Linguistic Frontiers • 8(3) • 2025
DOI: 10.2478/lf-2025-0020

Linguistic **Frontiers**

Sentiment analysis of cultural differences in online comments on popular news

Original Study

Kateryna Hordiienko, Libuše Kormaníková
Palacký University in Olomouc, Czech Republic
kateryna.hordiienko01@upol.cz (0000-0002-1049-7415), libuse.kormanikova01@upol.cz (0000-0001-8919-6748)

Received: 6 February 2025; Accepted: 26 November 2025

Abstract: The rapid growth of online communication through social networks has created new opportunities for understanding public opinion on socially relevant issues. This research examines how sentiment analysis (SA) can reveal cultural differences, specifically analyzing Czech and Ukrainian online comments on news topics including the war in Ukraine, political discussions, public health issues (tick-borne diseases, COVID-19), LGBTQ+ community matters, and natural disasters. By comparing three large language models (GPT-3.5-Turbo, Twitter-XLM-RoBERTa, and Zephyr 7B) with native speaker evaluations, we assess whether AI-based sentiment analysis can accurately capture culturally-specific emotional expressions in medium-resource languages. Our dataset comprises 6,085 comments (2,999 Czech from X/Twitter, 3,086 Ukrainian from Telegram) collected during 2023, focusing on socially relevant news coverage. We employed a hybrid methodology combining machine learning analysis with expert validation by native speakers. The study addresses a critical gap in cross-cultural sentiment analysis research, as no previous studies have compared Czech and Ukrainian linguistic patterns in this context. Results demonstrate significant performance differences among models depending on language: GPT-3.5-Turbo achieved highest accuracy for Czech ($p<0.001$), while all models performed comparably for Ukrainian. Both populations showed predominantly negative sentiment (Czech: 69.93%, Ukrainian: 68.93% via GPT-3.5), reflecting shared emotional responses to crisis events.

Keywords: sentiment analysis (SA), cultural aspect, artificial intelligence (AI), large linguistic models (LLMs), comments online

1. INTRODUCTION

The digital transformation of communication has fundamentally altered how public discourse unfolds, with social media platforms becoming primary spaces for opinion formation and expression. This shift presents both opportunities and challenges for understanding public sentiment across cultural boundaries. Online comments on news articles represent a particularly rich source of sentiment data, as they capture immediate emotional responses to socially relevant events and issues (Nemesh 2017; Yuna et al. 2022; Hordiienko, Joukl 2023).

The analysis of sentiment in online discourse has significant implications for cross-cultural understanding, particularly in an era marked by global conflicts, political polarization, and rapid information dissemination. Events such as the war in Ukraine, political debates, public health crises like tick-borne disease outbreaks, discussions surrounding LGBTQ+ rights, and coverage of war journalism generate intense emotional responses that vary considerably across cultural contexts. Misinterpreting these sentiment patterns can lead to inadequate policy responses, corporate missteps, and deepened cultural misunderstandings

(Yang et al. 2020; Wankhade et al. 2022; Robertson et al. 2023).

Sentiment analysis, as a computational approach to understanding emotional tone in text, has emerged as a critical tool for processing the vast volumes of opinion data generated on social media platforms (Liu 2012; Vidhya et al. 2021; Hordiienko, Joukl 2023). However, the effectiveness of current AI-based sentiment analysis tools across different languages and cultural contexts remains inadequately understood. This gap is particularly pronounced for medium-resource languages like Czech and Ukrainian, which have received less attention in sentiment analysis research compared to English.

Traditional approaches to sentiment analysis face specific challenges when applied to social media discourse during crisis events. Social media comments contain unique linguistic features including informal language, code-switching, irony, and culturally-specific expressions that standard natural language processing systems struggle to interpret accurately (Pang, Lee 2008; Sánchez-Rada, Iglesias 2019). During crisis events like wars, these challenges intensify as emotional expressions become more complex and culturally-charged. The rapid evolution of language during ongoing crises, emergence of new terminology, and heightened emotional states all contribute to the difficulty of accurate sentiment classification (Sánchez-Rada, Iglesias 2019; Hordiienko, Joukl 2023).

The present study addresses these challenges by comparing the performance of three large language models (GPT-3.5-Turbo, Twitter-XLM-RoBERTa, and Zephyr 7B) in analyzing sentiment in Czech and Ukrainian online comments on socially relevant news topics. By incorporating expert evaluation from native speakers, this research aims to assess the reliability of automated sentiment analysis across cultural contexts and identify the most effective approaches for cross-cultural sentiment analysis.

2. LITERATURE OVERVIEW

2.1. Related work and basics definitions

Sentiment Analysis (SA), also known as Opinion Mining (OM), automatically detects emotional polarity in text through computational methods (Liu 2012; Veselovská 2017). While often used interchangeably, SA specifically identifies and analyzes sentiment polarity (positive, negative, neutral), whereas OM focuses on extracting opinions about specific entities or aspects (Medhat et al. 2014). SA employs *natural language processing (NLP)* and *machine learning (ML)* techniques to extract emotional content from text data. Modern approaches utilize *Large Language Models (LLMs)* trained on vast corpora to understand context and nuance in human expression. These models process text through multiple layers of transformation, learning patterns that associate linguistic features with emotional states (Liu, Zhang 2012; Mao et al. 2024). Therefore, SA has become popular among researchers in *computational psycholinguistics* to process natural language, study the emotional state of the author, and assess the mood of society.

In SA, there are two main types of text classification: subjectivity detection and polarity detection (Veselovská 2017). *Subjectivity detection* involves determining opinion or fact, realism or unrealism of information, which is described in studies in psychology, communication and linguistics (Krippendorf 2004; Martin, White 2005; Hunston 2010). We define the expression of subjectivity in language as a positive, neutral, or negative opinion, referred to as *sentiment* (Taboada 2016), which is closely related to the field of emotion classification (surprise, anger, fear, sadness, disgust, joy, and so on (Bose et al. 2020; Hung, Alias 2023; Mao et al. 2024). However, our study focuses on emotional commentary texts and the cultural aspect of their evaluation, namely, the objectivity of texts within a particular culture. Therefore, we will not determine the subjectivity of the text of a particular subject, but will classify, that means, evaluate the polarity of the text in terms of positive, neutral or negative public opinion.

In contrast to subjectivity, *polarity* can be detected at the level of aspects, phrases, and sentences, which is appropriate for the size of the texts (commentaries) we investigate (Wankhade et al. 2022). Each sentence can contain multiple aspects, a polarity is assigned to all aspects, and the aggregate sentiment for the entire sentence is calculated (Schouten, Frasincar 2015). The term, at the phrase level, represents the demographic characteristics of individuals (gender, age, location) and desires, social position and personality, and cultural characteristics (Flek 2020). Sentence-level classification is related to subjective classification, and polarity will be determined with more training data and processing resources (Schouten, Frasincar 2015; Rao et al. 2018).

The scientific interest in the study of subjectivity and polarity has increased attention to the peculiarities of revealing opinions and emotions in *social media*, and the rapid development of the *internet* has generated a vast number of *online comments* (Taboada 2016; Mao et al. 2024). *Social data* (comments, age, nationality, etc.) is publicly available or semi-public information collected through an online application that allows users to share their mood, status, thoughts, experiences with their virtual social circle and the world (Boyd, Ellison 2007; Rosenthal et al. 2015; Pomytkina et al. 2021; Vidhya et al. 2021). That is, researchers confirm the idea of using the internet in the process of communication.

