IntechOpen

IntechOpen Series
Artificial Intelligence, Volume 39

Emotions in Code

The AI Frontier of Sentiment Analysis

Edited by Jinfeng Li

Emotions in Code - The AI Frontier of Sentiment Analysis

Edited by Jinfeng Li

Published in London, United Kingdom

Emotions in Code - The AI Frontier of Sentiment Analysis
http://dx.doi.org/10.5772/intechopen.1006407
Edited by Jinfeng Li

Contributors
Aditya Padmanabhan Varma, Andrej Brník, Bernard Mallia, Jinfeng Li, Lucia Furtáková, Marco A. Palomino, Rohan Allen, Simrandeep Kaur, Yan Li, Ľubica Janáčková

First published in London, United Kingdom, 2025 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 167-169 Great Portland Street, London, W1W 5PF, United Kingdom

For EU product safety concerns: IN TECH d.o.o., Prolaz Marije Krucifikse Kozulić 3, 51000 Rijeka, Croatia, info@intechopen.com or visit our website at intechopen.com.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Emotions in Code - The AI Frontier of Sentiment Analysis
Edited by Jinfeng Li
p. cm.

This title is part of the Artificial Intelligence Book Series, Volume 38
Topic: Machine Learning and Data Mining
Series Editor: Andries Engelbrecht
Topic Editor: Marco Antonio Aceves-Fernández

Print ISBN 978-1-83634-812-2
Online ISBN 978-1-83634-811-5
eBook (PDF) ISBN 978-1-83634-813-9
ISSN 2633-1403

If disposing of this product, please recycle the paper responsibly.

IntechOpen Book Series
Artificial Intelligence
Volume 38

Aims and Scope of the Series

Artificial Intelligence (AI) is a rapidly developing multidisciplinary research area that aims to solve increasingly complex problems. In today's highly integrated world, AI promises to become a robust and powerful means for obtaining solutions to previously unsolvable problems. This Series is intended for researchers and students alike interested in this fascinating field and its many applications.

Meet the Series Editor

Andries Engelbrecht received the Masters and Ph.D. degrees in Computer Science from the University of Stellenbosch, South Africa, in 1994 and 1999 respectively. He is currently appointed as the Voigt Chair in Data Science in the Department of Industrial Engineering, with a joint appointment as Professor in the Computer Science Division, Stellenbosch University. Prior to his appointment at Stellenbosch University, he has been at the University of Pretoria, Department of Computer Science (1998-2018), where he was appointed as South Africa Research Chair in Artifical Intelligence (2007-2018), the head of the Department of Computer Science (2008-2017), and Director of the Institute for Big Data and Data Science (2017-2018). In addition to a number of research articles, he has written two books, Computational Intelligence: An Introduction and Fundamentals of Computational Swarm Intelligence.

Meet the Volume Editor

Prof. Jinfeng Li is an Engineering Scientist on liquid crystal-based millimeter-wave technology for 5G/6G, and an authority on sentiment analysis. He received the Institution of Engineering and Technology (IET) Award, the Arthur Jarvis Prize, the Annals of Emerging Technologies in Computing (AETiC) Highly Cited Article Award, and three Best Paper Awards at IEEE, IOP, and IET conferences, respectively. He was a Cambridge Trust Scholar, an Emerging Technologist with Barclays UK, an Editorial Board member of three Science Citation Index journals, and a Session Chair of nine IEEE conferences. He was elected National-level Young Talent, Senior Member of the China Institute of Communications, Newton Prize reviewer for the UK National Commission for UNESCO, and Grants Reviewer for the Health and Social Care Delivery Research (HSDR) fund from the National Institute for Health Research (NIHR), UK. He teaches four undergraduate modules: (1) States of the Arts in Liquid Crystals Millimeter-wave Technology for 6G; (2) Frontiers and Progress of Electrical and Computer Engineering; (3) Frontiers of Electronic Science and Technology; and (4) States of the Arts in Piano.

Contents

Preface

In recent years, sentiment analysis has emerged as a critical tool in our increasingly data-driven world, bridging the gap between human emotion and artificial intelligence (AI). As machines become more adept at interpreting textual, vocal, and even visual emotional cues, new opportunities and challenges are arising at the intersection of technology, psychology, and society. *Emotions in Code – The AI Frontier of Sentiment Analysis* explores this rapidly evolving landscape through a curated collection of thought-provoking and technically rich chapters contributed by leading researchers and practitioners.

This book is organized into three sections, each reflecting a distinct dimension of the sentiment analysis domain, from foundational paradoxes to technical innovations and global applications.

The book opens with Section 1, "The AI – Emotion Paradox", which sets the theoretical stage. The single chapter in this section, *Introductory Chapter: Paradox of Decoding Emotion in the Age of AI*, explores the conceptual tensions inherent in training machines to understand and simulate emotions—elements of the human experience that are often fluid, context-dependent, and culturally nuanced. This chapter provides a philosophical and technological overview, examining the assumptions underlying the emotional capabilities of current AI systems and laying the groundwork for the more specialized studies that follow.

Section 2, "Algorithmic Advancements in Sentiment Analysis", delves into the computational core of the field. The first chapter in this section, *From Sentiment to Strategy: Machine Learning in Emotion-Based Asset Allocation*, demonstrates how sentiment analysis can be harnessed to inform financial decision-making. It highlights the integration of emotion recognition with quantitative models in the high-stakes world of asset management. The second chapter, *Perspective Chapter: The Faint Whispers of the Heart – Next-Generation Sentiment Analysis and "Empathy Algorithms"*, reflects on the emerging trend of embedding empathy into AI systems and discusses how these developments could transform user experience design and human–machine interaction. The third chapter, *Research on the Implementation of Algorithmic Thinking and Sentiment Analysis by Applying the Knowledge-Based System*, offers a practice-oriented examination of how knowledge-based frameworks can support sentiment analysis through systematic algorithmic reasoning, potentially improving interpretability and adaptability across various applications.

The final part of the volume, Section 3, "Global Diffusion of Sentiment Analysis Applications", explores how sentiment analysis technologies are being adopted and adapted in real-world settings. The opening chapter in this section, *Perspective Chapter: Artificial Intelligence in Slovak Radio Industry – The Present and the Future of Broadcasting*, investigates how sentiment-driven AI tools are shaping content creation and audience engagement in the broadcasting sector. This chapter provides insight into how emotion-aware systems are transforming traditional industries. The volume concludes with *An X Study of the Evolution of COVID-19-Related Sentiments in the UK*, which provides a data-driven case study illustrating how sentiment analysis can track public opinion and emotional responses during times of crisis, offering valuable lessons for public communication and health policy.

Together, these chapters offer a multidimensional perspective on sentiment analysis, highlighting both the technical breakthroughs and the broader societal implications of integrating emotional intelligence into artificial systems. The book is intended for researchers, graduate students, and professionals in AI, computational linguistics, psychology, communication studies, and related fields who are interested in the future of emotionally intelligent computing.

I hope this collection not only informs but also inspires further inquiry into how we decode, model, and perhaps one day empathize with emotion through machines.

Jinfeng Li
Beijing Institute of Technology,
Beijing, China

Section 1

The AI – Emotion Paradox

Chapter 1

Introductory Chapter: Paradox of Decoding Emotion in the Age of AI

Jinfeng Li

1. Introduction

In the ever-evolving landscape of artificial intelligence (AI) [1], sentiment analysis (SA) [2] has emerged as one of the most transformative—and paradoxical—technologies of our time. By 2025, the field will have reimagined and advanced far beyond rudimentary polarity detection (positive, negative, neutral) [3] into a multidimensional, almost empathetic understanding of human emotions. Businesses, governments, and individuals now wield tools capable of dissecting sentiment with unprecedented depth [4] and tighter integration between language models (e.g., GPT, Claude, etc. for nuanced textual understanding), behavioral data (e.g., eye-tracking, voice tone, biometrics for multimodal SA), and real-world context (e.g., situational awareness, cultural/subcultural cues for grounding interpretations in reality). As illustrated in **Figure 1**, this triad reflects the cutting edge of SA in 2025, where AI does not just read text but synthesizes how something is said, who is saying it, and why it might matter, yet this very power raises profound questions concerning ethics, authenticity, and the nature of emotion itself [5].

At the core of this chapter, the Diamond Sutra [6], a foundational Buddhist text, reminds us, "All conditioned phenomena are like dreams, illusions, bubbles, or shadows." This wisdom resonates eerily in the realm of AI-driven sentiment analysis. If emotions are transient and context-dependent, can machines—trained on vast but finite datasets—ever truly understand them? Or are they merely constructing elaborate simulations, reflections of human affect without true comprehension?

This chapter explores the dynamic interplay between technology and emotion, examining both the groundbreaking opportunities and the ethical, philosophical, and technical challenges that define sentiment analysis in 2025.

2. Evolution of sentiment analysis from keywords to contextual awareness and new opportunities in 2025: Emotion as a computational frontier

Sentiment analysis has undergone a radical transformation since its early reliance on keyword-based classifiers [7]. Today, multimodal AI systems [8] synthesize text, vocal tonality, micro-expressions, and even biometric signals to infer emotional states with startling precision. Large language models (LLMs) such as GPT-5 and Claude 4 demonstrate near-human contextual awareness, while federated learning [9] and privacy-preserving techniques allow for sentiment decoding without compromising individual confidentiality.

Yet, as the Diamond Sutra suggests, "Form is emptiness, emptiness is form." The very data that fuels these models—human expressions—are themselves fleeting,

IntechOpen

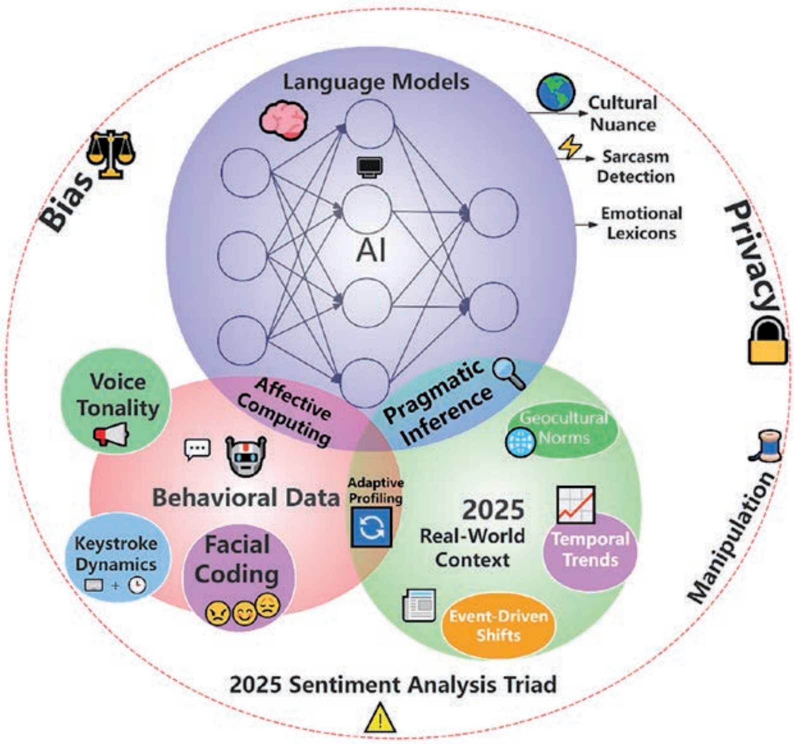


Figure 1.
The synergy triad driving modern sentiment analysis (circa 2025).

culturally contingent, and often illusory. Sarcasm [10], irony [11], and culturally nuanced emotions [12] remain formidable challenges, exposing the limitations of even the most sophisticated AI.

With an increased emphasis on leveraging emotion as a computational frontier, five developmental opportunities in 2025 are identified in **Figure 2**.

AI-driven sentiment analysis has ushered in an era of hyper-personalization, where systems dynamically adjust interactions—modulating tone, pricing, and engagement strategies—based on real-time detection of frustration, joy, or hesitation. Brands now deploy emotion-aware algorithms [13] to optimize customer experiences, from tailoring advertisements to altering service responses mid-conversation.

Yet this capability blurs the line between personalization and manipulation. If an AI can sense frustration and lower prices accordingly, does this constitute ethical adaptation or exploitative conditioning? The Diamond Sutra's teaching on the illusory nature of perceived reality ("all phenomena are illusory") invites us to question whether these "personalized" experiences are genuine or merely algorithmic constructs designed to maximize engagement.

Sentiment-aware technologies have also permeated mental health care [14], with chatbots and wearables analyzing linguistic patterns, vocal stress, and physiological signals to detect early signs of anxiety or depression. These tools offer timely interventions, from calming prompts to crisis hotline referrals, potentially saving lives.

However, a fundamental tension remains: can an AI, devoid of subjective experience, truly empathize [15]? While it can simulate compassion through curated

Figure 2.
The five lobes of a single leaf identified for emerging scopes of new advancements envisioned for computational emotion (machine emotion).

responses, the Diamond Sutra's insight—that all phenomena lack inherent, independent existence—suggests that AI's "empathy" is a mirage, a reflection of human programming rather than authentic understanding. This raises ethical concerns about reliance on machines for emotional support, particularly when they cannot grasp the depth of human suffering.

Governments and NGOs increasingly rely on sentiment analysis to monitor public opinion, predict civil unrest, and combat disinformation campaigns [3]. By analyzing social media, news trends, and even encrypted messaging platforms, AI models attempt to gauge collective emotions and anticipate societal shifts.

Yet the Diamond Sutra's caution against clinging to perceptions ("all phenomena are like dreams, illusions") indicates the fragility of such data. Emotional expressions online are often performative, distorted by anonymity, satire, or algorithmic amplification [16]. Can sentiment analysis, trained on these unstable signals, ever produce reliable forecasts? Or does it risk reinforcing echo chambers and misinterpreting the very emotions it seeks to decode?

Advances in explainable AI (XAI) and fairness-aware algorithms [17] aim to mitigate biases in sentiment analysis, ensuring models do not disproportionately misinterpret emotions across cultures, genders, or dialects.

Yet the Diamond Sutra's assertion that "form is emptiness" challenges the notion of a perfectly objective system. Linguistic and cultural diversity ensures that no model can be truly universal; emotions are expressed and interpreted in infinitely variable ways. Even the most rigorously audited AI may inadvertently privilege certain worldviews, revealing the inherent subjectivity embedded in all computational frameworks. The pursuit of ethical AI, then, is not a destination but a continuous negotiation—one that demands humility about the limits of machine interpretation [18].

3. Emerging challenges and concerns: The limits of machine emotion

Under the context of an ever-evolving landscape of AI and a lack of a unified standard for machine emotion, the longstanding and urgent concerns of sentiment analysis depicted in **Figure 3** cannot be overstated.

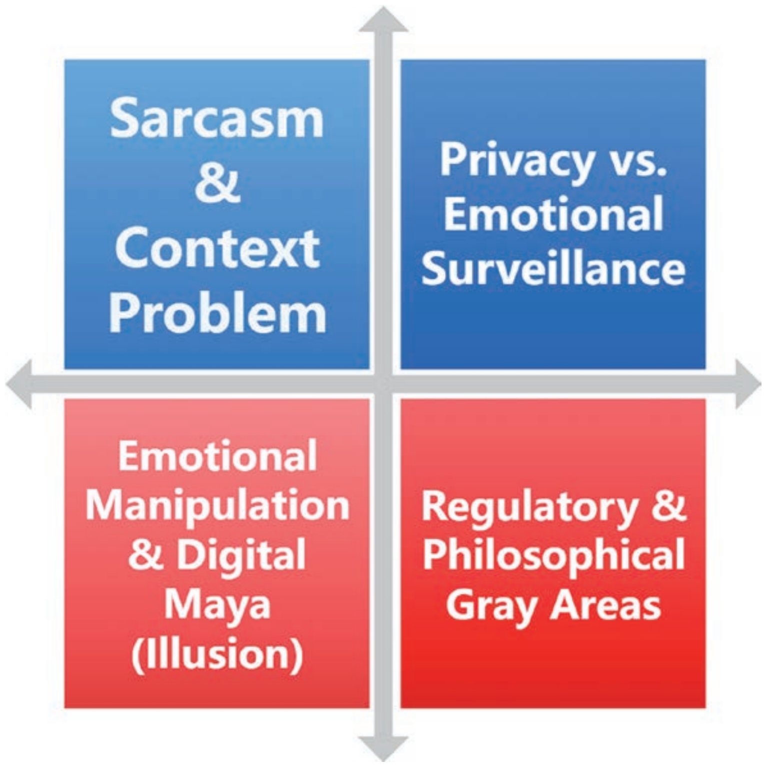


Figure 3.
Four pillars of concerns en route to machine emotion.

Despite their sophistication, even the most advanced AI models falter when confronted with irony [10], sarcasm [11], and culturally nuanced expressions [12]. A joke in Mumbai may read as sincere concern in Munich, and a politeness algorithm trained on American English might misinterpret respectful hesitation in Japanese communication as disinterest. These misclassifications carry real-world consequences—from misdirected customer service interventions to flawed hiring tools that misjudge candidates' enthusiasm. The Diamond Sutra's teaching that "all conditioned phenomena are like illusions" finds eerie resonance here: if human communication itself is inherently ambiguous and context-dependent, can any machine ever fully grasp its subtleties?

As sentiment analysis expands into real-time voice stress detection, emotion-tracking wearables, and even gaze analysis in virtual meetings, society faces an existential question: where should the boundary lie between beneficial emotional insight and dystopian surveillance? The same technology that helps therapists monitor patients' depressive episodes could also empower employers to penalize workers for "insufficient enthusiasm."

The Diamond Sutra's warning against attachment to form ("those who see me through form cannot perceive the truth") invites us to reconsider whether perpetual emotional data collection—no matter how well-intentioned—might ultimately distance us from authentic human connection [19].

The Diamond Sutra's concept of maya (illusion) becomes urgently relevant in an age where AI leverages sentiment analysis to craft hyper-persuasive propaganda, deepfake emotional appeals, and psychologically tailored scams. Political campaigns now deploy “emotionally optimized” messaging that bypasses rational critique by directly targeting voters’ subconscious fears or hopes [2], while advertisers use real-time frustration detection to nudge consumers toward purchases. When machines learn to exploit the gap between felt emotion and conscious decision-making, they risk reducing human autonomy to a series of algorithmic triggers [20]—a digital puppet show where our emotions pull strings we cannot see.

The legal and ethical frameworks governing sentiment analysis lag far behind its capabilities. Should AI-interpreted emotional states—such as a defendant’s “lack of remorse” inferred from vocal patterns—be admissible in court? Can corporations justify dynamic pricing that escalates costs when detecting customer frustration? These dilemmas expose deeper philosophical tensions: if emotions are transient and culturally constructed (as the Diamond Sutra suggests), can they ever serve as reliable metrics for justice or commerce? The lack of consensus reflects a broader societal struggle to reconcile technological possibility with human values—a struggle that will define whether sentiment analysis becomes a tool for empowerment or control.

4. Concluding remarks and the path forward: Navigating the illusion

This chapter does not seek definitive answers but instead invites readers to contemplate the fragile interplay between human emotion and machine interpretation, with no compromises on the scientific rigor. The Diamond Sutra teaches that all phenomena are interdependent and impermanent—a lens through which we might view AI sentiment analysis not as an infallible tool but as a mirror reflecting our own biases, aspirations, and limitations.

As we stand on the cusp of emotions in code (with a helpful dose of engineering injected), the challenge is not merely to build better algorithms but to wield them with wisdom, recognizing that emotion—like all things—is both data and illusion.

Author details

Jinfeng Li
Beijing Institute of Technology, Beijing, China

*Address all correspondence to: jinfengcambridge@bit.edu.cn

References

[1] Mitchell M. Debates on the nature of artificial general intelligence. Science. 2024;**383**(6689):eado7069. DOI: 10.1126/science.ado7069

[2] Li J. Advances in Sentiment Analysis - Techniques, Applications, and Challenges. London, UK: IntechOpen; 2024. DOI: 10.5772/intechopen.111293

[3] Guo X, Li J. A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency. In: Proceedings of the 2019 Sixth IEEE International Conference on Social Networks Analysis, Management and Security (SNAMS); 22-25 October 2019; Granada. New York: IEEE; 2019. pp. 472-477. DOI: 10.1109/SNAMS.2019.8931720

[4] Smatov N, Kalashnikov R, Kartbayev A. Development of context-based sentiment classification for intelligent stock market prediction. Big Data and Cognitive Computing. 2024;**8**(6):51. DOI: 10.3390/bdcc8060051

[5] Sedikides C, Schlegel R. Distilling the concept of authenticity. Nature Reviews Psychology. 2024;**3**:509-523. DOI: 10.1038/s44159-024-00323-y

[6] Gemmell W, editor. The Diamond Sutra: (Chin-Kang-Ching) or Prajna-Paramita. 1st ed. London: Routledge; 1912. DOI: 10.4324/9781003280446

[7] Baumgarten M, Mulvenna M, Rooney N, Reid J. Keyword-based sentiment mining using twitter. International Journal of Ambient Computing and Intelligence. 2013;**5**:56-69. DOI: 10.4018/jaci.2013040104

[8] Rodis N, Sardianos C, Radoglou P, Sarigiannidis P, Varlamis I, Papadopoulos G. Multimodal explainable artificial intelligence: A comprehensive review of methodological advances and future research directions. IEEE Access. 2024;**12**:159794-159820. DOI: 10.1109/ACCESS.2024.3467062

[9] Bansal S, Singh M, Bhadauria M, Adalakha R. Federated learning approach towards sentiment analysis. In: Proceedings of the 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS); 10-12 October 2022; Tashkent. New York: IEEE; 2022. pp. 717-724. DOI: 10.1109/ICTACS56270.2022.9987996

[10] Vitman O, Kostiuk Y, Sidorov G, Gelbukh A. Sarcasm detection framework using context, emotion and sentiment features. Expert Systems with Applications. 2023;**234**:121068. DOI: 10.1016/j.eswa.2023.121068

[11] Bedmar I, Bartolome S. Multimodal fake news detection. Information. 2022;**13**(6):284. DOI: 10.3390/info13060284

[12] Li M, Olsen K, Thompson W. Cross-cultural biases of emotion perception in music. Brain Sciences. 2025;**15**(5):477. DOI: 10.3390/brainsci15050477

[13] Gamage G, Silva D, Mills N, et al. Emotion AWARE: An artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis. Journal of Big Data. 2024;**11**:93. DOI: 10.1186/s40537-024-00953-2

[14] Ahsan S, Djenouri D, Haider R. Privacy-enhanced sentiment analysis in mental health: Federated learning with data obfuscation and bidirectional encoder representations

from transformers. Electronics. 2024;**13**(23):4650. DOI: 10.3390/electronics13234650

[15] Perry A. AI will never convey the essence of human empathy. Nature Human Behaviour. 2023;**7**:1808-1809. DOI: 10.1038/s41562-023-01675-w

[16] Jim J, Talukder M, Malakar P, Kabir M, Nur K, Mridha M. Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. Natural Language Processing Journal. 2024;**6**:100059. DOI: 10.1016/j.nlp.2024.100059

[17] Naveed S, Stevens G, Robin K. An overview of the empirical evaluation of explainable AI (XAI): A comprehensive guideline for user-centered evaluation in XAI. Applied Sciences. 2024;**14**(23):11288. DOI: 10.3390/app142311288

[18] Li J. Machine learning and digital twinning enabled liquid crystals mm-wave reconfigurable devices design and systems operation. In: Proceedings of the 2022 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP); 27-29 November 2022; Guangzhou. New York: IEEE; 2023. pp. 1-3. DOI: 10.1109/IMWS-AMP54652.2022.10107046

[19] Li J. Introductory chapter: The 2023 sentiment analysis roadmap. In: Li J, editor. Advances in Sentiment Analysis - Techniques, Applications, and Challenges. London, UK: IntechOpen; 2024. DOI: 10.5772/intechopen.112276

[20] Miah M, Kabir M, Sarwar T, et al. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. Scientific Reports. 2024;**14**:9603. DOI: 10.1038/s41598-024-60210-7

Section 2

Algorithmic Advancements in Sentiment Analysis

Chapter 2

From Sentiment to Strategy: Machine Learning in Emotion-Based Asset Allocation

Simrandeep Kaur

Abstract

The objective of this chapter is the integration of machine learning (ML) techniques and investor sentiment analysis for emotion-based asset allocation strategies. It addresses challenges in noisy and imbalanced sentiment data by highlighting the effectiveness of models of supervised learning, which include fuzzy support vector machines (FSVM) and new fuzzy support vector machines (NFSVM) proposed in Kaur et al. These models use fuzzy membership functions to denoise and better classify assets with greater growth potential. The crucial methodologies/trends include effective feature extraction techniques like principal component analysis (PCA), which can streamline sentiment data and capture investor emotions effectively. This chapter further shows how such ML-driven insights can be used to integrate them with Markowitz's mean-variance portfolio optimization model (MV model), with and without cardinality constraints, to create optimal portfolios. Experimental analysis based on datasets taken from the Nifty 50 and Euro Stoxx 50 indices reveals the fact that emotion-driven strategy, NFSVM-based, outperforms traditional SVM and FSVM models in terms of risk-adjusted returns and portfolio diversification. This work opens up possibilities for combining sentiment analysis and ML to navigate turbulent financial markets and offers new portfolio optimization frameworks. Future studies could involve expanding datasets, hybrid machine-learning techniques, and advanced deep-learning models that are applied in real-time applications.

Keywords: machine learning, supervised learning, classification, data imbalance, fuzzy support vector machines, feature extraction, investor sentiment analysis, financial investment, emotion recognition, portfolio optimization, shortlisting assets, prediction

1. Introduction

Sentiment analysis in finance, especially investor sentiment, has been highly demanded and used in financial research in recent years [1]. Investor sentiment was first proposed by De Long et al. in 1990 and considered very important for financial markets to have a great influence [2]. Later, using different approaches, Baker and Wurgler showed how investors' moods influence market behavior and asset prices [3]. This chapter expands on these fundamental notions by shifting the discussion "*from sentiment to strategy*," concentrating on the use of ML methods to create frameworks for asset allocation that are based on emotions. This work specifically attempts to use investor sentiment to handle issues including data imbalance, emotion recognition, and classification, ultimately directing financial investment and portfolio optimization decisions. This chapter demonstrates how high-powered supervised learning models can draw valuable insights from noisy sentiment data by using techniques like fuzzy support vector machines and their improved variant, the new fuzzy support vector machines, which entail solving a nonlinear quadratic programming problem (NLQPP). The proposed methodology attempts to convert investor sentiment analysis into useful tactics for asset shortlisting and prediction if applied with feature extraction techniques such as PCA. The ultimate aim is to ensure that there is a meaningful mechanism that makes use of new combinations of ML algorithms and emotion-driven analytics and addresses the challenges related to asset allocation in volatile markets.

This chapter highlights the need/use of machine learning (ML) algorithms, particularly FSVM, for asset preselection in portfolio optimization. ML models update continually with fresh data, changing along with market fluctuations, reducing risk while potentially increasing returns. By integrating advanced ML techniques, the decision surface is learned with different input points, which increases the asset selection process and optimizes portfolio performance.

The chapter is divided as follows: Section 2 introduces sentiment analysis and provides background about investor sentiment and some of the theoretical foundations in finance for sentiment analysis. It describes how investor sentiment analysis could be transformed into an emotion-based asset allocation technique. Section 3 presents experimental results using datasets from the Nifty 50 and Euro Stoxx 50 indices, demonstrating investor sentiments turning into strategy and comparing NFSVM to conventional SVM and FSVM models. A summary of the chapter's main conclusions and recommendations for future research, including the possibility of hybrid machines, is provided in Section 4. Section 5 concludes with a conclusion that elaborates on the relationship between emotions and computational processes mentioned in the appendix, offering a more comprehensive view of the chapter's theme, *Emotions in Code Appendix A1*.

2. Portfolio selection with cardinality constraints in the Markowitz mean-variance model: Investor sentiment analysis to strategy for emotion-based asset allocation

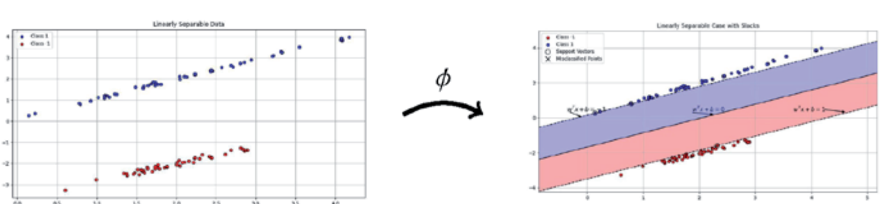
This section provides a brief description of the fundamental ideas and models associated with the theme of this chapter, that is, support vector machines (SVM), FSVM, and MV models. The notations and definitions employed in this chapter are taken from relevant previous studies, and hence, readers are advised to look at Kaur et al. [1]; Chandra et al., particularly Chapter 16 mentioned in Ref. [4]; and Lin et al. [5].

2.1 Support vector machines for binary data classification

SVMs are frequently applied to binary classification. Assuming a collection of l training points, each having a m feature vector and a class label $(+1 \text{ or } -1)$, if the data is linearly separable, there is some weight vector w and some bias term b such that for points with class label +1, the linear combination of the m features, plus the bias, is greater than or equal to +1, and for points with class label −1, the linear combination of the features, along with the bias, is less than or equal to −1. The SVM method seeks the best hyperplane that maximizes the $\text{margin}\left(\frac{2}{\|\mathbf{w}\|}\right)$ between two classes. For linearly separable data, the SVM formulates an optimization problem that minimizes the squared magnitude of the weight vector subject to the constraint that every training point is classified correctly. The weight vector w is perpendicular to the hyperplane and b is the bias term. The detailed theory and derivation behind the SVM model are mentioned in Ref. [6]. For more details, refer to **Figure 1**.

For nonlinearly separable data, it has a nonlinear function, say ϕ, mapping the input into high-dimensional feature space where it can become linearly separable. In this sense, slack variables ξ_i are introduced with a tolerance of misclassification. Therefore, the optimization problem minimizes the trade-off between the margin and the penalty for being misclassified and is controlled by some regularization parameter C. Determining C will balance between a maximizing margin and minimizing training error. For more details, refer to **Figure 2**.

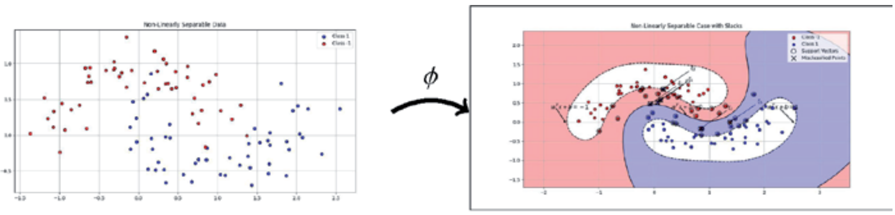
The Wolfe dual of the SVM optimization problem, as detailed by Chandra et al., specifically in chapter 8 mentioned in Ref. [4], maximizes the Lagrangian subject to constraints on its derivatives and non-negativity of Lagrange multipliers α_i. The Lagrangian dual reformulates the problem as maximizing a function of the Lagrange multipliers with constraints on their values as well as on the data labels. The proposed method defines the optimal hyperplane parameters in terms of the decision function expressed using the optimal multipliers, weights, and biases. The kernel function, which satisfies the Mercer condition, replaces the feature space inner product. The Radial Basis Function (RBF) kernel is commonly used for this purpose. For a better understanding of investor sentiment analysis after the shortlisted asset results from SVM, we apply the MV model



(a) Linearly Separable Case Data (Input Space)

(b) Decision Boundaries and Support Vectors in a Linearly Separable Case using Support Vector Machines (Feature Space)

Figure 1.
Support vector machine for linearly separable data.



(a) Non-Linearly Separable Case Data (Input Space)

(b) Decision Boundaries and Support Vectors in a Non-Linearly Separable Case using Support Vector Machines (Feature Space)

Figure 2.
Support vector machine for non-linearly separable data.

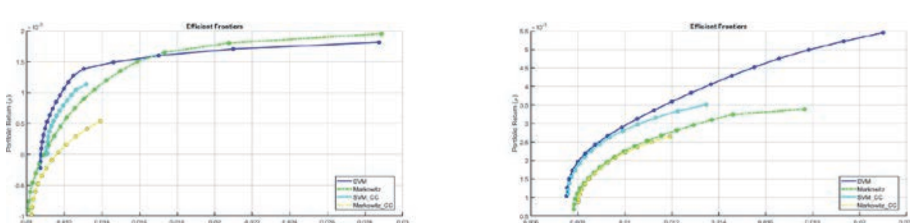
[1] to obtain the optimal Markowitz efficient frontier, which is depicted in **Figure 3** for Nifty 50 and **Figure 4** for Euro Stoxx 50 assets, respectively. Both figures clearly show the importance of using ML models like SVM for shortlisting assets rather than simply using the MV models for portfolio optimization.

2.2 Fuzzy support vector machine for binary data classification

SVMs are excellent for classification but might not be sensitive to noise and outliers. To overcome this, Lin et al. [5] presented Fuzzy SVM (FSVM), as illustrated graphically in **Figure 5**, which applies fuzzy membership values to the training samples, thus lessening the influence of minor points and outliers. FSVM is better than classical SVM because it employs fuzzy membership values to handle noisy data and outliers more effectively. The decision function predicts class labels by taking into account the influence of these weights, providing greater flexibility in classification. For a better understanding of investor sentiment analysis after the shortlisted results from FSVM, we apply the MV model for obtaining the optimal Markowitz efficient frontier depicted in **Figure 6** for Nifty 50 and **Figure 7** for Euro Stoxx 50 assets, respectively.

2.3 New fuzzy support vector machine for binary data classification

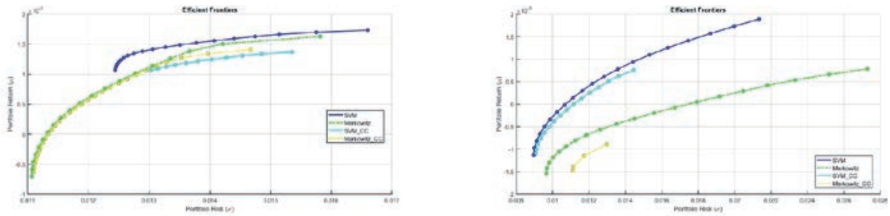
Tang [7] described an enhanced fuzzy membership function that utilizes the structural characteristics of feature space. The technique here again reduces the effect



(a) Rolling 1

(b) Rolling 2

Figure 3.
Efficient frontiers for the assets selected by SVM model for Nifty 50 dataset.



(a) Rolling 1 (b) Rolling 2

Figure 4.
Efficient frontiers for the assets selected by SVM model for Euro Stoxx 50 dataset.

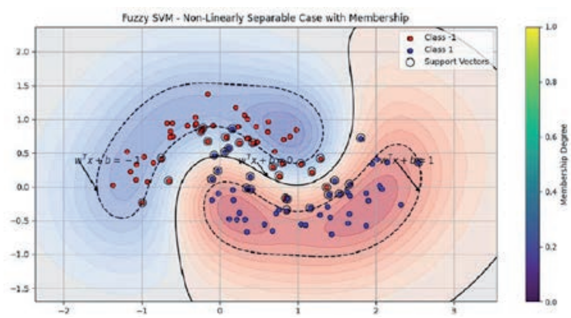
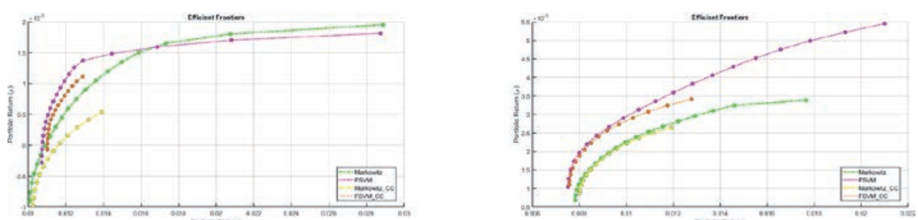
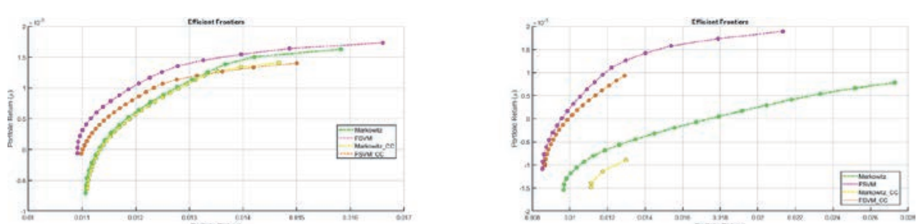


Figure 5.
Fuzzy support vector machine for non-linearly separable data.



(a) Rolling 1 (b) Rolling 2

Figure 6.
Efficient frontiers for the assets selected by FSVM model for Nifty 50 dataset.



(a) Rolling 1 (b) Rolling 2

Figure 7.
Efficient frontiers for the assets selected by FSVM model for Euro Stoxx 50 dataset.

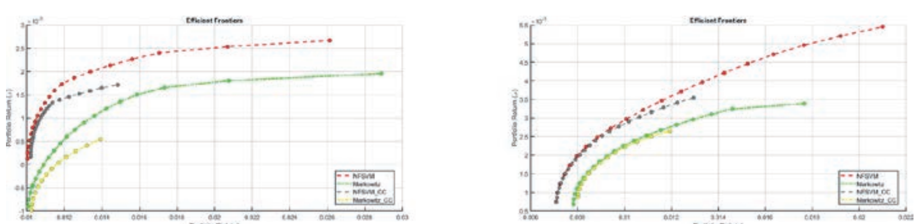
of outliers and further improves accuracy in classification. The memberships are computed in terms of relative distances from a sample point to the class prototypes with their significance adjusted at appropriate weights. This method is very efficient in the noisy scenario and superior to that developed earlier; refer to **Figures 8** and **9**. Details of its implementation are given in *Appendix A2*.

2.4 Optimal portfolio construction by Markowitz mean-variance model without using SVM

The classical Markowitz portfolio optimization model aims to minimize portfolio risk while considering expected returns. The model allocates weights to assets, ensuring the total weight equals one and that no short-selling occurs. The portfolio return is determined based on asset returns across various scenarios, with the expected return calculated as a weighted average. Variance represents portfolio risk, computed using asset covariances (refer to [1]).

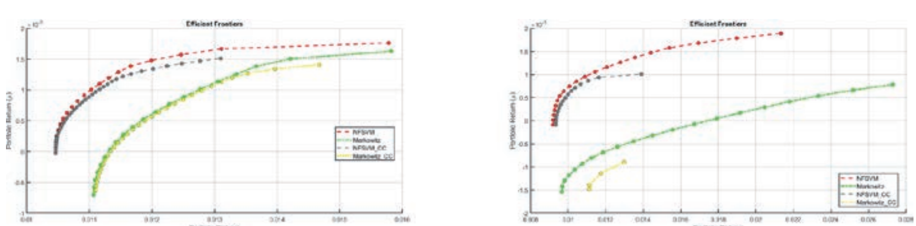
The optimization involves minimizing risk and maximizing expected returns. The investor chooses an expected return within a defined range. Additional constraints include limiting the number of selected assets and setting minimum and maximum allocation levels for each asset. Binary variables indicate whether an asset is included in the portfolio, ensuring flexibility in asset selection. These constraints refine the model to align with practical investment preferences.

For the first and second rolling windows, respectively, the assets for which the label $y_i = +1$ on the 249-th day and 339-th day data are shortlisted, and **Figures 10** and **11**



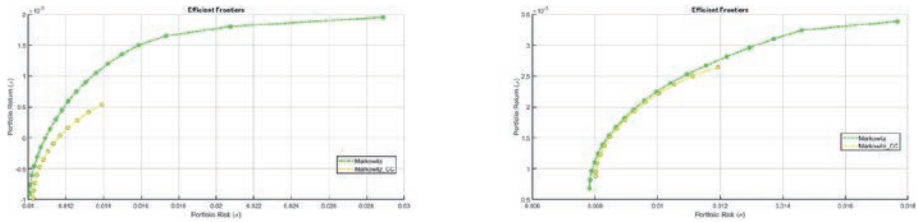
(a) Rolling 1 (b) Rolling 2

Figure 8.
Efficient frontiers for the assets selected by NFSVM model for Nifty 50 dataset.



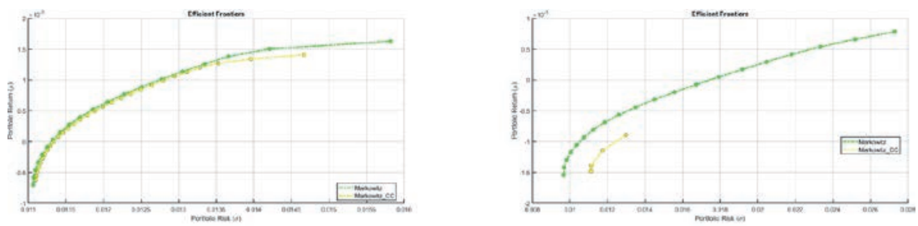
(a) Rolling 1 (b) Rolling 2

Figure 9.
Efficient frontiers for the assets selected by NFSVM model for Euro Stoxx 50 dataset.



(a) Rolling 1 (b) Rolling 2

Figure 10.
Efficient frontiers for the assets selected by the Markowitz model for the Nifty 50 dataset.



(a) Rolling 1 (b) Rolling 2

Figure 11.
Efficient frontiers for the assets selected by Markowitz model for Euro Stoxx 50 dataset.

illustrate the efficient frontier of short-listed assets for the 90 days of closing prices with (from 249 to 338 days) and (from 339 to 428 days). Thus, the investor sentiment can select a portfolio that minimizes risk while respecting both the desired expected return and the cardinality constraints.

3. Proof from emotion to action: Sentiment to portfolio strategy

This section demonstrates how investor emotion materializes into practical solutions of portfolio improvement. In light of this, the investigation takes feeds for the dates from January 1, 2021, to September 30, 2022, from the data feed of both the Nifty 50 and Euro Stoxx 50 indices by Thomson Reuters Eikon. The methodology used involves splitting the datasets into in-sample and out-of-sample data and using rolling window techniques for strong performance evaluation. The detailed steps are discussed in Ref. [1] and **Figure 12** where the study comes up with a strong methodology for incorporating sentiment analysis into asset allocation using an exhaustive pipeline. The Python code for feature extraction is discussed in *Appendix A1*. Through the systematic transformation of sentiment data into actionable insights, the approach enhances portfolio optimization and adapts to the dynamic nature of financial markets. This sentiment-driven framework bridges the gap between investor emotions and strategic portfolio construction, offering a powerful tool for improving risk-adjusted returns and advancing the field of emotion-based finance.

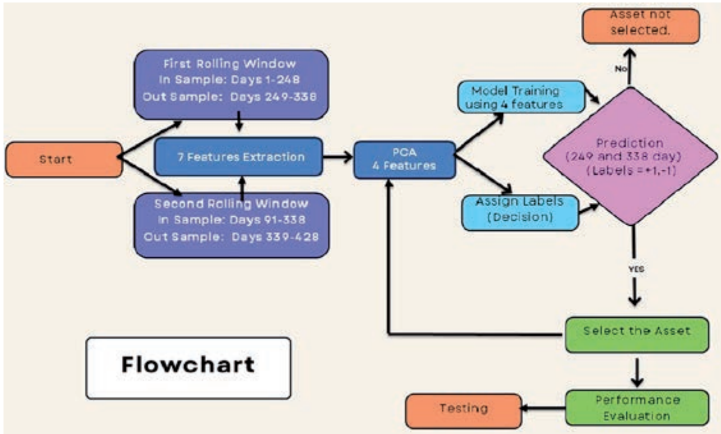
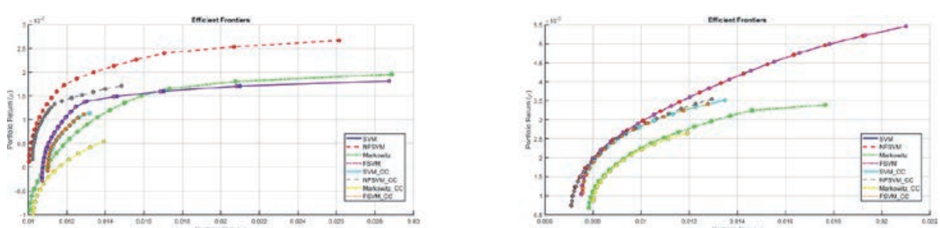


Figure 12.
Methodology/trend process flow diagram.

