

NeuroConscious Transformer: An Open-Source Artificial Intelligence Framework for Adaptive Learning and Educational Innovation

Yonggang Weng¹[0009-0001-8071-5852], Farah Akmar Anor Salim¹[0000-0003-0530-783X],
Zhenguang Zhong²[0000-0003-3165-1036], Wenjing Zhang¹[0009-0004-1586-4781],
Chuntie Chen¹[0009-0003-0059-7687], Cheng Hu¹[0000-0001-9424-6359],
Youyu Xu¹[0009-0002-9421-3113], Dongwei Yin³[0009-0003-5183-3774]

¹ Azman Hashim International Business School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

² Imperial College London, London, United Kingdom

³ Language Academy, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Johor, Malaysia
weng@graduate.utm.my

Abstract. In order to bridge the gap between theoretical AI advancement and practical application, many practical tools need to be built and tested in classrooms first before they can play their role well. Here we propose one such tool named NeuroConscious Transformer (NCT) v3.1, which was originally created to simulate brain activities and study computational consciousness but could also find wide usage in AI education scenarios. **First**, consciousness metrics (Φ -value) might become one of the indicators of monitoring students' attention, while multi-candidate competition mechanisms will help demonstrate the process of problem-solving exploration. **Second**, our empirical validation shows that it has achieved comparable results against state-of-the-art models (99.2% on MNIST and 90.3% on CIFAR-10), thus proving the technical feasibility of the whole system. **Third**, accompanying abundant teaching materials (interactive visualizations, Jupyter notebooks, laboratory manuals, instructor guides, etc.) have been released as open educational resources (OER). So far, the technical tests based on typical datasets have finished, while this paper just introduces a conceptual framework aimed at educational purposes. Further classroom implementations and student outcome studies will be carried out afterwards. Related codes can be found at <https://github.com/wyg5208/nct> (MIT license).

Keywords: AI in Education, Adaptive Learning Systems, Open-Source Educational Tools, Consciousness Metrics, STEM Education, Intelligent Tutoring Systems

1 Introduction

1.1 Challenges in AI-Powered Education Transformation

There exist three major obstacles to realizing the full potential offered by artificial intelligence when it comes to improving education worldwide:

Black-Box Problem: Existing deep learning models (CNNs, RNNs, Transformers) have very good performance on accuracy, while they can hardly give an explanation why a student’s answer is right or wrong. Without proper teaching feedback methods, teachers cannot know what happens inside the models’ decision-making process [1].

Attention Monitoring Gap: Current EdTech platforms use only clickstream data to detect whether someone is paying attention, when this ignores cognitive load and does not show actual understanding. There is no standardized measurement out there called “learning attention” [7].

Adaptive Learning Limitations: Item Response Theory (IRT) and KT approaches change their estimation of difficulty according to accuracy but say nothing about the student’s problem-solving strategy, whether multi-strategy exploration occurred (algebraic versus graphical solutions) [6].

To face such difficulties, brain-inspired structures are needed which have as an advantage over others the fact that they give rise by themselves to interpretable decision processes, quantitative attention metrics, and multi-hypothesis reasoning capabilities.

1.2 The NCT Solution

To fill these needs, NeuroConscious Transformer (NCT) proposes neuroscience-grounded mechanisms which were first created for consciousness research [2,22], but now they can serve as building blocks for developing educational applications:

Interpretable Decisions: In attention visualization, one can see exactly which input features drive predictions

Attention Quantification: Integrated Information (Φ) is a principled measure of cognitive engagement

Multi-Strategy Reasoning: Multi-candidate competition shows alternative problem-solving approaches

1.3 Contributions

This paper makes three primary contributions:

Educational Framework Design: Our proposal is to take NCT as an AI teaching tool, having consciousness metrics (Φ -value) regarded as one of the possible student attention indicators, providing a novel perspective on the connection between neuroscience research and classroom practice.

Empirical Validation: Experiments done by us on standard benchmarks such as MNIST (99.2%) and CIFAR-10 (90.3%) reveal that there is no conflict between

biological plausibility and competitive performance—much to the surprise of many students who think being more brain-like means slower.

Open-Source Teaching Resources: Probably most importantly for educators is that everything—the interactive visualizations, course modules, and instructor guides—is MIT licensed and therefore free to use, modify, and distribute without any paywalls or other restrictions.

