DOI: 10.66069/ojspub.1137260609

# Multimodal Diagnosis of AI Literacy Among University Teachers and Students under the Background of Big Data

**Ruiying Gao, Fengtian Kou**

Xi'an International University, Xi'an, Shaanxi, China

**Abstract:** ***In the macro-context of artificial intelligence reshaping social production and daily life, AI literacy has become a core competency and key competitive edge for individuals adapting to survival and development in the intelligent era. As educational evaluation shifts from a single outcome-oriented approach to all-process procedural diagnostics, traditional assessment paradigms have revealed certain limitations: most AI literacy assessments heavily rely on static scales, making it difficult to penetrate the deep mechanisms of individual cognitive processing. In view of this, constructing a multimodal diagnostic model through systematic thought experiments and theoretical deduction. The "Neuro-Behavior-Capability" ternary dynamic model aims to bridge three dimensions—neuroscience, behavioral patterns, and ability performance—systematically outlining the deep underlying mechanisms behind the formation of AI literacy. The model break through the surface-level limitations of current assessments and lay a theoretical foundation for the development of intelligent, precision-driven AI literacy assessment systems.***

**Keywords:** AI literacy, Multimodal assessment, Diagnostic theory, Thought experiment, Theoretical model.

## 1. Introduction

As we enter the third decade of the 21st century, artificial intelligence technology is embedding itself into every aspect of higher education at an unprecedented speed—ranging from intelligent retrieval and automated writing assistance to personalized learning recommendations. Interaction between students and AI systems has become a basic form of daily learning (Fitria, 2021). Meanwhile, big data technology has granted educational researchers the ability to collect, store, and analyze massive amounts of learning behavior data: operation logs from learning management systems, eye-tracking trajectories, physiological sensing signals, and even EEG data now constitute a multidimensional, high-temporal-resolution learning information stream (Samal & Hashmi, 2024). This technological background fundamentally alters the possibility boundaries of educational assessment—we are no longer restricted to the static information captured by paper-and-pencil tests, but can instead construct assessment systems that dynamically and continuously monitor the learning process.

However, a noticeable gap remains between this technological possibility and the academic understanding of AI literacy. AI literacy is not a single-dimensional concept, rather, it is a multidimensional capability system encompassing knowledge understanding, critical evaluation, practical application, and ethical judgment (Shiri, 2024). When individuals complete an AI-assisted task, the underlying cognitive mechanisms at work can be entirely different: a "shallow proficient user" might rely on memorized templates to generate surface-level fluent prompts, whereas a "deep critic" will repeatedly verify the underlying assumptions of the model's output and identify its limitations through counterfactual reasoning. While these two modalities may yield similar scores on a traditional static scale, they differ substantially in cognitive investment, strategic depth, and transfer ability—and distinguishing between them is the precise prerequisite for precise instructional intervention.

Existing research on AI literacy is abundant, but assessment tools mostly stall at the questionnaire and scale level, relying primarily on self-report surveys or multiple-choice knowledge questions (Lintner, 2024). Although such tools are highly efficient, they fail to reveal the depth and quality of cognitive engagement when individuals tackle complex AI tasks. In contrast, accumulations in the field of neuroscience provide powerful tools for exploring cognitive processes. Numerous studies have shown that prefrontal theta waves (4–8 Hz) serve as a crucial neurophysiological indicator of high-order cognitive control and executive functions (Basharpoor et al., 2021; Yao et al., 2025). Behavioral science has also revealed that prompt iteration patterns during interactions with Large Language Models (LLMs) reflect deep cognitive strategies and problem-solving styles (Wei et al., 2022). However, data from these neurobiological, behavioral, and capability Dimensions still lack a systematically integrated theoretical framework within AI literacy research.

