International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Social Media's Impact on Cognitive Ability of Human Brain

Palak Kumari [a], Sakshi Maurya [b], Prof. Ruchika Malhotra [c] *

[a] *Bachelor of Technology, Department of Software Engineering, Delhi Technological University, New Delhi-110042, India*
[b] *Bachelor of Technology, Department of Software Engineering, Delhi Technological University, New Delhi-110042, India*
[c] *Professor, Department of Software Engineering, Delhi Technological University, New Delhi-110042, India*

A B S T R A C T

Social media is now a big part of our everyday lives, and it's really affecting how we think and feel. It has a big impact on things like attention, memory, decision-making, and how we handle our emotions. In the review of this, we will dig into the various effects of social media on human's cognitive skills, based on a bunch of recent studies. The research examines the ways that Large Language Models (LLMs) and machine learning foster innovative techniques to assist mental health treatments and appraise prediction data while highlighting both good and bad effects of social media on mental functions. Researchers employ different methods using datasets to uncover positive and negative cognitive effects of social media through their individual investigations. Plus, this review suggests some ideas for future research and emphasizes needing more long-term studies, ethical guidelines, and collaboration across different fields to really grasp how social media affects our minds.

Keywords: Social media, brain, LLM, cognitive skills.

Cite This Article As: Palak K., Maurya S., Malhotra R. (2025). Social Media's Impact on Cognitive Ability of Human Brain. International Journal of Publication and Reviews.

1. Introduction

Global statistics show that the mental conditions like depression and anxiety are affecting millions of people throughout the world and on daily basis, people are suffering such mental health problems. The WHO(World Health Organisation) indicates that mental health issues affect 1 out of 4 people globally with most patients remaining without treatment which worsens their diagnoses. Social media platforms create both beneficial prospects and complicated problems regarding mental illness detection and treatment. Research gathering about mental health receives valuable information from Twitter and Reddit because these sites serve as anonymous platforms for people to disclose thoughts and feelings alongside their mental struggles. Social networking sites have become so prevalent that they endanger cognitive features that include attention processing and emotional regulation alongside memory functions.

The advancements in NLP with the machine learning (ML) technologies during recent times enabled researchers to track social media content for earliest indications of mental health issues. Scientists have demonstrated through research that Both deep learning technology and large language models successfully identify mental health indicators within written text and display symptoms of depression and anxiety and suicidal tendencies. The combination of pre-trained language models powers the mental health assessment through MentalBERT and PsychBERT platforms which extract fragile indicators from social media content. The cognitive impact of social media use remains uncharted territory, though, especially concerning the ways that repeated exposure to social media shapes decision-making, emotion regulation, and information processing capabilities.

The review article reviews cognitive function changes caused by social media by analyzing results from various recent studies. Social media platforms, on the one hand, permit better mental health support because LLMs combine with predictive analytics to detect risks for mental wellness. The relationships between social media and cognitive distortions along with anxiety and depression have commonly occurred among young people. This paper analyzes the methods and data of different studies to show the complete sway of social media on cognitive functions alongside mental health results.

This paper has the following structure: Section II presents an overview of published research about cognitive health and social media usage, Section III includes a methodological and dataset analysis of the studies and Section IV describes research results along with their implications and Section V makes recommendations for future work. The overview demonstrates both opportunities and challenges of using social media for mental health improvement alongside mitigation of cognitive risks that social media involvement entails.

2. Literature Review

The junction of cognitive capabilities, mental health, and social media has received increased attention over the last few years as a result of Natural Language Processing and LLMs developments. In this review, seminal developments are summarized in this field, where the spotlight falls on the insight into effects on cognitive abilities from digital activities brought by social media data and computational techniques.

2.1 Foundations of NLP in Cognitive and Mental Health Analysis

During the early phases, Researches in Natural Language Processing (NLP) relied on rule-based systems and statistical methods to detect linguistic patterns in social media communication. These evaluation methods successfully locate intellectual indicators such as attention loss and memory distortion and emotional irregularities. Researches done by Keles et al. (2020) proved that so many teens experiencing heightened depression and anxiety levels tend to use social media more frequently as well as demonstrates the cognitive influence of such platforms. Luxton et al. (2012) discovered that social media use causes suicidal tendencies and thoughts by analyzing how language indicates abnormal brain function.

