Mental Health Analysis Based on Social Media Data

Dr. Chandni Saxena
The Chinese University of Hong Kong, SAR China

Big-data-analytics in Astronomy, Science and Engineering
11th International Conference Theme - Data Science and Applications

Background of Mental Health

Dimensions of Health

Image Source: pdhpe.net

- "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity"

–W.H.O. (1948)

- "Health is the condition of being sound in body, mind, or spirit"

– Webster

Mental Health

- A good mental health is a state of overall well-being
- Mental health is a state of balance between body and mind.

Image Source: workstars.com

Mental Illness: Types and Effects

- According to American Psychiatric Association, there are hundreds of mental illness listed in the Manual of Mental Disorders
- Mental illness has a negative effect on the way an individual **thinks**, **feels** and **behaves**

Theresa Chiechi / Verywell

Source: verywellhealth.com

Mental Health: A Global Issue

NHS Providers says the amount spent on mental health care needs to rise from £14.3bn to at least £17.15bn from next year to help cope with demand. Photograph: Microgen Images/Science Photo Library/Getty Images

- As per reports released in August 2021[1], 1.6 million people in England were on waiting lists for mental health care
- Estimated 8 million people could not get specialist help as they were not considered sick enough to qualify
- According to the World Health Organization[2], a person commits suicide every 11.1 minutes in the US[1] and 23% of deaths in the world are associated with mental disorders

- The statistics above induce the need for automation in mental healthcare

[1] https://www.theguardian.com/society/2021/aug/29/strain-on-mental-health-care-leaves-8m-people-without-help-say-nhs-leaders
[2] https://suicidology.org/wp-content/uploads/2021/01/2019datapgsv2b.pdf

Agenda

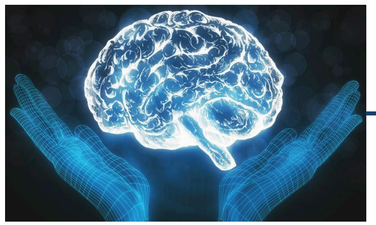
- Role of AI in mental health analysis (MHA)
- Features from social media data
- Text-based MHA
- NLP datasets for MHA
- Ethics and future directions

Role of AI in Mental Health Analysis (MHA)

AI for MHA

Traditional Clinical Methods

AI applied for automatic detection of Mental Health issues



- Clinical Interview
- Personal Journal
- Electronics Health Records
- Questionnaire
- Social Media Content
  - Reddit
  - Twitter
  - Weibo
  - Instagram
  - Facebook
  - ...

Why Social Media?

Social Media is a medium of self-disclosure

- Social media platform is a medium of self-disclosure with user-generated data (feelings, thoughts and emotions)
- 80% of patients do not undergo psychological treatments with mental health practitioners
- Self-disclosure is a significant curative component of social well-being[1]
- Research community emphasizes on the use of **Computational Intelligence Techniques** for Mental Health Analysis on Social Media

1 Jourard, S. M. (1959). Healthy personality and self-disclosure. Mental Hygiene. New York.

Social Media Data

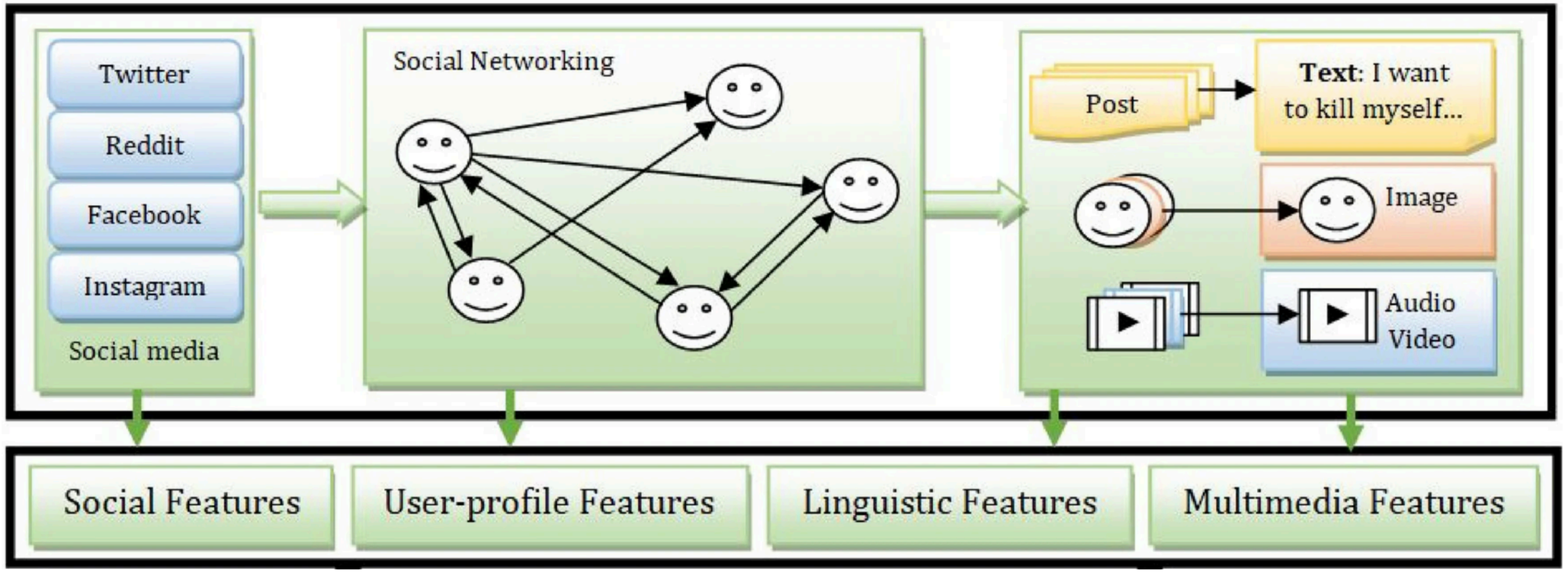
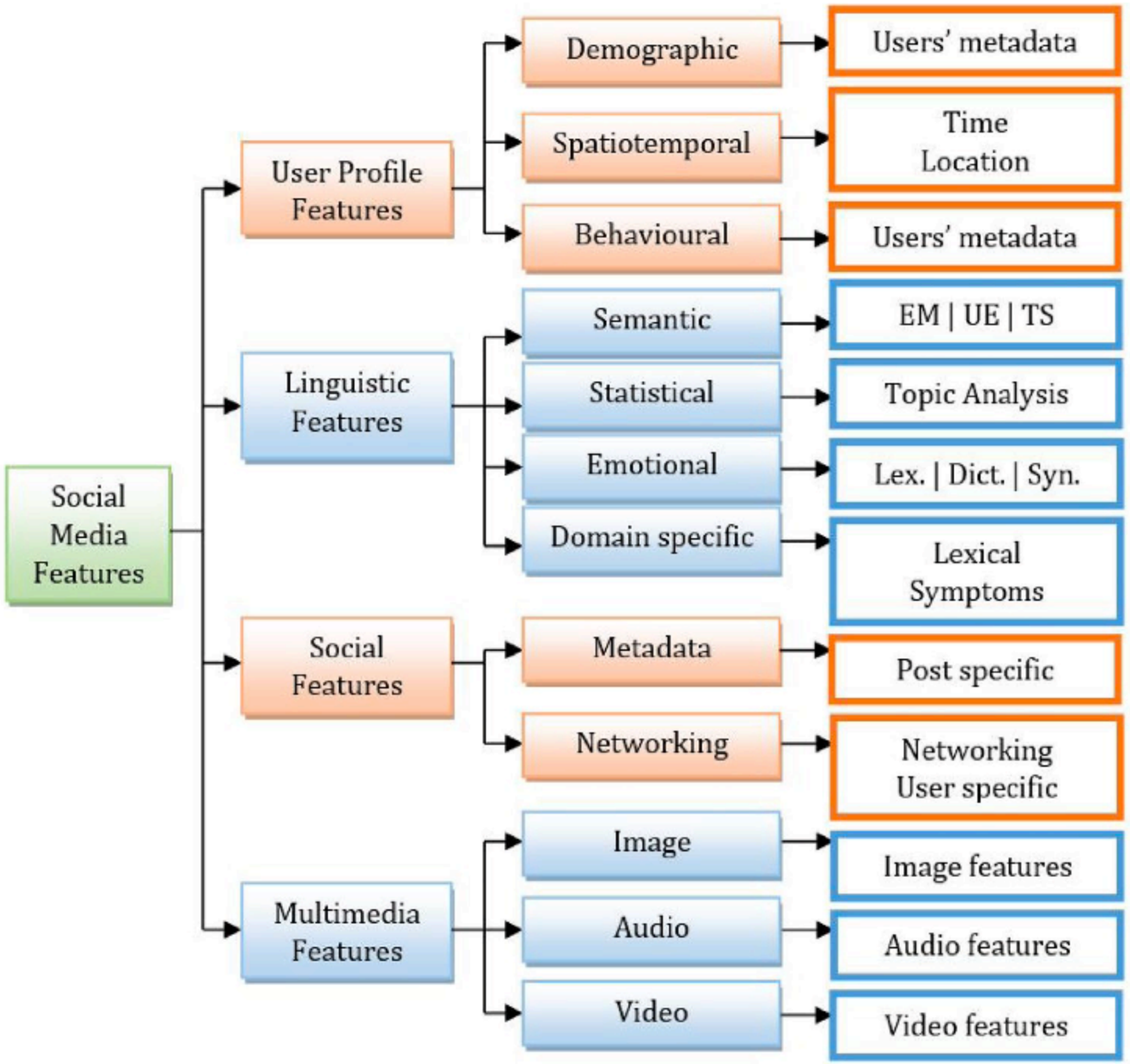


Image: Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819-1842.

