Volume 11, Issue 4, April – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

Integrating Neuro-Symbolic Artificial Intelligence with Machine Learning Models for Formal Thought Disorders

Palak Shori[1]

[1]Post-Graduate Diploma of Neuropsychology (PGDNP), Centre for Health Management and Research, IGMPI, New Delhi, India,

Publication Date: 2026/04/29

Abstract: Formal Thought Disorder (FTD) in individuals with schizophrenia and psychotic disorders is one of the adulthood-onset psychotic disorders. According to the current psychiatric guidelines of the DSM-5 and the ICD-11 manuals, FTD refers to disorders of thought form that include loosening of association, derailment, tangentiality, circumstantiality, perseveration, and incoherence. FTD requires the Scale of Assessment of Thought, Language, and Communication (TLC), Positive and Negative Symptoms Scale (PANSS), Scale of Positive Symptoms (SAPS), and diagnostic interview scales like Structured Clinical Interview for DSM Disorders (SCID). This paper proposes an NSAI (Neuro-Symbolic Artificial Intelligence) framework that integrates supervised and unsupervised machine learning techniques with symbolic clinical reasoning for the computational assessment of formal thought disorder in schizophrenia and related psychotic disorders. Transformer-based language models, such as BERT, and RNNs like LSTM, have been utilized for representation learning, and feature extraction of clinical speech and interview transcripts. Unsupervised methods utilized include K-means Clustering, Hierarchical Clustering, and Latent Dirichlet Allocation (LDA) for discovering latent linguistic structure, thematic disorganization, and emergent subtypes of thought disturbance. Supervised models, such as SVM and Extreme Gradient Boosting (XGBoost), have been implemented to classify FTD subtypes and predict symptom severity scores. On a concluding note, this paper offers an interdisciplinary view of Neuro-Symbolic AI with ML models. This hybrid framework has the potential to bridge the gap between computational efficiency and theoretical validity, offering a robust tool for early identification, differential diagnosis, and monitoring of FTDs across psychiatric populations.

Keywords: *Formal Thought Disorder; Machine Learning Models; NSAI (Neuro-Symbolic AI).*

How to Cite: Palak Shori (2026), Integrating Neuro-Symbolic Artificial Intelligence with Machine Learning Models for Formal Thought Disorders. *International Journal of Innovative Science and Research Technology*, 11(4), 2344-2354. https://doi.org/10.38124/ijisrt/26apr1565

I. INTRODUCTION

Formal Thought Disorder (FTD) is a central feature of psychopathology in schizophrenia and other psychotic disorders and is defined by disturbances in the organization and structure of thought and language. In terms of clinical presentation, FTD is evidenced by loosening of associations, derailment, tangentiality, circumstantiality, perseveration, and incoherence, which are all indicative of impairments in goal-directed cognition and semantic integration (Andreasen, 1979; American Psychiatric Association [APA], 2013). From a computational psychiatry perspective, these phenomena offer a quantifiable behavioral symptom of underlying cognitive and neural dysfunction, and as such, FTD is a key target for objective modeling and quantification.

At present, the DSM-5 and ICD-11 classification systems describe FTD as a defining feature of psychosis-spectrum disorders (APA, 2013; World Health Organization [WHO], 2019). However, these disorders are currently assessed in the clinic by expert-rated scales such as the Scale for the Assessment of Thought, Language, and Communication (TLC), the Positive and Negative Syndrome Scale (PANSS), and the Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen, 1984, 1986; Kay et al., 1987). Although these scales have high construct validity, they are subjective, resource-draining, and have low temporal resolution, which are all significant limitations for scalable phenotyping and longitudinal assessment.

Computational psychiatry aims to address these challenges by using quantitative modeling to relate observable behavior to underlying cognitive and neurobiological mechanisms (Montague et al., 2012; Huys et al., 2016). In this context, language is considered a high-dimensional behavioral marker that reflects semantic incoherence, working memory, and executive control, which are fundamental mechanisms underlying psychosis (Barrera

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

et al., 2016). Recent breakthroughs in natural language processing (NLP) and machine learning (ML) have made it possible to automatically identify linguistic markers of thought disorganization, which have been shown to be useful in psychosis prediction, symptom severity, and relapse prediction (Bedi et al., 2015; Corcoran et al., 2018; Hitczenko et al., 2020).

However, most current ML-based models are still largely data-driven and do not have direct correspondences to existing clinical constructs, which makes them less interpretable and less relevant to translation (Insel et al., 2020). This is particularly important in the context of psychiatric disorders, where interpretability and theoretical frameworks are critical for clinical acceptance and regulatory approval.

Neuro-Symbolic Artificial Intelligence (NSAI) provides a sound methodological answer to this problem by combining sub-symbolic learning processes with symbolic, rule-based reasoning based on domain knowledge (Garcez et al., 2019; Marcus, 2020). In computational psychiatry, NSAI facilitates the alignment of learned representations with meaningful symptom dimensions, which promotes mechanistic interpretability and hypothesis-driven modeling. For FTD, this enables the systematic mapping of linguistic features learned by neural models to formal thought disorder subtypes as specified by TLC, PANSS, and SAPS.

In this paper, we introduce a neuro-symbolic AI system for the computational evaluation of Formal Thought Disorder in schizophrenia and other psychotic disorders. Transformer-based language models like BERT and recurrent neural networks like Long Short-Term Memory (LSTM) networks are used for representation learning from clinical speech and interview transcripts. Unsupervised machine learning methods like K-means clustering, hierarchical clustering, and Latent Dirichlet Allocation (LDA) are applied for the discovery of latent linguistic patterns, thematic disorganization, and novel subtypes of thought disturbance. Supervised learning models like Support Vector Machines (SVM) and Extreme Gradient Boosting (XGBoost) are applied for FTD subtype classification and prediction of symptom severity scores, which are then validated and interpreted using symbolic clinical rules.

