ISSN 2616-5902 Vol. 8, Issue 3: 106-113, DOI: 10.25236/AJBM.2026.080312

Four-Dimensional Measurement and Intervention of Learning Efficacy

Chang Liu[1], Hongxia Xiao[2,*], Mingyu Lin[3], Xiaoju Yao[4]

[1]*School of Economics and Management, Hubei Business College, Wuhan, 430079, China*
[2]*School of Accounting, Hubei Business College, Wuhan, 430079, China*
[3]*School of Computer Science and Technology, Hubei Business College, Wuhan, 430079, China*
[4]*Academic Quality Assurance Office, Hubei Business College, Wuhan, 430079, China*
[*]*Corresponding author*

Abstract: *Aiming to solve core problems of ambiguous attention perception and coarse grained intervention in smart education this study proposes a four dimensional measurement and intervention empirical framework for learning efficacy. A layered perception system is constructed by integrating IR PoseNet for spatial orientation and Sub Eye Tracker for visual allocation along with Freq Physio for implicit cognitive load and GL Emotion for emotional utility to break through the limitations of single modal sensing. At the data governance layer the Efficiency Synchronization Algorithm eliminates spatiotemporal asynchrony caused by heterogeneous sampling frequencies of physiological and visual signals while a dynamic weighting mechanism adaptively adjusts modal contributions via scene adaptation factors to enhance robustness in complex environments. Multi scenario empirical results demonstrate that the system achieves high attention recognition accuracy and fake attention detection rates by leveraging the physiology behavior dual evidence chain which effectively resolves information asymmetry in educational supervision and realizes the shift from empirical supervision to data driven governance.*

Keywords: *Four-dimensional Layered Perception; Learning Efficacy; Fake Attention Recognition; Efficiency Synchronization Algorithm; Smart Education*

1. Introduction

Against the backdrop of educational digital transformation, intelligent learning intervention has emerged as a core pathway to enhance personalized learning outcomes [1]. Current learning state perception serves as a prerequisite for intelligent intervention and mostly relies on single behavioral modalities such as gaze trajectory [2] or physiological signals such as EEG [3]. This field faces critical issues such as fake attention detection failure [4] and poor scenario adaptability [5]. The former fails to distinguish the state of eye fixation combined with cognitive distraction due to the lack of a behavior and physiology dual evidence chain [6]. The latter suffers from insufficient reliability of cross modal feature fusion because multi modal data synchronization errors often exceed 100 milliseconds [7].

Although existing studies have attempted to integrate multi modal data such as 3D hand pose and gaze trajectory [8], they have not formed a complete closed loop of perception governance intervention and feedback [9]. Furthermore, they have not designed differentiated intervention strategies for typical scenarios such as online learning and AI study rooms [10]. To address these gaps, this study proposes a Four Dimensional Layered Perception and Intervention Closed Loop System. This system utilizes IR PoseNet coarse grained pose [11], Sub Eye Tracker fine grained gaze, Freq Physio implicit EEG [12], and GL Emotion emotional expression as core perception dimensions. By employing the Efficiency Synchronization Algorithm, data synchronization error is controlled within 50 milliseconds. The system achieves cross modal feature fusion through dynamic weighting and delivers light touch interventions such as virtual encouragement cards or physical virtual collaborative measures such as lighting adjustment and guided meditation for different scenarios. A one to three minute observation window is used to monitor physiological rhythm recovery rates to complete the feedback iteration.

2. Related Research and Solution Overview

This section systematizes the research scope of smart learning perception while elucidating the

ISSN 2616-5902 Vol. 8, Issue 3: 106-113, DOI: 10.25236/AJBM.2026.080312

measurement logic spanning from explicit behaviors to implicit physiological states. It also delineates the comprehensive technical roadmap for four dimensional layered acquisition and the intervention closed loop.

2.1 Full caliber Measurement from Explicit Behaviors to Implicit Physiology

This research focuses on the entire chain of learning state perception and intervention in the field of intelligent education to construct a full caliber multi dimensional learning state measurement system.

1) Full-caliber Measurement Scope. The research covers key links from low level feature extraction to cross scenario field verification. At the feature processing level, this study moves beyond the limitation of relying solely on single visual signals. Instead, it integrates IR PoseNet for infrared pose, Sub Eye Tracker for gaze tracking, Freq Physio for physiological frequency, and GL Emotion for emotional expression to enable comprehensive capture of both explicit behavioral features and implicit physiological states.