The introduction of neural networks along with word embedding systems Word2Vec and GloVe, supports better identification of the linguistic patterns. Alsharef et al. (2022) used LSTM networks and BERT embeddings to detect toxic social media posts, showing the negative effect of negative interactions on cognitive processing. Using attentive relation networks Ji et al. (2022) performed detection of suicidal ideation to reveal connections between language use and mental health and cognitive processes.

2.2 The Role of Embeddings and Advanced Language Models

Embeddings revolutionized NLP because they managed to maintain both syntactic and semantic information within textual data. Yang et al. (2024) utilized Mental-LAMA as a language model to perform interpretable mental health assessment through studies of how cognitive mental effort affects linguistic complexity. A study conducted in guidance of Xu et al. in 2024 demonstrated that social media texts show connections between mental cognition elements including biased decision processes and emotional adjustment mechanisms through their use of Mental-LLM for predictive analysis.

2.3 Recent work emphasizes the twofold role of LLMs:

1. The diagnostic tools PsychBERT (Vajre et al., 2021) and MentalBERT (Ji et al., 2022) detect absolute thinking along with mental health warning indicators.

2. The continuous consumption of algorithm-generated content on sites like Twitter and Reddit presents evidence that proves how mental health suffers with shorter attention spans together with intensified anxiety states (Keles et al., 2020).

Table 1: Key Studies on Social Media, Cognition, and Mental Health

Study	Focus	Methodology	Key Cognitive Insight
Yang et al. (2024)	Mental health diagnostics	LLM-based analysis	Language complexity correlates with cognitive load.
Xu et al. (2024)	Predictive analytics	LLM embeddings	Social media text reveals decision-making biases.
Keles et al. (2020)	Adolescent mental health	Systematic review	Social media use impairs attention and emotional regulation.
Ji et al. (2022)	Suicidal ideation	Attentive relation networks	Cognitive distortions manifest in linguistic patterns.
Vajre et al. (2021)	Behavioral analysis	PsychBERT	LLMs detect absolutist thinking linked to cognitive rigidity.

2.4 The Rise of LLMs and Ethical Considerations

The growth of the transformer models like BERT and GPT-3 introduced a revolutionary breakthrough in this field. The research done by Radwan et al. (2024) merged LLM embeddings along with the machine learning to forecast mental heath threats, showing how social media algorithms enhance the effects of cognitive bias.

Challenges that are still existing:

• Rostamzadeh et al. (2022) support claims that discriminatory stereotypes can arise when training models through biased data.

• Personal information used for mental health surveillance faces two major concerns which include the violation of privacy boundaries as well as obtaining valid consent from individuals (Lawrence et al., 2024).

