PCN Psychiatry and Clinical Neurosciences

RESEARCH LETTER

Dynamic cognitive differences between internal and external attention are associated with depressive and anxiety symptoms

doi:10.1111/pcn.13859

The internal/external attention framework distinguishes attention directed toward internal representations (e.g., emotions) from attention focused on perceptual stimuli.[1] Dysfunctions in this attention balance are associated with various psychiatric symptoms, including depression.[2] Emerging evidence suggests these imbalances may reflect dysfunction in neural systems, such as the fronto-parietal network.[2] While many researchers have emphasized the need to clarify dynamic differences in neural activity underlying various cognitive dimensions of internal/external attention and their associations with psychiatric symptoms, this area remains understudied.[3]

To investigate how cognitive processes differ across conditions in terms of brain representations, fMRI provides a useful and common approach. Techniques such as regression and representational similarity analysis (RSA) have helped examine the geometry of brain activity patterns and associations with specific behavioral features.[4] However, complex functions like internal and external attention emerge from the dynamic interplay of multiple cognitive processes (e.g., emotions, selective attention).[1] Traditional approaches cannot fully capture the temporal dynamics of these processes and their interrelations. While dynamic functional connectivity can characterize time-varying interregional coupling patterns,[5] it is less suited to tracking how specific cognitive representations diverge across conditions. Thus, a novel method is needed to disentangle how these dynamics differ across conditions.