2) Measurement Depth and Precision Objectives.The core objective is to overcome the limitation of insufficient information integrity from single signals. By synergizing global features such as pose and emotional expression with local features such as fine grained gaze and implicit brainwaves, the system captures both the overall performance of learning states and fine grained implicit fluctuations.

3) Application Value.The ultimate goal is to develop a highly reliable and scalable smart efficacy guardian efficiency protection system. Through precise measurement of learning states, it reduces blind investment in educational resources, resolves adaptation issues during technical deployment, and promotes the upgrading of educational informatization toward intensive and personalized services.

2.2 Four dimensional Layered Acquisition and Intervention Closed loop Roadmap

This study adopts a closed loop technical framework fully aligned with the four layer architecture covering multi modal perception, data governance, hierarchical intervention, and feedback iteration. The roadmap is structured as follows Figure 1.

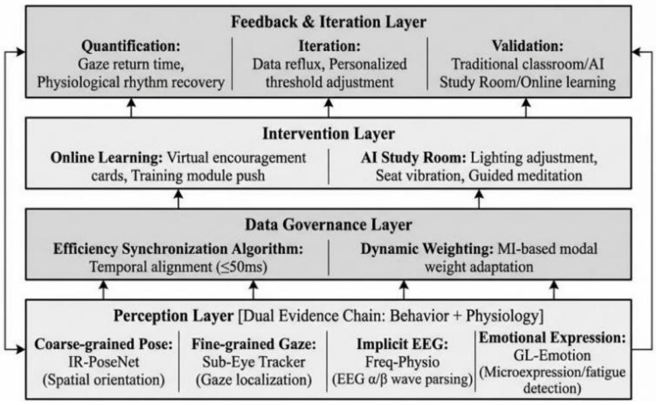

Figure 1: Four-Dimensional Layered Perception and Intervention Closed-Loop Architecture for Learning Efficacy.

1) Construction of the Four dimensional Layered Perception System. This system enables synchronized acquisition of explicit behavioral and implicit physiological signals to form a behavior physiology dual evidence chain. It leverages IR-PoseNet for coarse grained spatial orientation detection, Sub-Eye Tracker for sub pixel fine grained gaze localization, Freq-Physio for implicit brainwave rhythm parsing, and GL-Emotion for microexpression and fatigue detection to establish a layered feature foundation for comprehensive learning state perception.

2) Efficiency Synchronization and Dynamic Feature Alignment. To address sampling frequency discrepancies in multi source heterogeneous data such as brainwaves and visual signals, the efficiency synchronization algorithm is applied to achieve temporal alignment with errors controlled within 50 milliseconds. Additionally, dynamic weighting based on mutual information is implemented to adapt modal weights according to scenario conditions such as low light or occlusion to ensure high quality aligned features for subsequent fusion analysis.

3) Precise Triggering of Hierarchical Intervention Strategies. Based on scenario differences between

ISSN 2616-5902 Vol. 8, Issue 3: 106-113, DOI: 10.25236/AJBM.2026.080312

online learning and AI Study Rooms, the system triggers differentiated interventions dynamically according to the measured concentration index. For online learning, it delivers light touch interventions such as virtual encouragement cards and targeted training module pushes. For AI Study Rooms, it enables physical environment synergy including lighting adjustment, seat vibration prompts, and guided meditation to form an automated closed loop from state recognition to feedback correction.

4) Multi scenario Validation and Iterative Optimization. The system is deployed in diverse scenarios including traditional classrooms, AI Study Rooms, and online learning environments. Multi sample data are collected to quantitatively evaluate efficacy via core indicators such as gaze return time, physiological rhythm recovery, and fake attention mitigation. Data reflux and personalized threshold adjustment drive iterative system optimization to ultimately form a replicable standardized solution for educational efficacy enhancement.

3. Four-dimensional Measurement

3.1 Measurement Logic and Layered Strategy

In smart education scenarios, single modal perception has notable limitations as it cannot differentiate the diverse cognitive intentions behind superficially similar behaviors and particularly fails to identify fake attention where learners maintain gaze fixation but exhibit cognitive disengagement. This research therefore constructs a four dimensional collaborative measurement system integrating coarse grained pose, fine grained gaze, implicit EEG, and emotional expression data to form a behavior physiology dual evidence chain. This system addresses information asymmetry in educational supervision and rectifies misjudgments caused by fake attention.