The internet discourse of speech in the process of communication has a peculiar linguistic phenomenon — in other words, it contains characteristics of spoken language (spontaneity, linearity, direct nature of the speech act, contextual conditionality and others), a written form of fixation (for example, posts, tweets, and comments to them) and is a way of self-expression (slang, jargon, abbreviations, changes in generally accepted rules of punctuation, syntax, and spelling) (Nemesh 2017; Anderwald 2018). Let's define the difference between virtual and real communication: 1) *virtual communication (VC)*

is multimodal (different ways of interaction, often asynchronous [not coinciding in time]) or in real time ('computer-mediated communication'); 2) VC satisfies the need for communication and simplifies it, but may affect the loss of skills of effective interaction and establishing contacts in real life (emotions, views, intonation, and so on); 3) it eliminates communication and psychological barriers by creating a "Cyber-I" and a social mask for personal protection, self-expression, or reducing responsibility for one's own actions; 4) the presence of rapid flamewars in online communication, where participants are polarized and fixated on their points of view (increased criticality towards others and decreased self-regulation due to a sense of impunity) (Ellison et al. 2007; Dahlberg, Bagga-Gupta 2014; Nemesh 2017; Hordiienko 2021). These features should be taken into account when collecting data and interpreting the results.

SA on social media can be divided into news comment analysis (Krishnamoorthy 2018; Aslan 2023; Müngen et al. 2020), product and service comment analysis (Sharma, Mishra 2016; Shelke et al. 2017; Pavitha et al. 2022), and other types (Xu et al. 2019). We study the sentiment of comments on *popular* or *extraordinary news* in the Czech Republic and Ukraine on a specific topic. These are news that evoke an emotional response from commenters, for instance, about anthropogenic events (disasters, catastrophes, wars, epidemics and others), socio-political or cultural events, and make them react more actively by becoming popular (Aslan 2023; Hordiienko, Joukl 2024).

X (formerly Twitter) is one of the most popular social media in the world and among the Czech population, also with a more open policy, relative ease of access, used to disseminate activities related to unusual events (armed conflicts, epidemics, etc.). The use of SA to measure reactions to any event discussed on Twitter contributes to the growing literature on SA of tweets spread during crises (Aslan 2023; Ramos, Chang 2023) and is often used to teach LLMs through data collection via API X (Twitter) (Hassonah et al. 2020). In contrast, *Telegram* channels are popular among the Ukrainian-speaking population, which play a significant role in informing about events, shaping public perceptions and creating opportunities for communication, and supports the use of news channels for a diverse audience (Karakikes et al. 2024). Based on the above, and since our research is based on the analysis of comments from two different nationalities, we use a collection of Ukrainian comments on posts from popular Telegram channels and Czech comments on tweets from network X (Hordiienko 2024).

Thus, the analysis of past research and the identification of basic definitions allow us to proceed to a thorough theoretical analysis of the cultural aspect of SA.

2.2. Cultural aspect of sentiment analysis

Culture is one of the major factors that influence language use (Boroditsky 2011). One of the original concepts in this regard is Sapir and Whorf's theory of linguistic relativism, especially its soft version, which states that language influences our thinking and perception of the world. Since we also express (or not) our emotions through speech (whether in the form of speech or text), our cultural background has a great influence on our understanding of the sentiment of speech. For example, Western cultures often emphasize individualism, which is reflected in the expressive and explicit expression of emotions, whereas collectivistic cultures, such as the Japanese, prefer implicit communication, where emotional nuances can be expressed indirectly or through context (Tao et al. 2024).

The research of Krugmann and Hartmann (2024) emphasizes that LLMs trained on multilingual or general datasets may still align heavily with the norms of high-resource languages, such as English, potentially leading to misinterpretation of sentiment in culturally specific expressions. As the complexity of sentiment expressions increases, such as those involving cultural references or irony, the need for task-specific fine-tuning becomes evident. Generic LLMs often lack the specialized knowledge to interpret these subtleties accurately.

In the context of SA using large-scale LLMs, this should therefore mean that models must consider not only linguistic structure but also cultural dynamics in order to accurately interpret the emotions in the text and avoid cultural bias. For instance, a study by Havaldar et al. (2023) found that models such as XLM-RoBERTa tend to align emotional representations of other languages to English norms, which leads to ignoring cultural nuances. Such alignment causes the emotional meaning of expressions in other languages to be inaccurately interpreted in terms of English meaning. As an illustration, the term "frustration" in English and Chinese not only differs linguistically but also carries different cultural connotations – while frustration may be perceived as a temporary setback in Anglophone culture, in Chinese culture it may be associated with loss of face. Another example, from Slavic languages, cultural differences in how reviews are written (such as directness in Slovene versus more nuanced expressions in Bulgarian) impact how well models can generalize across languages (Thakkar et al. 2024).

Although intonation is mainly related to spoken language, in written text, emotional charge can be expressed through exclamation points, question marks, and word choice and syntactic structures that indicate the emotional state of the speaker. That is, through the representation, construction, and integration of the meaning of what is written (Field 2004) and the properties of the organization of internet discourse, where specific forms of communication are noticeable due to the use of computers and the peculiarities of the speech behavior of internet users, changes in generally accepted punctuation rules and unwritten rules are revealed (e.g., young people consider sentences with a period to be offensive). Therefore, in addition to the differences between cultures, we can also observe differences in the ways of using language in online forms (Anderwald 2018). Besides, in sentiment analysis, it was crucial to fine tune LLM or

dataset augmentation according to language, because low resource language indicates differences in analyzing sentiment (Thakkar et al. 2024; Prytula 2024). So far there are no studies comparing Ukrainian and Czech, thus, this paper is filling the gap in research in this direction.

Therefore, it is necessary to identify the basic results and problems of the pilot study to further understand the peculiarities of the cultural context and how we can address these issues in the main research.

2.3. Pilot study

Our pilot study (Hordiienko, Joukl 2024) employed the following methodology:

- The preliminary investigation reflects the initial procedures of the main research and confirms the feasibility of the study by evaluating the data inclusion and exclusion criteria (popular Czech and Ukrainian news in social media in 2023, comments on them), preparation of tools (Python, LLMs, social media), storage and testing of instruments used for measurements in the study (SA), interpretation of news and comments (positive, negative, neutral). The research stages and implementation, which included data collection, data cleaning, topical sorting (using GPT_TAG in Sheets and XLM-RoBERTa pipeline) with annotation scripts, system prompts and analysis code, are available in article[1].
- All data was anonymized, removing usernames, timestamps, and identifying information. Participants in the validation study provided informed consent.
- Accordingly, in the pilot study 2 models (Twitter-XLM-RoBERTa model and GPT-3.5-Turbo-0125 model) were tested for SA, which were powered by Python and processed 43 Czech Tweets from X and 38 Ukrainian posts from Telegram and 3000 Czech and Ukrainian comments each. Both LLMs evaluated the majority of Czech (GPT-3.5 – 69.93%, RoBERTa – 57.76%) and Ukrainian (GPT-3.5 – 68.93%, RoBERTa – 57.93%) comments as negative. The dominance of a negative tone is also evident in the results of the SA of Ukrainian news posts (GPT-3.5 – 65.8%, RoBERTa – 50%) and Czech tweets (GPT-3.5 – 81.4%, RoBERTa – 67.4%).
- The results of the pilot study indicate a moderate level of agreement between the model evaluations. Cohen's Kappa (0.467), Fleiss' Kappa (0.460), and Krippendorff's Alpha (0.461) coefficients indicate stable but not complete agreement between raters (Table 1). The highest agreement is observed for category -1, while in cases #N/A (compound comments) there is no agreement. Overall, the models show agreement in most cases; however, notable discrepancies remain, indicating the need for further refinement of evaluation criteria or model improvement.