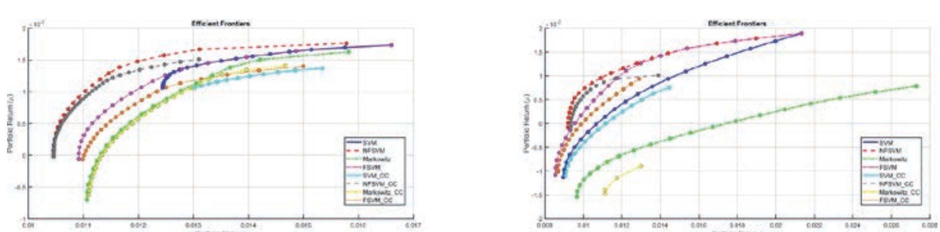
3.1 Asset selection across all models and optimal portfolio construction

This section presents portfolio optimization results using various models for the Nifty 50 and Euro Stoxx 50 datasets, with both cardinality-constrained and unconstrained efficient frontiers. **Figure 13** shows that for both windows (249–338 days and 339–428 days) of the Nifty 50 dataset, NFSVM outperforms FSVM, SVM, and Markowitz. In Window 1, NFSVM is leading, followed by FSVM and SVM, which are ahead of Markowitz. In Window 2, NFSVM, FSVM, and SVM outperform Markowitz. **Figure 14** reports similar



(a) Rolling 1 (b) Rolling 2

Figure 13.
Combined efficient frontiers of Markowitz, SVM, FSVM, and NFSVM model for Nifty 50 dataset.



(a) Rolling 1 (b) Rolling 2

Figure 14.
Combined efficient frontiers of Markowitz, SVM, FSVM, and NFSVM model for Euro Stoxx 50 dataset.

	Without cardinality constraints for rolling 1				With cardinality constraints for rolling 1			
	Markowitz	**SVM**	**FSVM**	**NFSVM**	**Markowitz**	**SVM**	**FSVM**	**NFSVM**
Out of sample return measures								
Average	−0.00175	−0.00127	−0.00118	*−0.00115*	−0.00175	−0.00119	*−0.00113*	−0.00116
Median	*−0.00129*	−0.00233	−0.00223	−0.00152	−0.00134	−0.00199	*−0.00117*	−0.00151
Min	*−0.04573*	−0.03807	−0.03723	−0.04063	*−0.04655*	−0.04169	−0.04081	−0.04067
Max	*0.03027*	0.03002	0.02933	0.02991	*0.03029*	0.02950	0.02867	0.02991
Out of sample risk measures								
SD	0.01300	0.01229	*0.01213*	0.01243	0.01308	0.01257	0.01244	*0.01243*
DD	0.01380	0.01261	*0.01248*	0.01278	0.01375	0.01292	0.01276	*0.01266*
VaR	0.02173	0.02307	0.02219	*0.02091*	0.02237	0.02261	0.02181	*0.02063*
CvaR	0.03019	0.02902	*0.02718*	0.02766	0.03062	0.02879	0.02841	*0.02698*
Maximum drawdown	0.10276	*0.04376*	0.04417	0.05463	0.09405	*0.04327*	0.04513	0.05500
Out of sample risk-adjusted return measures								
SR	−0.13447	−0.10349	−0.09712	*−0.09280*	−0.13388	−0.09506	*−0.09116*	−0.09347
STARR	−0.05791	−0.04382	−0.04335	*−0.04172*	−0.05719	−0.04149	*−0.03991*	−0.04308
Sortino	−0.12664	−0.11840	−0.09961	*−0.09605*	−0.12595	−0.09529	*−0.09027*	−0.09258
Out of sample statistical test measures								
t-statistics	NAN	1.00690	1.16650	*1.60920*	NAN	1.22950	1.27870	*1.60340*
p-value	NAN	0.31670	0.24650	*0.11110*	NAN	0.22210	0.20430	*0.11240*

Table 1.
Performance measures of an optimal portfolio generated by Nifty 50 dataset both with and without cardinality constraints for rolling 1.

	Without cardinality constraints for rolling 2				With cardinality constraints for rolling 2			
	Markowitz	**SVM**	**FSVM**	**NFSVM**	**Markowitz**	**SVM**	**FSVM**	**NFSVM**
Out of sample return measures								
Average	0.00082	0.00120	*0.00096*	0.00090	0.00086	0.00116	*0.00121*	0.00090
Median	*0.00162*	0.00278	0.00156	0.00137	0.00099	*0.00260*	0.00238	0.00222
Min	*−0.01935*	−0.02311	−0.01807	−0.01663	−0.01902	−0.02254	*−0.02372*	−0.01953
Max	*0.02236*	0.01904	0.01718	0.01720	*0.02174*	0.01853	0.01837	0.01683
Out of sample risk measures								
SD	0.00867	0.00881	*0.00785*	0.00801	0.00873	0.00868	0.00834	*0.00777*
DD	0.00767	0.00794	*0.00757*	0.00751	0.00754	0.00776	0.00754	*0.00747*
VaR	0.01619	0.01569	*0.01528*	0.01525	0.01588	0.01567	0.01544	*0.01525*
CvaR	0.02163	0.02287	*0.02137*	0.02187	0.02202	0.02173	0.02150	*0.02095*
Maximum drawdown	0.09319	0.09442	*0.08509*	0.09007	0.09269	0.09170	0.08991	*0.08478*
Out of sample risk-adjusted return measures								
SR	0.09442	0.13623	*0.12221*	0.11231	0.09758	0.13342	*0.14296*	0.11680
STARR	0.09047	0.12893	*0.11837*	0.10898	0.09021	0.12714	*0.13772*	0.11471
Sortino	0.09534	0.13853	*0.12324*	0.11332	0.09753	0.13564	*0.14396*	0.11897
Out of sample statistical test measures								
t-statistics	NAN	1.13130	1.22820	*1.27460*	NAN	1.17030	1.27780	*1.28220*
p-value	NAN	0.23440	0.20390	*0.15140*	NAN	0.22380	0.20500	*0.15070*

Table 2.
Performance measures of an optimal portfolio generated by Nifty 50 dataset both with and without cardinality constraints for rolling 2.

	Without cardinality constraints for rolling 1				**With cardinality constraints for rolling 1**			
	Markowitz	**SVM**	**FSVM**	**NFSVM**	**Markowitz**	**SVM**	**FSVM**	**NFSVM**
Out of sample return measures								
Average	−0.00064	−0.00051	−0.00059	*-0.00030*	−0.00064	−0.00051	−0.00059	*-0.00030*
Median	0.00117	0.00012	0.00083	*0.00047*	*0.00117*	0.00012	0.00083	0.00046
Min	−0.02814	*-0.04145*	−0.02709	−0.03124	−0.02814	*-0.04145*	−0.02709	−0.03124
Max	0.05134	*0.06359*	0.05857	0.05282	0.05133	*0.06359*	0.05857	0.05282
Out of sample risk measures								
SD	0.01294	0.01533	0.01255	*0.01215*	0.01293	0.01533	0.01255	*0.01214*
DD	0.01401	0.01497	0.01249	*0.01212*	0.01400	0.01497	0.01249	*0.01211*
VaR	0.02250	0.02455	0.02117	*0.01891*	0.02249	0.02455	0.02118	*0.01890*
CvaR	0.02639	0.03197	*0.02394*	0.02465	0.02639	0.03197	*0.02394*	0.02465
Maximum drawdown	*0.03299*	0.03585	0.07021	0.03916	*0.03298*	0.03586	0.07038	0.03911
Out of sample risk-adjusted return measures								
SR	−0.04982	−0.03353	−0.04718	*-0.02487*	−0.04981	−0.03353	−0.04717	*-0.02484*
STARR	−0.02441	−0.01607	−0.02473	*-0.01225*	−0.02441	−0.01608	−0.02473	*-0.01224*
Sortino	−0.04657	−0.03447	−0.04867	*-0.02422*	−0.04657	−0.03448	−0.04866	*-0.02449*
Out of sample statistical test measures								
t-statistics	NAN	0.15360	0.09400	*0.91330*	NAN	0.15330	0.09410	*0.91390*
p-value	NAN	0.87830	0.92530	*0.36360*	NAN	0.87850	0.92520	*0.36330*

Table 3.
Performance measures of an optimal portfolio generated by Euro Stoxx 50 dataset both with and without cardinality constraints for rolling 1.

	Without cardinality constraints for rolling 2				With cardinality constraints for rolling 2			
	Markowitz	**SVM**	**FSVM**	**NFSVM**	**Markowitz**	**SVM**	**FSVM**	**NFSVM**
Out of sample return measures								
Average	−0.00227	−0.00212	−0.00143	*−0.00128*	−0.00180	−0.00153	−0.00140	*−0.00092*
Median	−0.00217	−0.00283	−0.00215	*−0.00131*	*−0.00072*	−0.00164	−0.00230	−0.00190
Min	*−0.03703*	−0.02774	−0.02476	−0.02151	*−0.03418*	−0.02662	−0.02584	−0.02560
Max	0.03402	*0.03729*	0.03580	0.02601	0.02737	*0.03700*	0.03550	0.02211
Out of sample risk measures								
SD	0.01078	0.01033	0.00944	*0.00904*	0.01222	0.00993	*0.00951*	0.01045
DD	0.01141	0.01130	*0.00909*	0.00914	0.01314	0.00962	*0.00940*	0.01057
VaR	0.02174	0.02052	0.01525	*0.01466*	0.02400	0.01708	*0.01562*	0.01720
CvaR	0.02687	0.02401	*0.01867*	0.01888	0.03091	0.02074	*0.01896*	0.02230
Maximum drawdown	0.02778	0.02952	*0.02479*	0.02586	0.02677	0.02632	0.02463	*0.02380*
Out of sample risk-adjusted return measures								
SR	−0.21055	−0.20506	−0.15123	*−0.14149*	−0.14753	−0.15354	−0.14683	*−0.08828*
STARR	−0.08450	−0.08826	−0.07649	*−0.06778*	−0.05834	−0.07355	−0.07366	*−0.04138*
Sortino	−0.20992	−0.21345	−0.16750	*−0.14820*	−0.13807	−0.16309	−0.16219	*−0.09292*
Out of sample statistical test measures								
t-statistics	NAN	0.31370	1.21240	*1.28610*	NAN	0.62960	0.76650	*1.15650*
p-value	NAN	0.75450	0.22860	*0.20170*	NAN	0.53060	0.44540	*0.25060*

Table 4.
Performance measures of an optimal portfolio generated by Euro Stoxx 50 dataset both with and without cardinality constraints for rolling 2.

results for Euro Stoxx 50. In Window 1, NFSVM is leading, while FSVM outperforms SVM and Markowitz. In Window 2, NFSVM is still leading, while FSVM is ahead of SVM, and both models outperform Markowitz. In general, NFSVM consistently outperforms other models and is the first choice for portfolio optimization.

3.2 All-encompassing comparison and comparative analysis of the outcomes

The cardinality constraint drastically affects the performance of the models, which highlights their importance in determining the portfolio optimization results. Further, the fact that the market dynamics are different for the Nifty 50, representing the Indian market, as evident from **Table 1** for rolling window 1 and **Table 2** for rolling window 2, and the Euro Stoxx 50, representing the European market, as evident from **Table 3** for rolling window 1 and **Table 4** for rolling window 2, also affects the outcome. This again puts emphasis on the importance of accounting for market-specific factors in asset selection and portfolio construction.

For all the cases, the NFSVM model was a better performer than the Markowitz, SVM, and FSVM models. It showed better efficiency in portfolio construction irrespective of the dataset or constraints applied. It has been observed that SVM performed better than the Markowitz model and FSVM exhibited better performance than both SVM and Markowitz. This reinforces its performance in asset selection. From these, it is evident that machine learning models such as SVM, FSVM, and NFSVM support the Markowitz approach more effectively in portfolio construction with regard to asset selection. For a better analysis of results, refer to [1].

4. Conclusions

In the domain of fuzzy support vector machines, this chapter demonstrates its feasibility in the potential for exploiting investor sentiment into strategies applied to financial investment. Such ideas in the application may find scope in overcoming problems for noisy and imbalanced data using high-level feature extraction to drive optimization of exploitation and advance even further portfolio performance. This would present the application of machine learning to a discovery process that unveils hidden patterns in investor behavior and provides an alternative for dynamic financial markets to handle the situation. Future developments include more intense emotion recognition models, real-time sentiment analysis, and better reinforcement learning algorithms, among others, in the course of further refining the strategies for asset allocation and in raising the accuracy of forecasts about markets.

Thanks

I, Simrandeep Kaur, have been awarded the Short-Term Research Fellowship for doctoral candidates by Guru Gobind Singh Indraprastha University, New Delhi, India, as evidenced by letter number GGSIPU/DRC/2022/38/1269. This fellowship provides valuable support for my research endeavors.

Nomenclature

x	training points in the optimization problem
ξ	slack variable used to quantify the misclassification of sample points
$\phi(x)$	feature mapping function
$\mathbb{R}^n$	n-dimensional Euclidean space
α	lagrange multiplier
$f(x)/y(x)$	decision function that determines the predicted class label of a sample
s	Fuzzy membership value that represents the degree of belongingness of a sample to a specific class

Abbreviations

ML	machine learning
PCA	principal component analysis
NLQPP	nonlinear quadratic programming optimization problem
SVM	support vector machines
FSVM	fuzzy support vector machines
NFSVM	new fuzzy support vector machines
MV model	Markowitz mean-variance portfolio optimization model

A1. Python code for feature extraction

```
# Function to calculate percentage change in open prices
def PercentageChangeOpen(open_prices, N=1):
    """
    Calculate the percentage change in open prices.

    Args:
    open_prices (list): List of open prices.
    N (int): Lookback period.

    Returns:
    list: List of percentage changes in open prices.
    """
    percentage_changes = open_prices.copy()

    for i in range(N):
        percentage_changes[i] = 0

    for t in range(N, len(open_prices)):
        percentage_changes[t] = (open_prices[t] - open_prices[t -
    N]) / open_prices[t - N]
```

DOI: http://dx.doi.org/10.5772/intechopen.1009112

```
    return percentage_changes

# Function to calculate percentage change in closing prices
def PercentageChangeClose(closing_prices, N=2):
    """
    Calculate the percentage change in closing prices.

    Args:
    closing_prices (list): List of closing prices.
    N (int): Lookback period.

    Returns:
    list: List of percentage changes in closing prices.
    """
    percentage_changes = closing_prices.copy()

    for i in range(N):
        percentage_changes[i] = 0

    for t in range(N, len(closing_prices)):
        percentage_changes[t] = (closing_prices[t - 1] -
    closing_prices[t - N]) / closing_prices[t - N]

    return percentage_changes

# Function to calculate percentage change in high prices
def PercentageChangeHigh(high_prices, N=2):
    """
    Calculate the percentage change in high prices.

    Args:
    high_prices (list): List of high prices.
    N (int): Lookback period.

    Returns:
    list: List of percentage changes in high prices.
    """
    percentage_changes = high_prices.copy()

    for i in range(N):
        percentage_changes[i] = 0

    for t in range(N, len(high_prices)):
        percentage_changes[t] = (high_prices[t - 1] - high_prices[
    t - N]) / high_prices[t - N]

    return percentage_changes

# Function to calculate percentage change in low prices
def PercentageChangeLow(low_prices, N=2):
    """
```

```
    Calculate the percentage change in low prices.

    Args:
    low_prices (list): List of low prices.
    N (int): Lookback period.

    Returns:
    list: List of percentage changes in low prices.
    """
    percentage_changes = low_prices.copy()

    for i in range(N):
        percentage_changes[i] = 0

    for t in range(N, len(low_prices)):
        percentage_changes[t] = (low_prices[t - 1] - low_prices[t
    - N]) / low_prices[t - N]

    return percentage_changes

# Function to calculate percentage change in volume
def PercentageChangeVolume(volume, N=2):
    """
    Calculate the percentage change in trading volume.

    Args:
    volume (list): List of trading volumes.
    N (int): Lookback period.

    Returns:
    list: List of percentage changes in trading volume.
    """
    percentage_changes = volume.copy()

    for i in range(N):
        percentage_changes[i] = 0

    for t in range(N, len(volume)):
        percentage_changes[t] = (volume[t - 1] - volume[t - N]) /
    volume[t - N]

    return percentage_changes

# Function to calculate the fractional change in open prices
def FractionalChangeInOpen(open_prices, N=5):
    """
    Calculate the fractional change in open prices.

    Args:
    open_prices (list): List of open prices.
    N (int): Lookback period.
```

```
    Returns:
    list: List of fractional changes in open prices.
    """
    high_prices = [0 for i in range(len(open_prices))]
    low_prices = [0 for i in range(len(open_prices))]
    fractional_changes = [0 for i in range(len(open_prices))]

    for i in range(N + 1, len(open_prices) + 1):
        high_prices[i - 1] = max(open_prices[i - 1 - N:i - 1])
        low_prices[i - 1] = min(open_prices[i - 1 - N:i - 1])
        fractional_changes[i - 1] = ((open_prices[i - 1] -
    open_prices[i - 2]) / (high_prices[i - 1] - low_prices[i - 1])
    )

    return fractional_changes

# Function to calculate the fractional change in volume
def FractionalChangeInVolume(volume, N=5):
    """
    Calculate the fractional change in trading volume.

    Args:
    volume (list): List of trading volumes.
    N (int): Lookback period.

    Returns:
    list: List of fractional changes in trading volume.
    """
    high_prices = [0 for i in range(len(volume))]
    low_prices = [0 for i in range(len(volume))]
    fractional_changes = [0 for i in range(len(volume))]

    for i in range(N + 1, len(volume) + 1):
        high_prices[i - 1] = max(volume[i - 1 - N:i - 1])
        low_prices[i - 1] = min(volume[i - 1 - N:i - 1])
        fractional_changes[i - 1] = ((volume[i - 2] - volume[i -
    3]) / (high_prices[i - 1] - low_prices[i - 1]))

    return fractional_changes
```

A2. Python code for new fuzzy membership function

```
# Function to calculate fuzzy membership
def new_fuzzy_membership(X,y, delta):
    l = len(X)
    Q_plus = []
    Q_minus = []

    # Separate the samples into Q_plus and Q_minus classes
```

```
for i in range(l):
    if y[i] == 1:
        Q_plus.append(X[i])
    else:
        Q_minus.append(X[i])

l1 = len(Q_plus)
l2 = len(Q_minus)
Q_plus = np.array(Q_plus)
Q_minus = np.array(Q_minus)

def rbf_kernel(x1, x2, gamma):
    return np.exp(-gamma * np.linalg.norm(x1 - x2) ** 2)

# Calculate the class centers
phi_cen_plus = np.sum(Q_plus, axis=0) / l1
phi_cen_minus = np.sum(Q_minus, axis=0) / l2

def distance_to_class_center(xi, Q, gamma):
    l = len(Q)

    term1 = rbf_kernel(xi, xi, gamma)
    term2 = (2 / l) * np.sum([rbf_kernel(xi, xj, gamma) for xj
 in Q])
    term3 = (1 / (l ** 2)) * np.sum([rbf_kernel(xi1, xi2,
gamma) for xi1 in Q for xi2 in Q])

    return term1 - term2 + term3

distances_xi_plus_to_phi_cen_plus = [distance_to_class_center(
xi_plus, Q_plus, gamma) for xi_plus in Q_plus]

distances_xi_plus_to_phi_cen_minus = [distance_to_class_center
(xi_minus, Q_minus, gamma) for xi_minus in Q_minus]

distances_xj_minus_to_phi_cen_minus = [
distance_to_class_center(xi_minus, Q_minus, gamma) for
xi_minus in Q_minus]

distances_xj_minus_to_phi_cen_plus = [distance_to_class_center
(xj_plus, Q_plus, gamma) for xj_plus in Q_plus]

distances_xi_minus_to_phi_cen_minus = [
distance_to_class_center(xi_minus, Q_minus, gamma) for
xi_minus in Q_minus]

distances_xi_minus_to_phi_cen_plus = [distance_to_class_center
(xi_plus, Q_plus, gamma) for xi_plus in Q_plus]

ddsquare_xi_plus_xistar_minus = np.array([rbf_kernel(xi_plus,
xi_plus, gamma) - 2 * rbf_kernel(xi_plus, xj_minus,gamma)+
```

DOI: http://dx.doi.org/10.5772/intechopen.1009112

```
rbf_kernel(xj_minus, xj_minus, gamma) for xj_minus in Q_minus
for xi_plus in Q_plus])
 ddsquare_xi_plus_xistar_minus = np.min(
ddsquare_xi_plus_xistar_minus)

 ddsquare_xj_star_plus_xjstar_minus = np.array([rbf_kernel(
xj_star_plus, xj_star_plus, gamma) - 2 * rbf_kernel(
xj_star_plus, xi_minus, gamma) + rbf_kernel(xi_minus, xi_minus
, gamma) for xi_minus in Q_minus for xj_star_plus in Q_plus])
 ddsquare_xj_star_plus_xjstar_minus = np.min(
ddsquare_xj_star_plus_xjstar_minus)

 ddsquare_xi_star_plus_xistar_minus = np.array([rbf_kernel(
xi_star_plus, xi_star_plus, gamma) - 2 * rbf_kernel(
xi_star_plus, xj_minus, gamma) + rbf_kernel(xj_minus, xj_minus
, gamma) for xj_minus in Q_minus for xi_star_plus in Q_plus])
 ddsquare_xi_star_plus_xistar_minus = np.min(
ddsquare_xi_star_plus_xistar_minus)

 ddsquare_xj_minus_xjstar_plus = np.array([rbf_kernel(xj_minus,
  xj_minus, gamma) - 2 * rbf_kernel(xj_minus, xi_plus,gamma) +
rbf_kernel(xi_plus, xi_plus, gamma) for xi_plus in Q_plus for
xj_minus in Q_minus])
 ddsquare_xj_minus_xjstar_plus = np.min(
ddsquare_xj_minus_xjstar_plus)

 ddsquare_xi_minus_xistar_plus = np.array([rbf_kernel(xi_minus,
  xi_minus, gamma) - 2 * rbf_kernel(xi_minus, xj_plus,gamma)+
rbf_kernel(xj_plus, xj_plus, gamma) for xj_plus in Q_plus for
xi_minus in Q_minus])

 ddsquare_xi_minus_xistar_plus = np.min(
ddsquare_xi_minus_xistar_plus))


 ddsquare_xi_star_minus_xi_star_plus = np.array([rbf_kernel(
xi_star_minus, xi_star_minus, gamma) - 2 * rbf_kernel(
xi_star_minus, xj_plus, gamma)+ rbf_kernel(xj_plus, xj_plus,
gamma) for xj_plus in Q_plus for xi_star_minus in Q_minus])
 ddsquare_xi_star_minus_xi_star_plus = np.min(
ddsquare_xi_star_minus_xi_star_plus)


 dd_rel_xi_plus = (np.sqrt(ddsquare_xi_plus_xistar_minus) + 1)
/ (np.sqrt(ddsquare_xi_star_minus_xi_star_plus) + 1)
 dd_rel_xj_minus = (np.sqrt(ddsquare_xj_minus_xjstar_plus) + 1)
  / (np.sqrt(ddsquare_xj_star_plus_xjstar_minus) + 1)
 dd_rel_xi_minus = (np.sqrt(ddsquare_xi_minus_xistar_plus) + 1)
  / (np.sqrt(ddsquare_xi_star_plus_xistar_minus) + 1)


 s = np.zeros(1)  # Fuzzy membership array
```

```
    for i in range(l):
        if y[i] == 1:
            if np.sqrt(np.linalg.norm(
distance_xi_plus_to_phi_cen_plus)) <= np.sqrt(np.linalg.norm(
distance_xi_plus_to_phi_cen_minus)):
                s[i] = 1 / (dd_rel_xi_plus)
            else:
                s[i] = 1 - 1 / (dd_rel_xi_plus + eta)
        elif y[i] == -1:
            if np.sqrt(np.linalg.norm(
distance_xi_minus_to_phi_cen_minus)) <= np.sqrt(np.linalg.norm
(distance_xi_minus_to_phi_cen_plus)):
                s[i] = 1 / dd_rel_xi_minus
            else:
                s[i] = 1 - 1 / (dd_rel_xi_minus + eta)
    return s
```

A3. Data statistics of Nifty 50 and Euro Stoxx 50 assets

For a better understanding of Nifty 50 (A1-A48) and Euro Stoxx 50 (B1-B47) assets, please refer to [1] and below **Table 5**.

(a) Nifty 50 Assets							
Nifty 50							
Assets	**Mean**	**Std**	**Max**	**Min**	**Range**	**Kurtosis**	**Skewness**
A1	1727.85	692.34	3834.55	490.90	3343.65	0.92	0.78
A2	4029.18	771.15	5733.95	2409.25	3324.70	−0.83	−0.20
A3	735.60	79.37	970.25	496.80	473.45	1.93	−0.36
A4	3003.63	310.27	3576.30	2277.20	1299.10	−0.70	−0.44
A5	728.65	47.89	845.10	623.80	221.30	−0.79	−0.18
A6	3758.43	256.11	4295.05	3105.20	1189.85	−0.33	−0.15
A7	6444.16	903.44	7929.30	4479.60	3449.70	−1.22	−0.18
A8	1407.25	300.27	1907.66	835.46	1072.20	−1.34	−0.22
A9	384.08	42.61	462.52	296.05	166.47	−1.05	−0.17
A10	3577.03	204.20	4124.20	3094.50	1029.70	0.16	0.34
A11	646.09	85.55	799.90	498.39	301.51	−1.34	−0.36
A12	933.56	67.54	1115.85	755.15	360.70	0.08	−0.15
A13	166.57	27.71	238.85	124.50	114.35	−0.33	0.70
A14	4210.75	543.85	5372.15	3272.90	2099.25	−1.14	0.25
A15	2726.26	285.28	3748.75	2256.05	1492.70	2.30	1.47
A16	1518.51	196.39	1912.74	928.03	984.72	0.27	−0.65
A17	421.50	82.76	631.30	226.30	405.00	−0.04	0.08

(a) Nifty 50 Assets							
A18	1057.92	121.29	1358.20	883.15	475.05	−0.86	0.56
A19	1477.33	82.32	1688.70	1281.30	407.40	−0.32	−0.07
A20	2517.04	210.36	3000.85	2052.70	948.15	−0.70	0.05
A21	2382.59	183.22	2812.45	1943.95	868.50	−0.26	0.24
A22	2782.32	277.61	3571.38	2198.70	1372.68	0.44	0.52
A23	709.91	88.50	918.10	522.35	395.75	−0.36	0.11
A24	984.07	97.69	1264.10	777.65	486.45	−0.11	0.45
A25	1567.80	181.61	1939.50	1239.05	700.45	−1.15	0.09
A26	239.03	37.05	346.40	199.60	146.80	0.36	1.18
A27	627.42	101.44	767.85	366.85	401.00	0.52	−1.24
A28	1837.76	106.77	2210.95	1641.65	569.30	0.19	0.74
A29	1664.06	189.35	2069.40	1297.00	772.40	−1.07	−0.06
A30	895.43	157.49	1321.90	690.55	631.35	0.98	1.44
A31	7649.29	687.19	9401.85	6455.65	2946.20	−0.59	0.56
A32	18170.86	1076.42	20457.20	16094.25	4362.95	−1.01	0.08
A33	129.77	20.20	175.10	88.95	86.15	−0.98	0.11
A34	135.03	23.61	186.95	88.30	98.65	−1.14	0.09
A35	192.29	26.81	245.15	136.33	108.82	−1.20	−0.06
A36	4623.74	420.27	5575.70	3718.85	1856.85	−0.69	0.35
A37	2337.27	247.18	2819.85	1841.95	977.90	−1.14	−0.22
A38	451.49	65.40	574.05	275.65	298.40	0.02	−0.67
A39	778.54	107.43	948.65	564.35	384.30	−1.15	−0.38
A40	740.06	76.19	882.10	559.80	322.30	−0.70	−0.39
A41	390.22	81.36	530.15	186.50	343.65	−1.19	−0.17
A42	3393.52	268.21	4019.15	2894.30	1124.85	−1.05	0.34
A43	1208.98	245.08	1780.21	885.04	895.17	−1.13	0.50
A44	110.14	21.77	151.94	60.10	91.84	−0.53	−0.51
A45	2077.99	414.00	2737.60	1402.90	1334.70	−1.44	−0.23
A46	6737.80	691.24	8214.05	5177.30	3036.75	−0.71	−0.13
A47	720.92	85.29	853.55	469.30	384.25	0.16	−0.91
A48	530.45	96.91	721.50	388.10	333.40	−1.32	0.19
(b) Euro Stoxx 50 Assets							
B1	53.83	3.99	65.34	46.36	18.99	0.34	0.80
B2	244.20	54.74	336.25	118.88	217.37	−1.09	−0.52
B3	26.59	2.39	31.09	21.72	9.37	−1.08	−0.16
B4	132.48	7.99	150.60	113.36	37.24	−0.48	0.12
B5	105.13	8.38	120.40	83.11	37.29	−0.81	−0.38

(b) Euro Stoxx 50 Assets							
B6	201.37	15.71	231.95	159.62	72.33	−0.46	−0.49
B7	568.81	89.98	770.50	403.90	366.60	−0.84	0.33
B8	23.58	2.16	28.75	18.30	10.45	−0.03	0.09
B9	2.95	0.29	3.51	2.34	1.17	−0.91	−0.26
B10	59.69	9.98	72.61	38.85	33.76	−0.96	−0.71
B11	82.04	7.48	99.32	68.34	30.98	−0.86	0.21
B12	53.32	5.32	67.73	44.26	23.48	0.29	0.85
B13	5.00	0.57	6.28	3.71	2.57	−0.95	−0.01
B14	52.48	5.58	66.96	39.78	27.18	−0.10	0.29
B15	39.88	3.69	48.03	32.05	15.99	−0.94	−0.05
B16	56.02	3.36	65.00	47.48	17.52	−0.24	0.09
B17	47.46	8.29	60.87	30.00	30.87	−1.35	−0.11
B18	150.11	11.34	174.90	131.75	43.15	−0.97	0.41
B19	17.16	1.23	19.55	14.70	4.85	−0.85	−0.13
B20	6.89	1.23	8.95	4.22	4.72	−1.07	−0.30
B21	11.46	1.52	14.53	8.20	6.33	−0.95	0.05
B22	155.08	16.52	193.36	117.00	76.36	−0.59	0.13
B23	139.20	28.79	196.75	89.58	107.17	−1.25	−0.07
B24	1198.10	183.91	1675.50	842.60	832.90	−0.18	0.21
B25	10.11	0.59	11.76	8.38	3.38	−0.05	−0.10
B26	26.56	3.83	32.54	19.71	12.83	−1.24	−0.25
B27	32.42	5.18	43.46	21.07	22.40	−0.70	−0.14
B28	10.60	1.60	13.91	7.30	6.61	−0.90	0.16
B29	2.18	0.28	2.92	1.61	1.31	−0.85	−0.10
B30	612.58	89.15	792.10	442.45	349.65	−1.11	0.06
B31	357.14	33.39	431.40	290.10	141.30	−0.79	0.09
B32	631.74	58.70	758.00	491.05	266.95	−0.58	−0.42
B33	62.56	6.53	76.19	46.57	29.62	−0.24	−0.05
B34	242.31	14.98	280.20	209.15	71.05	−0.54	0.37
B35	4.59	0.62	5.61	3.15	2.47	−0.31	−0.92
B36	9.42	1.09	11.45	6.60	4.85	−0.06	−0.61
B37	184.05	14.45	216.40	151.95	64.45	−0.41	−0.23
B38	109.38	8.74	126.54	88.34	38.20	−0.68	−0.10
B39	88.04	6.92	105.67	74.96	30.71	−0.56	0.46
B40	107.39	13.29	128.28	81.06	47.22	−1.24	−0.22
B41	137.41	14.08	177.82	111.26	66.56	−0.49	0.44
B42	129.58	15.48	157.96	93.74	64.22	−0.67	−0.61
B43	14.99	1.97	19.14	11.20	7.94	−1.20	−0.01
B44	43.73	5.73	56.52	34.56	21.97	−1.19	0.23

(b) Euro Stoxx 50 Assets							
B45	90.49	4.23	103.00	76.58	26.42	0.87	−0.15
B46	178.49	30.73	246.55	120.64	125.91	−0.99	0.24
B47	44.96	9.32	56.64	21.03	35.61	−0.38	−1.00

Table 5.
Descriptive statistics of Nifty 50 and Euro Stoxx 50 assets.

Author details

Simrandeep Kaur†
University School of Basic and Applied Sciences, Guru Gobind Singh Indraprastha University, Delhi, India

*Address all correspondence to: simrandeepkaur107@gmail.com

†These authors contributed equally.

References

[1] Kaur S, Singh A, Aggarwal A. Mean-variance optimal portfolio selection integrated with support vector and fuzzy support vector machines. Journal of Fuzzy Extension and Applications. 2024;**5**(3):434-468. Available from: https://www.journal-fea.com/article_202834.html [Accessed: September 11, 2024]

[2] De Long J, Bradford Shleifer A, Summers LH, Waldmann RJ. Noise trader risk in financial markets. Journal of Political Economy. 1990;**98**(4):703-738. Available from: https://www.jstor.org/stable/2937765 [Accessed: January 15, 2024]

[3] Baker M, Wurgler J. Investor sentiment in the stock market. Journal of Economic Perspectives. 2007;**21**(2):129-151. Available from: https://www.jstor.org/stable/30033721 [Accessed: February 25, 2024]

[4] Chandra S, Jayadeva
Mehra A. Numerical Optimization with Applications. New Delhi, India: Narosa Publishing House; 2009. Available from: https://books.google.co.in/books?id=MogtPwAACAAJ [Accessed: January 7, 2022]

[5] Lin CF, Wang SD. Fuzzy support vector machines. IEEE Transactions on Neural Networks. 2002;**13**(2):464-471. DOI: 10.1109/72.991432 [Accessed: 28 May 2022]

[6] Rai P. Learning Maximum-Margin Hyperplanes: Support Vector Machines (SVMs). 2016. Available from: https://cse.iitk.ac.in/users/piyush/courses/ml_autumn16/771A_lec8_slides.pdf [Accessed: July 3, 2022]

[7] Tang WM. Fuzzy SVM with a new fuzzy membership function to solve the two-class problems. Neural Processing Letters. 2011;**34**:209-219. DOI: 10.1007/s11063-011-9192-y [Accessed: February 2, 2023]

Chapter 3

Perspective Chapter: The Faint Whispers of the Heart – Next-Generation Sentiment Analysis and "Empathy Algorithms"

Bernard Mallia

Abstract

Sentiment analysis has evolved from polarity detection into sophisticated multimodal systems—"empathy algorithms"—capable of capturing subtle, even subconscious emotional signals from diverse inputs such as text, facial micro-expressions, vocal intonation, and physiological indicators. Advances in transformer-based neural architectures, multimodal fusion strategies, and neural-symbolic approaches have enabled these systems to capture subtle affective signals, including sarcasm, subconscious emotional cues, and culturally inflected expressions, with fidelity that is significantly better than what has previously been possible. Yet, significant challenges persist, particularly in interpreting sarcasm, cultural diversity, linguistic nuances, and ensuring fairness across demographic groups. Concomitantly, explainability and fairness have emerged as critical design imperatives, driving model interpretation technique adoption, adversarial debiasing, and local rule engines to counteract cultural biases and ensure equitable performance across demographic groups. Researchers have begun integrating local rule-based modules that inject culturally specific knowledge into the model's predictions. Realising the full promise of empathetic AI also requires addressing the complex logistical and ethical challenges posed by massive, often sensitive datasets, making privacy-preserving paradigms paramount against the backdrop of evolving regulatory frameworks. In this evolving ecosystem, consilience across sensor miniaturisation, hardware accelerators, quantum computing, and self-supervised or few-shot multimodal learning offer promising avenues for real-time detection of intricate emotional states. In this spirit, the chapter provides a roadmap of current developments, spotlighting key open questions in scaling empathic, socially responsible sentiment analysis systems. Looking ahead, the consilience of several technologies at the software and hardware level promises further breakthroughs, potentially transforming sentiment analysis into empathic systems indistinguishable from human empathy.

Keywords: sentiment analysis, empathy algorithms, transformer models, multimodal integration, emotion recognition, attention mechanisms, sarcasm detection, GDPR, EU AI act, privacy, cultural diversity, ethical AI, self-supervised learning, federated learning, quantum computing

1. Introduction

Over the last two decades, Natural Language Processing (NLP) has undergone a number of profound transformations, elevating sentiment analysis from a rudimentary classification task – primarily differentiating text into positive, negative, or neutral categories – into a sophisticated discipline that incorporates neural architectures, attention mechanisms, transfer learning techniques and now even multimodality. This paradigm shift, which is by no means over, has been propelled by major breakthroughs in large-scale language modelling, exemplified by BERT,[1] which excels at contextualising word meanings within sentences, and GPT,[2] which aces the generation of coherent and contextually-nuanced text. Such models have propelled sentiment analysis far beyond simple polarity detection, and have thus enabled richer characterisations of sentiment across diverse languages and domains by capturing linguistic subtleties such as sarcasm and context-dependent usage.

Historically, early sentiment analysis systems leveraged relatively straightforward lexical or rule-based approaches. These included sentiment dictionaries and frequency-based methods that performed adequately for applications such as marketing insights, brand monitoring, basic social media analytics, and even financial leading indicator forecasting. Although tools such as SPSS, VADER, SentiWordNet, as well as early iterations of IBM Watson, offered valuable insights at a time when erratic manual processing was otherwise required, they frequently proved limited in parsing complex linguistic phenomena such as sarcasm, context-dependent word usage, and culturally-informed semantics, ultimately offering only a partial – if still very useful – understanding of human affect.[3] It became evident soon thereafter that human affect is far more intricate than

---

1 BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained deep neural network architecture that utilises a transformer-based mechanism to capture bidirectional contextual information from text. BERT was originally introduced by Devlin et al. [1], and is widely-acknowledged to have significantly advanced natural language processing tasks, not least sentiment analysis, by enabling more nuanced understanding of context and semantic relationships.

2 GPT (Generative Pre-trained Transformer) refers to a series of autoregressive language models developed by OpenAI. Beginning with the original GPT and conspicuously expanded by the time GPT-3 was released [2], these models generate coherent and contextually-relevant text by leveraging extensive pre-training on diverse textual datasets, thereby demonstrating advanced capabilities in text generation and language comprehension.

3 Lexicon-based methods, such as VADER or early dictionary-based approaches in SPSS, conferred a high degree of interpretability because each sentiment-bearing word used to be explicitly matched to a predefined polarity score. This clarity allowed analysts to trace classification outcomes back to the precise lexical items responsible for any given sentiment label. By contrast, neural-based systems – particularly those leveraging distributed embeddings and deep architectures – excel at capturing more subtle, context-dependent cues in language but tend to operate as the proverbial "black boxes" that has become

a unidimensional polarity label could ever be able to capture, and this eventually ushered in newer and more complex generations of tools and models that power them.

Indeed, recent advances in deep learning and pre-trained neural models have opened entirely new avenues for sentiment analysis. By way of a practical example, BERT's ability to reflect how word meanings shift within a sentence, combined with the fluency in text generation demonstrated by GPT-3 and its successors, facilitates a more refined interpretation of sentiment and emotional context. Moreover, multimodal emotion recognition, defined here as the integration of textual input with non-textual cues from microphone and video feeds such as vocal intonations, facial micro-expressions, and physiological signals, are equally significant, and offer valuable complementary insights into a person's emotional state at that point in time. Empirical studies suggest that subtle variations in vocal pitch may indicate rising stress levels or moral sentiment, minor changes in facial musculature can reveal unspoken emotional undercurrents, and eye-movement data may signify heightened cognitive load. Such cues, to which purely text-focused approaches are both blind and deaf, can be missed by even by the most empathic of human observers. This makes their systematic, real-time detection and analysis a significant enhancement of our understanding of affective communication. This joint embedding approach also provides robustness in that if one modality lacks clear affective signals, another can compensate, as in the phrase "I'm fine", which might be neutral or misleading in text alone, but is unequivocal when coupled with a quavering voice and downcast facial expression, something only a multimodal model can infer correctly as distress. In such cases, the visual and audio embeddings 'fall back' to reveal sentiment that the textual modality hides, illustrating the strength of fused representations. The evocative metaphor of the "faint whispers of the heart" in the title of this chapter underscores the relatively novel aspiration to design "empathy algorithms" – advanced AI-driven systems capable of decoding both explicit affective states and the subtler, even subconscious, signals individuals broadcast through language, vocal tone, micro-gestures, and physiological responses.[4] In this context, explicit technical mechanisms, such as attention mechanisms that focus on contextually-significant cues, are directly linked to improved detection of subliminal sentiment, thereby bridging the gap between observable linguistic patterns and the underlying emotional states. GPT-4 – a large language model who many are familiar with that was not explicitly trained for sentiment – can classify sentiment with high accuracy simply by being prompted with a few examples (the so-called 'few-shot' learning). This behaviour, first highlighted by Brown et al. [2] with GPT-3, shows that massive pretrained models can adapt to sentiment tasks on the fly, and in practical terms, one can feed GPT-4 a short prompt like: "Review: 'The product broke in a day. Terrible quality.' Sentiment:", and it will correctly continue with "Negative" without

---

synonymous with most of the AI development taking place today. While their capacity to model intricate linguistic phenomena is unparalleled, it often comes at the cost of direct interpretability, complicating both debugging and sometimes trust in automated sentiment decisions. This is an issue that has not been solved yet, but to which significant research budgets are being dedicated. The present author will, in the near future, be publishing a paper on this topic titled "A Sneak Peek Inside AI's Mind: Decoding Neural Decision Pathways with Causal Explanation Graphs" that proposes a new approach for better explainability using a new model called TRACE. The draft is still being worked on and has not yet been submitted for publication, so further details are unavailable.

[39] While recent literature refers to "Affective Computing" and "Artificial Empathy", the term "Empathy Algorithms" should be taken to refer to and to encapsulate the specific convergence of technologies aiming to decipher the subtle, and potentially latent, emotional states across multiple modalities.

requiring any fine-tuning whatsoever. Such capabilities underscore the paradigm shift toward versatile foundation models generally, but also in the area of sentiment analysis.

Tencent's utilisation of sentiment analysis within WeChat to identify and pre-empt user dissatisfaction, a *de facto* enabler of tailored interventions and enhanced user experience, serves as a good practical example of this,[5] identifying and pre-empting user dissatisfaction by detecting early negative sentiment in user interactions, and triggering tailored interventions to enhance user experience. Similarly, Meta's Facebook has developed systems to monitor written messages (evidently even the ones that are deleted before being sent,[6] although the specifics remain undisclosed), and researchers at Stanford University have demonstrated that AI-based sentiment models can detect early signs of depression in social media posts, offering timely opportunities for mental health interventions. Combined, these signals create an intricately-weaved tapestry of emotional communication that, when modelled effectively, enable algorithms to detect not only what a person says but also the emotions behind what they might not be saying overtly. These examples stress the potential for broader societal impact, ranging from personal well-being to large-scale community monitoring of emotional trends.

The globalisation of sentiment analysis introduces considerable technical challenges in sentiment analysis. While the English language has historically dominated NLP research, multilingual and code-switched contexts are increasingly the focus of study. For instance, Winata et al. [3] demonstrate how transfer learning can bolster Sino-English code-switching sentiment tasks, and Baidu's ERNIE 3.0 exemplifies model architectures specifically tailored to the nuances of Chinese text. Likewise, cutting-edge techniques are boosting sentiment accuracy in Chinese social media, with Li and Li [4] reporting that a graph neural network–enhanced model on Weibo comments achieved over 95% classification accuracy by leveraging syntactic dependency structures to better capture Chinese linguistic nuances – whichever way you look at it, a substantial improvement over prior sequential models.

Code-switching is also prevalent in many other bilingual communities like Italian-English, Spanish–English, Arabic–French, and in many other code-switching linguistic pairs. Notably, Winata et al. [3] provide a comprehensive survey of code-switching research in NLP, highlighting both the progress and persistent challenges in multilingual sentiment analysis. Their review underlines the need for culturally cognisant models that can seamlessly handle mixed-language inputs, consistent with the Sino-English transfer-learning improvements noted above. Such research underscores the importance of culturally-cognisant frameworks that recognise the interplay between language, context, and sociocultural norms. A comparative analysis of performance metrics between early and modern systems further reveals both the technical advancements and the remaining limitations in accurately capturing the full spectrum of human affect.