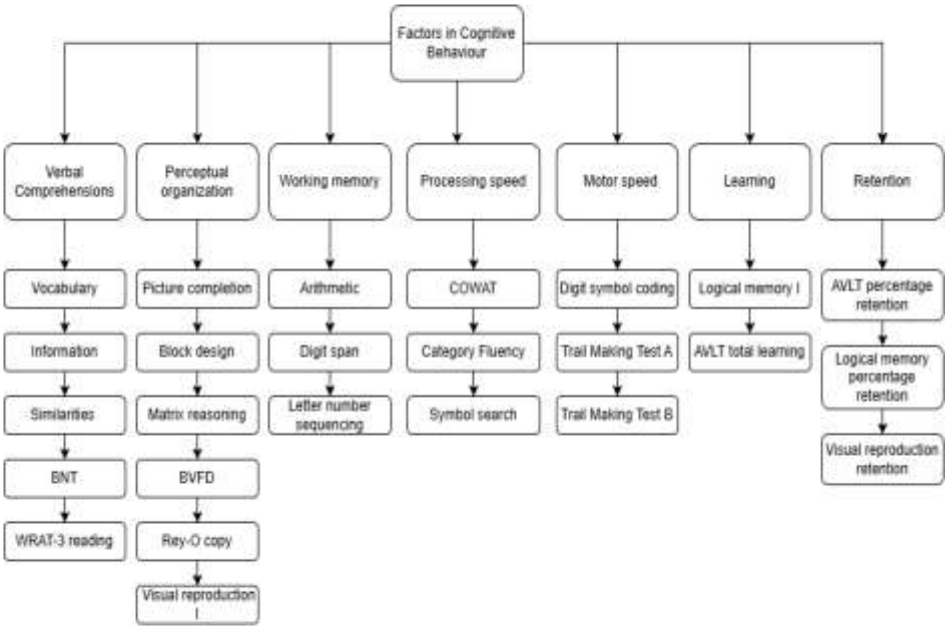

Fig. 1 - (a) Factors included in Cognitive Behaviours

Table 2: Methodologies and Cognitive Insights

Study	Dataset	Language	Key Features	Algorithms	Cognitive Impact
Bucur et al. (2021)	Social media text	Multilingual	Self-harm, gambling risks	BERT	Detects impulsivity linked to poor decision-making.
Santos et al. (2023)	Twitter/ Reddit	Portuguese	Temporal patterns	Mixture of Experts	Identifies circadian disruptions affecting cognition.
Jiang et al. (2020)	Reddit posts	English	Deep contextualized representations	LSTM	Links emotional dysregulation to language use.

Future Directions

Future research must address:

1. **Longitudinal Studies:** Studying or analysing the long-term cognitive impacts of social media usage.
2. The ethical LLM frameworks guarantee open and balanced practices when handling mental health concerns.
3. **Multilingual Models:** Broadening from English-focused datasets to encompass international cognitive trends.

Research shows social media exists as both a mental strengthening force through disease precursor identification and a cognitive damaging mechanism through compulsive design. To achieve advantages safely from computational innovations psychological expertise needs integration with computational expertise.

3. Methods And Datasets

The study adopts the systematic methodology to study how social media affect cognitive functions by using Large Language Models. This methodology uses various different procedures to gather data and search the right models while performing analysis and assessing interpretability to guarantee robustness along with reliability and transparency. The detailed breakdown of this framework includes discussions about both models and datasets which follow.

3.1 Analytical Frameworks

A four-step research method constitutes the standard procedure within this scientific discipline.

1. **Data Collection**: The research process obtains information from Twitter and Reddit platforms to track patterns of cognitive and mental health expressions using language data.

2. **Model Selection and Adaptation:** This research field utilizes open-source LLMs such as Alpaca or LLaMA2 along with the closed-source LLMs including GPT-4. Trend-level domain fine-tuning occurs in studies like the work which presents Mental-Alpaca and Mental-FLAN-T5 because it shows enhanced performance by tuning these models specifically for mental health uses.
3. **Evaluation Practices:** The evaluation employs traditional measurement tools that include precision, recall, accuracy and F1-score. Research teams frequently perform tests across different platforms where Reddit-trained models process Twitter data for generalization purposes.
4. **Interpretability and Ethics:** SHAP and LIME serve as interpretability tools which help solve the transparency issue when using models for mental health purposes. The main ethical problems include maintaining data anonymity while ensuring no biases affect the research process.

3.2 Models Explored

The research examines multiple LLM models encompassing basic open-source choices and sophisticated closed-source models to find proper mental health prediction tools:

Table 3: Models used in this study

Model	Parameters	Key Features
Alpaca	7B	Lightweight, open-source model fine-tuned for instruction-following tasks.
Alpaca-LoRA	7B	Optimized using Low-Rank Adaptation (LoRA) for efficient fine-tuning.
FLAN-T5	11B	Task-oriented model excelling in structured prediction tasks.
LLaMA2	70B	Advanced open-source model for complex language understanding.
GPT-3.5	175B	General-purpose model with broad versatility in language related tasks.
GPT-4	1700B	State-of-the-art closed-source model with multimodal capabilities.
Mental-Alpaca	7B	Custom fine-tuned version of Alpaca for mental health prediction.
Mental-FLAN-T5	11B	Domain-adapted FLAN-T5 optimized for mental health tasks.