Cohen's Unweighted kappa

			95% CI	
Ratings	Unweighted kappa	SE	Lower	Upper
Average kappa	0.467			
RoBERTa – GPT-3.5	0.467	0.014	0.439	0.495

Note. 3000 subjects/items and 2 raters/measurements. Confidence intervals are asymptotic.

Fleiss' kappa

			95% CI	
Ratings	Fleiss' kappa	SE	Lower	Upper
Overall	0.460	0.014	0.434	0.487
-1	0.508	0.018	0.472	0.544
0	0.407	0.018	0.371	0.443
1	0.446	0.018	0.410	0.482
#N/A	0.000	0.018	-0.036	0.036

Note. 3000 subjects/items and 2 raters/measurements. Confidence intervals are asymptotic.

Cohen's Unweighted kappa

			95% CI	
Method	Krippendorff's kappa	SE	Lower	Upper
Nominal	0.461	0.014	0.431	0.486

Note. 3000 subjects/items and 2 raters/measurements.

Table 1 *Inter-Rater Agreement Metrics between Models: Cohen's Kappa, Fleiss' Kappa, and Krippendorff's Alpha*

1 Hordiienko, K., Joukl, Z., 2024. Sentimental reflection of global crises: Czech and Ukrainian views on popular events through the prism of internet commentary. Jazykovedný Časopis, 75(1), 43–61, available at: https://doi.org/10.2478/jazcas-2024-0027.

- During the investigation, several problems were identified. At the stage of interpreting the results, it was found that Twitter-XLM-RoBERTa identified the same number of positive and neutral tweets in Czech tweets, which demonstrates the ambiguity of identifying tone by this model. At the same time, the model identified more positive and fewer negative responses compared to GPT-3.5-Turbo results, which may be due to different training data of these models (Twitter-XLM-RoBERTa has less training data than GPT-3.5-Turbo, which should be better at understanding the language). Other important errors found were that the GPT-3.5-Turbo model was trained on data up to 2021 at the time of the study (temporal aspect), and the Twitter-XLM-RoBERTa measurements revealed a problem with the interpretation of irony, humor and profanity (cultural aspect). The last problem in the pilot study is the interpretation of news and comments without considering the cultural characteristics of users, which can provide a general misunderstanding of sentiment, that is, it does not allow us to determine the reliability of the model for a particular language, taking into account the context.

The findings provided the impetus for studying the cultural aspect, delving deeper into the methodology and conducting the main research.

2.4. Research questions and objectives of the current study

This article seeks to answer the following research questions:

1. How do the sentiments of comments on social media differ from a sociocultural point of view between Czech and Ukrainian internet users? Which model best describes each sentiment?
2. Does manual classification of cultural media provide more information for text analysis? Can cultural context improve or impair SA?
3. Can a sentiment analyser based on AI and ML models be trusted to understand the current context and sentiment associated with an event? Which model is most accurate for diagnosing the sentiment of each culture?

Based on the defined questions, the goals of the research are the following answers corresponding to the individual stages of the research: 1) establishing the general results of the tonality of Czech and Ukrainian comments using three ML classifiers to compare the accuracy of calculations (Zephyr 7B, Twitter-XLM-RoBERTa, and GPT-3.5-Turbo); 2) determine the results of manual classification by experts of comments that received different tonality from 3 models at the 1st stage; 3) substantiation of the importance of the cultural aspect when using sentiment machine classifiers in SA and its comparison with the tonality of comments by expert speakers of Ukrainian and Czech culture in popular news.

3. METHODOLOGY

Based on the defined research questions and objectives, the approach for the experimental part was a hybrid strategy based on machine analysis of AI models and lexical analysis of a current comment by a native speaker of a particular culture. As we stated at the section 2.3 pilot study, the methodology of obtaining comments, data cleaning, anonymization and training of the models is detailedly described in previous article (Hordiienko, Joukl 2024). Briefly, in the initial pilot stage, over 3,000 Czech and Ukrainian comments (3,086 Ukrainian and 2,999 Czech) were manually collected from Twitter and Telegram, as previously mentioned. Irrelevant or sensitive data – such as names, dates, and punctuation – was removed to ensure anonymity, and only the original text content was retained. These texts were then classified by topic and keywords; in total, 81 news topics were processed and grouped into 41 thematic categories. SA was performed using three machine learning models: GPT-3.5-Turbo, Twitter-XLM-RoBERTa, and Zephyr 7B (i.e. *the machine analysis stage).*

- The first model (GPT-3.5-Turbo) was trained on a large amount of data from the web collection and social media in the period 2022–2023, so the study used the Google Sheets API through a function with three possible preset results and examples of comment scores, namely, tokenization and decoding (Kheiri, Karimi 2023; Ye et al. 2023).
- The second model (Twitter-XLM-RoBERTa), which is derived from BERT and trained on Twitter data, is used through Python library transformers, to be specific, pipeline and tokenizer, to obtain the score in an xlsx file normalised into three sentiment categories (positive [1], neutral [0], negative [-1]) (Barbieri et al. 2022; Matlach 2023; Tan et al. 2023).
- The third model, Zephyr 7B, although outperforming many of the 70B models for SA, was trained only in English, which prompted us to use the DeepL English translator first and then estimate sentiment through the model (Hordiienko 2024; Vergho et al. 2024).

3.1. Experimental Design

For the purposes of this study, particular attention was given to comments where the sentiment analysis models produced divergent results (e.g., GPT-3.5-Turbo rated a comment +1, Zephyr 7B rated it -1, and Twitter-XLM-RoBERTa rated it 0). Such ambiguous cases – where sentiment is difficult for models to evaluate – are often referred to in the literature as compound comments (Liu, 2012). In total, 157 such comments were identified in the Czech dataset and 94 in the Ukrainian dataset. From these, 20 comments per language were randomly selected to assess sentiment classification accuracy through human evaluation by native speakers. These comments were (again randomly) divided into two Google Forms (Forms A and B), which were then

distributed to native speakers for evaluation (i.e., *lexical analysis stage*).

In the first section of each form, participants were asked to provide basic demographic information, including confirmation of their native speaker status and whether they were long-term residents of the respective country. They also confirmed that they were of legal adult age and provided informed consent to participate in the study. No participants were excluded from either language group. Additional demographic data were not collected, as the primary objective of this study was not to analyze multiple variables, but rather to compare expert (i.e., native speaker) sentiment evaluations with the outputs of three different models in order to determine which model, if any, most closely aligned with human judgment.

Given the method of questionnaire distribution via social media, it is acknowledged that the participant pool may have been relatively homogeneous. However, due to the nature of the material being evaluated – namely, online comments – broad demographic diversity was not considered essential. Also, we did not post both Forms in the same groups or pages, and the participants were instructed not to fill the form twice.