- AI models are build on features extracted from data such as the handcrafted features, the statistical information, and automated features to name a few

Features from Social Media Data

Social Media Features

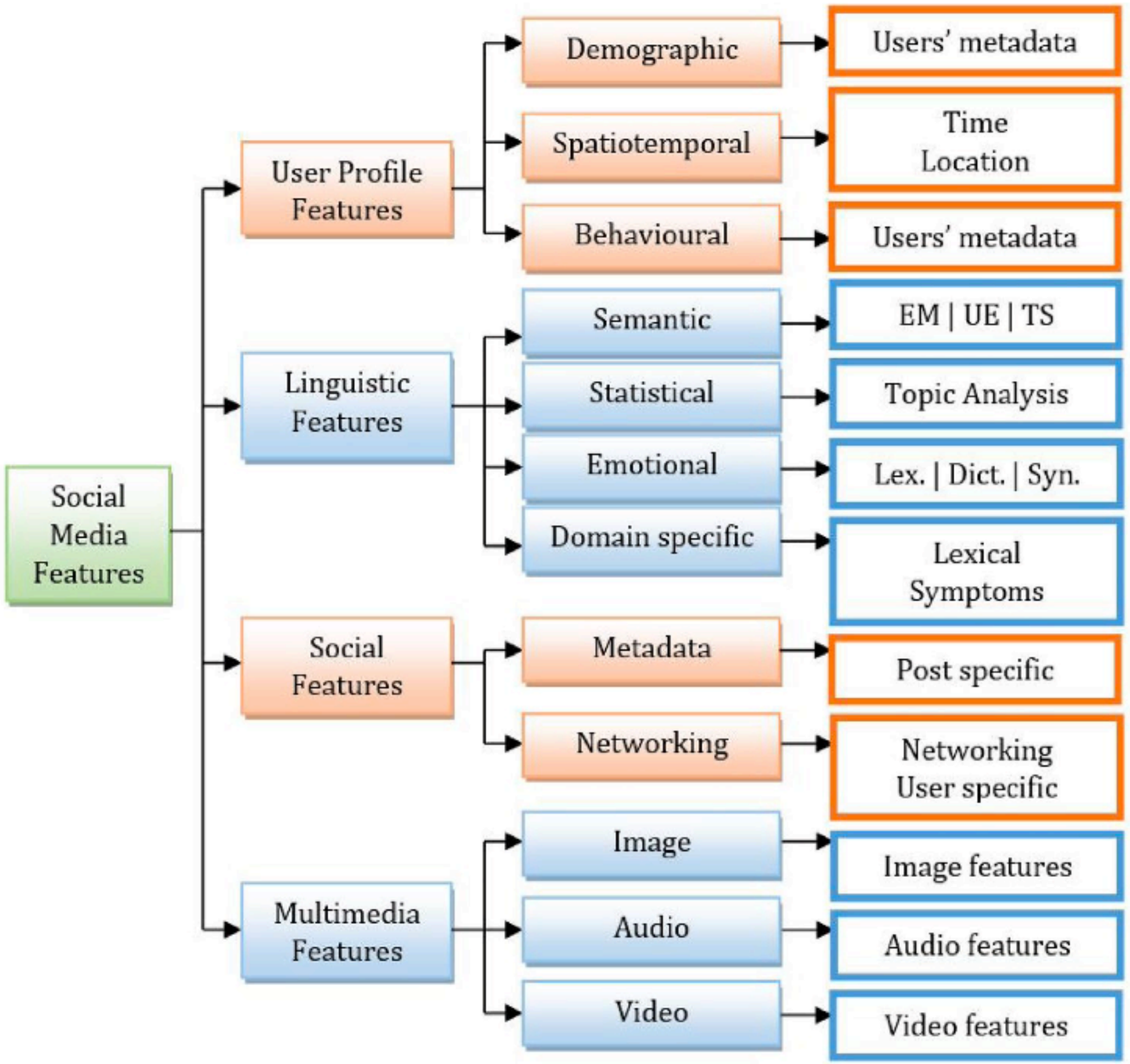


User Profile Feature Extraction

Category	Sub-category	Feature
Demographic	Users' Meta-data	Age
		Gender
		Education
		Occupation
	Users' Network	Follower
		Ego-network
Spatio-temporal	Temporal	Timeline
	Spatial	Location
Behavioural	Posting Behavior	Gen. Behaviour
		Ruminative
		Posting Time

Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819-1842.

Social Media Features

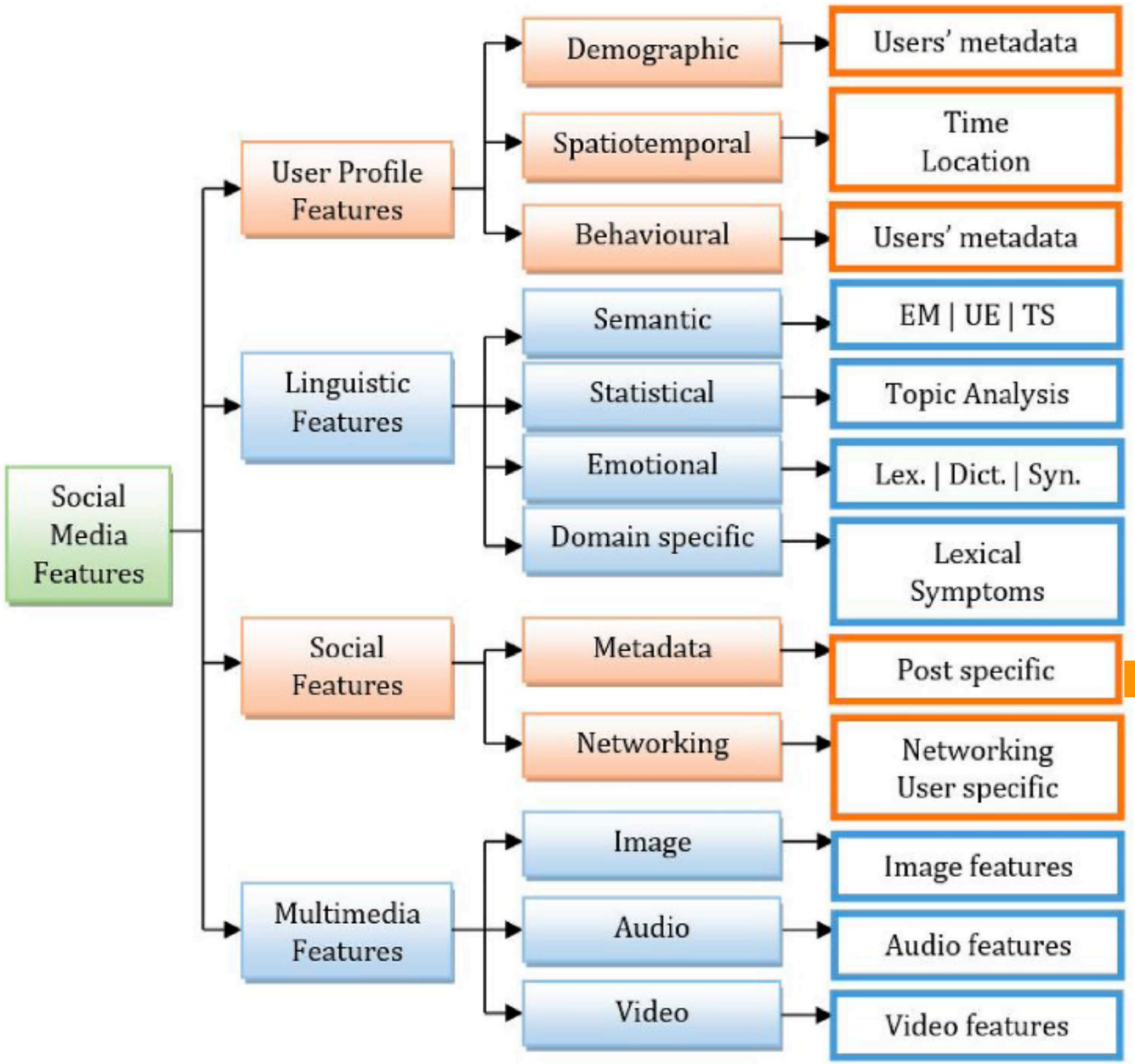


Linguistic Feature Extraction

Category	Sub-category	Feature
Emotional	Emotional Model	EmoBERT
		MentalBERT
	Textual Sentiments	Emoji
		Emoticons
		SentiWordNet
		SentiNet
Semantic	Topic Analysis	LDA
		Brown Clustering
Statistical	Lexical	TFIDF
		Text
		Morphological
		Stylometric
		n-gram
		Punctuation
	Dictionary	LIWC
		Suicide Dictionary
		ANEW
	Syntactical	POS Tagging
	Syntactical	POS Tagging
Domain Specific	Lexicon	Antidepressant
		TensiStrength
		Dictionaries
	Dep. Symptoms	DSM
		Plutchik
		VAD
		Affect and Intensity
		Big 5 Personality
		Anxiety, Anger, Dep.

Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819-1842.