As illustrated in Table 1 Three point Six, the system adopts a layered acquisition strategy across four complementary dimensions where each dimension has a distinct core role and dedicated technical implementation.

Table 1: Layered Acquisition Strategy of the Four-dimensional Fusion Perception Model

Dimension	Core Role	Method & Technology
Coarse-grained Pose	Spatial orientation & orientation cues	IR-PoseNet
Fine-grained Gaze	Real gaze target & gaze drift detection	Sub-Eye Tracker
Implicit EEG	Internal focus & anti-fake validation	Freq-Physio
Emotional Expression	Sentiment utility & fatigue detection	GL-Emotion

This research integrates four perception dimensions into a layered acquisition system where IR-PoseNet captures coarse-grained spatial orientation to establish the overall behavioral context of the learner. Simultaneously, the Sub-Eye Tracker identifies fine-grained gaze targets and drift to track visual allocation patterns. To address the challenge of fake attention, Freq-Physio analyzes implicit EEG rhythms to provide non-fakeable physiological evidence of cognitive engagement. Finally, GL-Emotion assesses sentiment utility and fatigue to capture emotional fluctuations correlating with sustained learning states. Together, these dimensions form a behavior-physiology dual evidence chain for the full-caliber measurement of learning efficacy.

This layered multi modal data framework not only resolves the ambiguity of fake attention but also enhances system robustness in extreme scenarios such as low light environments or partial occlusion to lay a solid foundation for reliable learning state assessment.

3.2 Multi-dimensional Behavioral Efficacy Quantification

This section integrates spatial action and emotional dimensions to standardize the quantification of learning productivity factors.

1) Spatial Orientation Measurement. Leveraging the IR PoseNet algorithm this layer addresses low light and occlusion challenges in natural learning environments. By modeling the head as an asymmetric rigid body with 2D Gamma distribution soft labels and a regularized divergence loss function it robustly quantifies macro head movements including pitch and yaw. This determines if learners face instructional content providing foundational spatial cues to identify fake attention and reduce supervision asymmetry.

2) Action Validity Measurement. Employing the 3D HandTrack algorithm with a Swin Transformer backbone and Adaptive Virtual Multi view technology this layer captures fine grained hand movements. It overcomes perspective limitations and occlusions to distinguish productive actions such as note taking or material consultation from distractions such as gadget use or pen spinning. This provides critical

ISSN 2616-5902 Vol. 8, Issue 3: 106-113, DOI: 10.25236/AJBM.2026.080312

behavioral evidence for evaluating learning efficiency.

3) Sentiment Utility Measurement. Using the GL Emotion dual branch CNN this layer analyzes emotional dynamics via global facial features and attention guided key region extraction focusing on the eyes and mouth. Integrated with deep product and residual correction strategies it accurately detects fatigue related microexpressions such as frowning or yawning. A mixed loss function enhances few shot recognition and generalization enabling the transition from basic state detection to emotion aware efficacy assessment.

3.3 Implicit-Visual Collaborative Calibration

This section acts as the core calibration layer of the measurement system to address misjudgments arising from the disconnect between explicit behaviors and implicit states. Such disconnects are exemplified by fake attention where learners maintain proper posture while lacking cognitive engagement. The system relies on spatiotemporal alignment and a dual modal technical pipeline to establish robust cross constraints between physiological and behavioral data.

Unifying the time and space dimensions of multi source data is a prerequisite for collaborative calibration. As detailed in Table 2, implicit EEG signals within the Freq-Physio module serve as the time reference at a native 250 Hertz sampling rate. Coarse grained pose, fine grained gaze, and emotional expression are upsampled to match this frequency while spatial mapping matrices unify various visual coordinate systems. All aligned modalities achieve a synchronization error of 50 milliseconds or less to ensure valid cross modal fusion.

The technical implementation of implicit-visual calibration follows the workflow illustrated in Figure 2. For implicit physiology, EEG signals are analyzed via feature extraction and power spectral density analysis to parse alpha and beta rhythms for quantifying actual brain activity as objective focus evidence. For explicit vision, the gaze estimation pipeline achieves sub pixel localization to measure attention distribution via gaze dispersion and fixation frequency. The system jointly judges attention decline so that true focus is only validated when both behavioral gaze fixation and physiological cognitive engagement are confirmed. Mismatches effectively trigger the identification of fake attention. This dual calibration mechanism establishes robust cross constraints to distinguish fake attention and enhance the rigor of the measurement framework in complex educational scenarios while mitigating information asymmetry.

