sample. Therefore, the results of the descriptive analyses of the data describe the target population without the need to generalize from a sample to the overall population. The sample size enabled enough power to detect even smaller effects, but inferential statistics were used to detect how relevant the effects are [72]. The Reddit platform is a public discussion forum where people interact on the platform through usernames. We retrieved the following information for the submissions: title, self-text, publication date, author username, the subreddit in which it was published, and the number of upvotes, downvotes, and comments it received.

Our search did not include identifiable or sensitive information, and AI is not considered a sensitive topic. We analyzed the submission texts from a larger scope to obtain insights from AI-related public discussion from a societal perspective. Thus, we do not report any individual-level information or results. Reddit's terms of use state the possible research purpose of the publicly available data [66], which fulfills the legal requirement of using its data [73]. To protect the authors of the submissions, we do not report any usernames or quotes from the users [66]. Prior ethical assessment is required when the data are not entirely public, which in this case the information is, and explicit consent is impossible to acquire [74]. Before collecting the data set, the local academic ethics committee stated that the study design does not include any ethical issues.

For human validation, we asked each participant ($N=1059$) to rate 20 texts from Reddit in a scale from -4 to 4 for negative–positive tone orientation. Of the sample texts, 250 referred to AI, 250 referred to a robot, and 500 were a random sample of Reddit comments. Each of the 1000 texts were given a mean score from human ratings. For comparison, we created three variables: human ratings mean score, LIWC Tone score, and VADER Compound score. The two other variables were rescaled to a scale from -1 to 1 to match the scale of VADER Compound score. Based on two reliability estimates for continuous variables, the three variables are measuring similar phenomenon: Cronbach's alpha ($\alpha=0.73$) and McDonald's omega ($\omega=0.74$) [75–76]. From the two sentiment analysis tools, the mean score of human ratings was slightly closer to VADER Compound score ($\alpha=0.60$; Pearson's correlation: $r=.43$; Concordance correlation: $r=.33$) than LIWC Tone score ($\alpha=0.55$; Pearson's correlation: $r=.38$; Concordance correlation: $r=.19$). Altman [77] recommends for Lin's concordance correlation coefficient similar interpretation than Pearson's coefficients where below 0.20 are considered poor and above 0.80 excellent.

Measures

The descriptive statistics of all study variables are shown in Table 1. For RQ1, we used the scores of two sentiment analysis tools as the dependent variables: the LIWC

Table 1 Descriptive statistics of the study variables: original continuous measures

Measure	<i>n</i>	M	SD	Min	Max
VADER compound	455,634.000	0.178	0.653	-1	1
LIWC tone	455,634.000	35.996	32.327	1	99
LIWC cognition	455,634.000	11.559	7.559	0	80
LIWC affect	455,634.000	5.537	5.168	0	75
LIWC analytic	455,634.000	66.426	27.726	1	99
LIWC emotion	455,634.000	1.393	2.488	0	50
Urban ¹	455,634.000	0.090	0.655	0	40
Art ¹	455,634.000	0.084	0.786	0	40
Hospital ¹	455,634.000	0.027	0.454	0	45.450
School ¹	455,634.000	0.052	0.572	0	40
Military ¹	455,634.000	0.166	0.888	0	37.500
Gaming ¹	455,634.000	1.309	2.609	0	66.670
Music ¹	455,634.000	0.062	0.678	0	50
Movies & books ¹	455,634.000	0.195	0.991	0	66.670
Word frequency	455,634.000	258.600	605.8405	2	11,562.000
Year	455,634.000	2016.347	1.695	2005	2018
Voting score ²	451,393.000	27.346	419.083	0	83,019.000
Num of comments ²	451,393.000	12.161	71.504	0	23,253.000

Note¹Custom dictionary created by the research team,
²Engagement with the Reddit post

lexicon for emotional tone and the VADER compound lexicon. For RQ2 we used LIWC cognition, affect, analytic, and emotion lexicons, and for RQ3 we used Reddit submissions' voting scores and numbers of comments as the dependent variables. Eight custom dictionaries (urban, art, hospital, school, military, gaming, music, and movies/books) were used as the main independent variables. Word frequency and year of the submissions were used as control variables. VADER and LIWC tools have been previously applied to study user engagement effects of linguistic features [e.g., 68, 78].

The analyzed texts included the submission's title, self-text, and subreddit. A submission on Reddit is given a title and can be either a link submission or a text submission called a *self-post*. A title of the submission can be up to 300 characters long. The character limit for self-posts used to be 40,000 for self-post-only subreddits and 15,000 for other subreddits, but in 2015, it was changed to 40,000 characters for all subreddits [79]. We analyzed VADER and LIWC's standard dictionaries based on the title and self-text parts of the submissions. In addition to the title and self-text, we included *subreddit* in the analyses of our custom dictionaries that aimed to capture the contextual theme of which the submission was written. In both cases, the two or three text parts were treated as one combined segment of text.

Tone positivity: VADER compound and LIWC tone

The first variable we used to measure the tone positivity is the unidimensional compound score of VADER, a rule-based model for general sentiment analysis [80]. This metric is calculated from valence scores given to the lexicon words by summing, adjusting by the rules, and finally normalizing the score from -1 to 1 [81]. This weighted and normalized composite measure is a recommended metric for analyzing sentiment on one dimension [81].

As a second dependent variable for tone positivity, we utilized LIWC-22's unidimensional variable labeled *emotional tone*, which measures the degree of positive (or negative) tone and is built upon research [82–83]. Instead of a percentage of total words such as in LIWC's standard dictionaries, the tone is a composite variable with a scale of 1–100 that consists of a percentile converted from calculated measures' standardized scores [84]. The tone variable combines LIWC's positive and negative tone dimensions into a summary measure where a higher score indicates higher positivity and a score below 50 indicates higher negativity [85]. LIWC-22's tone dictionaries reflect sentiment rather than emotion words [84]. As such, they are conceptually closer to and have a higher correlation with VADER measures than the LIWC emotion dimensions (e.g., Pearson's correlation coefficient for VADER compound and LIWC tone is $r=.48$, whereas for VADER compound and LIWC positive emotion, it is $r=.17$).

Cognitive, analytic, affective, and emotional orientations: four LIWC measures

To capture cognitive orientation of a submission, we used LIWC-22's variable labeled *cognition*, which is a percentual metric of cognitive words within a text with a scale of 0–100. LIWC cognition compares the words in a text with the lexicon words and calculates the percentage of lexicon words present in each text [84]. To capture more analytic orientation of a submission, we used LIWC-22's summary variable labeled *analytic thinking*, which is a metric of logical and formal thinking. Like the LIWC tone, it is a standardized composite variable with a scale of 1–100 and has been built upon research [36].

In addition to the tone summary variable, LIWC-22 includes nine affect dimensions. Affect, which is the overall dictionary for affect, includes tone categories for positive and negative tone, a swear category, and an emotion category. The emotion dimension further includes positive and negative emotion dimensions, from which the latter includes three further categories for negative emotions (anxiety, anger, and sad). From these, we utilized the lexicons labeled *affect* and *emotion*. The latter does not include tone categories (which are closer to our tone analysis for RQ1) and swear words, but it measures the emotionality by counting the percentage of emotional words used in each text. LIWC-22 can capture emoticons, that are an important part in conveying affect and emotion [86–87], and word stems.

Studies using a previous version of the LIWC program have, for example, compared cognitive processes with affect scores [1] or utilized analytic thinking along with cognitive processes to capture slightly different processes [40]. LIWC-22's new comprehensive cognition dictionary is currently the most general cognitive dimen-

sion that measures the words people use to refer to their thinking and includes the cognitive processes as a subcategory. We chose to include the more specific categories of analytic and emotion in addition to the more general cognition and affect in our analyses to obtain a comprehensive view of overall cognitive and affective language and its analytic and emotional content.