Against this backdrop, this study poses the question: Is it possible to construct a multimodal diagnostic model that integrates neurophysiological signals, behavioral interaction footprints, and explicit capability performance into a unified framework? Before embarking on high-cost empirical research, it is essential to first answer the logical feasibility of this question at a theoretical level.

## 2. Theoretical Foundations and Core Assumptions

The Extended Mind Theory (EMT) posits that cognitive processes do not occur solely within the biological brain, but are coupled dynamically with external tools, environments, and social systems to form a complete cognitive system (Clark & Chalmers, 1998). Based on this, the core assumption of the "Neuro-Behavior-Capability" ternary dynamic diagnostic model is that an individual's AI literacy level is not a single-dimensional linear quantity, but rather the synergistic output of three interconnected and mutually reinforcing systems: Measurable cognitive effort (neurophysiological dimension), Observable operational strategies (behavioral interaction dimension), and Assessable explicit performance (Capability Assessment dimension)

### 2.1 Neurophysiological Dimension: Proxy Indicators of Deep Cognition

Neuroscientific research provides key cross-Dimension measurement anchors for this model. A substantial body of research consistently demonstrates that when individuals engage in cognitive tasks requiring sustained working memory, executive control, and inhibitory regulation, theta wave (4–8 Hz) power in the prefrontal cortex (PFC) increases significantly (Basharpoor et al., 2021). This phenomenon has been interpreted as the "neural cost of cognitive effort," where the prefrontal cortex mobilizes more resources for active monitoring, conflict detection, and deep information integration. Specifically, this model invokes three sets of neural evidence:

**Strong correlation between prefrontal theta waves and executive function:** A decrease in theta power in the prefrontal poles (Fpz, Fp1, Fp2) is highly correlated with executive function impairment; conversely, an increase in theta power predicts deeper executive information processing.

**Dependence on the "cognitive load zone" of theta waves:** Theta waves hold the highest predictive value for learning outcomes within an "optimal cognitive zone"; excessively low theta activity implies that the task poses no real cognitive challenge to the individual (i.e., "shallow processing"), while excessively high activity may indicate cognitive overload rather than deep thinking.

**Causal role of theta waves in prefrontal emotional regulation and cognitive control:** Actively enhancing prefrontal theta oscillations can significantly improve task performance that requires cognitive control.

Based on this evidence, this model treats prefrontal theta power, particularly frequency band power in the Fz/F3/F4 electrode regions, as a neurophysiological proxy indicator for "deep cognitive effort." Crucially, theta waves are not a direct measurement of AI literacy, but rather a lateral reflection of cognitive depth; they hold diagnostic significance if and only if they are interpreted jointly with behavioral and capability Dimension data.

### 2.2 Behavioral Interaction Dimension: AI Tool Operational Strategies and Behaviors

While neurophysiological data reflect the depth of the cognitive process, behavioral interaction data record its path. During interactions with LLMs, user operations leave rich, quantifiable traces. This model focuses primarily on two types of behavioral metrics:

#### 2.2.1 Prompt Iteration Strategy (PIS)

Multi-round interaction patterns with LLMs profoundly reveal differences in user cognitive strategies. This model extracts three quantifiable PIS sub-metrics:

**Iteration Depth (ID):** The number of prompt revision rounds performed to complete a single task, reflecting whether the individual actively conducts critical examination and correction of the initial output.

**Revision Type Ratio (RTR):** The proportion of conceptual restructuring versus surface-level polishing within revision behaviors, used to distinguish strategic adjustments from habitual repetition.

**Cross-domain Prompt Transfer (CDT):** Whether prompt structures learned in one task are actively transferred and applied to new tasks, reflecting the generalization ability of cognitive schemas.

#### 2.2.2 Counterfactual Questioning Behavior (CQB)

Counterfactual reasoning is a core hallmark of critical thinking and an effective means of identifying limitations in AI outputs (Pereira & Saptawijaya, 2016). This model uses whether an individual actively poses counterfactual questions to the AI as a behavioral indicator of deep critique. The quantification of CQB includes the percentage of counterfactual questions out of total questions asked, as well as the frequency with which counterfactual questions trigger model output corrections. Both behavioral metrics can be automatically extracted via large-scale LLM interaction logs, offering high scalability and making them natural components of a multimodal learning analytics framework.

