SemEval-2026 Shared Task Proposal: Detection of Psychological Defense Mechanisms in Conversations

Hongbin Na[1], **Zimu Wang**[2], **Yining Hua**[3], **Rena Gao**[4],
Ling Chen[1], **Wei Wang**[2], **Shaoxiong Ji**[5], **John Torous**[3], **Sophia Ananiadou**[6]
[1]University of Technology Sydney [2]Xi'an Jiaotong-Liverpool University
[3]Harvard University [4]The University of Melbourne [5]Technical University of Darmstadt
[6]The University of Manchester

1 Overview

Psychological defense mechanisms are unconscious mental strategies individuals employ to cope with stress, regulate emotions, and manage complex interpersonal interactions (Freud et al., 2018). In emotional support conversations, help-seekers frequently and unconsciously exhibit these mechanisms, significantly affecting the dynamics and outcomes of supportive interactions. Although understanding and identifying defense mechanisms has long been essential in psychotherapy and psychological research (Cramer, 2015), computational methods designed specifically for detecting these nuanced psychological constructs within dialogues remain scarce in the NLP community.

To address this gap, we propose a SemEval-2026 shared task for detecting psychological defense mechanisms demonstrated by individuals seeking emotional support in conversational contexts. This task will employ the established Defense Mechanism Rating Scale (DMRS) (Perry et al., 1993) as the foundational annotation framework, applied to the publicly available conversation dataset (Liu et al., 2021). The proposed evaluation will feature multiple subtasks, including binary classification (presence or absence of defense mechanisms), multi-class classification (specific types of defense mechanisms), and multi-label classification, assessed both at sentence and dialogue levels.

While NLP has significantly advanced in sentiment analysis and emotion detection, detecting defense mechanisms involves deeper psychological insights and linguistic subtleties, which have not yet been adequately explored. With recent interest in integrating psychological theories into computational linguistics, especially leveraging powerful large language models (LLMs) (Hua et al., 2024a; Na et al., 2025), this task presents an unprecedented opportunity for interdisciplinary collaboration. It will likely attract researchers and practitioners from NLP, psychology, conversational AI, and mental health informatics communities.

This shared task aims to establish a novel and valuable dataset to inspire innovations for detecting complex psychological defense mechanisms in conversational data. By bridging psychological theory and computational linguistics, this initiative will not only foster interdisciplinary collaborations but also significantly advance NLP techniques in mental health applications. Outcomes from this task are expected to inform future research in therapeutic conversational systems, automated mental health support tools, and deeper linguistic modeling of psychological phenomena.

2 Data & Task Format

Annotation Process. We will use de-identified real-world public emotional support conversations. Building upon the valuable Emotional Support Conversation (ESConv) dataset (Liu et al., 2021), which comprises 1,300 conversations and covers 10 problem categories, including ongoing depression, relationship breakups, and job crises, we propose to add the sentence-level annotation to facilitate the analysis of dynamic psychological mechanisms through the conversations. We plan to annotate all 1,300 conversations and perform quality auditing to filter out samples with low quality. For the annotation process, we intend to use Label Studio[1], an open-source data labeling platform that supports various data types and offers a flexible interface suitable for our sentence-level annotation tasks.

During our pilot annotation phase, we developed an initial annotation manual to guide annotators in maintaining consistent standards. Table 1 provides representative examples of utterances annotated according to the DMRS, demonstrating various psychological defense mechanisms along with their respective labels. We plan to continually refine the

[1]labelstud.io/

DMRS Level	Example Utterance	Label
Highly adaptive	"I'm nervous, but hey, at least I'll have a funny story to tell later."	7
Obsessional	"I've analyzed all possible outcomes, so there's no reason for me to feel anxious."	6
Neurotic	"It's weird—I should feel angry, but I actually don't feel anything."	5
Minor image-distorting	"My new boss is absolutely perfect and knows everything."	4
Disavowal	"I know smoking is dangerous, but I'm young, nothing bad will happen."	3
Major image-distorting	"Everyone at my workplace is horrible and unfair."	2
Action-level	"I got angry, so I smashed a plate."	1
No defense	"I'm feeling a bit tired today, probably because I didn't sleep well."	0

Table 1: Examples of psychological defense mechanisms from the DMRS framework with corresponding labels.

annotation manual through regular discussion sessions, utilizing examples such as those presented in the table to clarify distinctions and enhance annotation consistency and overall data quality.