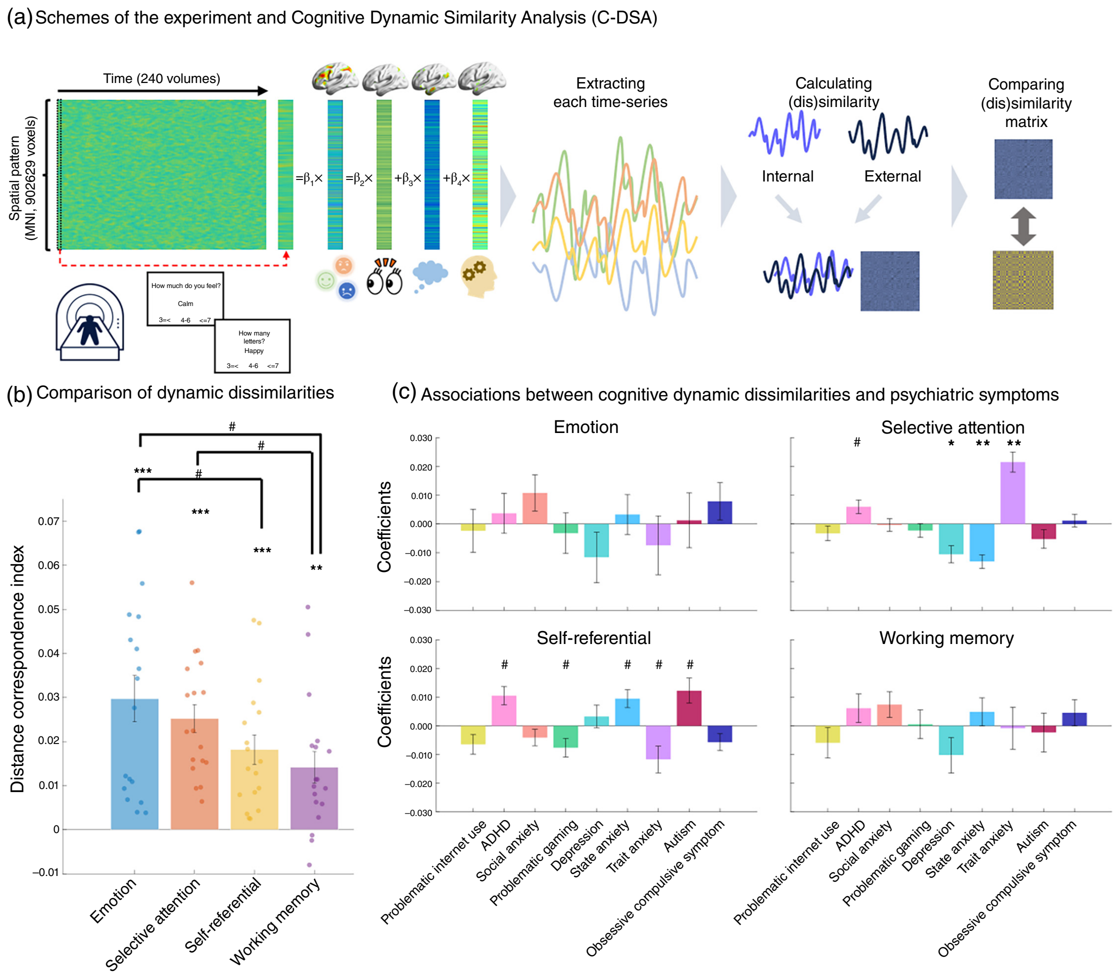

Fig. 1 Overview and results of Cognitive Dynamic Similarity Analysis (C-DSA). (a) Overview of C-DSA. Brain activation patterns were extracted using meta-analytic software and used as independent variables in a transposed Generalized Linear Model (GLM) based on cognitive dynamical estimation. Time series of each cognitive process were derived from regression coefficients, followed by the computation of cognitive dynamic dissimilarity matrices (CDM) using correlation-based measures. Finally, Distance Correspondence Indexes (DCI) were calculated by comparing the distance matrix across attentional types with the CDMs of each cognitive process. See Supporting Information Methods for further details. (b), (c) the results of the comparison of dynamic dissimilarities and their associations with psychiatric symptom scores. Error bars denote standard errors. # $P< 0.05$, uncorrected; *$P< 0.05$, **$P< 0.01$, ***$P< 0.001$, FDR-corrected for multiple comparisons over the number of statistical tests.

PCN Psychiatry and Clinical Neurosciences

Here, we sought to clarify the cognitive dynamic differences in internal and external attention using a new analytical method: Cognitive Dynamic Similarity Analysis (C-DSA, see Fig. 1a). Eighteen non-clinical adult participants completed the task. Internal/external attention was manipulated through an introspection task requiring participants to focus on each attentional dimension during emotion word-processing (see Table S1 and Fig. S1 for the methodology details and Fig. S3 for behavioral results). All fMRI scans underwent motion correction, realignment, and normalization to MNI space without spatial and temporal smoothing. A linear trend was removed from each time course. (See Supporting Information Methods for full preprocessing details.) Next, we extracted trial-wise cognitive dynamics of the two attention types using Cognitive Dynamics Estimation.[6] This method applies linear regression to an fMRI-derived two-dimensional matrix using spatial patterns of cognitive processes obtained from a meta-analytic platform as regression variables.[7] We focused on four *a priori* cognitive processes (*emotions*, *selective attention*, *self-referential thoughts*, and *working memory*; see Fig. S2 for full time series) while controlling for confounding factors (see Supporting Information Methods). We then computed dissimilarity indices between the two attentional types (internal *vs*. external) for each cognitive dimension. This approach allowed us to examine how independent cognitive processes differ dynamically between internal and external attention and to explore their relationship with psychiatric symptoms. Since difficulties in switching or differentiating between internal and external attention have been linked to psychiatric symptoms,[3] we hypothesized that greater similarity in cognitive dynamics across these attention types would correlate with symptom severity.

After computing the correlation matrix of each cognitive process, we obtained four distance-corresponding indexes (DCI), reflecting the difference in internal *vs*. external attention for each process (see Supporting Information Methods). All results were statistically significant (Wilcoxon signed-rank test, *emotions*: $Z = 3.72$, $p_{FDR} < 0.001$, $r = 0.88$; *selective attention*: $Z = 3.72$, $p_{FDR} < 0.001$, $r = 0.88$; *self-referential thoughts*: $Z = 3.72$, $p_{FDR} < 0.001$, $r = 0.88$; *working memory*: $Z = 3.33$, $p_{FDR} < 0.001$, $r = 0.79$, see Fig. 1b), indicating that cognitive dynamics significantly differed between attention types (see Fig. S4 for permutation tests with 1000 repetitions).