User engagement with the submissions

Voting score and the number of comments were used as the two dependent variables for measuring and predicting user engagement with the submissions. We retrieved the information from Reddit along with the submissions and both measures have missing values of $n=4,241$. We retrieved the total number of comments the submissions had received from other users up to our data collection cutoff (March 2019) and created a dichotomous variable for the final analysis (0=*no comments*; 1=*one or more comments*).

Logged-in users on Reddit can aid in ranking the submissions by upvoting or downvoting them [88–89]. Although individual users could treat the voting as simply an expression of liking or disliking a post, the official purpose of voting is stated to be an assessment of the value or contribution a submission or comment adds to the conversation or community [88]. Reddit’s frequently asked questions list states that the number of upvotes and downvotes are “fuzzed” to eliminate voting bias and are not reliable as such, and they have not been systematically available for retrieval either. For this reason, we selected the more reliable voting score as a proxy for user engagement through voting reactions. Unlike the voting score for comments, the voting score for submissions cannot have a negative number and, thus, all submissions receiving more negative than positive votes have the value 0. The default score of the submissions is 1 (the author is assumed to give a vote to their submission). We created a dichotomous dummy variable where the default score 1 was given a value 0 (= *no voting engagement from other users*) and scores of 0 and over 1 were given a value 1 (= *negative or positive voting engagement from other users*). This dummy variable was used in the final analysis. As an additional test, we also created a dummy variable where scores of 0 were combined with scores of 1 to compare negative or neutral scores with positive scores; the results with this variable remained the same, which adds robustness to our findings.

Contextual cues

We selected eight relevant contexts based on the literature on the future directions of AI [11–12, 16]: urban, art, hospital, school, military, gaming, music, and movies/books. For these contextual cues, we utilized LIWC-22’s dictionary workbench function to create custom dictionaries. After entering a few seed words for each category, the program allows a user to find and add similar words by using its built-in word network search. The final categories were then tested for internal consistency utilizing LIWC-22’s demo data [90], which includes Reddit data from the Pushshift platform [65]. The internal consistency results are presented in Appendix A. As recommended for text data, we looked at Kuder-Richardson formula 20 (KR-20) results and found

the internal consistencies of the new dictionaries to be good based on KR-20 values being >0.50 . The values of Cronbach's α were also assessed to be acceptable for natural language in text form (>0.30).

Control measures

Two other measures were used as control variables in the models: the word frequency and year of the submission. Word frequency was acquired using LIWC-22 program when running the content analyses, and it ranged from two to 11,562 words. The year variable was created from the timestamp information retrieved from Reddit. The first comment was created in 2005 and the last one in 2018.

Analysis methods

Table 1 reports descriptive results for the study variables, including frequencies, means, standard deviations, and minimum and maximum values. For the analyses, we computed linear regression models with robust estimation of variance and M robust estimators that reweight least squares iteratively [91]. For the content orientation analysis models and effect sizes we report unstandardized regression coefficients, their estimated standard errors, confidence intervals (CIs; 95%), and statistical significance. For the user engagement analyses of voting scores and the number of comments, we computed logistic regression models and report odds ratios, statistical significance, and average marginal effects (AME). The robust model coefficients and AMEs of the logistic regression models are also illustrated in regression coefficient plots with standardized independent variables (Figs. 1 and 2, and 3). As the linear regression models with continuous variables yielded similar results, for the user engagement analyses we report logistic regression models to avoid complexity with voting score interpretation. Two-way histograms in Fig. 4 report distributions of two LIWC lexicons with scales of 0–100 in fractions that scale the height of the histogram bars so that summing their heights is equal to 1.

The models in Figs. 1, 3, 4 and 5 included 455,634 observations, whereas the models in Fig. 2 predicting user engagement included 451,393 observations. All the variables are standardized for the regression models for comparison, except VADER compound, LIWC tone, and LIWC analytic, which are already standardized measures. Based on variance inflation factors, the multicollinearity of the variables in all the models is low (<1.28). We report all data exclusions, manipulations, and measures utilized in this study. We conducted the automated content analyses with the VADER sentiment analysis tool and LIWC-22 software. Follow-up statistical analyses were performed with Stata 17 software.

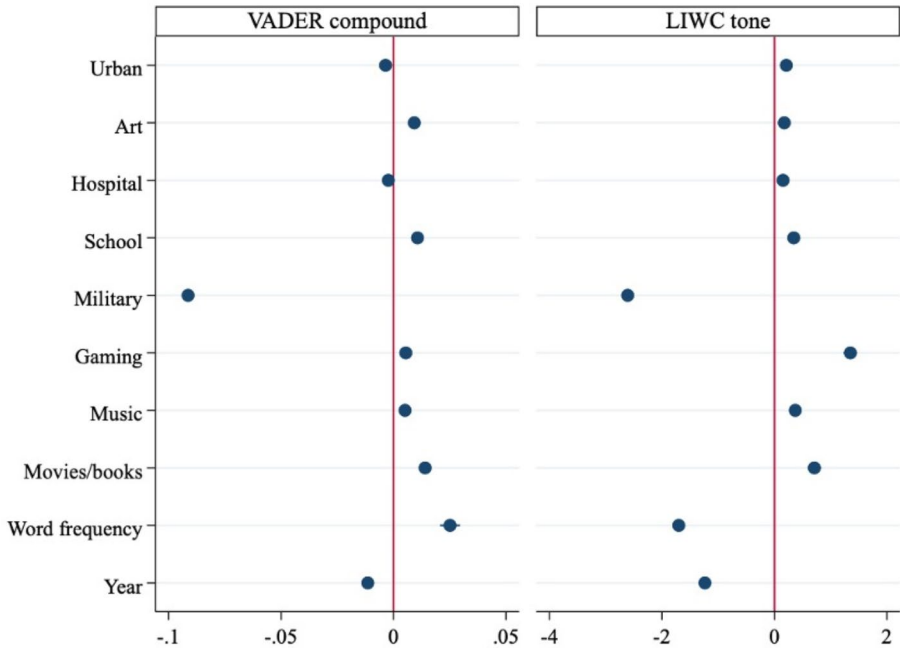


Fig. 1 Robust Regression Models with M Estimator and Robust Standard Errors Predicting Tone Positivity With VADER Compound and LIWC Tone. *Note:* 455,634 observations (Reddit submissions on AI). The independent variables were standardized for comparison

Results

Contextual cues and tone in Reddit submissions on AI

Figure 5 presents time trends of descriptive statistics for contextual cues across our study variables of VADER (compound) and LIWC (tone, cognition, affect, analytic, and emotion). The trends revealed that the tone (VADER compound and LIWC tone mean scores) in AI submissions have become more negative over the years within most of the analyzed contexts in 2005–2018, but based on LIWC tone, also slightly overall. The main trend in analytic and emotional scores have been slightly upwards, but no clear yearly trends were detected for the broader lexicons for cognitive and affective language in AI submissions.

The LIWC tone mean scores for all categories were more negative than the LIWC-22's reported mean score: $M=47.81$ [84]. Based on the means of the tone positivity variables (VADER compound and LIWC tone), AI submissions that consist of art and school contextual cues had the most positive tones on average. The language in AI submissions within the military context had the highest mean score for negative and affective words. Highest mean scores for analytic and emotional words and lowest for cognitive language were found for music contextual cues. Gaming context scored highest on cognitive and lowest on analytic language, on average.