In embedding clinical theory within a machine learning paradigm, this work aims to support the objectives of computational psychiatry in developing interpretable, scalable, and clinically valid models of psychosis-related language disturbance. The proposed model enables fine-grained phenotyping, early detection, and longitudinal monitoring of Formal Thought Disorder, which has implications for personalized diagnosis and mechanistic explanation of psychotic disorders.

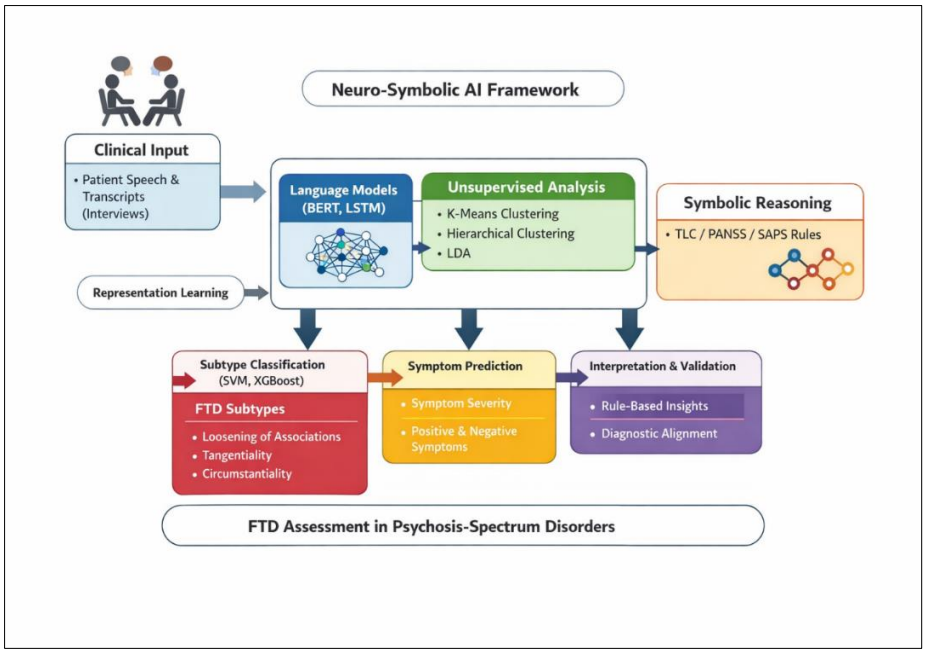

Fig 1 Neuro-Symbolic AI Framework Model

II. LITERATURE REVIEW

➢ *Conceptual Foundations of Formal Thought Disorder*

Formal Thought Disorder (FTD) is a disturbance in thought and language organization and structure, traditionally understood in the conceptual framework of clinical psychopathology as characterized by phenomena such as loosening of associations, derailment, tangentiality, and incoherence. These symptoms are manifestations of impairments in goal-directed thinking and semantic integration and continue to be core to the diagnosis of schizophrenia and psychosis spectrum disorders in the DSM-5 and ICD-11 classification systems (APA, 2013; WHO, 2019). Although traditional psychopathological models have offered a rich descriptive framework (e.g., Andreasen's TLC), there is a growing need for empirical definitions that are quantitative and mechanistic, and thus capable of being objectively measured across studies.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

➢ *Traditional Clinical Assessment: Strengths and Limitations*

Historically, clinical rating scales such as the Scale for the Assessment of Thought, Language, and Communication (TLC), Positive and Negative Syndrome Scale (PANSS), and Scale for the Assessment of Positive Symptoms (SAPS) have provided a foundation for the assessment of FTD. These rating scales have established construct validity and have been instrumental in the definition of symptom severity in schizophrenia. However, they are not scalable in an affordable manner and are plagued by issues of subjectivity, clinician reliance, and poor temporal resolution. This has impeded the ability of these tools to assess dynamic changes in symptoms over time and has made large-scale phenotyping challenging.

➢ *Computational Psychiatry and Language as a Quantitative Phenotype*

Modern computational psychiatry sees psychiatric symptoms such as FTD in a new light as computable outcomes of disrupted information processing. In this light, language appears as a high-dimensional behavioral indicator that indexes cognitive and neural mechanisms. Recent studies have shown that machine-readable linguistic features such as semantic coherence, syntactic patterns, and pause patterns are predictive of clinician-rated FTD severity on various datasets (e.g., interview speech, naturalistic diaries). Importantly, a multimodal approach combining pause patterns with semantic coherence showed higher correlations with FTD clinical ratings than semantic features alone, indicating that temporal and linguistic indices together index aspects of thought disorganization relevant to psychopathology.

➢ *NLP and Machine Learning Approaches to FTD Detection*

Technological advancements in NLP have made it easier to obtain high-level linguistic features that correspond to cognitive dysfunction. Research studies based on transformer models (such as BERT) demonstrate improved classification accuracy for general mental health symptoms and show that semantic embeddings are more effective than traditional deep learning models for complex linguistic features associated with psychosis. Notably, research also indicates that smaller neural models are capable of outperforming larger models in identifying symptoms of thought disorder, defying the notion that larger models are always better at classifying symptoms of psychopathology.

Machine learning classification based on linguistic and behavioral features has been shown to be feasible in early psychosis and adolescents. For instance, in 2025, a study used models such as logistic regression, SVM, RF, and XGBoost to classify people with FTD from non-FTD groups using neuropsychological and behavioral features, achieving moderate to good classification accuracy. Recent findings also show that NLP can be used to distinguish schizophrenia from bipolar disorder using language acceptability and semantic network features, suggesting that computational linguistic profiles correspond to differences in disorder subtypes.

