# Conceptualising Case Formulation as a Neurosymbolic AI Framework for Mental Health

**Geoffrey Chern-Yee Tan[1], Pandya Poorva Harshang[2], Elizabeth Yap Ning Xuan[3], Sharon Lu Huixian[4], Hong Ming Tan[2,5]**

*1 Institute of Mental Health, Department of Mood and Anxiety, Singapore*

*2 National University of Singapore, Institute of Operations Research and Analytics, Singapore*

*3 Institute of Mental Health, Central Region, Singapore*

*4 Institute of Mental Health, Department of Psychology, Singapore*

*5 National University of Singapore Business School, Department of Analytics and Operations, Singapore*

*Correspondence to: [Geoffrey Chern-Yee Tan] geoffrey.tan@nhghealth.com.sg*

## 1. Introduction

Clinical case formulation forms the basis of the delivery of psychotherapy. It organises an individual's history and presenting problems into a coherent account that guides decisions on suitability and right-siting.[1] In practice, case formulation is commonly guided by the biopsychosocial (BPS) model, which highlights the interaction of biological, psychological, and social factors in the development and maintenance of mental health difficulties.[2] It is complemented by the 4Ps framework, which categorises factors into four categories (predisposing, precipitating, perpetuating, protecting) according to their temporal and causal roles for a more explanatory narrative.[3] An alternative framework we have developed is the 13-Factor Vulnerability (13FV) model, a structured taxonomy of transdiagnostic vulnerability factors across biological, psychological and social domains that are highly relevant to psychotherapy.

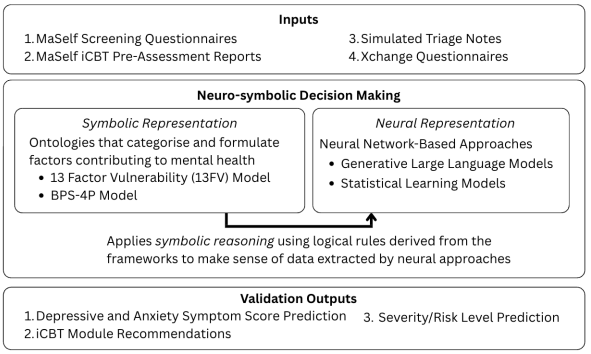
Constructing a case formulation is often cognitively demanding. As clinicians must synthesise large volumes of patient information from self-reported measures and case notes in a short period of time, there is potential value for AI approaches to support timely psychotherapy triage and assessment. For instance, large language models (LLMs) have shown growing promise in Information Extraction (IE) from unstructured electronic health record (EHR) data.[4] Meanwhile, supervised learning algorithms can predict symptom scores and vulnerability factor ratings from questionnaire data. However, adoption of purely neural approaches is limited by their "black box" nature, which reduces interpretability of output in contexts such as case formulation where transparent reasoning is essential to guide clinical decision-making processes.[5,6]

Neurosymbolic AI addresses this limitation by integrating neural network-based learning with symbolic knowledge-based reasoning, allowing domain knowledge from clinical frameworks to be explicitly embedded into AI models.[6,7] The neural component identifies patterns in patient data, while the symbolic component applies logical rules derived from the frameworks to map the extracted information to clinically meaningful constructs. This ensures outputs are not only informative but also interpretable in a way that mirrors human clinical reasoning.

In this paper, we conceptualise clinical case formulation as a neurosymbolic AI architecture. By embedding the BPS and 13FV models as symbolic knowledge layers and leveraging neural approaches to process clinical data, we propose an explainable AI-assisted approach that supports efficient case formulation for psychotherapy triage.

## 2. Methodology

Fig. 1: Case Formulation as a Neurosymbolic Architecture




We propose a neurosymbolic architecture that combines neural extraction and learning with structured symbolic reasoning to improve interpretability in AI-assisted case formulation (Fig. 1). The system integrates inputs from two modalities: unstructured clinical text (simulated triage notes created with reference to Xchange triage notes, and MaSelf internet-based Cognitive Behavioural Therapy (iCBT) pre-assessment reports) and structured data (item-level questionnaire responses from MaSelf and Xchange studies).

The symbolic layer encodes domain knowledge from two clinical frameworks: the broad BPS-4P model and the fine-grained 13FV model. Together, these ontologies provide explicit rule-based reasoning that guide neural approaches on what information is clinically relevant and how it should be organised in the factor space.

For the unstructured clinical text, a generative LLM extraction layer (GPT-5.2) performed instruction parsing, text segmentation, normalization of narrative variants into canonical clinical statements, schema alignment, and structured rendering, producing a mapping of items to the appropriate BPS-4P/13FV factors. The structured data was mapped directly onto the ontologies using explicit item-to-factor scoring rules, yielding factor score vectors.