Annotator Recruitment. We have successfully secured two annotators with backgrounds in psychology for the annotation of data points. These annotators have already completed a pilot annotation session designed to familiarize them with the annotation criteria and ensure a robust understanding of the task based on the defense mechanism.

Annotation Quality. In addition to the standard labels from 0 to 7, we have introduced an additional label (8) to capture instances where annotators are uncertain about the appropriate label. We plan to conduct regular discussion sessions throughout the annotation process to address instances labeled as 8. This approach serves to improve the accuracy of annotations for ambiguous samples and provides ongoing training to annotators, thereby enhancing the quality of future annotations. After the annotation process, 10% of samples will be randomly selected for expert evaluation, led by a professor of psychiatry, to assess consistency and ensure data quality.

Task Descriptions. Our shared task comprises three subtasks, structured according to the DMRS's seven-level classification of psychological defense mechanisms. Table 1 provides examples of defense mechanisms within the framework.

Specifically, the subtasks include:

- **Task A1: Binary Classification of Defense Mechanism Presence at Sentence-level**: Performing binary text classification to determine whether psychological defense mechanisms are present within individual utterances.
- **Task A2: Binary Classification of Defense Mechanism Presence at Dialogue-level**: Performing binary text classification to determine whether psychological defense mechanisms are present throughout an entire conversation.
- **Task B: Multi-Class Classification of Defense Mechanism Type at Sentence-level**: Classifying each utterance into a specific psychological defense mechanism category defined by the DMRS framework, such as denial, projection, or intellectualization.
- **Task C: Multi-Label Classification of Defense Mechanism Types at Dialogue-level**: Performing multi-label classification to identify all categories of defense mechanisms that appear throughout an entire conversation.

Evaluation Metrics. We will evaluate model performance using standard classification metrics appropriate for the tasks. For binary classification tasks (Tasks A1 and A2), we will employ accuracy, precision, recall, and F1-score. For the multi-class classification task (Task B), metrics such as macro precision, recall, and F1-score will be used to assess overall model performance. For multi-label classification (Task C), we will use metrics specifically designed for multi-label scenarios, including micro and macro precision, recall, and F1-score, as well as subset accuracy. These evaluation metrics will provide a comprehensive assessment of model capabilities in detecting and classifying defense mechanisms in conversational data.

Example Datapoints. Figure 1 shows an example of the annotated data. Each conversation consists of multiple utterances labeled by speaker type (`seeker` or `supporter`). Each seeker utterance is annotated with a psychological defense mechanism label (`dmrs_label`). Dialogue-level labels (`labels_dialogue_level`) indicate whether any defense mechanisms are present (binary classification) and what types of mechanisms are identified throughout the dialogue (multi-label classification).

```
{
  "dialogue_id": "conv_001",
  "utterances": [
    {"speaker": "seeker", "text": "Hi!", "dmrs_label": 0},
    {"speaker": "supporter", "text": "Hello. How are you feeling today?", "dmrs_label": null},
    {"speaker": "seeker", "text": "Well, I'm a little tired and anxious but I got use to it. And you?", "dmrs_label": 3},
    ...
    {"speaker": "seeker", "text": "I will, thank you so much.", "dmrs_label": 0}
  ],
  "labels_dialogue_level": {
    "binary_presence": 1,
    "multi_label_types": [3, 4, 7]
  }
}
```


Figure 1: A JSON-formatted example datapoint illustrating sentence-level and dialogue-level annotations.

Ethical Considerations. Our conversation data originates from the ESConv dataset, distributed under the CC BY-NC 4.0 license[2], permitting use, modification, and distribution with appropriate attribution for non-commercial, academic research purposes. The sole human-involved component of this study is the annotation of the existing conversation data. This research has been reviewed by the ethics committee of Xi'an Jiaotong-Liverpool University and is deemed to be exempt from ethics review, as it involves only secondary analysis of publicly available datasets and does not include primary data collection.