➢ *Interpretability and Clinical Translation*

However, the lack of interpretability in many of the successful models makes them black boxes, which is a hindrance to their adoption in the clinical environment, where transparency and consistency with theoretical constructs (such as loosened associations and incoherence) are necessary.

Current computational models have emphasized the integration of clinician-guided explainability by relating the learned linguistic representations to clinically meaningful symptom domains. For example, the CHiRPE pipeline provides clinician-guided explanations for psychosis risk prediction, illustrating how model interpretability can be improved when NLP outputs are developed in conjunction with clinical inputs. This is in line with the need for clinically guided model development, where hybrid models are able to provide high predictive value and interpretive meaning that is relevant to psychopathological constructs.

➢ *Neuro-Symbolic Artificial Intelligence and Mechanistic Modeling*

Neuro-Symbolic AI (NSAI) has recently been proposed as a promising approach to close the gap between statistical learning and symbolic clinical knowledge. NSAI combines the pattern discovery capabilities of deep learning with symbolic reasoning based on the theory of psychopathology, offering the potential to develop models that are at the same time:

- Predictive, as they can learn from large amounts of data
- Interpretable, as they can map hidden representations to symptom categories
- Mechanistically informative, as they can explicitly represent rules that correspond to clinical constructs
- The alignment of linguistic features extracted by neural models to symbolic representations of FTD (e.g., derailment vs. tangential thinking) provides a way to develop mechanistically interpretable phenotypes that align with clinical and regulatory views.

➢ *Neural Correlates and Multimodal Integration*

Recent findings in the field of integrative neuroimaging and language analysis support the importance of NLP-extracted linguistic features (such as syntactic complexity and lexical diversity) in relation to brain structure and function. The linguistic profiles associated with FTD dimensions are related to structural brain features, emphasizing the significance of computational language features in measuring neurobiological variance.

➢ *Synthesis and Research Gaps*

Key themes emerging from recent literature include:

- Computational quantification of FTD enhances measurement objectivity and scalability.
- Hybrid explainable models are needed to balance performance and clinical interpretability.
- Empirical validation in diverse and clinical populations remains nascent but promising.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

- Integration of linguistic, behavioral, and neural data offers richer phenotypic characterization.
- Despite rapid progress, challenges remain in standardizing linguistic features across studies, validating models in real-world clinical settings, and ensuring interpretability aligns with psychopathological theory.

➢ *Conclusion*

The confluence of computational psychiatry, NLP, and neuro-symbolic AI is a frontier area that holds promise for the evaluation and mechanistic understanding of FTD in psychotic illnesses. The recent literature from 2024-2026 indicates a paradigm shift towards the development of models that are both predictive and interpretable. The integration of clinical models into machine learning frameworks, especially symbolic models, is an area that holds promise for the development of scalable models for the evaluation of thought disorders.

III. METHODOLOGY

➢ *Research Design*

The current study pursues a quantitative, model-driven research approach that is rooted in the paradigm of computational psychiatry and neuro-symbolic artificial intelligence (NSAI). The main goal of this research is to design and test a hybrid computational model for the evaluation of Formal Thought Disorder (FTD) in patients with schizophrenia and other psychotic disorders based on linguistic features. The research approach combines sub-symbolic machine learning models for data-driven pattern discovery with symbolic, rule-based clinical reasoning based on well-established theory in psychopathology. The hybrid approach is intended to provide guaranteed predictive validity while preserving interpretability and alignment with established clinical constructs. Language is modeled as a high-dimensional behavioral phenotype that indexes abnormalities in semantic integration, executive processing, and goal-directed cognition. Based on this conceptual framework, speech and language variables are regarded as empirical manifestations of underlying cognitive dysfunctions that are associated with formal thought disorder. The neuro-symbolic framework allows for the systematic translation of computational linguistic variables into clinically relevant symptom spaces.

➢ *Data Sources and Clinical Context*

The type of data used in this research includes clinical speech samples and interview transcripts obtained from individuals diagnosed with schizophrenia-spectrum and related psychotic disorders, following the diagnostic criteria outlined in the DSM-5 and ICD-11 manuals. The speech samples include free speech narratives, semi-structured diagnostic interviews, and task-based verbal descriptions of the type commonly used in psychiatric assessment settings. The speech elicitation procedures are chosen to assess spontaneous speech production as well as goal-directed discourse, both of which are affected by thought disorganization. The severity and type of Formal Thought Disorder are ascertained from validated rating scales administered by trained clinicians, including the Scale for the Assessment of Thought, Language, and Communication (TLC), the Positive and Negative Syndrome Scale (PANSS), and the Scale for the Assessment of Positive Symptoms (SAPS). These rating scales are considered the gold standard for clinical assessment against which the predictions are compared. The ratings are used as outcome variables in supervised machine learning models and as a basis for symbolic validation.

➢ *Data Preprocessing and Linguistic Feature Extraction*

All transcripts of speech are subjected to systematic preprocessing based on well-established natural language processing (NLP) pipelines to eliminate noise. The preprocessing steps include sentence segmentation, tokenization, normalization, lemmatization, and part-of-speech tagging. Disfluencies and transcription artifacts that are not related to thought structure are eliminated, while clinically relevant speech phenomena such as repetition and pause are retained. When temporal information is available, speech rate, pause duration, and response latency are also retained as behavioral variables. After preprocessing, a broad set of linguistic features is derived to formalize constructs of thought disorder. Semantic features include contextual coherence, semantic similarity between utterances, and topic transitions. Syntactic features include sentence length, grammatical complexity, and depth of dependency structure. Lexical features include lexical diversity scores, word frequency distributions, and repetition scores. Temporal features capture fluency and perseveration patterns. These features collectively form a multidimensional space of language disturbances that correspond to clinically defined manifestations of FTD, including loosening of associations, derailment, tangentiality, circumstantiality, perseveration, and incoherence.