Social Media Features

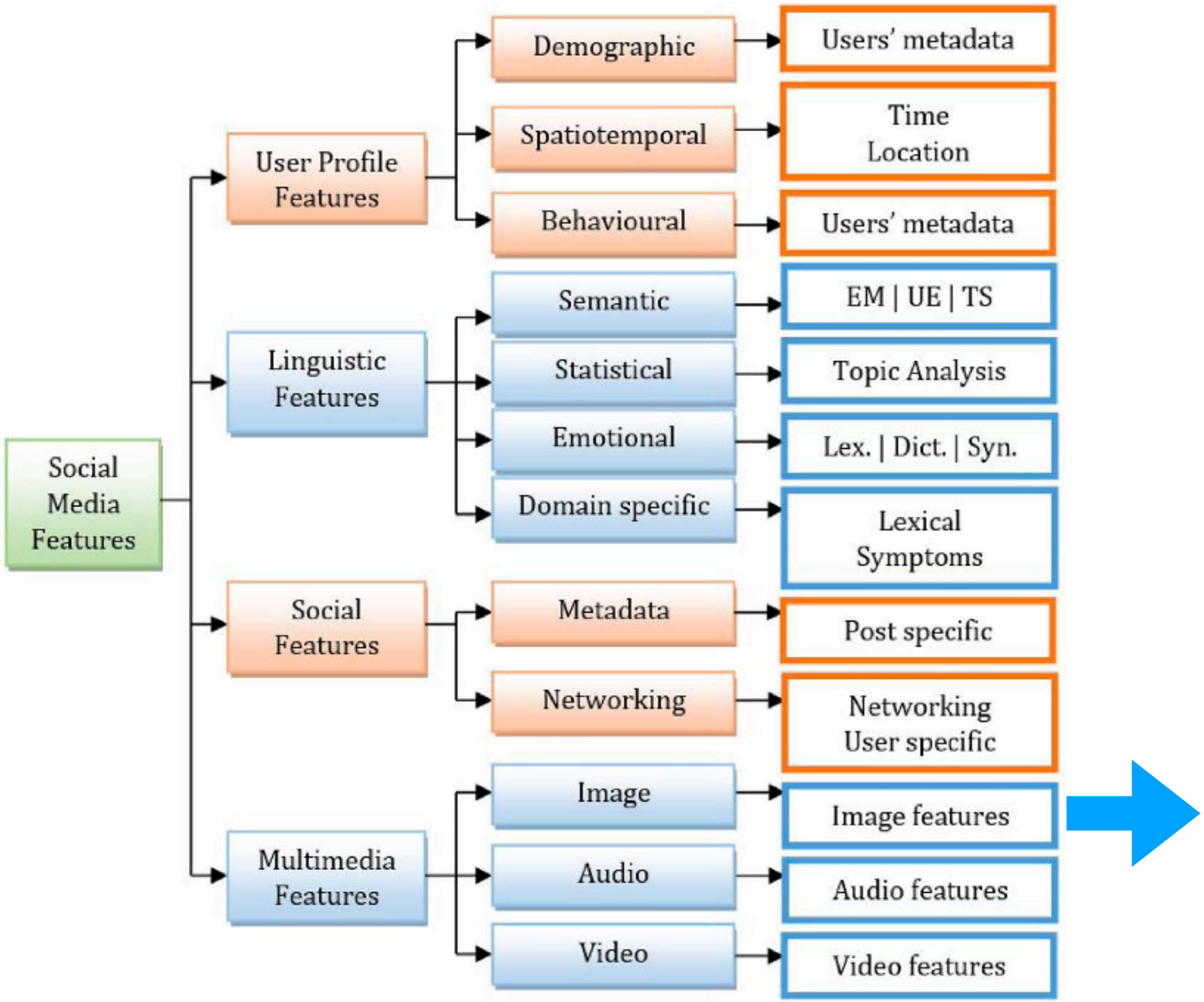


Social Feature Extraction

Category	Sub-category	Feature
Social Metadata	Post Specific	Length
		#(Hashtags)
		#(URL)
		Metadata
Social Network	Networking	Interactions
		At-Mentions
		Replied to
	User Specific	#(Favourites)
		#(Likes)
		#(Posts)
		#(Comments)
		#(ReTweet)

Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819-1842.

Social Media Features



Multimodal Feature Extraction

Category	Feature
Image	Colour Combinations
	Colour Ratio
	Brightness
	Saturation
	Convolution

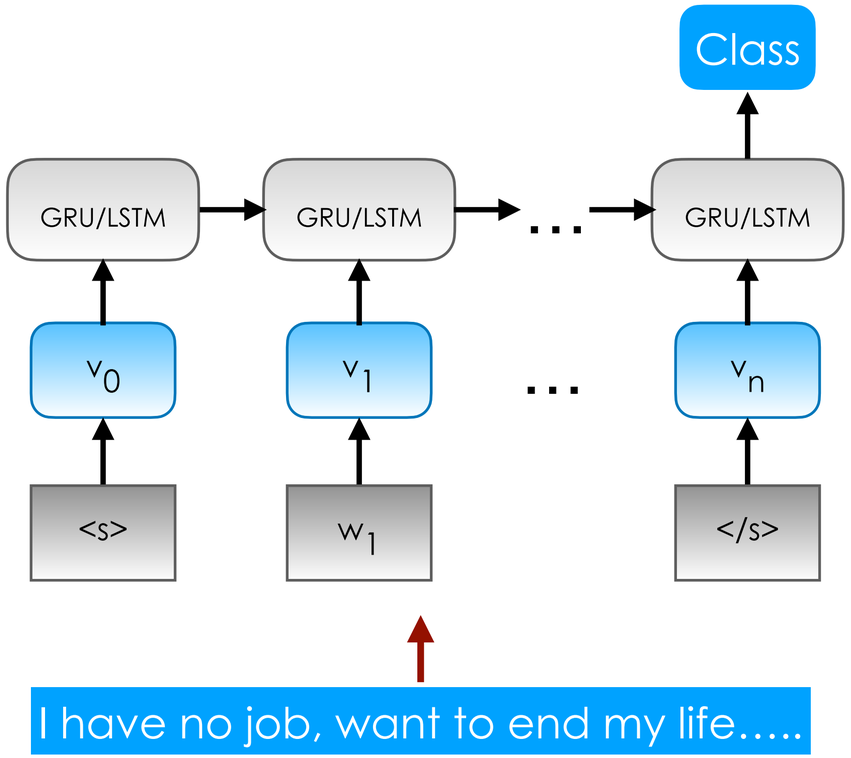
Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819-1842.

Text-based MHA

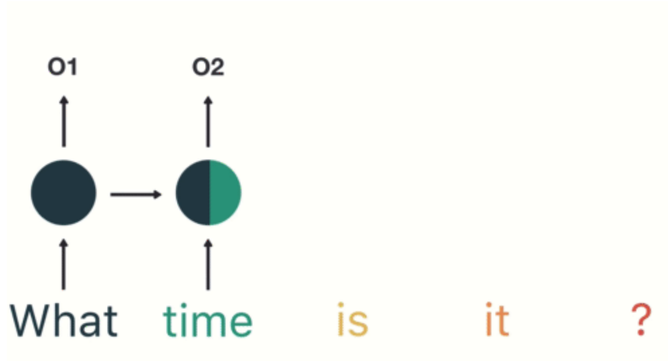
RNN-based Methods

- LSTM and GRU
  - LSTM with transfer learning
  - LSTM /GRU with multi-task learning
  - LSTM/GRU with reinforcement learning

LSTM/GRU Models

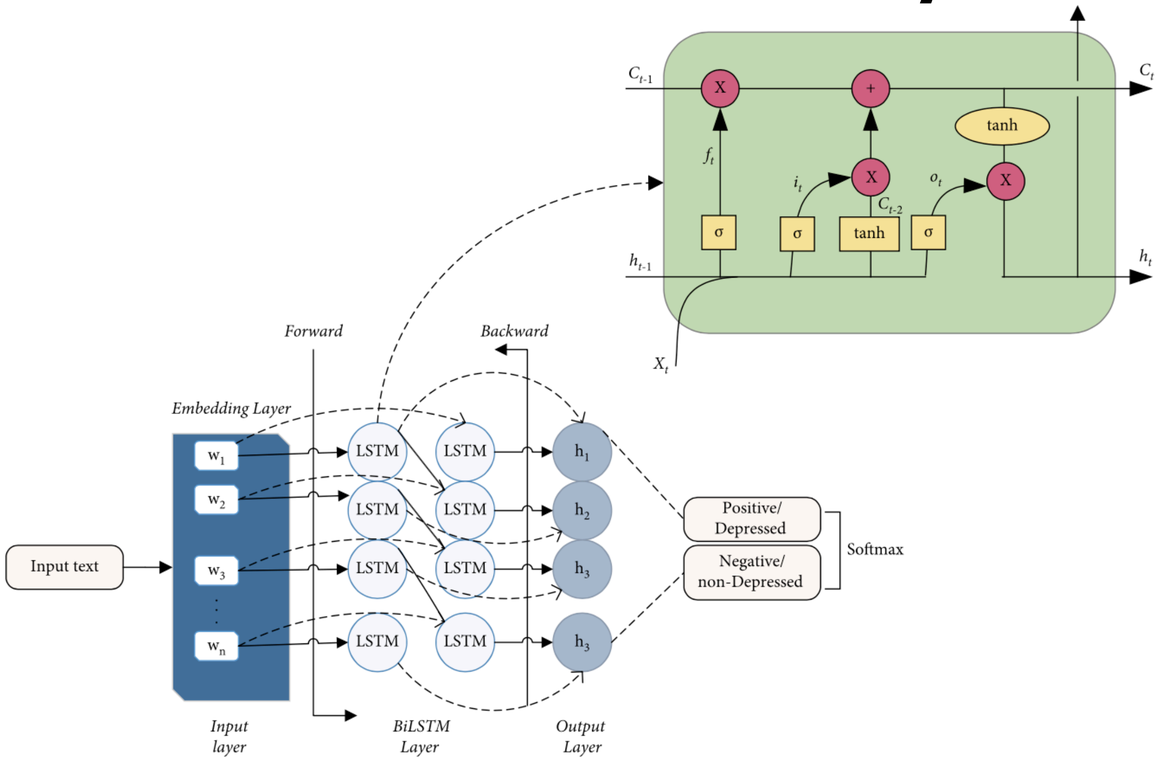


- LSTM/GRU architecture preserve context and capture sequential dependencies in the text



- Ghosh, S., & Anwar, T. (2021). Depression intensity estimation via social media: a deep learning approach. *IEEE Transactions on Computational Social Systems*, *8*(6), 1465-1474.
- Yao, X., Yu, G., Tang, J., & Zhang, J. (2021). Extracting depressive symptoms and their associations from an online depression community. *Computers in human behavior*, *120*, 106734.
- Gui, T., Zhang, Q., Zhu, L., Zhou, X., Peng, M., & Huang, X. (2019). Depression detection on social media with reinforcement learning. In *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18* (pp. 613-624). Springer International Publishing
- Gui, T., Zhu, L., Zhang, Q., Peng, M., Zhou, X., Ding, K., & Chen, Z. (2019). Cooperative multimodal approach to depression detection in twitter. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 110-117).
- Garg, M., Saxena, C., Krishnan, V., Joshi, R., Saha, S., Mago, V., & Dorr, B. J. (2022) CAMS: An Annotated Corpus for Causal Analysis of Mental Health Issues in Social Media Posts. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 6387-6396).

LSTM/GRU Models



(A bi-LSTM model)

Image: Zeberga et al. A novel text mining approach for mental health prediction using Bi-LSTM and BERT model. *Computational Intelligence and Neuroscience*, 2022.