We observed significant negative associations between the DCI of selective attention and depression and state anxiety scores ($\beta = -0.002$, $p_{FDR} = 0.023$; $\beta = -0.002$, $p_{FDR} = 0.002$, Fig. 1c). This suggests that greater similarity in selective attention dynamics across internal and external attention types corresponds to more severe symptoms. Conversely, there was a significant positive association between the DCI of selective attention and trait anxiety ($\beta = 0.002$, $p_{FDR} = 0.002$, see Table S3 for all statistical values). Though these findings may seem paradoxical, they provide mechanistic evidence supporting the role of attention-training techniques in psychiatric symptom improvement. Specifically, interventions targeting attentional control could be effective in reducing symptoms related to rumination and anxiety.[8] Moreover, the findings suggest that cognitive training or tailored neurofeedback protocols could be used to reinforce more adaptive patterns of cognitive dynamics. For example, individuals with excessive similarity in selective attentional dynamics could be trained to increase the differentiation between internal and external attention using decoded fMRI neurofeedback.

In sum, C-DSA offers a novel method for quantifying temporal differences across cognitive dynamics, enabling the study of complex psychological constructs like internal and external attention from an ontological perspective.[9] This allows investigating associations between dynamic cognitive processes and psychiatric symptoms. While previous research has explored neural differentiation in psychiatric symptoms,[10] these studies primarily focused on static brain structures and functions, potentially limiting direct clinical applicability. In contrast, C-DSA identifies specific cognitive processes involved in psychiatric symptom modulation. This method could open new avenues for understanding dynamic cognitive mechanisms underlying complex mental functions, such as hierarchical decision-making and problem-solving, but also link these processes to dysfunctions in psychiatric issues.

Acknowledgments

We thank Kaori Nakamura for participants' recruitment and data collection, and Naoyuki Okamoto for advice about methodology.

Disclosure statement

KDDI Corporation funded this study; however, KDDI had no role in the study design, conclusions drawn, or publication decision. There are no other disclosures to report.

Data availability statement

The current study has not been pre-registered. Data, relevant scripts for all analyses, and other Supporting Information can be found at the Open Science Framework repository (DOI:10.176,05/OSF.IO/CQKNZ).

References

1. Chun MM, Golomb JD, Turk-Browne NB. A taxonomy of external and internal attention. *Annu. Rev. Psychol.* 2011; **62**: 73–101.
2. Keller AS, Leikauf JE, Holt-Gosselin B, Staveland BR, Williams LM. Paying attention to attention in depression. *Transl. Psychiatry* 2019; **9**: 279.
3. Amir I, Bernstein A. Dynamics of internal attention and internally-directed cognition: The attention-to-thoughts (A2T) model. *Psychol. Inq.* 2022; **33**: 239–260.
4. Freund MC, Etzel JA, Braver TS. Neural coding of cognitive control: The representational similarity analysis approach. *Trends Cogn. Sci.* 2021; **25**: 622–638.
5. Luo L, Li Q, Wang Y *et al*. Shared and disorder-specific alterations of brain temporal dynamics in obsessive-compulsive disorder and schizophrenia. *Schizophr. Bull.* 2023; **49**: 1387–1398.
6. Koyama Y, Yamamoto T, Hirayama J-I, Jimura K, Sadato N, Chikazoe J. Cognitive dynamics estimation: A whole-brain spatial regression paradigm for extracting the temporal dynamics of cognitive processes. *bioRxiv* 2023. doi.org/10.1101/2023.06.12.543130
7. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 2011; **8**: 665–670.
8. Knowles MM, Foden P, El-Deredy W, Wells A. A systematic review of efficacy of the attention training technique in clinical and nonclinical samples. *J. Clin. Psychol.* 2016; **72**: 999–1025.
9. Poldrack RA, Yarkoni T. From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annu. Rev. Psychol.* 2016; **67**: 587–612.
10. Tamman AJF, Abdallah CG, Dunsmoor JE, Cisler JM. Neural differentiation of emotional faces as a function of interpersonal violence among adolescent girls. *J. Psychiatr. Res.* 2024; **172**: 90–101.