### 2.3 Capability Assessment Dimension: Structured Assessment of AI Literacy

The definition of AI literacy continuously expands alongside technological evolution. Early work defined it as the basic capability to use AI in a critical, safe, and responsible manner (Chee et al., 2025); with the popularization of generative AI, academia has begun to emphasize critical evaluation, creative application, and human-AI collaboration.

In this model, the capability Dimension refers to an individual's stable, general AI literacy traits—an independent, transferable core literacy that is decoupled from specific tasks. Referencing existing scales, it is divided into four dimensions: Basic Knowledge and Understanding, Critical Evaluation, Practical Application, and Ethical Judgment.

Its measurement is achieved through standardized tools, which can adopt self-assessment or peer-assessment scores from general AI literacy scales, or use the output quality of standardized tasks as a situational observation index in specific research contexts. It must be noted that the assessment of the capability Dimension is not strictly equivalent to specific task scores. The same task score may correspond to different capability Dimension levels; for instance, a "template-matching" individual might secure high task scores by applying a fixed template, yet their general critical evaluation capability remains low. This discrepancy is precisely why the ternary model requires cross-validation from neural and behavioral Dimensions.

### 2.4 Core Logic of Ternary Mutual Verification and the Hermeneutic Circle

The core logic of the ternary dynamic diagnostic model is not a simple aggregation of three independent indicators, but a relationship of mutual constraint and joint interpretation across the three data Dimensions. This can be understood through the concept of the hermeneutic circle, meaning that the interpretation of each Dimension of data relies on the context of the other two Dimensions, while simultaneously providing constraints for them.Consequently, eight typical combination modes and diagnostic types of ternary mutual verification are formed (see Table 1).

When the three Dimensions of data display specific combination modes, highly specific diagnostic inferences can be generated. For example, when the neural Dimension shows high theta wave activation, the behavioral Dimension exhibits high iteration depth and counterfactual questioning, and the capability Dimension yields high-quality output, a strong consistency pattern is formed, which can be inferred as “Deep Collaborative” literacy. Conversely, if theta wave activation is low while the capability score is high, the inconsistency between the two Dimensions helps distinguish between the “Highly Automated” type and the “Template-Matching” type. Figure 1 visually represents the non-linear relationship model among the three elements.

**Table 1:** Typical Combination Modes and Diagnostic Types of Ternary Mutual Verification

| Indicators | | | | |
|---|---|---|---|---|
| **Neural Dimension wave)** | **Behavioral (θDimension + CQB)** | **Capability (PISDimension (Score)** | **Diagnostic Type** | **Core Characteristics** |
| High | High | High | **Deep Collaborative** | Both cognitive and behavioral investments are high; capable of deep collaboration with AI to foster high-quality innovations. |
| High | High | Low | **Exploratory Restructuring** | Cognitive and behavioral investments are immense, but capability performance is not yet distinct; still in a stage of cognitive exploration and restructuring. |
| High | Low | High | **Cognitive Empowered** | Cognition is active and closed-loop; unaccustomed or unwilling to use AI tools to extend and test thinking. |
| Low | High | High | **Template-Matching** | Relies on fixed cognitive pathways; output is stable but lacks deep neural processing. |
| Low | Low | High | **Highly Automated** | Skills are highly internalized; cognitive load is extremely low, placing the user in an automated processing state. |
| Low | High | Low | **Interactive Inefficient** | Neural arousal is insufficient but behavioral investment is high; effort struggles to translate into effective output, leading to low efficiency. |
| High | Low | Low | **Cognitive Inefficient** | Neural arousal is high but behavioral organization is chaotic; improper resource allocation leads to low efficiency. |
| Low | Low | Low | **Fully Disengaged** | Lacks cognitive drive and behavioral investment; in a state of learning disengagement. |