3 Pilot Annotation

We conducted a pilot annotation with 30 dialogues to test the feasibility of our annotation framework, where two annotators with backgrounds in psychology independently annotated each utterance. Inter-annotator agreement was 0.61 (Cohen's kappa κ), indicating substantial annotation reliability. Psychological defense mechanisms were identified in 23 dialogues, while 7 dialogues exhibited no defensive mechanisms. Annotators commonly disagreed in cases distinguishing denial (label 3) from no defense (label 0), and major (label 2) from minor image-distorting mechanisms (label 4). Key insights include ambiguities primarily arising between similar mechanisms (e.g., label 0 vs. label 3, label 2 vs. label 4), necessitating more explicit guidelines. Context was found essential for accurate labeling, underscoring the need for clear contextual criteria. Furthermore, incorporating an "uncertain" category (label 8) effectively managed ambiguous cases, and regular discussions enhanced annotation consistency. Based on these findings, the annotation manual was refined with clearer definitions, contrasting examples, and structured decision trees to further enhance annotation quality.

[2] creativecommons.org/licenses/by-sa/4.0/

4 Organizers

The organization team possesses substantial expertise in the fields of mental health and LLMs, with extensive experience in organizing shared tasks and workshops, along with proficiency in other aspects for the current research proposal. H. Na and Z. Wang will lead this proposal, both of whom have strong expertise in mental health (Na, 2024; Na et al., 2024, 2025; Hua et al., 2024a,b; Qian et al., 2024; Ma et al., 2025) as well as LLM-based data annotation and evaluation (Wang et al., 2022; Peng et al., 2023; He et al., 2024; Li et al., 2024). S. Ji, J. Torous, and S. Ananiadou will act as advisory organizers. S. Ji specializes in language modeling and mental health data curation (Ji et al., 2018; Yang et al., 2023; Chen et al., 2024b; Zhang et al., 2024) and served as the organizer for Task 3 at SemEval-2025 (Vázquez et al., 2025). J. Torous brings broad expertise in psychiatry and has published papers on various prominent mental health journals, such as *World Psychiatry*, *The Lancet Psychiatry*, *JAMA Psychiatry*, and *npj Digital Medicine*. S. Ananiadou has contributed highly regarded research in mental health, including EmoLLMs (Liu et al., 2024b) and MentaLLaMA (Yang et al., 2024), and organized shared tasks in BioNLP (Goldsack et al., 2023) and FinNLP (Xie et al., 2024; Liu et al., 2025), as well as workshops in leading NLP conferences (Demner-Fushman et al., 2008, 2018, 2019, 2021, 2022, 2023, 2024a,b; Cohen et al., 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017; Ananiadou et al., 2016; Chen et al., 2025). Other team members, Y. Hua, R. Gao, L. Chen, and W. Wang, also bring a range of expertise in mental health and NLP, with experience in data annotation (Jiang et al., 2022; Zhou et al., 2022; Han et al., 2024; Liu et al., 2024a; Chen et al., 2024a) and workshop organization (Hahn et al., 2024).

References

Sophia Ananiadou, Riza Batista-Navarro, Kevin Bretonnel Cohen, Dina Demner-Fushman, and Paul Thompson, editors. 2016. *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*. The COLING 2016 Organizing Committee, Osaka, Japan.

Chung-Chi Chen, Antonio Moreno-Sandoval, Jimin Huang, Qianqian Xie, Sophia Ananiadou, and Hsin-Hsi Chen, editors. 2025. *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*. Association for Computational Linguistics, Abu Dhabi, UAE.

Jian Chen, Peilin Zhou, Yining Hua, Loh Xin, Kehui Chen, Ziyuan Li, Bing Zhu, and Junwei Liang. 2024a. FinTextQA: A dataset for long-form financial question answering. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6025–6047, Bangkok, Thailand. Association for Computational Linguistics.

Pinzhen Chen, Shaoxiong Ji, Nikolay Bogoychev, Andrey Kutuzov, Barry Haddow, and Kenneth Heafield. 2024b. Monolingual or multilingual instruction tuning: Which makes a better alpaca. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1347–1356, St. Julian's, Malta. Association for Computational Linguistics.

K. Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, John Pestian, Jun'ichi Tsujii, and Bonnie Webber, editors. 2009. *Proceedings of the BioNLP 2009 Workshop*. Association for Computational Linguistics, Boulder, Colorado.

K. Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, John Pestian, Jun'ichi Tsujii, and Bonnie Webber, editors. 2010. *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Uppsala, Sweden.

Kevin Cohen, Dina Demner-Fushman, Sophia Ananiadou, and Jun-ichi Tsujii, editors. 2014. *Proceedings of BioNLP 2014*. Association for Computational Linguistics, Baltimore, Maryland.

Kevin B. Cohen, Dina Demner-Fushman, Sophia Ananiadou, Bonnie Webber, Jun'ichi Tsujii, and John Pestian, editors. 2012. *BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Montréal, Canada.

Kevin Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, John Pestian, and Jun'ichi Tsujii, editors. 2013. *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Sofia, Bulgaria.

Kevin Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, John Pestian, Jun'ichi Tsujii, and Bonnie Webber, editors. 2011. *Proceedings of BioNLP 2011 Workshop*. Association for Computational Linguistics, Portland, Oregon, USA.

Kevin Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, and Jun-ichi Tsujii, editors. 2015. *Proceedings of BioNLP 15*. Association for Computational Linguistics, Beijing, China.

Kevin Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, and Jun-ichi Tsujii, editors. 2016. *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Berlin, Germany.

Kevin Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou, and Junichi Tsujii, editors. 2017. *BioNLP 2017*. Association for Computational Linguistics, Vancouver, Canada,.

Phebe Cramer. 2015. Understanding defense mechanisms. *Psychodynamic psychiatry*, 43(4):523–552.

Dina Demner-Fushman, Sophia Ananiadou, and Kevin Cohen, editors. 2023. *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*. Association for Computational Linguistics, Toronto, Canada.

Dina Demner-Fushman, Sophia Ananiadou, Kevin Bretonnel Cohen, John Pestian, Jun'ichi Tsujii, and Bonnie Webber, editors. 2008. *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*. Association for Computational Linguistics, Columbus, Ohio.

Dina Demner-Fushman, Sophia Ananiadou, Makoto Miwa, Kirk Roberts, and Junichi Tsujii, editors. 2024a. *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Bangkok, Thailand.

Dina Demner-Fushman, Sophia Ananiadou, Paul Thompson, and Brian Ondov, editors. 2024b. *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*. ELRA and ICCL, Torino, Italia.

Dina Demner-Fushman, Kevin Bretonnel Cohen, Sophia Ananiadou, and Junichi Tsujii, editors. 2018. *Proceedings of the BioNLP 2018 workshop*. Association for Computational Linguistics, Melbourne, Australia.

Dina Demner-Fushman, Kevin Bretonnel Cohen, Sophia Ananiadou, and Junichi Tsujii, editors. 2019. *Proceedings of the 18th BioNLP Workshop and Shared Task*. Association for Computational Linguistics, Florence, Italy.

Dina Demner-Fushman, Kevin Bretonnel Cohen, Sophia Ananiadou, and Junichi Tsujii, editors. 2021. *Proceedings of the 20th Workshop on Biomedical*

Language Processing. Association for Computational Linguistics, Online.

Dina Demner-Fushman, Kevin Bretonnel Cohen, Sophia Ananiadou, and Junichi Tsujii, editors. 2022. *Proceedings of the 21st Workshop on Biomedical Language Processing*. Association for Computational Linguistics, Dublin, Ireland.

Anna Freud et al. 2018. *The ego and the mechanisms of defence*. Routledge.

Tomas Goldsack, Zheheng Luo, Qianqian Xie, Carolina Scarton, Matthew Shardlow, Sophia Ananiadou, and Chenghua Lin. 2023. Overview of the biolaysumm 2023 shared task on lay summarization of biomedical research articles. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 468–477, Toronto, Canada. Association for Computational Linguistics.