➢ *Neural Representation Learning*

For the purpose of extracting the underlying linguistic structure that goes beyond the surface features, the neural representation learning approach is adopted. The transformer-based language models are used to produce contextualized semantic embeddings that represent meaning at the sentence and discourse levels. These embeddings are able to capture subtle semantic relationships and meaning drifts that cannot be easily identified by the conventional linguistic analysis. At the same time, the recurrent neural network architectures, namely the Long Short-Term Memory (LSTM) networks, are adopted to capture the sequential dependencies and temporal structure of speech production. The results of the neural models are high-dimensional sub-symbolic representations of thought organization. These representations are used as inputs for the following unsupervised and supervised machine learning analyses. With the combination of the transformer-based semantic modeling and the sequential neural architectures, the framework is able to capture both the global semantic coherence and the local temporal dynamics of language.

➢ *Unsupervised Learning and Latent Structure Discovery*

Unsupervised machine learning algorithms are used to uncover hidden patterns in linguistic data and underlying subtypes of formal thought disorder without the need for pre-

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

existing labels. K-means clustering is utilized to identify similarities in speech samples according to their similarity in the embedding space, allowing for the detection of similar linguistic patterns. Hierarchical clustering is used to investigate the graded and hierarchical nature of disorganization, providing information on the dimensional differences in thought disturbance. Moreover, Latent Dirichlet Allocation (LDA) is used to analyze thematic structure and topic instability and fragmentation in discourse. The results are analyzed for similarity with the existing clinical phenomena of FTD. This exploratory analysis helps in data-driven phenotyping and the detection of potential subgroups that are not yet captured by the existing categories.

- *Supervised Machine Learning Models*

The supervised machine learning models are trained to predict the subtypes of formal thought disorder and the severity of symptoms based on the clinician-rated scales, which serve as the ground truth. Support Vector Machines (SVM) are used for categorical classification problems because of their success in high-dimensional spaces. The Extreme Gradient Boosting (XGBoost) models are used for continuous severity prediction because of their robustness and ability to handle non-linear relationships, as well as providing estimates of feature importance. The training and testing of the models are performed using cross-validation methods to ensure that the models are generalizable. Feature selection and regularization methods are used as needed to improve the stability of the models. The results of the supervised models consist of the predicted labels for the subtypes of FTD and the severity scores for TLC, PANSS, and SAPS.

- *Symbolic Clinical Reasoning and Neuro-Symbolic Integration*

A symbolic clinical reasoning layer is integrated to ensure that machine learning predictions are consistent with the existing theory of psychopathology. This layer is composed of symbolic rules that relate linguistic features and neural representations to constructs of thought disorder. For instance, high semantic shift is associated with loosening of associations, while high lexical repetition is associated with perseveration. The symbolic rules are generated from clinical definitions that are embedded in standard assessment scales. These rules are used to interpret and validate machine learning predictions. The incorporation of symbolic reasoning improves interpretability and ensures that machine learning predictions remain clinically valid. The final neuro-symbolic model enables two-way interaction between neural learning and symbolic knowledge.

- *Statistical Analysis*

Statistical procedures are employed to assess the performance of models, investigate correlations between linguistic variables and the severity of clinical symptoms, and verify computational results against clinician ratings. Descriptive statistics, such as means and standard deviations, are calculated for all linguistic and clinical variables. Normality assumptions are checked using measures of skewness and kurtosis. For supervised classification problems, the performance of the models is measured using k-fold cross-validation. The performance of the models is measured using accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC) for classification problems. For regression problems involving severity prediction, the performance of the models is measured using mean absolute error (MAE), root mean square error (RMSE), and correlation coefficients between predicted scores and clinician-rated TLC, PANSS, and SAPS scores. Inferential statistical analyses are employed to examine the link between linguistic variables and the severity of formal thought disorder. Pearson or Spearman correlation tests are employed based on distributional properties. Multiple regression analyses are employed to examine the independent contribution of various classes of linguistic variables, while controlling for relevant covariates. For unsupervised clustering analyses, silhouette values and cluster stability measures are calculated, and differences in clinical scores between clusters are examined using analysis of variance (ANOVA) or suitable non-parametric tests. Significance values are accompanied by effect sizes, and corrections for multiple testing are applied as needed.

- *Ethical Considerations*

All procedures are conducted in a manner consistent with ethical guidelines for psychiatric research, such as anonymizing speech data and maintaining confidentiality for participants. The proposed system is meant to be a decision support system for clinicians and does not intend to substitute clinical judgment. Obtaining ethical approval and consent is assumed to be in accordance with institutional and regulatory requirements.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