Table 2: Spatiotemporal Parameters and Alignment Strategies of Each Modality.

Modality	Orig. Freq.	Aligned Freq.	Spatial Src.	Map Matrix	Sync Error
Coarse-grained Pose	30Hz	250Hz	Infrared Camera	M_p	≤50ms
Fine-grained Gaze	20Hz	250Hz	Monocular Camera	M_g	≤50ms
Implicit EEG	250Hz	250Hz	64-channel EEG Sensor	-	Native Synchronization
Emotional Expression	25Hz	250Hz	RGB Camera	M_f	≤50ms

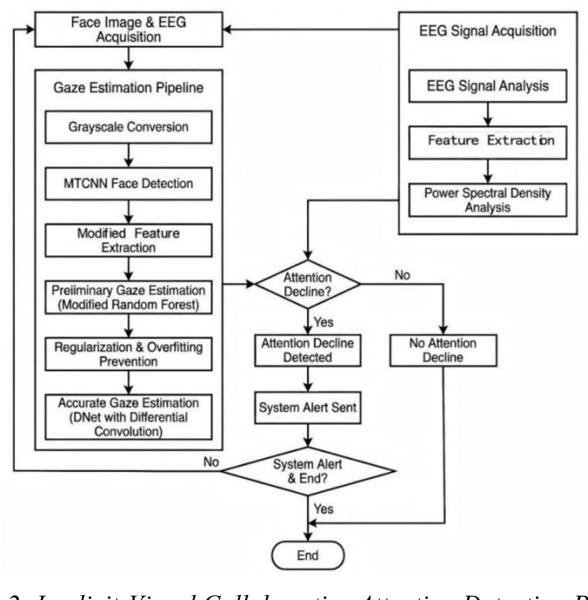

Figure 2: Implicit-Visual Collaborative Attention Detection Pipeline.

ISSN 2616-5902 Vol. 8, Issue 3: 106-113, DOI: 10.25236/AJBM.2026.080312

4. Data Governance and Efficacy Intervention Closed-loop

This section constructs a governance framework for multi source heterogeneous data in smart learning environments and builds an intervention closed loop based on the smart efficacy guardian. It eliminates measurement level information asymmetry to translate full caliber perceived data into targeted pedagogical interventions for optimizing educational resource allocation efficiency.

4.1 Efficiency Synchronization and Standardization

To resolve sampling frequency conflicts and spatio temporal asynchrony between physiological and visual signals, standardized data governance protocols are established. Temporally, the Efficiency Synchronization Algorithm takes 250 Hertz Freq Physio signals as the baseline to perform linear interpolation on low frequency modalities such as IR PoseNet, ensuring alignment with errors of 50 milliseconds or less on a unified timeline. Spatially, a unified 3 D coordinate system is built via device calibration with transformation matrices standardizing feature coordinates from diverse devices to support cross modal correlation modeling.

4.2 Dynamic Weight Allocation

A scenario adaptive dynamic weight allocation mechanism is introduced to enhance system robustness in complex environments. This mechanism adjusts fusion weights by calculating mutual information between global and local features. In fake attention scenarios, it reduces macro pose weight and elevates the contribution of non fakeable implicit EEG and fine grained gaze. For environment specific constraints such as low light or occlusion, the system optimizes modal weights in real time based on scenario adaptation factors.

Table 3 presents concrete weight allocation instances for three typical scenarios to illustrate the adaptive adjustment logic. EEG is prioritized at forty percent in low light online classrooms because it is immune to lighting, while emotion recognition is downweighted to ten percent. In occluded study rooms, gaze tracking becomes the core behavioral evidence at thirty percent, and pose weight is reduced to fifteen percent. For exams, balanced weights ensure comprehensive state assessment to maintain perception accuracy across diverse scenarios.