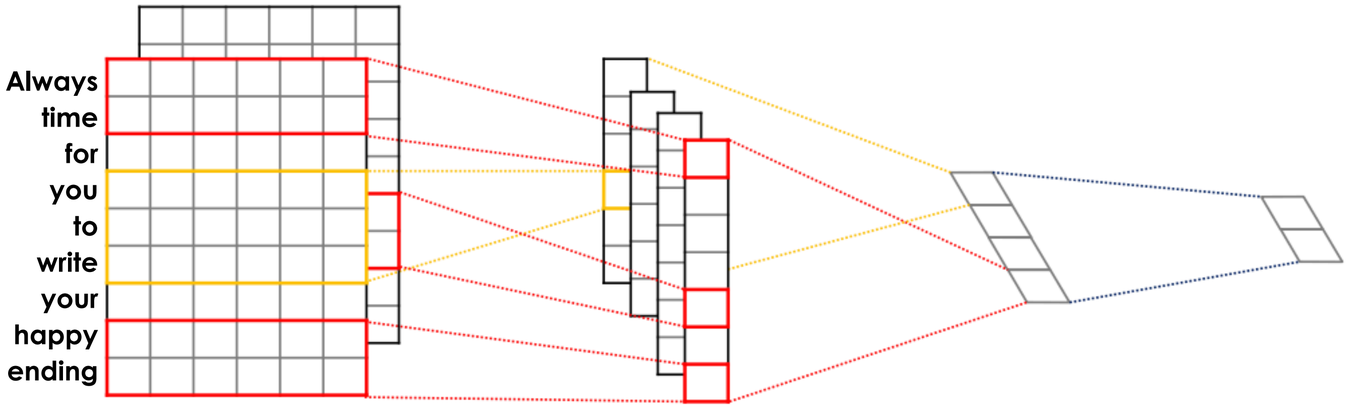
- Bi_LSTM preserves past and future context of long-term word relationship in social media data
- Integrating these architectures with multitask learning or reinforcement learning further enhances mental health analysis

- Ghosh, S., & Anwar, T. (2021). Depression intensity estimation via social media: a deep learning approach. *IEEE Transactions on Computational Social Systems*, *8*(6), 1465-1474.
- Yao, X., Yu, G., Tang, J., & Zhang, J. (2021). Extracting depressive symptoms and their associations from an online depression community. *Computers in human behavior*, *120*, 106734.
- Gui, T., Zhang, Q., Zhu, L., Zhou, X., Peng, M., & Huang, X. (2019). Depression detection on social media with reinforcement learning. In *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18* (pp. 613-624). Springer International Publishing
- Gui, T., Zhu, L., Zhang, Q., Peng, M., Zhou, X., Ding, K., & Chen, Z. (2019, July). Cooperative multimodal approach to depression detection in twitter. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 110-117).
- Garg, M., Saxena, C., Krishnan, V., Joshi, R., Saha, S., Mago, V., & Dorr, B. J. (2022) CAMS: An Annotated Corpus for Causal Analysis of Mental Health Issues in Social Media Posts. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 6387-6396).

CNN-based Methods

- Basic CNN
- GNN

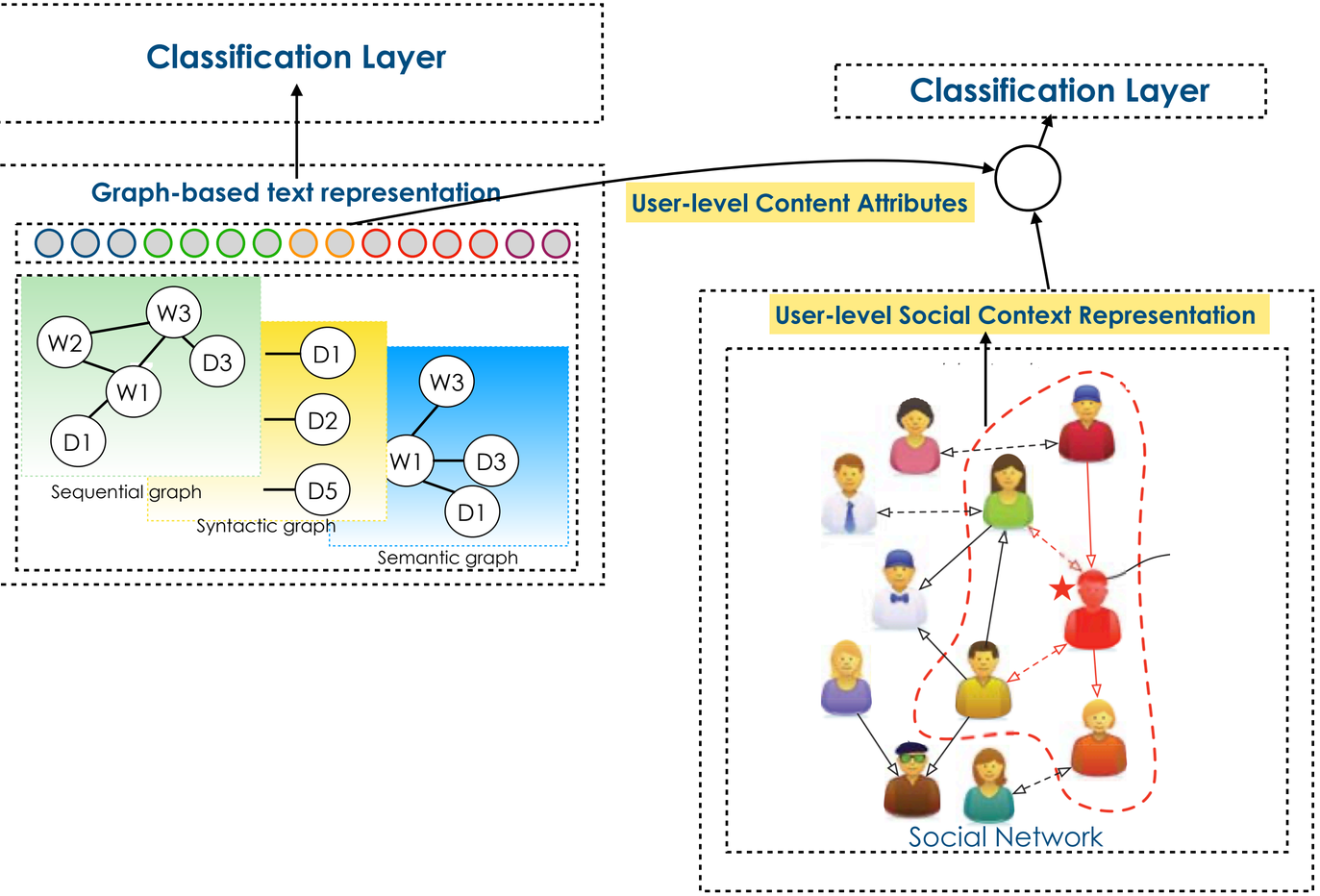
Basic CNN Model



- Typical CNN architecture consists of a convolutional layer, pooling layer, and fully connected layer
- Different approaches utilize unified hybrid models that also combine CNN

- Gaur, M. et al. Knowledge-aware assessment of severity of suicide risk for early intervention. In The World Wide Web Conference, pp. 514–525 (2019).
- Boukil, S., El Adnani, F., Cherrat, L., El Moutaouakkil, A. E. & Ezziyyani, M. Deep learning algorithm for suicide sentiment prediction. In International Conference on Advanced Intelligent Systems for Sustainable Development, pp. 261–272 (2018).
- Phan, H. T., Tran, V. C., Nguyen, N. T. & Hwang, D. A framework for detecting user's psychological tendencies on twitter based on tweets sentiment analysis. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 357–372 (2020).
- Wang, Y. -T., Huang, H. -H., Chen, H. -H. & Chen, H. A neural network approach to early risk detection of depression and anorexia on social media text. In CLEF (Working Notes) (2018).

GNN-based Methods



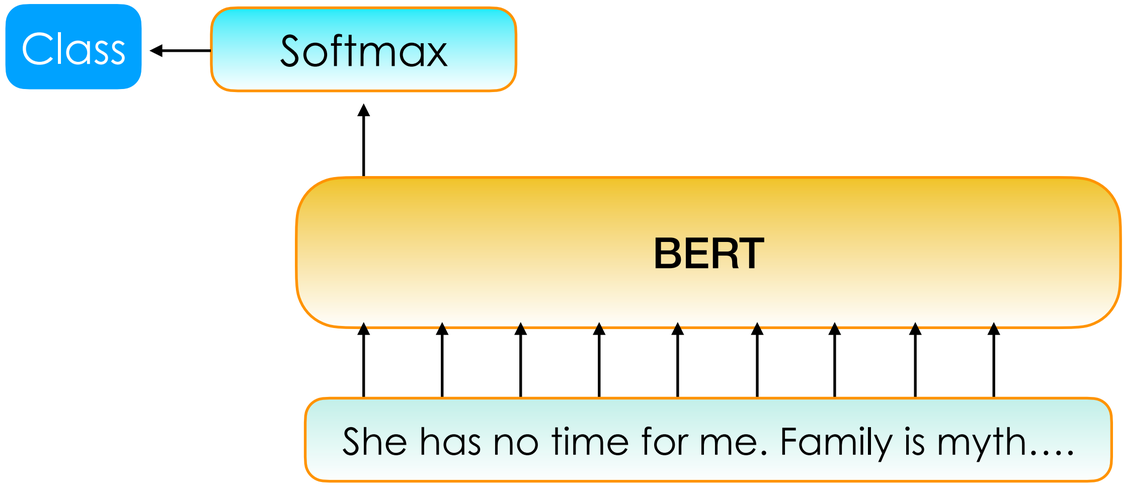
- GNNs operate on graph-structured data and can capture dependencies and relationships between words or entities in a text
- They also provide a flexible framework for learning diverse graph-based user context representations

- Naseem, U., Kim, J., Khushi, M., & Dunn, A. (2023, April). Graph-Based Hierarchical Attention Network for Suicide Risk Detection on Social Media. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 995-100
- Sawhney, R., Joshi, H., Shah, R., & Flek, L. (2021, June). Suicide ideation detection via social and temporal user representations using hyperbolic learning. In *Proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies* (pp. 2176-2190).