Supporting Information

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Taiki Oka, PhD ,[1,2] Yutaro Koyama, MSc,[3,4] Akihiro Sasaki, PhD ,[5] Misa Murakami, MA,[1] Nao Kobayashi, PhD[5] and Aurelio Cortese, PhD[1]
[1]The Department of Decoded Neurofeedback, Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute International, Kyoto, [2]Clinical Psychology, Graduate School of Human Sciences, Osaka University, Suita, Japan, [3]Neuroscience Training Program, University of Wisconsin–Madison, Madison, Wisconsin, USA, [4]SandBox Inc., Tokyo, and [5]Data-Driven Healthcare Research Group, Life Science Laboratories, KDDI Research Inc., Saitama, Japan
Email: morio.sinri@gmail.com
Received 20 February 2025; revised 22 May 2025; accepted 28 May 2025.

Diagnosing schizophrenia spectrum disorders: Large language models (LLMs) vs. leading international psychiatrists (LIPs)

doi:10.1111/pcn.13864

The availability of large language models (LLMs) as putative diagnostic aids has a tangible transformative potential in many areas of science. Unlike many fields of medicine where diagnostic processes heavily rely on laboratory tests and imaging, psychiatric assessment fundamentally depends on narrative understanding and linguistic analysis of patients' experiences, emerging as relevant fields to further test LLMs potential for diagnostic purposes. A recent work by Li and colleagues[1] provides the first comprehensive evaluation of LLMs in psychiatric assessment by testing GPT-4, Bard, and Llama-2 against both a standardized licensing examination and real clinical scenarios presented as multiple-choice questions. Their study demonstrated that certain LLMs, particularly GPT-4, could achieve diagnostic accuracy approaching that of expert psychiatrists.

Furthermore, a recent study by Urkin *et al.*[2] demonstrated concerningly suboptimal levels of diagnostic accuracy among Leading International Psychiatrists (LIPs). Notably, when tasked with ascribing the best diagnostic estimate to two real-world clinical vignettes of schizophrenia spectrum disorders (SSDs), only 33% correctly identified both test cases (an overt and a more sublet clinical presentation). The study, aimed at benchmarking diagnostic precision in psychiatry for SSD by quantifying the diagnostic performance of LIPs on simulated cases, raises critical questions about diagnostic reliability. Yet, it provides an opportunity to assess LLMs on the same diagnostic task, extending the work from Li *et al.* in a more challenging scenario where LLMs must generate diagnoses without predefined options, specifically focusing on SSDs, and using a broader range of LLMs, including both closed-source and open-source models.

For this reason, we tested multiple state-of-the-art LLMs on the two vignettes presented in the original benchmark study[2] (see: https://pmc.ncbi.nlm.nih.gov/articles/PMC11207759/). Among open-source LLMs, we included large-scale architectures (DeepSeek-V3, Llama 3.1 405B, and Mixtral-8\u00D722B) and lightweight ones (Phi-3.5-mini, Ministral-8B, and Llama-3.2-3B). We also evaluated closed-source LLMs including GPT-4o, Claude-3.5-Haiku, and Gemini Flash 1.5. Each model was prompted in a Chain-of-Thought (CoT) fashion[3] (i.e., prompting the model to reason step-by-step through symptoms and diagnostic criteria to reach a clinical conclusion, see Data S1 for the prompt template visualization) to provide diagnostic impressions. All models were used in inference mode by setting the temperature to zero to ensure that the outcomes (reported in the following Github repository: https://github.com/Fede-stack/LLMs-vs-LIPs) would be deterministic and thus reproducible. The experiments were conducted following TRIPOD-AI research guidelines.[4] No ethical approval was required for this study as it involved the use of anonymized clinical vignettes and computational analysis of publicly available large language models, without human participant involvement or access to personal data. The research adhered to established guidelines for artificial intelligence research (TRIPOD-AI).