---

5 Chen, 2023, Journal of AI Applications.

6 Internal studies revealed that 71% of users type content they never post, and Facebook's system silently tracks these instances, ostensibly to better understand user intent and sentiment. Vide https://www.wired.com/story/facebook-is-tracking-what-you-dont-do/#:~:text=Das%20and%20Kramer%20tracked%20the,not%20the%20keystrokes%20or%20content (accessed 13 Feb 2025); https://www.wired.com/story/whats-not-included-in-facebooks-download-your-data/?utm_source=chatgpt.com (accessed 13 Feb 2025) and Dr. Zeynep Tufekci's TED talk available here: https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads (accessed 13 Feb 2025).

These challenges are magnified when sentiment-analysis technologies are embedded in applications that potentially facilitate surveillance or manipulation, further emphasising the need for ethically-grounded frameworks that respect personal autonomy and societal norms.

Notwithstanding these developments, the integration of multimodal data and the expansion of linguistic coverage invariably come with profound ethical concerns and ramifications. Robust regulatory frameworks – exemplified in Europe by the European Union's AI Act, General Data Protection Regulation (GDPR), Digital Services Act (DSA), and Digital Markets Act (DMA); in China by its Personal Information Protection Law (PIPL) [5]; and in the USA through its now-revoked Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,[7,8] underscore the importance of privacy, data security, and consent in multimodal sentiment analysis. In parallel, cultural psychology findings like the one by Senft et al. [6] reveal substantial heterogeneity in how different heritage groups interpret and express emotions. Such between-group differences (documented among people of East Asian, European, and Latino heritage) highlight potential cultural mismatches – raising pressing questions about fairness when deploying a one-size-fits-all emotion recognition model. These ethical imperatives should directly inform the technical design choices of next-generation sentiment analysis systems.

In light of these considerations, this chapter provides a multidimensional overview of next-generation sentiment analysis. It interweaves the technical foundations of neural architectures and pre-trained language models with the complexities of multimodal input integration and cultural adaptation. Additionally, it explains the ethical imperatives that shape the responsible deployment of "empathy algorithms" by linking regulatory frameworks and empirical findings to design choices. By synthesising developments in transfer learning, multilingual modelling, and nuanced affect detection, this chapter offers a roadmap for researchers and practitioners seeking to develop AI systems capable of "hearing" the faint whispers of the human heart. We conclude by delineating unresolved research gaps – particularly concerning data heterogeneity, model interpretability, cross-cultural evaluation and better models – while emphasising the crucial interplay of ethics and technological innovation necessary for advancing AI's emotional intelligence responsibly.

---

[7] This was an Executive Order signed by USA President Biden on October 31, 2023 and revoked by subsequent USA President Trump on 20 Jan 2025 (*vide* https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/) (accessed 9 April 2025).

[8] Here, I am not taking into account State-Level Regulations like the California Consumer Privacy Act (CCPA) and its successor, the California Privacy Rights Act (CPRA), which have established robust frameworks to protect consumer data, influencing practices in AI and multimodal analysis. These regulations emphasise transparency, data minimisation, and consumer consent, which are particularly pertinent when collating and linking data from diverse sources and in different formats (text, audio, video), but which would be beyond the scope of this chapter.
Neither am I touching on Federal Guidelines, like, say the NIST's work on AI standards and data security that have been credited with shaping best practices and offering technical guidelines that influence how AI systems are developed and deployed. These guidelines, while not legally binding and thus outside of the scope of this discussion, still serve to inform both public and private sector approaches to data privacy and security, providing an important reference point for developing multimodal sentiment analysis systems that comply with evolving expectations in data handling.

2. From traditional sentiment analysis to "empathy algorithms"

2.1 Historical perspectives and foundational work

Classical sentiment analysis systems primarily relied on lexicon-based methods, such as lists of positive and negative words, or supervised machine-learning techniques that utilised manually-crafted features like part-of-speech tags and the presence of emoticons. Rule-based systems relied on sentiment lexicons – collections of words with pre-assigned polarity scores (like, for example, "good" = +1, "bad" = −1). Pioneering methods involved counting occurrences of positive and negative terms and offsetting their final tally to determine the overall sentiment in a document or sentence.

While these approaches achieved reasonable accuracy on early benchmark datasets, such as movie reviews and product sentiment corpora, their widespread use stemmed from the increasing availability of structured data, the relatively straightforward nature of these domains, and their ability to provide early insights into sentiment classification challenges. However, as with many other things in the realm of technology, adoption oftentimes exposes faults and infirmities which failure might otherwise have concealed from observation, and the complexity of human language quickly brought their limitations to light. Although easy to implement and easily interpretable, these intuitive methods had some severe inherent limitations, namely that they were unable to account for context-dependent meanings (for instance, "not good" vs. "good"); and struggled badly in the presence of linguistic phenomena like irony or semantic shift. Nuances such as sarcasm, irony, cultural references, and domain-specific jargon posed even more significant challenges, often leading to misclassification and incomplete interpretations. Indeed, shared task evaluations confirm this challenge, as in the EVALITA 2016 Italian Tweet Sentiment Analysis benchmark, even the top-performing system for irony detection achieved only about 0.54 F1-score (barely above a 0.47 baseline) [7], underscoring how elusive ironic sentiment remained. Such results illustrate that basic polarity models frequently misclassify sarcastic remarks, necessitating more sophisticated approaches [8, 9].

Statistical and machine-learning approaches slowly but surely took precedence, gradually supplanting the preceding rule-based systems. Techniques such as Naïve Bayes, Support Vector Machines, and Logistic Regression leveraged handcrafted features – Term Frequency-Inverse Document Frequency (TF-IDF), part-of-speech tags, presence of intensifiers, emoticons, and domain-specific keywords, improving overall accuracy on benchmark datasets such as movie reviews and product feedback, even though they only captured superficial aspects of emotional expression.

By the mid-2010s, the introduction of distributed representations, particularly Word2Vec,[9] marked a major paradigm shift in sentiment analysis. Word2Vec's dense

---

[9] Word2Vec is a machine learning model introduced by researchers at Google in 2013 (notably by Tomas Mikolov and his team) designed to transform words into dense vector representations in a continuous vector space. It has since become a foundational technique in natural language processing (NLP). Word2Vec originally proposed two distinct architectures, Continuous Bag of Words (CBOW) and Skip-gram, and these represent different learning paradigms.
Word2Vec encodes words into fixed-length vectors such that semantically similar words are positioned closer together in the vector space. These embeddings capture the contextual and semantic relationships between words, enabling the model to infer meanings and relationships even for words that may not have been seen together in the training data.

vector representation, in sharp contradistinction to traditional sparse representations, captured nuanced semantic and syntactic relationships between words thereby enabling models to infer context with far greater precision while being computationally efficient and significantly more effective in handling nuanced language. This advancement ushered in deeper neural architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which could automatically extract sentiment-relevant patterns from raw text. However, even with these remarkable improvements, a single text-based representation often proved insufficient to detect deeply-rooted or subconscious expressions of emotion. Despite the fact that even advanced RNN-based architectures (e.g. LSTMs) suffer from fundamental limitations in retaining long-range context due to issues like vanishing gradients and the sequential nature of their processing, this limitation was effectively overcome by Transformer models, whose self-attention mechanism allows them to capture dependencies across an entire sequence in parallel. The introduction of Transformers thus enabled sentiment models to consider the full context (mitigating the information decay in RNNs) and markedly improved performance on long, complex inputs [10]. These limitations ultimately proved to be the catalysing force for a plethora of research into richer models and alternative modalities that could provide a more holistic understanding of human sentiment.

The advent of word embeddings and distributed representations – including successors to Word2Vec like GloVe, and fastText – introduced a new paradigm between 2013 and 2015, one where instead of treating words as discrete tokens, words were embedded into continuous vector spaces that better captured semantic relatedness and contextual usage. This laid the groundwork for more complex neural architectures that came later and that would ultimately move the field toward deeper contextual analysis.

2.2 Limits of basic polarity detection

Basic polarity-based sentiment analysis often missed or misclassified the nuanced and multifaceted nature of human emotion. A single label such as "negative" ignores distinctions between sadness, anger, fear, anxiety, disappointment or despair, all of which manifest differently in language. Similarly, positive sentiment might take the form of relief, joy, excitement, elation or awe.

For instance, a blend of joy and surprise in a heartfelt message or the restrained disappointment in a politely-worded critique can elude such systems. Similarly, professional emails often mask frustration or sarcasm behind neutral phrasing as it is socially not acceptable to do otherwise in certain organisational cultures, and this presents additional challenges. These emotional categories carry vital information in fields such as mental health assessment, user experience research, and consumer-brand relationships, and yet they proved to be the most elusive under polarity detection frameworks.

Emotions often exist on a spectrum, encompassing states such as joy, surprise, and disgust, which cannot be reduced to simple binary categories. They can also shift fluidly within a single discourse, making static classifications in such circumstances at best inadequate and at worst unfit for the job. The need to detect subtler cues, handle mixed emotions within a single text, and address cultural nuances motivated a shift toward fine-grained and domain-focused approaches.

Word2Vec laid the foundation for subsequent advancements in NLP, including context-sensitive models like GloVe, fastText, and modern transformer-based models like BERT and GPT, which address Word2Vec's limitations, particularly with contextual embeddings and dynamic representation of words.

The domain dependence of language was no less challenging. Words or phrases can reverse valence in different contexts. For instance, "wicked" might be negative in a general context but can be positive slang in certain dialects or cultural subgroups. Similarly, the same adjective describing a child's behaviour ("naughty child") vs. describing a new tech device's capabilities ("naughty good camera!") can imply very different shades of sentiment. Early frameworks often missed these contextual subtleties, spurring researchers to develop more context-aware solutions.

Multilingual data adds another layer of complexity, particularly in low-resource languages where annotated datasets are scarce, and emotion expression often relies heavily on cultural and linguistic nuances. For example, the Japanese phrase 'shōganai' (しょうがない) conveys resignation or acceptance, while the German term 'doch' communicates subtle affirmation or contradiction. These cultural intricacies are very challenging for simplistic models. Taking Japanese as an illustrative example, it is a language that conveys emotion through nuanced phrasing that can be challenging to interpret even for non-native speakers, let alone simplistic models. Furthermore, implied or subtle sentiment, as seen in statements like "I suppose it could have been worse" or a carefully-worded professional emails that subtly convey disappointment, often eludes basic classifiers due to the ambiguity of linguistic markers, making these challenges particularly acute in professional and social media contexts. Even within the English language, linguistic bias is omnipresent, particularly when non-native speakers interpret words differently or assign different meanings deriving from their mother tongues in the first place relative to the meanings ascribed or inferred by native speakers. This is further compounded by regional variations among native English speakers themselves, such as those in Australia, the USA, and the UK, which can result in divergent interpretations of the same terms even among native English speakers.

Another set of challenges is presented by tonal languages like Chinese which have driven research in areas such as syntax-sensitive models and knowledge graph integration [11, 12]. Zhang et al. [13] emphasise how Chinese syntactic structures complicate sentiment classification, demonstrating how these linguistic intricacies have driven advancements in model architecture and data representation strategies while highlighting unique challenges, such as handling tonal languages and context-dependent characters. They also discuss the complexities of Chinese syntactic structures in sentiment classification, while Li and Li [4] explored sentiment analysis on Weibo, uncovering cultural nuances in emotional expression.

These challenges underscore the need for advanced systems capable of moving beyond surface-level sentiment to uncover the deeper emotional and psychological states embedded within language. Without these advanced systems, it would be very difficult to decode the cognitive, emotional and holistic mental states behind the linguistic manifestation being explored.

2.3 Neural architectures and pre-trained language models

The methodological evolution of sentiment analysis is intricately and inextricably comingled with the rapid advancements in neural network architectures and the advent of pre-trained language models. The lion's share of early approaches in sentiment analysis relied on statistical methods and shallow neural networks that utilised handcrafted features, such as Term Frequency-Inverse Document Frequency

(TF-IDF),[10] 'part-of-speech' tags, and sentiment lexicons, and were computationally-efficient and interpretable, but also inherently limited in capturing the rich contextual dependencies and semantic distinctions and gradations inherent in natural language.

Thereafter, the introduction of distributed representations ushered in a significant departure from traditional feature engineering. Methods such as Word2Vec [14] and GloVe [15] enabled words to be represented as dense vectors within continuous vector spaces, thereby capturing latent semantic relationships while facilitating the transition to deeper neural architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Notwithstanding the fact that these were capable of learning hierarchical representations of text, models built on recurrent architectures were themselves beset by challenges such as vanishing gradients[11] and computational inefficiencies when processing long-range dependencies.

A pivotal breakthrough came with the introduction of the Transformer architecture [16], which saw transformers departing from the sequential processing paradigm of RNNs by leveraging a self-attention mechanism, and which in turn enabled the model to consider all input tokens simultaneously and to capture long-range dependencies more effectively. Within the Transformer architecture, each token simultaneously generates three vectors: a query vector, used to compute attention scores over keys to decide where the model should focus; a key vector, evaluated by other tokens' queries to determine the relevance of each token; and a value vector, which carries the content that is aggregated based on attention weights. This explicit decomposition facilitates dynamic contextualisation that is key for advanced sentiment analysis and acts as the enabler of a self-attention mechanism that computes a set of attention weights which then go on to determine the importance of each token relative to others within the same sequence. This not only mitigates the issues of vanishing gradients, but also enhances parallelisability, as a corollary of which training times are significantly reduced.

Building on the Transformer framework, pre-trained language models such as BERT and GPT have redefined the capabilities of NLP systems.

BERT, pioneered by Devlin et al. [1], utilises a bidirectional self-attention mechanism that allows it to contextualise word representations by considering both preceding and succeeding tokens simultaneously, a bidirectional approach particularly advantageous in capturing context-dependent meanings and disambiguating polysemous words that enhances tasks such as sentiment analysis where subtle shifts in meaning can be highly critical. BERT's architecture consists of multiple stacked transformer encoder layers, each refining the contextual embeddings through successive layers of self-attention and feed-forward networks.

---

[10] Term Frequency–Inverse Document Frequency (TF-IDF) is a statistical measure that evaluates the importance of a word in a document relative to a larger corpus calculated as the product of the term frequency (the number of times a term appears in a document) and the inverse document frequency (a logarithmic measure that decreases the weight of terms occurring in many documents). It is considered to be an instrumental metric in information retrieval and text mining, as it downweights common words while emphasising terms that are more discriminative and contextually-significant.

[11] Vanishing gradients refer to a deep neural network phenomenon whereby gradients – computed during backpropagation – diminish exponentially as they are propagated through successive layers. This decay hampers effective weight updates, particularly in earlier layers, thereby impeding the learning process and often necessitating architectural modifications or specialised training strategies. Residual connections and normalisation layers in transformers can alleviate these issues by enabling better gradient flow, which is why transformers have supplanted LSTMs in several tasks.

Conversely, GPT models, starting with the original GPT and later significantly expanded in GPT-3 [2],[12] adopt an autoregressive approach based on transformer decoder blocks where the next token in a sequence is predicted by utilising a unidirectional attention mechanism that considers the sequence of the preceding tokens. Despite this unidirectionality, GPT models have demonstrated remarkable fluency and contextual coherence in text generation tasks. Their capability to generate extended passages of contextually-relevant text has rendered them particularly effective in applications requiring narrative coherence and the generation of nuanced sentiment expressions.

A noteworthy innovation within these transformer-based architectures is the attention mechanism itself, in that within the transformer model, attention is implemented through the computation of queries, keys, and values for each token, with the resulting dot-product attention scores subsequently normalised (typically using the softmax function[13]) to yield attention weights that allow the model to focus on parts of the input sequence that are most relevant for the current prediction dynamically, a capability that has proven essential in sentiment analysis, where the significance of particular words or phrases can vary dramatically depending on their context within the sentence or the document.

Furthermore, and without going into any of the inherent legal (particularly copyright and privacy laws) ramifications, the practice of pre-training on vast, heterogeneous corpora followed by fine-tuning on task-specific data has been instrumental in the success of both BERT and GPT, where pre-training has allowed these models to acquire a general understanding of language structure and semantics that can subsequently be specialised through fine-tuning on sentiment-specific datasets. This transfer learning paradigm has proven especially effective in low-resource settings and in applications involving multilingual data or domain-specific language, thereby broadening the applicability of advanced sentiment analysis techniques to a very significant extent.

The journey from early statistical methods to the sophisticated transformer-based models makes for an important, often underestimated paradigm shift in how sentiment is understood and processed. By harnessing the power of attention mechanisms and large-scale pre-training, contemporary models such as BERT and GPT have significantly advanced the field, enabling a more nuanced, context-aware, and robust interpretation of sentiment and its analysis that not only addresses the shortcomings of previous approaches, but also – and perhaps more importantly – lays the foundation for the development of “empathy algorithms” capable of capturing the subtle, often subconscious, emotional cues embedded in human cognition and communication.

---

[12] GPT-2 and GPT-3 incorporated significantly larger architectures and datasets compared to their predecessor GPT-1.

[13] The softmax function is a mathematical transformation that converts a vector of real-valued inputs into a probability distribution, ensuring that all output values are non-negative and sum to one, a feature achieved by exponentiating each input and normalising by the total sum of these exponentials, making it particularly useful in multi-class classification settings where outputs are interpreted as probabilities.

Given an input vector z = (z1, z2, … , zK), the softmax function σ maps z to a probability distribution over K classes by computing:

$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$ for i = 1, 2, … , K.

This formulation ensures that each output $\sigma(z)_i$ is in the interval (0,1) and that the sum of all outputs is equal to one, thus constituting a valid probability distribution function.

A practical example of Self-Attention Heatmaps (also referred to as "Transformer Lens") showing the raw text and its Self-Attention Heatmap is provided in **Figure 1**.

2.4 Multimodal integration techniques

The conflation of textual and non-textual data has emerged as a critical research frontier in sentiment analysis, driven by the recognition that human affect is conveyed through multiple channels.[14] Multimodal integration techniques endeavour to combine information from an increasingly diverse array of sources, such as textual content, eye-movements and pupil changes, audio signals (e.g., vocal intonations), visual cues (e.g., facial expressions, body language), and physiological signals (e.g., heart rate variability, galvanic skin response and brain area activity[15]) – to construct a more encompassing and holistic representation of emotional states.

---

[14] It is worth noting that because such processing can potentially (and most of the time does *de facto*) reveal sensitive information or serve predictive functions about an individual's affective states, most of the applications of this are likely to be illegal in several, if not most, of its forms unless there is full user consent under EU law, which places particular emphasis on whether personal data or special categories of data are involved under GDPR, and faces several additional and sometimes still ambiguous restrictions under the EU AI Act. It might also have implications under article 102 of the Treaty of the Founding of the EU (TFEU), as well as under China's, the USA's and several other jurisdictions' antitrust laws as it enables "surveillance pricing" in which companies use AI to advertise products at person-specific levels based on their browser activity and their emotional state at the time preceding the sale. This practice, commonly known in competition law parlance as 'personalised pricing' or 'behavioural price discrimination', occurs when AI analyses user-specific behavioural or emotional data to dynamically tailor individual pricing thus extracting the highest possible economic surplus from consumers (be they end-consumers or business clients), and it can be interpreted as a price abuse under circumstances of significant market power.
Additionally, under Article 6(1)(k) of the DMA, which explicitly addresses transparency obligations for gatekeepers when applying personalised pricing, designated 'gatekeepers' are required to clearly inform consumers if and how personalised pricing is being used. Article 5 also broadly mandates 'fairness', 'transparency', and 'non-discrimination', thereby indirectly addressing practices that exploit market power to unfairly discriminate between users.
Accordingly, platforms designated as gatekeepers under the DMA must clearly disclose personalised pricing practices, which directly constrains hidden or manipulative uses of behavioural or emotional profiling for differential pricing, and failure to comply can result in heavy fines and regulatory intervention, thus materially restricting the latitude previously available under the broader competition rules of Article 102 TFEU.
Lastly, while the DSA primarily targets content moderation, transparency, user protection, and targeted advertising on digital services, it also has some potential indirect relevance in that Articles 24 and 26 place stringent transparency requirements on recommender systems and personalised advertising that utilise behavioural or emotional data, which while not targeting pricing directly, indirectly constrain exploitative practices, including emotional-based price discrimination, if they rely on misleading or opaque consumer interactions. Article 25 also introduces obligations for Very Large Online Platforms (VLOPs) to disclose detailed parameters used in targeting and recommending content, making covert emotional profiling for pricing significantly more legally challenging.

[47] Whereas visual cues can be captured through computer vision using a simple camera, capturing heart rate, galvanic skin response and brain signals typically requires wearables or more invasive (or at least non-trivial) equipment like EEG, functional near-infrared spectroscopy, or fMRI.

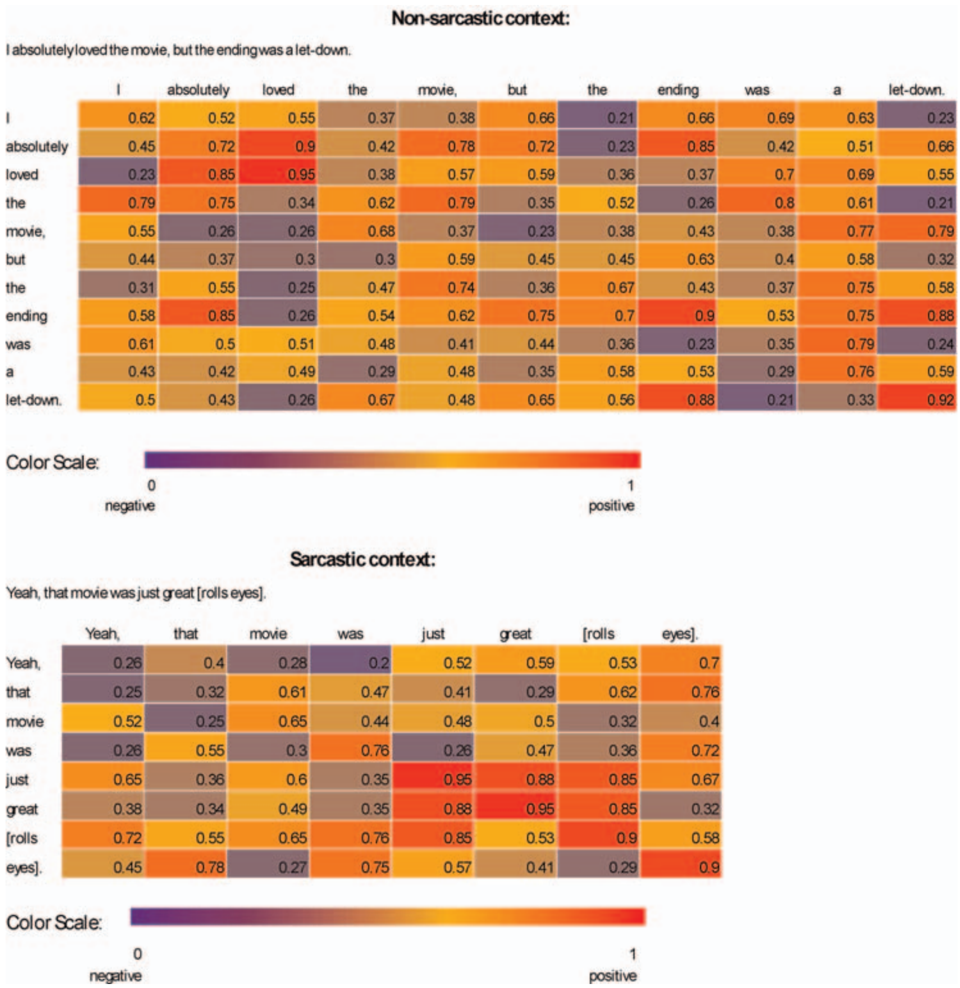


Figure 1.
An example of Self-Attention Heatmaps (also referred to as "Transformer Lens").

2.5 Conflation strategies

A variety of conflation strategies have been proposed to achieve effective multimodal integration, with early conflation approaches combining raw data or feature representations from different modalities at an initial processing stage. This strategy can facilitate the joint learning of cross-modal correlations but was found to be susceptible to what has become known in the field of computing as 'the curse of dimensionality'[16] and heterogeneous noise characteristics across modalities. In contrast, late fusion methods integrate the outputs or decisions of unimodal classifiers,

---

[16] The curse of dimensionality refers to the exponential increase in data sparsity and computational complexity as the number of dimensions in a dataset grows, often leading to challenges in model training, overfitting, and decreased performance in high-dimensional spaces. Neural methods often mitigate – but do not eliminate – some aspects of high dimensional spaces by learning lower-dimensional embeddings, but the potential for exploding parameter counts remains.

which allows each modality to be processed independently before combining the inferred sentiments together. Although late fusion is computationally simpler and often more robust to modality-specific noise, it may fail to capture intricate interdependencies that exist at the feature level.

Recent advances have popularised intermediate fusion techniques, which seek to integrate modalities at an intermediate layer within a deep neural network, thus benefiting from learning a joint representation space in which intermodal relationships are more effectively aligned. Multimodal transformers, for instance, have been proposed to extend the self-attention mechanism to simultaneously attend to both textual tokens and non-textual features, and empirical studies [17, 18] have demonstrated that such joint representations can significantly improve the accuracy of sentiment detection, particularly in settings where subtle emotional cues are distributed across modalities.

2.6 Representation learning and joint embedding spaces

Building joint embedding spaces that encapsulate the semantic and affective content of each modality is the undisputed underpinning of many multimodal integration techniques that project features from textual, auditory, visual, and physiological data into a common latent space, so as to be able to assess cross-modal interactions more directly. Techniques such as Canonical Correlation Analysis (CCA)[17] and its deep learning extensions have been employed to maximise the correlation between different modalities. More recent methods utilise adversarial learning to align embeddings from disparate modalities, ensuring that the joint space is invariant to modality-specific artefacts while retaining critical emotional signals. Empirical evidence suggests that models leveraging joint embedding spaces are particularly effective in contexts where one modality may be noisy or partially missing, as the shared latent representation provides a form of robustness not found in other models [19, 20].

It is important to note, in this regard, that traditional CCA assumes linear relationships between modalities, and that as a corollary, nonlinear generalisations such as Kernel CCA and more recent deep-learning variants like Deep CCA have been developed, enabling richer, nonlinear multimodal embeddings that capture more complex cross-modal interactions essential for nuanced sentiment analysis.

2.7 Attention mechanisms for modality-specific relevance

Another promising methodology involves the use of attention mechanisms tailored to multimodal data. In these architectures, attention weights are computed not only within each modality but also across modalities, allowing the model to dynamically assign greater significance to the cues that are most informative for sentiment inference. For instance, in a scenario where textual cues are ambiguous, the model may allocate more weight to visual expressions or variations in vocal intonation. Empirical

[17] CCA is a multivariate statistical technique that is particularly useful in multimodal data conflation, where it reveals underlying relationships between disparate data sources by capturing their shared variance and which identifies and quantifies the associations between two sets of variables by computing linear combinations – known as canonical variates – that are maximally correlated with one another. Linearity here refers to linearity-in-variable rather than to linearity-in-parameter, which essentially means that the canonical variates are constructed as linear combinations of the original variables from each dataset.

research has underscored the utility of such cross-modal attention mechanisms in enhancing sentiment recognition performance [21]. This approach enables the model to selectively focus on modality-specific features while simultaneously considering the broader context provided by complementary data sources.

2.8 Empirical findings and applications

Several empirical studies have studied the efficacy of multimodal integration in sentiment analysis. Indeed, studies integrating textual sentiment with vocal prosody and facial micro-expressions have reported significant improvements in the detection of subtle emotional shifts in social media and clinical settings [22, 23]. Moreover, the incorporation of physiological signals has been shown to enhance the detection of stress and affective arousal, particularly in real-time applications such as mental health monitoring [24]. These findings underscore the potential of multimodal approaches to capture a more nuanced spectrum of human emotion than is possible with text alone and make underlying model optimisations and improvements a very promising area of practical work and research.

In toto, these findings underscore the power of multimodal techniques to enrich our understanding of affect beyond what text-based systems can achieve alone. However, capturing more subtle or subconscious emotional signals – particularly those that might not even be articulated by the individual – remains a major challenge, not least because it requires equipment and/or wearables that are not always widely available. As research shifts toward a deeper characterisation of emotional and psychological states, as well as to how well a model built on specific statistically unrepresentative populations generalises to broader populations, the field has begun focusing on systems that attempt to emulate real empathic understanding. This outlook has given rise to the concept of "empathy algorithms," which aim to integrate multiple modalities and contextual factors to infer not just what a user explicitly expresses, but also the nuanced feelings they may be unable or unwilling to convey outright. The following section explores how these emerging systems expand traditional sentiment analysis into a more holistic, human-like interpretation of affect.

2.9 The emergence of "Empathy algorithms"

As neural network architectures became more sophisticated, researchers began to speculate about capturing underlying affect and "subconscious" emotional signals exclusively from text. However, purely text-based models soon demonstrated limitations in capturing the fuller emotional range, leading to an increased focus on multimodal analysis. Combining textual insights with behavioural or physiological cues can uncover a deeper "empathic" perspective. The term "empathy algorithms" symbolises this shift: AI is not merely labelling an emotion but attempting to infer how the user might actually feel, potentially even uncovering emotional states the user might be unaware of him/herself at the conscious level.

"Empathy algorithms", illustrated in **Figure 2**, embody a new category of sentiment analysis systems that aim to surpass the limitations of basic emotion detection by approximating a richer, more human-like understanding of affective states. Unlike traditional approaches, these algorithms attempt to estimate both conscious and subconscious emotional states through a holistic analysis of language, context, as well as verbal and non-verbal cues. Their effectiveness relies on three key capabilities, that all such systems have as a common denominator, namely:

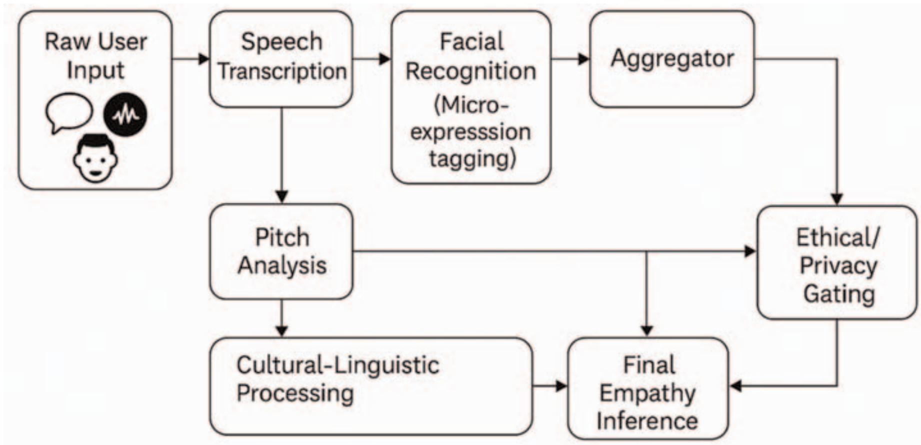


Figure 2.
Conceptual architecture of a multimodal empathy algorithm.

1. *Deep contextual understanding*: Leveraging large-scale pre-trained models and attention mechanisms, empathy algorithms can process long-range dependencies in text and detect subtle shifts in tone or subject, such as transitioning from optimism to sarcasm within a social media post. This allows them to interpret context-specific emotional nuances with unprecedented accuracy.

2. *Dynamic adaptation*: Incorporating zero-shot and few-shot learning techniques enables these systems to adapt to new emotional domains with minimal annotated data. For instance, a model trained on English-language sentiment data can infer sentiment in emerging contexts or other languages with limited retraining.

3. *Multimodality*: By integrating textual analysis with verbal and non-verbal signals such as facial expressions, vocal intonation, normalised pitch, and deviation from the Root-Mean-Square of the voice sound wave, as well as physiological cues, empathy algorithms provide a more comprehensive view of human emotion. For instance, algorithms might synchronise text sentiment analysis with real-time stress levels inferred from vocal pitch and facial micro-expressions captured on video, combining these insights into a unified emotional profile. This multimodal integration enables a richer and more accurate representation of affective states.

These “empathy algorithms” adopt advanced representation learning, multi-task objectives (e.g., predicting sentiment plus speaker traits), and robust domain adaptation methods (e.g., zero-shot learning). They often employ large pre-trained language models, leveraging billions of tokens from diverse text *corpora*, and then enhance these models with additional signals from eye tracking, facial expressions, or acoustic waveforms. In this sense, it might be in order to distinguish between “*operational empathy*” which deals with detecting and responding appropriately to inferred emotional states, and “*affective empathy*” that requires feeling what others feel. Although these empathy algorithms are still far from genuine human affective empathy, they represent a significant leap toward capturing the complex underpinnings of the wide multi-dimensional spectrum of human emotion and are thus already capable of operational empathy.

Subsequently in this chapter, wherever I refer to empathy algorithms, the reference will implicitly encompass empathy that is of the *operational* type as defined above. Together, these capabilities empower empathy algorithms to address the shortcomings of earlier systems by providing nuanced emotional insights. For instance, empathy algorithms have been used to monitor mental health through subtle sentiment shifts in patient communication, identify implicit emotional cues in professional communication, and highlight subtle shifts in affective states over time. These applications bridge the gap between artificial intelligence and human emotional intelligence while showcasing the transformative potential of these systems. As these systems continue to evolve, further advancements in multimodal data integration and contextual understanding promise to push the boundaries of affective computing, enabling applications that were once considered the domain of science fiction. From mental health monitoring and crisis intervention to personalised customer experiences and educational tools, empathy algorithms are reshaping the possibilities of affective computing across diverse fields, and while their risks, in the wrong hands are glaringly obvious, so is their vast potential in the right ones.

2.10 Ethical and legal considerations

Addressing ethical challenges is integral to the development of sentiment analysis and must be considered in tandem with technical and methodological advancements. The integration of multimodal data raises complex issues around privacy, consent, and data security, with regional specifics playing a decisive role. Ethical challenges, such as privacy in multimodal data and fairness in algorithmic design, require consideration and exploration alongside opportunities like neuroscience integration and applications in multilingual contexts with the aim of guiding future research in creating AI systems capable of truly "hearing" the faint whispers of the human heart while respecting cultural diversity and individual autonomy.

Real-world deployments of sentiment-driven algorithms are already impacting society. For instance, YouTube's recommendation AI has been found to amplify negative emotions – an audit study showed that if a user (or bot) exhibits an interest in angry or grievance-filled videos, the algorithm disproportionately feeds them more such emotionally charged content, reinforcing an "anger bubble" [25]. Likewise, TikTok's famed "For You" feed can rapidly spiral users into specific emotional atmospheres: The Wall Street Journal demonstrated in 2021 that a fresh account which paused on depressive content began receiving a flood of similar sad videos within minutes [26]. These examples highlight both the influence and the potential risks of sentiment-sensitive algorithms that can create echo chambers of particular feelings, thereby affecting user well-being and even societal discourse, while underscoring the need for careful design and oversight of "empathy algorithms".

The European Union's General Data Protection Regulation (GDPR), as well as the newer Digital Services, Digital Markets Act, and the AI Regulation, as well as China's Personal Information Protection Law (PIPL) [5], significantly impact multimodal data collection, necessitating compliance with stringent privacy regulations. Moreover, Senft et al. [6] highlight cultural biases in emotion recognition models applied to East Asian populations, reinforcing the importance of designing systems that account for cultural and linguistic diversity referred to hitherto.

By embedding these ethical considerations into the technical narrative, developers and systems architects can ensure a more comprehensive approach to advancing sentiment analysis that does not fall foul of the law within the jurisdiction(s) in which such systems are intended to be deployed. These include concerns around privacy, consent, and data security, as well as potential misuse of sensitive data in applications like surveillance or manipulation. The integration of multimodal data, including facial expressions, eye movements, and physiological signals, make for pressing concerns around privacy, consent, and data security, as well as the generalisability of underlying models and the extent to which a model trained in, say the USA, can be used to target citizens in Europe or China. The potential misuse of sensitive data – particularly when combined with powerful AI for surveillance or manipulation – has been the mainstay of regulators' calls for robust regulatory frameworks and ethical guidelines. It is also worth noting that the interpretative nature of these systems introduces risks of bias and misrepresentation, especially when applied across culturally- and linguistically- diverse populations. These considerations, though beyond the direct scope of this chapter, remain critical to ensuring that empathy algorithms are developed and deployed in a manner that respects individual autonomy and societal values. These ethical challenges, which are beyond the scope of this chapter, must be considered alongside the technical advancements explored in this book to ensure that AI systems progress safely and responsibly toward unprecedented machine emotional intelligence.

2.11 Challenges and future directions

Despite these advances, several challenges remain to be dealt with, and these include the heterogeneity of data sources that poses significant issues in terms of synchronisation, noise handling, and the design of unified architectures capable of processing diverse data types, with the latter being one of the holy grails not only for sentiment analysis but also for the fast-accelerating AI scientific research programme more broadly.[18] Furthermore, the computational complexity of multimodal models, especially those employing attention across multiple channels, necessitates further research into optimisation techniques, semiconductors and scalable architectures, even though advances in these fields and in computing power available are happening at break-neck speed, and are thus months rather than years or decades away. Future work will likely also focus on refining joint embedding methods, enhancing cross-modal attention, and exploring novel sensor modalities to further improve the robustness and interpretability of multimodal sentiment analysis systems.

Today, multimodal integration techniques represent a critical evolution in sentiment analysis, offering a pathway to better, more comprehensive and context-sensitive models of human affect by effectively combining textual data with auditory, visual, and physiological signals. These techniques and methodologies enable the provision of richer and more accurate insights into the multifaceted nature of emotional expression, than would otherwise be possible in their absence.

Looking ahead, wearable technologies, miniaturised sensors and semiconductors, advancements in Internet of Things (IoT) technologies, as well as declining hardware costs and increased accessibility to hardware that was only available, in the past, to

[18] Joint Embedding Predictive Architectures (JEPAs), which Focus on learning robust, invariant latent representations through a self-supervised predictive task and Liquid Neural Networks (LNNs), which concentrate on modelling dynamic temporal processes and adapting to continuous changes in input data are at the time of writing of this chapter the most promising avenues, but this does not mean that this will be the case for long as novel and improved architectures proliferate at an ever-increasing speed.

well-funded research institutions and hospitals, are expected to become the enablers of next-generation "empathy algorithms". Imagining a scenario where users wear lightweight eyeglasses equipped with discreet cameras and neural chipsets capable of processing *their own* micro-expressions, voice intonation, and biometric signals in near-real time, in the near-future is no longer in the realm of science fiction, but one that is much closer to becoming reality thanks to Meta's recent launch of Orion and Aria Gen 2 glasses and to the continuous advancements being made by Meta, MediaTek and Qualcomm who produced the underlying technologies for Orion and Aria. As these glasses scan the environment, they would be able to seamlessly integrate audio cues, facial movements, and even subtle physiological changes (such as heart rate variability or galvanic skin response) to infer complex emotional states of the wearer and of other people the wearer is interacting with more accurately than text-based models alone, and thanks to Meta FAIR's recent breakthroughs in generalising the decoding of brain wave signals with a more-than-decent accuracy of 80%,[19] we are also one step closer to non-invasive brain wave decoding if the technology can be miniaturised sufficiently to fit onto a pair of glasses, cap or headset.

At the same time and equally importantly, advances in multilingual modelling will allow these sentiment systems to adapt instantly to code-switched conversations or lesser-resourced languages, broadening their impact beyond Anglophone domains. In a global workplace setting, such a system could facilitate intercultural virtual meetings by translating not just speech but also meaning as well as underlying emotional undercurrents, flagging moments of rising tension or confusion so that interventions can be made proactively and respectfully. The other side of the coin is also true, as these technologies could instead make such meetings more difficult if the technology is asymmetrically deployed giving a dominant advantage to one party in the meeting over another. Over time, these real-time, context-aware "empathy algorithms" could be applied in the milieu of some very diverse applications, from personalised learning tools that respond to student frustration cues, to remote healthcare solutions offering customised real-time emotional support. By synthesising wearable-device data, cross-cultural analytics, and deep neural representations, sentiment analysis is soon very likely to evolve from a purely descriptive framework into an interactive, empathetic assistant that aids both individuals and societies at scale. In what comes next, we shall delve deeper into the technical side of the cutting-edge techniques that have made these developments possible.

3. State-of-the-art NLP sentiment analysis methods

3.1 Deep learning architectures

Sentiment analysis has been profoundly transformed by Deep Learning architectures that have enabled models to capture intricate emotional patterns in text, from handling negation and sarcasm to understanding complex relationships between words and phrases. These methods have addressed challenges that were previously

[19] *Vide* https://ai.meta.com/research/publications/brain-to-text-decoding-a-non-invasive-approach-via-typing/, https://ai.meta.com/research/publications/from-thought-to-action-how-a-hierarchy-of-neural-dynamics-supports-language-production/ and https://ai.meta.com/blog/brain-ai-research-human-communication/ (all accessed on 21 Feb 2025).

insurmountable, while emerging techniques in NLP, such as zero-shot and few-shot learning, further enhance these capabilities by enabling models to adapt to new tasks with minimal annotated data while expanding their applicability across different domains.

3.1.1 Recurrent neural networks (RNNs)

Early RNN architectures, such as Elman Networks (simple recurrent networks with context units feeding back from hidden layers) and Jordan Networks (with recurrent feedback from the output layer), laid the foundational principles for sequential data modelling, and despite their simplicity, have gone to great lengths at highlighting the necessity for specialised architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) to effectively capture longer-range temporal dependencies without suffering from vanishing gradients. LSTM and GRU networks have been extensively tested for capturing the evolution of sentiment over the sequence of words or phrases. LSTMs, for instance, can effectively handle sentences with negation, such as "I don't dislike the product", where the sentiment changes due to the position of "do not" in relation to "dislike" [27, 28]. Similarly, GRUs can capture shifts in tone, such as when a sentence transitions from neutral to positive sentiment, like "The start was slow, but the ending was fantastic". They excel in tasks where the position of a word in a sentence significantly modifies the overall meaning, such as negation, and although they were overshadowed in recent years by transformer-based models, RNNs remain relevant where data or computational constraints make transformer-based fine-tuning less feasible. In fact, RNNs are still commonly deployed in low-resource environments, such as embedded systems or IoT devices, where computational efficiency is critical, even though this survival is likely to be short-lived given continuing advancements in semiconductors that are making edge-device microchips increasingly computationally capable and energy efficient.

3.1.2 Convolutional neural networks (CNNs)

In sentiment analysis, 1-Dimensional CNNs have proven beneficial for identifying local n-gram-like features indicative of sentiment. CNNs typically require fewer parameters than RNNs or transformers, enabling more efficient training and giving them an advantage in situations where computational resources are limited or when working with smaller datasets, especially where overfitting might be a concern. By way of an example, CNNs are particularly effective in detecting sentiment in tweets or product reviews, where local patterns often dominate the text, but their tendency to capture local features is both a strength (for domain-specific phrasing) and a limitation (since context far from a relevant phrase may be missed).

3.1.3 Transformer-based models

The introduction of attention mechanisms – particularly in the form of the Transformer architecture – has managed to unleash significant progress. Bidirectional Encoder Representations from Transformers (BERT), RoBERTa, XLNet, and more recent variants can be fine-tuned to excel in sentiment tasks. As already discussed, BERT introduced bidirectional attention to understand context from both preceding and succeeding words, while RoBERTa refined this with optimised training strategies for improved accuracy. XLNet, on the other hand, leveraged permutation-based training to model sentence structure more effectively, making it particularly adept at handling nuanced or complex sentiment relationships. The capacity of attention mechanisms to handle long-range dependencies and provide contextual embeddings for each token yields more accurate and nuanced sentiment predictions, while recent advancements, such as lightweight transformers like DistilBERT and multilingual pre-training, are also making transformer-based architectures more accessible and effective for low-resource, mobile, edge and fog applications. These innovations enable real-time applications, such as mobile customer feedback analysis and voice assistant sentiment recognition. Building on this foundation, transfer learning and domain adaptation extend the capabilities of these models by allowing them to adapt to new domains and tasks with minimal additional data or computational resources.

When it comes to transfer learning, the common practice is to use models already pre-trained on massive corpora (e.g., English Wikipedia, large news datasets) and then fine-tune on the target sentiment dataset. This drastically reduces the labelled data needed to achieve strong performance and has also democratised access to advanced NLP, enabling smaller organisations to leverage these models without requiring extensive computational resources. It has also given rise to risks that are only now coming to the fore. They include misinformation and disinformation material specifically targeted to AI foundational models by leveraging these datasets to bias results.

When it comes to domain adaptation, on the other hand, domain-specific variants, like BERTweet, RoBERTweet, TweetEval BERT, Legal-BERT, PatentBERT, BioBERT, SciBERT, BlueBERT and ClinicalBERT, provide additional performance enhancements in specialised tasks by continuing pre-training on domain-relevant corpora [29–37].