For the Czech dataset, Form A received 40 responses and Form B received 42. For the Ukrainian dataset, Form A received 41 responses and Form B received 42. Participants were instructed to classify each comment as positive, neutral, or negative. Formal ethics committee approval was not required for this study design. Ethical standards were maintained through several measures: complete anonymization of both comment content and participant data, voluntary participation with explicit consent, transparent communication about research purposes, and provision of researcher contact information for any participant questions or concerns.

The resulting participant data were compared with the model outputs, and sentiment evaluation accuracy was calculated as a percentage of agreement. Descriptive statistics were used for the distribution of SA by tone, the coincidence of sentiments of comments in different models and for comparing assessments of Czech- and Ukrainian-speaking participants' assessments with LMMs' results. The description of SA tone through all models was done to help visualize overall SA for the readers. The ANOVA was used to understand the influence of factors (models and comments) on sentiment determination for Czech and Ukrainian, the refinement method Post-hoc tests to determine differences between models.

4. RESULTS

4.1. Descriptive Statistics: Sentiment analysis of all comments

The analysis of Czech comments on popular news from X revealed that GPT-3.5-Turbo rated 2,097 comments as negative, representing the highest count among the models. The second model favouring negative sentiment was Twitter-XLM-RoBERTa, which classified 1,738 comments as negative. Zephyr 7B, however, stood out by assigning the highest number of neutral ratings (1,309 comments) and a lower number of negative ratings (1,209 comments). Besides, the analysis of Ukrainian comments from Telegram showed that GPT-3.5-Turbo classified 2,142 comments as negative, which is the highest rate among the models. Twitter-XLM-RoBERTa classified 1,860 comments as negative, and Zephyr 7B classified 1,538 comments, which is slightly more than the number of neutral comments (1,241). Notably, none of the models showed a preference for positive sentiment over neutral sentiment. GPT-3.5-Turbo in Czech comments (only 280 comments), as well as in the Ukrainian comments (only 212 comments). A visualisation of the results is presented in Table 2 (i.e. *machine analysis stage*).

Sentiment	**Czech Comments**			**Ukrainian Comments**		
Model	negative	neutral	positive	negative	neutral	positive
GPT-3.5-Turbo	2097	620	280	2142	720	212
Twitter-XLM-RoBERTa	1738	665	594	1860	881	344
Zephyr 7B	1209	1309	479	1538	1241	307

Table 2 *Distribution of SAs by tone and LLMs of Czech and Ukrainian comments*

However, despite this low score of positive sentiment, the other SA models for comments demonstrated a percentage agreement with the GPT-3.5-Turbo in its identification. All three models agreed on positive sentiment in 156 cases (Czech comments), and in 92 cases (Ukrainian comments). The Czech data further indicate that GPT-3.5-Turbo aligns more closely with Twitter-XLM-RoBERTa (1,166 comments) than with Zephyr 7B (1,291 comments). Moreover, Twitter-XLM-RoBERTa and Zephyr 7B often disagree on sentiment classification (673 matches in total), highlighting that Zephyr 7B diverges significantly in its sentiment scoring compared to the other two models. In the SA of the Ukrainian comments, GPT-3.5-Turbo and Twitter-XLM-RoBERTa are the most similar in positive (61 comments) and negative (539 comments), while Zephyr 7B is the same as Twitter-XLM-RoBERTa and GPT-3.5-Turbo in neutral ratings. Possible explanations for Zephyr 7B's tendency to assign neutral ratings as the most

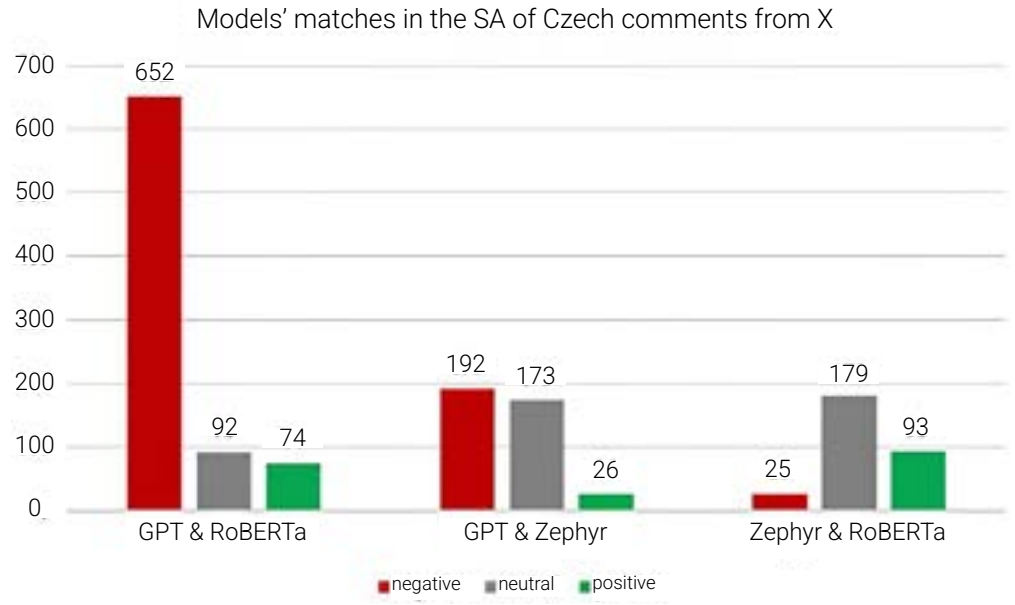

Figure 1 *Sentimental models' matches of Czech comments on popular news X (Twitter) network*

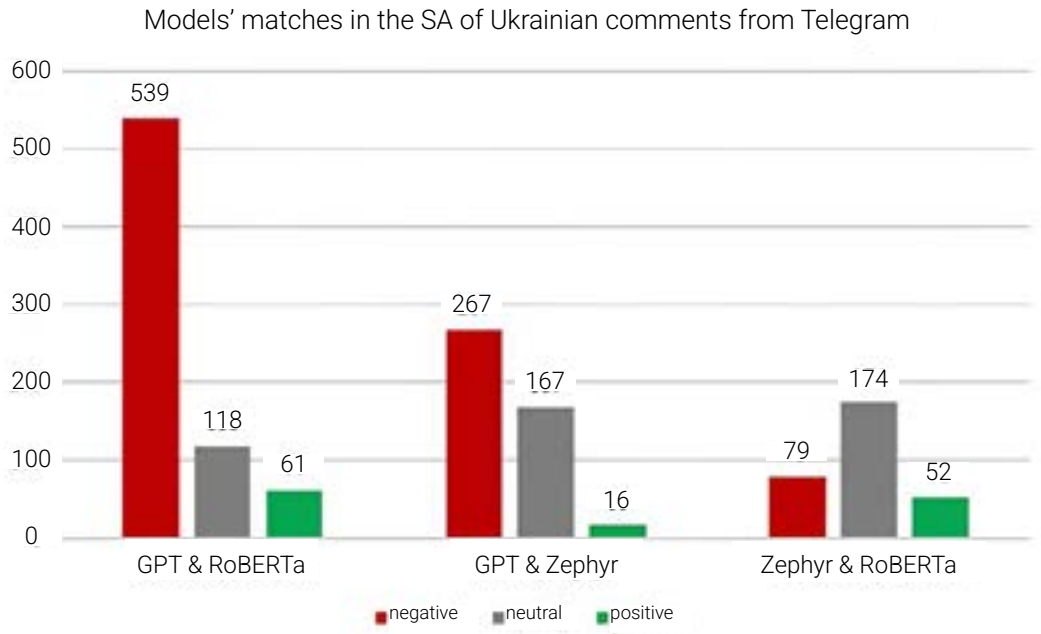

Figure 2 *Sentimental models' matches of Ukrainian comments on popular news Telegram network*

frequent classification will be explored in the Discussion section. Figures 1 and 2 illustrate the data obtained.