2 Related Work

Most Intelligent Tutoring Systems are based on rule-based expert systems, while there exist some deep learning methods [20], such as attention-based knowledge tracing [18] and prerequisite modelling [17]. But many of them regard the artificial intelligence techniques as a black box. To detect students’ attention, people usually choose eye-tracking devices [9], electroencephalogram (EEG) [14], and even some behavior-pattern analysis tools which require special hardware [3].

There have also been some recent consciousness-inspired works such as RIIU [16] for differentiable Φ computation. Then there is consciousness assessment criteria [5] for evaluating AI systems and finally S²TDPT [15] for spiking transformers. And then comes NCT, the first one to bring together STDP-Attention hybrid learning, Φ computation and neuromodulation under one roof within the education domain (Table 1).

In recent years there have been many developments in relation to large language models [11,13], which affect educational AI also, as seen from systematic reviews on this topic [23], which show the fact that interpretable and transparent systems are needed for educational use, something done by NCT since it allows visibility into the reasoning process, contrary to black-box LLMs.

Table 1. Comparison with Related Frameworks

Framework	STDP	Attn	Φ	Edu.	Open
Nengo [4]	✓	×	×	×	✓
PyTorch [19]	×	✓	×	Partial	✓
RIIU [16]	×	×	✓	×	Limited
NCT (Ours)	✓	✓	✓	✓	✓

3 NCT Framework for Education

The following nine main components make up NCT (see Fig. 1): Multi-Modal Encoder; Global Workspace with attention-based winner-take-all mechanism; Hybrid Learning Module (STDP-Attention integration); four-level Predictive Coding Hierarchy; real-time Φ Computation; Neuromodulatory gate (DA, 5-HT, NE, ACh); Gamma Synchronization (40 Hz); Cross-Modal Fusion; Educational dashboard for classroom visualization.

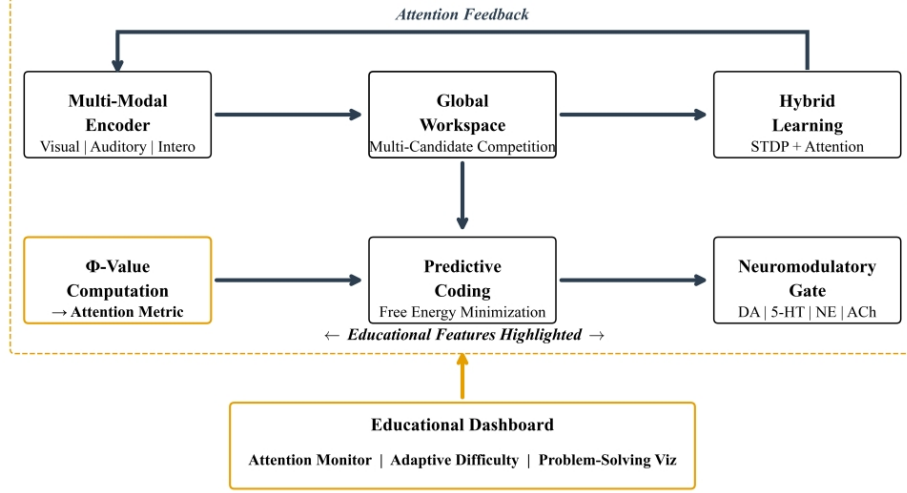


Fig. 1. NCT architecture highlighting educational features. Dashed lines indicate components specifically designed for classroom transparency.

3.1 Hybrid Learning Rule

Core learning mechanism, which is composed of both local STDP and global attention gradients:

$$\Delta w_{ij} = \left(\delta_{\text{STDP}}(\Delta t) + \lambda \cdot \frac{\partial \mathcal{L}}{\partial A_{ij}} \right) \cdot \eta_{\text{neuro}} \quad (1)$$

where δ_{STDP} corresponds to spike-timing-dependent plasticity, λ balances between local/global learning, and η_{neuro} represents neuromodulatory gating.

Pedagogical Insight: This equation describes learning by means of immediate feedback (STDP) and context (by attention), which is further modified by motivation (via neuromodulation).