Michael Hahn, Alexey Sorokin, Ritesh Kumar, Andreas Shcherbakov, Yulia Otmakhova, Jinrui Yang, Oleg Serikov, Priya Rani, Edoardo M. Ponti, Saliha Muradoğlu, Rena Gao, Ryan Cotterell, and Ekaterina Vylomova, editors. 2024. *Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*. Association for Computational Linguistics, St. Julian's, Malta.

Wenhan Han, Meng Fang, Zihan Zhang, Yu Yin, Zirui Song, Ling Chen, Mykola Pechenizkiy, and Qingyu Chen. 2024. MedINST: Meta dataset of biomedical instructions. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 8221–8240, Miami, Florida, USA. Association for Computational Linguistics.

Jianfei He, Lilin Wang, Jiaying Wang, Zhenyu Liu, Hongbin Na, Zimu Wang, Wei Wang, and Qi Chen. 2024. Guardians of discourse: Evaluating llms on multilingual offensive language detection. *Preprint*, arXiv:2410.15623.

Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li, Hongbin Na, Yi han Sheu, Peilin Zhou, Lauren V. Moran, Sophia Ananiadou, Andrew Beam, and John Torous. 2024a. Large language models in mental health care: a scoping review. *Preprint*, arXiv:2401.02984.

Yining Hua, Hongbin Na, Zehan Li, Fenglin Liu, Xiao Fang, David Clifton, and John Torous. 2024b. Applying and evaluating large language models in mental health care: A scoping review of human-assessed generative tasks. *Preprint*, arXiv:2408.11288.

Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long. 2018. Supervised learning for suicidal ideation detection in online user content. *Complexity*, 2018(1):6157249.

Hang Jiang, Yining Hua, Doug Beeferman, and Deb Roy. 2022. Annotating the Tweebank corpus on named entity recognition and building NLP models for social media analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7199–7208, Marseille, France. European Language Resources Association.

Ruosen Li, Zimu Wang, Son Quoc Tran, Lei Xia, and Xinya Du. 2024. Meqa: A benchmark for multi-hop event-centric question answering with explanations. In *Advances in Neural Information Processing Systems*, volume 37, pages 126835–126862. Curran Associates, Inc.

Fenglin Liu, Zheng Li, Hongjian Zhou, Qingyu Yin, Jingfeng Yang, Xianfeng Tang, Chen Luo, Ming Zeng, Haoming Jiang, Yifan Gao, Priyanka Nigam, Sreyashi Nag, Bing Yin, Yining Hua, Xuan Zhou, Omid Rohanian, Anshul Thakur, Lei Clifton, and David A. Clifton. 2024a. Large language models are poor clinical decision-makers: A comprehensive benchmark. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13696–13710, Miami, Florida, USA. Association for Computational Linguistics.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. In *ACL*.

Zhiwei Liu, Keyi Wang, Zhuo Bao, Xin Zhang, Jiping Dong, Kailai Yang, Mohsinul Kabir, Polydoros Giannouris, Rui Xing, Seongchan Park, Jaehong Kim, Dong Li, Qianqian Xie, and Sophia Ananiadou. 2025. FinNLP-FNP-LLMFinLegal-2025 shared task: Financial misinformation detection challenge task. In *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*, pages 271–276, Abu Dhabi, UAE. Association for Computational Linguistics.

Zhiwei Liu, Kailai Yang, Qianqian Xie, Tianlin Zhang, and Sophia Ananiadou. 2024b. Emollms: A series of emotional large language models and annotation tools for comprehensive affective analysis. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, page 5487–5496, New York, NY, USA. Association for Computing Machinery.

Jiayuan Ma, Hongbin Na, Zimu Wang, Yining Hua, Yue Liu, Wei Wang, and Ling Chen. 2025. Detecting conversational mental manipulation with intent-aware prompting. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9176–9183, Abu Dhabi, UAE. Association for Computational Linguistics.

Hongbin Na. 2024. CBT-LLM: A Chinese large language model for cognitive behavioral therapy-based mental health question answering. In *Proceedings of*

the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 2930–2940, Torino, Italia. ELRA and ICCL.

Hongbin Na, Yining Hua, Zimu Wang, Tao Shen, Beibei Yu, Lilin Wang, Wei Wang, John Torous, and Ling Chen. 2025. A survey of large language models in psychotherapy: Current landscape and future directions. *Preprint*, arXiv:2502.11095.