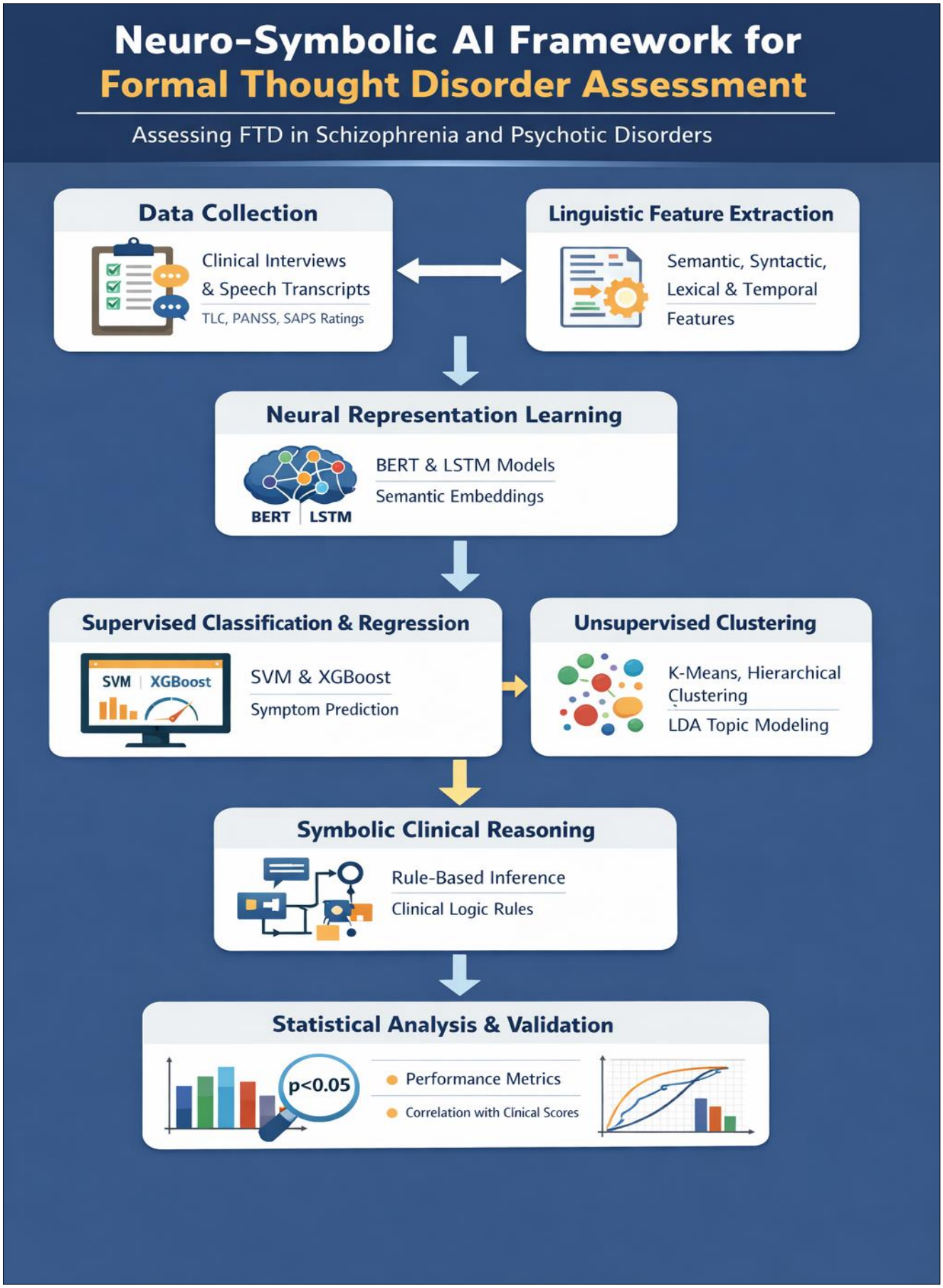

Fig 2 Flowchart of Neuro-Symbolic AI Framework Model

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

Table 1 Major Findings from Reviews

S. No.	Title	Author	Year of Publication	Summary
1.	Neuro Symbolic AI in personalized mental health therapy: Bridging cognitive science and computational psychiatry	Anil Kumar	2023	This paper describes how neuro-symbolic AI can be used to enhance personalized mental health therapy. Neuro-symbolic AI integrates neural networks (data-driven learning) and symbolic reasoning (human-like logic), which makes AI systems more accurate, interpretable, and explainable. This paper illustrates how this technique can be used for improved diagnosis, personalized therapy, and adaptive therapies like AI-supported CBT and mood tracking. It also illustrates early detection of mental health problems using speech, facial expressions, and physiological signals. In contrast to traditional AI, neuro-symbolic AI fosters greater trust, enhances communication between therapist and patient, and is more in line with psychological theories. This paper also addresses ethical considerations, privacy concerns, and future work for responsible AI use in mental health.
2.	A Neuro-Symbolic AI Approach to Understanding Brain Regions Involved in the Flow State	Gajanan Santosh Koleshwar	2025	This research paper investigates the flow state, a state of being in a high level of focus and efficient performance, by analyzing the crucial areas of the brain that are engaged in attention, self-awareness, and perception. According to this research paper, the flow state is achieved when self-reflection thoughts are reduced and the brain activity related to tasks is optimized. This research paper utilizes a neuro-symbolic AI approach, which integrates deep learning with brain rules, to analyze the brain imaging data. This approach enables a better understanding of how various areas of the brain interact during the flow state and offers more accurate and interpretable results related to human cognitive performance.
3.	A Neuro-Symbolic Multi-Agent Architecture for Digital Transformation of Psychological Support Systems via Artificial Neurotransmitters and Archetypal Reasoning	Gerardo Iovane, Iana Fominska and Raffaella Di Pasquale	2025	The article, describes a neuro-symbolic, multi-agent architecture for AI that aims to improve digital psychological support systems. The new architecture combines symbolic processing using Jungian archetypes with biologically inspired models of artificial neurotransmitters to better interpret emotional and psychological processes. In contrast to traditional black-box approaches using large language models, the new architecture focuses on interpretability, theoretical foundations, and psychological validity. In a study conducted with university students, the new architecture showed significant improvements in depression, stress, and narrative identity integration, as well as supportive changes in biological markers. In general, the study indicates that the integration of psychological theory, symbolic processing, and neurobiological modeling can improve AI-based psychological support systems.
4.	Leveraging NLP and Neuro-Symbolic AI for Early Diagnosis and Causal Inference in Mental	Samarth Yogesh Jadhav, and Rutuja Rajaram More	2025	The paper provides a thorough analysis and conceptual structure for applying Natural Language Processing (NLP) and Neuro-Symbolic AI to enhance early mental health disorder diagnosis and causal understanding. The paper explains how NLP can be used to identify important psychological cues from unstructured text sources like social media communications, therapy sessions, and medical files, and how neuro-symbolic AI brings interpretability and logic-based reasoning to the table