3.1.4 Hybrid architectures

Combining CNNs, RNNs, and transformers capitalises on each model's strengths. For instance, hybrid architectures have demonstrated measurable improvements in tasks like sarcasm detection within customer feedback. Traditional models often fail to identify layered sentiment, but combining CNNs for phrase-level detection, RNNs for sequential dependencies, and transformers for refined attention across paragraphs has achieved significant gains. These combinations have also been applied in industries such as healthcare for analysing patient emotions in clinical notes and e-commerce for assessing customer feedback [27].

3.2 Advanced representation learning

3.2.1 Contextualised word Embeddings

Whereas earlier embeddings like Word2Vec and GloVe generated a single static vector per word, contextualised embeddings like ELMo and BERT-based embeddings, produce representations that shift based on surrounding context. This improves handling of polysemous words (e.g. "bank" as a financial institution vs. "bank" of a river) and domain-specific language usage significantly.

3.2.2 Knowledge-infused Embeddings

Technical studies shed light on how incorporating structured knowledge from resources such as WordNet, DBpedia, or domain-specific ontologies can refine sentiment models, especially for tasks like sarcasm detection or domain adaptation. By linking words or named entities to external knowledge graphs, an algorithm can glean external context in the form of historical sentiment usage, real-world relationships, or concept hierarchies that a purely text-based approach may miss.

3.3 Few-shot and zero-shot learning

Despite the web's massive data availability, high-quality labelled sentiment data for specific domains or emotions remain scarce and pose strict limits on what can be done using such datasets. This is where few-shot and zero-shot learning approaches that can drastically reduce the need for manual annotation come in. These usually take the form of either Meta-Learning, where models are trained on multiple tasks to learn a "learning-to-learn" capability, enabling quick adaptation to new tasks or categories from a handful of examples and Prompt Engineering, usually deployed in large-scale pre-trained language model settings where meticulously-crafted "prompts" can guide the model to classify or interpret new sentiment categories without explicit re-training. A practical example of this would be a sentiment analysis prompt phrased in natural language: "Classify the sentiment in the following text," along with a few exemplars.

Such methods open new opportunities for dynamic or on-demand sentiment analysis, where an analyst might need to label emergent emotional states (e.g. "uncertainty," "cautious optimism") in a newly-evolving environment, and have become easier to use with the rapid proliferation of easily- and cheaply- available high-quality Large Language Models.

3.4 Explainable AI (XAI) in sentiment analysis

As sentiment models are increasingly applied in domains with very high stakes, like healthcare and telemedicine, where diagnostic hints or triaging decisions may rely in part on inferred patient mood and emotional state; workplace evaluations, where sentiment models assess employee feedback, performance, and team social dynamics and where misinterpretations of tone or context could lead to unfair reviews, biased promotions, or workplace surveillance concerns; recruitment and workplace hiring, where misclassifications can reinforce biases or unjustly affect career trajectories; criminal justice and law enforcement, where sentiment-driven risk assessments or automated profiling could impact bail, sentencing, or investigative practices; finance and insurance, where creditworthiness assessments or fraud detection might incorporate sentiment-based risk indicators; Education and Student Evaluation, where automated tutoring systems or performance analytics risk misjudging learner engagement and emotional well-being; social media moderation and misinformation control, where large-scale sentiment analyses can influence which content is flagged, served or suppressed, with very serious societal ramifications and implications; politics, where sentiment analysis influences campaign strategies, voter targeting, and policy decisions and where misclassifications or biases can distort electoral outcomes, amplify political polarisation, or undermine trust in governance; and mental health screening, where AI-driven sentiment detection in chat logs, voice data, or social media aids early intervention and where classification errors could lead to missed warning signs or unnecessary escalations, affecting resource allocation and patient outcomes.

In all of the above, which are by no means meant as an exhaustive list of high-stake domains, explainability has emerged as one of the top priorities [38].

Two main categories of approach that are being adopted in this space and which are still far from perfect, are:

1. *Inherently interpretable architectures*: Models with built-in attention mechanisms that can 'visualise' which tokens or phrases most influence a classification decision, and with some sentiment systems producing "heat maps" of the text, underscoring fundamental emotional triggers like negative adjectives and intensifiers; (other approaches like concept bottleneck models aim for interpretability by design) [39, 40]; and

2. *Post-hoc explanation methods*: Techniques such as LIME[20] or SHAP[21] that approximate the decision boundary around a local instance, providing a set of features or words that most contributed to the model's output, and which while not always fully accurate, encourage user trust and facilitate model debugging, particularly in cases where biased or spurious correlations are likely to emerge [41, 42].

Transformer models inherently provide interpretability through attention weight visualisations, enabling analysts to directly observe which tokens most strongly influence sentiment decisions. Such visualisations complement *post-hoc* methods like LIME and SHAP by offering intrinsic, model-specific explanations.

4. Beyond text: The multimodal emotion recognition revolution

4.1 The rationale for multimodal analysis

Human emotional expression is rarely confined to text. Vocal intonation, pauses, pitch, facial micro-expressions, and gestures interweave to convey an affective message. Studies in psychology and cognitive science have consistently underscored the importance of non-verbal signals in emotional communication. By integrating these signals, AI systems stand a better chance of identifying complex or latent emotions – for instance, a user who claims to feel "fine" but exhibits stress in voice tremors and fleeting facial tightening. In real-world tasks such as mental health evaluation, user experience research, or advanced marketing analytics, multimodal sentiment analysis has steadily gained traction. An illustrative schema of multimodal emotion recognition analysis in action is provided in **Figure 3** on page 23.

4.2 Attention-tracking for deeper emotional insight

Eye-tracking has evolved from a niche tool in psycholinguistics to a widely used technology across consumer research, gaming, website design, and several other areas. Systems like Tobii which were popularised through their integration with Alienware gaming systems at around 2016, provide robust measurements of features like:

[20] Local Interpretable Model-Agnostic Explanations (LIME) creates simplified surrogate models around individual predictions, illustrating which features contribute most in that local neighbourhood.

[58] SHapley Additive exPlanations (SHAP) draws on Shapley values from cooperative game theory to quantify each feature's overall contribution to the final prediction in a more globally consistent manner.

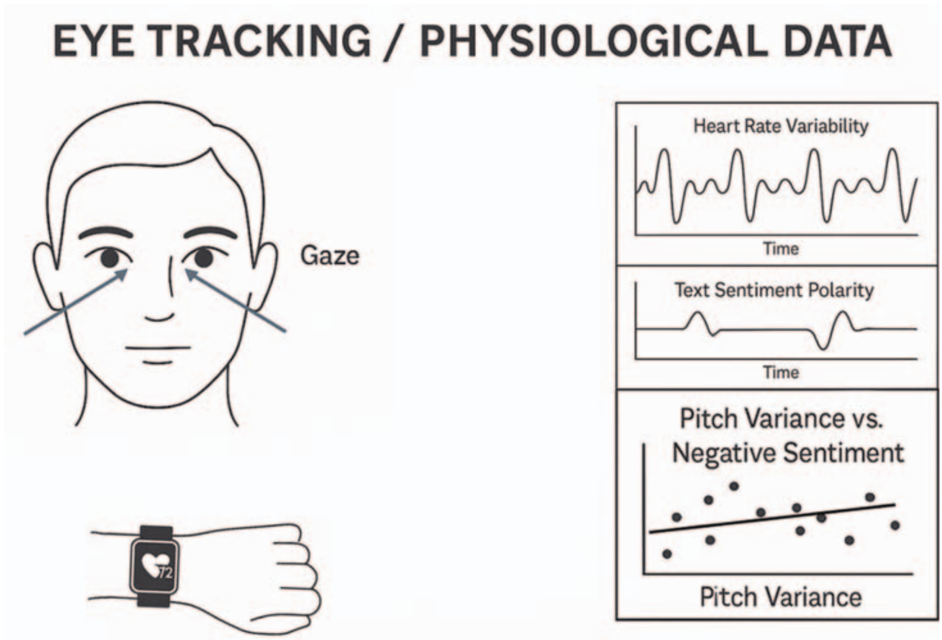


Figure 3.
Multimodal emotion recognition analysis in action.

- *Fixation duration*: Longer fixations may indicate deeper cognitive processing or heightened emotional reaction (e.g. shock, confusion, or fascination);
- *Saccadic patterns*: The rapid movements between fixations can reflect scanning behaviour, avoidance, or exploratory curiosity;
- *Pupil dilation*: Although more complex to track reliably, changes in pupil size can be associated with autonomic arousal, which in turn correlates with emotional intensity.

Other popular platforms, such as Hotjar, offer similar capabilities for web interfaces by tracking mouse movements, scroll depth, and click patterns to infer user engagement and cognitive load. While these lack the detailed physiological insights provided by Tobii's dedicated hardware or integrated camera-based computer vision eye-tracking, they still illustrate the growing accessibility of gaze-based and behavioural data for both commercial and research applications, with the gap being on a trajectory of closing fast with recent advancements in single-shot deflectometry [43, 44]. Recent advances in deep learning have also enabled increasingly reliable eye-tracking even through standard consumer-grade webcams with as low a resolution as 480p, significantly democratising access to multimodal sentiment analysis by coming very close to – and thus eliminating the need for – specialised and costly peripheral hardware. A key benefit of eye-tracking for sentiment analysis is its relative objectivity compared to self-reported measures: it can capture "subconscious" interest or repulsion that might not emerge in textual, spoken or mouse hovering/mobile zoning responses, which combined with text-based analysis, like correlating which words or phrases a user lingers on with subsequent typed or spoken responses, can provide researchers with more subtle insights into real human emotional states.

4.3 Facial expressions and body language

Facial expression recognition, especially that powered by very deep convolutional networks, is capable of identifying not just broad categories like happy, sad, or angry but also micro-expressions that last only fractions of a second that often betray genuine emotions that a person may consciously try to hide or moderate, and which s/he might have received training on to hide. In a similar fashion, body language analysis offers complementary cues like posture, hand gestures and head tilts that oftentimes clarify or at least disambiguate some otherwise ambiguous facial or textual signals.

Multimodal systems often face the technical challenge of aligning data streams in time, like, for example, in ensuring that a user's textual statement at time t is synchronised with the facial expression or gaze recorded at the very same moment, with contemporary approaches leveraging data-fusion techniques such as attention-based fusion layers or gating mechanisms that ascribe a weight to each modality's contribution depending on context. By way of an illustrative example, in an e-commerce setting, a user's negative statement about a product might be confirmed by an eye gaze fixation on a disliked product feature or by a subtle facial grimace upon seeing its image.

4.4 Voice waveform and speech analysis

Voice-based sentiment analysis integrates features like pitch (F0), intensity, and spectral characteristics with standard Automatic Speech Recognition (ASR) outputs and makes use of the fact that the fundamental frequency of a person's voice can shift with excitement, anger, or fear, and formants (resonances of the vocal tract) can change with stress. By analysing the generated wave form and after denoising it, modern techniques allow direct ingestion of raw waveforms into convolutional or recurrent layers, circumventing the need for manually-engineered audio features [45].

In this respect, it is worth noting that a key advantage of voice-based signals is their potential for real-time monitoring. For instance, advanced call-centre analytics tools combine sentiment-laden transcripts with pitch tracking to suggest interventions for stressed customers or to assess overall agent performance, whereas in telehealth or counselling contexts, voice features can alert practitioners to emotional states that might not be explicit in textual transcripts, especially for patients who struggle to articulate their feelings.

5. Handling complexity in emotional content

5.1 Emotion intensity and classification

Human emotions are not merely binary or ternary categories, as they exhibit considerable granularity in both type (e.g. fear, joy, disgust) and intensity (e.g. mild annoyance vs. anger vs. blinding rage). Datasets such as GoEmotions provide annotation for more than two dozen discrete emotional labels, each with varied intensity levels, while datasets like AffectNet and EmoBank incorporate both categorical and continuous emotion annotations, and CrowdFlower and SEMEVAL 2007 Task 14 offer emotion-labelled text data for diverse applications. Some models rely on regression-based methods to predict continuous intensity scores, while others employ multi-label classification to account for overlapping emotional states. One increasingly-popular practical application is in mental health research, where the intensity of sadness or hopelessness harvested from online forum posts may help identify individuals at risk. Likewise, brand marketing professionals might measure the intensity of excitement or satisfaction in product reviews to guide product strategy decisions.

It is worth noting that these tasks demand models capable of not only labelling an emotion but also of estimating how strongly it is expressed, or how likely it is to co-occur with other emotional states. The improvements made in this space over the past decade have been nothing short of astounding.

5.2 Sarcasm, irony, and figurative language

Sarcasm and irony remain some of the toughest challenges for sentiment analysis, often because they rely on pragmatics, context, or shared cultural references that are difficult to model in a generalisable way. Scholars have proposed a range of solutions, with each having its pros and cons, namely:

- *Context-aware embeddings*: Sarcasm can often be identified by comparing the literal sentiment of a phrase to the overall conversational or situational context. Consider, by way of a practical example, a user exclaiming "Wonderful!" in response to an obvious mishap.

- *User modelling*: Some approaches incorporate user history where past patterns govern interpretation and where if a user frequently posts sarcastic remarks, the algorithm can adapt its interpretation accordingly.

- *Metadata and pragmatic cues*: Emojis, punctuation (e.g., "!!!" or "?!?!?"), or capitalisation can hint at sarcasm or specific emphasis. In spoken contexts, prosodic cues or a distinctive intonation can be informative.

Nevertheless, despite advances in neural techniques and context-aware embeddings, detection accuracy for sarcasm and irony remains significantly lower than for straightforward sentiment classification tasks, as even with advanced neural methods, detection accuracy for sarcasm and irony often lags behind that for straightforward sentiment, underlining the need for continued innovation and better annotated corpora to deal with this linguistic complexity.

One promising approach is to integrate commonsense knowledge via knowledge graphs. For the sake of an illustrative example, consider the sarcastic remark "Oh, Microsoft has been so reliable lately". A sentiment model augmented with a knowledge graph can recognise Microsoft as a software company frequently associated with bugs or crashes and this contextual linkage reveals the dissonance between 'reliable' (positive literal sentiment) and the factual reality, cueing the model that the praise is likely insincere and sarcastic. By grounding language in real-world facts (like company reputations, well-known events and associated features), such neuro-symbolic systems improve sarcasm detection that purely text-based models would misclassify.

5.3 Subjectivity and pragmatics

Subjectivity analysis separates factual from subjective (and thus opinion-based) statements. This is vital for tasks such as news summarisation or scientific literature review, where conflating the authors' opinions with objective facts can mislead readers, irrespectively of whether this is intentional or consequential. In the realm of social media or product reviews, however, harnessing subjective statements is precisely the defined objective, which is exactly why a targeted pipeline might first filter out factual segments to focus purely on subjective content, thereby reducing noise in downstream sentiment classification. Contemporary research [46–52] suggests that advanced Transformers fine-tuned on domain-specific subjectivity corpora can reliably split subjective statements from objective descriptions, resulting in more precise sentiment detection on subjective segments.

Another important development in this area is uncertainty quantification in sentiment predictions, with modern sentiment classifiers capable of being equipped with Bayesian neural network layers or ensemble methods to estimate their confidence in a given prediction [53]. In practice, this means an empathy algorithm could say not just "this post is negative", but could also add "with 95% statistical confidence". Such measures of certainty allow the system to flag low-confidence judgements for human review, refrain from acting on ambiguous cases or try to solicit additional information until the confidence level reaches a target level (say 68%, 95% or 99%), thereby improving reliability and safety in sensitive deployments [54]. Ongoing research surveys methods for capturing model uncertainty in deep networks, aiming to integrate these techniques into sentiment analysis pipelines [55, 57].

6. Challenges and biases in empathy algorithms

6.1 Cross-lingual and multilingual sentiment analysis

With global social media usage rising in both penetration rate and also average user time spent on social media, robust sentiment analysis must accommodate a plethora of languages. This gives rise to an issue where many languages are "low-resource", lacking large, labelled *corpora*. Multilingual BERT[22] or Cross-lingual Language Model – XLM (particularly XLM-R)[23] show promise by learning shared latent representations that transfer across languages. One approach is cross-lingual transfer learning, whereby a model trained on English data is further tuned on a small corpus of the target language, but given its inherent and obvious limitations, effective, cultural idiosyncrasies, unique idioms, and morphological complexities in certain languages (e.g. agglutinative forms) remain a formidable obstacle to perfect transfer that still requires further research and testing work to overcome. **Table 1** on page 23 shows some common practical examples of output misclassifications and why they happen.

[22] This is a variant of BERT (described earlier) trained on multiple languages using Wikipedia *corpora* that learns a shared representation across languages, enabling zero-shot cross-lingual transfer for tasks such as sentiment analysis, even for languages with limited labelled data.

[23] An extension of the XLM model, XLM-R (also referred to as XLM-RoBERTa) is a transformer-based language model trained on over 100 languages using a massive multilingual *corpus* that improves cross-lingual understanding by leveraging large-scale masked language modelling, outperforming multilingual BERT in many multilingual NLP tasks. XLM-R improves cross-lingual transfer by employing large-scale masked language modelling without explicit cross-language alignment objectives, differentiating it from earlier cross-lingual models like XLM.

Example	Misclassified output	Interpretability explanation
Sarcasm with exclamation: "I LOVE waiting in traffic. Just fantastic!"	Positive sentiment	Model overweights 'LOVE' and 'fantastic', missing sarcastic tone
Code-switched phrase: "The party was really chévere!"	Positive sentiment	Model does not recognise 'chévere' as regional slang (Spanish)
Dialect or slang: "Gonna swim down at the holler day."	Holiday	Misinterprets 'holler day' as 'holiday' due to phonetic similarity
Culturally specific pun: "Life is rice, do not waste it!"	Food	Model interprets 'rice' literally, missing pun on 'life is nice'

Table 1.
Output misclassifications.

6.2 Bias, fairness, and ethical considerations

Large Language Models often inherit biases present in their training data, which have survived various attempts at eradicating them. On that showing, certain demographics might be systematically associated with negative or stereotypical language, and in sentiment analysis, such biases can lead to disproportionate misclassifications, especially if the model was predominantly trained on data from a single culture or region. This arises because of the very well-known statistical phenomenon of sample selection bias and the biasedness of the estimators subsequently computed thereon which make them both 'biased' and 'inconsistent' in a technical statistical sense. Popular approaches to mitigate biases of this type thus far include:

- *Rebalancing training data*: Ensuring balanced representation of demographic groups of interest such that the addressable statistical population is better represented by the sample (which is arguably the simpler, more straightforward solution);

- *Adversarial training*: Introducing a secondary objective to discourage the model from encoding protected attributes;

- *Bias and fairness metrics*: Quantifying and minimising disparate impact or false positive/negative rates across user groups.

While discussing deep ethical implications is beyond this chapter's scope, it is crucial to note that empathic or emotion-detecting systems carry inherent risks if developed or deployed irresponsibly and without consideration to possible underlying biases, and this includes potential privacy violations and misuse for manipulative ends.

6.3 Data quality, labelling, and annotation costs

For "empathy algorithms" to be trustworthy, they need large, carefully-curated datasets reflecting real and genuine emotional expressions. This usually requires very laborious, time-consuming and costly manual labelling of sentiment or emotions, can be considerably subjective, and is often plagued by inconsistent inter-annotator agreement, especially for subtle categories like irony or confusion. In practice, crowdsourcing solutions might sometimes help, but more often than not introduce self-defeating inconsistencies, unless annotation guidelines are meticulously defined, validated and rigorously followed and enforced – a Herculean task in itself. This complexity rises exponentially in multimodal settings, where synchronised annotation of text, audio, and facial data can be expensive and time-consuming, and where strategies like semi-supervised learning, active learning (where uncertain samples are prioritised for human annotation), and synthetic data generation are ongoing areas of research (thus far with different degrees of success) aimed at reducing annotation bottlenecks and costs.

7. Practical applications and use cases

7.1 Social media monitoring and mental health

One of the first large-scale applications of sentiment analysis was real-time social media monitoring (most often Twitter, Facebook and Orkut) for brand management. “Empathy algorithms”, however, expand the scope to detecting emotional shifts in populations or identifying individuals who may be exhibiting signs of distress. Research focusing on Twitter data (prior to its rebranding to X) has shown that certain emotional markers – such as changes in language style or negative intensifiers – can correlate with higher self-reported depression or anxiety. Coupled with data from speech or face analytics, as in video-based telehealth sessions, for example, next-generation sentiment systems could provide early warnings to mental health professionals.

7.2 Customer experience and interface design

Eye-tracking data, combined with textual user feedback, allows anyone deploying this technology to pinpoint interface elements or processes that elicit confusion or frustration. A website’s sign-up page might yield negative textual feedback in surveys, but eye-tracking (and to a lesser extent web interface tracking – mouse movements, scroll depth, and click patterns) can reveal precisely where end-users lose orientation, lose interest or become stressed. Integrating these signals with heatmap visualisations clarifies how UI changes impact user sentiment, and similarly, voice-based sentiment analytics in customer service calls can detect when a client’s frustration is escalating, prompting timely intervention by a human operator or adaptive workflow.

7.3 Financial market analysis

Financial sentiment analysis was spurred by the shocking discovery, a couple of decades ago, that a significant percentage of daily stock movements was based on news items, and was thus initially centred on aggregated textual signals from news articles, reports, and social media posts. This was eventually also reported on by the US National Bureau of Economic Research [58].[24] The addition of voice-based cues

[64] Available from: https://www.nber.org/digest/jun13/which-news-moves-stock-prices (accessed 19 March 2025).

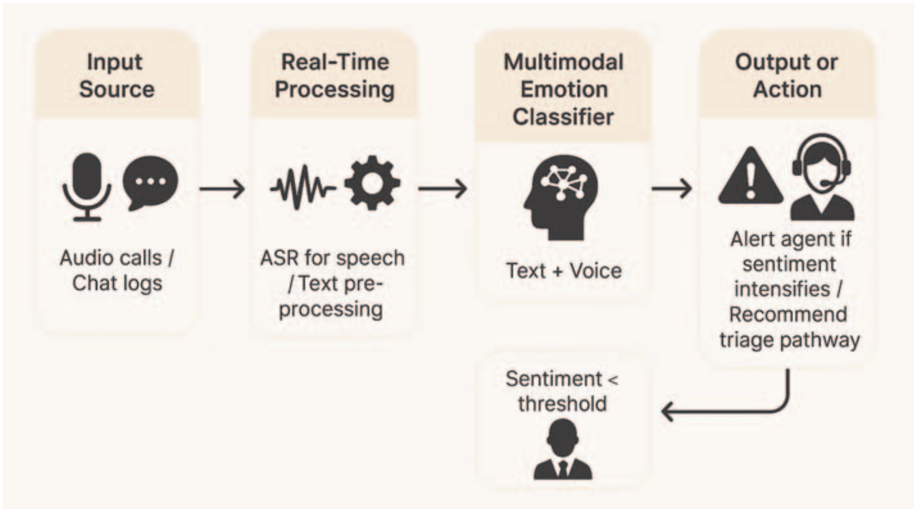


Figure 4
Workflow for real-time multimodal (Text + Voice) sentiment analysis with action triggering.

from corporate earnings calls or CEO interviews can refine risk and trading models even further given that multimodal sentiment analysis has demonstrated potential for improved market volatility predictions, as well as the fact that research has indicated that subtle changes in vocal pitch or hesitation during earnings calls may correlate with future stock performance anomalies. In this regard, an "empathy algorithm" capturing these cues might provide an informational edge in high-stakes financial trading. The other side of the coin is also true, and it likely will not be very long before we get algorithms to smooth out vocal pitches so as to filter out underlying emotional cues and make them undetectable by voice analysis algorithms.

A practical workflow example for real-time multimodal sentiment analysis with action triggering over text and voice is provided in **Figure 4**.

8. Best practices and evaluation

8.1 Benchmark datasets and appropriate metrics

Generally speaking, evaluation demands clarity in tasks, datasets, and metrics. In evaluating empathy algorithms, common metrics include accuracy, F1-score, and ROC-AUC for binary or multi-class classification, but tasks such as multi-label emotion recognition may require macro- or micro-averaged F1, or ranking-based metrics. Resources like the SemEval competitions (over various years) offer complex tasks on sentiment, sarcasm, and emotion detection across multiple languages, while TweetEval provides a unified benchmark for tweet classification [59]. Google's GoEmotions, with 27 emotion labels, is often cited as the a very prominent and widely used benchmark for fine-grained emotion classification, and Multimodal Opinion-level Sentiment Intensity (MOSI) was for some time the most popular used dataset to evaluate multimodal sentiment intensity detection in video data. This has now been superseded by its successor, Multimodal Opinion

Sentiment and Emotion Intensity (MOSEI), a larger and extremely popular benchmark, sometimes preferred for its scale and inclusion of discrete emotion labels alongside sentiment and Interactive Emotional Dyadic Motion Capture (IEMOCAP), which is also very popular, especially for discrete emotion recognition in dialogue.

Researchers and practitioners would be well-advised to document experimental details, from model parameters, to training epochs and hardware configurations with a view to ensuring perfect reproducibility, particularly in view of the fact that where technically- and commercially- feasible, releasing pre-trained models and code fosters transparency and comparability across different approaches.

8.2 Reproducibility and reporting

Deep learning models are path-dependent and can be highly sensitive to random initialisations, slight hyperparameter configuration changes, and data splits. As compute availability and large-scale pre-trained architectures and cross-domain fine-tuning increasingly become the norm, reproducibility can suffer if code and hyperparameters are not meticulously recorded or frozen through a snapshot. It is worth noting that many top-tier conferences and journals in NLP now either mandate or encourage the submission of reproducibility checklists detailing data usage, random seeds, and ablation studies, while open-source toolkits like Hugging Face Transformers further help standardise protocols.

8.3 Deployment considerations

Real-time sentiment detection requires efficient inference, and LLMs are computationally heavy, prompting research into model compression, pruning, and quantisation. Knowledge distillation (popularised with the advent of DeepSeek) can, for example, yield "lighter" student models that preserve much of the teacher model's accuracy. Edge-based sentiment applications – like mobile chatbots or AR/VR devices providing real-time emotional feedback – stand to benefit from such optimisations and have the potential of radically transforming a number of fields like education and on-the-job training. Scalability is also an issue, and systems ingesting high-volume Twitter streams or video feeds must be architected for parallel processing and robust streaming pipelines.

8.4 Data integrity and annotation protocols

High-quality sentiment analysis depends on the integrity of the data and its metadata, as well as on the consistency of annotations, so in the event that you are building your own dataset, detailed labelling guidelines must be developed, and annotators should receive thorough training with clear examples to ensure that subjective judgements are as consistent across the dataset as humanly possible. Supervision and double-checking, though costly, usually result in higher-quality datasets and subsequently better model accuracy. Alternatively or additionally, depending on the case, reporting inter-annotator agreement using metrics like Cohen's kappa or Krippendorff's alpha is essential for demonstrating the reliability of the annotations. Incorporating active learning strategies can further improve data quality by identifying and prioritising ambiguous samples for additional review.

8.5 Robust data splitting and cross-validation

Effective evaluation also involves using robust data-splitting strategies. Rather than relying solely on random splits, stratified partitioning is recommended to ensure that key attributes (such as sentiment distribution and demographic variables) are evenly represented in training, development, and test sets. In dynamic domains like social media, temporal splits that reflect chronological data are usually particularly useful. Furthermore, evaluating model performance on out-of-domain or cross-lingual datasets can provide insights into the model's ability to generalise beyond its training context.

8.6 Fairness and subgroup analysis

To address the potential for bias in sentiment analysis systems, it is important to disaggregate performance metrics by demographic features and categories, such as language, ethnicity, demography, dialect, or gender. This detailed analysis helps to identify any disparities in model performance that may lead to unintended bias. When imbalances are detected, strategies such as rebalancing the training data, employing adversarial training techniques, or using bias and fairness metrics to adjust the model can be applied to at least mitigate these issues.

8.7 Qualitative error analysis

Beyond quantitative metrics, conducting a qualitative analysis of errors can reveal recurring issues, such as difficulties in detecting sarcasm, handling code-switching, or interpreting cultural references, and these are not always evident from numerical scores alone. Detailed case studies of misclassified examples can provide valuable insights into model weaknesses that can then inform and govern further iterative refinements. Incorporating user feedback loops in deployed systems can also help to identify and rectify such errors in a real-world setting, flagging them for correction later.

8.8 Model calibration and threshold optimisation

It is also critical to assess whether the confidence scores produced by a model are well-calibrated against true outcome probabilities. Techniques such as Platt scaling or isotonic regression can be employed to improve model calibration, and optimising decision thresholds -especially in binary or multi-label classification tasks – ensures that the balance between precision and recall is appropriate for the application at hand, particularly in high-stakes domains such as the ones delineated earlier.

8.9 Continual and lifelong learning

Finally, as language evolves and new trends emerge, sentiment analysis models must adapt accordingly. Continual or lifelong learning strategies enable models to update incrementally without 'forgetting' previously learned information. Regular monitoring for model drift and periodic retraining on fresh data are essential practices to maintain performance over time, but at the same time care must be taken to avoid 'catastrophic forgetting', where new data disrupts the model's performance on earlier tasks.

8.10 Ethical and transparency documentation

Across all stages of development and deployment, ethical considerations and transparency are paramount, especially in jurisdictions that mandate this legally. Model cards or "fact sheets" can be created to detail the model's architecture, training data characteristics, and known limitations, thereby fostering trust among users, and if sensitive or personal data are involved – particularly in areas like mental health or social media analysis – clear documentation of privacy safeguards, controls and user consent protocols is essential to comply with regulatory standards and ethical norms [60].

9. The consilience of advanced technologies for next-generation sentiment analysis

The progression towards truly empathic AI hinges on the *consilience* (a "coming together") of diverse technological strands that have historically developed in relative isolation: sensor technologies emerged from human-computer interaction research, advanced neural architectures arose in the deep learning community, and privacy-preserving frameworks grew from security and cryptography fields. The next leap in sentiment analysis and empathy-related AI will likely arise from their intentional and purposeful convergence, and to illustrate how these strands can be woven together in practice, Section 9.1 will detail a specific architectural approach integrating several key advancements. These converging technological strands include:

1. *Sensor innovation and miniaturisation*: Wearable eye trackers, discreet physiological sensors, and smartphone-based facial recognition (as well as potential future miniaturised EEGs) can capture myriad micro-signals of user stress, boredom, excitement, or engagement. As these devices become more affordable, less intrusive and more portable, the volume and fidelity of available data will increase drastically.

2. *Neural-symbolic integration*: A new wave of research seeks to merge the strengths of data-driven neural networks with symbolic AI's interpretability and rules-based reasoning. Such hybrid systems might, for instance, learn user-specific emotional patterns from data but also apply logical constraints ensuring that improbable emotional transitions are flagged or reconsidered.

3. *Federated and privacy-preserving learning*: One of the biggest fundamental challenges at the moment is how to responsibly handle the sensitive nature of emotional data. In this milieu, emerging frameworks like federated learning enable user devices to locally train partial models, only sharing anonymised updates with a central server or distributed ledger. Differential privacy adds noise to these updates to protect user-level data, and these may become more crucial for large-scale adoption of empathic systems without compromising individual privacy [61].

4. *Quantum and high-performance computing*: Preliminary explorations of quantum computing for NLP, while still nascent because of the state of the art of quantum technology, could in the not-too-distant future facilitate faster training or much more advanced feature representation. Similarly, advanced Graphics Processing

Units (GPUs) and custom accelerators (Neural Processing Units [NPUs], Tensor Processing Units [TPUs], Field-Programmable Gate Arrays [FPGAs], etc.) are enabling real-time processing of multimodal data streams, making operational empathic computing feasible in time-critical applications like crisis helplines, stock trading, or immersive virtual reality settings. Quantum computing offers promising future possibilities for NLP, particularly in accelerated training of complex transformer models, novel quantum-inspired embeddings, and advanced optimisation methods, and hold significant potential to improve computational efficiency and possibly to even enable deeper contextual understanding in sentiment analysis, and despite still being experimental, Quantum Natural Language Processing (QNLP) has already been applied to sentiment tasks. Recent work by Ganguly et al. [62] demonstrated a QNLP model achieving perfect accuracy on a small sentiment dataset under simulation, hinting at potential quantum advantages in the future. While real quantum hardware results are more modest, these explorations suggest that quantum computing could eventually handle complex sentiment inference in ways classical models might not be able to.

5. *Multimodal fusion with domain-specific knowledge*: The next generation of empathy algorithms might fuse not only raw sensor data but also curated knowledge repositories and domain-specific rules. A mental health application might, for instance, integrate a knowledge graph about symptoms and treatments, enabling the system to cross-reference a user's emotional expressions with known clinical patterns, whereas a financial system could incorporate real-time economic indicators. This layered approach helps ensure that the sentiment analysis is both context-aware and practically actionable.

The consilience of diverse fields is shaping sentiment analysis beyond classification-based solutions into realms of AI that can respond adaptively to user emotions, potentially offering solace, guidance, or personalised experiences in near real-time. Section 9.1 provides a concrete example of how such consilience could look like in practice by detailing a proposed Neural-Symbolic Federated Learning architecture designed specifically for multimodal sentiment analysis and which aims to directly address the critical needs for interpretability, privacy, and robustness by combining several of the principles just outlined.

9.1 Neural-symbolic federated learning for multimodal sentiment analysis as an example of a consilient architecture under development

A growing body of work on multimodal sentiment analysis has shown that combining text, audio, and visual cues can substantially enrich the detection of subtle affective states, and yet, modern systems continue to grapple with three interrelated demands, *viz.*:

i. *interpretability*, so that key model decisions can be explained and justified;

ii. *privacy compliance*, given that data such as facial images, voice recordings, and textual diaries are often considered to be sensitive data, and fall under strict data-protection rules under some jurisdictions; and

iii. *robustness* in handling complex or ambiguous emotions such as sarcasm, code-switching, or culturally-nuanced expressions.

This section describes a unified neural-symbolic federated architecture that is currently being worked on by the present author to address these imperatives by:

1. Integrating *symbolic reasoning* – through knowledge graphs or ontologies – alongside powerful data-driven neural modules, which provides transparency and allows direct incorporation of domain- or culture- specific rules;

2. Employing *Federated Learning (FL)* to ensure that raw user data (including text, voice, and images) never leaves the local device, thereby reducing legal and ethical risks by preserving GDPR-compliant privacy-by-design; and

3. Emphasising *multimodal pipelines* to capture a broader spectrum of emotional signals than text-only or unimodal frameworks can achieve.

Although past research and existing code have explored each of these strands in isolation, few studies have explicitly combined them to produce a single, deployable system that is transparent, privacy-preserving, and capable of detecting fine-grained affective states.

9.2 Related work and motivations

Building on the principle of consilience, the practical realisation of next-generation empathy algorithms requires architectures specifically designed to overcome persistent challenges in interpretability, privacy, and robustness, especially when processing sensitive multimodal data [19, 23]. Current Deep Learning approaches, while powerful, often struggle to provide transparent justifications for their classifications [63, 64], face significant hurdles in complying with stringent data protection regulations like GDPR and PIPL when handling facial, vocal, or textual data [5], and frequently misinterpret complex affective states such as sarcasm [65], code-switching [3], or culturally-inflected expressions [6]. To address these intertwined demands, the work detailed in this section centres on a proposed unified Neural-Symbolic Federated Learning architecture currently under development that seeks to synergistically combine the strengths of symbolic reasoning – leveraging knowledge graphs or ontologies for transparency and rule-based grounding [66] – with the pattern-recognition capabilities of neural networks, all deployed within a Federated Learning (FL) framework [67, 68]. This federated structure is key to enhancing privacy by ensuring raw user data remains localised on the user's device [67–69]. While prior research has explored neural-symbolic methods [65, 66] and federated learning [67, 69, 70] individually, and sometimes in hybrid forms for other domains [71–75], their specific integration to create a transparent, privacy-preserving, and robust system tailored for multimodal affect detection remains a relatively nascent area. What follows next outlines the motivations, design principles, and core components of this envisioned architecture, detailing how local feature extraction, hybrid reasoning, and privacy-preserving aggregation mechanisms are combined to address the complexities of advanced sentiment analysis.

Neural-symbolic methods have been increasingly applied in various NLP domains [66]. Within sentiment analysis, knowledge graphs or lexical ontologies have been used to interpret ambiguous or context-dependent emotional expressions more reliably. By providing a stable set of symbolic constraints (e.g., domain rules, cultural norms, or cause–effect relationships), such systems facilitate interpretability and help disambiguate difficult cases like sarcasm [65]. However, these solutions are often developed in a centralised environment, where raw data from all users are pooled into a single repository, thereby raising legal and operational concerns about personal-data handling when large user populations are involved.

- *Federated Learning* (FL) has recently emerged as a compelling alternative to centralised training by transmitting only local model updates, in contradistinction to raw data, to a coordinating server [69, 76]. This decentralised paradigm is especially pertinent for tasks such as video-based or audio-based emotion recognition, where personal and potentially biometric information is at play [70]. Although purely neural FL-based approaches protect data privacy, they often inherit the interpretability limitations of Deep Learning. Their gradient-sharing protocols can also be vulnerable to adversarial attacks if carefully designed security measures like differential privacy and encrypted gradients are not implemented.

- *Hybrid federated approaches* are a response to early attempts at merging knowledge-based logic and federated architectures that have been mostly domain-agnostic, or limited to relatively simple classification tasks [73, 75], with concrete solutions that harness neurosymbolic FL to tackle the intricacies of multimodal affect detection (text, facial expressions, and vocal cues) being relatively rare. Additionally, the question of how best to maintain or update an ontology in a federated environment remains a tough nut to crack, as emotional conventions, gestures, and colloquialisms are usually regional and also evolve over time.

9.3 The architecture envisaged

With a view to be able to address the challenges and leverage the opportunities identified in the previous section, the proposed Neural-Symbolic Federated Learning architecture operationalises these principles through a carefully orchestrated workflow unfolding across several interconnected stages, beginning with localised data handling on the user's device to ensure privacy, moving through a hybrid reasoning process that combining data-driven insights with symbolic rules for enhanced interpretability, and finally culminating in a privacy-preserving aggregation mechanism for building global knowledge without centralising sensitive information. The following parts of this section flesh out these core components: local feature extraction and symbolic pre-processing, the hybrid reasoning layer where neural and symbolic information converge, and the federated aggregation strategy for secure knowledge sharing.

- *Local feature extraction and symbolic pre-processing*: Each user's device locally captures raw textual, audio, or video streams, thereby ensuring that personal data never leaves the user's control. Lightweight neural backbones, such as DistilBERT for text or small convolutional networks for facial gestures, convert these inputs into lower-dimensional embeddings. Concurrently, a mini-rule engine references a local emotional ontology that encodes relevant symbolic rules (e.g., *"High pitch often indicates negative valence"* or *"Smiling face with raised eyebrows denotes strong positivity"*) that can be tailored to local or cultural contexts, effectively forming a "symbolic cache" that preserves domain expertise within each device.

- *Hybrid reasoning layer*: The system then merges the neural embeddings with these rule-based signals to produce a “reasoning vector”, capturing both data-driven and symbolic cues. During inference, symbolic constraints intervene whenever contradictory or ambiguous evidence is detected (e.g., a text snippet that appears strongly positive, while the user’s micro-expressions reflect distress), thus flagging inconsistencies and allowing the framework to address a key shortcoming of purely neural systems, namely the opaque nature of learned features. Additionally, these symbolic rules or constraints can be updated dynamically if local data repeatedly contradict a prior assumption in the ontology (e.g., discovering that a local community uses certain facial gestures to mean surprise rather than annoyance).

- *Federated aggregation for global knowledge*: Once local training or rule updates are completed, each device sends *encrypted model gradients* or *high-level rule adjustments* to a coordinating server. Robust approaches to differential privacy and secure gradient sharing achieved through cryptographic means like homomorphic encryption minimise the risk of reconstructing personal data from these updates. Homomorphic encryption and similar techniques now enable computations to be performed on encrypted data, meaning federated models can aggregate user contributions without ever decrypting personal information: in practice, each device sends only encrypted gradient updates to the server; the server performs mathematical aggregation on these encrypted values, and the final model update is decrypted after aggregation. This ensures that even if intercepted, the individual sentiments or features remain indecipherable (as demonstrated in recent federated learning frameworks).

A weighted averaging process integrates parameters from multiple devices into a global model, effectively transferring knowledge across distributed populations while minimising privacy leakage. Certain symbolic constraints that prove consistently beneficial (e.g., newly validated rules about sarcasm) can be elevated into a *global knowledge base* that may be held centrally or on a distributed ledger, whereas regionally-idiosyncratic or culturally specific rules remain local to preserve context-sensitive accuracy.

Through this synthesis of neural and symbolic representations—encompassed in a distributed, privacy-preserving learning framework—the architecture aspires to retain interpretability while addressing the ethical and legal imperatives of large-scale affective computing.

A simplified overview of the architecture is provided in **Figure 5**.

9.4 Prospects and challenges

While the proposed Neural-Symbolic FL architecture presents a promising pathway toward more interpretable, privacy-preserving, and robust multimodal sentiment analysis, its practical implementation and widespread adoption hinge on navigating several significant prospects and inherent challenges. The potential advantages, particularly regarding enhanced interpretability and alignment with strict data protection regulations, are considerable, but realising this potential requires confronting key issues related to computational demands on local devices, the complexities of managing evolving symbolic knowledge across diverse user groups, and the persistent security vulnerabilities associated with distributed learning systems. Examining these

DOI: http://dx.doi.org/10.5772/intechopen.1010889

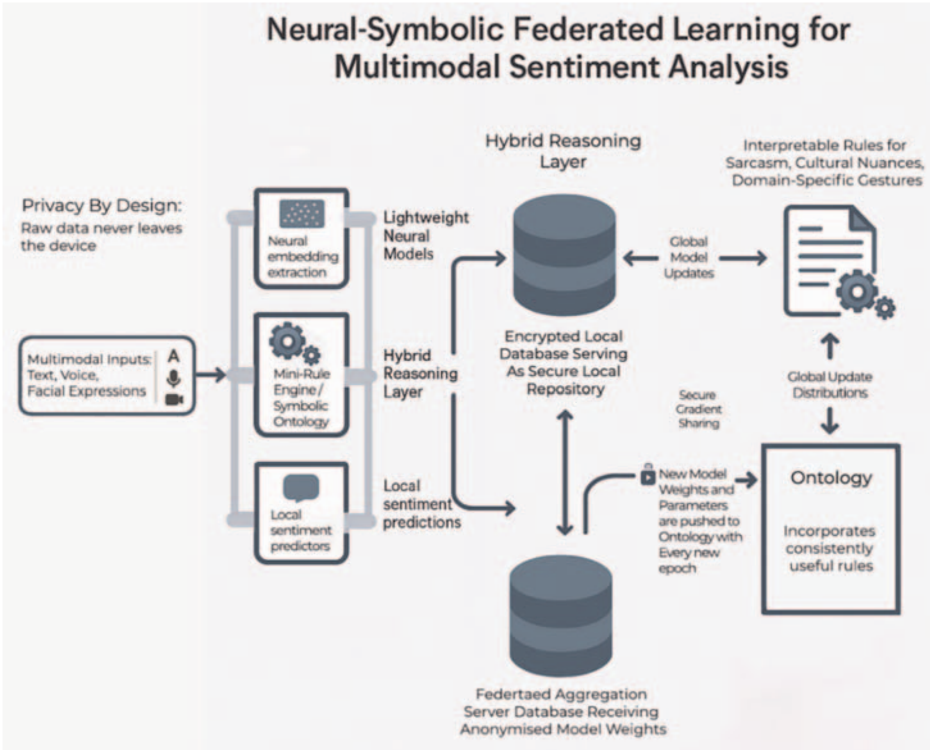


Figure 5.
System schema for neural-symbolic federated learning of multimodal sentiment analysis.

facets in detail is crucial for understanding the architecture's real-world viability as well as for guiding future development efforts, starting with its core strength in interpretability and trust.

By design, neural-symbolic systems produce *explanations* via symbolic inferences. This feature is vital when errors or misclassifications may have important real-world consequences, like misreading a user's stress level in a mental health app or pausing a learning activity due to inferred stress (an activity that is in any case likely to be prohibited in the EU under its AI Act, even though this is not fully clear at the time of writing of this chapter and is excluded if done for research purposes). Attaching each final sentiment label or emotion category to an explanatory rule path can bolster user and stakeholder trust, a vital consideration in a wide variety of fields from telehealth all the way to corporate HR analytics.