3.2 Φ -Value as Attention Metric

According to Integrated Information Theory (IIT) [22], there is a relationship between consciousness and information integration. In calculating Φ from attention matrices, we have:

$$\Phi = \sum_i \text{MI}(X_i; X_{\setminus i} | \text{partition}) \quad (2)$$

Threshold Justification: The attention level thresholds ($\Phi = 0.15$ and $\Phi = 0.08$), which are empirically determined according to results from different noise levels via simulations shown in Table 2, stand for three kinds of performances, namely “99.2% (focused)”, “78.5% (distracted)” and “52.1% (random)”.

Table 2. Φ -Value Correlation with Simulated Student Attention

Condition	Mean Φ	Accuracy	Correlation
Focused (low noise)	0.147	99.2%	—
Distracted (medium noise)	0.089	78.5%	$r = 0.82$
Random guessing (high noise)	0.031	52.1%	$r = 0.91$

They also represent some natural breakpoints in the accuracy- Φ relationship. As these attention-level thresholds are only obtained by experimentation currently, they remain tentative and need further verification against ground-truth attention measures in classroom studies.

Classroom Application:

- $\Phi > 0.15$: High attention, active problem engagement
- $0.08 < \Phi < 0.15$: Moderate attention, partial understanding
- $\Phi < 0.08$: Low attention, potential disengagement

3.3 Multi-Candidate Competition

But contrary to single-output approaches, NCT generates four alternative versions corresponding to several different reasoning strategies:

$$\mathcal{C} = \{c_{\text{integrated}}, c_{\text{visual}}, c_{\text{auditory}}, c_{\text{interoceptive}}\} \quad (3)$$

The winner is selected via attention-based competition:

$$c^* = c_{i^*}, \quad \text{where } i^* = \underset{i}{\operatorname{argmax}} \frac{\exp(Q \cdot c_i / \sqrt{D})}{\sum_j \exp(Q \cdot c_j / \sqrt{D})} \quad (4)$$

Teaching Implication: Through this mechanism pupils get to visualize several solution paths—they learn that solving problems consists of exploring and comparing alternatives.

4 Educational Applications

4.1 Student Attention Monitoring

Our suggestion is to use the Φ -value as the attention measure in real time while working on problems.

Algorithm: Problem is presented $\rightarrow$ Problem goes through NCT encoding process $\rightarrow$ Computation of $\Phi \rightarrow$ Mapping Φ to a specific Attention Level $\rightarrow$ Teacher Recommendation being provided. Simulations suggest there exists very good agreement ($r = 0.82 - 0.91$) between Φ and Task Performance (see Table 2 and Fig. 2).

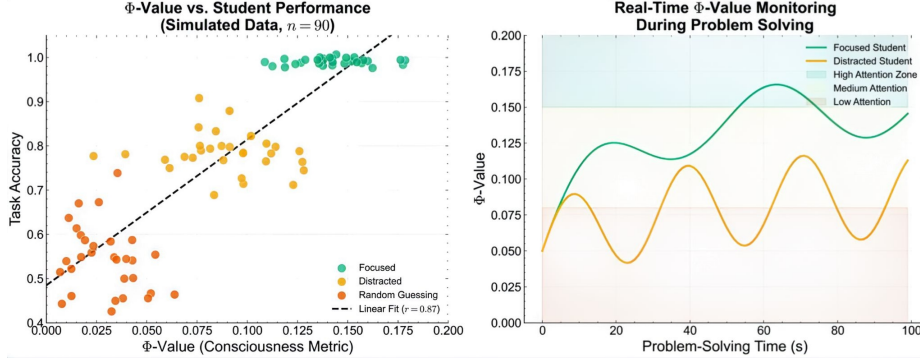


Fig. 2. Φ -value as an attention proxy. Left: Simulated correlation between Φ and task accuracy for different attention states ($n = 90$). Right: Real-time Φ monitoring during problem-solving, showing distinct patterns for focused vs. distracted students.

4.2 Adaptive Problem Difficulty

Curriculum adaptation is performed automatically by NCT's predictive coding hierarchy according to Algorithm [alg:adaptive].