Transformer-based Methods

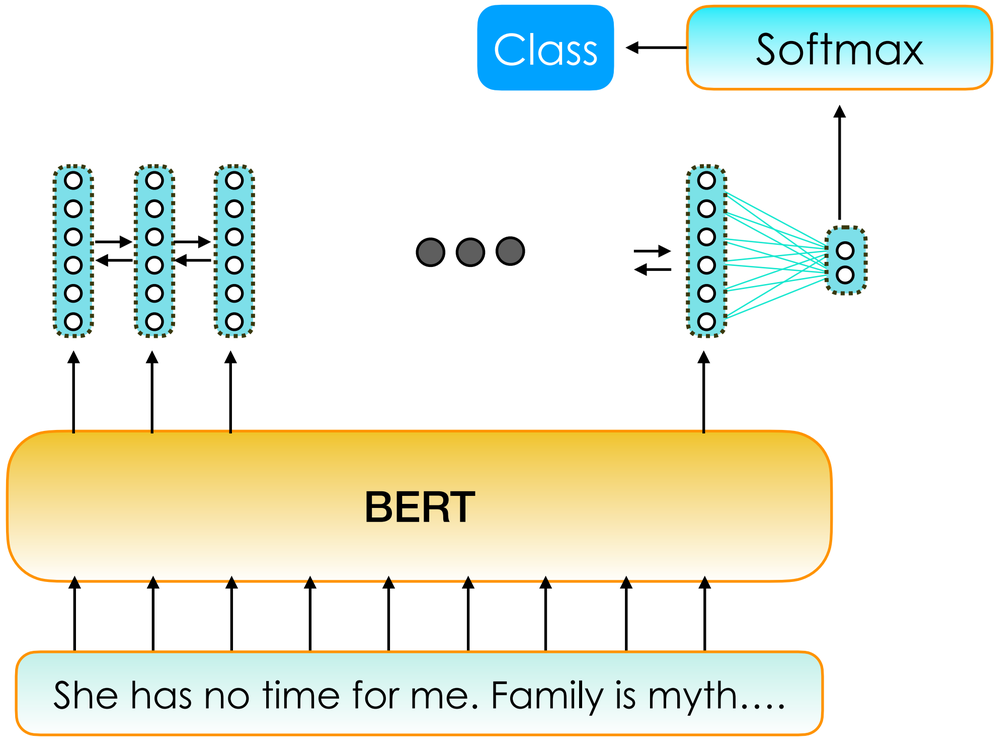
- BERT-based model
- Large Language Models (LLMs)

BERT-based Models

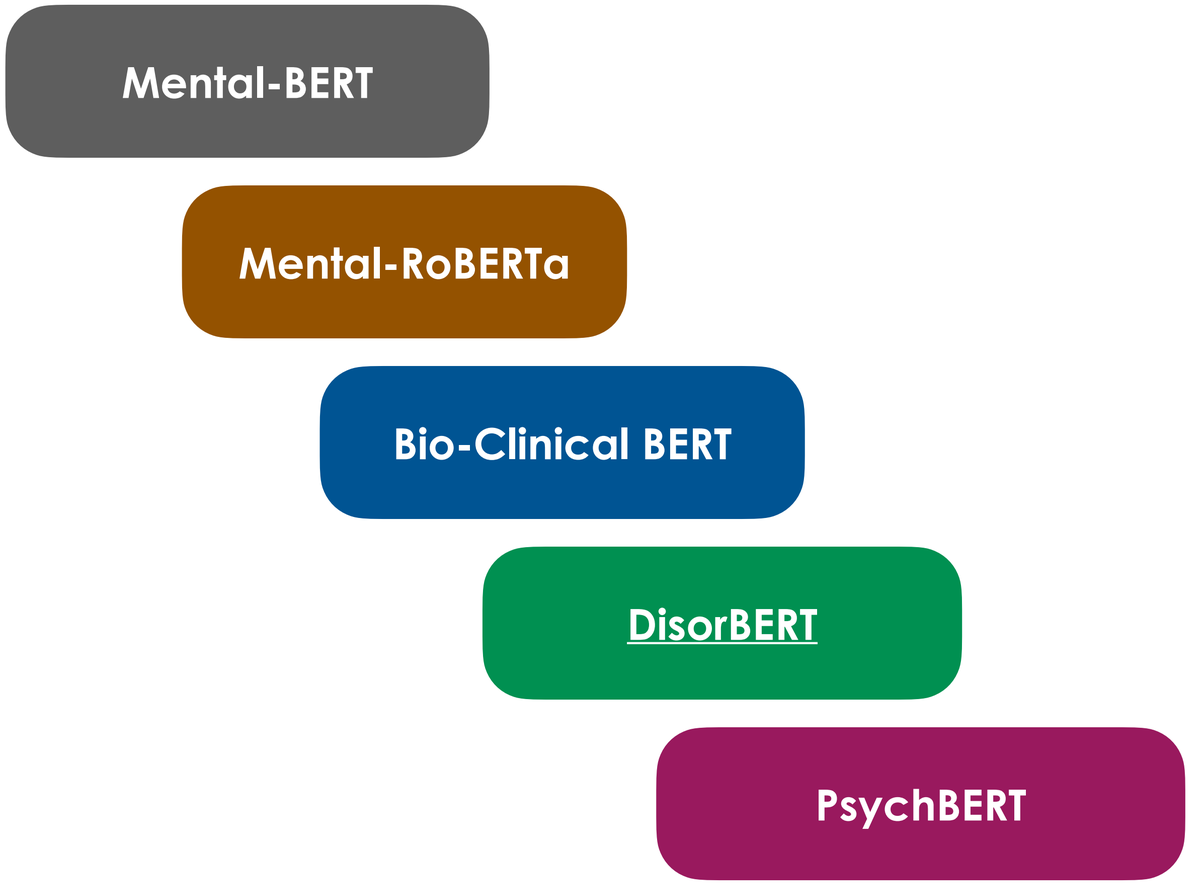


- One popular approach is to use BERT as a classifier, or fine-tune the BERT model on the specific text classification task

- Feature representations obtained from BERT can be further processed through additional recurrent or convolutional layers to capture additional text patterns and dependencies



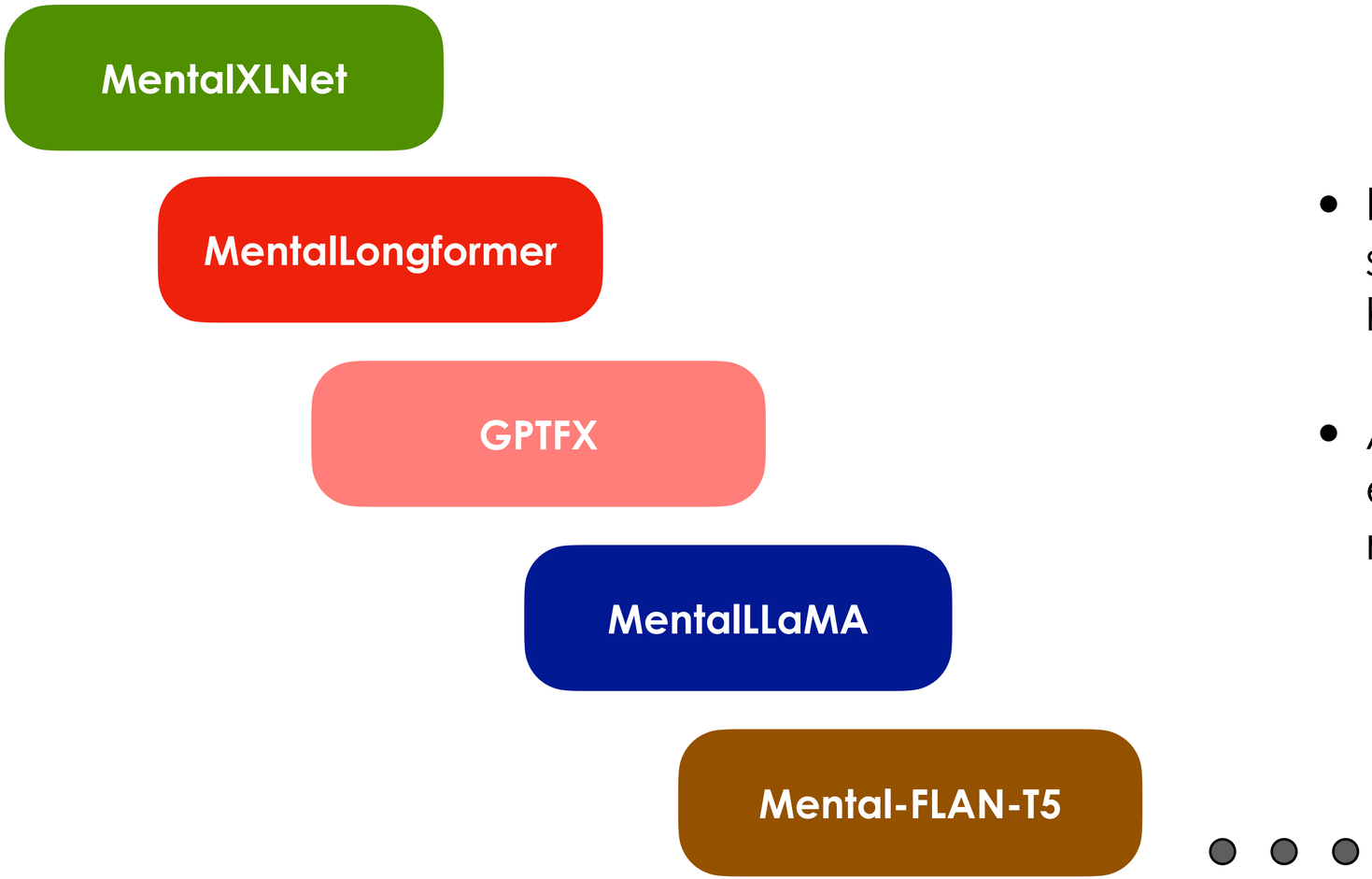
Domain-adapted BERT Models



- Domain specific, fine-tuned BERT models excel in mental health classifications, showcasing their versatility and effectiveness in analyzing text data pertaining to mental health

- Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2022, June). MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 7184-7190)
- Kshatriya, B. S. A., Nunez, N. A., Resendez, M. G., Ryu, E., Coombes, B. J., Fu, S., ... & Wang, Y. (2021). Neural language models with distant supervision to identify major depressive disorder from clinical notes. *arXiv*
- Aragon, Mario, et al. "DisorBERT: A Double Domain Adaptation Model for Detecting Signs of Mental Disorders in Social Media." *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2023.
- Vajre, Vedant, et al. "PsychBERT: a mental health language model for social media mental health behavioral analysis." *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021.