A federated approach means that local voice recordings, facial images, or text diaries are not transmitted. This has direct advantages under stringent legal frameworks such as GDPR and PIPL, and pre-emptively addresses proposed restrictions on emotion-recognition AI in documents like the EU AI Act. However, gradient inversion or membership inference attacks might still be a risk if updates are not rigorously protected.

Synchronising a local rule engine with neural embeddings may increase the computational overhead, especially on mobile devices, whereas model compression, pruning, and hardware acceleration can mitigate these issues, with fine-tuning and verifying rule sets being non-trivial in real-time scenarios.

A persistent open question is how to manage the drift or evolution of symbolic rules in globally distributed environments, as symbolic constraints that are valid in one culture or user group may not hold for others. In this regard, automated or semi-automated rule-updating pipelines – where potential new rules are cross-validated on a subset of devices – provide an interesting way forward, but implementing them reliably at scale is still somewhat challenging.

Even with federated updates, malicious actors may attempt to infer local data distributions from encrypted gradients, necessitating protocols that combine secure multiparty computation, differential privacy, and robust model aggregation to safeguard properly, which in turn strongly suggests that these will remain an active research area, due to the sensitive nature of emotionally-revealing data that might be prone to abuse in the wrong hands.

9.5 Use cases

Some examples of the use cases of this framework would include (but are by no means limited to) the following:

- *Mental health monitoring*: A smartphone-based app evaluates typed journal entries, vocal stress markers, and micro-expressions at the network edge on the device itself, with any potential sign of deepening depression or anxiety triggering an *interpretable flag* based on symbolic constraints drawn from validated psychiatric handbooks. Anonymised data collected from the device is sent to a federated learning model to make the model better and updates are pushed to the device for better-performing models when they become available, while privacy is respected, as no raw personally-identifiable data leave the device, aligning well with medical data protection regulations, including GDPR which is the most stringent;

- *Global call centres*: A call-centre solution analyses real-time audio (pitch and tone) alongside agents' textual summaries, while the local symbolic engine spots sarcasm if certain textual phrases and intonational cues contradict each other, and because sensitive recordings never leave local nodes, the system complies with cross-border data protection laws and responds better to users' requests as it otherwise struggles to capture these sarcastic undertones;

- *E-learning across cultures*: Students' webcams and textual responses are analysed locally to gauge engagement or confusion in an online course. A universal set of rules (e.g., standard facial expressions) coexists with region-specific rules capturing cultural nuances in gestures and facial micro-expressions, while the devices periodically share model updates, thus improving the global model while safeguarding personal data, thus leading to better learning outcomes and further tutor interventions where and when necessary;

- *Rehabilitation and assistive care*: In settings involving physical rehabilitation or long-term care, wearable sensors and local video streams are deployed to measure a patient's recovery progress, gauging frustration or motivational lapses via facial micro-expressions and vocal changes. A local rules-based module, enriched with medical ontologies, issues timely alerts to caregivers for personalised interventions and provides the federated model with totally anonymous data, ensuring sensitive biometric data remain on-device, thereby complying with stringent medical-data privacy regulations, while providing the federated model with anonymised data to still learn from anonymous data and ensure the long-term amelioration evolution of the model;

- *Corporate performance reviews and team dynamics*: In multinational corporations, employee one-on-one meetings are analysed for subtle cues of stress, burnout, or disengagement (as manifested in changes in pitch, negative sentiment in typed feedback, and the like). A distributed system of micro-clients on each participant's device captures these signals locally and shares aggregated representations thus helping supervisors provide generalised team-level support while adhering to labour regulations and preventing centralised storage of sensitive personal data;

- *Smart home environments*: Home-assistant systems equipped with cameras and microphones detect household occupants' emotional states on edge smart devices, such as escalating anger or heightened stress, and adapt the environment accordingly through smart home systems by, for example, dimming lights, adjusting temperature, or playing calming, soothing music. Symbolic constraints encode user preferences or medical guidelines (like, for instance, avoiding certain triggers for individuals with anxiety), while feeding federated models with optimisation data and receiving updates to allow the global model to improve progressively without uploading raw home-audio or video feeds to external servers;

- *Personalised language tutoring*: Language-learning platforms incorporating speech recognition analyse linguistic fluency and detect emotional cues that indicate confusion or frustration while local audio processing look for signals that reveal anxious hesitation, and facial tracking spots signs of embarrassment. The symbolic reasoning layer integrates specific pedagogical rules, like repeating grammar exercises if stress indicators rise above a certain threshold, thereby helping learners overcome specific language obstacles, and because all raw data processing occurs remotely on local devices, personal privacy is maintained all the time;

- *Driver well-being and traffic safety*: On-board systems in autonomous or semi-autonomous vehicles locally process voice tone, facial tics, and steering behaviour to infer drowsiness or road rage. In the meantime, a symbolic ontology encodes domain-specific rules like "frequent yawning plus drifting from lane indicates high fatigue and an increased likelihood of an imminent accident" and intervene by issuing audible alerts or temporarily engaging autonomous features. Federated learning allows the vehicle fleet's aggregated experience to improve detection of dangerous emotional or cognitive states, while preserving the privacy of individual drivers;

- *Political sentiment and opinion tracking*: Citizens using civic-engagement platforms record short video testimonies on local devices describing experiences with public services or policy where analysis of micro-expressions, vocal stress, and choice of words underscore deep-seated concerns or trust issues. Local symbolic knowledge then incorporates known rhetorical figures or culturally-tinged political references, refining the emotion inferences, while only model parameters are merged centrally, ensuring direct compliance with the more stringent personal data protection laws and mitigating concerns of political surveillance;

- *Retail and Augmented Reality (AR) shopping:* AR shopping apps overlay product information while simultaneously tracking the user's gaze fixation and vocal responses to recommended items, and if a user exhibits signs of irritation or hesitation, like, say, repeated re-scans of the same product coupled with negative utterances, the local symbolic engine can adapt the recommendation strategy. Federated updates allow improved recommendation logic across many users without transferring personal purchase preferences or biometric data to a central server; and

- *Virtual therapy and counselling*: Online counselling platforms can integrate real-time facial sentiment detection, voice waveform analysis, and typed statements, with a local symbolic rule set correlating specific patterns (e.g., mention of hopelessness, tearfulness in voice, downcast gaze) with clinical guidelines for depression or anxiety. Rather than uploading raw session recordings, the system aggregates symbolic rule outcomes for the global model, achieving improved accuracy through collective learning while maintaining the patient's confidentiality and adhering to mental-health data regulations.

9.6 Future potential

By weaving together the principles of neural-symbolic reasoning, federated learning, and multimodal sentiment analysis, this approach could be a promising path for next-generation "empathy algorithms". It addresses key functional demands like cultural adaptability, interpretability, and accuracy across diverse data types, while remaining sensitive to privacy and ethical constraints that are tightening worldwide, with the notable exception of the USA lacking a comprehensive Federal law (though a complex patchwork of state laws like California's CCPA/CPRA, Illinois' BIPA, Virginia's VCDPA, Colorado's CPA, Connecticut's CTDPA, Utah's UCPA, Washington's My Health My Data Act, and others impose significant obligations).

Important questions remain. These may be summed up into the 3 categories that follow:

1. *Quantitative gains*: How significantly does the hybrid approach outperform purely neural or purely symbolic pipelines on publicly available benchmarks (e.g., CMU-MOSEI, MISA)?

2. *Ontology maintenance*: Which mechanisms best handle real-time updates when contradictory rules emerge in local user data when you do not have individual-level variable data to use as discriminators?

3. *Robust security*: Can advanced encryption, differential privacy, and secure aggregation guarantee resilience against emerging adversarial attacks?

Progress on these fronts could catalyse a fundamentally new generation of interpretably "empathic" AI systems—ones that help us navigate sensitive emotional domains while ensuring privacy and incremental model amelioration over time.

DOI: http://dx.doi.org/10.5772/intechopen.1010889

10. Future directions and open challenges

10.1 Unified multimodal fusion and self-supervised learning

Despite notable advances, several of which will be covered in this chapter, most current multimodal frameworks still treat each modality in separate, parallel streams before merging outputs in a late-fusion step. More sophisticated approaches for early or joint fusion, especially ones that exploit self-supervised learning, may allow a model to learn cross-modal representations from massive unlabelled datasets. This happens, for example, when mapping user eye-tracking signals and speech prosody onto a shared embedding space that usually yield a richer, more interconnected representation of emotional states [77].

Emerging self-supervised multimodal frameworks, notably Contrastive Language-Image Pre-training (CLIP) and A Large-scale ImaGe and Noisy-text embedding (ALIGN) have shown remarkable capabilities in learning powerful multimodal representations without extensive manual annotations, and their many successes provide a promising pathway for sentiment analysis systems to leverage vast quantities of unlabelled multimodal data effectively. Additionally, Meta AI's ImageBind model which can jointly embed six modalities (including text, images, audio, and even depth sensor data) into a unified representation space[25] blog.roboflow.com, enabling cross-modal understanding without explicit supervision and Google's recent multimodal model *PaLM-E* (562B parameters), which combines vision and language, and has managed to achieve state-of-the-art performance on a visual question-answering benchmark *without* task-specific fine-tuning[26] while retaining strong language capabilities foreshadow a new generation of empathy algorithms that natively integrate diverse data sources (language, vision, audio) within one AI system.

10.2 Evolving "operational empathy": Toward personalised understanding

True "empathy" (affective empathy, that is) requires acknowledging individual differences, and a statement that is generally perceived as neutral might be deeply upsetting for someone who has experienced a related personal trauma. In the same way, cultural backgrounds strongly influence the reading of certain emotive cues. Personalisation strategies where the model gradually learns a user's baseline emotional patterns could also drastically increase accuracy, but must be balanced against privacy concerns and the risk of overfitting to idiosyncratic and *sui generis* user data. This is indeed impossible to achieve today while respecting privacy.

10.3 Lifelong and continual learning in sentiment analysis

Social language, slang, and dominant platforms shift rapidly, and traditional static models can fast become obsolete when new contexts, terminologies, or cultural references emerge. This makes Lifelong or continual learning techniques, in which an AI system updates its knowledge incrementally without forgetting previously learned tasks (a topic covered in Section 8.9), an interesting avenue to explore in

---

[77] *Vide* https://blog.roboflow.com/what-is-imagebind (accessed 10 April 2025).

[26] *Vide* https://research.google/blog/palm-e-an-embodied-multimodal-language-model/ (accessed 10 April 2025).

endeavouring to ensure that "empathy algorithms" keep pace with evolving linguistic and cultural landscapes. However, 'catastrophic forgetting', where new knowledge overwrites old knowledge, remains a major headache in incremental model updates and thus presents a considerable and as-yet unresolved challenge.[27]

A different approach could come in the form of human-in-the-loop strategies designed to make sentiment systems more adaptive by establishing active learning loops where an "empathy algorithm" consults a human moderator or domain expert when its confidence is low or when facing novel slang, and then learns from that feedback. Over time, these curated corrections – for example, a clinician confirming whether an ambiguous social media post indicates distress – become additional training data that expand the model's emotional understanding, ensuring that the AI continues to evolve with emerging language and cultural shifts, guided by human expertise to prevent drift or misinterpretation.

10.4 Addressing the data bottleneck

Obtaining large-scale, high-quality, ground-truth data for nuanced emotional states is often prohibitively expensive. Synthetic data generation, advanced data augmentation strategies in the form of Generative Adversarial Networks for creating additional labelled samples, and robust semi-supervised pipelines may help bridge this gap despite being only an imperfect substitute, while innovative annotation protocols like iterative human-in-the-loop setups that refine model predictions could further reduce the cost and improve the granularity of emotion-laden datasets.

Recent advances in Generative AI methods – such as Generative Adversarial Networks (GANs), diffusion models, and Variational Autoencoders (VAEs) – have demonstrated significant potential for generating realistic synthetic emotional datasets.

These methods help mitigate data scarcity, providing diverse and accurately-labelled samples for training robust emotion recognition systems. However, exclusive reliance on synthetic data can lead to overfitting 'distributional drift' or 'mode collapse' (typically observed after a few, typically 4 or 5, epochs of training) resulting in models that generalise poorly to real emotional data, and which can perform worse with every successive epoch. To counteract this, synthetic datasets should ideally complement real-world data rather than replace it, maintaining regular validation against real-world samples to ensure robust, generalisable performance (**Figure 6**).

10.5 Uncertainties under new regulatory frameworks

Next-generation sentiment analysis and empathy algorithms face an increasingly complex regulatory landscape. Indeed, despite the complexity of developing such systems, it may be said with quite a degree of confidence that it is child's play relative to trying to wade through the complex maze of international regulation, particularly that emanating from the European Commission. Increasingly, recent laws and

[27] While numerous techniques, including regularisation methods like EWC, rehearsal/replay methods, parameter isolation techniques, and generative replay have been developed to mitigate catastrophic forgetting, none have thus far completely solved the problem in a general, efficient, and scalable way without trade-offs, and in fact, this remains one of the most significant hurdles in developing truly adaptable and continuously learning AI systems. Finding a perfect balance between retaining old knowledge (stability) and acquiring new knowledge (plasticity) is an active and complex area of research.

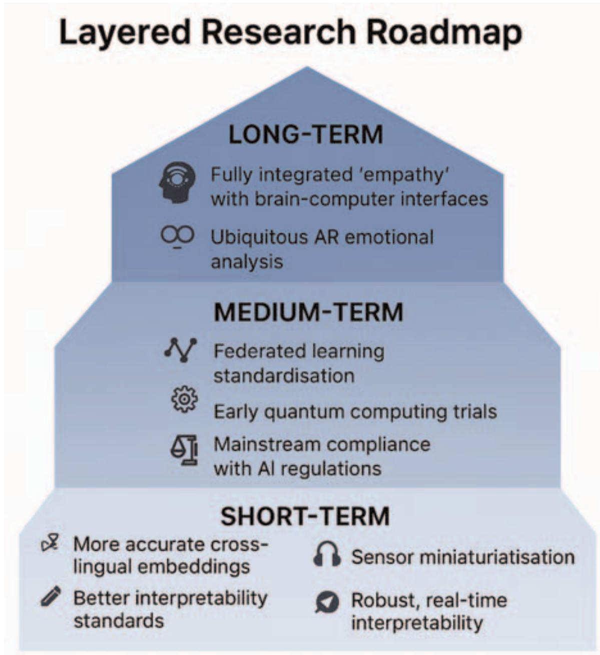


Figure 6.
The research roadmap.

proposals around the world, but perhaps most notably the European Union's AI Act, as well as data protection laws like the GDPR in the EU and PIPL in China, emerging U.S. state-level AI laws, and other international initiatives, are reshaping how such AI systems can be developed (or not developed, if they fall under the prohibited category under the EU AI Act) and deployed.

Strict data protection laws in the EU and beyond directly impact how sentiment and empathy AI systems collect, process, and store user data from text, voice, facial expressions, physiological signals, and other modalities. In the EU, the GDPR requires a lawful basis (such as informed consent) for processing personal data and enforces principles like data minimisation and purpose limitation, and while emotional or sentiment data can be highly sensitive, the GDPR does not explicitly list "emotion" as a special category of sensitive data – an omission that creates considerable legal ambiguity. Developers of affective computing systems struggle with this uncertainty, since emotion-related data might still fall under sensitive categories depending on context (think about, for example, facial images or vocal patterns could be treated as biometric data, or heart rate which could be classified as health data). The problem with this is that although Article 9 of the GDPR does not explicitly classify emotional data as sensitive, emotional data can be indirectly covered when it intersects with explicitly protected categories, particularly biometric or health data and thus, sentiment analysis systems employing facial expressions, physiological signals, or voice biometrics frequently find themselves subject to heightened GDPR compliance requirements, including explicit consent and rigorous data handling standards, while being subject to hefty fines if there is any breach in the process.

In practice, organisations usually either err on the side of caution by treating many forms of multimodal sentiment data as sensitive personal information, requiring either heightened protection and explicit user consent under GDPR's framework, or even total elimination from dataset and workflow pipelines, or else they disregard GDPR completely and shroud what they do in secrecy, taking any risks that might materialise in that operating space as a cost of doing business. GDPR also gives individuals rights over their data (access, deletion, objection, etc.), meaning systems that continuously monitor emotions must have robust mechanisms to respect these rights – by allowing users, for example, to opt out of emotional tracking and ensuring data used for sentiment analysis is securely stored and promptly deleted when no longer needed.

China's Personal Information Protection Law (PIPL), which broadly applies to any personal information handling, which includes virtually all AI services processing user data, imposes similar, stringent constraints. Under PIPL, sentiment analysis providers must obtain consent to collect and use personal data, with separate, explicit consent for sensitive attributes like facial or voice biometrics. They also face data localisation and transfer restrictions – large volumes of personal data or sensitive data (potentially including emotion-related biometrics) generally cannot be exported without stringent government security assessments, further complicating cross-border sentiment data collection. A company training an empathy model on Chinese users' facial expression data, for instance, may be legally required to keep that data on servers in China and undergo security reviews for any cross-border AI model deployment. Both GDPR and PIPL create a cautious environment around biometric and physiological data use, increasing compliance overhead for projects using webcams, microphones, or wearables to gauge emotions in a context where even seemingly benign data like text chat logs can contain personal sentiments or mental health cues that demand privacy safeguards, and with the net effect being that businesses and researchers must build flexible and robust privacy by design into sentiment AI systems – incorporating strong encryption, anonymisation/pseudonymisation of data, strict access controls, and data governance policies to meet these evolving standards. Yet, a high degree of uncertainty remains in how regulators will interpret these laws for novel emotion-sensing technologies, as evident from ongoing debates about whether "emotion data" should be considered on par with health or biometric data in terms of protection, hampering the growth of these technologies, which while risky also have some very socially beneficial use cases.

In addition to general data privacy laws, new AI-specific regulations are introducing transparency and fairness obligations that directly affect sentiment analysis algorithms. The EU's AI Act is on the so-called 'bleeding-edge': it represents the world's first comprehensive AI law employing a risk-based approach which will not be clear for decades as only snail-paced guidelines and legal judgements can provide ultimate clarity in a domain where technology is moving at the speed of light, while regulation still moves at the speed of an 1880's locomotive. Under the AI Act, many sentiment and emotion recognition systems (excepting research where it is still not clear whether this will encompass all research or public research) could, at best, be classified as "high-risk" AI, especially if used in sensitive domains like employment, education, or healthcare. Indeed, emotion recognition systems – AI that infers emotions or intentions from biometric data – are explicitly addressed in Article 3(54) of the Act (previously 3{39}), which defines these systems broadly (covering AI analysing facial expressions, voice intonation, physiological signals, etc. to deduce emotional states). Due to concerns about scientific validity and potential bias, the Act prohibits the use of emotion recognition in certain high power asymmetry settings: for example, it bans outright emotion AI in workplaces and schools (with some very narrow exceptions for safety or medical purposes). This means that an employer or educator in the EU cannot deploy a system to continuously analyse employees' or students' emotions, eliminating some use-cases of sentiment analytics entirely in those contexts, even where they could be potentially useful and highly beneficial. For other use-cases, the AI Act stops short of a blanket ban but still treats emotion recognition and advanced sentiment analysis as "high-risk" technologies, subjecting them to very stringent requirements that make them uneconomical to develop and impractical to deploy due to the resulting high compliance burden. Annex III of the Act lists high-risk systems and is expected to include many affective AI applications (employment, education, law enforcement, migration, access to essential services, and the like), reflecting regulators' worries that errors or biases in these systems could seriously affect people's lives, and triggering the high-risk classification if not already prohibited by Article 5.

Because AI regulations are evolving unevenly across regions, businesses and researchers deploying sentiment analysis systems globally face significant cross-jurisdictional compliance challenges. A system that is legal and acceptable in one jurisdiction may be restricted or expose the provider to liability in another. The upshots of this will be a highly-fragmented sentiment analysis market that will probably adopt the most stringent regulatory framework in a system's operating jurisdictions and adopt that, if GDPR is any guide, and provided further that the regulations do not conflict and do not ban the development of a system outright (in which case it would need to be developed elsewhere or not at all). Similarly, an emotion-aware wearable device that tracks a user's stress levels via heart rate and facial cues must heed GDPR/PIPL consent requirements in Europe and China, while also steering clear of any U.S. state where biometric or health data rules demand additional disclosures. Reconciling these varying obligations is non-trivial, and in practice with wearables moving on the tailwinds of international trade, they will probably be very difficult to enforce entirely. Be that as it may, anyone working on emotion recognition systems underlied by empathy algorithms might need to implement region-specific compliance programs: for instance, geofencing certain features (disabling or modifying them in countries where they are restricted), maintaining separate data storage silos to satisfy localisation laws, and tailoring user consent flows and privacy notices to each legal regime. One can also foresee problems arising where localised rules conflict with other jurisdictions' localised rules. Under GDPR, for instance, personal data (including emotional data) can only be exported outside the EU if the destination offers "adequate" protection or if special safeguards like Standard Contractual Clauses are in place. China's laws conversely require that data collected within China (especially large-scale personal data or anything deemed critical) stay within China's borders unless strict government approvals are obtained. This means a multinational project on sentiment analysis might have to maintain separate datasets and models per region, undermining the common practice of aggregating diverse global data to improve AI accuracy. As if though this was not complex enough, the EU AI Act's record-keeping requirement for high-risk AI, which could compel storing detailed model inferences and decision logs also conflicts with EU's own privacy laws that push for minimising stored personal data (and this is within the same jurisdictional bloc), effectively leaving developers in a situation where complying with one legislation puts them at odds with the other.

Steering through such jurisdictional conflicts is legally delicate – organisations often seek counsel on a case-by-case basis, and in some scenarios may decide to cease offering a service in a particularly restrictive jurisdiction to avoid non-compliance, and this is something that has already happened with some AI systems being restricted in Europe due to legal lack of clarity. Federated Learning models might provide a middle-of-the-way solution, but legal ramifications are still hazy. At the end of the day, until global standards harmonise (if they ever do), navigating the mosaic of AI laws will remain a significant uncertainty that researchers and practitioners must carefully manage as part of the development and deployment process.

11. Conclusion

Sentiment analysis has, in a very short time, evolved from simple polarity detection to a multi-faceted endeavour aimed at uncovering rich, and sometimes latent, non-conscious or subconscious emotional signals. Under the umbrella of "empathy algorithms", cutting-edge research incorporates transformer-based NLP, advanced representation learning, few-shot adaptation, and crucially, multimodal data signals including eye tracking, facial expressions, and vocal intonations, challenging the boundaries of how deeply AI can discern emotional states, offering nuanced benefits in domains as varied as mental health, user experience, social media monitoring, and financial market analysis. Nevertheless, despite the giant strides made in such a short period of time, numerous hurdles remain: sarcasm, irony, cultural and linguistic diversity, and subjectivity continue to pose intricate problems, while bias and fairness considerations also persist, reminding practitioners and researchers alike that empathic systems can be misled by skewed training data and may inadvertently perpetuate harmful stereotypes. Moreover, capturing and interpreting physiological or subconscious emotional cues in an automated manner raises vital questions about privacy and ethics.

Looking ahead, the consilience of sensor miniaturisation, advanced neural architectures, knowledge-based reasoning, privacy-preserving learning, and high-performance hardware points to an exciting future; one filled with discoveries and breakthroughs that will be foundational and that will enable successive generations of ever-more capable empathy algorithms. We can envision empathic AI agents that not only classify sentiment with a high level of accuracy, but also contextualise and interpret subtle cues in real time, even personalising these insights for individual users. Although true human-like (affective) empathy is still beyond the foreseeable horizon, with some researchers arguing that this might never actually happen, such systems will significantly deepen AI's capacity to "hear the faint whispers of the heart", potentially transforming how we communicate, care for one another, and navigate our increasingly complex digital lives.

DOI: http://dx.doi.org/10.5772/intechopen.1010889

Author details

Bernard Mallia[1,2,3]

1 Equinox Group, Malta

2 Institute for Research and Improvement in Social Sciences, Malta

3 Mediterranean Institute for Innovation, Communications and Technology, Malta

*Address all correspondence to: bernard.mallia@equinoxadvisory.com

References

[1] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019. pp. 4171-4186. DOI: 10.18653/v1/N19-1423

[2] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems 33. Vancouver, Canada: Curran Associates, Inc; 2020. pp. 1877-1901. Available from: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[3] Winata GI, Aji AF, Yong ZX, Solorio T. The Decades Progress on Code-Switching Research in NLP: A Systematic Survey on Trends and Challenges. arXiv preprint arXiv: 2212.09660; 2022. Available from: https://arxiv.org/abs/2212.09660

[4] Li Y, Li N. Sentiment analysis of Weibo comments based on graph neural network. IEEE Access. 2022;**10**:85468-85478. DOI: 10.1109/ACCESS.2022.3198100

[5] Shen H. China's Personal Information Protection Law: Implications for AI and Global Data Governance. SSRN Scholarly Paper No. 3965051; 2021. DOI: 10.2139/ssrn.3965051

[6] Senft N, Doucerain MM, Campos B, Shiota MN. Within- and between-group heterogeneity in cultural models of emotion among people of European, Asian, and Latino heritage in the United States. Emotion. 2023;**23**(1):1-15. DOI: 10.1037/emo0001184

[7] Barbieri F, Basile V, Croce D, Nissim M, Patti V, and Sprugnoli R. Subjectivity, polarity, and irony classification of Italian tweets. In: Chair NCC, editor. Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016). Naples, Italy: Accademia University Press; 2016. pp. 12-19. Available from: http://www.evalita.it/2016/proceedings

[8] Barbieri F, Camacho-Collados J, Espinosa-Anke L, Neves L. Sarcasm Detection in Tweets with Neural Networks. EVALITA 2016: 6th Evaluation of NLP and Speech Tools for Italian; 2016

[9] Basile P, Patti V, Cutugno F, Nissim M, Sprugnoli R. EVALITA goes social: Tasks, data, and community at the 2016 edition. Italian Journal of Computational Linguistics. 2017;**3**(1):93-127. DOI: 10.4000/ijcol.441

[10] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE; 2016. pp. 770-778. Available from: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf

[11] Sun Y, Wang S, Li Y, Feng S, Chen X, Zhang H, et al. ERNIE 3.0: Large-Scale Knowledge Enhanced Pre-Training for Language Understanding and

Generation. arXiv preprint arXiv: 2107.02137; 2021. Available from: https://arxiv.org/abs/2107.02137

[12] Wang H, Qiu X, Tan X. Multivariate graph neural networks on enhancing syntactic and semantic for aspect-based sentiment analysis. AApplied Intelligence. 2024;**54**:10185-10202. DOI: 10.1007/s10489-024-05802-6

[13] Zhang W, Li X, Deng Y, Bing L, Lam W. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering. 2022;**35**(10):10019-10038. DOI: 10.1109/TKDE.2022.3188690

[14] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in Neural Information Processing Systems 26. Lake Tahoe, NV, USA: Curran Associates Inc; 2013. pp. 3111-3119. Available from: https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

[15] Pennington J, Socher R, Manning CD. GloVe: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014. pp. 1532-1543 10.3115/v1/D14-1162

[16] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems 30. Long Beach, CA, USA: Curran Associates Inc; 2017. pp. 5998-6008. Available from: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[17] Hazarika D, Zimmermann R, Poria S. MISA: Modality-invariant and -specific representations for multimodal sentiment analysis. In: Proceedings of the 28th ACM International Conference on Multimedia (MM '20). Seattle, WA, USA: Association for Computing Machinery; 2020. pp. 1122-1131. DOI: 10.1145/3394171.3413672

[18] Tsai Y-HH, Liang PP, Zadeh A, Morency L-P, Salakhutdinov R. Multimodal transformer for unaligned multimodal language sequences. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019). Florence, Italy: Association for Computational Linguistics; 2019. pp. 6558-6569. DOI: 10.18653/v1/P19-1656

[19] Baltrušaitis T, Ahuja C, Morency L-P. Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018;**41**(2):423-443. DOI: 10.1109/TPAMI.2018.2798607

[20] Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY. Multimodal deep learning. In: Getoor L, Scheffer T, editors. Proceedings of the 28th International Conference on Machine Learning (ICML 2011). Bellevue, WA, USA: Omnipress; 2011. pp. 689-696. Available from: https://icml.cc/2011/papers/387_icmlpaper.pdf

[21] Feng J, Xu H, Joo J. Learn2Smile: Learning non-verbal interaction through observation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE;

2020. pp. 960-969 10.1109/CVPRW50498.2020.00126

[22] Poria S, Cambria E, Hazarika D, Majumder N, Zadeh A, Morency L-P. Context-dependent sentiment analysis in user-generated videos. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada: Association for Computational Linguistics; 2017. pp. 873-883

[23] Zadeh AB, Liang PP, Poria S, Cambria E, Morency L-P. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics; 2018. pp. 2236-2246. DOI: 10.18653/v1/P18-1208

[24] Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems. 2005;**6**(2):156-166. DOI: 10.1109/TITS.2005.848368

[25] Habib H, Nithyanand R. YouTube Recommendations Reinforce Negative Emotions: Auditing Algorithmic Bias with Emotionally-Agentic Sock Puppets. arXiv; 2019. Available from: https://arxiv.org/abs/2501.15048v1

[26] Sanchez K. Go Watch This WSJ Investigation of TikTok's Algorithm. The Verge; 2021. Available from: https://www.theverge.com/2021/7/21/22587113/tiktok-algorithm-wsj-investigation-rabbit-hole

[27] Khandelwal A, Britto R, Yuan X. Neural multi-task learning for biomedical negation and speculation detection. Bioinformatics. 2020;**36**(11): 4435-4442. DOI: 10.1093/bioinformatics/btaa123

[28] Khandelwal A, Britto R. Multitask learning of negation and speculation using transformers. In: Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis (LOUHI 2020). Association for Computational Linguistics; 2020. pp. 79-87. Available from: https://aclanthology.org/2020.louhi-1.9/

[29] Barbiero P, Ciravegna G, Giannini F, Lià P, Gori M, Melacci S. Entropy-based logic explanations of neural networks. Proceedings of the AAAI Conference on Artificial Intelligence. 2022;**36**(6):6046-6054. DOI: 10.1609/aaai.v36i6.20551

[30] Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics; 2019. pp. 3613-3618. Available from: https://arxiv.org/abs/1903.10676

[31] Chalkidis I, Fergadiotis M, Aletras N. LEGAL-BERT: The muppets straight out of law school. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Toronto, Canada: Association for Computational Linguistics; 2020. pp. 2898-2904. Available from: https://arxiv.org/abs/2010.02559

[32] Huang K, Altosaar J, Ranganath R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. arXiv preprint arXiv:1904.05342; 2019. Available from: https://arxiv.org/abs/1904.05342

[33] Lee J, Hsiang J, Huang CR. PatentBERT: Patent Classification with Fine-Tuning a Pre-Trained BERT Model. arXiv preprint arXiv:1906.02124; 2019. Available from: https://arxiv.org/abs/1906.02124

[34] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;**36**(4):1234-1240. DOI: 10.1093/bioinformatics/btz682

[35] Nguyen DQ, Vu T, Nguyen AT. BERTweet: A pre-trained language model for English tweets. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics; 2020. pp. 9-14. Available from: https://arxiv.org/abs/2005.10200

[36] Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: An evaluation of BERT in clinical element extraction. In: Proceedings of the 18th BioNLP Workshop and Shared Task. Florence, Italy: Association for Computational Linguistics; 2019. pp. 58-65. Available from: https://arxiv.org/abs/1906.05474

[37] Taiatu I, Avram AM, Cercel DC, Pop F. RoBERTweet: A BERT language model for Romanian tweets. arXiv preprint arXiv:2306.06598; 2023. Available from: https://arxiv.org/abs/2306.06598

[38] Seuß D. Bridging the Gap Between Explainable AI and Uncertainty Quantification to Enhance Trustability. arXiv preprint arXiv:2105.11828; 2021. Available from: https://arxiv.org/abs/2105.11828

[39] Koh PW, Nguyen T, Tang YS, Mussmann S, Pierson E, Kim B, et al. Concept bottleneck models. In: Daumé H III, Singh A, editors. Proceedings of the 37th International Conference on Machine Learning (ICML 2020), Proceedings of Machine Learning Research, 119. Vienna, Austria: PMLR; 2020. pp. 5338-5348. Available from: https://proceedings.mlr.press/v119/koh20a.html

[40] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE; 2017. pp. 618-626. Available from: https://openaccess.thecvf.com/content_ICCV_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html

[41] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems 30. Long Beach, CA, USA: Curran Associates Inc; 2017. pp. 4765-4774. Available from: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[42] Ribeiro MT, Singh S, Guestrin C. "why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). San Francisco, CA, USA: Association for Computing Machinery; 2016. pp. 1135-1144. DOI: 10.1145/2939672.2939778

[43] Wang J, Wei M, Yang J, Jia K, Kolb A, Zhang J, et al. Accurate eye tracking from dense 3D surface reconstructions using single-shot deflectometry. Nature

Communications. 2025;**16**:Article 178. DOI: 10.1038/s41467-025-56801-1

[44] Wang J, Wei M, Yang J, Jia K, Kolb A, Zhang J, et al. Accurate eye tracking from dense 3D surface reconstructions using single-shot deflectometry. Nature Communications. 2024;**15**: Article 2687. DOI: 10.1038/s41467-024-46801-1

[45] Lowerre B, Reddy R. The HARPY speech recognition system: Performance with large vocabularies [abstract]. The Journal of the Acoustical Society of America. 1976;**60**(S1):S10-S11. DOI: 10.1121/1.2003027

[46] Barrón-Cedeño A, Da San Martino G, Yu B, Majithia S, Nakov P, editors. Proceedings of the 13th Workshop on Computational Approaches to Subjectivity. Sentiment & Social Media Analysis. Toronto, Canada: Association for Computational Linguistics; 2023. Available from: https://aclanthology.org/2023.wassa-1.0/

[47] Bjerva J, Bos J, Haagsma H. The unreasonable effectiveness of pretrained features for learning symbolic meaning. In: Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020). Barcelona, Spain: International Committee on Computational Linguistics; 2020. pp. 5536-5541. Available from: https://aclanthology.org/2020.coling-main.487/

[48] Chandar R, Li X, Kumar S. An empirical study of cross-lingual transferability of contextual embeddings for low-resource languages. In: Findings of the Association for Computation Linguistics: EACL 2023. Dubrovnik, Croatia: Association for Computational Linguistics; 2023. pp. 1139-1145. Available from: https://aclanthology.org/2023.findings-eacl.82/

[49] Chong W, Erickson H, Kaur P. Evaluating large language models in identifying argumentative claims. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italy: European Language Resources Association and International Committee on Computational Linguistics; 2024. pp. 16439-16447. Available from: https://aclanthology.org/2024.lrec-main.1431/

[50] De Clercq O, Barriere V, Barnes J, Klinger R, Sedoc J, Tafreshi S. Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. Torino, Italy: Association for Computational Linguistics; 2024. Available from: https://aclanthology.org/2024.wassa-1.0/

[51] Ranjan P, Chen J, Balakrishnan V. Leveraging linguistic structure for cross-lingual annotation projection. In: Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing (DeepLo 2022). Dublin, Ireland: Association for Computational Linguistics; 2022. pp. 101-111. Available from: https://aclanthology.org/2022.deeplo-1.11/

[52] Zhang H, Wang S, Ng V. Subjectivity cues for multilingual stance detection: A transformer-based study. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics; 2023. pp. 590-600. DOI: 10.18653/v1/2023.acl-main.35

[53] Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, et al. A survey of uncertainty in deep neural networks. Artificial Intelligence Review. 2023;**56**

(Suppl 1):1513-1589. DOI: 10.1007/s10462-023-10562-9

[54] Kochenderfer MJ. Decision Making under Uncertainty: Theory and Application. Cambridge, MA, USA: MIT Press; 2015. Available from: http://web.stanford.edu/group/sisl/public/dmu.pdf

[55] Rußwurm M, Ali M, Zhu XX, Gal Y, Körner M. Model and data uncertainty for satellite time series forecasting with deep recurrent models. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE; 2020. pp. 7025-7028. DOI: 10.1109/IGARSS39084.2020.9323890

[56] Schum DA, Tecuci G, Marcu D, Boicu M. Toward cognitive assistants for complex decision making under uncertainty. Intelligent Decision Technologies. 2014;**8**(3):231-250. DOI: 10.3233/IDT-140201

[57] Seoni S, Jahmunah V, Salvi M, Barua PD, Molinari F, Acharya UR. Application of uncertainty quantification to artificial intelligence in healthcare: A narrative review. Computers in Biology and Medicine. 2023;**153**:Article 107441. DOI: 10.1016/j.compbiomed.2023.107441

[58] National Bureau of Economic Research. Which News Moves Stock Prices? [NBER Digest]; 2013. Available from: https://www.nber.org/digest/jun13/which-news-moves-stock-prices

[59] Barbieri F, Camacho-Collados J, Espinosa-Anke L, Neves L. TweetEval: Unified benchmark for tweet classification. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics; 2020. pp. 1644-1650. Available from: https://arxiv.org/abs/2010.12421

[60] Yu H, Miao C, An B, Shen Z, Leung C. Reputation-aware task allocation for human trustees. In: Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014). Paris, France: International Foundation for Autonomous Agents and Multiagent Systems; 2014. pp. 357-364. DOI: 10.5555/2615731.2615791

[61] Xu Y, Rao K, Wu H. Differentially private correlated representation learning for cross-modal sentiment analysis. IEEE Transactions on Affective Computing. 2024;**15**(1):872-884. DOI: 10.1109/TAFFC.2023.3319076

[62] Ganguly S, Morapakula SN, Pozo Coronado LM. Quantum Natural Language Processing based Sentiment Analysis using Lambeq Toolkit. arXiv; 2023. Available from: https://arxiv.org/abs/2305.19383

[63] Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI—Explainable artificial intelligence. Science. Robotics. 2019;**4**(37):eaay7120. DOI: 10.1126/scirobotics.aay7120

[64] Xu F, Uszkoreit H, Du Y, Fan W, Zhao D, Zhu J. Explainable AI: A brief survey on history, research areas, approaches and challenges. In: Huang C-C, Zhang M, Jagadeesh GR, editors. Natural Language Processing and Chinese Computing. NLPCC 2019. Dunhuang, China: Lecture Notes in Computer Science. Vol. 11839. Springer; 2019. pp. 563-574. DOI: 10.1007/978-3-030-32236-6_51

[65] Scotte S, De Silva LC. Explainable neuro-symbolic AI for sarcasm detection. In: Barrón-Cedeño A, Da San Martino G, Yu B, Majithia S, Nakov P, editors. Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment &

Social Media Analysis. Toronto, Canada: Association for Computational Linguistics; 2023. pp. 137-146. Available from: https://aclanthology.org/2023.wassa-1.13/

[66] Kocoń J, Baran J, Gruza M, Janz A, Kajstura M, Kazienko P, et al. Neuro-symbolic models for sentiment analysis. In: Sloot PMA, Dongarra JJ, van Albada GD, editors. Computational Science – ICCS 2022. Lecture Notes in Computer Science. Vol. 13350. London, UK: Springer; 2022. pp. 667-681. DOI: 10.1007/978-3-031-08754-7_69

[67] Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, et al. Advances and open problems in federated learning. Foundations and Trends in Machine Learning. 2021;**14**(1–2):1-210. DOI: 10.1561/2200000083

[68] McMahan B, Moore E, Ramage D, Hampson S, Aguera y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: Singh A, Zhu J, editors. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017), Proceedings of Machine Learning Research, 54. Fort Lauderdale, FL, USA: PMLR; 2017. pp. 1273-1282. Available from: https://proceedings.mlr.press/v54/mcmahan17a.html

[69] Vasanthi P, Vankadara M. Federated learning for multimodal sentiment analysis. IEEE Access. 2024;**12**:175218-175239. DOI: 10.1109/ACCESS.2024.3424374

[70] Wu Z, Zhang Q, Miao D, Yi K, Fan W, Hu L. HyDiscGAN: A Hybrid Distributed cGAN for Audio-Visual Privacy Preservation in Multimodal Sentiment Analysis. arXiv preprint arXiv:2404.11938; 2024. Available from: https://arxiv.org/abs/2404.11938

[71] Fiosina J. Explainable federated learning for taxi travel time prediction. In: Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2021). Prague, Czech Republic: SciTePress; 2021a. pp. 670-677. Available from: https://pdfs.semanticscholar.org/478f/054c0eca4f9170bf5f3f7c02e4b3d53bcb00.pdf

[72] Fiosina J. Interpretable privacy-preserving collaborative deep learning for taxi trip duration forecasting. In: Brooks AL, Kim SK, Håkansson E, editors. Intelligent Computing. Lecture Notes in Networks and Systems. Vol. 499. Cham, Switzerland: Springer; 2021b. pp. 392-411. DOI: 10.1007/978-3-031-17098-0_20

[73] Xing P, Lu S, Yu H. Federated Neuro-Symbolic Learning (FedNSL). arXiv preprint arXiv:2308.15324; 2024. Available from: https://arxiv.org/abs/2308.15324

[74] Yang F, Abedin MZ, Hájek P. An explainable federated learning and blockchain-based secure credit modeling method. European Journal of Operational Research. 2024;**317**(2):449-467. DOI: 10.1016/j.ejor.2023.09.037

[75] Zhang Y, Yu H. LR-XFL: Logical reasoning-based explainable federated learning. Proceedings of the AAAI Conference on Artificial Intelligence. 2024;**38**(19):21788-21796. DOI: 10.1609/aaai.v38i19.30179

[76] Singh A, Chandrasekar S, Saha S, Sen T. Federated meta-learning for emotion and sentiment aware multi-modal complaint identification. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 16091–16103). Association for Computational

DOI: http://dx.doi.org/10.5772/intechopen.1010889

Linguistics. https://aclanthology.org/2023.emnlp-main.999

[77] Sobal V, Zhang W, Cho K, Balestriero R, Rudner TGJ, LeCun Y. Learning from Reward-Free Offline Data: A Case for Planning with Latent Dynamics Models. arXiv preprint arXiv: 2310.17641; 2024. Available from: https://arxiv.org/abs/2310.17641

Chapter 4

Research on the Implementation of Algorithmic Thinking and Sentiment Analysis by Applying the Knowledge-Based System

Yan Li

Abstract

Artificial intelligence technology is becoming increasingly mature and is being widely implemented and applied in academia and industry to solve practical problems. Due to the gradual growth and high development of AI, a large number of universities lack artificial intelligence literacy teaching for students and faculty members, especially the lack of educational innovation and exploration of algorithmic thinking and sentiment analysis for researchers. AI literacy is an emerging field that aims to equip individuals with the knowledge and skills to understand, interact with, and make informed decisions about AI technologies. This study aims to practice and reform the BERT-based model to develop an algorithmic thinking education innovation exploration curriculum and conduct sentiment analysis. To improve AI literacy and the efficiency of undergraduate students' programming ability, reduce the burden of data retrieval work on students, and enhance the efficiency of students' programming learning, this research has developed a series of algorithmic thinking courses based on the BERT model and aims to develop a knowledge graph-based question-answer system to enable undergraduate students to understand the operating rules and basic syntax of programming. The study finds that the system collects data from student and teacher portals and uses these data for sentiment analysis to optimize the system and allow researchers to efficiently obtain information for self-directed learning.

Keywords: AI literacy, algorithmic thinking, sentiment analysis, machine learning, knowledge engineering

1. Introduction

In the higher education context, universities are encouraged to provide specialized AI workshops and broader computer science courses that include hands-on practice of AI tools, ethical discussions, and interdisciplinary applications. This research aims to practice and reform the algorithmic thinking education innovation exploration courses based on the BERT model. It aims to create a question and answer system based on the BERT model and collect relevant data from students and teachers to perform sentiment analysis on the system. According to research by scholars and university strategies, developing undergraduate AI literacy is necessary for the research and learning cycle [1]. The teaching team should provide practical and direct teaching, systematically improve the learning efficiency of researchers, and regard the artificial intelligence literacy of students and researchers as the goal of lifelong learning [2]. Teaching content could include ethical considerations for AI, or more workshops could be held to raise public awareness. In the study, a BERT-based system has been proposed and applied to practical teaching for undergraduates in Chinese-style education and UK-style education. This research also focuses on reforming a series of teaching activities for undergraduates, developing more algorithmic thinking courses, artificial intelligence literacy courses, and roundtable discussions. This study provides an in-depth exploration of approaches to improving undergraduate students' AI literacy in higher education settings based on different teaching and research styles, applying the "classroom dialog" [3] as the basic idea and also as a guide for developing AI literacy for students and researchers.