Table 3: Weight Allocation Instances in Typical Scenarios

Scenario Type	Scenario Adaptation Factor	Modality Weight
Online Classroom (Low-light)	Pose:0.8, Gaze:0.6, EEG:1.0, Emotion:0.5	Pose:30%, Gaze:20%, EEG:40%, Emotion:10%
Offline Study Room (Occlusion)	Pose:0.5, Gaze:0.9, EEG:1.0, Emotion:0.8	Pose:15%, Gaze:30%, EEG:35%, Emotion:20%
Exam Scenario (High-focus)	Pose:0.9, Gaze:0.9, EEG:0.8, Emotion:0.9	Pose:25%, Gaze:25%, EEG:20%, Emotion:30%

4.3 Smart-Efficacy Guardian System Development

Based on the governance logic the Smart Efficacy Guardian system is developed for full scenario deployment. The online focus monitoring module adopts a lightweight architecture integrating the Sub Eye Tracker and Freq Physio to calculate focus indices and push light touch interventions such as encouragement cards or training modules. The AI Study Room module links physical devices with digital media triggering lighting adjustment and seat vibration or guided meditation for severe distraction to form a virtual physical collaborative intervention loop.

4.4 Intervention-Feedback Closed-loop and Iteration

The intervention framework relies on continuous feedback and iteration to ensure precision and efficiency. Post intervention a one to three minute observation window quantifies effectiveness by monitoring gaze return time and physiological rhythm recovery. Multi source data reflux drives personalized model iteration where the system adjusts intervention thresholds such as lighting intensity or vibration amplitude based on long term user profiles. Aligning these parameters with individual attention patterns realizes the transformation from precise perception to effective governance.

5. Experimental Verification and Efficiency Evaluation

5.1 Experimental Design and Deployment

This section conducts empirical research to verify the practical utility of the measurement framework and intervention closed loop. The research covers three distinct scenarios including traditional mixed teaching classrooms, AI Study Rooms, and uncontrolled online learning environments by adopting a combination of offline hardware deployment and online lightweight acquisition. Diverse subjects such as undergraduates, postgraduate entrance exam candidates, and remote learners were recruited to verify system robustness against neighbor interference, the efficacy of virtual physical collaborative intervention, and the adaptability of lightweight hardware. By deploying panoramic cameras, linking intelligent devices, and integrating IR PoseNet and Freq Physio acquisition chains, a full caliber data backflow mechanism covering pose, gaze, brainwaves, and expressions was established.

5.2 Validation of Core Efficacy Indicators

The core efficacy was verified through scenario specific indicator testing with key results summarized in Table Three point Twelve and Table Three point Thirteen.

1) Online Learning Scenario Verification. To address single camera constraints and background interference, the proposed system was compared with the traditional binocular vision scheme. Table 4 details the performance of core perception and intervention indicators. Data indicate that the proposed system achieved an attention recognition accuracy of 92.7 percent, which is significantly higher than the traditional scheme. Notably, in solving the challenge of fake attention where learners stare but are distracted, the detection rate reached 89.3 percent through brainwave frequency parsing and gaze trajectory fusion, representing a 22.1 percentage point improvement over traditional methods. Furthermore, the background interference false alarm rate was halved and the late night intervention response rate increased to 78 percent, confirming the adaptability of the lightweight acquisition chain in uncontrolled environments.

Table 4: Index Validation in Online Learning Scenarios

Indicator	Webcam	Traditional Binocular Scheme	Improvement
Attention Accuracy	92.7%	85.3%	7.4%
Fake Attention Rate	89.3%	67.2%	22.1%
Bg Interference False Alarm	7.8%	15.6%	50% (Relative)
Late-night Response Rate	78%	63%	15%

2) AI Study Room Scenario Verification. As shown in Table 5, in the AI Study Room, the comparison between the virtual physical collaborative intervention group and the single physical intervention group validated the superiority of the governance scheme. The collaborative approach increased daily deep focus duration from 1.2 hours to 2.3 hours, reflecting a 91.7 percent improvement. Distraction recovery time was shortened from 92 seconds to 55 seconds, a reduction of 40.2 percent. The misjudgment rate for reasonable relaxation was reduced by over 60 percent to balance intervention intensity and user experience. An environmental intervention response rate of 85 percent further proves the reliability of the virtual physical closed loop.

