Intelligent Cognitive System for Computational Psychotherapy with a Conversational Agent for Attitude and Behavior Change in Stress, Anxiety, and Depression

Tine Kolenik
Institute of Synergetics and Psychotherapy Research, University Hospital of Psychiatry, Psychotherapy and Psychosomatics, Paracelsus Medical University, Salzburg, Austria
E-mail: tine.kolenik@gmail.com

Thesis Summary

Keywords: generative artificial intelligence, attitude and behavior change support systems, digital mental health, intelligent cognitive conversational agent, artificial cognitive architecture, machine learning

Received: March 28, 2025

The paper summarizes a Doctoral Thesis presenting a computational psychotherapy system for stress, anxiety, and depression (SAD) using a conversational agent. The first contribution is a novel panel dataset combining quantitative diagnostic-level questionnaires and qualitative daily diaries. The second contribution is the system itself, built upon a cognitive architecture simulating Theory of Mind through an ensemble of user, machine learning, and knowledge models for SAD prediction, forecasting, and personalized intervention.

Povzetek: Članek povzema doktorsko disertacijo, ki predstavlja sistem računalniške psihoterapije za stres, anksioznost in depresijo (SAD) z uporabo pogovornega agenta. Prvi prispevek je nova panelna podatkovna zbirka, ki združuje kvantitativne diagnostične vprašalnike in kvalitativne dnevne zapise. Drugi prispevek je sam sistem, zgrajen na kognitivni arhitekturi, ki simulira teorijo uma z ansamblom različnih uporabniških, umetnointeligenčnih in ekspertnih modelov za zaznavanje, napovedovanje in personalizirano intervencijo pri uporabnikih s SAD.

1 Introduction

The growing burden of mental health issues, particularly stress, anxiety, and depression (SAD), necessitates innovative and accessible support systems [1]. Computational psychotherapy, utilizing tools like intelligent conversational agents (ICAs), offers a promising avenue. However, current state-of-the-art (SOTA) systems often lack sophisticated user modeling, personalization, adaptation, and forecasting capabilities, limiting their effectiveness for dynamic mental health needs [2]. Many also struggle with safe and domain-specific language generation.

This thesis addresses these gaps by proposing a novel computational psychotherapy system centered around an ICA with Theory of Mind (ToM). It makes two core contributions: 1) The creation of a unique, high-quality, mixed-methods panel dataset specifically designed for developing and evaluating such systems. 2) The design and implementation of the system's cognitive architecture (CogA) which distinctively simulates ToM [3]. ToM simulation allows the system to model user mental states (beliefs, emotions, intentions) and respond adaptively and persuasively, aiming to predict, forecast, and positively influence the user's SAD state.

2 Dataset and cognitive architecture

A key contribution was the creation of a novel "golden standard" dataset. Data was collected using Ecological Momentary Assessment (EMA) via the Synergetic Navigation System (SNS) application from 50 participants over approximately four weeks [4]. This resulted in a panel dataset of 1495 instances, each containing daily quantitative scores from an 18-item SAD symptom questionnaire and qualitative text diary entries (~150 words minimum) describing the user's mood, experiences, and thoughts. Big Five (B5) personality traits [5] and demographics were also collected.

The system's core is its CogA designed to simulate ToM. It comprises three main modules (Figure 1):

Natural Language Processing (NLP): Handles user text input via Dialog Management (tracking conversation state), Natural Language Understanding (sensibility filtering), and Feature Extraction (using LIWC and VADER to create numeric representations of mental dimensions [4]).

