Article

# Effects of different AI-driven Chatbot feedback on learning outcomes and brain activity

Check for updates

Jiaqi Yin[1,2], Haoxin Xu [1,2], Yafeng Pan [3] ✉ & Yi Hu [4] ✉

Artificial intelligence (AI) driven chatbots provide instant feedback to support learning. Yet, the impacts of different feedback types on behavior and brain activation remain underexplored. We investigated how metacognitive, affective, and neutral feedback from an educational chatbot affected learning outcomes and brain activity using functional near-infrared spectroscopy. Students receiving metacognitive feedback showed higher transfer scores, greater metacognitive sensitivity, and increased brain activation in the frontopolar area and middle temporal gyrus compared to other feedback types. Such activation correlated with metacognitive sensitivity. Students receiving affective feedback showed better retention scores than those receiving neutral feedback, along with higher activation in the supramarginal gyrus. Students receiving neutral feedback exhibited higher activation in the dorsolateral prefrontal cortex than other feedback types. The machine learning model identified key brain regions that predicted transfer scores. These findings underscore the potential of diverse feedback types in enhancing learning via human-chatbot interaction, and provide neurophysiological signatures.

The interactive learning features and flexibility of artificial intelligence (AI)-driven chatbots, in terms of time and location, have made their use increasingly popular in education[1]. Chatbots can provide learners with personalized content, instant feedback, and one-to-one guidance[2,3]. To enhance learning through chatbots, providing adequate feedback is essential, enabling learners to engage in self-regulated learning. Feedback helps learners reduce uncertainty in the learning process[4]. Feedback is a crucial element that affects learning across various environments and serves diverse purposes, such as assisting learning, clarifying expectations, reducing discrepancies, detecting error, and increasing motivation[5]. Feedback draws learners' attention to gaps in understanding and supports them in gaining knowledge and competencies[6]. Moreover, feedback aids learners in regulating their learning[7].

There is growing recognition that the lack of assessments and absence of feedback mechanisms have hindered the success of chatbots[8]. Advances in AI and natural language processing have prompted researchers to explore the effectiveness of cognitive feedback in chatbot-based learning. For instance, one study found that chatbot-based corrective feedback could promote the English learning experience[9], while another showed that it enhanced retention performance[10]. In addition, recent research has revealed

that, compared to teacher-based feedback, chatbot-based cognitive feedback can boost learning interest, perceived choice, and value, while alleviating pressure and cognitive load and improving mastery of applied knowledge[11]. However, the role and possible differences among different types of non-cognitive feedback in chatbot-based learning remain largely unclear. In previous studies, feedback primarily served to evaluate students' learning outcomes. In this study, however, feedback is used to facilitate the learning process itself. Previous assessments of chatbot-based feedback relied on tests or self-reports, lacking a neuroscience perspective. By incorporating brain measures of learning processes along with post-test outcome measures, this approach can deepen our understanding of the mechanisms behind different types of feedback design, providing insights into why a specific type of feedback may be more effective. It also provides insights for designing and optimizing various types of feedback for different educational scenarios.

In chatbot-based learning, students are encouraged to actively participate and manage their own learning. However, students are often reported to lack motivation and struggled with self-regulation, leading to poor performance in online learning environments[12,13]. Feedback is a crucial factor in promoting behavioral change, as it provides information on the gap between current achievements and desired goals[5,14]. Unlike cognitive feedback, which

[1]Shanghai Institute of Artificial Intelligence for Education, East China Normal University, Shanghai, 200062, China. [2]School of Computer Science and Technology, East China Normal University, Shanghai, 200062, China. [3]Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, 310058, China. [4]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China. ✉e-mail: yafeng.pan@zju.edu.cn; yhu@psy.ecnu.edu.cn

provides direct instructional content, metacognitive feedback supports "learning to learn" by guiding learners to develop self-regulation skills[15]. This form of feedback conveys strategies for effective learning behaviors, such as planning, monitoring, regulating, and reflecting[16]. Research has shown that metacognitive feedback can deepen understanding and support the transfer of knowledge to new contexts[17,18]. However, understanding exactly how metacognitive feedback impacts learning, particularly in chatbot-based environments, remains limited.