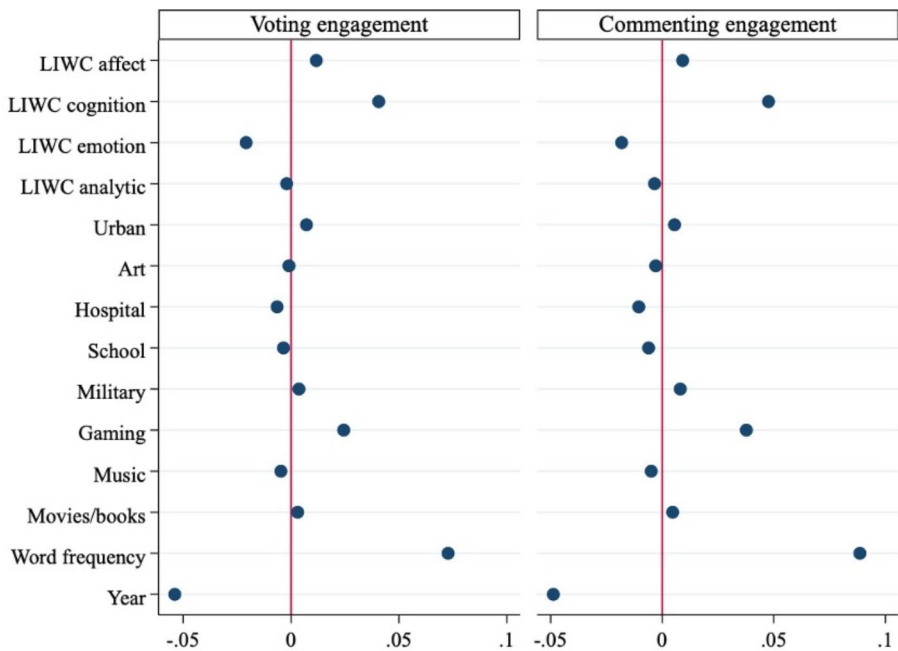


Fig. 2 Logistic Regression Model AMEs Predicting Voting and Commenting Engagement. *Note.* 451,393 observations (Reddit submissions on AI). Average marginal effects in the x-axis. Dummy variables for voting and commenting engagement (0=no engagement; 1=engagement from other users) as the dependent variables

The robust regression model coefficients in Fig. 1 predicting AI submissions' tone positivity (VADER compound and LIWC tone scores, respectively) showed that higher scores in contextual cues of gaming ($b=0.005$, $p<.001$; $b=1.350$, $p<.001$), school ($b=0.011$, $p<.001$; $b=0.343$, $p<.001$), art ($b=0.009$, $p<.001$; $b=0.175$, $p<.001$), music ($b=0.005$, $p<.001$; $b=0.371$, $p<.001$), and movies/books ($b=0.014$, $p<.001$; $b=0.709$, $p<.001$) were associated with positive tone, and military ($b=-0.091$, $p<.001$; $b=-2.610$, $p<.001$) contextual words with negative tone. The results between VADER compound and LIWC tone scores for urban ($b=-0.004$, $p<.001$; $b=0.211$, $p<.001$) and hospital ($b=-0.002$, $p<.001$; $b=0.153$, $p=.001$) contexts were mixed and thus less clear. In addition, an earlier submission year was associated with more positive scores ($b=-1.237$, $p<.001$; $b=-0.011$, $p<.001$), and longer submissions were associated with more positive VADER compound scores but more negative LIWC tone scores ($b=0.025$, $p<.001$; $b=-1.107$, $p<.001$).

Cognitive–affective orientation of Reddit submissions on AI

Descriptive analyses of LIWC's analytic and emotion lexicons revealed that Reddit submissions on AI consist of cognitive language ($M=11.559$; $SD=7.559$; range 0–80) more often than affective language ($M=5.537$; $SD=5.168$; range 0–75; see Fig. 4). A similar division with a larger difference was found between analytic language ($M=66.426$; $SD=27.726$; range 1–99) and emotional language ($M=1.393$;

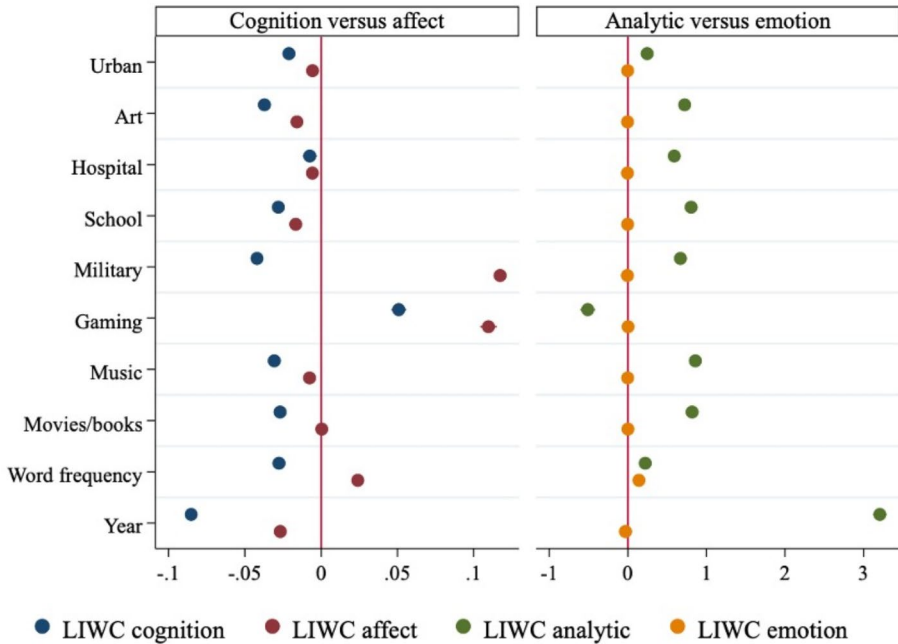


Fig. 3 Robust Regression Models with M Estimator and Robust Standard Errors Predicting Cognitive, Affective, Analytic, and Emotional Orientation of Posts on AI. Note. 455,634 observations (Reddit submissions on AI). Unlike the other three dependent variables, LIWC analytic is a standardized composite measure with possible values ranging from 1 to 100. LIWC's cognition, affect, and emotion measures' possible values range from 0 to 100. The latter three dependent variables and the independent variables were standardized for comparison

$SD=2.488$; range 0–50), although these two metrics are not as comparable in descriptive statistics as the two former ones due to unequal calculations behind the metrics. These descriptive results reflect the natural tendencies in the English language [32].

The cognition mean score ($M=11.56$) was slightly lower and the affect mean score ($M=5.54$) was slightly higher than the mean scores reported in LIWC-22's manual ($M=11.76$ and $M=5.35$, respectively) [84]. The analytic mean score ($M=66.43$) was significantly higher and the emotion mean score ($M=1.39$) was slightly lower than the LIWC-22 manual's reported mean scores ($M=49.52$ and $M=1.88$, respectively) [84]. These descriptive comparisons show that Reddit submissions on AI were more analytic in language use compared to average.