The main research objectives of this study are as follows. The study will develop a series of algorithmic thinking innovation courses for undergraduate students based on the BERT model, aiming to improve artificial intelligence literacy from different schools and majors to better adapt to the digital age. Secondly, this study plans to design an algorithm model for knowledge base question and answer, divide the question and answer module into two subtasks, and propose two corresponding models (kmBert twin tower model and ken-KGQA model). In addition, this study aims to conduct experiments and analysis on a self-built question and answer data set in the programming field to test and confirm whether the above two models are effective compared to the classic model. Finally, this research aims to design a knowledge question and answer system to improve undergraduates' artificial intelligence literacy and programming learning efficiency, collect relevant data from teachers and students for sentiment analysis, and optimize the system. The system allows students and researchers to ask questions to the system independently when they encounter programming problems, query similar information that occurred in history, and obtain corresponding answers. It improves the independent learning efficiency and enthusiasm of undergraduates, thereby achieving the goal of popularizing and cultivating undergraduates' artificial intelligence literacy.

2. Related work

AI literacy has become more indispensable in all aspects of life and work with the continuous development of AI technology, and AI literacy has gradually transformed into an increasingly important ability in the field of education. AI-driven tools such as BERT-based question-answering systems provide opportunities to achieve the goal of lifelong learning and improve students' proficiency in learning effectiveness. This study aims to explore how BERT-based systems can be used to improve the AI literacy of undergraduate students. Some scholars believe that BERT from Transformers has revolutionized the approach to natural language processing (NLP) tasks. BERT models are pre-trained on a large corpus of text and then fine-tuned for specific tasks. In question-answering systems, BERTs excel at understanding the context of queries and deriving the most relevant answers from a corpus of text [4]. A significant advantage of BERT-based question-answering systems is their ability to provide a precise answer to the question by understanding subtle differences in the meaning of words due to context. This feature makes BERT useful as an advanced educational tool and helps to help undergraduates understand complex material in greater depth.

This study aims to propose a new BERT model to improve AI literacy and conduct sentiment analysis among undergraduate students. By interacting with a BERT-based question-answering system, learners can actively participate in AI technology, demystifying how it works and what it can do. Some scholars believe that using AI systems for inquiry-based, hands-on learning can help deepen understanding and memory. The use of a question and answer system can enhance critical thinking and problem-solving skills to a certain extent. For example, the learner asks a question that leads to the best answer. It encourages them to think critically about the structure of questions and answers and how AI interprets queries, leading to better problem-solving skills [5]. In addition, BERT-based systems can be tailored to the learner's level of knowledge, providing a more personalized educational experience. By analyzing the responses to questions, the system is adapted to provide more targeted educational content that appropriately challenges learners [6]. The system serves as a guide to lifelong learning goals. In the context of lifelong learning, the system can collect the learners' feedbacks over a period of time to help learners develop better search and prompting skills, which are essential for effective information literacy. Users learn to navigate through large amounts of information quickly and efficiently, identifying high-quality sources [7]. As for academic use, the system can accelerate the development of abstract terms, concepts, or theories by quickly and efficiently extracting relevant information and summarizing findings from large paper datasets. This ability strengthens the learning memory of students and researchers, not only saving time but also enhancing the breadth and depth of reviewable meaning [8]. The system collects direct feedback from both the teacher and the students' sides to construct the model of sentiment analysis [9].

2.1 Named entity recognition

With the rise of machine learning in recent years, statistical machine learning-based methods have gradually replaced dictionary- and rule-based methods. Different from the previous method, which has poor portability and high cost, the method based on statistical machine learning is to manually label the corpus and then train the annotated corpus to obtain a language model suitable for this task. The annotated corpus does not need to have rich knowledge, the cost is not high, and it only needs to replace the training corpus when transplanting to a new field, so it is portable. There are four main named entity recognition methods based on statistical machine learning: Support Vector Machine (SVM), Hidden Markov Model (HMM), Maximum Entropy (ME), and Conditional Random Fields (CRF) [10]. Among these methods, ME is compact and versatile, but it is expensive due to the high complexity of the training time. SVM should be used as a supervised learning method to effectively handle the classification of high-dimensional feature spaces and to solve the problem of small samples by relying only on a small number of samples when making decisions.

2.2 Pre-trained language models

BERT is composed of a multi-layer Transformer Encoder, which adopts the attention mechanism [11], which abandons the traditional RNN and CNN and converts the distance between two words at any position into 1 through the attention mechanism, which effectively solves the thorny long-term dependency problem in NLP. BERT jointly adjusts the context in all layers to train a depth bidirectional representation, which can be fine-tuned with an additional output layer. The success of the model is mainly due to the two training methods it proposes: MLM (Masked Language Model) and NSP (Next Sentence Prediction), which can help BERT learn vector representations of words from the vast linguistic corpus available. The notation learned by BERT has been shown to be well generalized to downstream tasks.

3. Methodology

3.1 Theoretical framework

This study applies the Dialog Framework [3] as a teaching method for the curriculum design of the AI literacy workshop and a basic principle for engineering pedagogy. The framework is often applied to the design or evaluation of instructional activities, particularly in advanced technology-enhanced learning environments. The framework serves as a tool to guide the development of teaching and learning processes, emphasizing the importance of dialog (interaction between students and teachers) and interaction in knowledge construction. It analyzes the different stages of the learning process. Each stage aims to connect the dialogs in order to bridge the gap between theory and practice. It shares a feedback loop throughout the learning process to enhance undergraduates' understanding of subject knowledge, and teachers can use feedback to refine their teaching methods and fill gaps in student understanding. The framework provides students with the role of active learners. In order to strengthen teaching, some of the teaching courses will be relocated to the Wonder Lab. Each workshop consists of 60 minutes of lectures and 30 minutes of exercises, with small class sizes in practice and large classes in the teaching of the basics, and instruct students to write and publish in engineering pedagogy-related journals and conferences every year and attend some of the top conferences in the field of AI literacy so that they can keep up with the latest research trends.

3.2 Rationale

Developing AI literacy and conducting sentiment analysis among undergraduate students and researchers:

The widespread adoption of AI is reshaping the academic landscape of research and learning, transforming them from traditional to innovative approaches. This shift has had a significant impact on learning and scientific inquiry methods in higher education institutions. More and more universities around the world are involved in AI research. In order to improve researchers' AI literacy and use AI technology to transform academic support, Github has established and designed an AI tutorial website to cultivate researchers' AI fundamentals, including AI coding assistants and BERT-based systems, designed to help researchers start self-directed learning.

Building an interdisciplinary artificial intelligence working group and artificial literacy workshop:

Since AI is a multidisciplinary field of study that requires a variety of skills, it is possible to establish cross-university and interdisciplinary AI working groups and conduct more AI literacy workshops or digital job visits so that academics, administrators, and students from different backgrounds can discuss and progress together.

3.3 Key research problems

This study aims to address the following key research questions:

1. How to make good use of existing resources to conduct the sentiment analysis, integrate the BERT model into the existing digital curriculum for undergraduates, and cultivate their algorithmic thinking?

2. What is the performance of the proposed system?

3. To what extent can the system improve the AI literacy or programming learning efficiency of undergraduate students?

This study intends to solve the following four basic technical problems:

1. Based on the construction of a data collaborative representation model in complex data space, the vertical island problem of data knowledge is solved, and the knowledge model of multi-source heterogeneous data is constructed.

2. Realize the extraction, association, and integration from local knowledge to global knowledge.

3. The deep learning algorithm is used to solve the situation recognition of complex product collaborative workflows, and the adaptive service recommendation in complex scenarios is realized through the service recommendation algorithm.

4. Research on collaborative decision-making methods and construct collaborative interactive decision-making models based on BERT.

4. Overall pipeline

The dictionary and rule-based method is an early method used in the named entity recognition task, which often requires experts to construct a rule template in the corresponding field, select specific features including position words, central words, direction words, and keywords, and make corresponding matches according to the specified matching rules and entity dictionaries to identify entities. Dictionary- and rule-based methods have a better recognition effect than machine learning-based methods when the rules and dictionaries are more detailed and accurate, but the formulation of these rules often requires domain experts to make corresponding control of domain knowledge and text style, and the compilation process is time-consuming and costly, and when the language style and domain involved are different, the method performance will be greatly reduced. Therefore, this method is not only time-consuming and labor-intensive but also has poor portability and a low recognition rate of new words. This model solves the problem that LSTM only obtains information in one direction and can obtain forward and reverse features at the same time, making the acquisition of contextual information more complete. However, the above models require a large number of datasets and have requirements for dataset quality, so people have begun to think about how to obtain

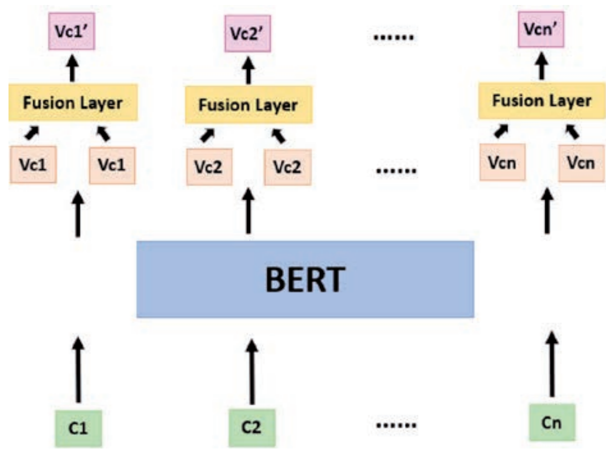


Figure 1.
Overall working principle of the BERT model.

a priori semantic information from a large number of unlabeled texts and apply it to subsequent training. Based on this idea, the Transformer model was born. Since then, a variety of pre-trained models derived from Transformer have emerged, and BERT is one of them (**Figure 1**).

4.1 Sentiment analysis

The system review data serves as learners' direct feedback on the system experience and contains rich information value. The study uses Python crawler technology to collect comment data, stores the collected data results in the SQLite database, uses a Python third-party library to process and analyze the comment data, and uses Python's Flask framework to build a Web application to achieve dynamic visual display of comment data. Through in-depth mining of student and teacher comment data, we can understand students' evaluations of different knowledge points, emotional tendencies, popular questions, popular knowledge points, etc., providing a scientific basis for the optimization of the question and answer system and public opinion monitoring. In terms of data visualization (**Figure 2**), the systematic review data analysis system uses a variety of chart forms, such as bar charts, pie charts, and word cloud charts, to intuitively display the analysis results. Through these charts, students and teachers can easily understand the distribution of popular issues and knowledge point data, the trend changes of new knowledge points, and other information. In this study, students' and teachers' feedbacks were collected as the data, and some typical algorithms (e.g., KNN Classifier, Naive Bayes, and SVM Classifier) were imported and applied in the experimental process. Feature vectorization and the Naive Bayes classification are applied during the data dealing process, and the accuracy of the model during training is around 83%.

4.2 Experimental setup and result analysis

The entity recognition experiment in this article is constructed using the PyTorch deep learning framework and the Python 3.6.5 programming language. The specific software- and hardware-related parameters are shown in **Table 1**.

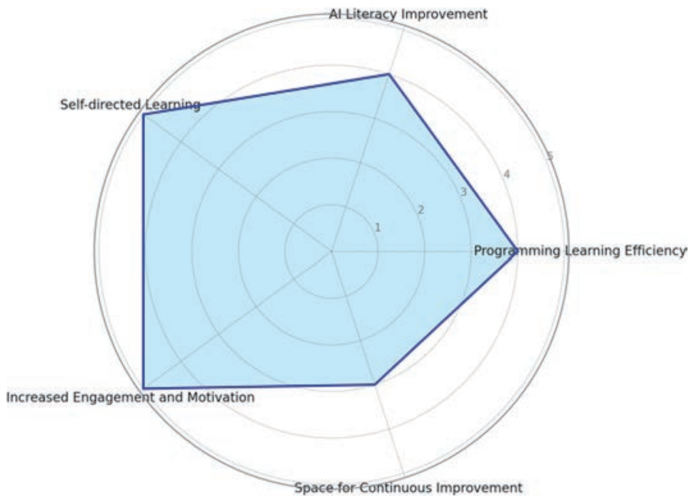


Figure 2.
Key aspects of students' feedbacks.

Software and hardware	Related parameters
GPU	NVIDIA 3080TI 32 GB
RAM	32 GB
Deep learning framework	PyTorch
Programming language	Python 3.6.5

Table 1.
Software- and hardware-related parameters.

The settings of hyperparameters have a great impact on the performance of the model. The hyperparameter settings of our model are as follows. For the knowledge graph embedding stage, the pre-trained model loaded by BERT is Chinese-BERT-base, and the output dimension is 770. Dropout processing is performed on BERT's parameter learning, and the dropout probability is set to 29%. The dimension of the embedding vector output after passing through a fully connected layer is 66 dimensions. The Adam optimizer with a learning rate of 0.0001 is used for optimization. When the batch size is set to 130, the training epoch is 50 rounds. The specific parameter settings of the knowledge graph embedding model are shown in **Table 2**.

In the text encoding stage, the dimension of the output vector of each character in the BERT model is 770, the pre-trained model loaded by BERT is Chinese-BERT-base, and the dropout rate is set to 50%. The dimension of the embedding vector output after passing through the fusion layer is still 64 dimensions. In the label encoding stage, the BERT model used by the label encoder and the BERT model used by the text encoder are the same model and share parameters. The embedding vector of the label is the 768-dimensional vector output by [CLS] at the top level of the BERT model. The dimension of the embedding vector output after passing through the fusion layer is still 66 dimensions. Using the Adam optimizer with a learning rate of 0.00005, the training phase of named entity recognition was trained for 60 epochs with a batch size of 65. The specific parameter settings of the model are shown in **Table 3**.

Name	Related parameters
BERT pre-training model	Chinese-BERT-base
BERT maximum input sequence length	50
BERT output dimension	770
BERT dropout rate	0.29
Fusion layer output dimension	66
Learning rate	1e – 4
Optimizer	Adam
Batch size	130
Epoch	50

Table 2.
Related parameters knowledge graph embedding model (former).

Name	Related parameters
BERT pre-training model	Chinese-BERT-base
BERT maximum input sequence length	50
BERT output dimension	770
BERT dropout rate	0.5
Fusion layer output dimension	66
Learning rate	5e – 5
Optimizer	Adam
Batch size	65
Epoch	60

Table 3.
Related parameters knowledge graph embedding model (later).

4.3 Result analysis

For the model in this article, we conducted experiments on a self-built data set. We conducted comparative experiments between the model in this article and several classic models in the field of named entity recognition to verify the effectiveness of the model: (1) the LSTM-CRF model. The embedding layer uses pre-trained Word2Vec for word vector embedding, and then LSTM-CRF performs entity extraction. (2) The Bi-LSTM-CRF model is a classic model in the NER field. It uses pre-trained Word2Vec as the output of the embedding layer and then inputs it into the Bi-LSTM-CRF network for encoding and predicting labels. (3) BERT-CRF model, using the BERT pre-trained language model for word vector embedding and then decoding by the CRF layer to calculate the optimal annotation sequence.

The results are shown in **Table 4**. From the experimental results of each model, it can be found that the accuracy, recall rate, and F1 value of the model used in this article all achieve the best results compared with the other three comparison models. The F1 value of the model in this article reached 95.9%, while the F1 values of the other three baseline models, LSTM-CRF, Bi-LSTM-CRF, and BERT-CRF, were 82.6%,

Model	Precision	Recall	F1-score
LSTM-CRF	85.1	80.3	82.6
Bi-LSTM-CRF	87.2	83.2	85.1
BERT-CRF	95.6	95.2	95.4
KLBERT (our)	96.1	95.8	95.9
KLBERT-t	94.8	94.3	94.5
KLBERT-l	95.2	94.7	94.9

Table 4.
Results on the students' dataset.

85.1%, and 95.4%, respectively. It can be seen that the recognition effects of the first two models are far behind BERT-CRF. The reason may be that the BERT pre-trained language model can better learn the universal features of language and has better generalization effects. The recognition effect of the KLBERT model in this article is slightly better than the BERT-CRF model. The reason may be that the KLBERT model incorporates lexical information, learns more language features, and produces better performance. As shown in the table, our model performs better than other models, which proves the validity of the idea in this paper.

To further analyze our model, this paper conducts ablation experiments to systematically study the impact and contribution of different components. KLBERT-t and KLBERT-l are variants of our full model KEm-KGQA. Here we briefly introduce these two model variants.

KLBERT-t: In the label encoding stage, the label is not expanded or converted into a description.

KLBERT-l: In the text encoding stage and label encoding stage, the encoding of text and labels does not fuse lexical information. After the text and labels pass through the BERT model, they no longer enter the fusion layer but directly perform similarity calculations.

The results of the ablation experiment are shown in **Table 4**. Based on the comparison of results between the model in this paper and its variants, we can clearly

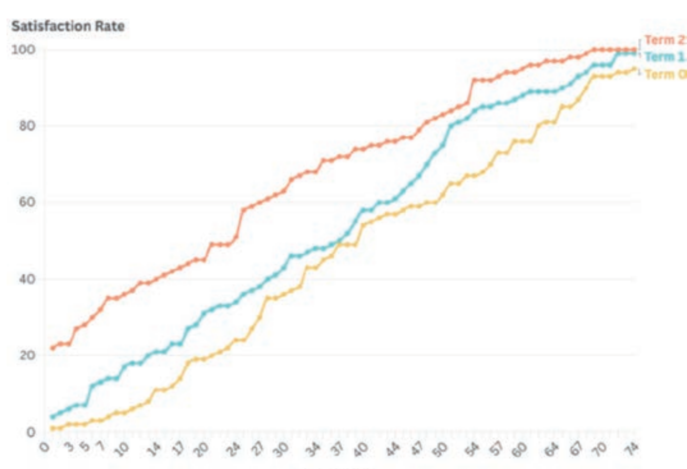


Figure 3.
Satisfaction rate of students.

conclude that the semantic expansion of labels and the incorporation of lexical information have an impact on the results. The comparison between KLBERT-t and our complete model shows that if the labels are semantically expanded during the label encoding stage, the F1 value of the named entity recognition model can be improved by 1.4%. This shows that the semantic expansion of tags plays a role in named entity recognition. When our full model is compared with KLBERT-l, we can see that the F1 value of the named entity recognition model can be improved by 1.0% if the corresponding lexical information is incorporated into the encoding of text and tags, which proves the importance of incorporating lexical information. Within the 2 years, the system witnessed nearly 1000 students' growth. **Figure 3** shows the students' satisfaction rate in three terms. It demonstrates that the satisfaction rate has successfully increased by 29% compared to before, and convinced of the effectiveness to some degree.

5. Future development

The dictionary and rule-based method is an early method used in the named entity recognition task, which often requires experts to construct a rule template in the corresponding field, select specific features including position words, central words, direction words, and keywords, and make corresponding matches according to the specified matching rules and entity dictionaries to identify entities. Dictionary- and rule-based methods have a better recognition effect than machine learning-based methods when the rules and dictionaries are used. This study conducts in-depth research on knowledge graphs and question and answer systems based on the field of higher education. With the advent of the era of artificial intelligence and big data, the demand for question and answer systems based on knowledge graphs will only continue to increase. Although the current research on question and answer systems combined with knowledge graphs is still in its preliminary stages, with the increase in market demand, research progress will inevitably advance by leaps and bounds, and the related technologies of question and answer systems combined with knowledge graphs will also continue to innovate. The question and answer model proposed in this article can currently only answer simple questions; it is still difficult to solve complex relational questions.

Future work can be carried out from the following aspects:

1. Improve the system's ability to understand complex relational questions. With the development of the times, the amount of data will inevitably skyrocket in the future, and the relationship between information will become more complex.

   Therefore, research on the understanding of complex relationship questions is imperative.

2. Complete the knowledge graph in the field of computer science in colleges and universities. At present, the knowledge graph mainly focuses on students' basic knowledge questions and answers, and running event information. For questions that are beyond the scope of the knowledge graph, the system will be overwhelmed. In the future, we can integrate more information into the knowledge graph, develop an automatic data collection function, regularly search for data in

open source websites and databases, perform automatic collection and processing, and import it into the knowledge graph.

3. Enrich the functions of the question and answer system. Most of the question and answer systems currently on the market have some recommendation function. Therefore, we can also make personalized recommendations based on the historical data of user operations. In addition, we can provide voice recognition and image recognition functions, and users can also recognize and answer input voice.

6. Conclusion

This study designs and implements a question and answer system in the field of higher education with the purpose of improving the collaboration and interactivity of teachers and students in universities. This system can be used to ask basic information about students' programming abilities as well as information about abnormal events in the system's operation. This study introduces the overall architecture of the question and answer system in universities and the design and implementation of each functional module. The algorithm model of the question and answer system is integrated through the Python back-end framework Flask; the front-end display is completed using the React front-end language and front-end frameworks such as Echarts and Ant Design to implement functional modules such as entity search, entity recognition, and question and answer in the field of university computer science. The question and answer system is based on the knowledge graph in the information field of universities constructed in this article, and its back-end algorithm model is processed by the model in this article. The system will collect basic information from teachers and students during use for sentiment analysis, thereby optimizing existing scenarios to a certain extent and allowing students and teachers to understand the system's operating information in real time. When encountering problematic events, students and teachers can ask questions to the Q&A system and query information about similar operating events that have occurred in history, including basic computer knowledge, operating status, event resolution measures, etc. The information can be used for self-study and to improve the programming efficiency of students and teachers.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive suggestions and feedback. This research is supported by the research project JGBA2024536, in part by a grant from the Zhejiang Provincial Department of Education.

Author details

Yan Li
University of Nottingham Ningbo, Ningbo, China

*Address all correspondence to: sylvie.li@nottingham.edu.cn

References

[1] Southworth J, Migliaccio K, Glover J, Reed D, McCarty C, Brendemuhl J, et al. Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence. 2023;**4**:100127

[2] Alamäki A, Nyberg C, Kimberley A, Salonen AO. Artificial intelligence literacy in sustainable development: A learning experiment in higher education. Frontiers in Education. 2024;**9**:1343406

[3] Gregory MR. A framework for facilitating classroom dialogue. Teaching Philosophy. 2007;**30**(1):59-84

[4] Sultani M, Daneshpour N. Extracting urgent questions from MOOC discussions: A BERT-based multi-output classification approach. Arabian Journal for Science and Engineering. 2024;**50**(2):1169-1190

[5] Xu Y, Shieh CH, van Esch P, Ling IL. AI customer service: Task complexity, problem-solving ability, and usage intention. Australasian Marketing Journal. 2020;**28**(4):189-199

[6] Suryanarayana KS, Kandi VP, Pavani G, Rao AS, Rout S, Krishna TSR. Artificial intelligence enhanced digital learning for the sustainability of education management system. The Journal of High Technology Management Research. 2024;**35**(2):100495

[7] Zheng J, Qiu S, Shi C, Ma Q. Towards lifelong learning of large language models: A survey. ACM Computing Surveys. 2024;**57**(8):1-35

[8] Rani N, Chu SL, Williamson YG. Supporting lifelong learning. Educational Technology & Society. 2024;**27**(2):256-269

[9] Yao L, Zheng N. Sentiment analysis based on improved transformer model and conditional random fields. IEEE Access. 2024;**12**(1):90145-90157

[10] Pradhan A, Yajnik A. Parts-of-speech tagging of Nepali texts with bidirectional LSTM, conditional random fields and HMM. Multimedia Tools and Applications. 2024;**83**(4):9893-9909

[11] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in Neural Information Processing Systems. 2017;**30**(1):1-15

Section 3

Global Diffusion of Sentiment Analysis Applications

Chapter 5

Perspective Chapter: Artificial Intelligence in Slovak Radio Industry – The Present and the Future of Broadcasting

Lucia Furtáková, Ľubica Janáčková and Andrej Brník

Abstract

The chapter addresses the implementation of artificial intelligence in the radio industry in Slovakia and analyses its impact on various aspects of broadcasting. Artificial intelligence (AI) is increasingly used to automate processes in the media, as radio stations are using AI to generate news, personalise content, or create synthetic presenters. The chapter traces the historical development of AI in the radio environment and its current use in Slovak radio stations, highlighting both the benefits and risks associated with AI. Particular attention is paid to text generation by generative AI and listener reactions to artificially generate voices, with results showing that despite technological advances, the public still prefers human presenters. The future of AI in Slovak radio is also discussed, underscoring dependence on regulations, technological advances, and the media's ability to find a balance between AI efficiency and the human factor.

Keywords: artificial intelligence, automated journalism broadcasting, generative AI, radio, radio broadcasting, radio industry, Slovakia, Slovak radio industry

1. Introduction

The rapid advancement of artificial intelligence (AI) has significantly transformed various industries, including media and broadcasting. Radio, as one of the oldest and most resilient forms of mass communication, is also undergoing a technological shift driven by AI. From early attempts at simple, rule-driven automated music playback to today's sophisticated generative models, the implementation of AI models is reshaping the way radio stations operate, engage audiences, and optimise content distribution. However, it should be noted that the radio stations have been using artificial intelligence long before the arrival of ChatGPT in late 2022 (see **Table 1**). Although some of these technologies were originally developed mainly abroad, in recent years they have also entered Slovak broadcasting practice; thus, it is crucial to understand the evolution of AI in radio that has shaped its development. The following timeline highlights the most important advances in AI applications in radio (**Table 1**).

Year	Radio station	Country	AI technology	Main tasks
1979	NBC	USA	RCS Selector	Automated playback and playlist creation
1980s	Multiple local stations	USA	RCS Selector	Basic "smart" music rotation, automated ads insertion
2000	Pandora (internet radio)	USA	Music Genome Project (ML* + human oversight)	Personalised "stations" by musical attributes
2011	KROV-FM	USA (San Antonio)	Digital automation with basic AI elements	Recording of texts, answering phone calls, checking emails, searching the web, and scheduling meetings
2017	iHeartMedia "SmartAudio"	USA	Analysis + ML (SmartAudio)	Predictive analytics for ad targeting and content optimisation based on audience data
2020	BBC "Beeb"	UK	BBC voice assistant	Voice control of content, personalisation, British dialect recognition
2022	Futuri Media TopicPulse	USA (global)	AI analysis of trending topics	Real-time monitoring of social networks and news to suggest discussion topics for broadcast
2023	RadioGPT	USA, Canada	Futuri's TopicPulse + AudioAI	AI-generated "DJ", combining breaking news, weather, and local information

**ML—machine learning.*
Source: own processing, 2025 according to Ref. [1–7].

Table 1.
Timeline of AI use in radio broadcasting.

Since Futuri launched the first AI-powered local radio [7], artificial intelligence has begun to play an important role in radio programmes around the world. Slovak radios—whether public service, commercial, or Internet—can be inspired by the above-mentioned proven solutions from abroad. The aim of this chapter is therefore to explore how artificial intelligence is already shaping the Slovak radio space today, what its advantages and limitations are in different areas, and what can be expected in the future.

2. Literature review

Artificial intelligence in the radio environment is proving to be a good servant but a bad master. It can efficiently process the routine performances of radio workers but on the other hand, it can cause complications that lead to both confusing and bizarre situations. Therefore, it is important to approach the categorisation of the advantages and disadvantages of using these tools in such a way that can predict the possibilities, threats or limits of the use and implementation of artificial intelligence in broadcasting, as well as the processes related to it.

The use of AI tools by media organisations in newsgathering can be broadly categorised into two main areas: optical character recognition (OCR), speech-to-text, and text extraction. This encompasses the use of AI tools to automate transcription, extract text from images, and structure data after gathering. Additionally, trend detection and news discovery represent another key application of AI in this field. These AI applications are capable of sifting through vast quantities of data and detecting patterns, such as data mining [8]. Furthermore, the radio industry is concerned with extracting service texts from big data that could have processed into an understandable form in a shorter amount of time. With AI, it is easy to create records of fictional events, but in which real people—politicians, artists or celebrities—are featured. They all look realistic, and the lay user may not be able to tell whether the videos are generated by artificial intelligence or edited with its help [9].

The (r)evolution in AI can be expected to change the labour market. In the field of journalism, it can be used for several purposes; for example, in the areas of creating questions, generating texts, using artificial voices, and there is speculation that AI could partially replace journalists. For now, this functions as so-called automated journalism, at least for producing routine stories that make journalists' work more efficient. Another type of AI utilisation is AI presenters, for several reasons: cost savings, the possibility of continuous broadcasting, immediate reaction to situations that arise, but this is also linked to the fact that fact-checking is needed, and that this type of presenter can be used for propaganda purposes [9]. AI algorithms are revolutionising radio personalisation by analysing listening habits to create tailored experiences. These systems generate content and recommendations aligned with individual preferences, covering news, music, and sports. AI also facilitates cost-effective content creation and distribution while enhancing ad targeting using listener data such as location and demographics [10]. As Furtáková and Janáčková state:

> *"AI can help the creative team to create texts, find topics for the programme, find additional sources of information, prepare different genres of radio journalism easily and quickly, compile and formulate a complete programme script or record an advertising spot according to the given parameters, and even speak or interpret the text according to how we want it to sound." [11, p. 102]*

According to Amponsah and Atianashie, AI has made journalism more efficient by automating tasks such as data collection and sorting. This has allowed journalists to focus on more complex aspects of storytelling. AI can process large amounts of data, which helps journalists find trends, and patterns. It also creates news content based on user preferences, which makes people more engaged. AI has also changed automated reporting, which generates reports for data-driven stories such as financial summaries, and election results, allowing journalists to explore complex topics [12].

While there are advantages to using AI in the media space, it is also important to consider the threats and disadvantages of using it. This is due to the training of the language model, the representation of relevant data, and other technological aspects. Furthermore, it depends on which version of the language model is used. For example, our previous research has shown that ChatGPT 3.5 errors significantly more often than technologically advanced ChatGPT 4 (or ChatGPT 4o) [cf. 13, 14]. Thus, if one desires to use the tools, the check of the outputs is always required. Moreover, AI is yet unable to distinguish between humour and relevant information. This is proved by occasional nonsense output [13]. Manipulating real material can diminish the credibility of government officials as well as politicians and thus undermine the democratic processes in the state. The same is true in the case of war conflicts as well as fraudulent actions on the internet (involving deceptive advertising). Among the threats of AI is the obfuscation of human-to-human communication. Even today, we already are witnessing an apparent shortening of words, sentences, the younger generation expressing themselves in slogans, and limiting their interpersonal contact, which will be further exacerbated by the COVID-19 pandemic in the post-2020 era.

In this context, the so-called relational artificial intelligence, which describes the service of artificial intelligence to simulate a friend, also comes to the surface. For many users of relational artificial intelligence, a situation may arise where they consider the relationship with the digital person to be real, which may in the long term promote toxic relationships, inappropriate relational stereotypes, unrealistic ideas, etc. [9]. In this case, it is also a disadvantage from a radio broadcasting perspective, as it is generally accepted that radio broadcasting should be lively, friendly, and conversational with a given real presenter who would represent a companion for the listeners.

The integration of AI into radio production presents several challenges. First, AI algorithms may reflect biases present in their training data, which can compromise the impartiality of the content. Additionally, the lack of transparency about AI-generated content may erode listener trust and raise concerns about its credibility. As AI becomes more proficient, there is also the risk of job displacement, with human roles in scriptwriting, editing, and audio production potentially being replaced, thus devaluing human expertise. AI can create convincing fake audio, which could spread misinformation. If we rely too much on AI for content, it could create echo chambers where only content that agrees with what people like is shown. AI also does not understand cultural nuances, humour, and politics, which can lead to mistakes in the content. This proves the need to find a balance between AI's efficiency and human creativity, ethics, and cultural awareness in radio production [12].

It is anticipated that the AI's abilities will be honed, depending on how it communicates in each language, working with intonation, stress, or melody in the voice. All of the mentioned can be implemented in radio broadcasting. Nevertheless, AI-related processes are also linked to the so-called back-office, that is digital advertising or music dramaturgy workers (**Table 2**).

AI can generally also help with translation, generate text, and analyse it but when it comes to source and fact-checking, and creating stories, it is often filled with misinformation or the stated facts are inaccurate [14]. Media organisations that have adopted AI tools in their newsrooms face a number of limitations. One of the biggest questions facing the industry is how are journalists able to adhere to the principles of their profession while working with algorithms that are likely to change journalistic practices [15]. This information will eventually evolve and change in the future as AI and its tools develop. Although for now it summarises the basic opportunities and threats current AI radio broadcast environment offers. Moreover, new employees can be trained through AI, through routine exercises, or simulated interviews for reporters.

The use of new technologies can bring both positives and negatives. As we have already encountered many examples of both in practice, it is necessary to look for an ideal formula of how artificial intelligence could be implemented in real radio production and in such a way that it does not deprive listeners of the necessary "proximity", and intercession. Moreover, so it does not cause an outflow and reduction of professional staff, which, as the practice shows, the industry needs, because they have experience, the ability to think operationally, and also have knowledge of local or regional events, which need to be connected in a logical way to be broadcast in the radio.

At the same time, however, it is necessary to adapt the possibilities artificial intelligence offers to practice in a way that effectively trains existing staff in radio as

Pros	Cons
Automation of content preparation: AI can generate questions, texts, and prepare news.	**Loss of authenticity**: Listeners expect live presenters who are friendly and spontaneous.
Cost savings: Using AI presenters reduces expenses on human staff.	**Lack of emotional interaction**: AI cannot replicate human emotions or live reactions.
Continuous broadcasting: AI presenters can operate 24/7 without breaks.	**Risk of propaganda**: AI presenters could be manipulated to spread disinformation.
Quick response to current events: AI can instantly process information and integrate it into broadcasts.	**Insensitivity to humour and irony**: AI struggles to distinguish between humour and serious content.
Efficiency in processing archival content: OCR and text extraction from recordings for quick reuse.	**Errors in content**: Older models may generate inaccurate information.
Improved accessibility: Speech-to-text aids in creating captions and archives.	**Loss of originality**: Automated reports may feel monotonous and mechanical.
Preparation of automated reports: AI handles routine tasks, easing journalists' workload.	**Ethical issues**: Manipulating audio recordings can erode trust in broadcasts.
Flexibility in broadcasting: AI can be deployed in emergencies or for rapid updates.	**Limitation of creativity**: AI might hinder the development of creative radio formats.

Source: own processing, 2025.

Table 2.
The pros and cons of using AI in radio broadcasting.

well as students who are just coming into contact with new knowledge. It is important to simplify and streamline routine tasks, as well as teach artificial intelligence what is time-consuming for radio workers. In radio, these can be mechanical tasks that AI can learn, such as marking the chorus of songs, the so-called hook, or the intro and outro of songs, as well as taking advantage of the possibilities of creating and placing advertisements, either directly on the radio or in online form.

3. Methodology

In order to complete this chapter, we conducted a critical reflection and analysis of the current state of AI use in Slovak radio industry by reviewing both national and international literature, including relevant academic monographs and journals, as well as individual radio station websites. For this reason, theoretical concepts are also presented in the Results section, together with the relevant sources.

Subsequently, we use content analysis to check whether the generated text meets the criteria of a radio text (or radio message). In content analysis, the text is examined by the categories that are anchored in the theoretical conceptions of radio news creations. These categories include, for example, genre, length of the content, number of words in a sentence, logical sequence of information, lexis and stylistics, etc. (for more, see Refs. [13, 16, 17]). In our previous research, we focused on how to implement these criteria in prompts to design a "universal" prompt that radio news presenters could input into any generative AI tool in order to facilitate the rewriting of news piece from news agencies into the form of radio report. The final version of the prompt is as follows:

> *Rewrite the text in angle brackets "< >" in the form of a short read radio report. The generated text must meet the following criteria: the length of the report must be*

> *no more than 6 sentences; the number of words per sentence must be no more than 20; alternate between simple sentences and clauses; the sentences must be simple; the information in the report must follow logically; the dates in the report must be recorded for the day of the week with respect to today's date; you must not directly address the listener; and there must be no editorial person present in the report. Within the production, adhere to the principles of news objectivity, brevity, and clarity; professional tone and style; and focus on the most important information. [18]*

This prompt, along with the news agency's text, was then fed into five generative AI tools: ChatGPT4-o1, Microsoft CoPilot, Gemini Advanced 1.5 Pro, Claude 3.5 Haiku, and Perplexity Free.

The ongoing research on listener perceptions of AI-generated voices used in radio broadcasting took place in The Laboratory of Neuromarketing Studies—NEUROLAB at FMK UCM (for more, see Ref. [19]) in September 2024. The preliminary study involved 16 respondents, aged 18–70, who listened to recordings of four different radio presenters. Using a micro-emotion tracking system, we monitored facial reactions and recorded the number of frames where a probability threshold was crossed for each emotion analysed.

4. Results

4.1 The creation of radio texts through generative AI

In the recent years, there has been a critical breakthrough in which systems based on artificial intelligence principles can build a sufficiently complex language model based on the learning process to generate results linguistically close to human-produced text [20]. This technology started to assist journalists in various aspects of their work, offering opportunities for creativity, innovation, and improved efficiency. Generative AI tools learn from large datasets (including images, text, audio, and video) and use that knowledge to generate new samples that are similar to the training data. They can then generate new instances of data that possess similar characteristics, structure, and distribution to the training data [21]. Content generation involves the use of AI algorithms to automatically create news articles, reports, summaries, and other journalistic content. This dimension emphasises the ability of AI systems to generate coherent and contextually relevant texts, mimicking human-like writing styles [21]. Alawida et al. [22] point out that although generative AI tools (such as ChatGPT, Gemini, or Copilot) can help with many tasks, particularly those with high levels of repetition and redundancy, they are not a replacement for human intelligence. They explain that human intelligence is capable of understanding context, interpreting meaning, and making connections. They specified the differences between human-generated text and AI-generated text as follows:

- human-generated text is characterised by the ability to convey meaning and intent, which also reflects cultural and emotional intelligence;
- human-generated text uses figurative language, idiomatic expressions, and cultural references, which vary across settings and audiences;

- AI-generated text is based on patterns and stored data, and therefore may struggle to take into account and understand cultural and social contexts;
- AI-generated text may also contain grammatical, stylistic, and word-formation errors;
- AI-generated text may not take into account or recognise language that is considered exclusive or derogatory based on social norms and other cultural circumstances.

These findings are particularly evident in texts that are generated in a language other than English. One example is an error in Norway newspaper. The AI-generated text said Norwegian football star Erling Haaland was shot [23]. In this case, it was a problem of translation, because the AI tool interpreted the English word "shoot" not as "photographed", but "shot". The problem of the language barrier was also confirmed by our previous research [18, 24], specifically that current AI tools cannot generate Slovak radio news fit for broadcast without further editing by a news presenter.

According to the 2023 global survey, 90% of the 105 participating news and media organisations from 46 different countries (excluding Slovakia) said they use generative AI (such as ChatGPT) in their news production phase [8]. As far as Slovakia is concerned, according to Tinák and Gáliková Tolnaiová [25], the interest in AI in newsrooms is undoubtedly growing, but the actual implementation process is happening at a slow pace, and the use of AI is still in the process of initial exploration and experimentation. Mikušová [26] adds that many Slovak journalists are unfamiliar with AI tools. Unlike global counterparts, Slovak newsrooms lack formalised policies to guide the adoption and integration of AI into their workflows. Current AI applications in Slovak journalism are limited to routine tasks, and advanced uses such as data-driven reporting or automated content creations remain unexplored. The Institute for Public Affairs stated that 57% of Slovaks reported experience with different versions of ChatGPT, 18% have used or tried CoPilot, 12% Gemini, and under 5% have experience with Claude or similar AI tools [27]. Out of the most used generative AI tools in Slovakia—using the mentioned "universal" prompt—the best rewriting of news piece from news agencies into the form of radio report was made by Claude (**Table 3**).

Although there are still problems with the Slovak language within the generated texts, the standard of these texts is not worse than what can be heard in the current radio broadcasts of Slovak commercial stations. The most problematic is the use of verbs in the passive voice. We assume that this problem comes from the fact that the primary language of these tools is English and not Slovak, so they cannot distinguish between the verb form in the "active voice" and the "passive voice". The same is valid for technical expressions. Concurrently, a significant linguistical shift in the development of generative AI tools can be observed. A research from 2023 states that these tools are not applicable in the Slovak radio industry, as multiple commands were needed to generate the basic text, which would still have to be additionally edited by the news presenter [24]. However, the current results show that these (single-prompt) generated news reports—after minor and quick adjustments by the news presenter—could be used in any radio station in Slovakia.

Criteria for the radio text*	Generative AI tools				
	ChatGPT	CoPilot	Gemini	Claude	Perplexity
Length of the radio report (maximum 6 sentences)	✓	✓	x (7 sentences)	✓	✓
Number of words in each sentence (maximum 20 words)	✓	✓	✓	✓	✓
Alternate simple sentences and clauses	✓	✓	✓	✓	✓
Simple clauses	✓	✓	✓	✓	~
The information must follow logically	~	x	✓	~	~
Answering the 5 Ws	✓	✓	✓	✓	✓
Without direct speech	✓	✓	✓	✓	✓
The verbs must be in the active voice	~	~	~	~	~
Without evaluative words	✓	✓	✓	✓	✓
Without technical/ academical words, foreign words, slang, jargon, double negatives	✓	x (technical words)	x (technical words)	✓	x (technical words)
Without filler words	~	~ (excessive repetition)	✓	✓	~ (excessive repetition)
Using colloquial tone	✓	✓	✓	✓	✓
Rounding off figures	x	x	x	x	x
Replacing the date per day	~	✓	x (kept both)	✓	x (kept both)

**Fulfilment of criteria: ✓—yes; X—no; ~—partially.*
Source: own processing, 2025.

Table 3.
Ability to rewrite the news piece from news agencies into radio report according to the criteria of Slovak "radio language" by the five most used generative AI tools in Slovakia.

4.2 The use of voice-based AI and its reception by listeners

Voice-based artificial intelligence is revolutionising the landscape of radio broadcasting, enhancing the way content is delivered and consumed. Advances in natural language processing (NLP) and deep learning have enhanced the fluency, and realism of AI-generated speech, raising questions about listener reception, trust, and usability [28].

Trust is crucial in radio broadcasting, as the credibility of the voice delivering information can significantly affect listener perception. Liu [29] argues that perceived credibility depends on the level of agency assigned to AI voices—whether they function autonomously or are controlled by humans. Transparent AI voice systems that clarify their artificial nature tend to foster more trust than those that obscure it. Becker et al. [30] suggest that the naturalness of an AI-generated voice plays a pivotal role in compliance and trust. Their findings indicate that more human-like voices encourage users to engage longer and perceive the system as more reliable. However, Patel et al. [31] warn that highly realistic AI voices may cause ethical dilemmas, as users may struggle to differentiate AI from human speech, potentially leading to misinformation or manipulation. Moreover, accent and pronunciation influence trust in AI voices. Pycha and Zellou [32] found that listeners rated AI voices with standard accents as more credible than those with regional or foreign accents. These findings suggest an inherent bias in listener perceptions, reinforcing the need for more inclusive and culturally diverse AI voice models.

Emotional resonance remains a challenge for AI-generated voices. While AI systems can generate speech with high accuracy, conveying emotion and nuanced intonation remain difficult. Campanilea et al. [33] found that emotionally expressive AI voices significantly improve listener engagement compared to monotonous voices. Their study emphasised that incorporating pitch modulation, rhythm variation, and prosody enhances the perceived naturalness of AI speech. According to Wang's [34] study, AI anchors can only convey the literal meaning of the text in the transmitted handwriting, and cannot convey the deeper meaning of the language and the text, particularly in terms of flexibility in tone, pitch, and pauses. The mechanical nature of AI presenters can compromise the effectiveness of communication with the audiences, especially in large live broadcasts. These presenters may lack the necessary communication skills and in-depth thinking to engage and interact with listeners, especially during large live broadcasts.

Despite these challenges, Slovak radio Europa 2 introduced a robotic news presenter "Eva" in 2019 to deliver the news block on weekdays at 4 pm. The response from listeners was largely unfavourable, with feedback such as "It lacked a human touch", "It failed to evoke emotion", and "It sounded like a text-to-speech conversion via Google Translate". In the Czech Republic, they created a position for artificial intelligence in the role of a presenter. This was done at Rádio Express FM, which is owned by Seznam.cz. They are the first in the Czech Republic to use an AI-generated synthetic voice. It comments on current events, but also on songs played on the air, and works during the night broadcast. The AI presenter was named "Hacsiko", and the original voice was provided by the presenter of the morning show Morning Club, Bára Hacsi. Similar to the Slovak audiences, the Czech listeners reacted rather negatively to the AI presenter: "I do not like Hacsiko", "There's a certain artificiality", "I've heard it, I have not warmed up to it yet", "The voice is the same, but the sentences are always in the same intonation, […] boring". One comment said that he would not have known the difference between human and AI broadcasting without prior warning, and he only saw the difference in the machine not being able to make mistakes and typos, which are "the spice of broadcasting" [11]. On the other hand, Kim et al. [35] investigated individuals' perceptions and reactions to a weather broadcast presented by an AI newscaster compared to a human newscaster. The results showed that while people perceived a human newscaster as more credible than an AI newscaster, their reactions to the news content did not differ.