Figure 1 presents the results of the LLMs diagnostic evaluation of the two vignettes: closed-source and large-scale open-source LLMs correctly identified the diagnosis for both vignettes, while open-source lightweight LLMs showed selective competence, correctly diagnosing the first—most flamboyant SSD presentation—case, but misinterpreting the second—more attenuated SSD presentation—as various forms of depressive disorder.

While Li *et al.*[1] demonstrated that earlier LLMs could approach psychiatrists' performance in structured multiple-choice scenarios, our evaluation of more recent models shows that state-of-the-art LLMs are comparable to the top-performing 33% of diagnostically precise LIPs who correctly identified both SSD cases, without predefined diagnostic options. They effectively recognize both classical (e.g., David vignette) and subtle

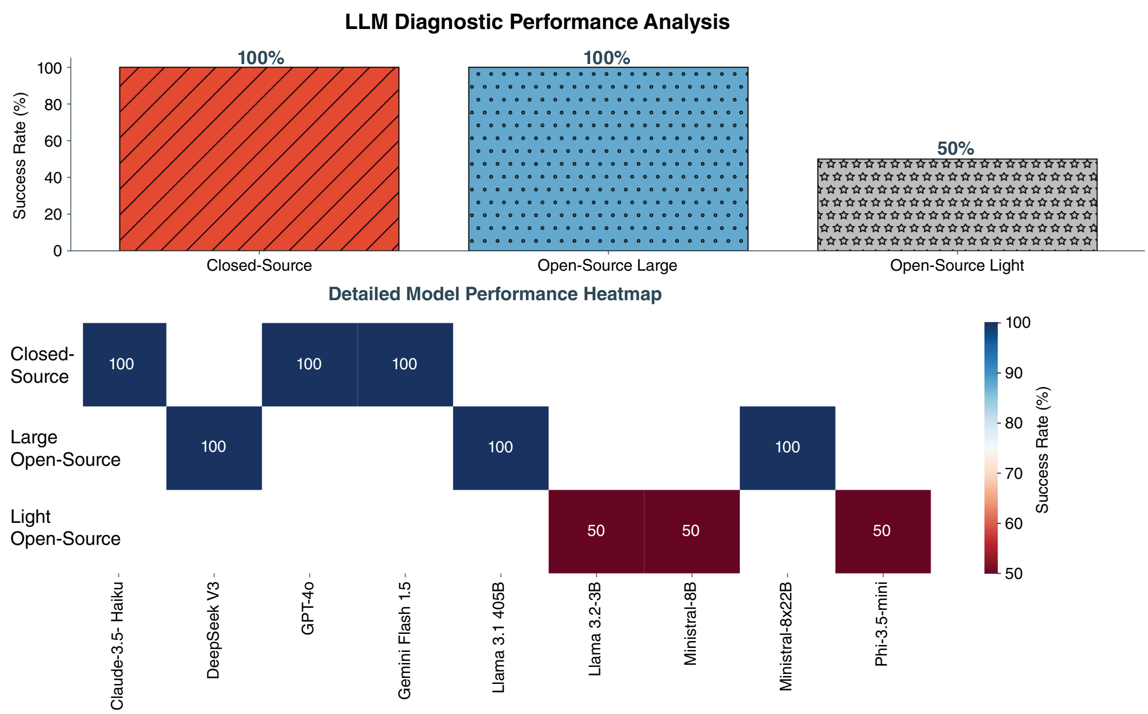

Fig. 1 Diagnostic ascriptions given by different closed- and open-source large language models (LLMs) on the two clinical vignettes previously used to benchmark leading international psychiatrists (LIPs) diagnostic accuracy. The barplot (up) shows the success rate by model category, while the heatmap (bottom) shows performance for each LLM considered.