Figure 3 presents linear robust regression models examining how the scores in the eight contextual cues were associated with cognitive versus affective and analytic versus emotional orientation of the language in the Reddit submissions on AI. High scores in all the other contextual cues were connected to higher analytic orientation ($b=0.246-0.859$, $p<.001$), except the gaming context ($b=-0.511$, $p<.001$). High scores in gaming ($b=0.110$, $p<.001$) and military ($b=0.117$, $p<.001$) contexts were connected to higher affective language in the AI submissions. A later submission year was linked to higher analytic and lower cognitive, affective, and emotional language ($b=3.208$, $b=-0.085$, $b=-0.027$, $b=-0.030$, $p<.001$, respectively). Frequency of

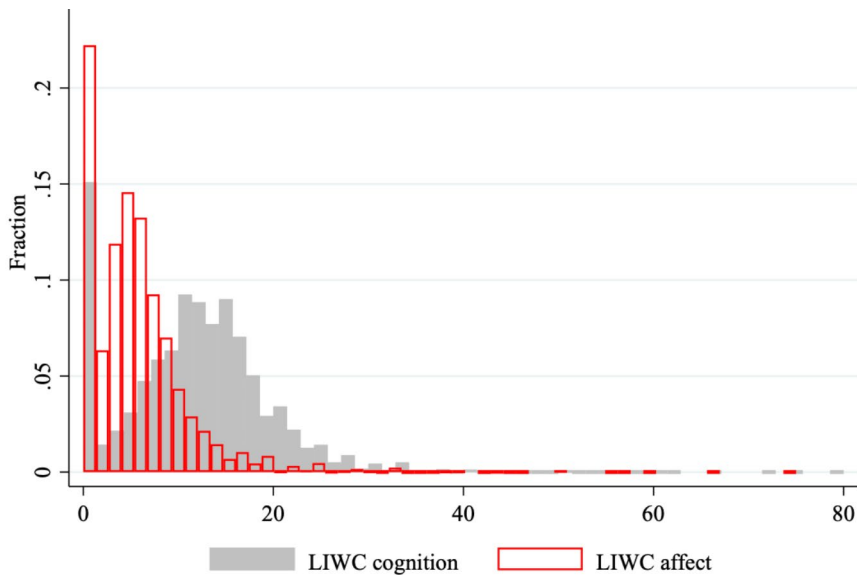


Fig. 4 A Two-Way Histogram of Cognitive and Affective Content in Posts on AI. *Note.* 455,634 observations (Reddit submissions on AI). Possible values for both variables range from 0 to 100. Fraction: the sum of the bars' heights equals 1

words used in the submission was associated with more emotionally oriented language ($b=0.141, p<.001$).

User engagement with Reddit submissions on AI

Figure 2 presents AMEs of the logistic regression results for user engagement with the Reddit submissions on AI in terms of voting score and the number of comments. The odds of Reddit submissions on AI to receive voting engagement from other users were slightly higher if more cognitive language was used ($OR=1.026, p<.001$) than if more affective language was used ($OR=1.011, p<.001$). We found a similar difference between analytic and emotional content that were both negatively associated with voting engagement ($OR=0.990, p<.001$; $OR=0.960, p<.001$). Figure 2 shows similar results for commenting engagement.

Higher scores for contextual cues of urban, gaming, military, and movies/books were positively associated with the odds of the AI submission to receive voting or commenting engagement from other users ($OR=1.015$ – $1.230, p<.001$). Higher scores for hospital, music, and school contextual cues were negatively associated with voting and commenting engagement ($OR=0.944$ – $0.983, p<.001$), and art was negatively associated with commenting ($OR=0.984, p<.001$) but not voting engagement ($OR=0.995, p=.130$). Longer submissions were positively ($OR=1.430$ – $1.627, p<.001$) and later submission year negatively ($OR=0.765$ – $0.767, p<.001$) associated with both types of user engagement. The latter is expected because the votes and comments of a submission are likely to accumulate over time.

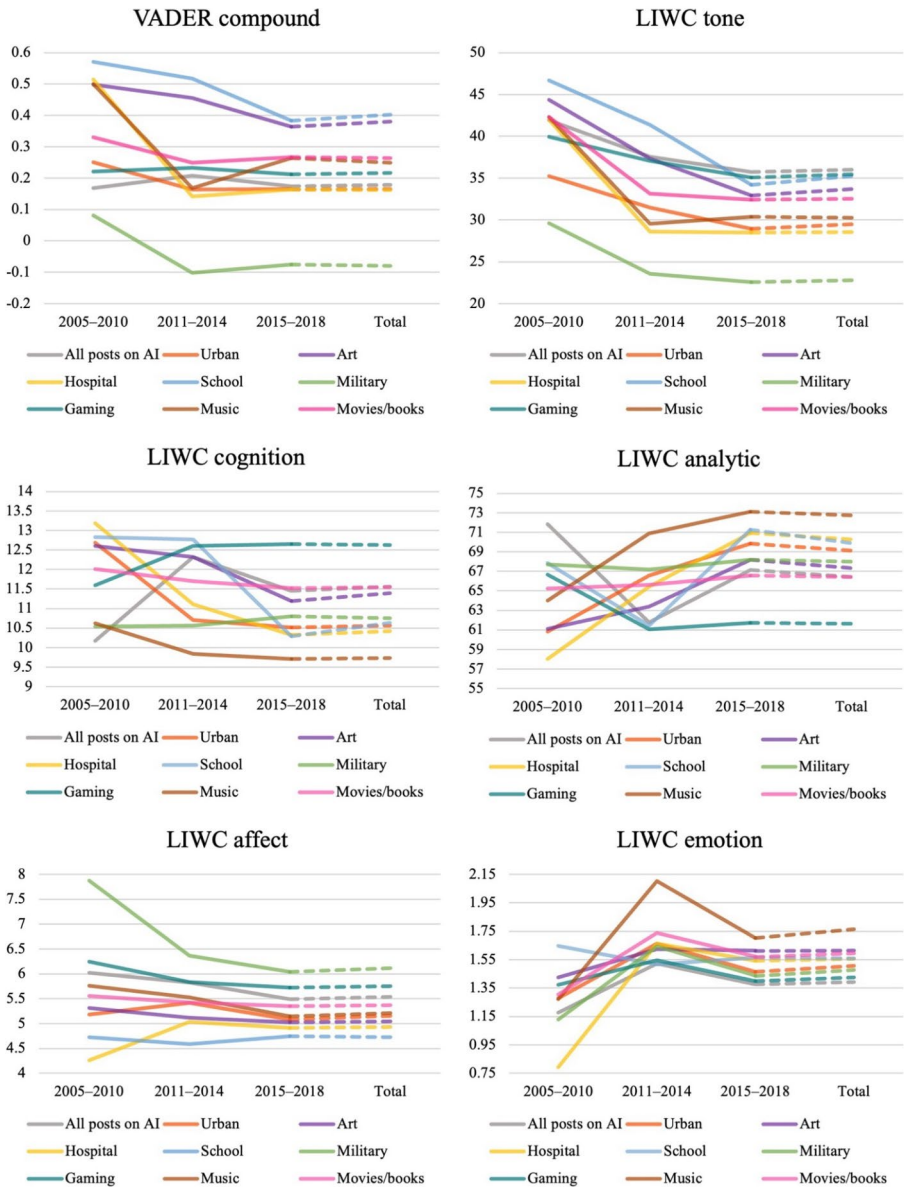


Fig. 5 Mean Scores of VADER and LIWC Measures by Contextual Cues in Posts on AI (2005–2018). *Note.* 455,634 observations (Reddit submissions on AI). Each context is measured by having >0 word matches for the context in question. VADER compound is a normalized and standardized measure on a scale from -1 to 1. LIWC tone and analytic are standardized composite measures on a scale from 1 to 100. LIWC cognition, affect, and emotion are percentual measures on a scale from 0 to 100. In addition to three time intervals, the figure shows mean scores of the total time frame, separated from the time trends with dashed lines

Discussion

We investigated social media posts on AI from Reddit utilizing emotional–analytic appeals of content characteristics approach to persuasion to examine affective–cognitive language. We collected all Reddit submissions on AI in 2005–2018 to analyze their tone positivity and affective versus cognitive language use with LIWC and VADER content analysis tools. Furthermore, we investigated how these linguistic contents are associated with our contextual cues of interest and engagement behavior from other users. In addition to the practical contribution of public opinion on AI, the results contribute to the theoretical discussions on affective and cognitive language in public social media discussions.