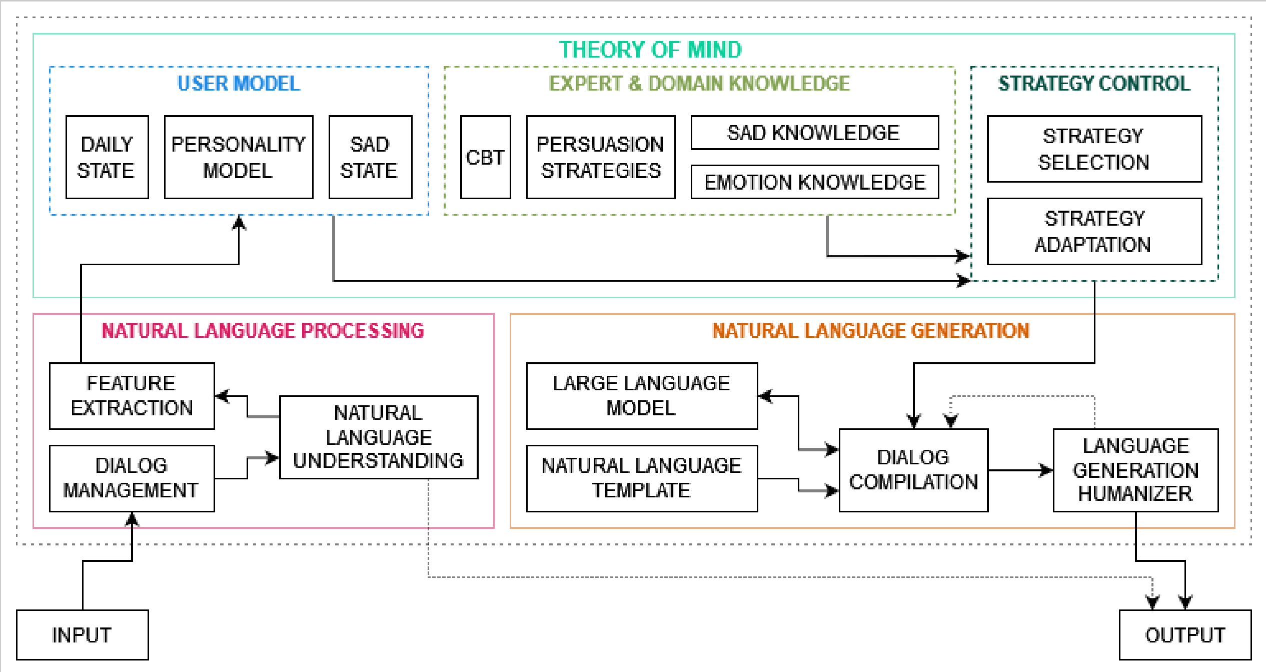

Figure 1: The system's cognitive architecture.

Theory of Mind: The central module, processing features to model the user and determine strategy. It includes:

- User Model: Simulates the user's state via submodules for Daily State (current multi-dimensional mental properties), Personality Model (long-term B5 traits), and SAD State (ML models for detecting current SAD levels/symptoms and forecasting them up to 7 days ahead from text).
- Expert & Domain Knowledge: Contains ontologies linking user states to interventions, including Cognitive Behavioral Therapy (CBT) techniques (difficulty-adapted), Persuasion Strategies (based on Cialdini's Principles mapped to B5 via a Domain Mapping Matrix (DMM)), SAD Knowledge (mapping issues to CBT via DMM), and Emotion Knowledge (guiding output tone) [4].
- Strategy Control: Selects and adapts strategies. Strategy Selection chooses appropriate CBT based on SAD state/topic and wraps it in a personalized persuasion strategy based on the User Model and DMM. Strategy Adaptation refines strategies based on effectiveness (using Ratio Formulas for learning) [4].
- Natural Language Generation (NLG): Compiles the final text output. It uses Natural Language Templates, enriches the text stochastically using a Large Language Model (LLM), and employs a Language Generation Humanizer (risk thresholding) to filter potentially harmful outputs before presenting the text to the user [4].

3 Evaluation and results

The system was evaluated through computational experiments and an empirical interventional study.

1.1 Computational experiments

Machine learning models within the SAD State submodule were trained and evaluated using 10-fold subject-wise cross-validation on the novel dataset (see Section 2).

Detection: SAD levels and 15 symptoms (inability to relax, nervousness, fear, tightness in chest, lightheadedness, feeling hot or cold, trembling, pounding heart, sadness, self-hatred, anhedonia, hopelessness, indecisiveness, fatigue, emotional detachment, suicidality) were detected from single text diary entries. Using kNN (chosen for explainability), accuracies ranged from 72.58% to 91.41%. This surpassed SOTA systems reviewed [4,6], which detected fewer categories, often with lower or non-comparable accuracy [4].

Forecasting: The system forecasted SAD levels and 15 symptoms 7 days ahead from single text entries, with kNN accuracies ranging from 71.54% to 87.68%. SOTA systems reviewed lacked this forecasting capability. Forecasting from 21 days of quantitative questionnaire data was also performed, achieving high accuracy (e.g., 93.14% for anxiety using Logistic Regression).

1.2 Empirical interventional study

A study involving 42 participants compared this work's system ("Our system") against Woebot [7] in a simulated daily check-in scenario. SAD levels were measured using Single Item Screening Questions (SISQs) before and after the interaction. User experience was measured using the User Experience Questionnaire (UEQ).