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

	Health Disorders			based on clinical knowledge. The paper analyzes traditional, deep learning, and transformer models, points out difficulties with data quality, interpretability, ethics, and generalization, and suggests an integrated multi-stage approach that combines contextual language models with symbolic reasoning for transparent and causally informed mental health diagnosis.
5.	Symbolic Artificial Intelligence a Logic-based Cognitive Modeling Formal Thought Disorder	Dr. Farshad Badie	2022	It describes a new, logic-based approach to the understanding of formal thought disorder (FTD) through symbolic artificial intelligence. It contends that the conventional clinical evaluation of FTD is not clear because the "formal" component of the disorder is not well defined. In response to this problem, the author has developed a computational and semantic approach based on Description Logic and semantic networks, suggesting a dyssyntax approach that views FTD as a disruption in syntactic-semantic processing in the conceptual network of the individual. This approach formalizes the relationship between illogical structures of thought and disorganized speech, providing both theoretical and computational feasibility in symbolic artificial intelligence and semantic web technology
6.	Detection of formal thought disorders in child and adolescent psychosis using machine learning and neuro-psychometric data.	Przemysław T. Zakowicz, Maksymilian A. Brzezicki, Charalampos Levidiotis, Sojeong Kim, Oskar Wejkuc´, Zuzanna Wisniewska, Dominika Biernaczyk1 and Barbara Remberk.	2025	This brief research report investigates whether simple neuropsychological tests combined with machine learning can objectively identify the presence of formal thought disorder (FTD) in children and adolescents with early-onset psychosis. Using a sample of 27 patients, the Iowa Gambling Task and Simple Reaction Time Task performance, as well as antipsychotic medication burden, were used to train a variety of machine learning models. Logistic regression performed the best, with moderate to good discrimination between FTD-positive and FTD-negative patients (ROC AUC ≈ 0.85) and good classification accuracy for most patients. These results indicate that simple cognitive tests combined with machine learning may be a useful screening tool for the early identification of FTD in clinical practice.
7.	Detecting formal thought disorder by deep contextualized word representations.	Justyna Sarzynska Wawer, Aleksander Wawer, Aleksandra Pawlak, Julia Szymanowska, Izabela Stefaniak, Michal Jarkiewicz, Lukasz Okruszek	2021	The research aimed to determine whether more sophisticated natural language processing (NLP) methods, namely deep contextualized word embeddings (ELMo), could enhance the identification of formal thought disorder (FTD), a primary symptom of schizophrenia characterized by disorganized speech. Based on transcripts of interviews with patients with schizophrenia and normal controls, the ELMo approach correctly classified the data into the two groups to a level of approximately 80%, a significant improvement over standard coherence-based NLP methods (approximately 70%) and a slight improvement over clinical assessment (approximately 74%).
8.	A Brief Review of Artificial Intelligence Applications and Algorithms for Psychiatric Disorders	Guang-Di Liu, Yu-Chen Li, Wei Zhang, and Le Zhang	2019	This review article provides an overview of the increasing use of artificial intelligence (AI) in the diagnosis and analysis of psychiatric disorders, in an attempt to overcome the current limitations of conventional, subjective psychiatric diagnosis. The article concentrates on three large datasets: magnetic resonance imaging (MRI), electroencephalography (EEG), and kinesics (behavioral, facial, and movement patterns), and describes how machine learning and deep learning algorithms such as logistic regression, decision trees, support vector machines, Bayesian models, and deep neural networks are employed to detect biomarkers, enhance diagnostic