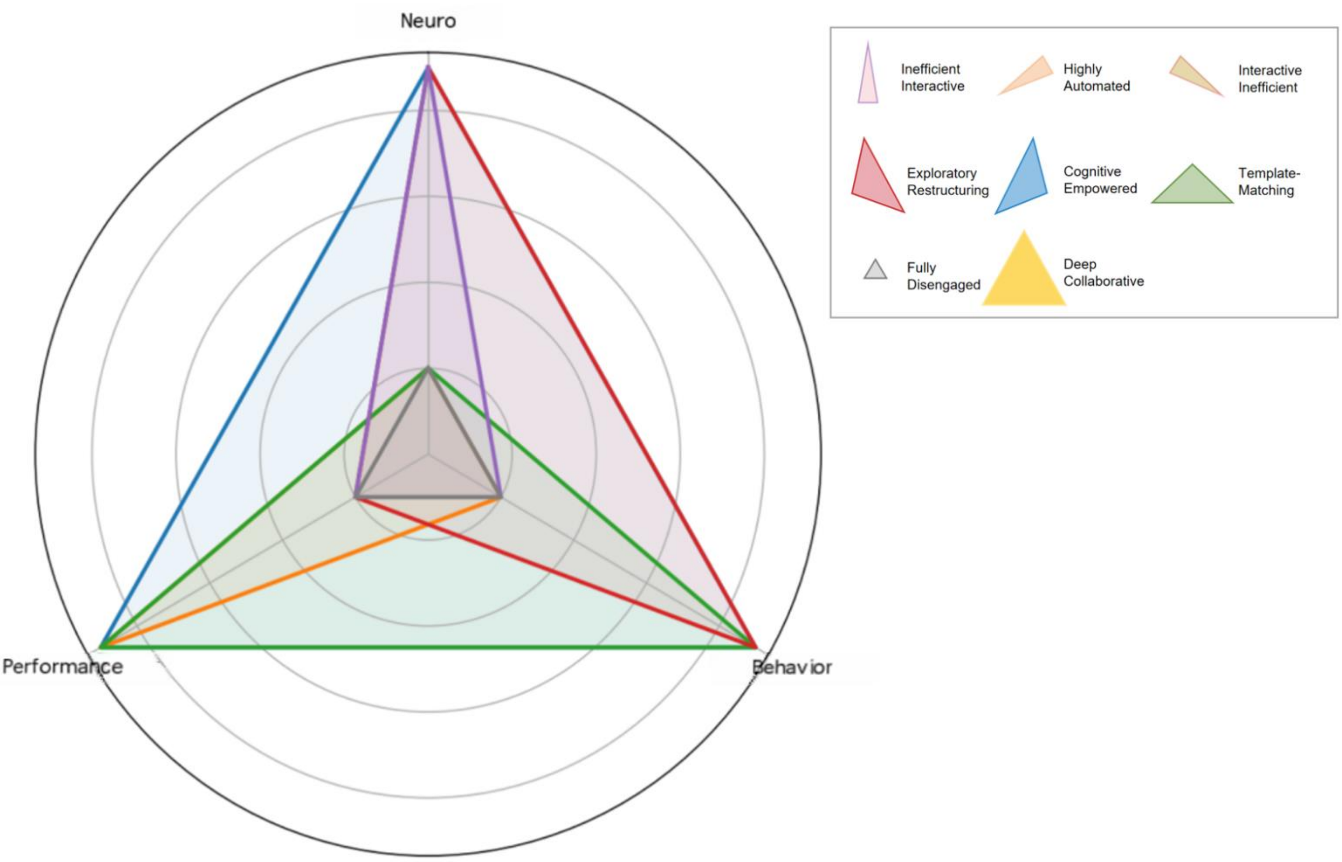


**Figure 1:** Diagnostic Radar Chart for Learner’ Status by Ternary Mutual Verification

## 3. Thought Experiment Model Deduction

To validate logical soundness prior to empirical studies, this research follows a unified paradigm of "Scenario Setting — Subject Pre-configuration — Data Deduction — Validity Testing" to design three sets of structured thought experiments, verifying the diagnostic boundaries of the ternary model Dimension by Dimension.