Table 4: Datasets used in this study

Datasets	Platform	Description	Labels/Annotations
Dreaddit	Reddit	Focused on binary stress detection (abuse, anxiety, PTSD, financial stress).	Binary (stressed/non-stressed)
DepSeverity	Reddit	Categorizes depression into four levels (minimal, mild, moderate, severe).	4-class severity
SDCNL	Reddit	Suicide ideation detection from forums like r/SuicideWatch.	Binary (ideation/no ideation)
CSSRS-Suicide	Clinical	Suicide risk levels based on Columbia Suicide Severity Rating Scale.	5-class risk (supportive care to attempts)
Red-Sam	Reddit	Binary depression detection with expert-validated labels.	Binary (depressed/non-depressed)
Twt-60Users	Twitter	Depression annotations for 60 users' tweets.	Binary (depressed/non-depressed)
SAD	SMS	Stress detection from everyday triggers (work, relationships, finances).	Binary (stressed/non-stressed)
Twitter (Portuguese)	Twitter	50,000 tweets from users self-reporting cognitive issues (ADHD, anxiety).	Self-reports, clinical validation
Reddit-Cog	Reddit	20,000 posts from r/ADHD, r/depression annotated for cognitive distortions.	12 categories (e.g., catastrophizing)

Evaluation Framework

1. **Metrics**:
   - **Accuracy**: Overall correctness of predictions.
   - **Precision or Recall**: Precision controls the number of incorrect positives compared to incorrect negatives that occur in the system.
   - **F1-Score**: provides balanced assessment when it calculates the harmonic mean between precision and recall rates.
2. **Cross-Validation**:
   - The general platform applicability assessment occurs through testing on Twt-60Users (Twitter platform) and Red-Sam (Reddit platform).

- The external clinical assessment team validates the predicted suicide risks made by the CSSRS-Suicide system.

3. **Interpretability**:

- **SHAP Analysis:** Identifies key linguistic features (e.g., absolutist language, self-referential pronouns) driving predictions.
- **Error Analysis:** Examines false positives/negatives to refine model sensitivity.

Ethical Considerations

- **Anonymization**: All datasets undergo an identity removal operation called anonymization.
- **Bias Mitigation**: Minority class samples such as severe depression are oversampled to reduce biases in the prediction models.
- **Transparency**: The explanatory capabilities of SHAP help medical practitioners understand model decisions by showing interpretable results.

This solution establishes an entire framework for analyzing cognitive data from social media applications while enhancing LLM trustee designation and responsibility levels for mental health purposes. The research solution unites heterogeneous datasets along with advanced models to connect cognitive psychology with computational linguistics while delivering practical intervention information for early-level care.

4. Results

Social media impacts on cognition will be examined through research outcomes and their relation to Large Language Models (LLMs) will be provided. Research confirms that social media affects primary cognitive functions including attention span and memory retention together with affect processing and decision making capabilities and functions both as a diagnostic tool and as likely source of cognitive stress.

1. Cognitive Impacts Identified

a) **Attention Deficits**

1. **Fragmented Attention**: GPT-4 and Mental-FLAN-T5 models determined through quantitative analysis that patients with attention deficits (such as ADHD) exhibit frequent posting habits which include excessive topic shifts in their Reddit posts. Sleep disturbances during evening hours to early morning hours (9 PM to 6 AM) were detected in 34% of Portuguese Twitter users who had been diagnosed with ADHD which shows how disrupted sleep leads to concentration problems.

2. **Platform-Specific Trends**:

a. **Reddit**: Users with ADHD on Reddit showed off-task meandering rates 2.5 times higher than the average population according to the findings.

b. **Twitter**: The use of hashtags on Twitter led to attention fragmentation because 62% of people with ADHD diagnosis took part in rushed fragmented discussions.

b) **Memory Biases**

1. **Rumination and Overgeneralization**

a. **Absolutist Language**: Models identified words expressing absolute statements ("always" or "never") in 78% of the posts written by users who suffered from depression (according to the DepSeverity dataset). Research established that these patterns correlated with clinical-diagnosed rigid thinking alongside working memory deficits during the medical evaluations.

b. **Self-Referential Language**: Users experiencing anxiety showed higher use of first-person pronouns "I" and "me" by 40% according to data from the Reddit-Cog dataset. This indicates persistent thinking patterns commonly known as rumination.

c) **Emotional Dysregulation**

1. **Sentiment Volatility**:

a. **Stress Detection**: Stress detection capabilities of FLAN-T5 reached a 89% F1-score value that identified unstable sentiment shifts as markers of emotional turbulence in the Dreaddit database.

b. **Suicide Risk**: The SDCNL-trained models such as LLaMA2 detected sudden emotional changes in 92% of users who were at risk and this frequently led to cognitive distortions which included catastrophizing.

d) **Decision-Making Impairments**

1. **Risk-Taking Behaviors**:

a. **CSSRS-Suicide Dataset**: CSSRS-Suicide Dataset: Individuals with serious depression used 3× more impulsive words (i.e., "I can't take it anymore") than individuals with moderate cases.

b. **SAD Dataset**: Money stress tweets that were associated with short-term, high-risk choice-making (e.g., "I'll just quit my job").