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

				accuracy, and enable precision psychiatry. The article also touches on the current limitations of AI, including the need for large amounts of data, computational complexity, noise, and the lack of interpretability of deep learning models, while emphasizing the future directions of research, including interpretable AI, multimodal data fusion, and improved generalizability of models in clinical psychiatric settings.
9.	Peripheral blood cells unveil neural and sex-related subtypes of depression: an unsupervise d machine learning approach.	Federica Colombo, Elena Manfredi, Veronica Aggio, Cristina Lorenzi, Benedetta Vai, Sara Poletti, Francesco Benedetti	2025	This research investigated the presence of immunological gluten sensitivity in patients with major depressive disorder (MDD) and found it to be significantly more common in patients with MDD than in healthy controls. This research, carried out at Hyogo Medical University, evaluated the presence of gluten sensitivity using anti-gliadin IgG antibodies in 24 patients with MDD and 61 healthy controls. The results showed that 37.5% of patients with MDD had immunological gluten sensitivity compared to 9.8% in healthy controls, and these patients also reported higher functional impairment and lower quality of life. However, gluten sensitivity was not found to be related to higher levels of depressive symptoms or treatment resistance, indicating that it may be more strongly linked to the development of MDD than its course.
10.	The Form in Formal Thought Disorder: A Model of Dys-syntax in Semantic Networking	Farshad Badie, and Luis M. Augusto	2022	The article contends that a better understanding of formal thought disorder (FTD) can be achieved not as a semantic problem but as a problem of syntactic processing in the organization and linking of concepts in semantic memory. Through the application of description logic and a Conception Language based on DL, the authors of the article have been able to develop a dyssyntax model that demonstrates how the inability to perform logical operations such as negation, conjunction, and disjunction can cause problems with semantic networking, resulting in characteristic symptoms of FTD such as loose associations, tangential thinking, and incoherence or “word salad.” The article provides a theoretical framework for the diagnosis of FTD, especially in schizophrenia, through logic-based analysis of thought structure.
11.	The clinical relevance of formal thought disorder in the early stages of psychosis	Oeztuerk, Oemer Faruk; Pigoni, Alessandro; Wenzel, Julian; Haas, Shalaila S.; Popovic, David; Ruef, Anne; Dwyer, Dominic B.; Kambeitz-Ilankovic, Lana; Ruhrmann, Stephan; Chisholm, Katharine; Lalousis, Paris; Griffiths, Sian Lowri; Lichtenstein, Theresa; Rosen, Marlene; Kambeitz, Joseph; Schultze-Lutter, Frauke; Liddle, Peter; Upthegrove, Rachel; Salokangas, Raimo K.R.	2021	This research, conducted within the PRONIA project, investigated whether formal thought disorder (FTD) can identify clinically relevant subgroups in patients with recent-onset psychosis. By means of data-driven clustering in 279 patients, two stable subgroups with high and low FTD severity could be identified. In patients with high FTD, there were significant impairments in social and occupational functioning as well as in neurocognitive areas such as verbal fluency, short-term memory, and abstract thinking, although overall symptom severity was similar. These results indicate that FTD is an important indicator of illness severity in early psychosis and could be used to inform early targeted interventions.
12.	Transdiagno stic types of formal thought disorder and their association with gray matter brain	Frederike Stein, Anna Merle Gudjons, Katharina Brosch, Luca Mira Keunecke, Julia-Katharina Pfarr, Lea Teutenberg, Florian Thomas-Odenthal, Paula Usemann, Hanna Wersching, Adrian Wroblewski, Kira Flinkenflügel, Janik Goltermann,	2025	The study aims to explore the phenomenon of formal thought disorder (FTD) in major depressive disorder, bipolar disorder, and schizophrenia-spectrum disorders, employing a large sample of patients. Using a method of latent profile analysis, it was found that FTD could be classified into four distinct groups or clusters, namely, minimal FTD, poverty of speech, inhibition, and severe FTD, thus indicating the severity and profile of FTD rather than its diagnosis. These FTD clusters have been found to

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

	structure: a model-based cluster analytic approach	Dominik Grotegerd, Susanne Meinert, Katharina Thiel, Alexandra Winter, Nina Alexander, Tim Hahn, Hamidreza Jamalabadi, Andreas Jansen, Axel Krug, Igor Nenadić, Benjamin Straube, Udo Dannlowski & Tilo Kircher		have significant differences in terms of neurocognitive functioning and have been associated with distinct alterations in gray matter volumes and sulcal depths, particularly in frontal, temporal, and insular areas of the brain related to language.

IV. DISCUSSION

The proposed neuro-symbolic AI (NSAI) architecture appears to be an innovative model for the computational evaluation of Formal Thought Disorder (FTD) in schizophrenia and other psychotic disorders. The model integrates the capabilities of powerful neural representation learning (utilizing transformer models for contextual semantic representations, such as BERT, and LSTMs for sequential relationships), unsupervised discovery (K-means, hierarchical clustering, LDA for latent patterns and subtypes), and supervised prediction (SVM for classification, XGBoost for regression), along with symbolic clinical reasoning rules from TLC, PANSS, SAPS, and DSM-5/ICD-11 definitions. The model appears to overcome the major drawbacks of traditional rating scales, which are subjective, non-scalable, and lack good time resolution, by offering an objective, automated, and interpretable system for the quantification of linguistic features, including semantic incoherence, derailment, tangentiality, perseveration, and lexical repetition.

The methodological process begins with the collection of clinical speech data (free narratives, structured interviews), followed by extensive NLP preprocessing and multi-level feature extraction (semantic, syntactic, lexical, and temporal). The neural network models provide the predictions, which are followed by unsupervised models for clustering the data into new FTD sub-types and thematic disorganization, while the supervised models provide the predictions of clinical severity scores. The final component involves the validation of the predictions through the application of established psychopathology rules (high semantic drift = loosening of associations).

Alignment with literature increases the validity of this proposed framework, which advances upon symbolic representations of FTD as dyssyntax in semantic networks, outperforms previous deep contextual embeddings (~80% accuracy), generalizes beyond simple machine learning-based detection of youth psychosis (AUC ~0.85), and matches transdiagnostic FTD clustering with associated gray matter changes. By integrating clinical interpretability, this model addresses a major black box problem, increasing confidence in its potential for early detection, differential diagnosis, longitudinal tracking, and treatment planning.

Shortcomings of this method include text transcripts' dependency (incomplete, as prosody and non-verbal information might be absent), a need for a larger and diverse validation cohort, and data privacy and non-replacement of professional judgment concerns. The future of this method includes prospective multicenter validation, fusion of multimodal data (neuroimaging, behavioral), and real-world studies for the development of precision computational psychiatry in FTD.

V. CONCLUSION

The proposed neuro-symbolic artificial intelligence (NSAI) framework offers a powerful and hybrid approach to the computational assessment of Formal Thought Disorder (FTD) in schizophrenia and other psychotic disorders. By leveraging the strengths of both transformer-based (BERT) and recurrent-based (LSTM) models in learning the representation of linguistics, unsupervised models like K-means, hierarchical clustering, and LDA in discovering the pattern and sub-type of FTD, and supervised models like SVM and XGBoost in classification and prediction of the severity of FTD, and a symbolic model based on TLC, PANSS, SAPS, and DSM-5/ICD-11 in the assessment of FTD, the model has successfully addressed the primary limitations of the conventional expert-based rating scales in the assessment of FTD, which are subjective, non-scalable, and lack resolution in the temporal dimension.