Listeners are not emotionally cold to AI-generated voices, as confirmed by our ongoing research into listener perceptions of radio programmes. During the research, participants were told that some of the voices they heard were generated by artificial intelligence. Based on electrodermal activity (EDA) data and emotion analysis, we can observe significant changes in participants' emotional responses, suggesting that this disclosure had an emotional impact. **Figure 1** shows a peak in electrodermal activity (GSR peaks) around the 10th second, which correlates with the moment

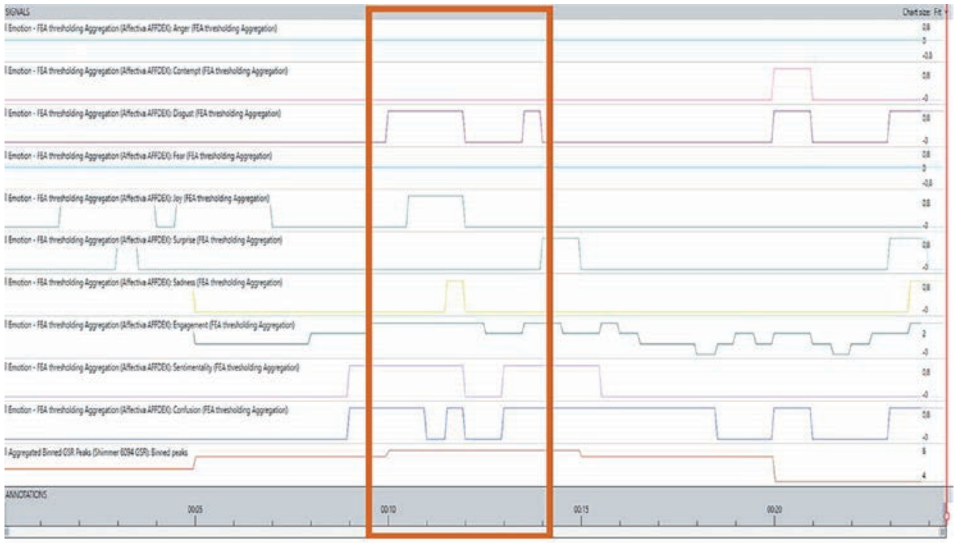


Figure 1.
Participant reaction on revealing information about the use of AI-generated voices during research. Source: own processing from unpublished research, 2025.

when participants received the information about the use of AI, which may be related to surprise or excitement.

At the same time, an increase in emotions such as disgust and sadness is evident, suggesting that some participants reacted negatively to this information. They may have been surprised that the voices they perceived as authentic were not real. However, there was also a slight increase in levels of joy and surprise, suggesting that some participants may have found the situation amusing, possibly due to suspicion that the voices were not human, or were surprised by the quality of the synthesised voices. Engagement remained at a relatively high level throughout the research, suggesting that participants remained engaged with the content and situation until the end of the test. Despite the relatively long duration (35 minutes on average), we consider this to be a positive aspect of conducting this type of survey. This may also be of interest for future research that focuses on people's reactions to synthesised voices versus real voices.

With the growing prevalence of AI-generated voices, ethical concerns surrounding their use have become more pressing. Mirek-Rogowska et al. [36] discuss the risks associated with AI-generated misinformation, emphasising that AI voices can be manipulated to spread false narratives. This ability to manipulate AI voices for deceptive purposes poses risks for credibility in media. Broadcasters must navigate these challenges by implementing robust ethical guidelines to ensure responsible use of AI technology, thus safeguarding their audiences from potential manipulation.

4.3 The ethical aspects of using AI in radio broadcasting

The ethical aspects of the use of AI are a major topic; in the visual domain, we encounter numerous examples of people being abused or manipulated. Similar situations are not avoided in the radio industry either. For example, the Polish radio station OFF Radio Kraków broadcast an interview conducted by an AI-generated presenter who, with the help of an artificially created voice, pretended to interview the Polish poet and Nobel Prize winner Wislawa Szymborska, who died in 2012. The radio planned to broadcast a similar interview with Polish statesman Józef Piłsudski, who died in 1935 [37]. The radio had previously fired a number of presumably redundant editors, which resulted in such mistakes, for example hallucination of AI. This may be a clear signal that manpower and skills are still needed and irreplaceable. Not only for their knowledge, but also for their ethical compass, and responsible approach to journalistic values.

While artificial intelligence (AI) has the potential benefits for human beings, it also gives rise to a number of ethical considerations that must be given due attention. As these systems are increasingly deployed across a range of sectors, AI has the potential to exert a significant impact on a number of key areas, including credit, employment, education, and competition [38]. However, without ethical considerations integrated into the design of AI algorithms, it is challenging to ensure that AI will not facilitate actions by certain actors that result in more harm than good. With the broader implementation of AI in recent years, it has been employed to undermine public trust in information, and has been implicated in perpetuating discrimination in the delivery of services and in creating unfavourable profiles of segments of the population, raising numerous other ethical concerns [39].

The absence of a standardised approach to the utilisation of AI in newsrooms has facilitated the advent of innovative and adaptable initiatives. While each news organisation has pursued its own projects, the absence of a universally accepted set of standards has resulted in the development of disparate ethical frameworks, which could result in potentially creating new challenges [40]. The publicly available generative AI tools have pre-set ethical limits; thus, they do not generate absolutely everything for the user, and do not process some inputs (prompts). Concurrently, it should be noted that it is very important how the prompts are entered. In general, however, tools that use large language model (LLM) should not provide content that promotes racism, xenophobia, anti-semitism, homophobia, sexism, or any other form of discrimination, or content that is outright illegal. Those can be topics related to the realm of crime—drugs, guns, murder, child pornography [39]. In this case, however, AI must not only know the realities of the area in which the text or information is processed, but also the legislative definition of specific areas. AI should be designed and used in a way that respects human dignity and privacy, and is not a tool to discriminate or harm individuals. Algorithms should be unbiased, and data should be managed with an eye to the presence of biases that could lead to unfair outcomes [41].

The ethical aspects of radio journalism are summarised in the internal codes of ethics of individual radio stations. This includes, for example, the formulations for citing a source, or obtaining information from multiple information relevant sources. This is an important aspect of the use of AI, that no source is mentioned when generating texts, the same applies to graphics. The ethical aspects of working with AI are also related to the legal framework, the cases mentioned above are related to copyright law, and it is debatable whether content created artificially can be considered copyrightable at all. However, there are also other cases that need to be dealt with—cloning of existing people and their voice, reviving the voice of the dead, but also plagiarism in a school environment [9]. Šantavý [42] says in the context of growing confidence in AI systems, we can and must set conditions without which the deployment of AI systems in the real world would not be possible. In general, AI systems must be: legal—comply with the required standards, laws, and regulations;

ethical—meet the required ethical criteria; safe—achieve the required standards of security and robustness. The integration of AI in journalism necessitates a multifaceted approach. It requires the adaptation of existing ethical frameworks, the establishment of new standards tailored to journalism, the enhancement of AI literacy, the maintenance of editorial control, the assurance of data governance, the addressing of biases, and the fostering of industry-wide collaboration. Through these measures, the journalism industry can effectively harness the benefits of AI while upholding its integrity and maintaining public trust [12]. This was summed by Kim [40] and is displayed in **Table 4**.

The ethical compliance of AI journalism, as represented by the balance among transparent practices, bias mitigation, and the influence of AI, is a growing concern in academic and professional circles. The existing literature emphasises the significance of transparency in the context of AI-driven news processes, underscoring the

Category	Subcategory	Description
Transparency and ethical standards	Public transparency	Release as many details as possible, ensure they do not jeopardise privacy or competitive advantages.
	Bias analysis	Regularly analyse models for bias.
		Standardise questions for reporters and editors to identify bias.
		Involve diverse groups of people (e.g., race, gender, income level, job positions).
	Editorial and ethical standards	Clearly communicate expectations between engineers and journalists about how algorithms are built.
		Emphasise the importance of ethical standards, especially for automated content generation.
Story discovery	Training data maintenance	Regularly evaluate the data used to train algorithms.
		Assign someone to check for outdated data based on the project duration.
	Evaluation of algorithm outputs	Create a process to fact-check and critically assess outputs generated by algorithms.
Story production	Context review	Regularly check if algorithms accurately reflect the context.
		Temporarily deactivate algorithms if they produce out-of-context content and require improvement.
Story distribution	Third-party data awareness	Understand what data third parties collect when using their algorithms or tools.
		Evaluate if the benefits of using such tools outweigh the risks of sharing private or competitive data.
	Reader awareness of algorithm use	Ensure readers understand where and how algorithms were used in the reporting or idea generation process.
		Attribute stories generated by algorithms appropriately and include details about their creation and functionality.

Source: Own processing, 2025 according to Ref. [40].

Table 4.
Steps in the context of using AI in media.

necessity for transparency and clarity in the manner by which news is algorithmically curated. Another notable challenge pertains to the potential for algorithmic bias. The possibility that AI systems may perpetuate biases, if left unchecked, has the potential to compromise the integrity of journalism. As the role of AI in newsrooms continues to expand, scholars have highlighted the importance of maintaining editorial independence and safeguarding journalistic ethics in the context of rapidly advancing technology [12].

In this context there are listed several core principles for ethical AI use in media, in which belong:

- *Ethical foundations* which link AI guidelines to core journalism values like trustworthiness and accuracy.

- *Transparency* which means to clearly communicate AI usage, specifying its role in content production.

- Disclose AI contributions in articles or content pieces to build *public trust*.

- *Human oversight* where we can assign clear responsibility for monitoring AI systems to ensure accuracy.

- *Defined boundaries* with specified acceptable and unacceptable uses of AI aligned with organisational values.

- *Data protection* and *algorithmic fairness*, where address AI biases by leveraging locally relevant systems and robust oversight.

- *Education and engagement* in which trained staffs help audiences understand AI.

- *Collaboration* where we should foster internal and external partnerships to ensure ethical AI use.

- *Dependency management* means to clarify and plan for reliance on third-party systems, exploring custom solutions.

- *Enforceability* is about establishing mechanisms and consequences for guideline violations.

- *Media diversity* stands for preserve distinct editorial voices and prevents homogenisation [43].

Brník et al. [44] claim that the work processes in editorial rooms have changed significantly. These changes include the shortening of reports, the incorporation of infotainment elements into broadcasting, the diversification and adherence to journalistic genre standards, and shifts in news preparation and fieldwork processes, such as transitioning between editorial rooms. For instance, a presenter who previously served as an announcer is now tasked with producing news, recording material, and occasionally creating new genres, such as assemblages or reportage. AI can effectively assist with all these tasks.

The integration of AI in journalism necessitates a multifaceted approach. It requires the adaptation of existing ethical frameworks, the establishment of new standards tailored to journalism, the enhancement of AI literacy, the maintenance of editorial control, the assurance of data governance, the addressing of biases, and the fostering of industry-wide collaboration. Through these measures, the journalism industry can effectively harness the benefits of AI while upholding its integrity and maintaining public trust [12].

The only thing that remains is to wait and see how the radio industry is able to adapt to the emerging needs of ethical frameworks in the context of AI, and also whether this powerful tool can influence the future of human potential in the broadcasting environment. Listeners, who are the pillars of listenership and preference, will have the last word anyway; if they prefer humanity and intercession, artificial intelligence will not be a competitor for presenters and editors, but if listenership perception slips to the level of adjacency and below-average attention is paid to the broadcast, artificial intelligence will be abundantly sufficient.

4.4 Description of the current situation in the Slovak radio industry in relation to the use of AI in practice

The Slovak radio industry is currently undergoing a technological transformation, with artificial intelligence at the core of these changes. Although traditional radio broadcasting still dominates, a number of Slovak radio stations have started to introduce AI tools into various areas of the broadcasting process to improve their operations. These range from browsing online articles and recommending personalised content, to automated news reporting, or broadcasting content through AI-generated voices.

As mentioned above, AI was first used in Slovak radio broadcasting in December 2019, when radio Europa 2 introduced a female news presenter, "Eva", who reported the news every weekday at 4 pm. However, it should be remembered that AI technology and the tools were not as well developed before now, and as a result, reactions to the AI news presenter were also largely negative, and the broadcaster discontinued the project fairly quickly. Since the launch of ChatGPT in late 2022, many radio stations around the world have started to implement AI tools in various areas, from playlist creation to AI DJs, or news anchors.

Slovak public service and private broadcasters have also recognised the potential of AI. One of the first broadcasters to use an AI DJ was the private station Rádio Expres. In early 2024, they introduced a night-time DJ called "Robo". According to Marian Staráček (Innovation Specialist at Bauer Media Slovakia, which owns Rádio Express, Rádio ROCK, Rádio Melody, and Europa 2), the company's goal was not to create a perfect copy of a human being, but only to fill the night time, which would not be hosted [45]. In addition to the AI DJ, the company is currently using AI voices to commercials. The second most listened to private station—Fun rádio—has decided to use AI in a different way. As part of a temporary image campaign, they trained AI DJs (based on the voices of real DJs) to create personalised audio messages for listeners. Roman Janajev (Creative Tech Lead of the company that worked with Fun rádio on this project) said, "The AI DJs' comments are generated using GPT-4 with fine-tuned prompts that include the personality traits and presentation style of specific Fun rádio DJs. The result of the generated prompts is further transferred to ElevenLabs with the rehearsed DJ voices. Their model also preserves the specifics of each DJ's diction" [46]. The programme director of Fun rádio, Marek Mikúšek, added that in general they are using AI very cautiously so far, but that they are using it mainly for tasks related to information retrieval, preparing creative texts for broadcasting, advertising texts, or creating various sound elements such as jingles or podcasts [47].

As for public service Slovak radio (as an organisational component of Slovak Television and Radio, which consists of nine stations), the first radio to include AI in its broadcasts was Radio Slovakia International (RSI). RSI broadcasts information about Slovakia in six languages (English, French, German, Russian, Spanish, and Slovak) for listeners around the world and for foreigners living in Slovakia. The newsroom has started using AI to dub interviewees into French. Kristína Hanáková (French editor) explained that in the past they used to get help from people in the newsroom or external collaborators, which was difficult to organise. However, "Mathilde and François (AI voices) are immediately available and their performance in French is hardly distinguishable from a human voice" [48]. The most listened public service radio station—Rádio Slovensko—took a different approach. It invited its listeners to send 10 words to the editing room, which were then inserted into ChatGPT to create an original 2-minute short story. These are then narrated by Slovak actors and the selected short stories are broadcast on the radio as part of the morning prime-time show [49].

In addition to the radio stations with national reach, several regional radio stations have also started to use AI tools. Trnavské rádio introduced an AI editor "Ivana Šnajder" in early 2024, whose role is to cover regional content that is no longer within the capacity of human resources. According to the newsroom, this works in practice as follows: the field editor collects material and prepares a report, which the AI then fine-tunes textually, and prepares as an audio output for the newsroom [50]. Dobré rádio, on the other hand, has included a segment in which they "take ChatGPT on air". According to DJ Michaela Kicková, the way it works is that they ask ChatGPT a question and it then creates an answer through its AI voice, which they broadcast. At the same time, they also use it in music, for example, asking it to sing a song by Ema Drobná (a Slovak singer) in the voice of Michael Jackson [47].

On the other hand, not all Slovak radio stations are currently using AI in broadcasting. One example is Viva radio. DJ Dávid Schun said that they had tried it in their editorial room, but the generated texts were inaccurate, and it was necessary to check the facts again. He added that he could see AI being very useful in the pre-production phase of radio in the future, but only if it provided 100% accurate information [47].

Despite these advances, the introduction of AI in Slovak radio is still at an early stage compared to global trends. Ethical concerns about authenticity, displacement of jobs, and credibility of content remain topics of debate in the industry. However, as AI technology develops and regulatory frameworks evolve, its role in shaping the future of Slovak radio broadcasting is expected to grow.

5. Discussion and conclusion

The use of artificial intelligence in radio broadcasting has undergone a significant transformation, especially in the recent years. AI now enables more efficient information processing, personalisation of content, and automation of routine tasks, leading to increased efficiency and reduced costs in radio production. However, in the Slovak context, there is still caution about the broader implementation of AI in the radio environment. Part of this may be due to the fact that AI tools are not currently able to work in Slovak at the level required by the "Slovak radio language" criteria. Another reason may be that listeners perceive AI voices as unnatural and "lacking a human touch". The future of AI in Slovak radio will therefore depend on the ability to combine modern technology with maintaining the credibility and authenticity of the broadcast.

We expect AI to continue to improve and gradually permeate all aspects of radio production—from news generation and advertising, to interactive voice interfaces, and personalised AI presenters. Developments in neural voice synthesisers suggest that in the near future AI presenters may be almost indistinguishable from human ones. However, this trend raises the question of the extent to which listeners will prefer human interaction to artificial intelligence.

Another important factor is the regulation and ethical framework for the use of AI in the media. Given the potential for content manipulation and the spread of disinformation, it will be essential to establish transparent rules for the use of AI in broadcasting. At the same time, a balance will need to be struck between the efficiencies that AI brings, and human creativity and journalistic ethics.

The future of radio will therefore depend not only on technological advances, but also on the ability to adapt to new challenges, and listener preferences. AI has the potential to significantly innovate the radio environment, but at the same time it is important not to lose the unique atmosphere and authenticity that traditional radio offers—as mentioned above—"the machines are not able to make the mistakes and typos that are the spice of live broadcasting". The key to success will therefore be to find the optimal symbiosis between artificial intelligence and the human factor to ensure that radio remains a relevant and trusted medium in the era of digital transformation.

Acknowledgements

This chapter was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V05-00004.

Author details

Lucia Furtáková*, Ľubica Janáčková and Andrej Brník
University of Ss. Cyril and Methodius in Trnava, Trnava, Slovakia

*Address all correspondence to: furtakova1@ucm.sk

References

[1] RCS. We are RCS [Internet]. Available from: https://www.rcsworks.com/company/about-us/ [Accessed: January 27, 2025]

[2] Jacobs F. Radio's most innovative: Andy economos' selector [Internet]. 2014. Available from: https://jacobsmedia.com/radios-most-innovative-dr-andrew-economos/ [Accessed: January 27, 2025]

[3] MacArthur A, Laporte L. NET@NIGHT 6: Tim Westergren of Pandora Media [Internet]. 2006. Available from: https://twit.tv/shows/netnight/episodes/6 [Accessed: January 27, 2025]

[4] Roach J. Non-human DJ gets radio gig [Internet]. 2011. Available from: https://www.nbcnews.com/news/world/non-human-dj-gets-radio-gig-flna121286 [Accessed: January 27, 2025]

[5] O'Brien K. iHeartMedia adds smartaudio digital data Ad product to its programmatic offerings [Internet]. 2017. Available from: https://www.thedrum.com/news/2017/03/16/iheartmedia-adds-smartaudio-digital-data-ad-product-its-programmatic-offerings [Accessed: January 27, 2025]

[6] BBC. OK Beeb: BBC voice assistant will learn regional accents [Internet]. Available from: https://www.bbc.com/news/technology-52891155 [Accessed: 2025 January 27]

[7] Futuri. Futuri launches RadioGPT™, the world's first AI-driven localized radio content [Internet]. 2023. Available from: https://futurimedia.com/futuri-launches-radiogp/ [Accessed: January 27, 2025]

[8] Beckett C, Yaseen M. Generating change: A global survey of what news organisations are doing with AI [Internet]. 2023. Available from: https://static1.squarespace.com/static/64d60527c01ae7106f2646e9/t/656e400a1c23e22da0681e46/1701724190867/Generating+Change+_+The+Journalism+AI+report+_+English.pdf [Accessed: January 10, 2025]

[9] Kopecký K, Szotkowski R, Mackenzie K. Rizika spojena s generativní umělou inteligencí. Olomouc: Faculty of Education at Palacký University Olomouc; 2024. 98 p

[10] Kumari M, Prakash R. To study the role of artificial intelligence technology in community radio. Indian Journal of Legal Review (IJLR). 2024;**4**(1):1278-1289

[11] Furtáková L, Janáčková Ľ. AI in radio: The game changer you did not hear coming. In: Prostináková Hossová M, Martovič M, Solík M, editors. Marketing Identity: AI—The Future of Today. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius; 2023. pp. 95-106. DOI: 10.34135/mmidentity-2023-09

[12] Amponsah P, Atianashie A. Navigating the new frontier: A comprehensive review of AI in journalism. Advances in Journalism and Communication. 2024;**12**(1):1-17. DOI: 10.4236/ajc.2024.121001

[13] Furtáková L, Bôtošová Ľ, Vrabec N. Radio reportage: The aspects that shape its style. In: Prostináková Hossová M, Labudová L, Martovič M, editors. Marketing & media identity 2022: Metaverse je nový vesmír. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius; 2022. pp. 14-28

[14] Bullard G. Smart ways journalists can exploit artificial intelligence [Internet]. 2023. Available from: https://niemanreports.org/articles/artificial-intelligence-newsrooms/#:~:text=AI%20is%20helping%20newsrooms%20reach,more%20likely%20to%20click%20on [Accessed: January 10, 2025]

[15] Alwadi N. Case study: AI and algorithmic journalism in newsroom [Internet]. 2023. Available from: https://www.academia.edu/117800398/Case_Study_AI_and_Algorithmic_Journalism_in_Newsroom [Accessed: January 10, 2025]

[16] Bôtošová Ľ, Furtáková L. Rozhlasová reportáž. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius; 2023. 131 p

[17] Jenča I. Rozhlasové spravodajstvo. Ružomberok: Verbum; 2018. 274 p

[18] Furtáková L, Janáčková Ľ. News at the speed of AI: Automating journalism through text generator. In: Prostináková Hossová M, Solík M, Martovič M, editors. Marketing Identity: Human vs. Artificial. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius; 2024. pp. 166-174. DOI: 10.34135/mmidentity-2024-16

[19] Krajčovič P, Darázs T. Neuromarketing—A new possibility in marketing research at FMK UCM in Trnava. Communication Today. 2021;**12**(2):196-197

[20] Macková V, Mařík R. A robotic reporter still lacks creativity, but it can already replace human journalists in several areas. Acta Fakulty filozofické Západočeské univerzity v Plzni/West Bohemian Review of Social Sciences & Humanities. 2023;**15**(1-2):23-37. DOI: 10.24132/actaff.2023.15.1-2.3

[21] Gondwe G. Exploring the multifaceted nature of generative AI in journalism studies: A typology of scholarly definitions. SSRN. 2023:1-11. DOI: 10.2139/ssrn.4465446

[22] Alawida M, Merji S, Mehmood A, Chikhaoui B, Abiodun OI. A comprehensive study of ChatGPT: Advancements, limitations, and ethical considerations in natural language processing and cybersecurity. Information. 2023;**14**(8):462. DOI: 10.3390/info14080462

[23] Bode B. Künstliche Intelligenz verkündet Tod von Fußballstar Erling Haaland [Internet]. 2023. Available from: https://www.basicthinking.de/blog/2023/09/29/kuenstliche-intelligenz-erschiesst-fussballstar-erling-haaland/ [Accessed: 2025 January 20]

[24] Furtáková L. Can ChatGPT replace a radio news anchor? In: Francistyová B, Furtáková L, editors. Quo Vadis 2023 <Generated by AI>. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius; 2023. pp. 11-29

[25] Tinák T, Gáliková TS. Media metamorphosis in the digital age: Interaction with artificial intelligence and its consequences in journalism. In: Prostináková Hossová M, Solík M, Martovič M, editors. Marketing Identity: Human vs. Artificial. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius; 2024. pp. 191-200. DOI: 10.34135/mmidentity-2024-19

[26] Mikušová S. AI in Slovak journalism: A threat or an opportunity? In: Prostináková Hossová M, Solík M, Martovič M, editors. Marketing Identity: Human vs. Artificial. Trnava: Faculty of Mass Media Communication, University of Ss. Cyril and Methodius;

2024. pp. 487-497. DOI: 10.34135/mmidentity-2024-50

[27] IVO. Skúsenosti s umelou inteligenciou má na Slovensku každý piaty človek [Internet]. 2024. Available from: https://www.ivo.sk/8938/sk/aktuality/skusenosti-s-umelou-inteligenciou-ma-na-slovensku-kazdy-piaty-clovek [Accessed: January 21, 2025]

[28] Hancock JT, Naaman M, Levy K. AI-mediated communication: Definition, research agenda, and ethical considerations. Journal of Computer-Mediated Communication. 2020;**25**(1):89-100. DOI: 10.1093/jcmc/zmz022

[29] Liu B. In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. Journal of Computer-Mediated Communication. 2021;**26**(6):384-402. DOI: 10.1093/jcmc/zmab013

[30] Becker D, Braach L, Clasmeier L, Kaufmann T, Ong O, Ahrens K, et al. Influence of robots' voice naturalness on trust and compliance. ACM Transactions on Human-Robot Interaction. 2024;**14**(2):1-25. DOI: 10.1145/3706066

[31] Patel R, Price N, Bahr R, Bedrick S, Bensoussan Y, Bélisle-Pipon J-C, et al. The Bridge2AI-voice consortium, Powell M. Summary of keynote speeches from the 2024 voice AI symposium, presented by the Bridge2AI-voice consortium. Frontiers in Digital Health. 2024;**6**:1484503. DOI: 10.3389/fdgth.2024.1484503

[32] Pycha A, Zellou G. The influence of accent and device usage on perceived credibility during interactions with voice-AI assistants. Frontiers in Computer Science. 2024;**6**:1411414. DOI: 10.3389/fcomp.2024.1411414

[33] Campanilea L, De Fazio R, Di Giovanni M, Marrone S, Marulli F, Verde L. Inferring emotional models from human-machine speech interactions. Procedia Computer Science. 2023;**225**:1241-1250. DOI: 10.1016/j.procs.2023.10.112

[34] Wang X. Research on the training model of broadcasting and hosting talents under the background of AI anchors. Academic Journal of Humanities & Social Sciences. 2021;**4**(5):74-77. DOI: 10.25236/AJHSS.2021.040513

[35] Kim J, Xu K, Merrill K. Man vs. machine: Human responses to an AI newscaster and the role of social presence. The Social Science Journal. 2022:1-13. DOI: 10.1080/03623319.2022.2027163

[36] Mirek-Rogowska A, Kucza W, Gajdka K. AI in communication: Theoretical perspectives, ethical implications, and emerging competencies. Communication Today. 2024;**15**(2):24-38. DOI: 10.34135/communicationtoday.2024.Vol.15.No.2.2

[37] Uličná K. V poľskom rádiu nahradila umelá inteligencia moderátorov. AI urobila rozhovor s mŕtvou umelkyňou [Internet]. 2024. Available from: https://refresher.sk/169616-V-polskom-radiu-nahradila-umela-inteligencia-moderatorov-AI-urobila-rozhovor-s-mrtvou-umelkynou [Accessed: January 23, 2025]

[38] Višňovský J, Francistyová B, Dúbravská O. Artificial intelligence - risks and opportunities in higher education in media and communication studies. In: Sałaty E, Ziębakowska-Cecot K, editors. Aktualne wyzwania współczesnej edukacji. Radom: Uniwersytet Radomski im. Kazimierza Pulaskiego; 2023. pp. 71-80

[39] USAID. Artificial intelligence (AI) ethics guide [Internet]. 2023. Available from: https://www.usaid.gov/sites/default/files/2023-12/_USAID%20AI%20Ethics%20Guide_1.pdf [Accessed: January 23, 2025]

[40] Kim H. AI in journalism: Creating an ethical framework AI in journalism: Creating an ethical framework. 2019. Available from: https://surface.syr.edu/cgi/viewcontent.cgi?article=2095&context=honors_capstone [Accessed: January 23, 2025]

[41] Kopecký K, Szotkowski R. Umělá inteligence—Rizika a odpovědnost. Olomouc: Faculty of Education at Palacký University Olomouc; 2024. 67 p

[42] Šantavý P. Umelá inteligencia—dobrý sluha a zlý pán? Bratislava: Rímskokatolícka cyrilometodská bohoslovecká fakulta Univerzity Komenského v Bratislave; 2023. 309 p

[43] Guidelines for using generative AI media organisations [Internet]. Available from: https://www.mediamonitoringafrica.org/wp-content/uploads/2024/04/AI_Media_Guidelines.pdf [Accessed: January 23, 2025]

[44] Brník A, Bôtošová Ľ, Kapec M. Changes in broadcasting to keep a listener in commercial radios in Slovakia. In: CER Comparative European Research 2021. London: Sciemcee Publishing; 2021. pp. 21-25

[45] Rádio Expres odohrá nonstop 2024 hitov, v noci bude hitparádu moderovať umelá inteligencia [Internet]. 2024. Available from: https://www.radia.sk/spravy/4877_radio-expres-odohra-nonstop-2024-hitov-v-noci-bude-hitparadu-moderovat-umela-inteligencia [Accessed: January 26, 2025]

[46] Fun rádio prináša AI moderátorov Sajfu, Ajku a Juniora s Marcelom [Internet]. 2024. Available from: https://www.radia.sk/spravy/4967_fun-radio-prinasa-ai-moderatorov-sajfu-ajku-a-juniora-s-marcelom [Accessed: January 26, 2025]

[47] Kufová M. Osobnosť moderátora v Rozhlasovom vysielaní [unpublished thesis]. Trnava: University of Ss. Cyril and Methodius; 2025

[48] Umelá inteligencia je už aj vo vysielaní Slovenského rozhlasu. Ako prvé ju využíva Radio Slovakia International [Internet]. 2024. Available from: https://www.radia.sk/spravy/4983_umela-inteligencia-je-uz-aj-vo-vysielani-slovenskeho-rozhlasu-ako-prve-ju-vyuziva-radio-slovakia-international [Accessed: January 26, 2025]

[49] Žofčin M. Umelá inteligencia napísala poviedky pre Slovenský rozhlas. Vypočujte si jednu z nich [Internet]. 2024. Available from: https://zive.aktuality.sk/clanok/0xTNQzv/umela-inteligencia-napisala-poviedky-pre-slovensky-rozhlas-vypocujte-si-jednu-z-nich/ [Accessed: January 26, 2025]

[50] Umelá inteligencia pomáha v rádiu. Prichádzajú "lietajúci" AI reporter [Internet]. 2024. Available from: https://trnavske.radio/2024/02/18/umela-inteligencia-pomaha-v-radiu-prichadzaju-lietajuci-ai-reporteri/ [Accessed: January 26, 2025]

Chapter 6

An X Study of the Evolution of COVID-19-Related Sentiments in the UK

Marco A. Palomino, Rohan Allen and Aditya Padmanabhan Varma

Abstract

An outbreak of SARS-CoV-2 caused the World Health Organisation (WHO) to declare a public health emergency of international concern on 30 January 2020. As the emergency escalated, the WHO declared it a global pandemic on 11 March 2020, triggering a parallel outbreak of fear and depression throughout the world, which negatively impacted the wellbeing of the public and healthcare workers alike. While helping to accelerate mental health diagnoses, we explored the use of sentiment analysis, a powerful tool for understanding opinions. We developed a machine learning classifier to detect depression, a common COVID-19-related mood disorder. To examine the shifting emotional landscape of the public discourse surrounding COVID-19, we studied two X—formerly known as Twitter—collections: one from 2020 and another one from 2022. We complemented our work with the utilisation of an off-the-shelf classifier and concluded that, over a span of two years—between 2020 and 2022—fear was the most dominant emotion attached to COVID-19 and depression the most dominant mood. Our practical insights can help to design strategic choices concerning the wellbeing of people worldwide.

Keywords: sentiment analysis, emotion recognition, X (formerly known as twitter), social media, machine learning

1. Introduction

In December 2019, an outbreak of pneumonia of unknown origin was reported in Wuhan, Hubei Province, China. The disease became known as *COVID-19* and unfolded rapidly, reaching all countries and causing a large number of deaths, which led the *World Health Organisation* (WHO) to declare a pandemic on 11 March 2020 [1]. As a result of this pandemic, adult men were at greater risk of developing severe illnesses and comorbidities, but various reports indicated that the groups whose mental health was most severely affected were women, young adults, and unemployed people [2–5]. Regrettably, these groups also developed physiological and behavioural symptoms associated with distress more frequently than others, because of an increase in their heart rate, sleeping disorders [5], and depression [6]. In the case of the UK, which is where we carried out our investigation, suicidal ideation grew considerably after the *first wave* of COVID-19 [5]—that is, between 31 March 2020 and 9 April 2020.

Symptoms of anxiety, defeat, and entrapment decreased across waves. However, women, young people between 18 and 29 years of age, socially disadvantaged and vulnerable groups, and those with pre-existing mental health conditions reported worse emotional wellness during the pandemic [5]. Generally, people observed a decrease in psychological wellbeing due to COVID-19, and higher levels of psychiatric symptoms were found among healthcare workers [7]. However, those seeking help experienced serious delays in their treatment, because of the overall shrinkage of resources.

To expedite mental health diagnoses, we became interested in developing assistive technologies. We delved deeper into the subjects of *emotion recognition* [8] and *sentiment analysis* [9], as they have become indispensable to mine people's opinions [10]. Words are reflective of our attitudes and emotions, and can also explain how we think [11, 12]. Moreover, words can convey people's emotions, making them a useful tool for examining people's attitudes and sentiments towards an issue [13]. This is where our main contribution lies: analysing the language people use and the sentiments expressed in their words.

In April 2020, we launched an investigation about the sentiment expressed on social media with regards to COVID-19. We concentrated on *X*, then known as *Twitter* [14], the micro-blogging platform. X users publish short messages—so-called "tweets"—to report their thoughts and actions, comment on breaking news, and engage in discussions. Tweets comprise reviews, praise, and criticism [15]. X users interact with one another, with groups, and with the public at large. When conversations emerge, they are frequently witnessed by an audience beyond the immediate participants. X is an inexhaustible stream of data to uncover emotions and experiment with sentiment analysis [16].

To assess how the sentiment about COVID-19 evolved over time, we gathered a second collection of tweets in March 2022, precisely when the second anniversary of the announcement of the start of the UK's first nationwide lockdown took place [17]. We then processed both collections to extract insights into the feelings and emotions expressed by X users. These allowed us to address the following research questions, which will be discussed in detail in this chapter:

RQ1 (Prevailing sentiment): What was the prevailing sentiment expressed by the public during the COVID-19 pandemic in the UK? While the measures exercised by the Government—including lockdowns and vaccination campaigns—were eventually effective in curbing the virus's spread, they triggered intense debates on social media, which were often exacerbated by fake news and fuelled negative emotions [18]. We were interested in confirming how negative such emotions were and how long they prevailed.

RQ2 (Most prominent emotion): What was the most prominent emotion associated with COVID-19 in the UK? We were not only interested in determining the polarity—whether positive, neutral, or negative—of the sentiments around the pandemic, but we were also keen on discovering a wider range of emotions expressed by the public—anger, fear, sadness, etc.—and which one was the most prominent.

RQ3 (Sentiment change and evolution): How did the public sentiment about COVID-19 change over time? We aimed to determine if the sentiment and emotions expressed about COVID-19 varied throughout the pandemic and how specifically they evolved in the UK.

The insights gained from our research can support strategic decision-making regarding the mental health of the population.

DOI: http://dx.doi.org/10.5772/intechopen.1008866

The reminder of our contribution is organised as follows: We summarise the related work on sentiment analysis in Section 2. Section 3 offers a snapshot of the two X collections used in our study. Section 4 presents our results, and Section 5 discusses such results. Finally, Section 6 states our conclusions.

2. Related work

Emotions are critical for the interaction of human beings, as they enable people to reveal their perceptions and express their reactions to *stimuli* [19]. The idea of employing computers to recognise emotions was introduced by Picard's seminal work in 1997 [20]. According to Picard, if we want computers to be genuinely intelligent and interact naturally with us, we must give them the ability to identify, understand, and even express emotions. Regrettably, it was only recently when the advances in machine learning boosted *natural language processing* (NLP) to reach human-level performance.

Whilst most of the earlier work on sentiment analysis focused on allocating a positive or negative rating—that is, a *polarity*—to a piece of free text [21], the latest body of literature on the subject aims to recognise a range of emotional categories. The recent availability of large language datasets has allowed scientists to improve their understanding of mental health issues through the study of words [12, 22]. Sentiment analysis has been widely used to analyse evaluations and attitudes towards products, services, and topics [9, 23]. In the past decade, recognising the emotions expressed in text has also earned significance as an alternative to assess people's wellbeing. One example of this has been published in Europe to address suicide prevention [24].

2.1 Emotion vs. mood

There is a difference between *emotion* and *mood*, despite the terms being often used interchangeably. As per the *Oxford English Dictionary*, an emotion is "any strong mental or instinctive feeling, as pleasure, grief, fear, etc., deriving especially from one's circumstances or relationship with others" [25]. An emotion is an intense feeling typically directed at a source. Conversely, a *mood* tends to be less intense than an emotion and does not necessarily need a contextual stimulus. The Oxford English Dictionary defines a mood as "a prevailing but temporary state of mind or feeling; a person's humour, temper, or disposition at a particular time" [26]. Emotions are brief—they last anywhere from seconds to minutes at most [27]—whereas moods last longer.

Emotional *categories*, which divide the emotions into discrete classes [28], and emotional *dimensions*, which represent the classes in a 2D or 3D space [29], have been used to classify emotions. Two of the most notable works on this field are *Ekman's* basic emotion model [27] and *Plutchik's* bipolar emotion model [30]. Ekman defined six universal emotions: anger, disgust, fear, joy, sadness, and surprise [27], whereas Plutchik considered eight pairwise, contrasting emotions: joy vs. sadness, trust vs. disgust, fear vs. anger, and surprise vs. anticipation [31].

To illustrate his ideas, Plutchik created a *wheel of emotions* in 1980 [31]. This model links the idea of an emotion circle and a colour wheel. Like colours, Plutchik's pairwise emotions can be expressed at different intensities and can mix with one another to form other emotions. An example of Plutchik's wheel of emotions is displayed in **Figure 1**. The emotions with no colour represent combinations of two primary

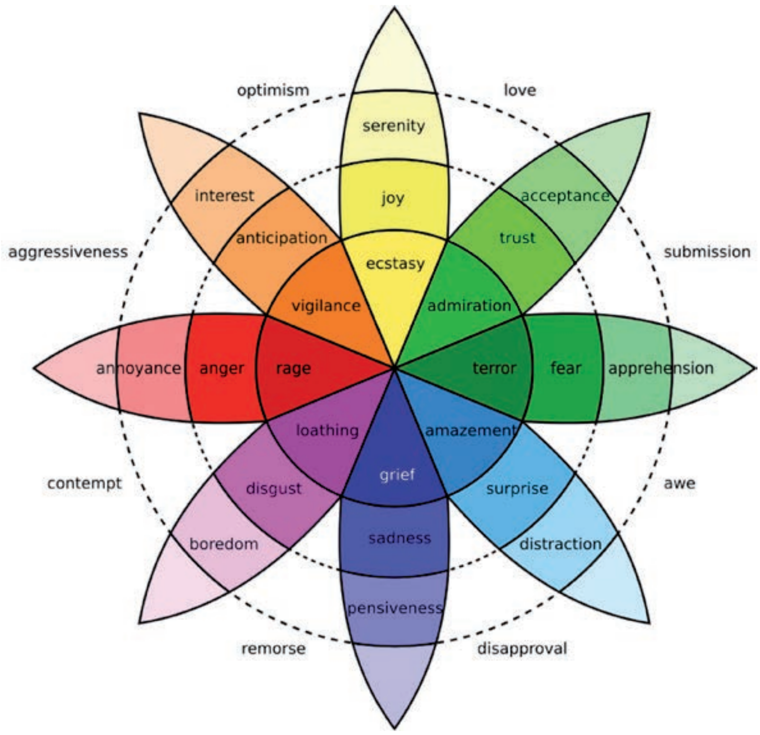


Figure 1.
Plutchik's Wheel of Emotions—Attribution: Machine Elf 1735, public domain, via Wikimedia Commons [32].

emotions. For instance, anticipation and joy combine to form optimism, and joy and trust combine to form love.

Despite the importance of Ekman's and Plutchik's work for sentiment analysis, we opted for a third option: a classifier capable of detecting depression. This is because depression is critical to our investigation, as it is a mood disorder frequently associated with COVID-19 [33, 34]. Hence, rather than Ekman's and Plutchik's model, we implemented the *Profile of Mood States* (POMS) [35], a psychological test for assessing an individual's mood [36], which was formulated by McNair *et al.* [37] and has lately been tested for NLP. We favoured POMS over other alternatives, and we will describe our work with POMS below.

2.2 Supervised machine learning

Most of the recent literature has adopted supervised machine learning to identify specific sets of emotions. For instance, Chaffar and Inkpen created a training set composed of emotion-annotated news headlines, fairy tales, and blogs [38]. Using such a set to compare a few classifiers, they reported a significant performance improvement when using a *support vector machine* (SVM) [39]. Furthermore, their SVM classifier demonstrated strong generalisation capabilities when tested on unseen examples [38].

While we recognise the significance of SVMs, we aimed to move beyond traditional approaches to determine the sentiment expressed in text, which rely on lexicons and bag-of-words models [40]. The results achieved by researchers who have utilised sequences of characters—bypassing pre-processing and feeding the sequences directly into a *recurrent neural network* (RNN)—have continued to improve over time and we wanted to test this option. Colnerič and Demšar [41] have used this approach, and we followed their example.

2.3 Public health

Social media platforms, such as X, are promising instruments to monitor and foster health policies among the population. In 2012, the WHO launched its global eHealth strategy to encourage the promotion, development, and evaluation of actions that involve social media [42]. The impetus for eHealth came from the increasing use of information and communication technologies in support of health services—both in developed and developing countries. It has been proven that social media can encourage citizen participation, be an interactive space for science dissemination, and promote healthy behaviours [43]. Moreover, the discovery of new trends and its appropriate monitoring may lead to the enhancement of early warning systems [43].

Social media has also helped to increase citizens' awareness while allowing people to take a more active and better-informed role in their communities [44]. For instance, the analysis of X content has shown a significant participation in discussions related to childhood obesity [45] and approaches to manage alcohol consumption [46]. Social media is likely to have a successful impact on vulnerable populations, low-income sectors, and minority ethnic groups [47].

3. Materials and methods

As of the beginning of 2022, the number of daily, active X users amounted to 237.80 million globally [48] and 500 million tweets were sent out each day [49], which equates to nearly 6000 tweets per second. Generally, *hashtags* allow X to group discussions about specific subjects. Hashtags include the sign '#' followed by a word or phrase that groups tweets around a particular theme [50]. Hashtags often signal aspects of a tweet's meaning or its intended audience. For example, the hashtag #NHSEngland ostensibly groups tweets related to the *National Health Service (NHS)*—the publicly funded healthcare system in England.

3.1 Experimental corpus

Daily press briefings highlighting the strategy to manage COVID-19 were held by the UK Government between 16 March 2020 and 23 June 2020. The briefings provided the British public with updates on statistics, news and safety measures put in place by the Government—at the time, COVID-19 had already been characterised as a pandemic by the WHO [51].

To address our research questions, we gathered 409,761 tweets about COVID-19 on Wednesday 22 April 2020, and we will refer to this corpus hereafter as the *2020 Corpus*. We selected this date because it was when Dominic Raab, the then UK Foreign Secretary, held a press briefing highlighting, for the first time, the use of vaccines as part of the Government's strategy. The British press began reporting on a COVID-19 vaccine in early April 2020, coinciding with the launch of the first human trials in Europe [52], which received substantial investment.

We anticipated that the Foreign Secretary's briefing on 22 April 2020 would ignite discussions. Therefore, this was an opportunity to capture tweets expressing strong sentiments, whether in the form of criticism of the Government or concerns about the ongoing situation—the full transcript of the Foreign Secretary's briefing is available at GOV.UK [53].

Given that the Foreign Secretary's briefing was announced at 16:30, we began collecting tweets a couple of hours beforehand and continued for a couple of hours afterwards. Specifically, the first tweet was captured at 14:24:39, and the last one at 18:56:27, covering a total of 4 hours, 31 minutes, and 48 seconds.