We validated this neurosymbolic architecture through three outputs that can inform case formulation and psychotherapy triage. Firstly, we assessed predictive validity by fitting univariate linear regressions within each questionnaire dataset to predict PHQ-9 and GAD-7 outcomes. Secondly, we evaluated severity-weighted extraction coverage from the triage notes. Extracted items were canonicalized through semantic matching to form a union patient information library, before GPT 5.2 assigned each item a severity label (1–3) based on its expected relevance to depression and anxiety outcomes. A clinician then reviewed the labels, resolved disagreements and finalised the severity annotations used for subsequent severity-weighted coverage evaluation. Lastly, we evaluated treatment planning by generating iCBT module recommendations from MaSelf questionnaires and pre-assessment reports using explicit factor-to-module mappings. The performance was evaluated through comparing agreement scores with clinician recommendations.

## 3. Results

We assessed predictive validity by regressing screening scores onto PHQ-9 and GAD-7 outcomes, summarising fit with $R^2$ and RMSE, alongside F-statistics and p-values for evidence of association. Screening scores showed strong, statistically robust associations with both symptom measures, with high explained variance. (Table 1)

Table 1: Associations between Screening Scores with PHQ-9 and GAD-7 scores

| Outcome | □ | $R^2$ | p | RMSE | F |
|---|---|---|---|---|---|
| Depressive Symptoms (PHQ-9) | 1.02 | 0.80 | 7.78 $e^{-18}$ | 2.14 | 183 |
| Anxiety Symptoms (GAD-7) | 1.04 | 0.76 | 8.00 $e^{-17}$ | 2.46 | 145 |

We also evaluated module-selection agreement as a multi-label matching problem using the F1 score (precision-recall balance) and Hamming loss (label wise mismatch rate). Overall, recommendations demonstrated moderate to good concordance with clinician selections, indicating that the factor-to-module mapping produced actionable treatment planning signals. (Table 2)

Table 2: Agreement between model-generated iCBT module recommendations and clinician expert recommendations

| Outcome | F1 Score | Hamming Loss |
|---|---|---|
| M | 0.73 | 0.30 |
| SD | 0.16 | 0.17 |

## 4. Conclusion

While other studies have attempted to use neurosymbolic AI to support the selection of interventions[8], to our knowledge, we are the first to use neurosymbolic AI to conceptualise case formulation in a way that mirrors clinicians' thought process. Using these constructs, we were able to predict depressive and anxiety symptoms with an R2 of 0.76-0.80 and clinician expert recommendations for personalised iCBT module selection with an F1 score of 0.73.

## Acknowledgments

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported by the National Research Foundation, Singapore, and the Agency for Science Technology and Research (A*STAR), Singapore, under its Prenatal/ Early Childhood Grant (Grant No. H22P0M0005). This research is also supported by the National Health Innovation Centre, Singapore, under its I2D Grant (Reference No. NHIC-I2D-2502370). The authors also thank the study participants of the XChange study.

## References

[1] Godoy A, Haynes SN. Clinical Case Formulation. European Journal of Psychological

Assessment. 2011 Jan;27(1):1–3. https://doi.org/10.1027/1015-5759/a000055
[2] Engel GL. The Clinical Application of the Biopsychosocial Model. Journal of Medicine and Philosophy. 1981 Jan 1;6(2):101–24. doi.org/10.1093/jmp/6.2.101
[3] Bolton JW. Case Formulation After Engel—The 4P Model: A Philosophical Case Conference. Philosophy, Psychiatry, & Psychology. 2014;21(3):179–89. 10.1353/ppp.2014.0027
[4] Li L, Zhou J, Gao Z, Hua W, Fan L, Yu H, et al. A scoping review of using Large Language Models (LLMs) to investigate Electronic Health Records (EHRs) [Internet]. arXiv.org. 2024. https://arxiv.org/abs/2405.03066
[5] Garcez, Gori M, Lamb LC, Serafini L, Spranger M, Tran SN. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. 2019 May 1;6(4):611–31.
[6] Vidya R. Leveraging Neurosymbolic Ai And Explainable Ai For Enhanced Disease Prediction In Electronic Health Records Using Multimodal Datasets. In: Ramana S, Chandar VP, Kumar GM, Nair R, editors. AI and it's Applications. Innovation Online Training Academy; 2024. p. 145–57.
[7] Sheth A, Roy K, Gaur M. Neurosymbolic AI - Why, What, and How [Internet]. arXiv.org. 2023. https://arxiv.org/abs/2305.00813
[8] Anil Kumar. Neuro Symbolic AI in personalized mental health therapy: Bridging cognitive science and computational psychiatry. World Journal of Advanced Research and Reviews [Internet]. 2023 Aug 30 [cited 2025 Nov 13];19(2):1663–79. https://wjarr.com/sites/default/files/WJARR-2023-1516.pdf