### 3.1 Thought Experiment I: Deep Diagnosis of Critical Understanding

3.1.1 Task Scenario Setting

University students are selected as research subjects. They are required to use a large language model to analyze a position paper titled "The Application of AI-Assisted Decision Making in Assignment Evaluation" and write a critical analysis report that identifies argumentative loopholes and potential biases within the paper. This task is designed to require high-order operations of the critical understanding dimension, corresponding to the critical evaluation dimension of AI literacy.

3.1.2 Participant Pre-configuration

Participant A: Possesses a strong habit of counterfactual reasoning; tends to view AI-assisted content as hypotheses to be verified rather than authoritative conclusions; actively initiates challenging questions during interactions; remains vigilant toward the model output's underlying assumptions.

Participant B: Mastered the discourse of critical thinking and can utilize terms like "bias" and "logical fallacy"; however, critical behaviors remain at the phrasing level, lacking deep verification of the core argumentative structure; tends to directly adopt the critical analysis templates provided by the AI.

3.1.3 Theoretical Deduction of Ternary Data

Neurophysiological Dimension Deduction

Participant A: Continuously performs text comparison, contradiction detection, and conflict processing, invoking working memory updating and executive control functions. It is anticipated that prefrontal theta waves (Fz/F3/F4 electrodes) will strengthen significantly and exhibit transient peaks when argumentative contradictions are detected.

Participant B: Only identifies surface-level critical vocabulary, resulting in lower cognitive load. It is anticipated that theta wave activation will remain at a low-to-medium level, with a stable waveform and no task-related transient activation peaks.

**Behavioral Interaction Dimension Deduction**

Participant A: 1) High iteration depth, such as multi-round prompt revisions to question the core assumptions of the paper; 2) Revision types are dominated by conceptual restructuring, such as asking, "If the personalized recommendation algorithm of this system uses grade maximization as a single optimization goal, what is the underlying educational philosophy?"; 3) High proportion of counterfactual questioning (CQB), such as, "If the subjects of this study were university students instead of primary school students, would the conclusion still hold?".

Participant B: 1) Low iteration depth, mostly accepting the AI's first-round output with only surface-level polishing; 2) Revisions are dominated by superficial edits; 3) The proportion of counterfactual questioning (CQB) is extremely low or zero.

**Capability Dimension Deduction**

This experiment utilizes the quality of the critical analysis report as a situational observation metric for the "Critical Evaluation" dimension (the core dimension of the capability Dimension).

Participant A: Scores high in identifying deep argumentative structure defects and proposing counterfactual refutations, though they might not stand out in format regularity or the density of critical terminology usage. This maps their general critical evaluation capability to a high level.

Participant B: The report may score high in terminology density and structural completeness, but low in the depth of critique. This maps their general critical evaluation capability to a low level.

**Testing of Model Diagnostic Validity**

Under traditional single-capability scoring, Participant A and Participant B might receive similar total scores due to mutual cancellation (e.g., high depth score + formatting deductions vs. formatting scores + depth deductions), failing to differentiate the essence of their literacy. The ternary model achieves precise diagnosis through combination pattern matching:

**Table 2:** Comparison of Participants' Ternary Data and Theoretical Patterns in Thought Experiment I

| Participant | Neural Dimension Features | Behavioral Dimension Features | Capability Dimension Features | Diagnostic Type |
|---|---|---|---|---|
| A | High θ activation + transient peaks | High iteration depth + high CBQ | High depth dimension | Deep Collaborative |
| B | Low θ activation + flat waveform | Low iteration depth + low CBQ | High regularity + low depth dimension | Template-Matching |