Humans adapt their behavior not only by observing the consequences of their actions but also by internally monitoring their performance, a capacity referred to as metacognitive sensitivity[19,20]. Also known as meta-cognitive accuracy, this refers to an individual's ability to distinguish between their correct and incorrect judgments[21]. Signal detection theory analysis suggests that metacognitive sensitivity correlates with task performance. According to meta-reasoning theory, metacognitive monitoring, as reflected in metacognitive sensitivity, is essential for thinking and reasoning by influencing metacognitive control – decisions such as whether to respond, switch strategies, or give up[22]. Ackerman and Thompson further argued that meta-reasoning involved the ability to reflect on and evaluate one's reasoning processes[22]. Enhancing students' metacognitive sensitivity can foster autonomous learning and support long-term learning development. In this study, we use post-answer confidence as a measure to examine whether different types of feedback can improve metacognitive sensitivity.

In addition, research grounded in cognitivism and social-cultural theory highlights the importance of considering the feedback recipient's needs, including encouragement[23-25] and attention to factors like locus of control and self-esteem[26]. Affective feedback, which aims to sustain students' interest, attention, or motivation, addresses these needs[27]. Affective feedback also helps mitigate the impact of negative experiences on motivation[28], and can stimulate student engagement by encouraging participation and learning[29,30]. Many AI-driven educational tools have integrated affective feedback, such as intelligent tutoring systems and chatbots, to induce positive emotions or alleviate negative emotions, with the goal of enhancing student motivation and engagement[31]. For example, various forms of affective feedback have been incorporated into intelligent tutoring systems to address students' frustration through congratulatory, encouraging, sympathetic, and reassuring messages[32]. Chatbots, by offering social interaction and affective cues like praise and emotional sharing, have been shown to elicit positive emotions during tasks like reading comprehension[33]. These affective cues have been linked to improved motivation, self-regulated learning, and academic performance[34]. Despite these advances, however, the effects of affective feedback on learning outcomes, particularly its neural underpinnings, remain insufficiently understood. This study aims to fill this gap by exploring the neural mechanisms associated with different types of feedback, including affective feedback, and their effects on learning outcomes.

Neuroimaging tools such as fNIRS (functional near-infrared spectroscopy), EEG (electroencephalography), and fMRI (functional magnetic resonance imaging) can capture different aspects of brain activity. fNIRS is an optical brain monitoring technique that uses near-infrared light to measure changes in the concentration of oxygenated and deoxygenated hemoglobin in the blood[35,36]. While fNIRS has lower spatial resolution and cannot measure deep brain activity, it offers better temporal resolution than fMRI[37], Its main advantages include being relatively low-cost, portable, safe, low in noise, and easy to operate. Unlike EEG and magnetoencephalography, fNIRS data are also less affected by electrical noise[38]. For these reasons, we designed an fNIRS experiment to study the neural mechanisms underlying AI-driven chatbot feedback.

Neuroimaging studies have highlighted the critical role of the Mirror Neuron System (MNS) in various socially relevant functions, including observation of others' actions, imitation learning, and social communication[39-41]. The MNS comprises a complex network involving frontal, temporal, and parietal areas[40,42]. Among these, activation in the superior temporal gyrus (STG) has been linked to social interaction and perception[43-45]. Feedback processing, on the other hand, is primarily

associated with frontoparietal brain regions, including the anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC), and parietal lobules[46-50]. In adults, the parietal cortex is more heavily relied upon than the ACC for processing informative and efficient feedback, aiding in performance adjustment and error correction[47,49,50].

Considerable progress has been made in understanding the neural mechanisms of metacognition[51,52]. However, most research has focused on cognitive neuroscience paradigms, such as perceptual decision-making[51-54], memory judgements[55-57], or problem-solving tasks[58,59]. Few studies have examined the neural associations between metacognition and learning process within educational settings. The PFC plays an important role in cognitive activity[60], serving as the brain region responsible for working memory, decision-making, and coping with novelty. Metacognition, as a high-order brain function, is strongly depends on the PFC[54,61]. Research on adult neural correlates of metacognitive monitoring across different tasks indicates a consistent involvement of a frontoparietal network[51,52]. Furthermore, emerging evidence highlights the involvement of the right rostrolateral PFC[53,62] and the anterior PFC (aPFC)[63] in metacognitive sensitivity.

Speaking of affective processing and feedback, research has highlighted the central roles of the amygdala and insula in the emotional perception and processing[64-66]. Also, higher-order cortical structures, including the medial prefrontal cortex and the rostral anterior cingulate cortex, support affective information[67,68].The frontal regions are widely implicated in emotional processing[69]. Furthermore, the temporo-parietal junction (TPJ) has been consistently identified as playing a key role in emotional functions[70-72].