Understanding the results of the descriptive statistics and the general tendencies of the LLMs' evaluation of all comments gave rise to a more specific analysis of compound comments, in other words, those with different sentiment results for LLMs according to the interference statistics.

4.2. Inferential statistics

Sentiment analysis of compound comments (Czech dataset)

As shown in Table 3, the choice of model plays a crucial role in determining sentiment in Czech. The results were highly significant ($p<0.001$). In contrast, the factor Comments was not significant ($p=1.000$), indicating no meaningful differences among the various comment contents. Post-hoc analyses (Table 4) were conducted to identify the most appropriate model for sentiment determination. These comparisons revealed that GPT-3.5-Turbo was significantly different from the other two models:

- GPT-3.5-Turbo vs. Twitter-XLM-RoBERTa ($p<0.001$)
- GPT-3.5-Turbo vs. Zephyr 7B ($p<0.001$)

In contrast, no significant difference was found between Twitter-XLM-RoBERTa and Zephyr 7B ($p=0.681$).

Post-hoc analyses revealed:

	SS	df	F	p	η^2p
ANOVA Omnibus tests					
Model	1.41	21	1.33	0.217	0.424
Comments	6.67e-7	19	6.95e-7	1.000	0.000
LLM	1.41	2	13.97	<.001	0.424
Residuals	1.92	38			
Total	3.33	59			

Table 3 *Results of ANOVA for Czech*

- GPT-3.5-Turbo vs. Twitter-XLM-RoBERTa: p<0.001, Cohen's d = 0.84 (large effect)
- GPT-3.5-Turbo vs. Zephyr 7B: p<0.001, Cohen's d = 0.91 (large effect)
- Twitter-XLM-RoBERTa vs. Zephyr 7B: p=0.681, Cohen's d = 0.12 (negligible effect)

The boxplot analysis, combined with the results of the ANOVA, provides insights into the performance variability and effectiveness of the evaluated models. The ANOVA results confirm that GPT-3.5-Turbo significantly outperforms the other two models, Twitter-XLM-RoBERTa and Zephyr 7B, in terms of the evaluated metric (p<0.001). This finding is consistent with the boxplot visualization (Figure 3), which highlights GPT-3.5-Turbo's higher median and wider interquartile range, indicating both superior performance and greater variability in outcomes. The presence of outliers in certain models underscores specific instances where performance deviates from the expected range, possibly reflecting edge cases or scenarios where the models underperform. These observations suggest that GPT-3.5-Turbo may be the most effective model for the given task. Twitter-XLM-RoBERTa could be an option if consistency (smaller spread of results) is preferred.

Comparison								
LLM	**LLM**	**Difference**	**SE**	**t**	**df**	**p**	**$p_{bonferroni}$**	**p_{tukey}**
Post Hoc Comparisons – LLM								
GPT-3.5	TWITTER-XLM-RoBERTa	0.3390	0.0711	4.770	38.0	<.001	<.001	<.001
GPT-3.5	Zephyr 7B	0.3096	0.0711	4.356	38.0	<.001	<.001	<.001
TWITTER-XLM-RoBERTa	Zeohyr 7B	-0.0294	0.0711	-0.414	38.0	0.681	1.000	0.910

Table 4 *Post-hoc tests of differences between models (CZ)*

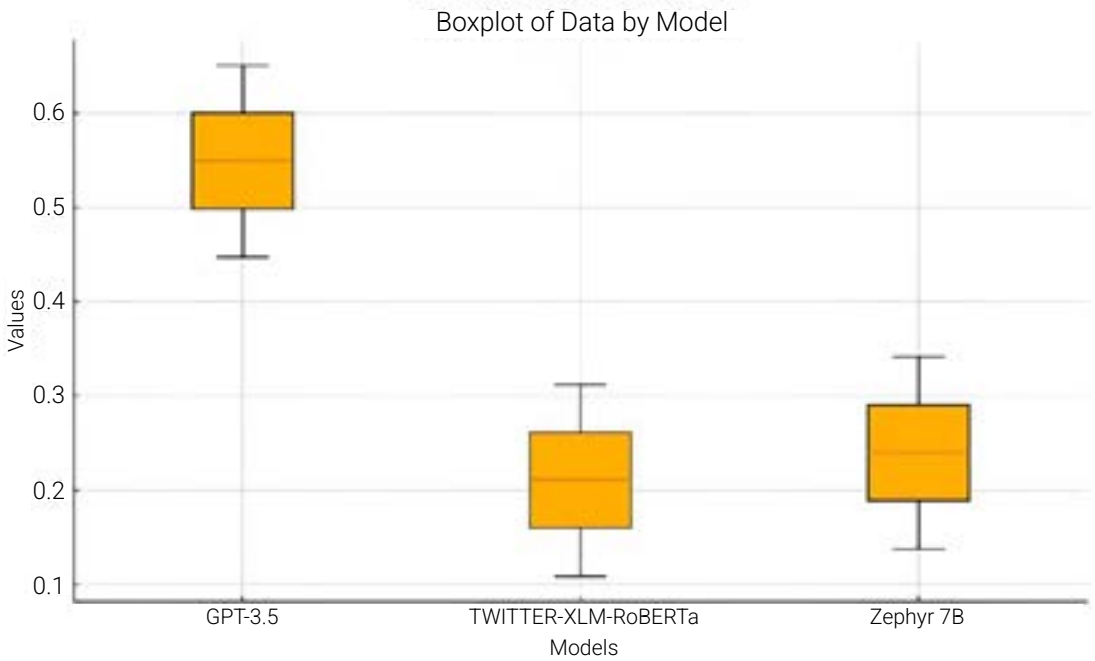

Figure 3 *Boxplot of model performance (CZ)*

Sentiment analysis of compound comments (Ukrainian dataset)

In contrast to the Czech dataset, the statistical analysis revealed no significant differences in performance among the evaluated language models: GPT-3.5-Turbo, Twitter-XLM-RoBERTa, and Zephyr 7B. As you can see in the Table 5, the ANOVA results indicated that the choice of LLM had no meaningful impact on the outcomes (p=0.916). Post-hoc comparisons supported this finding, showing no statistically significant differences between any pair of models (Table 6). The factor Comments did not show any significant effect (p=1.000), indicating that the content of the comments does not influence the observed outcomes.

Statistical analysis of the Ukrainian dataset revealed no significant performance differences among the three language models (p=0.916). Post-hoc comparisons confirmed no statistically significant differences between any model pairs.