### 3.2 Thought Experiment II: Typological Diagnosis of Human-AI Collaboration Efficiency

3.2.1 Task Scenario Setting

University teachers are selected as research subjects. They are required to utilize AI tools to complete a research proposal with real-world constraints within 30 minutes, including a literature review section and a methodology section. The task design simultaneously activates basic knowledge, critical evaluation, and practical application dimensions within the capability Dimension, involving real-time human-AI task allocation decisions.

3.2.2 Participant Pre-configuration

**Participant C Characteristics:** Possesses a clear sense of cognitive division of labor; plans human-AI task boundaries in advance; selectively adopts AI outputs; maintains deep human reflection at critical nodes.

**Participant D Characteristics:** Tends to outsource all segments to the AI; lacks task allocation strategies; repeatedly re-prompts when AI outputs fail to meet standards, falling into a "question-acceptance" loop; cognitive investment and output efficiency are severely mismatched.

3.2.3 Theoretical Deduction of Ternary Data

**Neurophysiological Dimension Deduction**

Participant C: Displays a rhythmic fluctuating pattern. Theta waves are highly activated during human-dominated phases such as critical integration and architectural decision-making, and recede during AI-assisted phases like text polishing and formatting, reflecting dynamic switching between active cognitive investment and cognitive outsourcing.

Participant D: Theta waves maintain a low-to-medium level throughout without rhythmic changes, reflecting continuous evasion of deep thinking, replacing active cognitive processing with a passive "question-acceptance" mode.

**Behavioral Interaction Dimension Deduction**

Participant C: 1) Moderate total number of prompt rounds; 2) Prompt structures are differentiated according to task stages (refined conceptual prompts during strategic stages, concrete task-oriented prompts during execution stages); 3) High cross-domain prompt transfer (CDT), capable of transferring prompt structures learned during the literature review stage to the methodology section.

Participant D: 1) High total number of prompt rounds, but mostly repetitive questions; 2) Low proportion of conceptual restructuring; 3) Low cross-domain prompt transfer, with behaviors characterized by a "low-quality iterative loop."

**Capability Dimension Deduction**

This experiment uses the quality of the research proposal as a situational observation metric for the "Practical Application" and "Critical Evaluation" dimensions.

Participant C: The produced research proposal is expected to have a clear logical structure, noticeable traces of human judgment, and high overall quality.

Participant D: The output volume is large, but the logical coherence and depth of original thinking are low, exhibiting a distinct "copy-paste" or fragmented feel.

**Testing of Model Diagnostic Validity**

Thought Experiment II further highlights the core diagnostic advantage of the ternary model. If relying solely on coarse-grained overall capability scores, Participant D might secure a relatively high score due to volume advantages, and the score gap between C and D could be limited, failing to distinguish quality variations in human-AI collaboration. Through three-Dimension combination pattern matching, the model achieves refined typological diagnosis.

**Table 3:** Comparison of Participants' Ternary Data and Theoretical Patterns in Thought Experiment II

| Participant | Neural Dimension Features | Behavioral Dimension Features | Capability Dimension Features | Diagnostic Type |
|---|---|---|---|---|
| C | θ wave rhythmic fluctuations | Moderate rounds + differentiated + high CDT | High overall quality | Deep Collaborative |
| D | Low θ activation +no rhythm | High rounds + low-quality loop + low CDT | Low logical coherence | Inefficient Interactive |

### 3.3 Thought Experiment III: Explanatory Power Testing of Contradictory Data

3.3.1 Theoretical Significance of Contradictory Scenarios

The first two experiments validated the model's differentiation capability in data-consistent scenarios. However, real-world multimodal assessments frequently encounter "high-input, low-output" contradictory combinations. Traditional assessments easily misinterpret this as a lack of capability, whereas the ternary model can parse its internal mechanisms through cross-Dimension mutual verification. This experiment focuses on the "Exploratory Restructuring" sub-type to test the model's explanatory power over "high effort-low performance" contradictions.