We retrieved the tweets using *Tweepy* [54], an open-source, Python library that simplifies real-time tweet collection by facilitating the interaction with the *X RESTful API*—including the *X Streaming API*—while managing authentication and connection [55, 56]. The gathering of tweets was carried out in Plymouth (UK), using *Colaboratory*, or *Colab* [57, 58], Google's environment for interactive development. To ensure our data captured COVID-19-related information, we only considered tweets that used specific hashtags. These hashtags are listed in **Table 1**, along with the number of tweets we found for each hashtag.

The values in **Table 1** do not sum up to the total number of tweets in the 2020 Corpus—which is 409,761—because many tweets in the corpus contain two or more of the hashtags listed in **Table 1**. Additionally, some of the tweets in the 2020 Corpus may not explicitly contain the hashtags listed in **Table 1**, but the Streaming API still captures them because these hashtags appear in the URLs or metadata attached to the tweets [56].

Figure 2 displays the number of tweets retrieved every 30 minutes. On average, the retrieval proceeded at a rate of 81,952 tweets per hour between 14:24 and 19:24, but the rate exceeded 90,000 tweets per hour during the first three hours. Notably, in the 10-minute window between 16:20 and 16:30—that is, right before the briefing began—16,043 tweets were retrieved.

According to the *term frequency–inverse document frequency* (TF-IDF) metric [59], the most characteristic nouns found in the experimental corpus were covid19, pandemic, coronavirus, people, stayhome and death. All the characteristic keywords discovered using TF-IDF are depicted in **Figure 3** using a word cloud—the higher the TF-IDF score, the larger the font of the keyword.

Hashtag	Number of tweets
#covid19	238,432
#coronavirus	116,557
#stayhome	31,820
#covid_19	11,068
#socialdistancing	6510
#covid-19	4636
#covid2019	2341
#flattenthecurve	2124
#coronavirusoutbreak	2058
#sarscov2	1861
#virus	1211

Table 1.
Hashtags used to retrieve the 2020 Corpus on 22 April 2020.

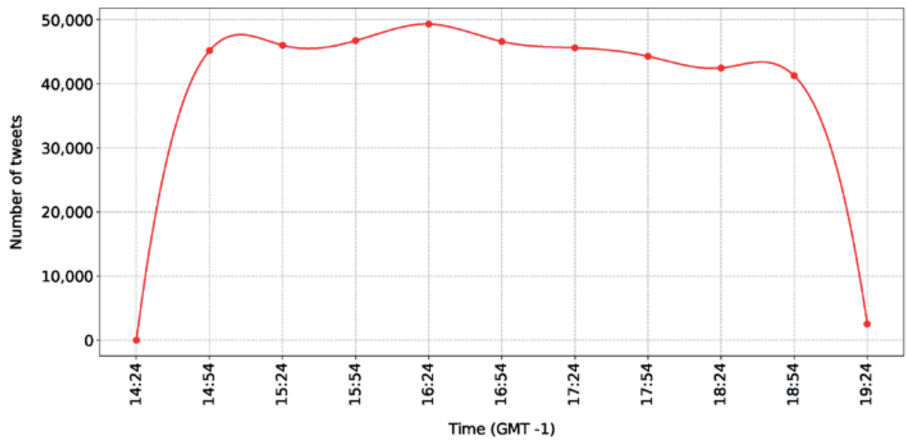


Figure 2.
Distribution of tweets in the 2020 Corpus retrieved on 22 April 2020.

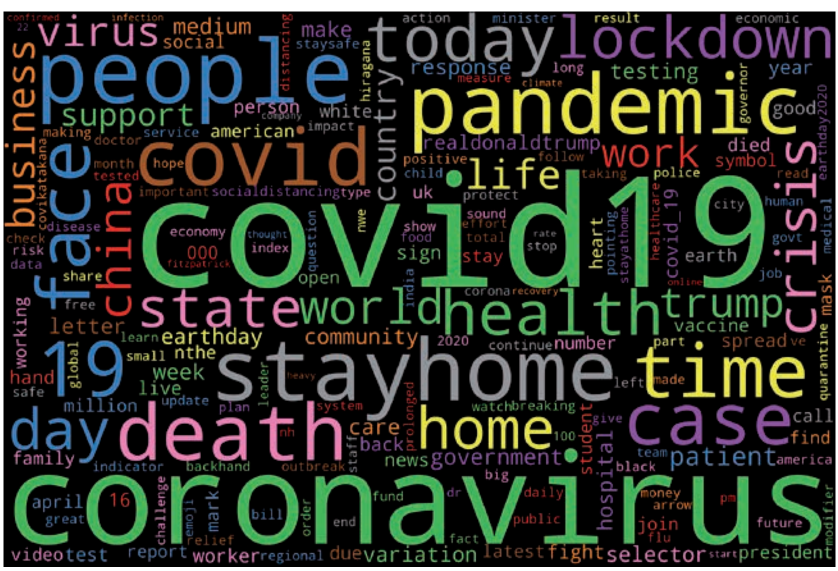


Figure 3.
TF-IDF word cloud corresponding to the 2020 Corpus retrieved on 22 April 2020.

Some of the most commonly occurring X users in the corpus belong to politicians—namely, Donald Trump (@realdonaldtrump), the then President of the United States; Matt Hancock (@matthancock), the Secretary of State for Health and Social Care in the UK at the time; and Narendra Modi (@narendramodi), the Prime Minister of India since 2014. Other commonly occurring users were the WHO (@who); the Director for the Centers for Disease Control and Prevention (@cdcdirector); and Tedros Ghebreyesus (@drtedros), the then Director-General of the WHO—Ghebreyesus currently serves a second five-year term as Director-General. **Table 2** lists the 10 most mentioned X users in the 2020 Corpus.

Originally, we planned to gather in 2022 as many tweets as we did in 2020, so that we could have two similar collections to compare. However, COVID-19 was no

X user	Number of occurrences
@realdonaldtrump	9272
@who	4685
@actorvijay	2182
@narendramodi	2051
@drtedros	1808
@ncdcgov	1524
@secpompeo	1457
@potus	1053
@cdcdirector	1043
@matthancock	1033

Table 2.
Most common users in the 2020 Corpus retrieved on 22 April 2020.

longer a trending topic when we started the retrieval of tweets in 2022. Of course, COVID-19 remained a popular topic, especially as new variants of the virus kept surging and updated vaccines were produced—COVID-19 was not the third-leading cause of death anymore, but the disease still ranked among the top 10 causes of death. Regardless, we were unable to gather as many tweets in 2022 as we did in 2020 within the same amount of time.

In 2022, we collected the first tweet on 24 March 2022 at 23:30:05 and the last one on 29 March 2022 at 00:30:07, and we refer to this collection as the *2022 Corpus*. While retrieving the 2020 Corpus took 4 hours, 31 minutes, and 48 seconds, and gathered 427,639 tweets, the retrieval of the 2022 Corpus took 4 days, 1 hour, and 2 seconds, and gathered only 265,108 tweets. Although the size of the 2022 Corpus is just 62% of the size of the 2020 Corpus, it provides sufficient tweets to monitor the evolution of the subject between 2020 and 2022.

To compile the 2022 Corpus, we collected tweets containing the hashtags listed in **Table 3**. This second time, we focused on hashtags that had not been considered in 2020 but had since emerged as the most popular keywords when referring to COVID-19. Thus, we included #longcovid, #covidisnotover, #omicron, #vaccinated, and #covidiots, among others. The values for the numbers of tweets retrieved for each hashtag in **Table 3** do not sum up to the total number of tweets in the 2022 Corpus—which is 265,108—because many tweets in the corpus contain two or more of the hashtags listed in **Table 3**. Additionally, some of the tweets in the 2022 Corpus may not explicitly contain the hashtags listed in **Table 3**, but the Streaming API still captures them because these hashtags appear in the URLs or metadata attached to such tweets [56].

Figure 4 shows the number of tweets that we retrieved every 12 hours between 24 and 29 March 2022. We captured 66,277 tweets per day on average. The most prolific day was 25 March 2022—when we captured 73,615 tweets—which coincided with the release of the *COVID-19 Infection Survey in the UK* [60].

As in the case of the 2020 Corpus, we used TF-IDF to identify the most characteristic terms in the 2022 Corpus. All the characteristic terms are depicted in **Figure 5** using a word cloud. Unsurprisingly, terms such as *vaccine*, *long COVID*, and *omicron* became more characteristic in 2022 than the terms in **Table 1**. Still, terms such as COVID and COVID19 remained the most predominant ones.

Hashtag	Number of tweets
#covid	182,686
#covid19	125,434
#longcovid	34,565
#covidisnotover	25,010
#omicron	15,608
#coronavirus	9487
#covid-19	8443
#pandemic	8429
#mask	6627
#sarscov2	2950
#stayhome	2118
#virus	2032
#vaccinated	1320

Table 3.
Hashtags used to gather the 2022 Corpus, which was retrieved between 24 March 2022 and 29 March 2022.

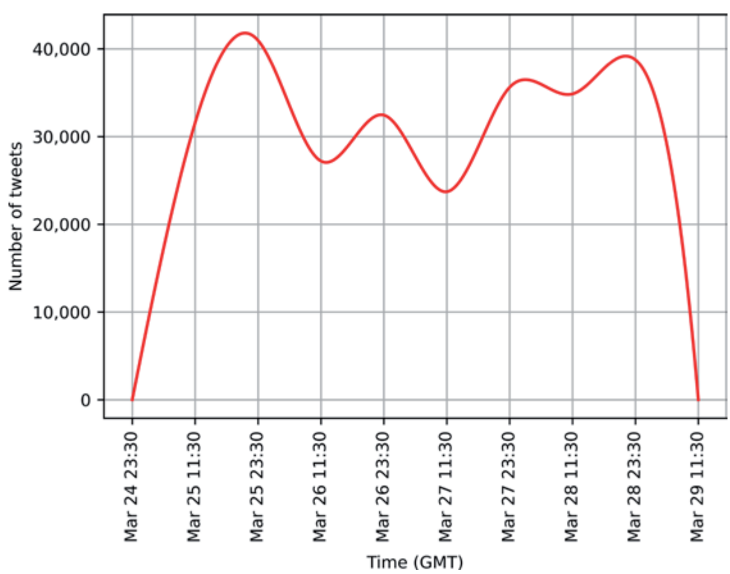


Figure 4.
Distribution of the tweets in the 2022 Corpus, which was retrieved between 24 March 2022 and 29 March 2022.

To archive the tweets we collected and analysed them later in detail, we uploaded them into two databases: a MySQL database and a *Neo4j* graph database [61]. While MySQL provided data persistence, Neo4j was largely used to analyse specific subsets of tweets separately.

Figure 6 displays a random sample of 967 tweets—the blue circles—and the 518 X users who posted them—the green circles. Green and blue circles are linked to show which specific user posted which specific tweet. The red circles represent hashtags

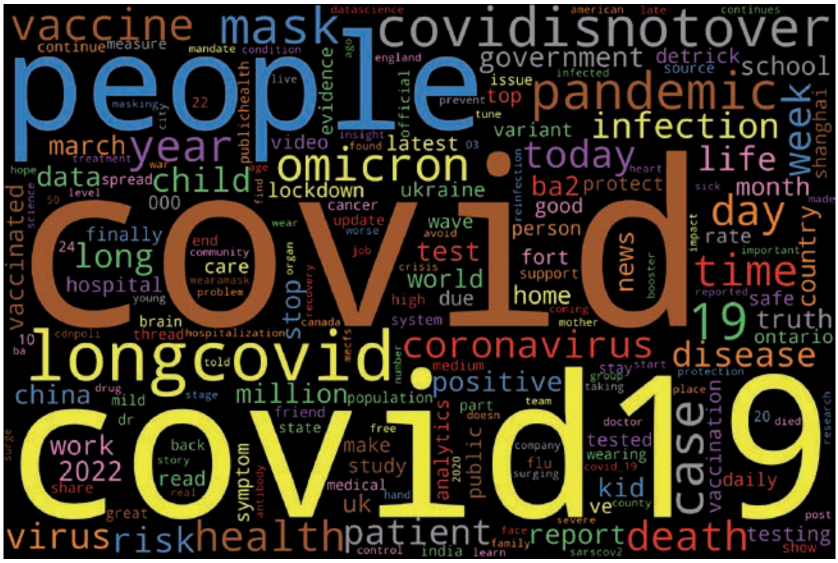


Figure 5.
Word cloud corresponding to the 2022 Corpus, retrieved between 24 March 2022 and 29 March 2022.

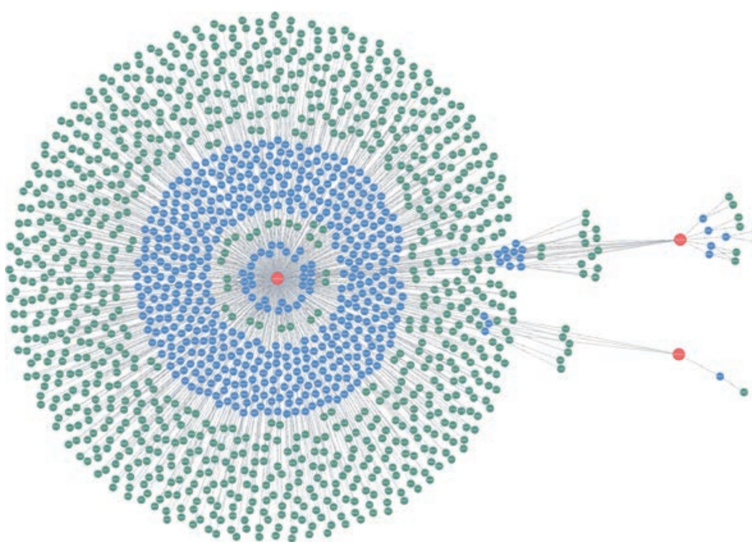


Figure 6.
Neo4j graph to analyse a random sample of 967 tweets.

#socialdistancing, #flattenthecurve and #coronavirusoutbreak. Hashtag #socialdistancing tagged most of the tweets in the sample; thus, it appears surrounded by many tweets. Hashtags #flattenthecurve and #coronavirusoutbreak, which appear on the right, tagged a small number of tweets from the sample.

3.2 Profile of mood states (POMS)

POMS is a test to measure a person's mood [62]. It lists 65 adjectives which are rated by individuals taking the test. Each adjective is associated with one of seven moods: anger, confusion, depression, fatigue, friendliness, tension, and vigour. POMS has been adapted lately to work with textual corpora, showing its benefits for determining moods in NLP research [63, 64]. To implement our own POMS classifier, we experimented with logistic regression [65], long short term memory (LSTM) [66], Naïve Bayes [67], random forests [68]—the number of trees was selected using linear search—and SVM [39].

We trained our classifier using Colnerič and Demšar's training set, which is based on a large dataset of 73 billion tweets labelled with emotions using a method called *distant supervision* [41]. Such a dataset was collected over more than six years, and it is split into training (60%), validation (20%) and test (20%) sets. Although the tweets are not recent, they were gathered over a long period of time, which means they are not influenced by temporal variations. Also, Colnerič and Demšar's corpus is much larger than other options available, such as Mohammad and Kiritchenko's corpus [69].

The random forest model was very slow to train. Hence, we could only build forests with a maximum of 100 trees. Even with this limit, it took more than a day to train a 100-tree forest using bi-grams on Colab. Our work was conducted using the scikit-learn library [70], with all parameters kept at their default settings.

Each tweet was treated as a sequence of characters that were input sequentially into the classifier. The classifier was then responsible for aggregating these characters into a suitable representation and predicting the corresponding mood—the benefit of doing this is that we did not require to do any pre-processing. We followed Colnerič and Demšar's suggestions [41] to train the classifier. We restricted our character set to characters appearing 25 times or more in the training data, resulting in a set of 410 characters. While POMS encompasses seven mood states, we excluded *friendliness*. This decision was informed by the findings of Norcross *et al.* [35], who concluded that the adjectives associated with friendliness were too weak to guarantee reliable results.

Table 4 shows the adjectives that we employed to identify each of mood under consideration. For instance, adjectives such as *worthless*, *hopeless* and *helpless* used to

Mood state	Adjectives
anger	angry, peeved, grouchy, spiteful, annoyed, resentful, bitter,
	ready to fight, deceived, furious, bad tempered, rebellious
confusion	forgetful, unable to concentrate, muddled, confused,
	bewildered, uncertain about things
depression	sorry for things done, unworthy, guilty, worthless,
	desperate, hopeless, helpless, lonely, terrified,
	discouraged, miserable, gloomy, sad, unhappy
fatigue	fatigued, exhausted, bushed, sluggish, worn out, weary,
	listless
tension	tense, panicky, anxious, shaky, on edge, uneasy, restless,
	nervous
vigour	active, energetic, full of pep, lively, vigorous, cheerful,
	carefree, alert

Table 4.
Mapping between adjectives and moods for POMS.

describe a person's feelings help to categorise her mood as falling within *depression* [71]. To gain further intelligence, we used text2emotion [72], which is a tool based on the work by Diaz *et al.* [19] and classifies the emotions embedded in text as happy, angry, sad, surprise, and fear.

4. Results

A metric for gauging sentiment is the *net sentiment rate* (NSR). Though originally designed for digital marketing, the NSR has proven effective in other domains—for example, Palomino *et al.* [45] applied it in a public health study.

Essentially, the NSR calculates the difference between the number of positive conversations—positive tweets—and the number of negative conversations—negative tweets—and then divides this difference by the total number of conversations—total number of tweets:

$$NSR = \frac{Positive\ tweets - Negative\ tweets}{Total\ number\ of\ tweets}. \tag{1}$$

We utilised *SentiStrength* [73] to gauge the overall sentiment conveyed on the experimental corpus. SentiStrength estimates the intensity of positive and negative sentiment in short texts, such as tweets, exploiting the informal grammar and spelling styles associated with social media [74]. SentiStrength has been able to predict positive emotion with 60.60% accuracy and negative emotion with 72.80% accuracy [73].

The NSR values for both the 2020 Corpus and the 2022 Corpus are negative: -0.32% and -0.37%, respectively, reflecting the negative nature of the corpora. **Figure 7** displays the percentages of positive, negative and neutral tweets in the 2020 Corpus and the 2022 Corpus, according to SentiStrength. Because the size of the 2020 Corpus differs from the size of the 2022 Corpus, **Figure 7** displays percentages, as opposed to absolute values. As shown in **Figure 7**, the distribution of the polarities of the 2020 Corpus and the 2022 Corpus is very similar.

For each tweet in the experimental corpus, we assigned a probability for the occurrence of each of the emotions according to text2emotion. **Figure 8** displays the

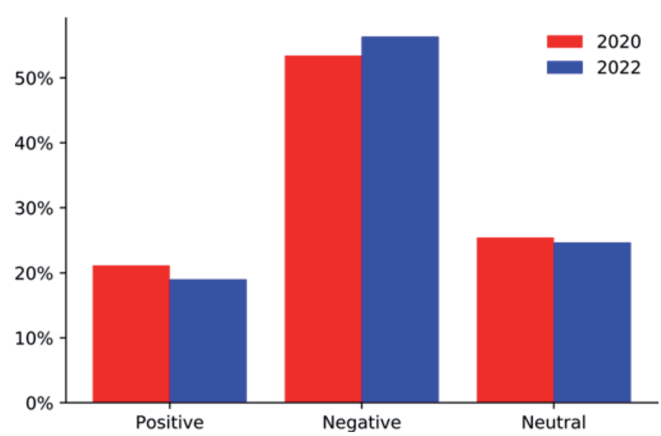


Figure 7.
Polarity of the experimental corpus according to SentiStrength.

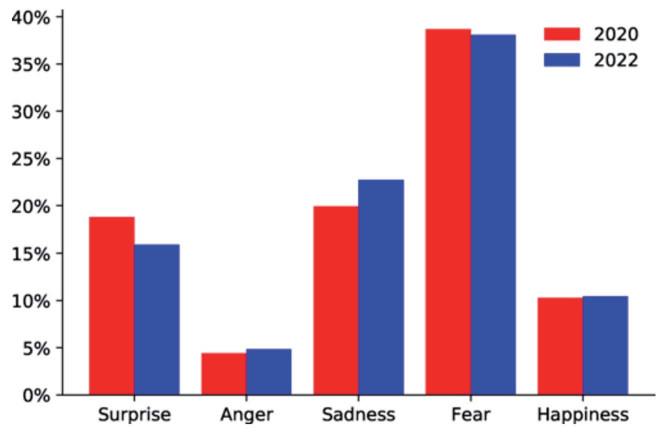


Figure 8.
Emotions in the experimental corpus according to text2emotion.

aggregate probability for each emotion to occur across all tweets. However, to normalise for the differing sizes of the 2020 and 2022 corpora, **Figure 8** presents percentages instead of absolute values. As it can be seen, fear is the most dominant emotion. Furthermore, the percentage distribution of all the emotions, including depression, remains strikingly similar across both years.

For each tweet in the experimental corpus, we assigned a probability for the occurrence of each mood according to our POMS classifier. **Figure 9** displays the aggregate probability for each mood to occur across all tweets.

To normalise for the differing sizes of the 2020 and 2022 corpora, **Figure 9** presents percentages instead of absolute values. As it can be seen, depression dwarfs all the other moods in the experimental corpus. Indeed, the presence of the other moods is minimal in comparison. Furthermore, the percentage distribution of all the moods, including depression, remains strikingly similar across both years.

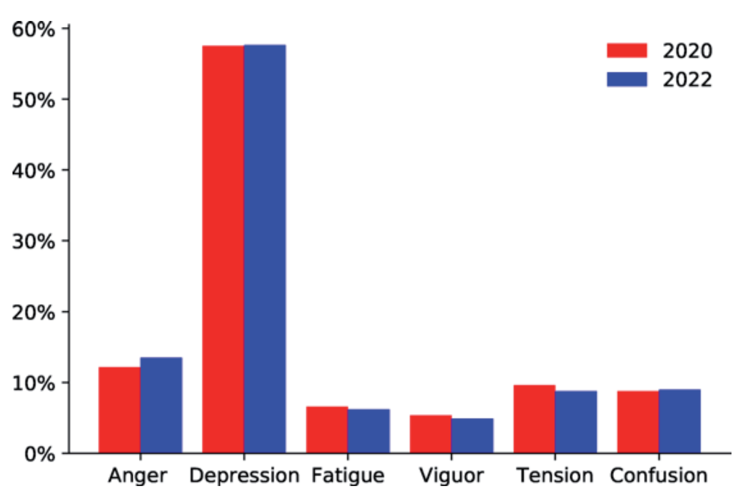


Figure 9.
Moods in the experimental corpus according to POMS.

5. Discussion

On 5 May 2023, the WHO announced that COVID-19 was no longer a public health emergency of international concern—more than three years after declaring the pandemic, the world achieved a high level of population immunity [75].

As lockdown restrictions eased across the UK by the end of 2021, social media users gave vent to their anger in various forms [18]. We witnessed a great deal of exasperation in response to the news. For instance, civil servants expressed their reticence about returning to their offices, while members of the public were in favour of ending the work-from-home restriction [76].

Eventually, violence erupted and clashes between policemen and protesters were reported in the UK and abroad [77]. There were grievances against the Government delays in relaxing the rules [78], complaints about patient safety in hospitals [79], and public outrage piqued when the Prime Minister was accused of not abiding by the same restrictions set for everyone else. Consequently, we expected anger to be the most dominant emotion expressed about COVID-19 in 2022. However, both text2emotion and POMS tell a different story. According to text2emotion, fear is the most dominant emotion, and according to POMS, depression is the most dominant mood.

5.1 Fear

Fear, as an emotion consistently experienced in relation to COVID-19, may be explained through evolutionary theory. Throughout evolution, defence mechanisms against potential dangers to survival, such as COVID-19, have been developed by humans [80]. The existing literature concentrates on the *behavioural immune system* (BIS), an evolutionary development consisting of affective, cognitive, and behavioural functions that have evolved to detect and respond to possible sources of infection [81].

The BIS is activated through sensory methods, invoking emotional and cognitive processes that lead to prophylactic behaviours. During BIS activation, the cognitive process can encourage future decisions to avoid danger-inducing situations, such as contact with the disease. Research has highlighted that avoidance plays into the development and maintenance of a psychopathology involving fear [82]. Indeed, fear has been characterised as an adaptive behavioural response and successful defence mechanism to predict and cope with dangers and threats [80, 83]. Expressions of fear can occur in multiple ways, such as physiological changes like sweating, heart rate, and blood pressure [84], as well as distress and behavioural avoidance [85]. Moreover, fear and its responses can be acquired through different means—for example, previous experience, verbal instruction, or observation [84]. Hence, there seems to be a rational explanation behind the fear of COVID-19 and its prevalence over time.

5.2 Depression

According to the *International Classification of Diseases* (ICD-10) [86] and the *DSM-V* [87], a potential symptom of depression is sadness [88]. Thus, it is not surprising that the second most dominant emotion and the most prominent mood felt were sadness and depression, respectively—see **Figure 8** and **Figure 9**. The occurrence of sadness and depression in the public discourse about COVID-19 may have started because of the lack of "everyday life" experienced by the population [89]. Research into the public perception of time during the lockdown showed that a feeling of slower passage of time was due to a significant increase in boredom and sadness [89].

An increase in depression and other mental and behavioural disorders commonly accompany large-scale disasters, whether traumatic, natural or environmental [90]. In the case of COVID-19, the UK saw a clear increase in the public perception of sadness and depression, and such perception was voiced in our experimental corpus, as determined by text2emotion and POMS.

The continuation of sadness and depression surrounding COVID-19 for two years is synonymous with previous research suggesting that the consequences of sadness and depression on mental health can last long after the event has occurred. Some researchers have estimated up to three years [90–92].

Recent reports have documented the increase in inflation, rising consumer prices, and the UK Government borrowing more than in previous years since the Second World War because of COVID-19 [93, 94]. It is, therefore, possible for the UK public to anticipate economic burdens, which is a potential risk factor for poor mental health in a post-COVID era [92].

6. Conclusions

Words can be easily misinterpreted or amended to convey different messages. For instance, writing "kl urself" instead of "kill yourself" is enough to circumvent the detection of profane words [95]. However, emotions are hard to conceal. Therefore, if we focused on identifying emotions, as opposed to simply reading text, we would be able to discover features and trends that are harder to spot otherwise. This is what we have attempted to do. We have applied machine learning to analyse tweets related to COVID-19.

The NSR values for our experimental corpus are negative. Thus, the answer to our first research question (RQ1) confirms that the prevailing sentiment expressed by the public during the COVID-19 pandemic is negative.

Academics have found that the rates of depression doubled when the pandemic began [96, 97]. Such mental health pattern was supported by prior studies analysing the psychological impact of previous epidemics and natural disasters [90]. This has been confirmed in our study. We employed POMS, which is capable of recognising depression, to support our investigation. Using POMS, we confirm that the most prominent emotion associated with COVID-19 in the UK is depression, and this answers our second research question (RQ2).

Finally, our third research question (RQ3) has led us to confirm the continuation of sadness and depression surrounding COVID-19 for two years. Former research has suggested that the consequences of sadness and depression on mental health can last long.

Author details

Marco A. Palomino[1*†‡], Rohan Allen[2†‡] and Aditya Padmanabhan Varma[3†‡]

1 School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, Scotland, UK

2 Whiting School of Engineering, Johns Hopkins University, Baltimore, USA

3 Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

*Address all correspondence to: marco.palomino@abdn.ac.uk

†These authors contributed equally.

‡‡Work undertaken while at the University of Plymouth.

References

[1] Ciotti M, Ciccozzi M, Terrinoni A, Jiang W-C, Wang C-B, Bernardini S. The COVID-19 pandemic. Critical Reviews in Clinical Laboratory Sciences. 2020;**57**(6):365-388

[2] Imran N, Zeshan M, Pervaiz Z. Mental health considerations for children & adolescents in COVID-19 pandemic. Pakistan Journal of Medical Sciences. 2020;**36**(COVID19-S4):S67

[3] Pfefferbaum B, North CS. Mental health and the COVID-19 pandemic. New England Journal of Medicine. 2020;**383**(6):510-512

[4] Benjamin A, Kuperman Y, Eren N, Rotkopf R, Amitai M, Rossman H, et al. Stress-related emotional and behavioural impact following the first COVID-19 outbreak peak. Molecular Psychiatry. 2021;**26**(11):6149-6158

[5] O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, et al. Mental health and well-being during the COVID-19 pandemic: Longitudinal analyses of adults in the UK COVID-19 mental health & wellbeing study. The British Journal of Psychiatry. 2021;**218**(6):326-333

[6] Shader RI. COVID-19 and depression. Clinical Therapeutics. 2020;**42**(6):962-963

[7] Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain, Behavior, and Immunity. 2020;**89**:531-542

[8] Koolagudi SG, Rao KS. Emotion recognition from speech: A review. International Journal of Speech Technology. 2012;**15**(2):99-117

[9] Liu B. Sentiment Analysis and Opinion Mining. Switzerland, AG: Morgan & Claypool Publishers; 2012

[10] Sun S, Luo C, Chen J. A review of natural language processing techniques for opinion mining systems. Information Fusion. 2017;**36**:10-25

[11] Holtgraves T. Social psychology and language: Words, utterances, and conversations. In: Handbook of Social Psychology. Hoboken, NJ; 2010

[12] Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology. 2010;**29**(1):24-54

[13] Eggly S, Manning MA, Slatcher RB, Berg RA, Wessel DL, Newth CJ, et al. Language analysis as a window to bereaved parents' emotions during a parent-physician bereavement meeting. Journal of Language and Social Psychology. 2015;**34**(2):181-199

[14] Murthy D. Twitter. Cambridge: Polity Press; 2018

[15] Mertiya M, Singh A. Combining naive Bayes and adjective analysis for sentiment detection on twitter. In: Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT). Vol. 2. Coimbatore, India: IEEE; 2016. pp. 1-6

[16] Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. In: International Conference on Language Resources and Evaluation. Valletta, Malta: European Language Resources Association (ELRA); 2010. pp. 1320-1326

[17] Johnson B. Prime Minister's Statement on Coronavirus (COVID-19): 23 March 2020. Government Digital Service (GDS); 2022. Published on 23 March 2020. Available from: https://www.gov.uk/government/speeches/

[18] Luan YL, Sun L, Luo F, Stillwell D. Public emotional responses to crisis: The COVID-19 pandemic in Wuhan and London. Social and Personality Psychology Compass. 2023;**17**(8):e12773

[19] Díaz SS et al. Intelligent execution of behaviors in a NAO robot exposed to audiovisual stimulus. In: IEEE Colombian Conference on Robotics and Automation (CCRA). Barranquilla, Colombia: IEEE; 2018. pp. 1-6

[20] Picard RW. Affective Computing. Cambridge, MA: MIT Press; 2000

[21] Calvo RA, D'Mello S. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing. 2010;**1**(1):18-37

[22] Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology. 2003;**54**(1):547-577

[23] Gonçalves P, Araújo M, Benevenuto F, Cha M. Comparing and combining sentiment analysis methods. In: Proceedings of the ACM Conference on Online Social Networks. Boston, Massachusetts, USA: Association for Computing Machinery (ACM); 2013. pp. 27-38

[24] Desmet B, Hoste V. Emotion detection in suicide notes. Expert Systems with Applications. 2013;**40**(16):6351-6358

[25] O. E. Dictionary. Emotion, N. Oxford University Press; 2024. DOI: 10.1093/OED/3434102633 [Accessed: 30 October 2024]

[26] O. E. Dictionary. Mood, N. (1). Oxford University Press; 2024. DOI: 10.1093/OED/3434102633 [Accessed: 30 October 2024]

[27] Ekman P. An argument for basic emotions. Cognition & Emotion. 1992;**6**(3-4):169-200

[28] Alm CO, Roth D, Sproat R. Emotions from text: Machine learning for text-based emotion prediction. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Vancouver, Canada: Association for Computational Linguistics; 2005. pp. 579-586

[29] Hakak NM, Mohd M, Kirmani M, Mohd M. Emotion analysis: A survey. In: International Conference on Computer, Communications and Electronics (COMPTELIX). Jaipur, India: IEEE; 2017. pp. 397-402

[30] Plutchik R, Kellerman H. Theories of Emotion. Vol. 1. Amsterdam, The Netherlands: Academic Press; 2013

[31] Plutchik R. A general psychoevolutionary theory of emotion. In: Theories of Emotion. Amsterdam, The Netherlands: Elsevier; 1980. pp. 3-33

[32] Machine Elf 1735. Robert Plutchik's Wheel of Emotions. 2011. Public domain via Wikimedia Commons. Available from: https://commons.wikimedia.org/wiki/File:Plutchik-wheel.svg

[33] Renaud-Charest O, Lui LM, Eskander S, Ceban F, Ho R, Di Vincenzo JD, et al. Onset and frequency of depression in post-Covid-19 syndrome: A systematic review.

Journal of Psychiatric Research. 2021;**144**:129-137

[34] Johns G, Samuel V, Freemantle L, Lewis J, Waddington L. The global prevalence of depression and anxiety among doctors during the Covid-19 pandemic: Systematic review and meta-analysis. Journal of Affective Disorders. 2022;**298**:431-441

[35] Norcross JC, Guadagnoli E, Prochaska JO. Factor structure of the profile of mood states (POMS): Two partial replications. Journal of Clinical Psychology. 1984;**40**(5):1270-1277

[36] Berger BG, Motl RW. Exercise and mood: A selective review and synthesis of Research employing the profile of mood states. Journal of Applied Sport Psychology. 2000;**12**(1):69-92

[37] McNair DM, Lorr M, Droppleman LF. Manual Profile of Mood States. San Diego, CA: Educational & Industrial Testing Service; 1971

[38] Chaffar S, Inkpen D. Using a heterogeneous dataset for emotion analysis in text. In: Butz C, Lingras P, editors. Advances in Artificial Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. pp. 62-67

[39] Noble WS. What is a support vector machine? Nature Biotechnology. 2006;**24**(12):1565-1567

[40] Rudkowsky E, Haselmayer M, Wastian M, Jenny M, Emrich Š, Sedlmair M. More than bags of words: Sentiment analysis with word embeddings. Communication Methods and Measures. 2018;**12**(2-3):140-157

[41] Colnerič N, Demšar J. Emotion recognition on twitter: Comparative study and training a unison model. IEEE Transactions on Affective Computing. 2018;**11**(3):433-446

[42] World Health Organization (WHO). Global Diffusion of eHealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on eHealth. 2016. Available from: https://iris.who.int/bitstream/handle/10665/252529/9789241511780-eng.pdf

[43] Wilson K, Brownstein JS. Early detection of disease outbreaks using the internet. Canadian Medical Association Journal. 2009;**180**(8):829-831

[44] de Zúñiga HG. Social media use for news and individuals' social capital, civic engagement and political participation. Journal of Computer-Mediated Communication. 2012;**17**(3):319-336

[45] Palomino M, Taylor T, Göker A, Isaacs J, Warber S. The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to "nature-deficit disorder". International Journal of Environmental Research and Public Health. 2016;**13**(1):1-23

[46] Stautz K. Reactions on twitter to updated alcohol guidelines in the UK: A content analysis. BMJ Open. 2017;**12**(4):e015493

[47] Welch V, Petkovic J, Pardo JP, Rader T, Tugwell P. Interactive social media interventions to promote health equity: An overview of reviews. Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice. 2016;**36**:63

[48] Dixon SJ. X/twitter: Number of monetizable daily active users worldwide 2017-2022. Stat. 2024. Available from: https://www.statista.com/

[49] Internet Live Stats. Twitter Usage Statistics. 2022. Available from: https://www.internetlivestats.com/twitter-statistics/

[50] Milstein S, O'Reilly T. The Twitter Book. Farnham, Surrey, United Kingdom: O'Reilly Media; 2009

[51] World Health Organization (WHO). Coronavirus Disease (COVID-19) Pandemic. 2020. Available from: https://www.who.int/europe/emergencies/situations/covid-19

[52] Walsh F. Coronavirus: First patients injected in UK vaccine trial. BBC News Services. 2020. Available from: https://www.bbc.co.uk/news/health-52394485

[53] GOV.UK. Slides, Datasets and Transcripts to Accompany Coronavirus Press Conferences. 2020. Available from: https://www.gov.uk/government/collections/

[54] Roesslein J. Tweepy Documentation. 2020. Available from: http://docs.tweepy.org/en/v3.5.0/

[55] Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In: 7th International AAAI Conference on Weblogs and Social Media. Cambridge, MA: The AAAI Press; 2013. pp. 400-408

[56] Twitter Inc. Filter Realtime Tweets. 2020. Available from: https://developer.twitter.com/en/docs/tweets/filter-realtime/overview

[57] Bisong E. Google Colaboratory. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Berkeley, CA: Springer; 2019. pp. 59-64

[58] G. Research. Google Colaboratory. 2022. Available from: https://colab.research.google.com/

[59] Schütze H, Manning CD, Raghavan P. Introduction to Information Retrieval. Cambridge, United Kingdom: Cambridge University Press; 2008

[60] Yapp R, Fuller E, Carbury R. Coronavirus (COVID-19) Infection Survey, UK: 25 March 2022. Office for National Statistics; 2022. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot

[61] Miller JJ. Graph database applications and concepts with Neo4j. Proceedings of the Southern Association for Information Systems Conference. 2013;**2324**(36):141-147

[62] Curran SL, Andrykowski MA, Studts JL. Short form of the profile of mood states (POMS-SF): Psychometric information. Psychological Assessment. 1995;**7**(1):80

[63] Pepe A, Bollen J. Between conjecture and memento: Shaping a collective emotional perception of the future. In: AAAI Spring Symposium: Emotion, Personality, and Social Behavior. Palo Alto, CAL: AAAI Press; 2008. pp. 111-116

[64] Bollen J, Mao H, Pepe A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proceedings of the AAAI Conference on Web and Social Media. Barcelona, Spain: AAAI Press; 2011. pp. 450-453

[65] Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. Logistic Regression. New York, NY: Springer; 2002

[66] Nowak J, Taspinar A, Scherer R. LSTM recurrent neural networks for short text and sentiment classification. In: Artificial Intelligence and Soft

Computing. Zakopane, Poland: Springer; 2017. pp. 553-562

[67] Berrar D. Bayes' theorem and naive Bayes classifier. In: Ranganathan S, Cannataro M, Asif M, editors. Encyclopedia of Bioinformatics and Computational Biology. Amsterdam, The Netherlands: Elsevier (In Press); 2025

[68] Biau G. Analysis of a random forests model. The Journal of Machine Learning Research. 2012;**13**(1):1063-1095

[69] Mohammad SM, Kiritchenko S. Using hashtags to capture fine emotion categories from tweets. Computational Intelligence. 2015;**31**(2):301-326

[70] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. Journal of Machine Learning Research. 2011;**12**:2825-2830

[71] Mackenzie B. Scoring for POMS. 2022. Available from: https://www.brianmac.co.uk/pomscoring.htm

[72] Python Software Foundation. text2emotion 0.0.5. 2022. Available from: https://pypi.org/project/text2emotion/

[73] Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology. 2010;**61**(12):2544-2558

[74] Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology. 2012;**63**(1):163-173

[75] Farid Rahimi MD, Abadi ATB. 'The end'—Or is it? Emergence of SARS-CoV-2 EG.5 and BA.2.86 subvariants. Future Virology. 2023;**18**(13):823-825

[76] Zeffman H. Civil servants angry at COVID return to office but public in favour. The Times. 2022. Published on 24 January 2022. Available from: https://www.thetimes.co.uk/

[77] BBC News—Europe. Netherlands: Police Fire Warning Shots at COVID Protesters. 2021. Published on 20 November 2021. Available from: https://www.bbc.co.uk/news/

[78] Parsley D, Blackall M, Barrie J. 'We canot wait much longer... We have spent thousands getting ready': Venues angry at COVID-19 reopening delay. iNews: Associated Newspaper Limited. 2021. Published on 14 June 2021. Available from: https://inews.co.uk/news

[79] Smyth L. My Dad Was Not Ready to Die: Family Angry at Covid Death in Infected NI Hospital. 2020. Published on 7 September 2020. Available from: https://www.belfasttelegraph.co.uk/

[80] Öhman A, Mineka S. Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychological Review. 2001;**108**(3):483

[81] Zhao X, Shi W, Li X, Li W, Li C. Autistic traits and negative emotions in the general population during COVID-19: Mediating roles of the Behavioural immune system and COVID-19 risk perception. Psychiatry Research. 2021;**300**:113918

[82] Shook NJ, Oosterhoff B, Terrizzi JA, Clay R. Disease Avoidance: An Evolutionary Perspective on Personality and Individual Differences. In: The SAGE Handbook of Personality and Individual Differences. London, UK; 2018

[83] Dymond S, Dunsmoor JE, Vervliet B, Roche B, Hermans D. Fear generalization in humans: Systematic review and implications for anxiety

disorder Research. Behavior Therapy. 2015;**46**(5):561-582

[84] Delgado MR, Olsson A, Phelps EA. Extending animal models of fear conditioning to humans. Biological Psychology. 2006;**73**(1):39-48

[85] Presti G, McHugh L, Gloster A, Karekla M, Hayes SC. The dynamics of fear at the time of COVID-19: A contextual behavioral science perspective. Clinical Neuropsychiatry. 2020;**17**(2):65

[86] Fung KW, Xu J, Bodenreider O. The new international classification of diseases 11th edition: A comparative analysis with ICD-10 and ICD-10-CM. Journal of the American Medical Informatics Association. 2020;**27**(5):738-746

[87] Frances A. Whither DSM–V? The British Journal of Psychiatry. 2009;**195**(5):391-392

[88] Mouchet-Mages S, Baylé FJ. Sadness as an integral part of depression. Dialogues in Clinical Neuroscience. Taylor & Francis. 2008;**10**(3):321-327

[89] Droit-Volet S, Gil S, Martinelli N, Andant N, Clinchamps M, Parreira L, et al. Time and COVID-19 stress in the lockdown situation: Time free, "dying" of boredom and sadness. PLoS One. 2020;**15**(8):e0236465

[90] Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: The need for prevention and early intervention. JAMA Internal Medicine. 2020;**180**(6):817-818

[91] Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. The Lancet. 2020;**395**(10227):912-920

[92] Vadivel R, Shoib S, El Halabi S, El Hayek S, Essam L, Bytyçi DG, et al. Mental health in the post-COVID-19 era: Challenges and the way forward. General Psychiatry. 2021;**34**(1):e100424

[93] Giles C, Strauss D. COVID Response Pushes UK Borrowing to Highest since Second World War. 2021. Available from: https://www.ft.com/content/68804db6-11a1-4d6e-97b2-ba83cb476e43

[94] Leaker D. Labour Market Overview, UK: June 2022. 2022. Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/june2022

[95] Blackwell L, Dimond J, Schoenebeck S, Lampe C. Classification and its consequences for online harassment: Design insights from HeartMob. In: Proceedings of the ACM on Human-Computer Interaction. NY, USA: ACM New York; 2017. pp. 1-19

[96] Khubchandani J, Sharma S, Webb FJ, Wiblishauser MJ, Bowman SL. Post-lockdown depression and anxiety in the USA during the COVID-19 pandemic. Journal of Public Health. 2021;**43**(2):246-253

[97] Seal A, Schaffner A, Phelan S, Brunner-Gaydos H, Tseng M, Keadle S, et al. COVID-19 pandemic and stay-at-home mandates promote weight gain in US adults. Obesity. 2022;**30**(1):240-248

Edited by Jinfeng Li

What happens when machines start to understand—maybe even feel—our emotions? *Emotions in Code - The AI Frontier of Sentiment Analysis* explores the fascinating intersection of artificial intelligence and human emotions. As algorithms become more sophisticated, they are learning not just to process data but to, in effect, read between the lines—interpreting tone, mimicking affect, and reacting to our moods in real-time. In a world increasingly shaped by AI, grasping how these systems interpret sentiment is not just useful—it is vital. The call for emotional intelligence in our technologies echoes louder than ever.

- Addressing this urgent need, the book provides:
- A layered exploration of sentiment analysis, from paradoxical challenges to emerging "empathy" algorithms.
- A look at the machine learning methods behind emotional recognition in language, voice, and behavior.
- Case studies and insights from across the globe, spanning media, healthcare, finance, and more.
- Reflections on the ethical stakes of emotional AI—what it means for trust, communication, and the human experience.

Blending technical depth with real-world relevance, this book caters to researchers, developers, and anyone curious about how machines are becoming smarter at reading us, and what that means for the future.

Andries Engelbrecht, Artificial Intelligence Series Editor

ISSN 2633-1403
ISBN 978-1-83634-812-2
9 781836 348122

Published in London, UK

IntechOpen