SAD Reduction: Paired t-tests showed "Our system" significantly reduced participant stress (M_diff = -0.263, p = 0.048) and anxiety (M_diff = -0.263, p = 0.040). Woebot showed no significant change in stress (p = 0.484) or anxiety ($p = 0.509$). Neither system significantly changed depression levels ($p > 0.6$).

User Experience: "Our system" was rated significantly more supportive than Woebot (*M_ours* = 5.368 vs *M_woebot* = 4.261, $p = 0.041$). There was no significant difference in perceived novelty ($p = 0.084$). Qualitative feedback indicated users appreciated the depth of assessment in "Our system" but preferred Woebot's user interface and friendly personality.

3 Discussion and conclusion

The system's performance was benchmarked against relevant SOTA systems identified in the literature review [5], ChatGPT [8], and through empirical comparison. Computationally, the CogA demonstrated superior detection capabilities, identifying a broader range of SAD levels and specific symptoms (18 categories) from single text entries with high accuracy (up to 91.41% using kNN) compared to the fewer categories and often lower or non-comparable accuracies reported by reviewed systems and recent evaluations of ChatGPT. Crucially, the system introduced a novel 7-day forecasting capability from text data, achieving accuracies up to 87.68%, a capability absents in the reviewed SOTA systems. The empirical interventional study directly compared "Our system" to Woebot, a prominent SOTA agent. While Woebot excels in user interface and personality, "Our system," leveraging its ToM simulation, achieved statistically significant short-term reductions in user-reported stress and anxiety, which Woebot did not. Participants also perceived "Our system" as significantly more supportive, although they noted Woebot's superior interface. These comparative results highlight the advantages of the developed CogA in assessment depth, predictive power, and intervention effectiveness for specific negative states.

This thesis therefore successfully developed and evaluated a novel computational psychotherapy system based on simulating Theory of Mind, supported by a unique mixed-methods panel dataset. The computational results demonstrate SOTA performance in detecting a wide range of SAD states from text and introduce novel forecasting capabilities. The empirical study confirmed the hypothesis that simulating ToM can lead to a system that performs comparably or better than established SOTA systems like Woebot.

Limitations include potential dataset biases, reliance on LLMs' capabilities, and the quasi-experimental nature of the interventional study. Future work involves user interface development, refining ML models, expanding the dataset, conducting larger and longitudinal clinical evaluations, and comparing different LLMs within the CogA framework. Overall, the work demonstrates the potential of integrating AI, cognitive science, and behavioral science through ToM simulation to create more effective and personalized digital mental health interventions.

Acknowledgements

I acknowledge the support of my supervisors Prof. Dr. Matjaž Gams and Prof. Dr. Günter Schiepek, the Jožef Stefan Institute, the Slovenian Research and Innovation Agency (ARIS) (research core funding No. P2-0209 and Young researchers postgraduate research funding), and the study participants.

References

[1] T. Kolenik, M. Gams, "Persuasive technology for mental health: One step closer to (mental health care) equality?" IEEE Technology and Society Magazine, 40(1):80–86, 2021. DOI: 10.1109/MTS.2021.3056288

[2] T. Kolenik, M. Gams, "Intelligent cognitive assistants for attitude and behavior change support in mental health: State-of-the-art technical review," Electronics, 10(11):1250, 2021. DOI: 10.3390/electronics10111250

[3] A. M. Leslie, O. Friedman, T. P. German, "Core mechanisms in 'theory of mind'," Trends Cogn Sci, 8(12):528–533, 2004. DOI: 10.1016/j.tics.2004.10.001

[4] T. Kolenik, G. Schiepek, M. Gams, "Computational Psychotherapy System for Mental Health Prediction and Behavior Change with a Conversational Agent,", Neuropsychiatr Dis Treat, 20:2465-2498, 2024. DOI: 10.2147/NDT.S417695

[5] B. Rammstedt, O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," J Res Pers, 41(1):203–212, 2007. DOI: 10.1016/j.jrp.2006.02.001

[6] T. Kolenik. Methods in digital mental health: Smartphone-based assessment and intervention for stress, anxiety and depression. In Integrating Artificial Intelligence and IoT for Advanced Health Informatics, C. Comito, A. Forestiero, and E. Zumpano, Eds., Springer, 2021. DOI: 10.1007/978-3-030-91181-2_7

[7] K. K. Fitzpatrick, A. Darcy, M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial," JMIR Ment Health, 4(2): e19, 2017. DOI: 10.2196/mental.7785

[8] B. Lamichhane, "Evaluation of chatgpt for nlp-based mental health applications," arXiv:2303.15727 [cs.CL], 2023. DOI: 10.48550/arXiv.2303.15727