To answer RQ1, we investigated how the tone of the Reddit posts on AI were associated with different contextual cues. To consider the context-dependency [1, 10], we selected eight salient contexts for AI discussions based on literature on the future directions of AI [11–12, 16]: urban, art, hospital, school, military, gaming, music, and movies/books. The results revealed that contextual cues of gaming, school, art, music, and movies/books were associated with positive and military contextual words and a later submission date were associated with negative tone. The results for urban and hospital contexts were mixed between the two utilized tools for tone positivity analyses. Although researchers have reported optimistic trends for AI discussions within fiction, health care, and education [21] and for robot attitude surveys within medicine, traffic, and military [17–18, 20] during the last decade, our findings for 2005–2018 Reddit discussions on AI did not show similar positive trends. Our conclusion was closer to previous findings on AI and robots showing negativity or no positive trend for military topics compared to, for example, entertainment [17–18, 21–22].

For RQ2, we examined whether Reddit posts on AI consist of more cognitive and analytic than affective and emotional content. The language in Reddit posts on AI was less emotionally than analytically oriented. This finding reflects the natural tendency of language to consist of less than 5% emotional words [32]. However, the tendency for lower emotional and especially higher analytic language was stronger for Reddit's AI posts than the reported average scores for LIWC variables. The affective–cognitive orientation was also dependent on the more specific context, highlighting the previous research finding on the context-dependency stating that the cognitive–emotional language ratio in public discussions is dependent on the topic [1]. Gaming context predicted higher emotional and affective language than analytic and cognitive language. The other examined contexts predicted higher analytic language compared to emotional, affective, and cognitively oriented language. In general, AI posts on Reddit have become more analytic and less cognitive and affective in language use.

To answer RQ3, we investigated whether affective and emotional Reddit posts on AI were more likely to receive engagement (voting and follow-up comments) from other users than posts high in cognitive and analytic language. Our results revealed that other users were more likely to engage with AI posts that were high in cognitive compared to affective content, which also predicted user engagement positively. A similar relationship was found for analytic versus emotional language, although both

of these content characteristics predicted user engagement negatively. User engagement with AI posts was also affected by the more specific context. Other users were more likely to engage with posts on AI in the contexts of gaming, urban, military, and movies/books, whereas the opposite was found for hospital, music, school, and art contexts. These post engagement behavior observations occurred systematically for both voting score and the number of comments. Although we did not find that emotional appeals would only predict passive user engagement as Dolan et al. [2] did, our results were in line with their finding that rational appeals in social media predict higher engagement in terms of both active and passive engagement.

Theoretical and practical implications

The results have implications for the theoretical discussions on affective and cognitive language in public social media discussions focusing on the topic of AI. Dolan et al.'s [2] study demonstrated how rational appeals in social media predict higher engagement in terms of both active and passive engagement. Our findings verify that this is also the case in AI discussions on Reddit, specifically in terms of cognitive content characteristics but not in terms of analytic language. Content scoring high on the analytic thinking variable we used tends to be linked with high reasoning skills but also viewed as less intuitive, friendly, and personal [36–38]. Personal and informal communication style has been reported to predict engagement in social media [92], which could explain why analytic language did not yield more user engagement in our study. In addition to words on cognitive processes, the cognition lexicon we used consists of memory vocabulary and all-or-none words such as “never” or “always” [84]. This gives more insight to the type of cognitive language that predicted more user engagement in our study implying that users reacted more, for example, to posts using cognitive language to state something more absolute rather than offering a more formal analysis on it. These conclusions would be in line with research reporting that certainty in language predicts more consumer engagement [93].

Findings in prior literature suggest that higher emotional content in social media messages predicts more sharing behavior [53, 55]. Dolan et al. [2] further demonstrate how emotional appeals predicted only passive and not active type of user engagement (e.g., liking vs. re-posting). Our results do not support the notion of emotional appeals being a more mobilizing factor than cognitive language in the case of AI discussions on Reddit. Comparing the results between affect lexicon and its sub-lexicons offer a little more insight on this result. Further analysis on the sub-dictionaries reveals that AI posts that express positive emotions or anger, sadness, anxiousness, and swearing are connected to more user engagement, which could explain why the overall affect lexicon predicts higher user engagement (not as strongly as cognition). Therefore, other negative emotional expressions seem to explain the negative connection of emotion lexicon with user engagement. This could imply that people do not engage with AI posts consisting of emotions that express, for example, tiredness or negative judgements (e.g., weird). Our mixed finding for emotional appeals could also be due to topic or platform-specific differences [94]. Our results are in line with prior research on visual social media formats highlighting context-dependency and reporting less engagement for emotional compared to rational content [95].

Our study strengthens the importance of context-dependency [1, 10] and confirms that the cognitive–emotional language ratio in public discussions depends on the specific topic [1]; for example, discussions referring to AI. The findings also offer practical contributions to AI discussions and public opinion on relevant and timely life domains. If one wishes to avoid reading negative discussions, focusing on topics such as entertainment and education instead of military could be beneficial. In addition to offering other users more content in discussion-opening messages and avoiding specific negative emotional expressions, the use of cognitively oriented language is the most recommended strategy if one wishes to engage other users in social media discussions on AI. Affective language in general might not hinder this aim, but highly emotional negative content could cause other users to avoid engaging with the message. Our study calls for more research on the effects of affective or emotional (vs. cognitive or analytic) language on different types of user engagement.

Strengths, limitations, and future research direction

Our study includes more than one lexicon for each examined construct to strengthen the reliability of the findings. We utilized recently updated LIWC automated content analysis lexicons and measure reliability diagnostic tools to test their utility and further the research avenues of computational social sciences. This study used a total population sample (all Reddit submissions on AI). Therefore, the results of the descriptive analyses of the data describe the target population without the need to generalize from a sample to the overall population. The sample size enabled enough power to detect even smaller effects, but inferential statistics were used to detect how relevant the effects are [72]. In addition to descriptive statistics, we report inferential models that control for the potential effects of different contexts and the length and year of the submissions.

In addition to the tendency of social media users to be highly educated and wealthy, Reddit users specifically are more likely to be young and male [96]. Although our study design is not as sensitive to potential missing observations as studies focusing on demographic, longitudinal, or network analyses [97], caution should be exercised when making generalized statements based on social media data from Reddit or other platforms. Although we did not find a reliable way to incorporate topic-information of subreddits to analyze our dataset, subreddits can be considered for future studies beside or instead of custom dictionary workbench approaches. Future research plans should consider comparing the target topic with discussions of other topics and implement this aspect to the research designs. While controlling for the potential effects of context-dependency, future studies should continue to investigate the effects of cognitive–affective language on discussions of different topics. The influence of analytic vs. emotional appeal on social media engagement from other users specifically needs more attention in future research.

Conclusions

Utilizing content characteristics persuasion approach in examining affective–cognitive language, we investigated social media posts on AI utilizing 2005–2018 data

from Reddit. The results revealed that contextual cues of school, art, gaming, music, and movies/books are associated with positive and military contextual words and later submission year with negative tone. The general trends revealed that Reddit posts on AI have become increasingly analytic and negative in tone and decreasingly cognitive and affective in language over the years. Although the language in AI posts was more analytically than emotionally oriented in general, the affective–cognitive language ratio was context-dependent. Other users were more likely to engage with AI social media posts high in cognitive and analytic content compared to affective and emotional content. Avoiding emotional language and favoring cognitive language seems to be the most beneficial strategy for those looking to engage other users. In addition to offering practical contributions to AI discussions and public opinion on relevant and timely life domains, the results have implications for the theoretical discussions on affective and cognitive language in public social media discussions.