PCN Psychiatry and Clinical Neurosciences

(e.g., Michael vignette) SSD presentations. These findings suggest that complex LLMs excel in psychiatric pattern recognition, particularly for nuanced clinical states, achieving expert-level diagnostic precision on textual case material, and outperforming simpler LLMs models. This corroborates their potential as decision-support tools to enhance diagnostic accuracy, reduce time-to-treatment in real-world settings and marks a significant step toward clinically implementable AI assistance in psychiatric diagnosis.

Limitations: (i) For comparability reasons, the clinical testing material is limited to the two SSD vignettes previously used to benchmark LIPs diagnostic performance; (ii) we tested a subset of LLMs, although representative of current state of the art. (iii) While our focus was on diagnostic performance, we acknowledge that implementation of LLMs in psychiatric practice would require addressing considerations around potential algorithmic biases, privacy, and appropriate integration with clinical judgment. While broader testing incorporating a range of diverse clinical presentations is desirable to further validate LLMs performance across other diagnostic categories, the current study confirms their potential for diagnostic precision. Closed-source and large-scale open-source LLMs achieve a level of accuracy comparable to that of the top-performing LIPs, whereas lightweight LLMs, though suboptimal, still align with the performance of a substantial fraction of LIPs. Our findings highlight the potential of LLMs in psychiatric education and clinical support, especially in recognizing complex diagnostic patterns. Large-scale training programs[5] could benefit from AI-guided decision support. Future research should investigate how LLMs might complement human expertise in psychiatric diagnosis while acknowledging the irreplaceable role of clinical experience and human judgment. Additionally, evaluating LLMs' diagnostic adaptability across different scenarios warrants further exploration.

Disclosure statement

The authors declare no conflicts of interest, financial or otherwise, related to the content of this study.

Acknowledgment

Open access publishing facilitated by Universita della Svizzera italiana, as part of the Wiley - Universita della Svizzera italiana agreement via the Consortium Of Swiss Academic Libraries.

Data availability statement

These data were derived from the following resources available in the public domain: Github Repository and the data that support the findings of this study are available in Github Repository at https://github.com/Fede-stack/LLMs-vs-LIPs.

References

1. Li DJ, Kao YC, Tsai SJ *et al*. Comparing the performance of ChatGPT GPT-4, bard, and Llama-2 in the Taiwan psychiatric licensing examination and in differential diagnosis with multi-center psychiatrists. *Psychiatry Clin. Neurosci.* 2024; **78**: 347–352.
2. Urkin B, Parnas J, Raballo A, Koren D. Schizophrenia Spectrum disorders: An empirical benchmark study of real-world diagnostic accuracy and reliability among leading international psychiatrists. *Schizophr. Bull. Open* 2024; **5**: sgae012.
3. Wei J, Wang X, Schuurmans D *et al*. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* 2022; **35**: 24824–24837.
4. Collins GS, Moons KGM, Dhiman P, Logullo P. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; **385**: e078378.
5. Bajaj S. How students and grandparents could solve the global mental-health crisis. Available from URL: https://www.nature.com/articles/d41586-024-03757-9?utm_source=Live+Audience&utm_campaign=47421e5c3c-nature-briefing-daily-20241119_COPY_01&utm_medium=email&utm_term=0_b27a691814-47421e5c3c-50614976.

Supporting Information

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Andrea Raballo, MD, PhD [1,2] Federico Ravenda, Msc[3] and Antonietta Mira, PhD[4,5]

[1]REthinking MEntal Health Through Clinical and Data Intelligence (REMEDI) Lab, Faculty of Biomedical Sciences, Euler Institute, Università della Svizzera Italiana, Lugano, [2]Cantonal Sociopsychiatric Organisation, Public Health Division, Department of Health and Social Care, Repubblica e Cantone Ticino, Mendrisio, [3]Department of Informatics, Università della Svizzera Italiana, [4]REthinking MEntal Health Through Clinical and Data Intelligence (REMEDI) Lab, Department of Economics, Euler Institute, Università della Svizzera Italiana, Lugano, Switzerland, and [5]Department of Science and High Technology, University of Insubria, Como, Italy

Email: andrea.raballo@usi.ch

Received 15 April 2025; revised 26 May 2025; accepted 20 June 2025.