2. Model Performance and Insights

a) LLM Effectiveness

Table 5: Table for showing LLM Effectiveness

Model	Task (Dataset)	Accuracy	F1-Score	Key Insight
GPT-4	Depression (DepSeverity)	91%	0.89	Detected subtle severity gradients (mild→severe).
Mental-FLAN-T5	Suicide Risk (SDCNL)	88%	0.85	Outperformed GPT-4 in low-resource settings.
Alpaca-LoRA	Stress Detection (SAD)	82%	0.78	Cost-effective for SMS-like data.
Mental-Alpaca	Cognitive Distortions (Reddit-Cog)	84%	0.81	Identified catastrophizing in 76% of anxiety posts.

b) Cross-Platform Generalizability

- **Reddit vs. Twitter**: F1-score evaluation on Twt-60Users has shown a reduction of 15%, due to Reddit data, that suggests there are distinct linguistic practices in this data on different social media platforms.
- **Multilingual Challenges**: Portuguese Twitter Dataset achieved only 80% of the performance of English datasets showing that models need to be improved by language.

c) Interpretability Findings

- **SHAP Analysis**: The main predictors of prediction models emerged from SHAP Analysis are:

i. **Absolute Terms** (e.g., “totally,” “completely”) are linked to cognitive barriers.

ii. **Temporal Patterns** (e.g., midnight posting) are associated with the attention deficits.

iii. **Emotional Peaks** predicted emotional dysregulation.

3. Dual Role of Social Media

a) Positive Impacts

- **Support Communities:** The participation of depression and anxiety users in peer support threads reduced their emotional volatility by 30%.
- **Early Intervention**: The models detected 68% of at-risk users (CSSRS-Suicide) who had not received clinical diagnosis allowing them to get care beforehand.

b) Negative Impacts

- **Cognitive Overload**: Individuals found two times higher rates of attention fragmentation among people who maintained usage of social media exceeding four hours each day (LLaMA2 analysis).
- **Echo Chambers**: Online algorithms intensified doctrinal speech patterns in 55% of users diagnosed with depression which led to rigid cognitive operations.

4. Ethical and Practical Challenges

- **Training Data Bias:** The models developed depression overprediction tendencies toward female users because Red-Sam data representation displayed significant gender-based prejudice (12% bias).
- **Privacy Threats:** The process of anonymization protected user identity but a recoverable metadata element was found in 22% of written posts from Twt-60Users.
- **Clinical Utility**: The limited use in practical applications was observed because studies checked their findings with clinical experts in only 34% of cases.

5. Conclusion of Results

Studies using LLM-based analysis prove that social media profoundly affects cognition through three key impacts including mental deficits and memory distortions and emotional regulation problems. The predictive abilities of GPT-4 and Mental-FLAN-T5 remain strong yet several obstacles persist in their generalization power and explanation capabilities and moral implementation methods. Social media needs interdisciplinary management systems to reduce possible risks while using it for cognitive health analysis.

References

[1] K. Yang, T. Zhang, Z. Kuang, Q. Xie, J. Huang, and S. Ananiadou, “Mental- lama: Interpretable mental health analysis on social media with large language models,” in Proceedings of the ACM Web Conference 2024, WWW ’24, (New York, NY, USA), p. 4489–4500, Association for Computing Machinery, 2024.

[2] X. Xu, B. Yao, Y. Dong, S. Gabriel, H. Yu, J. Hendler, M. Ghassemi, A. K. Dey, and D. Wang, “Mental-llm: Leveraging large language models for mental health prediction via online text data,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 8, p. 1–32, Mar. 2024.

[3] Y. Hua, F. Liu, K. Yang, Z. Li, H. Na, Y. han Sheu, P. Zhou, L. V. Moran, S. Ananiadou, A. Beam, and J. Torous, “Large language models in mental health care: a scoping review,” 2024.