Appendix A

Custom Dictionary Internal Consistency Statistics and Word Entries

Urban	Art	Hospital	
Cronbach's α : 0.35	Cronbach's α : 0.36	Cronbach's α : 0.49	
KR-20: 0.57	KR-20: 0.50	KR-20: 0.62	
Mean (SD): 0.08 (0.19)	Mean (SD): 0.08 (0.18)	Mean (SD): 0.05 (0.15)	
urban*	art	hospital	oncology
architecture	arts	doctor	care unit
city	artist*	nurse	nursing
cities	artmaker*	clinic	attending physician
public place*	creativ*	health care center	dermatology
public environment*	artwork*	physician	paramedics
street*	sculpture*	operating room	in the emergency department
traffic	sculptor*	urgent care	orthopedic clinic
transport*	drawing*	pharmacist	general surgery
towns	painting*	surgical p.a.	paramedic
neighborhoods	painter*	patient	hospital setting
highways	poetry	pediatrician	number of patients
the city	poem*	emergency room	surgical unit
town	galler*	physical therapist	cardiac
commuters	portray*	pediatric	memory care
buildings	sketch*	nurse practitioner	skilled nursing facility
roads	ceramics	nursing home	colonoscopy
	mural*	surgeon	medical assistant
	handcraft*	surgery	nurse aide
	aesthetics	icu	clinical setting
	watercolour	patient's room	psychiatrist

Urban	Art	Hospital	
	watercolor	clinic setting	dermatologist
	acrylic*	emergency department	rehab
	fresco*	intensive care unit	oncologist
	coloring	ambulance	local clinic
	calligraphy	pharmacy	general medicine
	statue*	care facility	medical center
	graphic*	doctors	physician assistant
	photograph*	dentist	radiologist
	picture*	memory care unit	neurologist
		nurses	biopsy
		gastroenterologist	lab assistant
		patient care assistant	cardiologist
		waiting room	surgery center
		outpatient	a physician
		internal medicine	

School	Military	Gaming	Music	Movies / Books
Cronbach's α : 0.45	Cronbach's α : 0.38	Cronbach's α : 0.32	Cronbach's α : 0.47	Cronbach's α : 0.38
KR-20: 0.58	KR-20: 0.58	KR-20: 0.63	KR-20: 0.66	KR-20: 0.72
Mean (SD): 0.16 (0.28)	Mean (SD): 0.05 (0.16)	Mean (SD): 0.21 (0.32)	Mean (SD): 0.13 (0.27)	Mean (SD): 0.24 (0.33)
universit*	military	game*	music*	movie*
school*	army	gaming	performer*	film*
education	defence force	player*	orchestra*	tv-show
teacher*	marine corps	contest*	composer*	actor
student*	navy	tournament*	performance*	actress
	armed forces	campaign*	song*	tv
	troops	match	concert*	cinema
	war	championship*	symphon*	television
	marines	battle*	philharmonic*	big screen
	air force	opponent*	singing	movie scene
	militia	videogam*	lyrics	tv-episode
	forces	multiplayer	rhythm	movie script
	infantry	mmo	band	tv-series
	soldiers	rpg	bands	novels
		level	playlist	novel
		dlc	sing	books
		expansion pack*	soundtrack*	book
		teampay	cd*	story*
		co-op	choir	script
		walkthrough	instrument	character
		winning	choral	characters
		losing	chorus	writer*
		avatar*	harmonies	reader*
		arcade	singer*	casting
		e-sport	instruments	narrative

School	Military	Gaming	Music	Movies / Books
		esport	melodies	documentar*
		duel*		fiction
		quest*		heroine
		instance*		hero
		guild*		author*
		loot*		nonfiction
		in-game		typewriter*
		console*		trilogy
				protagonist*

Funding The authors have no conflicts of interest to declare. This research received funding from Kone Foundation (UrbanAI project, 2021–2024, PI Atte Oksanen, grant 202011325). David Garcia and Max Pellert received funding from the Vienna Science and Technology Fund (WWTF) [<https://doi.org/10.47379/VRG16005>]

Open Access funding provided by LUT University (previously Lappeenranta University of Technology (LUT)).

Data availability The data that support the findings of this study are available in the pushshift.io at <https://doi.org/10.1609/icwsm.v14i1.7347> [65].

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bail, C. A., Brown, T. W., & Mann, M. (2017). Channeling hearts and minds: Advocacy organizations, cognitive-emotional currents, and public conversation. *American Sociological Review*, 82(6), 1188–1213. <https://doi.org/10.1177/0003122417733673>
2. Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J., & Goodman, S. (2019). Social media engagement behavior: A framework for engaging customers through social media content. *European Journal of Marketing*, 53(10), 2213–2243. <https://doi.org/10.1108/EJM-03-2017-0182>
3. Pei, J., Lu, Z., & Yang, X. (2022). What drives people to repost social media messages during the COVID-19 pandemic? Evidence from the Weibo news microblog. *Growth and Change*, 53(4), 1609–1626. <https://doi.org/10.1111/grow.12573>
4. Devedzic, V. (2022). Identity of AI. *Discover Artificial Intelligence*, 2(1), Article23. <https://doi.org/10.1007/s44163-022-00038-0>
5. Eliasoph, N., & Lichterman, P. (2003). Culture in interaction. *The American Journal of Sociology*, 108(4), 735–794. <https://doi.org/10.1086/367920>

6. Moscovici, S. (1988). Notes towards a description of social representations. *European Journal of Social Psychology*, 18(3), 211–250. <https://doi.org/10.1002/ejsp.2420180303>
7. Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 391–445). McGraw-Hill.
8. Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>
9. Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. https://doi.org/10.1007/978-1-4612-4964-1_1
10. Kieslich, K., Lünich, M., & Marcinkowski, F. (2021). The Threats of Artificial Intelligence Scale (TAI) development, measurement and test over three application domains. *International Journal of Social Robotics*, 13, 1563–1577. <https://doi.org/10.1007/s12369-020-00734-w>
11. Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., & Teller, A. (2022). Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence. *arXiv Preprint*. arXiv:2211.06318
12. Jordan, M. I. (2019). Artificial intelligence—the revolution hasn’t happened yet. *Harvard Data Science Review*, 1(1), 1–9. <https://doi.org/10.1162/99608f92.f06c6e61>
13. Grand View Research (2022). *Artificial Intelligence market size, share & trends analysis report by solution, by technology (deep learning, machine learning), by end-use, by region, and segment forecasts, 2023–2030*. Report GVR-1-68038-955-5. <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market/>
14. Ipsos (2022). Global opinions and expectations about AI. January 2022.
15. Myers, J. (2022). Jan 5, 5 charts that show what people around the world think about AI. World Economic Forum. <https://www.weforum.org/agenda/2022/01/artificial-intelligence-ai-technology-trust-survey/>
16. European Union (2020). *Europeans and Artificial Intelligence*. Standard Eurobarometer 92. <https://ec.europa.eu/eurobarometer/surveys/detail/2255>
17. Savela, N., Turja, T., & Oksanen, A. (2018). Social acceptance of robots in different occupational fields: A systematic literature review. *International Journal of Social Robotics*, 10(4), 493–502. <https://doi.org/10.1007/s12369-017-0452-5>
18. Taipale, S., De Luca, F., Sarrica, M., & Fortunati, L. (2015). Robot shift from industrial production to social reproduction. In (Eds.), *Social robots from a human perspective*, 11–24. https://doi.org/10.1007/978-3-319-15672-9_2
19. Takayama, L., Ju, W., & Nass, C. (2008, March). Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 25–32). <https://doi.org/10.1145/1349822.1349827>
20. Gnambs, T., & Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Computers in Human Behavior*, 93, 53–61. <https://doi.org/10.1016/j.chb.2018.11.045>
21. Fast, E., & Horvitz, E. (2017, February). Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI conference on artificial intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10635>
22. Javaheri, A., Moghadamnejad, N., Keshavarz, H., Javaheri, E., Dobbins, C., Momeni-Ortner, E., & Rawassizadeh, R. (2020). Public vs media opinion on robots and their evolution over recent years. *CCF Transactions on Pervasive Computing and Interaction*, 2, 189–205. <https://doi.org/10.1007/s42486-020-00035-1>
23. Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2021). Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews. *International Journal of Hospitality Management*, 98, 103032. <https://doi.org/10.1016/j.ijhm.2021.103032>
24. Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, 40(1), 21–41. <https://doi.org/10.1177/0261927X20967028>
25. Gasiorek, J., Gallois, C., Pierson, H., Nussbaum, J. F., & Harwood, J. (2018). Advancing theory in language, communication, and intergroup relations. In J. Harwood, J. Gasiorek, H. D. Pierson, J. F. Nussbaum, & C. Gallois (Eds.), *Language, Communication, and Intergroup relations* (pp. 291–305). Routledge.
26. Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 269–322). McGraw-Hill.