This study aims to explore how different types of feedback affect learning processes, which in turn affect learning outcomes and metacognitive sensitivity. Therefore, we designed and developed three types of educational chatbots providing metacognitive, affective, or neutral feedback. This study primarily focuses on metacognitive and affective feedback types and neutral feedback serves as a control condition. Metacognitive feedback guides students to think and reflect on the learning process through targeted questions while affective feedback contains praise and encouragement statements. Neutral feedback is not judgmental and not evaluative[73,74]. It does not interfere with the learning process and serves only as a break prompt. Although it does not resemble typical feedback, it serves as a control condition, helping us understand the effects of the other feedback types. Using behavioral measures and brain imaging, the study addresses two central questions. First, within a chatbot-based learning environment, do metacognitive, affective, and neutral feedback have distinct effects on learning outcomes and metacognitive sensitivity? While several behavioral studies have demonstrated the effectiveness of cognitive feedback in improving students' learning performance[75], self-efficacy[76], and self-regulated learning[77], it remains unclear whether different feedback types, such as metacognitive, affective, and neutral feedback, yield varying effects. Second, how do these feedback types affect the learning process? Specifically, what are the neurocognitive mechanisms underlying students' responses to different feedback conditions? Previous behavioral research has not addressed these questions, and little is known about how the brain responds to varying feedback conditions. Adopting a neurocognitive approach to these questions may provide valuable insight into optimizing the integration of diverse feedback types into educational chatbots for specific learning goals.

We addressed these questions by comparing the learning process (measured by brain activity in targeted areas), learning outcomes (assessed via retention and transfer tests), and metacognitive sensitivity of students who interacted with educational chatbots providing metacognitive, affective, or neutral feedback. Importantly, this study aimed at providing a neurobiological testbed for different types of feedback. Our goal was to characterize the human-chatbot interaction from a neurophysiological perspective, and explore whether and how different feedback types might facilitate these neurophysiological processes. Guided by prior research on brain activity in social, feedback-related, and metacognitive brain areas, we placed fNIRS probes on the PFC and right temporoparietal areas of the
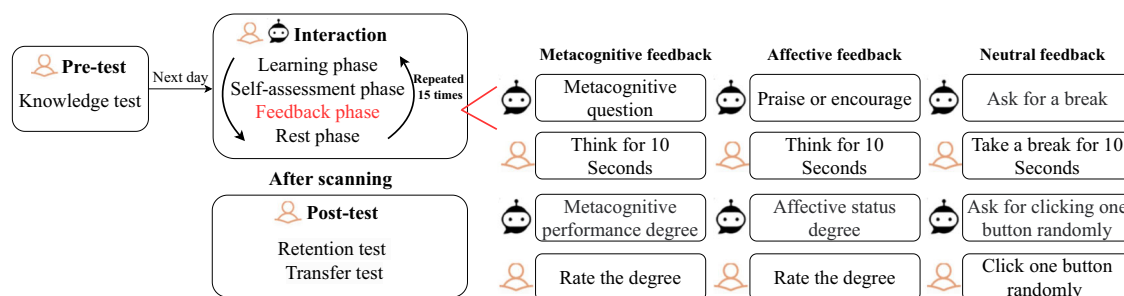
**Fig. 1 | Visualization of experimental sessions.** There are three sessions, entailing pre-test, human-chatbot interaction, and post-test.

brain. We expected that metacognitive feedback would enhance learning outcomes, especially on the transfer tests, relative to other types of feedback. This feedback was also anticipated to increase students' brain activity, particularly in prefrontal and parietal regions. Additionally, we hypothesized that affective feedback would improve retention outcomes and enhance brain activity in the prefrontal regions and TPJ.

Participants' brain activity was acquired using fNIRS as they interacted with different chatbots. These human-chatbot interactions encompassed the delivery of biology learning content, self-assessment tasks, and feedback provision. Behavioral results revealed the different effects of different feedback on retention, transfer, and metacognitive sensitivity. The differences of brain activation observed during the human-chatbot interaction was used to infer the underlying processes of different feedback types, providing insights into their effectiveness. The results may have important implications for optimizing chatbot-based feedback and advancing AI-supported pedagogical applications in educational settings.

## Results

We recruited a group of college students to investigate how different types of feedback affect learning outcomes, metacognitive sensitivity, and brain activity. Participants were randomly assigned to one of three groups: the Metacognitive Feedback group (MF group), the Affective Feedback group (AF group), or the Neutral Feedback group (NF group). Each group underwent three sessions (Fig. 1, for more details, see Methods): (1) a pre-test session (initial visit) to assess participants' baseline knowledge; (2) a human-chatbot interaction phase consisting of 15 trials, each including learning, self-assessment, and feedback phases. In the feedback phase, the MF group received metacognitive feedback, the AF group received affective feedback, while the NF group received neutral feedback (for detailed feedback descriptions, see Methods); (3) a post-test session to evaluate the effects of feedback on retention (using the same questions as the pre-test) and transfer (using new questions).