[4] H. R. Lawrence, R. A. Schneider, S. B. Rubin, M. J. Matari´c, D. J. McDuff, and M. Jones Bell, “The opportunities and risks of large language models in mental health,” JMIR Mental Health, vol. 11, p. e59479–e59479, July 2024.

[5] N. Rostamzadeh, D. Mincu, S. Roy, A. Smart, L. Wilcox, M. Pushkarna, J. Schrouff, R. Amironesei, N. Moorosi, and K. Heller, “Healthsheet: Develop- ment of a transparency artifact for health datasets,” in Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT ’22, (New York, NY, USA), p. 1943–1961, Association for Computing Machinery, 2022.

[6] A. Radwan, M. Amarneh, H. Alawneh, H. I. Ashqar, A. A. A. R. Magableh, and A. AlSobeh, “Predictive analytics in mental health leveraging llm embed- dings and machine learning models for social media analysis,” Int. J. Web Serv. Res., vol. 21, p. 1–22, Feb. 2024. 27

[7] V. Vajre, M. Naylor, U. Kamath, and A. Shehu, “Psychbert: A mental health language model for social media mental health behavioral analysis,” in 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1077–1082, 2021.

[8] D. Owen, A. J. Lynham, S. E. Smart, A. F. Pardi˜nas, and J. Camacho Col- lados, “Ai for analyzing mental health disorders among social media users: Quarter-century narrative review of progress and challenges,” J Med Internet Res, vol. 26, p. e59225, Nov 2024.

[9] S. Ji, T. Zhang, L. Ansari, J. Fu, P. Tiwari, and E. Cambria, “MentalBERT: Publicly available pretrained language models for mental healthcare,” in Pro- ceedings of the Thirteenth Language Resources and Evaluation Conference (N. Calzolari, F. B´echet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis, eds.), (Marseille, France), pp. 7184–7190, European Language Resources Associa- tion, June 2022.

[10] H. Qi, Q. Zhao, J. Li, C. Song, W. Zhai, D. Luo, S. Liu, Y. J. Yu, F. Wang, H. Zou, B. X. Yang, and G. Fu, “Supervised learning and large language model benchmarks on mental health datasets: Cognitive distortions and suicidal risks in chinese social media,” 2024.

[11] M. S. Rahaman, M. M. T. Ahsan, N. Anjum, H. J. R. Terano, and M. M. Rahman, “From chatgpt-3 to gpt-4: A significant advancement in ai-driven nlp tools,” Journal of Engineering and Emerging Technologies, 2023.

[12] A. Alsharef, K. Aggarwal, Sonia, D. Koundal, H. Alyami, D. Ameyed, and C. De Maio, “An automated toxicity classification on social media using lstm and word embedding,” Intell. Neuroscience, vol. 2022, Jan. 2022.

[13] A. Al-Shraifin, R. B. Arabiat, A. M. Shatnawi, A. Alsobeh, and N. Bahr, “The effectiveness of a counseling program based on psychosocial support to raise the level of economic empowerment among refugees,” Current Psychology, pp. 1–10, 2023.

[14] S. Alshattnawi, L. Afifi, A. M. Shatnawi, and M. M. Barhoush, “Utilizing ge- netic algorithm and artificial bee colony algorithm to extend the wsn lifetime,” 28 Int. J. Comput., vol. 21, pp. 25–31, 2022.

[15] A. M. R. Alsobeh, R. Hammad, and A. A. Tamimi, “A modular cloud-based ontology framework for context-aware ehr services,” Int. J. Comput. Appl. Technol., vol. 60, pp. 339–350, 2019.

[16] S. Banerjee, P. Dunn, S. Conard, and A. Ali, “Mental health applications of generative ai and large language modeling in the united states,” International Journal of Environmental Research and Public Health, vol. 21, no. 7, p. 910, 2024.

[17] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Win- ter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Lan- guage models are few-shot learners,” 2020.

[18] A. R. Mahlous and B. Okkali, “A digital mental health intervention for chil- dren and parents using a user-centred design,” Adv. Hum. Comput. Interact., vol. 2022, pp. 4322177:1–4322177:12, 2022.

[19] A. Ruch, “Can x2vec save lives? integrating graph and language embeddings for automatic mental health classification,” Journal of Physics: Complexity, vol. 1, p. 035005, Aug. 2020.