Additionally, the estimated marginal means for the models were similar (GPT-3.5-Turbo 0.350, Twitter-XLM-RoBERTa 0.332, Zephyr 7B 0.318), with overlapping confidence intervals further reinforcing the lack of significant differences (visualization in boxplot Figure 4).

	SS	df	F	p	η^2p
ANOVA Omnibus tests					
Model	0.0101	21	0.00841	1.000	0.005
Comments	2.00e-6	19	1.85e-6	1.000	0.000
LLM	0.0101	2	0.08832	0.916	0.005
Residuals	2.1626	38			
Total	2.1727	59			

Table 5 *Results of ANOVA for Ukrainian*

Comparison								
LLM	**LLM**	**Difference**	**SE**	**t**	**df**	**p**	$p_{bonferroni}$	p_{tukey}
Post Hoc Comparisons – LLM								
GPT-3.5	TWITTER-XLM-RoBERTa	0.0175	0.0754	0.231	38.0	0.818	1.000	0.971
GPT-3.5	Zephyr 7B	0.0317	0.0754	0.420	38.0	0.677	1.000	0.908
TWITTER-XLM-RoBERTa	Zeohyr 7B	0.0142	0.0754	0.188	38.0	0.852	1.000	0.981

Table 6 *Post-hoc tests of differences between models*

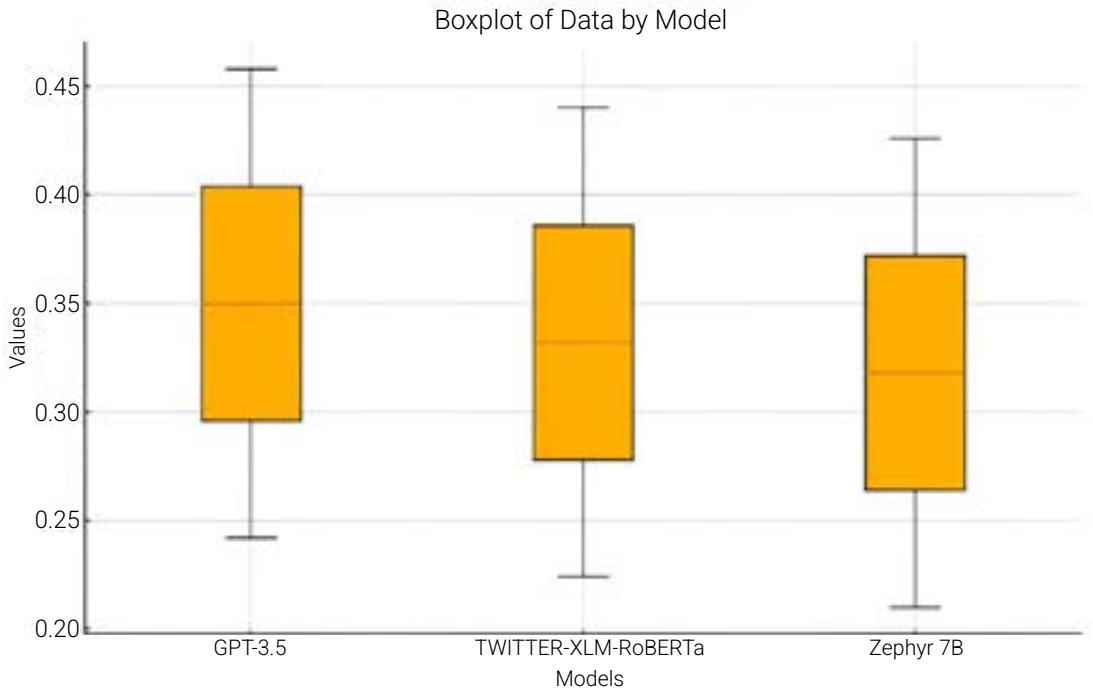

Figure 4 *Boxplot of model performance (UA)*

5. DISCUSSION

The performance differences observed across models and languages reveal important insights into cross-cultural sentiment analysis challenges and opportunities. Despite being designed for social media sentiment analysis and trained on Twitter data, Twitter-XLM-RoBERTa did not demonstrate superior performance in either language context. While fine-tuned versions exist for specific languages, we employed the standard multilingual version to assess whether social media training provided inherent advantages – a hypothesis our results do not support, particularly for Czech content.

Zephyr 7B's requirement for translation to English introduced potential complications, as emotional nuances and idiomatic expressions may be lost in translation. However, the model's performance was not as compromised as anticipated. Its architecture appears well-suited for informal, short texts typical of online comments. With 7 billion parameters compared to GPT-3.5-Turbo's 175 billion, Zephyr 7B offers computational efficiency advantages (Tunstall et al. 2023). Research suggests that Zephyr 7B tends to classify ambiguous emotional content as neutral due to limited emotional range compared to larger models, which may explain its high neutral classification rates in our study (Amirizaniani et al. 2024; Shaik et al. 2024).

However, on the Czech dataset, GPT-3.5-Turbo significantly outperformed the other models in terms of sentiment analysis (SA) scores, which were more closely aligned with user assessments. This discrepancy could also stem from translation effects, where emotional nuances and idiomatic expressions may be lost. The superior performance of GPT-3.5-Turbo can likely be attributed to its larger training corpus and stronger multilingual capabilities, which proved advantageous for SA tasks on the Czech data. Although the model is primarily tuned for broad, general-purpose applications, its sentiment scores surpassed those of the other two models, demonstrating closer alignment with human evaluations. Furthermore, its advantage may also lie in the incorporation of reinforcement learning from human feedback (RLHF) during the training process.

The contrasting results between Czech and Ukrainian datasets highlight the importance of linguistic and cultural factors in sentiment analysis. For Ukrainian, no significant model differences emerged, suggesting that all three models face similar challenges with this language. Ukrainian's status as a lower-resource language in many NLP contexts means all models, including GPT-3.5-Turbo, likely received limited exposure to Ukrainian during training. The cultural dimension of these findings may reflect different patterns of emotional expression and online communication norms. Czech online discourse might contain more complex linguistic constructs, cultural references, or subtle emotional cues that require deeper linguistic understanding, giving larger, more sophisticated models an advantage. Alternatively, Ukrainian online communication might employ more direct emotional expression patterns, reducing the advantage of more complex models. These findings challenge Krugmann and Hartmann's (2024) assertion about GPT-3.5-Turbo's inadequacy for social media sentiment analysis, at least for certain language contexts. As Havaldar et al. (2023) highlighted, models like XLM-RoBERTa often align emotional representations in other languages with English norms, leading to cultural misinterpretations and inaccuracies in emotional meaning.

Furthermore, the observation that only 94 (3%) Ukrainian comments versus 157 (5.2%) Czech comments generated model disagreement suggests possible differences in linguistic complexity or emotional expression patterns. The relative absence of ironic content (based on expert assessment) and potential platform-specific communication norms may also contribute to these patterns. The study reveals that sentiment analysis effectiveness is highly dependent on both language-specific factors and cultural communication patterns. For practitioners, this suggests that model selection should consider the specific linguistic context rather than relying on ostensibly specialized tools that may not perform better for target languages.

Due to copyright and ethical considerations, we provide aggregated results rather than individual comments. Researchers requiring access to specific examples must request permission, ensuring compliance with platform terms of service and user privacy.