- *Implications*

Enables early, objective detection of FTD, supports differential diagnosis across psychosis-spectrum conditions, facilitates longitudinal symptom tracking, and informs personalized treatment planning (e.g., targeting specific disorganization patterns in cognitive or pharmacological interventions). Advances mechanistic understanding by linking quantifiable linguistic disturbances to underlying cognitive and neural processes (e.g., semantic network dyssyntax, frontal-temporal alterations), promotes hypothesis-driven modeling, and bridges data-driven ML with established psychopathological theory. Enhances scalability for large-scale phenotyping and real-world monitoring, increases clinician trust through interpretable outputs, and supports regulatory acceptance of AI-assisted tools in mental health by combining predictive power with domain-grounded explainability. Prospective multicenter validation, integration with neuroimaging and multimodal behavioral data, and deployment studies in diverse populations are needed to fully realize its potential for precision psychiatry.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1565

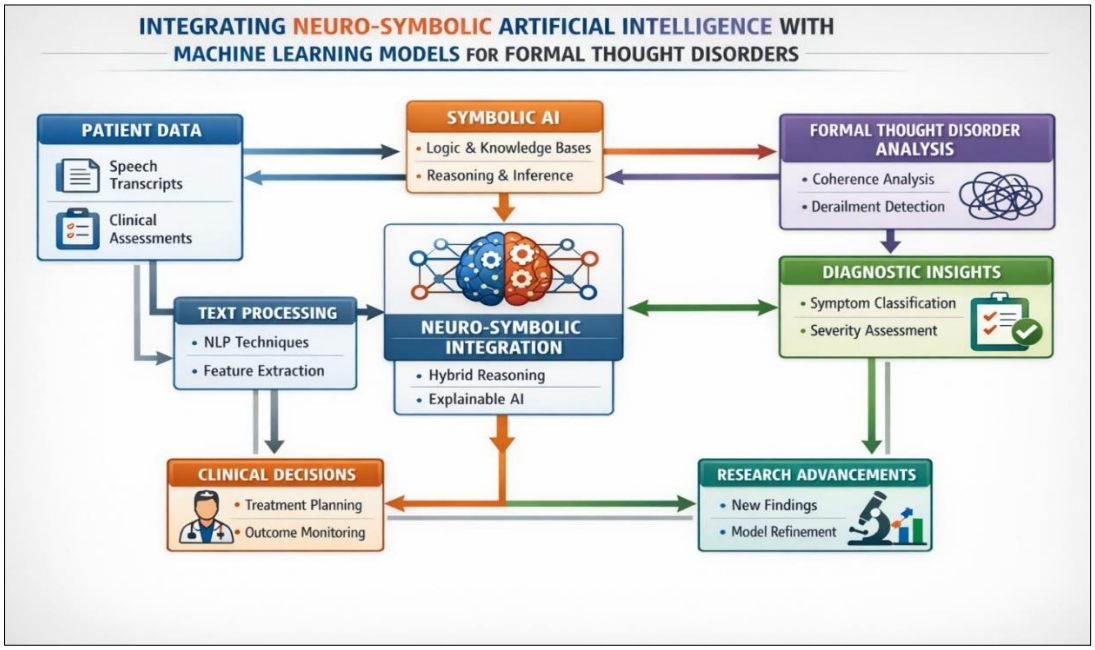

Fig 3 Shows how integrating Neuro-Symbolic AI with Machine Learning Models can be beneficial for Formal Thought Disorders

REFERENCES

[1]. Andreasen, N. C. (1979). Thought, language, and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. Archives of General Psychiatry, 36(12), 1315–1321. https://doi.org/10.1001/archpsyc.1979.01780120045006

[2]. Andreasen, N. C. (1986). Scale for the Assessment of Thought, Language, and Communication (TLC). Schizophrenia Bulletin, 12(3), 473–482.

[3]. Badie, F., & Augusto, L. M. (2022). The form in formal thought disorder: A model of dyssyntax in semantic networking. AI, 3(2), 353–370. https://doi.org/10.3390/ai3020022

[4]. Bedi, G., et al. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. npj Schizophrenia, 1, Article 15030. https://doi.org/10.1038/npjschz.2015.30

[5]. Colombo, F., et al. (2025). Peripheral blood cells unveil neural and sex-related subtypes of depression: An unsupervised machine learning approach.

[6]. Garcez, A. d'Avila, & Lamb, L. C. (2019). Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning.

[7]. Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophrenia Bulletin, 13(2), 261–276. https://doi.org/10.1093/schbul/13.2.261

[8]. Kumar, A. (2023). Neuro symbolic AI in personalized mental health therapy: Bridging cognitive science and computational psychiatry. World Journal of Advanced Research and Reviews, 19(2), 1663–1679. https://doi.org/10.30574/wjarr.2023.19.2.1516

[9]. Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. Trends in Cognitive Sciences, 16(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

[10]. Oeztuerk, O. F., et al. (2021). The clinical relevance of formal thought disorder in the early stages of psychosis: Results from the PRONIA study. World Psychiatry, 20(1), 120–129.

[11]. Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., & Okruszek, L. (2021). Detecting formal thought disorder by deep contextualized word representations. Psychiatry Research, 304, Article 114135. https://doi.org/10.1016/j.psychres.2021.114135

[12]. Zakowicz, P. T., et al. (2025). Detection of formal thought disorders in child and adolescent psychosis using machine learning and neuro-psychometric data.