Table 5: Index Validation in AI Study Room Scenarios

Indicator	VP Collaborative Group	Single Physical Intervention	Improvement
Deep Focus Duration (H/D)	2.3	1.2	91.7%
Distraction Recovery Time (Sec.)	55	92	40.2%
Relaxation Misjudgment Rate	4.8%	12.5%	61.6% (Relative)
Env. Intervention Response Rate (Light/Vib.)	85%	67%	26.9%

3) Traditional Mixed Teaching Scenario Verification. In traditional classrooms, the system achieved a group attention synchronization rate of 81 percent. The collective distraction response delay was reduced from 35 seconds under manual supervision to 12 seconds, representing a 65.7 percent decrease. This result significantly minimizes the lag of manual supervision and verifies the scalability and resource allocation optimization capabilities of the system in group teaching scenarios.

5.3 Empirical Summary and Productivity Optimization

This research pioneers a four dimensional collaborative measurement system integrating IR PoseNet and 3D HandTrack as well as GL Emotion with Freq Physio and Sub Eye Tracker to achieve full caliber learning efficacy measurement from explicit behaviors to implicit physiology. Validated via the Efficiency Synchronization Algorithm for precise spatiotemporal alignment, the Smart Efficacy Guardian intervention closed loop delivered exemplary performance across multiple scenarios by enhancing attention recognition and significantly increasing deep focus duration. These results mitigate information asymmetry and drive the shift from empirical supervision to data driven governance while optimizing educational resource allocation and human capital output. Notwithstanding these breakthroughs, limitations remain such as a narrow sample demographic and insufficient algorithmic robustness in extreme environments including backlighting or occlusion. Future work will focus on transfer learning across diverse groups and scenarios alongside adversarial training and dynamic personalized intervention models to establish a universal standardized pathway for smart educational technologies.

6. Conclusions

To address the pain points of attention perception and fake attention recognition in smart education, this study constructs a four-dimensional learning efficacy measurement and intervention system. Empirical results verify its excellent performance, promote data-driven educational governance. Future work will optimize the model’s adaptability.

Acknowledgment

This study was supported by the Provincial Teaching Research Project for Undergraduate Universities in Hubei Province (Grant No. 2024548), the Science and Technology Innovation Team Project of Hubei Province (Grand No. T201838).

References

[1] Ding, J. "The Theoretical Model, Practice Logic and Educational Value of Multimodal Collaborative Learning Analysis." Journal of Distance Education (1672-0008), 2 (2023).

[2] Xu, Hui, et al. "Analyzing students' attention by gaze tracking and object detection in classroom teaching." Data Technologies and Applications, 57, 5 (2023): 643-667.

[3] Yang, Y, et al." Research on Deep Learning Monitoring Based on Multimodal Data in Intelligent Learning Environment." e-Education Research, 43, 6 (2022):68-76.

[4] Su, Caizhen, et al. "Using synchronized eye movements to predict attention in online video learning." Education Sciences, 14, 5 (2024): 548.

[5] Yin, Rui, et al. "Research Progress and Prospects of Multimodal Learning Analytics Based on Systematic Literature Review." Modern Distance Education Research, 34, 06(2022):54-63.

[6] Liu, Qingtang, et al. "Developments and Prospects of Empirical Research on Multimodal Learning Analysis." e-Education Research, 43, 01(2022):71-78+85.

[7] Kao, Y., Pan, B., Xu, M., Lyu, J., Zhu, X., Chang, Y., ... & Lei, Z.. “Toward 3d face reconstruction in perspective projection: Estimating 6dof face pose from monocular image.” IEEE Transactions on Image Processing, 32, (2023): 3080-3091.

[8] Wan X, Ju J, Tang J, et al. “MPCTrans: Multi-Perspective Cue-Aware Joint Relationship Representation for 3D Hand Pose Estimation via Swin Transformer.” Sensors, 24, 21 (2024): 7029.

[9] Wang Y, Zheng Y. "Intelligent Learning Intervention: Realistic Dilemma, Original Clarification and Proper Form." Open Education Research, 29, 02(2023):103-111.

[10] Vu, Nhung Kim Thi, Thanh Chi Phan, and Thanh Tu Ngo. "Enhance teaching method of self-study competence in the digital environment." Journal of Education and Teaching (JET), 4, 3 (2023): 294-307.

[11] Liu, Hai, et al. "ARHPE: Asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction." IEEE Transactions on Industrial Informatics, 18, 10 (2022): 7107-7117.

[12] Wu, Jade, Jingwen Dou, and Sofia Utoft. "Refining Human-Data Interaction: Advanced Techniques for EEGEyeNet Dataset Precision." International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland, (2024).