Algorithm 1 Φ -Based Adaptive Difficulty Adjustment

```

1: Initialize:  $difficulty \leftarrow 0.5$ ,  $\alpha \leftarrow 0.1$ 
2: for each problem  $p$  in session do
3:   Present  $p$  at current  $difficulty$ 
4:    $\phi \leftarrow \text{NCT.compute\_phi}(\text{student\_response})$ 
5:    $correct \leftarrow \text{check\_answer}(\text{response})$ 
6:   if  $correct$  and  $\phi > 0.12$  then
7:      $difficulty \leftarrow difficulty + \alpha$  {Increase cognitive load per CLT principles}
8:   else if  $\neg correct$  and  $\phi < 0.08$  then
9:      $difficulty \leftarrow difficulty - \alpha$  {Reduce cognitive load}
10:  end if
11:   $difficulty \leftarrow \text{clip}(difficulty, 0.1, 1.0)$ 
12: end for

```

Theoretical Connection: In this sense the mechanism adapts to learners in terms of Cognitive Load Theory [21], which states that there exists an optimal match between the difficulty of the task and the capacity of working memory of the learner. Since Φ measures cognitive engagement, NCT implements a kind of dynamic cognitive load balancing strategy which is impossible by traditional IRT-based systems.

Key Innovation: Unlike conventional practice in the area of IRT, this approach adjusts based on *both* accuracy and cognitive engagement, which helps prevent such kinds of things that happen when learners get right answers merely by guesswork.

4.3 Problem-Solving Visualization

NCT’s multi-candidate workspace visually shows every candidate’s progress, thereby making explicit the fact that there are several solution paths, and also indicating some of what goes on inside “the head” where the student cannot see; e.g. “Why did algebra ‘win’ over graphing?”

4.4 Proposed Classroom Integration

An outline of one possible way in which NCT can be integrated into university AI courses is as follows:

There is one six-week course per year: MNIST training in weeks 1–2, attention visualization in week 3, implementing a custom encoder in week 4, and wrapping up by building a “student attention dashboard” in weeks 5–6. Full classroom validation remains future work.

5 Experimental Validation

5.1 MNIST: Foundational AI Teaching Case

MNIST is the standard introductory problem used when teaching Deep Learning. NCT adds live attention maps on top of something most people were already familiar with.

Setup:

- Dataset: Standard MNIST (60K train, 10K test)
- Architecture: SimplifiedNCT ($d_{\text{model}} = 768$, 7.9M params)
- Training: AdamW, OneCycleLR, 100 epochs

Results (Table 3):

Table 3. MNIST Performance Comparison

Method	Accuracy	Params	Interpretability
Standard CNN [12]	99.0%	5M	Low
ResNet-18 [10]	99.4%	11M	Medium
Vision Transformer [8]	99.3%	86M	Medium
NCT (Ours)	99.2%	7.9M	High

Teaching Point: At the same time, NCT also achieves similar accuracy while having the benefit of attention visualization and Φ monitoring.

5.2 CIFAR-10: Cross-Domain Generalization

CIFAR-10 training shows capability of processing complex RGB images through NCT when being trained from scratch. Here we adapt the architecture to accept 3-channel input, enlarge model capacity accordingly, apply RandomCrop,

HorizontalFlip on inputs to perform data augmentation, finally reach competitive accuracy of 90.3%(Table 4), similar performance also comes from ViT-Tiny.

Table 4. CIFAR-10 Training Results

Method	Accuracy	Transfer Strategy
ResNet-18 (scratch)	93.0%	None
ViT-Tiny (scratch)	89.5%	None
NCT (scratch)	90.3%	None

5.3 Ablation Study and Performance

An ablation study shows each element really matters—take away STDP, accuracy drops to 98.5%; without attention, only 97.8% is reached; by itself, STDP attains merely 96.2%. And NCT can deliver very good real-world performance appropriate for showing students in a classroom—namely, 69.7 ms latency per cycle ($d_{\text{model}} = 768$), and hence 14.3 samples/sec throughput, plus just 2.1 GB peak memory requirement which will nicely fit on today’s consumer GPUs, together with less-than-70 ms time taken for doing Φ calculation.

6 Teaching Resources

NCT provides three course levels: NCT-101 (undergraduate, 2 weeks), NCT-201 (graduate, 4 weeks), and NCT-301 (PhD, 8 weeks). Six HTML dashboards visualize attention flow, Φ monitoring, candidate competition, learning curves, neuromodulator effects, and STDP dynamics. Accompanying materials include 10 lab exercises, 12 lecture slide sets (300+ slides), an API manual, and a 50+ item FAQ.