3.3.2 Description of Contradictory Data

Participant E: At the neural Dimension, prefrontal theta waves are continuously highly activated, equivalent to the "deep critic" Participant A. At the behavioral Dimension, prompt iteration depth and the proportion of counterfactual questions are both high, also matching Participant A. At the capability Dimension, the score for the critical understanding dimension is low, but the score for the basic knowledge dimension is high.

3.3.3 Diagnostic Inferences of the Ternary Model

In this context, the ternary model provides the following explanatory inference chain:

**Cognitive Stage Inference:** High neural activation combined with high behavioral investment indicates that Participant E is currently in a phase of cognitive exploration and strategy restructuring. They have not yet formed a stable, highly efficient critical pattern, rather than being unwilling to put in effort or lacking basic ability.

**Bottleneck Localization Inference:** The capability Dimension's "high basic knowledge + low critical score" rules out gaps in fundamental knowledge. High data in neural and behavioral Dimensions rules out a lack of motivation or strategy. The bottleneck points directly to "insufficient strategic effectiveness," such as deviations in counterfactual questioning angles or incorrect iterative directions.

**Intervention Direction Inference:** Targeted strategy feedback must be provided—such as optimizing the counterfactual questioning framework and adjusting iteration paths—rather than repeating basic knowledge lectures or merely encouraging hard work.

3.3.4 Speculative Conclusion: Theoretical Advantages and Diagnostic Potential of the Model

Synthesizing the three sets of thought experiments, the core theoretical advantages of the "Neuro-Behavior-Capability" ternary dynamic diagnostic model compared to traditional single-Dimension assessments can be summarized as follows:

**Depth:** It penetrates surface scores of the capability Dimension through neural Dimension data, reaching the physiological imprints of cognitive processes, rendering process diagnostics rather than outcome evaluation possible.

**Heterogeneity:** It identifies entirely different mechanistic sources behind identical capability scores through the combination patterns of three data Dimensions, breaking the dilemma where a single score shadows individual differences.

**Interpretability:** When data inconsistencies arise, the model does not discard them as measurement errors. Instead, it treats the inconsistency itself as a diagnostic signal, generating instructionally meaningful intervention guidance.

## 4. Future Research Outlook

Based on the theoretical deduction of this study, future empirical work should revolve around two main tracks: system development and hypothesis testing, progressively driving the model from theoretical conception to practical application.

### 4.1 Prototype Development of Multimodal AI Literacy Diagnostic System (MALDS)

Future endeavors can rely on existing technical architectures to develop the "Multimodal AI Literacy Diagnostic System" in phases:

**Data Collection End:** Integrate lightweight EEG headbands, LLM API log interfaces, and task management platforms to achieve millisecond-level synchronization of neural and behavioral data.

**Algorithm Fusion End:** Explore multimodal fusion algorithms based on attention mechanisms, dynamically weighting the feature contributions of different levels.

**Application Output End:** Develop a visual diagnostic dashboard to present individual literacy profiles and sub-type attributions in real time, automatically pushing personalized learning resources.

### 4.2 Empirical Validation Path for Theoretical Propositions

The three sets of core propositions extracted in this study offer an explicit framework for subsequent quantitative verification.