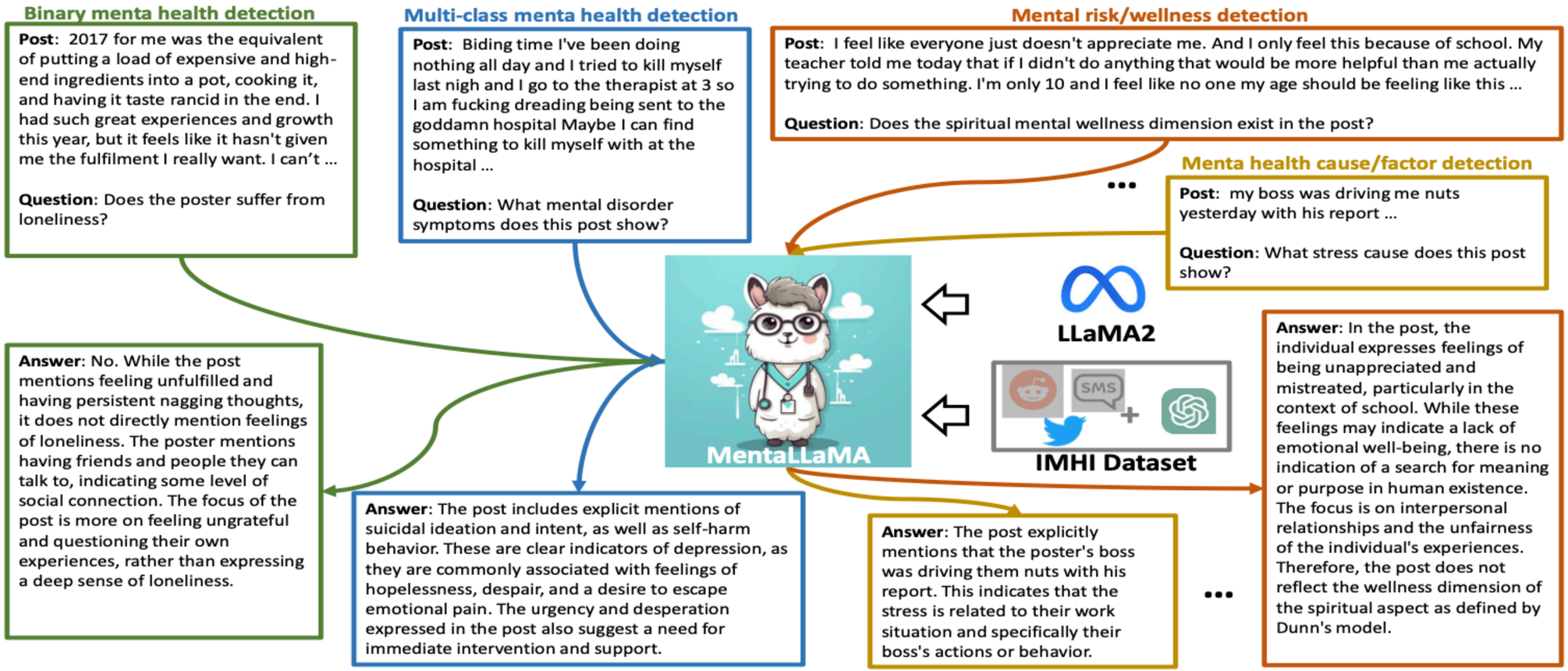
Large Language Models



- Large language models (LLMs) achieve state-of-the-art performance for the mental health prediction task
- Additionally, they fine-tuned for generating explanations related to the predictions made by these machine learning models

- Zhang, Tianlin, Kailai Yang, and Sophia Ananiadou. "Sentiment-guided Transformer with Severity-aware Contrastive Learning for Depression Detection on Social Media." *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*. 2023
- Mazumdar, Hirak, et al. "GPTFX: A Novel GPT-3 Based Framework for Mental Health Detection and Explanations." *IEEE Journal of Biomedical and Health Informatics* (2023).
- Aragon, Mario, et al. "DisorBERT: A Double Domain Adaptation Model for Detecting Signs of Mental Disorders in Social Media." *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2023.
- Xu, Xuhai, et al. "Leveraging large language models for mental health prediction via online text data." *arXiv preprint arXiv:2307.14385* (2023).

Large Language Models



- LLMs, owing to their capability for generating explanations, achieve remarkable proficiency in producing high-quality explanations

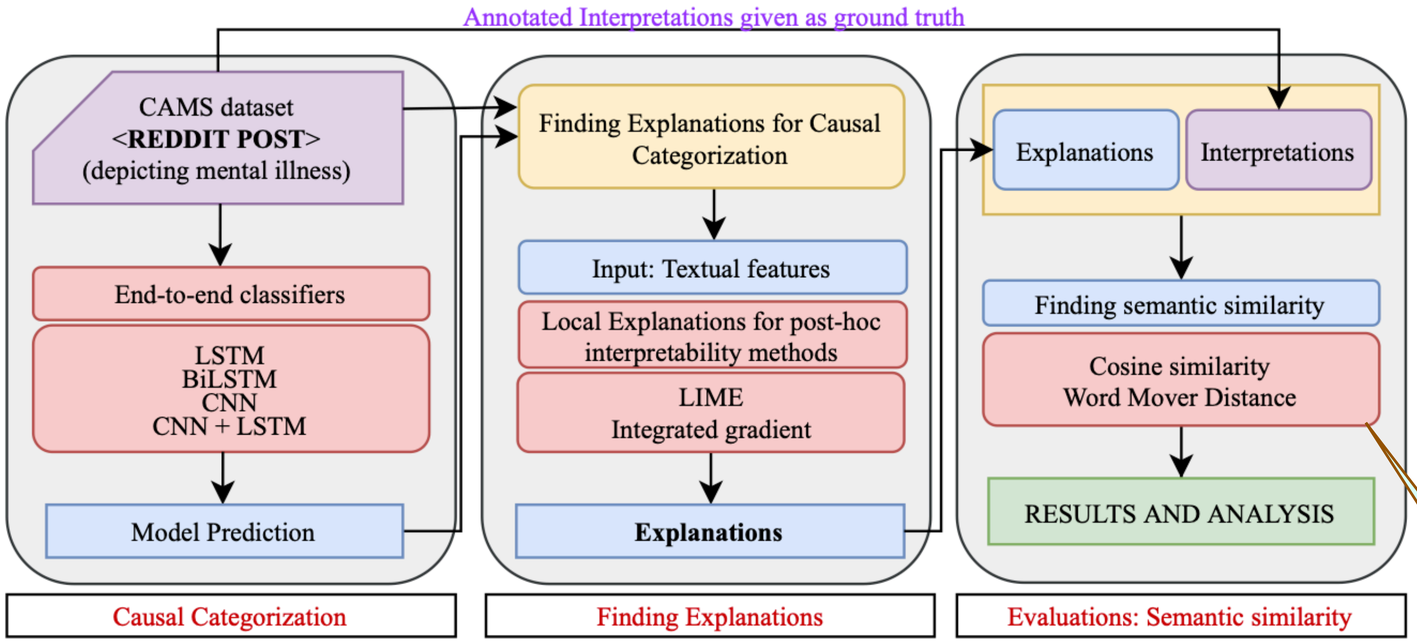
Image Source: Yang, Kailai, et al. "Mentalllama: Interpretable mental health analysis on social media with large language models." *arXiv preprint arXiv:2309.13567* (2023).

https://github.com/SteveKGYang/MentalLLaMA

Interpretable Mental Health Analysis

- Explainable Causal Analysis
- Generative LLMs for Explanation

Explainable Causal Analysis



- We proposed a framework for explainable causal analysis of mental health on social media data
- The framework is divided into three phases- Phase 1: Causal categorization, Phase 2: Finding explanations, Phase 3: Semantic similarity

Saxena, Chandni, Muskan Garg, and Gunjan Ansari. "Explainable causal analysis of mental health on social media data." *International Conference on Neural Information Processing*. Cham: Springer International Publishing, 2022.

Explainable Causal Analysis

Classifier	F1:C0	F1:C1	F1:C2	F1:C3	F1:C4	F1:C5	Accuracy
LSTM	0.55	0.30	0.36	0.45	0.55	0.25	0.4514
BiLSTM	0.59	0.25	0.53	0.44	0.58	0.43	0.5054
CNN	0.57	0.26	0.53	0.54	0.58	0.35	0.4919
CNN-LSTM	0.57	0.17	0.38	0.46	0.48	0.52	0.4784

Performance evaluation of multi-class classifiers for causal categorization of mental illness on social media data where F1:C0, F1: C1, F1:C2, F1:C3, F1:C4 and F1:C5 defines F1-score for 6 categories: cause 0: 'No reason', cause 1: 'Bias or abuse', cause 2: 'jobs and careers', cause 3: 'medication', cause 4: 'relationships', and cause 5: 'alienation', respectiely

Method used	Class0	Class1	Class2	Class3	Class4	Class5
LSTM+LIME	0.787	0.825	**0.889**	0.751	0.881	0.854
LSTM+IG	0.723	0.779	**0.870**	0.701	0.869	0.813
BiLSTM+LIME	0.784	0.821	**0.881**	0.751	0.867	0.857
BiLSTM+IG	0.716	0.773	**0.866**	0.709	0.865	0.814
CNN+LIME	0.776	0.835	**0.898**	0.822	0.894	0.861
CNN+IG	0.729	0.765	**0.863**	0.689	**0.863**	0.818
CNN-LSTM+LIME	0.781	0.831	**0.878**	0.811	0.868	0.852
CNN-LSTM+IG	0.728	0.789	0.851	0.690	**0.870**	0.815

Values obtained for semantic similarity among resulting top-keywords and human-annotated inferences using Cosine Similarity: The distance lies between 0 and 1

$$Sim(Doc_1, Doc_2) = \frac{Doc_1 \cdot Doc_2}{||Doc_1||\,||Doc_2||} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

where A_i and B_i represent the components of vectors Doc_1 and Doc_2, respectively.