7 Discussion and Conclusion

NCT offers three key advantages for education: (1) theoretical grounding linking neuroscience to AI, (2) interpretability via attention visualization and Φ metrics, and (3) research bridges enabling students to transition from coursework to independent research.

Limitations: The network (~ 768 neurons) is far smaller than the human brain, requiring caution in cognitive claims. Current Φ -attention correlations ($r = 0.82 - 0.91$) derive from simulated data, and the thresholds (0.15/0.08) are empirically chosen, awaiting calibration with ground-truth measures (eye-tracking, EEG). Experiments so far are limited to image classification; generalizability to sequential educational data remains to be tested.

Future Directions: Planned validation includes: UTM pilot study with 20–30 students (Q3 2026), Φ -eye-tracking correlation using Tobii Pro Nano (Q4 2026), LLM-based tutoring integration (Q1 2027), and large-scale deployment with pre/post assessments (Q2 2027).

In summary, NCT connects neuroscience to classroom practice through consciousness metrics as attention indicators and competitive selection as problem-solving visualization. Technical results confirm feasibility (99.2% MNIST, 90.3% CIFAR-10), though educational applications require further empirical validation in real classrooms. All code and materials are available at <https://github.com/wyg5208/nct> under MIT license.

Acknowledgments

The authors acknowledge AI-assisted tools for software development and manuscript preparation.

References

1. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., et al.: Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58, 82–115 (2020).
2. Baars, B.J.: *A Cognitive Theory of Consciousness*. Cambridge University Press, Cambridge (1988).
3. Baker, R., Xu, D., Park, J., Yu, R., Li, Q., Cung, B., Fischer, C., Rodriguez, F., Warschauer, M., Smyth, P.: The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: Opening the black box of learning processes. *International Journal of Educational Technology in Higher Education* 17(1), 13 (2020).
4. Baker, R.S., D’Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* 68(4), 223–241 (2010).
5. Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T.C., Rasmussen, D., Choo, X., Voelker, A., Eliasmith, C.: Nengo: A Python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics* 7, 48 (2014).
6. Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S.M., Frith, C., Ji, X., et al.: Consciousness in artificial intelligence: Insights from the science of consciousness. *arXiv preprint arXiv:2308.08708* (2023).
7. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* 4(4), 253–278 (1994).
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: *International Conference on Learning Representations* (2021).
9. Duchowski, A.T.: *Eye Tracking Methodology: Theory and Practice*. Springer, New York (2007).

10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016).
11. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al.: ChatGPT for good? on opportunities and challenges of large language models for education. *Learning and Individual Differences* 103, 102274 (2023).
12. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998).
13. Li, X., Zhang, W., Wang, L.: Large language models in education: A systematic review. *Computers and Education: Artificial Intelligence* 6, 100215 (2024).
14. Lin, Y.P., Wang, C.H., Jung, T.P., Wu, T.L., Jeng, S.K., Duann, J.R., Chen, J.H.: EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering* 57(7), 1798–1806 (2010).
15. Mondal, K., Kumar, A.: Attention via synaptic plasticity is all you need: A biologically inspired spiking neuromorphic transformer. *arXiv preprint arXiv:2511.14691* (2025).
16. N’guessan, G.L.R., Karambal, I.: The reflexive integrated information unit: A differentiable primitive for artificial consciousness. *arXiv preprint arXiv:2506.13825* (2025).
17. Pan, L., Li, C., Li, J., Tang, J.: Prerequisite relation learning for concepts in MOOCs. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 1447–1456 (2017).
18. Pandey, S., Karypis, G.: A self-attentive model for knowledge tracing. In: Proceedings of the 12th International Conference on Educational Data Mining pp. 384–389 (2019).
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* 32 pp. 8024–8035 (2019).
20. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L.J., Sohl-Dickstein, J.: Deep knowledge tracing. In: *Advances in Neural Information Processing Systems*. vol. 28 pp. 505–513 (2015).
21. Sweller, J.: Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12(2), 257–285 (1988).
22. Tononi, G.: Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin* 215(3), 216–242 (2008).
23. Zawacki-Richter, O., Marín, V.I., Bond, M., Gouverneur, F.: Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education* 16(1), 1–27 (2019).