**Table 4:** Empirical Validation Path for Core Theoretical Propositions

| Theoretical Proposition | Core Hypothesis | Key Variable Operationalization | Recommended Statistical Method |
|---|---|---|---|
| Proposition I: Neuro-Behavioral Covariation | Neural activation of high-order holders is strongly positively correlated with behavioral depth. | IV: Prefrontal theta power (EEG)<br>DV: Prompt conceptual restructuring ratio<br>CV: Task difficulty, basic knowledge | Partial correlation analysis; Hierarchical Linear Modeling (HLM) |
| Proposition II: Profile Discriminant Validity | Ternary data can cluster into 8 theoretical sub-types. | Indicators: Neural + Behavioral + Capability<br>Criteria: Complex project performance, peer evaluation | K-means / Spectral Clustering; Discriminant analysis; ANOVA group difference testing |
| Proposition III: Incremental Predictive Validity | The ternary model predicts complex task performance better than single-Dimension models. | Predictors: Neural/Behavioral/Capability (Single vs. Fusion)<br>Outcome: Comprehensive score of open AI projects | GBDT / MLP modeling; F-test (nested model comparison); Incremental Validity analysis |

## 5. Conclusion

Using thought experiments as its core methodology, this study systematically constructed and argued a "Neuro - Behavior - Capability" multimodal AI literacy diagnostic model at a theoretical level prior to empirical studies. Through three structured thought experiments, the following four dimensions of theoretical work were completed:

Firstly, we defined the four-dimensional capability structure of AI literacy, providing structural anchors for capability Dimension measurements and positioning them as the explanatory objective for capability and behavioral data.

Secondly, we established prefrontal theta wave activation as a proxy indicator of deep cognitive effort, clarifying its role in the "ternary mutual verification" framework as the neural anchor that exerts arbitrating efficacy alongside behavioral and capability Dimension data.

Tirdly, we proposed Prompt Iteration Strategy (PIS) and Counterfactual Questioning Behavior (CQB) as core measurement dimensions of the behavioral Dimension, systematically integrating LLM interaction logs—a highly scalable data source in the big data era—into the literacy diagnostic framework.

Finally, we demonstrated the model's theoretical differentiation capability across various literacy types through three thought experiments, distilling these deductions into three testable theoretical propositions, providing concrete directions for future researchers in neuroeducation, learning analytics, and AI literacy assessment to test, amend, and expand with empirical data.

## References

[1] Fitria, T. N. (2021). Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. *Prosiding seminar nasional & call for paper STIE AAS*. 4(1): 134-147.
[2] Samal, P., & Hashmi, M. F. (2024). Role of machine learning and deep learning techniques in EEG-based BCI emotion recognition system: a review. *Artificial Intelligence Review, 57*(3), 50.
[3] Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning, 9*(1), 50.
[4] Yao, Y., Cui, Y., Wang, K., Chen, Y., & Xu, R. (2025). Electroencephalography in China: Spread, Development and Prospects. *Clinical EEG and Neuroscience*, 15500594251343170.
[5] Basharpoor, S., Heidari, F., & Molavi, P. (2021). EEG coherence in theta, alpha, and beta bands in frontal regions and executive functions. *Applied Neuropsychology: Adult, 28*(3), 310-317.
[6] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems, 35*, 24824-24837.
[7] Clark, A., & Chalmers, D. (1998). The extended mind. *analysis, 58*(1), 7-19.
[8] Pereira, L. M., & Saptawijaya, A. (2016). Counterfactuals in critical thinking with application to morality. In *Model-Based Reasoning in Science and Technology: Logical, Epistemological, and Cognitive Issues* (pp. 279-289). Cham: Springer International Publishing.
[9] Chee, H., Ahn, S., & Lee, J. (2025). A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway. *British Journal of Educational Technology, 56*(5), 2146-2182.
[10] Shiri, A. (2024). Artificial intelligence literacy: a proposed faceted taxonomy. *Digital Library Perspectives, 40*(4), 681-699.

## Funding

This work was supported by the Belt and Road International Land Port Logistics Joint Research Center of Shaanxi Province (Grant No. GH202509), and the 14th Five-Year Plan for Education Science of Shaanxi Province (Project No. SGH25Q639).