27. Gruber, T., Bazhydai, M., Sievers, C., Clément, F., & Dukes, D. (2022). The ABC of social learning: Affect, behavior, and cognition. *Psychological Review*, 129(6), 1296–1318. <https://doi.org/10.1037/rv0000311>
28. Robbins, T. W. (2011). Cognition: The ultimate brain function. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 36(1), 1–2. <https://doi.org/10.1038/npp.2010.171>
29. Niven, K. (2013). Affect. In M. D. Gellman, & J. R. Turner (Eds.), *Encyclopedia of behavioral medicine*. Springer. https://doi.org/10.1007/978-1-4419-1005-9_1088
30. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>
31. Ortony, A., Clore, G. L., & Collins, A. (2022). *The cognitive structure of emotions* (2nd ed.). Cambridge University Press.
32. Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
33. Perlovsky, L. (2009). Language and cognition. *Neural Networks*, 22(3), 247–257. <https://doi.org/10.1016/j.neunet.2009.03.007>
34. Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 2). MIT Press.
35. Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24(6), 425–432. <https://doi.org/10.1177/0963721415604610>
36. Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLOS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0115844>. Article e115844.
37. Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science*, 26(5), 570–582. <https://doi.org/10.1177/0956797614566658>
38. Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, 201811987. <https://doi.org/10.1073/pnas.1811987116>
39. Ta, V. P., Boyd, R. L., Seraj, S., Keller, A., Griffith, C., Loggarakis, A., & Medema, L. (2022). An inclusive, real-world investigation of persuasion in language and verbal behavior. *Journal of Computational Social Science*, 5(1), 883–903. <https://doi.org/10.1007/s42001-021-00153-5>
40. Seraj, S., Blackburn, K. G., & Pennebaker, J. W. (2021). Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7), Article e2017154118. <https://doi.org/10.1073/pnas.2017154118>
41. Abe, J. A. A. (2012). Cognitive–affective styles associated with position on war. *Journal of Language and Social Psychology*, 31(2), 212–222. <https://doi.org/10.1177/0261927X12438532>
42. Davalos, S., Merchant, A., Rose, G. M., Lessley, B. J., & Teredesai, A. M. (2015). The good old days’: An examination of nostalgia in Facebook posts. *International Journal of Human-Computer Studies*, 83, 83–93. <https://doi.org/10.1016/j.ijhcs.2015.05.009>
43. Savela, N., Oksanen, A., Pellert, M., & Garcia, D. (2021). Emotional reactions to robot colleagues in a role-playing experiment. *International Journal of Information Management*, 60. <https://doi.org/10.1016/j.ijinfor.2021.102361>. Article 102361.
44. Savela, N., Garcia, D., Pellert, M., & Oksanen, A. (2021). Emotional talk about robotic technologies on Reddit: Sentiment analysis of life domains, motives, and temporal themes. *New Media & Society*. <https://doi.org/10.1177/14614448211067259>. Advance online publication.
45. Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology*, 57, 345–374. <https://doi.org/10.1146/annurev.psych.57.102904.190034>
46. Flanagin, A. J., Winter, S., & Metzger, M. J. (2020). Making sense of credibility in complex information environments: The role of message sidedness, information source, and thinking styles in credibility evaluation online. *Information Communication & Society*, 23(7), 1038–1056. <https://doi.org/10.1080/1369118X.2018.1547411>
47. Galperti, S. (2019). Persuasion: The art of changing worldviews. *American Economic Review*, 109(3), 996–1031. <https://doi.org/10.1257/aer.20161441>
48. Kumkale, G. T., & Albarracin, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin*, 130(1), 143–172. <https://doi.org/10.1037/0033-2909.130.1.143>

49. Peng, W., Lim, S., & Meng, J. (2023). Persuasive strategies in online health misinformation: A systematic review. *Information Communication & Society*, 26(11), 2131–2148. <https://doi.org/10.1080/1369118X.2022.2085615>
50. Jasser, J., Garibay, I., Scheinert, S., & Mantzaris, A. V. (2022). Controversial information spreads faster and further than non-controversial information in Reddit. *Journal of Computational Social Science*, 5(1), 111–122. <https://doi.org/10.1007/s42001-021-00121-z>
51. Ali, K., Li, C., & Muqtadir, S. A. (2022). The effects of emotions, individual attitudes towards vaccination, and social endorsements on perceived fake news credibility and sharing motivations. *Computers in Human Behavior*, 134, 107307. <https://doi.org/10.1016/j.chb.2022.107307>
52. Chen, S., Mao, J., Li, G., Ma, C., & Cao, Y. (2020). Uncovering sentiment and retweet patterns of disaster-related tweets from a spatiotemporal perspective—A case study of Hurricane Harvey. *Telematics and Informatics*, 47, 101326. <https://doi.org/10.1016/j.tele.2019.101326>
53. Chen, J., Yan, Y., & Leach, J. (2022). Are emotion-expressing messages more shared on social media? A meta-analytic review. *Review of Communication Research*, 10, 59–79. <https://doi.org/10.12840/ISSN.2255-4165.034>
54. Dujeancourt, E., & Garz, M. (2023). The effects of algorithmic content selection on user engagement with news on twitter. *The Information Society*. Advance online publication. <https://doi.org/10.1080/01972243.2023.2230471>
55. Weismueller, J., Harrigan, P., Coussement, K., & Tessitore, T. (2022). What makes people share political content on social media? The role of emotion, authority and ideology. *Computers in Human Behavior*, 129, 107150. <https://doi.org/10.1016/j.chb.2021.107150>
56. Xu, W. W., & Zhang, C. (2018). Sentiment, richness, authority, and relevance model of information sharing during social crises—the case of #MH370 tweets. *Computers in Human Behavior*, 89, 199–206. <https://doi.org/10.1016/j.chb.2018.07.04>
57. Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>
58. Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1
59. Petty, R. E., & Brinol, P. (2008). Persuasion: From single to multiple to metacognitive processes. *Perspectives on Psychological Science*, 3(2), 137–147. <https://doi.org/10.1111/j.1745-6916.2008.00071>
60. Ruiz, S., & Sicilia, M. (2004). The impact of cognitive and/or affective processing styles on consumer response to advertising appeals. *Journal of Business Research*, 57(6), 657–664. [https://doi.org/10.1016/S0148-2963\(02\)00309-0](https://doi.org/10.1016/S0148-2963(02)00309-0)
61. Rosselli, F., Skelly, J. J., & Mackie, D. M. (1995). Processing rational and emotional messages: The cognitive and affective mediation of persuasion. *Journal of Experimental Social Psychology*, 31(2), 163–190. <https://doi.org/10.1006/jesp.1995.1008>
62. Shin, D. (2023). Embodying algorithms, enactive artificial intelligence and the extended cognition: You can see as much as you know about algorithm. *Journal of Information Science*, 49(1), 18–31. <https://doi.org/10.1177/0165551520985495>
63. Shin, D. (2022). The perception of humanness in conversational journalism: An algorithmic information-processing perspective. *New Media & Society*, 24(12), 2680–2704. <https://doi.org/10.1177/1461444821993801>
64. Savela, N., Turja, T., Latikka, R., & Oksanen, A. (2021). Media effects on the perceptions of robots. *Human Behavior and Emerging Technologies*, 3(5), 989–1003. <https://doi.org/10.1002/hbe2.296>
65. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020, May). The push-shift reddit dataset. *Proceedings of the international AAAI conference on web and social media*, 14, 830–839. <https://doi.org/10.1609/icwsm.v14i1.7347>
66. Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and Ethics. *Social Media + Society*, 7(2), 1–14. <https://doi.org/10.1177/205630512111019004>
67. Pew Research Center (2021, April 7). Social Media Fact Sheet. <https://www.pewresearch.org/internet/fact-sheet/social-media/#panel-d6a80215-f4d8-4e90-b53d-909ef851e4b3-chart>
68. Aldous, K. K., An, J., & Jansen, B. J. (2022). What really matters? Characterising and predicting user engagement of news postings using multiple platforms, sentiments and topics. *Behaviour & Information Technology*. Advance online publication. <https://doi.org/10.1080/0144929X.2022.2030798>