6. CONCLUSION

This research demonstrates that sentiment analysis performance varies significantly across cultural and linguistic contexts, with implications for both theoretical understanding and practical applications. This study contributes to understanding cross-cultural sentiment analysis by demonstrating that model effectiveness cannot be assumed to generalize across languages, even within related language families.

Responding to our research questions: (1) Sentiment patterns showed consistent negativity dominance across both Czech and Ukrainian comments, with GPT-3.5-Turbo and Twitter-XLM-RoBERTa better detecting negative sentiment while Zephyr 7B showed preference for neutral classifications. Model effectiveness varied by language, with significant differences emerging for Czech but not Ukrainian content. (2) Manual classification by cultural experts provided crucial insights into subtle emotional expressions, cultural context, and pragmatic elements (sarcasm, irony, implicit meanings) often missed by automated systems. However, expert evaluation requires careful consideration of inter-rater reliability and representative sampling. (3) Sentiment analyzer reliability depends critically on cultural and linguistic context. For Czech content, GPT-3.5-Turbo demonstrated superior performance, while Ukrainian analysis showed no significant model differences, suggesting potential suboptimal performance across all models for this language.

Future research should explore fine-tuned models specifically trained for target languages and cultural contexts. Additionally, expanding the expert evaluation

component with larger samples and inter-rater reliability assessments would strengthen methodological rigor. Investigation of other Slavic language pairs could further illuminate patterns observed in Czech-Ukrainian comparisons. Our study focused on two Slavic languages during a specific crisis period. Generalization to other languages or non-crisis contexts requires further research.

ACKNOWLEDGEMENTS

1. The authors would like to express their sincere thanks to the anonymous participants in the manual evaluation of comments (culture experts, native speakers) for their useful contribution to the research. The authors are also grateful for the support of the Department of General Linguistics, Faculty of Arts, Palacký University, and for expert comments from the editors and reviewers for revising this article.
2. This publication was made possible thanks to targeted funding provided by the Czech Ministry of Education, Youth and Sports for specific research, granted in 2023 to Palacký University Olomouc (IGA_FF_2023_029).
3. The authors would like to thank the IGA project mentor Mgr. Kateřina Lesch, Ph.D., Mgr. Vladimír Matlach, Ph.D., who kindly provided us with a Python script that uses the multilingual Twitter-XLM-RoBERTa-base model, PhDr. Daniel Dostál, PhD., who kindly consulted the statistics and M.A. Israel Chávez, Ph.D for his valuable insights.

REFERENCES

Amirizaniani, M., Martin, E., Sivachenko, M. et al., 2024, October. Can LLMs reason like humans? Assessing theory of mind reasoning in LLMs for open-ended questions. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, pp. 34–44.

Anderwald, L., 2018. Language change and cultural change: The grammaticalization of the GET-passive in context. Language & Communication, 62, 1–14.

Aslan, S., 2023. A deep learning-based sentiment analysis approach (MF-CNN-BILSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. Applied Soft Computing, 143, 110404, available at: https://doi.org/10.1016/j.asoc.2023.110404.

Barbieri, F., Espinosa Anke, L., Camacho-Collados, J., 2022. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, pp. 258–266, available at: https://aclanthology.org/2022.lrec-1.27.

Boroditsky, L., 2011. How language shapes thought. Scientific American, 304(2), 62–65.

Boyd, D. M., Ellison, N. B., 2007. Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210–230, available at: https://doi.org/10.1111/j.1083-6101.2007.00393.x.

Bose, R., Dey, R. K., Roy, S., Sarddar, D., 2020. Sentiment analysis on online product reviews. In: Tuba, M., Akashe, S., Joshi, A. (Eds.), Information and Communication Technology for Sustainable Development (Advances in Intelligent Systems and Computing, Vol. 933). Singapore: Springer, pp. 559–569, available at: https://doi.org/10.1007/978-981-13-7166-0_56.

Dahlberg, G. M., Bagga-Gupta, S., 2014. Understanding global learning spaces: An empirical study of languaging and transmigrant positions in the virtual classroom. Learning, Media and Technology, 39(4), 468–487, available at: https://doi.org/10.1080/17439884.2014.931868.

Ellison, N., Steinfield, C., Lampe, C., 2007. The benefits of Facebook "friends": Exploring the relationship between college students' use of online social networks and social capital. Journal of Computer-Mediated Communication, 12(3), article 1, available at: http://jcmc.indiana.edu/vol12/issue4/ellison.html.

Field, J., 2004. Psycholinguistics: The key concepts. Routledge.

Flek, L., 2020. Returning the N to NLP: Towards contextually personalized classification models. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7828–7838, available at: https://doi.org/10.18653/v1/2020.acl-main.700.

Havaldar, S., Singhal, B., Rai, S. et al., 2023, July. Multilingual language models are not multicultural: A case study in emotion. In: Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Association for Computational Linguistics, pp. 202–214, available at: https://doi.org/10.18653/v1/2023.wassa-1.19.

Hassonah, M. A., Al-Sayyed, R., Rodan, A. et al., 2020. An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter. Knowledge-Based Systems, 192, 105353, available at: https://doi.org/10.1016/j.knosys.2019.105353.

Hordiienko, K., 2024. AI and human sentiment analysis with cultural aspect. Poster presented at the 21st International Congress of Linguists, Poznań, Poland.

Hordiienko, K., 2021. Moral disengagement as a factor of cyberbullying among students. Habitus, (29), 147–151, retrieved from: http://nbuv.gov.ua/UJRN/habit_2021_29_27.

Hordiienko, K., Joukl, Z., 2023, August 11–20. Sentiment analysis of different nationalities' internet comments on extraordinary news: cultural aspect. České Budějovice, Czech Republic: Summer School of Linguistics, retrieved from: https://ssol.ff.cuni.cz/summer-school-of-linguistics/ssol-2023/.

Hordiienko, K., Joukl, Z., 2024. Sentimental reflection of global crises: Czech and Ukrainian views on popular events through the prism of internet commentary. Jazykovedný Časopis, 75(1), 43–61, available at: https://doi.org/10.2478/jazcas-2024-0027.

Hung, L. P., Alias, S., 2023. Beyond sentiment analysis: A review of recent trends in text-based sentiment analysis and emotion detection. Journal of Advanced Computational Intelligence and Intelligent Informatics, 27(1), 84–95, available at: https://doi.org/10.20965/jaciii.2023.p0084.

Hunston, S., 2010. Corpus approaches to evaluation: Phraseology and evaluative language. Routledge.

Karakikes, A., Alexiadis, P., Kotis, K., 2024. Bias in X (Twitter) and Telegram-Based Intelligence Analysis: Exploring Challenges and Potential Mitigating Roles of AI. SN Computer Science, 5(5), 574, available at: https://doi.org/10.1007/s42979-024-02935-w.

Kheiri, K., Karimi, H., 2023. SentimentGPT: Exploiting GPT for advanced sentiment analysis and its departure from current machine learning. arXiv preprint, arXiv:2307.10234.

Krippendorf, K., 2004. Content analysis: An introduction to its methodology. Sage Publications.

Krishnamoorthy, S., 2018. Sentiment analysis of financial news articles using performance indicators. Knowledge and Information Systems, 56(2), 373–394, available at: https://doi.org/10.48550/arXiv.1811.11008.

Krugmann, J. O., Hartmann, J., 2024. Sentiment analysis in the age of generative AI. Customer Needs and Solutions, 11(3), available at: https://doi.org/10.1007/s40547-024-00143-4.