Hongbin Na, Tao Shen, Shumao Yu, and Ling Chen. 2024. Multi-session client-centered treatment outcome evaluation in psychotherapy. *Preprint*, arXiv:2410.05824.

Hao Peng, Xiaozhi Wang, Jianhui Chen, Weikai Li, Yunjia Qi, Zimu Wang, Zhili Wu, Kaisheng Zeng, Bin Xu, Lei Hou, and Juanzi Li. 2023. When does in-context learning fall short and why? a study on specification-heavy tasks. *Preprint*, arXiv:2311.08993.

J Christopher Perry, Marianne E Kardos, and Christopher J Pagano. 1993. The study of defenses in psychotherapy using the defense mechanism rating scales (dmrs). *The concept of defense mechanisms in contemporary psychology: Theoretical, research, and clinical perspectives*, pages 122–132.

Lu Qian, Yuqi Wang, Zimu Wang, Haiyang Zhang, Wei Wang, Ting Yu, and Anh Nguyen. 2024. Domain-specific guided summarization for mental health posts. *Preprint*, arXiv:2411.01485.

Raúl Vázquez, Timothee Mickus, Elaine Zosa, Teemu Vahtola, Jörg Tiedemann, Aman Sinha, Vincent Segonne, Fernando Sánchez-Vega, Alessandro Raganato, Jindřich Libovický, Jussi Karlgren, Shaoxiong Ji, Jindřich Helcl, Liane Guillou, Ona de Gibert, Jaione Bengoetxea, Joseph Attieh, and Marianna Apidianaki. 2025. SemEval-2025 Task 3: Mu-SHROOM, the multilingual shared-task on hallucinations and related observable overgeneration mistakes.

Xiaozhi Wang, Yulin Chen, Ning Ding, Hao Peng, Zimu Wang, Yankai Lin, Xu Han, Lei Hou, Juanzi Li, Zhiyuan Liu, Peng Li, and Jie Zhou. 2022. MAVEN-ERE: A unified large-scale dataset for event coreference, temporal, causal, and subevent relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 926–941, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Qianqian Xie, Jimin Huang, Dong Li, Zhengyu Chen, Ruoyu Xiang, Mengxi Xiao, Yangyang Yu, Vijayasai Somasundaram, Kailai Yang, Chenhan Yuan, Zheheng Luo, Zhiwei Liu, Yueru He, Yuechen Jiang, Haohang Li, Duanyu Feng, Xiao-Yang Liu, Benyou Wang, Hao Wang, Yanzhao Lai, Jordan Suchow, Alejandro Lopez-Lira, Min Peng, and Sophia Ananiadou. 2024. FinNLP-AgentScen-2024 shared task: Financial challenges in large language models - FinLLMs. In *Proceedings of the Eighth Financial Technology and Natural Language Processing and the 1st Agent AI for Scenario Planning*, pages 119–126, Jeju, South Korea. -.

Kailai Yang, Shaoxiong Ji, Tianlin Zhang, Qianqian Xie, Ziyan Kuang, and Sophia Ananiadou. 2023. Towards interpretable mental health analysis with large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6056–6077, Singapore. Association for Computational Linguistics.

Kailai Yang, Tianlin Zhang, Ziyan Kuang, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. 2024. Mentallama: Interpretable mental health analysis on social media with large language models. In *Proceedings of the ACM Web Conference 2024*, WWW '24, page 4489–4500, New York, NY, USA. Association for Computing Machinery.

Tianlin Zhang, Kailai Yang, Shaoxiong Ji, Boyang Liu, Qianqian Xie, and Sophia Ananiadou. 2024. SuicidEmoji: Derived emoji dataset and tasks for suicide-related social content. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1136–1141.

Peilin Zhou, Zeqiang Wang, Dading Chong, Zhijiang Guo, Yining Hua, Zichang Su, Zhiyang Teng, Jiageng Wu, and Jie Yang. 2022. Mets-cov: A dataset of medical entity and targeted sentiment on covid-19 related tweets. In *Advances in Neural Information Processing Systems*, volume 35, pages 21916–21932. Curran Associates, Inc.