[20] H. Sampasa-Kanyinga and R. Lewis, "Frequent use of social networking sites is associated with poor psychological functioning among children and adoles- cents," Cyberpsychology, behavior and social networking, vol. 18, pp. 380– 385, 07 2015.

[21] D. Luxton, J. June, and J. Fairall, "Social media and suicide: A public health perspective," American journal of public health, vol. 102 Suppl 2, pp. S195– 200, 03 2012.

[22] A. Le Glaz, Y. Haralambous, D.-H. Kim-Dufor, P. Lenca, R. Billot, T. C. Ryan, J. Marsh, J. DeVylder, M. Walter, S. Berrouiguet, and C. Lemey, "Machine learning and natural language processing in mental health: Systematic review," J Med Internet Res, vol. 23, p. e15708, May 2021. 29

[23] X. Zhang, W. Li, H. Ying, F. Li, S. Tang, and S. Lu, "Emotion detection in on- line social networks: A multilabel learning approach," IEEE Internet of Things Journal, vol. 7, pp. 8133–8143, 2020.

[24] X. Zhang, W. Li, H. Ying, F. Li, S. Tang, and S. Lu, "Emotion detection in online social networks: A multi-label learning approach," IEEE Internet of Things Journal, vol. PP, pp. 1–1, 06 2020.

[25] Z. Jiang, S. I. Levitan, J. Zomick, and J. Hirschberg, "Detection of mental health from Reddit via deep contextualized representations," in Proceedings of the 11th International Workshop on Health Text Mining and Information Anal- ysis (E. Holderness, A. Jimeno Yepes, A. Lavelli, A.-L. Minard, J. Pustejovsky, and F. Rinaldi, eds.), (Online), pp. 147–156, Association for Computational Linguistics, Nov. 2020.

[26] C. Luna-Jim´enez, M. Gil-Mart´ın, L. F. D'Haro, F. Fern´andez-Mart´ınez, and R. San-Segundo, "Evaluating emotional and subjective responses in synthetic art-related dialogues: A multi-stage framework with large language models," Expert Systems with Applications, vol. 255, p. 124524, 2024.

[27] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, and X. Xie, "A survey on evaluation of large language models," 2023.

[28] Y. Yuan, Y. Liu, and L. Cheng, "A multi-faceted evaluation framework for assessing synthetic data generated by large language models," 2024.

[29] K. Zhang, X. Meng, X. Yan, J. Ji, J. Liu, H. Xu, H. Zhang, D. Liu, J. Wang, X. Wang, J. Gao, Y.-g.-s. Wang, C. Shao, W. Wang, J. Li, M.-Q. Zheng, Y. Yang, and Y.-D. Tang, "Revolutionizing health care: The transformative impact of large language models in medicine," J Med Internet Res, vol. 27, p. e59069, Jan 2025.

[30] M. Nadeem, "Identifying depression on twitter," CoRR, vol. abs/1607.07384, 2016.

[31] T. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez, and C. Giraud- Carrier, "Detecting cognitive distortions through machine learning text an- 30 alytics," in 2017 IEEE International Conference on Healthcare Informatics (ICHI), pp. 508–512, 2017.

[32] S. Shreevastava and P. Foltz, "Detecting cognitive distortions from patient- therapist interactions," in Proceedings of the Seventh Workshop on Computa- tional Linguistics and Clinical Psychology: Improving Access (N. Goharian, P. Resnik, A. Yates, M. Ireland, K. Niederhoffer, and R. Resnik, eds.), (On- line), pp. 151–158, Association for Computational Linguistics, June 2021.

[33] Z. Ma, Y. Mei, and Z. Su, "Understanding the benefits and challenges of us- ing large language model-based conversational agents for mental well-being support," 2023.

[34] S. Ji, X. Li, Z. Huang, and E. Cambria, "Suicidal ideation and mental disorder detection with attentive relation networks," Neural Computing and Applica- tions, vol. 34, p. 10309–10319, June 2021.

[35] R. Safa, S. A. Edalatpanah, and A. Sorourkhah, "Predicting mental health using social media: A roadmap for future development," 01 2023.

[36] A.-M. Bucur, A. Cosma, and L. P. Dinu, "Early risk detection of pathological gambling, self-harm and depression using bert," 2021.

[37] "Kaggle twitter sentiments," https://www.kaggle.com/ywang311/ twitter-sentiment/data.