Liu, B., 2012. Sentiment analysis and opinion mining. Morgan & Claypool Publishers, available at: https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

Liu, B., Zhang, L., 2012. A survey of opinion mining and sentiment analysis. In: Aggarwal, C. C., Zhai, C. (Eds.), Mining text data. Springer, pp. 415–463, available at: https://doi.org/10.1007/978-1-4614-3223-4_13.

Mao, Y., Liu, Q., Zhang, Y., 2024. Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University – Computer and Information Sciences, 36(4), Article 102048, available at: https://doi.org/10.1016/j.jksuci.2024.102048.

Martin, J. R., White, P. R., 2005. The language of evaluation, vol. 2. Basingstoke: Palgrave Macmillan.

Matlach, V., 2023. Úvod do zpracování dat 1 [Introduction to data processing 1]. Olomouc: Univerzita Palackého v Olomouci, VUP.

Medhat, W., Hassan, A., Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113, available at: https://doi.org/10.1016/j.asej.2014.04.011.

Müngen, A. A., Aygün, İ., Kaya, M., 2020. Finding the relationship between news and social media users' emotions in the COVID-19 process. Sakarya University Journal of Computer and Information Sciences, 3(3), 250–263, available at: https://doi.org/10.35377/saucis.03.03.830867.

Nemesh, O. M., 2017. Virtual activity of personality: Structure and dynamics of psychological content. Slovo.

Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.

Pavitha, N., Pungliya, V., Raut, A. et al., 2022. Movie recommendation and sentiment analysis using machine learning. Global Transitions Proceedings, 3(1), 279–284, available at: https://doi.org/10.1016/j.gltp.2022.03.012.

Pomytkina, L., Podkopaieva, Y., Hordiienko, K., 2021. Peculiarities of manifestation of student youth' roles and positions in the cyberbullying process. International Journal of Modern Education and Computer Science, 13(6), 1–10, available at: https://doi.org/10.5815/ijmecs.2021.06.01.

Prytula, M., 2024. Fine-tuning of BERT, DistilBERT, XLM-RoBERTa, and Ukr-RoBERTa models for sentiment analysis of reviews in the Ukrainian language. Machine Learning, 3, 4.

Ramos, L., Chang, O., 2023. Sentiment analysis of Russia-Ukraine conflict tweets using RoBERTa. Uniciencia, 37(1), 421–431, available at: http://dx.doi.org/10.15359/r.37-1.23.

Rao, G., Huang, W., Feng, Z., Cong, Q., 2018. LSTM with sentence representations for document-level sentiment classification. Neurocomputing, 308, 49–57, available at: https://doi.org/10.1016/j.neucom.2018.04.045.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K. et al., 2023. Negativity drives online news consumption. Nature Human Behaviour, 7(5), 812–822, available at: https://doi.org/10.1038/s41562-023-01538-4.

Rosenthal, S., Nakov, P., Kiritchenko, S. et al., 2015. SemEval-2015 Task 10: Sentiment analysis in Twitter. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), pp. 451–463.

Sánchez-Rada, J. F., Iglesias, C. A., 2019. Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. Information Fusion, 52, 344–356, available at: https://doi.org/10.1016/j.inffus.2019.05.003.

Schouten, K., Frasincar, F., 2015. The benefit of concept-based features for sentiment analysis. In: Gandon, F., Cabrio, E., Stankovic, M., Zimmermann, A. (Eds.), Semantic Web Evaluation Challenges: SemWebEval 2015. Springer, vol. 548, pp. 273–287, available at: https://doi.org/10.1007/978-3-319-25518-7_19.

Shaik, Z. H., Prasanna, D., Jahnavi, E. et al., 2024, June. FeedForward at SemEval-2024 Task 10: Trigger and sentext-height enriched emotion analysis in multi-party conversations. In: Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pp. 745–756.

Sharma, P., Mishra, N., 2016. Feature level sentiment analysis on movie reviews. In: Proceedings of the 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 306–311. IEEE.

Shelke, N., Deshpande, S., Thakare, V., 2017. Domain independent approach for aspect-oriented sentiment analysis for product reviews. In: Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications: FICTA 2016, vol. 2, Springer Singapore, pp. 651–659.

Taboada, M., 2016. Sentiment analysis: An overview from linguistics. Annual Review of Linguistics, 2(1), 325–347, available at: https://doi.org/10.1146/annurev-linguistics-011415-040518.

Tan, K. L., Lee, C. P., Lim, K. M., 2023. RoBERTa-GRU: A Hybrid Deep Learning Model for Enhanced Sentiment Analysis. Applied Sciences, 13(6), 3915, available at: https://doi.org/10.3390/app13063915.

Tao, Y., Viberg, O., Baker, R. S., Kizilcec, R. F., 2024. Cultural bias and cultural alignment of large language models. PNAS Nexus, 3(9), 346, available at: https://doi.org/10.1093/pnasnexus/pgae346.

Thakkar, G., Preradović, N. M., Tadić, M., 2024. Examining sentiment analysis for low-resource languages with data augmentation techniques. Eng, 5(4), 2920–2942.

Tunstall, L., Beeching, E., Lambert, N. et al., 2023. Zephyr: Direct distillation of LM alignment. arXiv preprint, arXiv:2310.16944.

Vergho, T., Godbout, J. F., Rabbany, R., Pelrine, K., 2024. Comparing GPT-4 and Open-Source Language Models in Misinformation Mitigation. arXiv preprint, arXiv:2401.06920, available at: https://doi.org/10.48550/arXiv.2401.06920.

Veselovská, K., 2017. Sentiment analysis in Czech. Ústav formální a aplikované lingvistiky.

Vidhya, R., Gopalakrishnan, P., Vallamkondu, N. K., 2021. Sentiment analysis using machine learning classifiers: Evaluation of performance. In: Proceedings of the First International Conference on Computing, Communication and Control System (I3CAC 2021), June 07–08. Chennai, India: Bharath University, available at: https://doi.org/10.4108/eai.7-6-2021.2308565.

Wankhade, M., Rao, A. C., Kulkarni, C., 2022. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731–5780, available at: https://doi.org/10.1007/s10462-022-10144-1.

Xu, G., Meng, Y., Qiu, X. et al., 2019. Sentiment analysis of comment texts based on BiLSTM. IEEE Access, 7, 51522–51532, available at: https://doi.org/10.1109/ACCESS.2019.2909919.

Yang, T., Majó-Vázquez, S., Nielsen, R. K., González-Bailón, S., 2020. Exposure to news grows less fragmented with an increase in mobile access. Proceedings of the National Academy of Sciences of the United States of America, 117(46), 28678–28683, available at: https://doi.org/10.1073/pnas.2006089117.

Ye, J., Chen, X., Xu, N. et al., 2023. A comprehensive capability analysis of GPT-3 and GPT-3.5 series models. arXiv preprint, arXiv:2303.10420, available at: https://doi.org/10.48550/arXiv.2303.10420.

Yuna, D., Xiaokun, L., Jianing, L., Lu, H., 2022. Cross-cultural communication on social media: Review from the perspective of cultural psychology and neuroscience. Frontiers in Psychology, 13, 858900, available at: https://doi.org/10.3389/fpsyg.2022.858900.