Method used	Class0	Class1	Class2	Class3	Class4	Class5
LSTM+LIME	1.029	0.854	0.857	0.896	**0.838**	0.889
LSTM+IG	1.097	0.890	0.870	0.926	**0.867**	0.906
BiLSTM+LIME	1.029	0.880	0.865	0.886	**0.852**	0.876
BiLSTM+IG	1.117	0.900	0.898	0.919	**0.870**	0.908
CNN+LIME	1.042	0.820	0.831	**0.817**	0.823	0.843
CNN+IG	1.123	0.907	0.882	0.912	**0.880**	0.913
CNN-LSTM+LIME	1.018	0.843	**0.831**	0.848	0.851	0.863
CNN-LSTM +IG	1.117	0.913	**0.869**	0.918	0.874	0.890

Values obtained for semantic similarity among resulting top-keywords and human-annotated inferences using Word Mover Distance: More distance indicates less similarity among two different texts.

WMD is computed using the cost-matrix having x_i and x_j be embedding of word i and j. The cost matrix $CM \in \mathbb{R}^m \times \mathbb{R}^m$ is the distance of embeddings, such that $CM_{ij} = ||x_i - x_j||^2$ as referred to in Eq. 2. The distance between two documents Doc_1 and Doc_2 is the optimum value of the following problem:

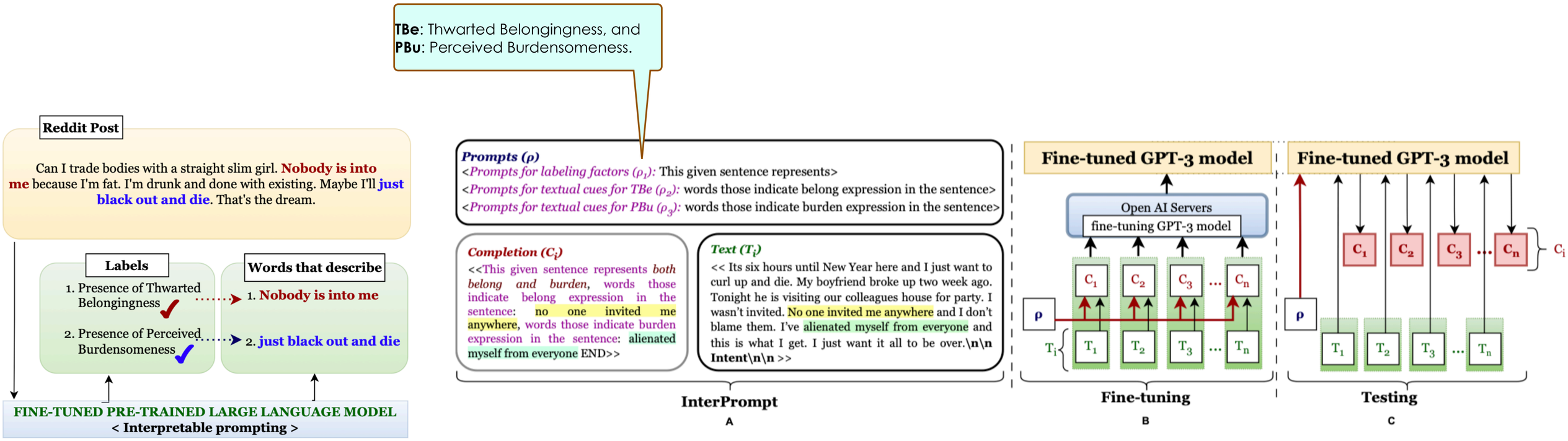
$$\min_{P \in \mathbb{R}^{m \times m}} \sum_{ij} CM_{ij}\, P_{ij} \quad (2)$$

such that $P_{ij} \geq 0$ Intuitively, P_{ij} represents the amount of word i that is transported to word j. WMD is defined as the minimum total distance to convert one document to another document.

Saxena, Chandni, Muskan Garg, and Gunjan Ansari. "Explainable causal analysis of mental health on social media data." *International Conference on Neural Information Processing*. Cham: Springer International Publishing, 2022.

Generative LLMs for Explanation

- We enhanced the performance of GPT-3 model and tailored it to a downstream task of explainable identification of IRFs



(Overview of the InterPrompt-driven fine-tuning of GPT-3 model)

Sathvik, M. S. V. P. J., Surjodeep Sarkar, Chandni Saxena, Sunghwan Sohn, and Muskan Garg. "InterPrompt: Interpretable Prompting for Interrelated Interpersonal Risk Factors in Reddit Posts." *arXiv preprint arXiv:2311.12404* (2023).

Explainable Causal Analysis

Model	Thwarted Belongingness				Perceived Burdensomeness			
	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
BERT	69.70	76.97	72.30	68.97	56.47	53.00	52.20	72.56
RoBERTa	71.23	73.54	71.35	68.97	67.27	37.52	45.51	74.93
DistilBERT	70.24	74.08	71.15	68.50	51.15	31.89	36.93	71.71
MentalBERT	77.97	77.40	76.73	75.12	64.22	65.75	62.77	78.33
OpenAI+LR	79.00	**83.59**	81.23	78.62	82.66	63.08	71.55	84.58
OpenAI+RF	79.06	80.68	79.86	77.48	**83.33**	49.23	61.90	81.36
OpenAI+SVM	81.31	80.34	**80.83**	**78.90**	79.15	74.77	**76.90**	**86.19**
OpenAI+MLP	**81.40**	75.56	78.37	76.92	72.08	**77.85**	74.85	83.92
OpenAI+XGB	81.22	79.83	80.52	78.62	80.36	68.00	73.67	85.05
GPT-3 Zero-shot	63.78	21.54	32.21	51.63	27.56	16.28	20.47	61.42
GPT-3 One-shot	61.15	84.57	70.98	63.12	34.84	86.05	49.60	46.67
GPT-3 Few-shot	57.42	94.68	71.49	59.72	32.16	98.14	48.45	36.31
Fine-tuned GPT-3 [A]	86.14	83.93	85.02	83.63	76.72	79.08	77.88	86.19
Fine-tuned GPT-3 [B]	**86.38**	84.21	**85.12**	**84.29**	**81.24**	80.69	80.96	87.08
Fine-tuned GPT-3 [C]	85.17	**87.35**	86.24	84.58	79.94	80.92	80.43	**87.89**
Fine-tuned GPT-3 [D]	84.59	83.01	83.80	82.83	80.12	**84.17**	**82.13**	87.54

GPT-3 variants:

ada → [A],
babbage → [B]
curie → [C]
davinci → [D].

Comparison of InterPrompt driven fine-tuned GPT-3 variants with baseline models over IRF dataset

Explanation using LIME and SHAP

Generated Explanation

Model Name	Thwarted Belongingness				Perceived Burdensomeness			
	Rouge-1	Rouge-L	BLEU-1	EM	Rouge-1	Rouge-L	BLEU-1	EM
MentalBERT+LIME	0.2202	-	0.1509	-	0.2425	-	0.1706	-
MentalBERT+ SHAP	0.2415	-	0.1593	-	0.2601	-	0.1718	-
Fine-tuned GPT-3 [A]	0.6597	0.6591	0.6373	0.5787	0.7738	0.7738	0.7556	0.6993
Fine-tuned GPT-3 [B]	0.6631	0.6628	0.6412	**0.5816**	0.7832	0.7831	0.7693	0.7163
Fine-tuned GPT-3 [C]	0.6637	0.6633	0.6377	0.5603	**0.7989**	**0.7989**	**0.7846**	**0.7348**
Fine-tuned GPT-3 [D]	**0.6809**	**0.6805**	**0.6532**	0.5801	0.7771	0.7771	0.7579	0.6965

Performance evaluation of generated explanations through similarity measures

Sathvik, M. S. V. P. J., Surjodeep Sarkar, Chandni Saxena, Sunghwan Sohn, and Muskan Garg. "InterPrompt: Interpretable Prompting for Interrelated Interpersonal Risk Factors in Reddit Posts." *arXiv preprint arXiv:2311.12404* (2023).

NLP Datasets for MHA

Publicly Available Datasets

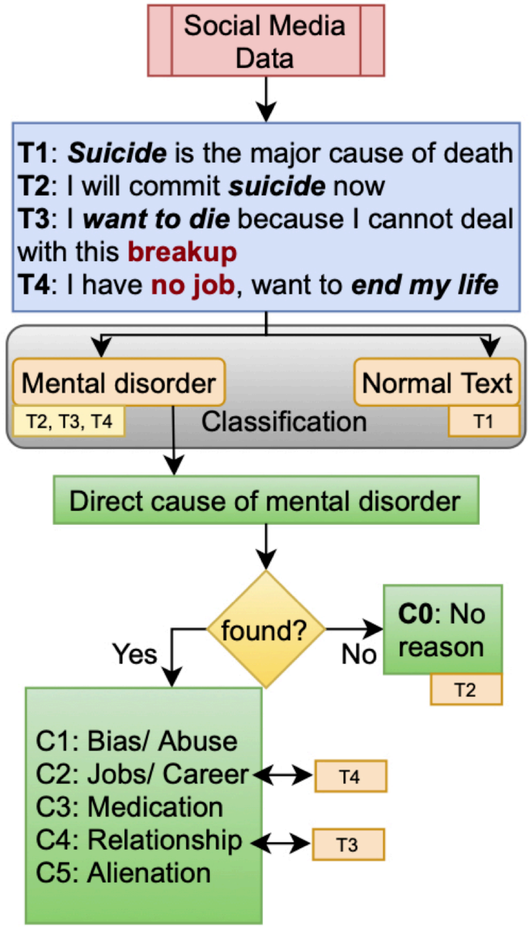
Dataset	Task	Avail.
CLPsych (Coppersmith et al., 2015)	Depression detection for suicide risk	S
MDDL (Shen et al., 2017)	Depression candidate detection (D1, D2, D3)	A
RSDD (Yates et al., 2017)	Depression detection from Reddit data	ASA
SMHD (Cohan et al., 2018)	Multi-task mental illness from Reddit data	ASA
eRISK (Losada et al., 2018)	Early risk detection: CLEF	A
Pirina18 (Pirina and Çöltekin, 2018)	Depression detection from Reddit data	A
Ji18 (Ji et al., 2018)	Suicide risk detection from Reddit data	AR
Aladag18 (Aladağ et al., 2018)	Suicide risk detection	AR
Sina Weibo (Cao et al., 2019)	Identifying candidates with suicide risk	AR
SRAR (Gaur et al., 2019)	Suicide risk from Reddit posts	ASA
Dreaddit (Turcan and McKeown, 2019)	Stress detection from Reddit posts	A
UMD-RD (Shing et al., 2020)	Suicide risk detection from Reddit data	ASA
SDCNL (Haque et al., 2021)	Suicide v/s depression from Reddit	A
CAMS (Ours)	Interpretable Causal analysis from Reddit	A