69. Davis, J. L., & Graham, T. (2021). Emotional consequences and attention rewards: The social effects of ratings on Reddit. *Information Communication & Society*, 24(5), 649–666. <https://doi.org/10.1080/01369118X.2021.1874476>
70. Kim, J., Han, J., & Choi, D. (2023). Predicting continuity of online conversations on Reddit. *Telematics and Informatics*, 79, Article 101965. <https://doi.org/10.1016/j.tele.2023.101965>
71. Yadav, J., Misra, M., & Singh, K. (2022). Sensitizing Netizen's behavior through influencer intervention enabled by crowdsourcing—a case of reddit. *Behaviour & Information Technology*, 41(6), 1286–1297. <https://doi.org/10.1080/0144929X.2021.1872705>
72. Breur, T. (2016). Statistical Power Analysis and the contemporary crisis in social sciences. *Journal of Marketing Analytics*, 4(2–3), 61–65. <https://doi.org/10.1057/s41270-016-0001-3>
73. Ahteensuu, M. (2019, October 23). Do you use social media data in your research? Responsible Research. <https://vastuullinentiede.fi/en/doing-research/do-you-use-social-media-data-your-research>
74. Finnish National Board on Research Integrity TENK (2019). *The ethical principles of research with human participants and ethical review in the human sciences in Finland*. Publications of the Finnish National Board on Research Integrity TENK 3/2019. Finnish National Board on Research Integrity TENK guidelines 2019. https://www.tenk.fi/sites/tenk.fi/files/lhmistieteiden_eettisen_ennakkoarvioinnin_ohje_2019.pdf
75. DeVellis, R. F. (2005). Inter-rater reliability. In (Ed. Kempf-Leonard, K.) *Encyclopedia of Social Measurement*, (pp. 317–322). <https://doi.org/10.1016/B0-12-369398-5/00095-5>
76. Shaw, B. (2021). *OMEGACOEf: Stata module to calculate the omega reliability coefficient*. Statistical Software Components S458861. Boston College Department of Economics.
77. Altman, D. (1991). Inter-rater agreement. *Practical Statistics for Medical Research* (pp. 403–409). Chapman and Hall.
78. Lee, J., Kim, Y., & Zhu, X. (2023). Liked and shared tweets during the pandemic: The relationship between intrinsic message features and (mis) information engagement. *Behaviour & Information Technology*, Advance online publication. <https://doi.org/10.1080/0144929X.2023.2222192>
79. Simpson, B. (2015, June 11). Selfpost character limit is now 40,000 for all subreddits. Reddit. https://www.reddit.com/r/change/comments/39hf9x/reddit_change_selfpost_character_limit_is_now/
80. Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, 8(1), pp. 216–225.
81. Hutto, C. J. (n.d.). VADER-Sentiment-Analysis. README. <https://github.com/cjhutto/vaderSentiment/blob/94fb74594ac0714dfc0e55bd7415994b24020cf/README.rst>
82. Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15(10), 687–693. <https://doi.org/10.1111/j.0956-7976.2004.00741.x>
83. Monzani, D., Vergani, L., Pizzoli, S. F. M., Marton, G., & Pravettoni, G. (2021). Emotional tone, analytical thinking, and somatosensory processes of a sample of Italian tweets during the first phases of the COVID-19 pandemic: Observational study. *Journal of Medical Internet Research*, 23(10). <https://doi.org/10.2196/29820>. Article e29820.
84. Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22* (pp. 1–47). University of Texas at Austin.
85. Pennebaker Conglomerates (n.d.-a). (liwc-22-demo-data [Data set]. Retrieved February 9, 2023 from <https://www.liwc.app/help/workbench>
86. Provine, R. R., Spencer, R. J., & Mandell, D. L. (2007). Emotional expression online: Emoticons punctuate website text messages. *Journal of Language and Social Psychology*, 26(3), 299–307. <https://doi.org/10.1177/0261927X06303481>
87. Riordan, M. A. (2017). Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology*, 36(5), 549–567. <https://doi.org/10.1177/0261927X1704238>
88. Reddit (2023a). How does voting work on Reddit? Reddit Help, Reddit 101, Basics. Retrieved May 6, 2023, from <https://support.reddithelp.com/hc/en-us/articles/7419626610708-How-does-voting-work-on-Reddit->
89. Reddit (2023b). POST/api/vote. API Documentation. Retrieved May 6, 2023, from https://www.reddit.com/dev/api#POST_api_vote
90. Pennebaker Conglomerates (n.d.-b). *LIWC Analysis*. <https://www.liwc.app/help/liwc#Tone>
91. Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 799–821.

92. Schamari, J., & Schaefer, T. (2015). Leaving the home turf: How brands can use webcare on consumer-generated platforms to increase positive consumer engagement. *Journal of Interactive Marketing*, 30(1), 20–33. <https://doi.org/10.1016/j.intmar.2014.12.001>
93. Pezzuti, T., Leonhardt, J. M., & Warren, C. (2021). Certainty in language increases consumer engagement on social media. *Journal of Interactive Marketing*, 53(1), 32–46. <https://doi.org/10.1016/j.intmar.2020.06.005>
94. Kwon, S., & Park, A. (2023). Examining thematic and emotional differences across Twitter, Reddit, and YouTube: The case of COVID-19 vaccine side effects. *Computers in Human Behavior*, 144, 107734. <https://doi.org/10.1016/j.chb.2023.107734>
95. Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The role of social media content format and platform in users' engagement behavior. *Journal of Interactive Marketing*, 53(1), 47–65. <https://doi.org/10.1016/j.intmar.2020.05.001>
96. Hollingshead, W., Quan-Haase, A., & Blank, G. (2022). Representativeness and Bias in Social Media Research: Quantitative and qualitative approaches to Sampling. *The SAGE Handbook of Social Media Research Methods* (2nd ed., pp. 79–89). SAGE.
97. Lizama-Mué, Y., & Suárez, J. L. (2022). The Reddit data analysis pipeline for researchers. In A. Quan-Haase, & L. Sloan (Eds.), *The SAGE handbook of social media research methods* (2nd ed., pp. 546–567). SAGE.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.