Different mental health datasets and their availability. A: Available, ASA: Available via Signed Agreement, AR: Available on Request for research work

More available datasets ……

Dataset	Task
loneliness	loneliness detection
MultiWD	Wellness dimensions detection
IRF	interpersonal risk factors detection

CAMS Datasets for Causal Analysis



Architecture of the causal analysis for mental health in social media posts

T1: I am done with my life after two years with no job
T2: Panic attacks and insomnia is deteriorating my mental peace
T3: She has no time for me. Family is a myth. Feeling lonely.
T4: Unable to cope up with my grades and university exams. Hopeless
T5: Help me or I ll do something to myself
T6: Life is useless.

Cause	CC	Train_S	Test_S	CAMS
No reason	292	332	70	694
Bias or abuse	122	194	35	351
Jobs/careers	399	181	48	628
Medication	410	170	43	623
Relationship	956	297	91	1344
Alienation	976	340	92	1408
Total	**3155**	**1517**	**379**	**5051**

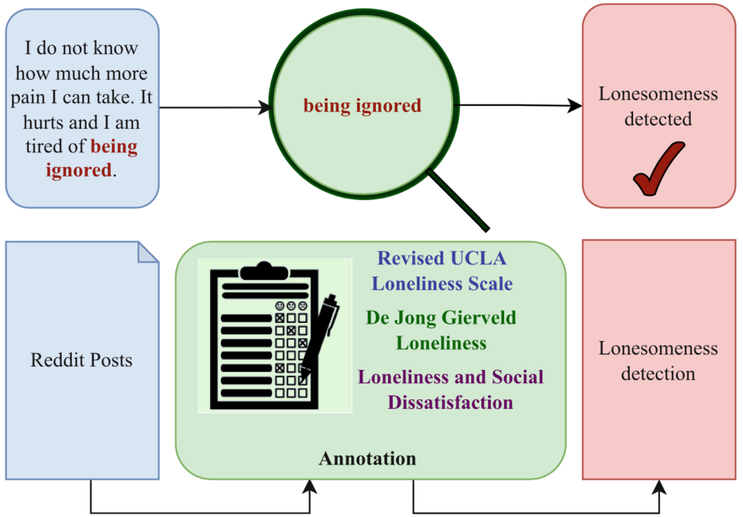
Sample distribution of the CAMS dataset for dif- ferent causes where CC is Crawled Corpus, *Train S* is the Training data of SDCNL dataset, *Test S* is the Test data of SDCNL dataset, and *CAMS* column contains the total number of samples in the dataset for each cause

Classifier	F1: C0	F1: C1	F1: C2	F1: C3	F1: C4	F1: C5	Accuracy
LR	**0.63**	**0.28**	0.54	**0.46**	0.46	**0.53**	**0.5013**
SVM	0.54	0.23	**0.56**	0.44	**0.48**	0.45	0.4670
LSTM	0.54	**0.27**	0.52	0.46	0.42	**0.51**	0.4595
CNN	0.56	**0.27**	0.51	0.42	0.46	0.38	0.4378
GRU	0.51	**0.27**	0.54	**0.47**	0.48	0.42	0.4541
Bi-LSTM	0.55	0.12	0.41	0.23	0.44	0.50	0.4351
Bi-GRU	**0.57**	0.14	**0.55**	0.46	0.49	0.39	0.4568
CNN+GRU	0.51	0.14	0.49	0.36	0.27	0.45	0.4027
CNN+LSTM	0.54	0.22	0.54	**0.47**	**0.54**	0.47	**0.4778**

Architecture of the causal analysis for mental health in social media posts

Garg, M., Saxena, C., Krishnan, V., Joshi, R., Saha, S., Mago, V., & Dorr, B. J. CAMS: An Annotated Corpus for Causal Analysis of Mental Health Issues in Social Media Posts.

LonXplain: Lonesomeness as a Consequence of Mental Disturbance in Reddit Post



Overview of classifying lonesomeness in Reddit Posts

Text	Label	Exp.
Just a sense of impending doom, this year is going to be shit. I'm starting to think things never actually do get better. All of my friends are out partying right now ← (Consequence) and I'm at home getting lectured by my family← (Cause) on my negative attitude. Anyway, happy new year I guess	Present	my friends are out partying
All of us on here are probably feeling alone and lonely← (Consequence) and depressed and like everyone else out there is having an awesome time← (Cause) except us, so why don't we have our own"party"? (In a way). Let's get to know each other! What is something really funny to you guys? It can be a joke/a meme/a video/a story of yours/whatever. Let's help each other feel less alone	Present	feeling alone and lonely
There is literally no point in life. We live and we die. And life has been hell to me← (Cause) so far so why should I even bother finishing. I am almost at the point where I am about to say **** it and quit	Absent	-

Overview An annotated dataset example illustrates causes (blue colored text-span) and lonesomeness as a consequence (red colored text-span) of mental disturbance in Reddit posts. Not all posts contain information about cause and/or consequences.

Model	Absent			Present			Accuracy
	P	R	F	P	R	F	
Word2Vec + RF	0.61	0.55	0.58	0.66	0.72	0.69	0.64
GloVe + LSTM	0.60	0.82	0.70	0.81	0.58	0.68	0.69
GloVe+BiLSTM	0.80	0.59	0.68	0.74	0.89	0.81	0.76
GloVe + GRU	0.72	0.80	0.75	0.83	0.76	0.79	0.77
Glove + BiGRU	0.70	0.84	0.76	0.85	0.73	0.79	0.78

Comparison of SOTA baselines. Score of each metric is averaged over 10-folds

Model	ROUGE-1 P	ROUGE-2 R	ROUGE-1 F
LSTM	0.50	0.12	0.18
BiLSTM	0.58	0.15	0.22
GRU	0.53	0.14	0.21
BiGRU	0.55	0.14	0.21

Performance evaluation of explanations obtained through LIME

Garg, M., Saxena, C., Samanta, D., & Dorr, B. J. (2023, June). LonXplain: Lonesomeness as a Consequence of Mental Disturbance in Reddit Posts. In International Conference on Applications of Natural Language to Information Systems (pp. 379-390).

Ethics in MHA, and Future Directions

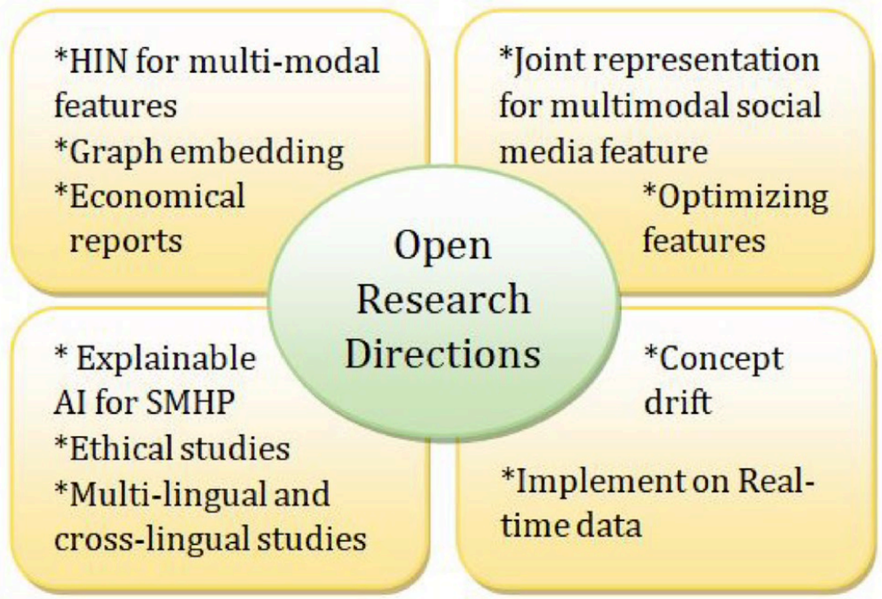
Ethics in MHA

- **Ethical considerations:** Prioritize discussing ethical concerns when using mental health-related textual data. Privacy and security of personal data, especially health data, are crucial. Follow strict protocols to protect privacy and prevent psychological distress.

- **Institutional approvals:** Obtain ethical approvals from institutional review boards and human research ethics committees when using publicly available data. This ensures compliance with ethical standards and protects the rights and well-being of research participants.

- Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M., & De Choudhury, M. (2019, January). A taxonomy of ethical tensions in inferring mental health states from social media. In Proceedings of the conference on fairness, accountability, and transparency (pp. 79-88).
- Young, S. D., & Garett, R. (2018). Ethical issues in addressing social media posts about suicidal intentions during an online study among youth: case study. *JMIR mental health*, *5*(2), e8971.
- Garg, M., Saxena, C., Naseem, U., & Dorr, B. J. (2023). NLP as a Lens for Causal Analysis and Perception Mining to Infer Mental Health on Social Media. *arXiv preprint arXiv:2301.11004*.

Future Directions



- Open-access datasets
- Unified tasks definition
- Multi-lingual, cross-lingual and language-independent approach
- Incremental learning from streaming data
- Interpretability and explainability
- More collaboration with clinical psychologist
- ……

Thank you!
Q & A?