### Behavioral results

**Retention scores.** After controlling the influence of pre-test knowledge test scores, students' retention scores showed a significant difference in the feedback type ($F = 7.51$, $p = 0.001$). The effect size was 0.15, and the statistical power was 0.94. Bonferroni post-hoc comparisons revealed that retention scores in the MF group ($M = 13.03$, $SD = 2.28$) were significantly higher than in the NF group ($M = 10.86$, $SD = 2.81$; *corrected p = 0.001*). It was found that retention scores in the AF group ($M = 12.34$, $SD = 2.14$) were significantly higher than in the NF group (*corrected p = 0.042*). However, retention scores were comparable between the MF and AF groups (*corrected p = 0.580*).

**Transfer scores.** After controlling the influence of pre-test knowledge test scores, students' transfer scores showed a significant difference in the feedback type ($F = 6.75$, $p = 0.002$). The effect size was 0.14, and the statistical power was 0.91. Bonferroni post-hoc comparisons revealed that transfer scores in the MF group ($M = 9.38$, $SD = 2.88$) were significantly higher than in the NF group ($M = 6.72$, $SD = 3.34$; *corrected*

$p = 0.002$) and the AF group ($M = 7.59$, $SD = 2.97$; *corrected p = 0.050*). However, transfer scores were comparable between the AF and NF groups (*corrected p = 0.739*).

**Metacognitive sensitivity.** After controlling the influence of pre-test metacognitive sensitivity, students' metacognitive sensitivity showed a significant difference in the feedback type ($F = 5.48$, $p = 0.006$). The effect size was 0.12, and the statistical power was 0.84. Bonferroni post-hoc comparisons revealed that metacognitive sensitivity in the MF group ($M = 0.71$, $SD = 0.11$) was significantly higher than in the NF group ($M = 0.63$, $SD = 0.12$; *corrected p = 0.018*) and the AF group ($M = 0.63$, $SD = 0.10$; *corrected p = 0.014*). However, metacognitive sensitivity was comparable between the AF and NF groups (*corrected p > 0.999*).

### fNIRS results of learning module

**Brain activation patterns in each group.** One sample t-test (above zero) with FDR corrections was applied to 20 channels of brain activation data ($\triangle$HbO) per group to investigate which channels were activated during the human-chatbot interaction period. Heat maps of the brain area with significant activation in the MF group are shown in Fig. 2. The corresponding activated brain areas include dorsolateral prefrontal cortex (DLPFC) (channel 07; $t = 2.84$, *corrected p = 0.028*), frontopolar area (FP) (channel 08; $t = 2.74$, *corrected p = 0.028*), supramarginal gyrus (SMG) (channel 12; $t = 2.77$, *corrected p = 0.028*), and middle temporal gyrus (MTG) (channel 19; $t = 3.11$, *corrected p = 0.028*).

As shown in Fig. 3, in the AF group, the corresponding activated brain areas include DLPFC (channel 07; $t = 3.95$, *corrected p < 0.001*), SMG (channel 11; $t = 3.30$, *corrected p = 0.03*), and STG (channel 17; $t = 2.67$, *corrected p = 0.050*).

As shown in Fig. 4, in the NF group, the corresponding activated brain areas include the DLPFC (channel 01; $t = 2.82$, *corrected p = 0.044* and channel 04; $t = 2.71$, *corrected p = 0.044*).

**Group differences in brain activation.** The results showed significant group differences in brain activation (Table 1). Brain activation was significantly higher in the MF group than in the AF group and NF group in channel 8 (frontopolar area) and channel 19 (middle temporal gyrus). In addition, brain activation was significantly higher in the AF group than in the MF and NF groups in channel 11 (supramarginal gyrus). In contrast, brain activation was significantly higher in the NF group than in the MF and AF groups in channel 1 and channel 4 (dorsolateral prefrontal cortex).

**Neural-behavioral relationships.** First, the Pearson correlation analysis indicated that channel 8 activation during the whole learning module was positively correlated with post-learning metacognitive sensitivity in the MF group ($r = 0.42$, $p = 0.023$), an association not seen in the other groups. Furthermore, channel 19 activation during the whole module was positively correlated with retention scores in the MF group ($r = 0.45$, $p = 0.014$), an effect not observed in the other groups. However, no significant correlations were observed between activated channels and
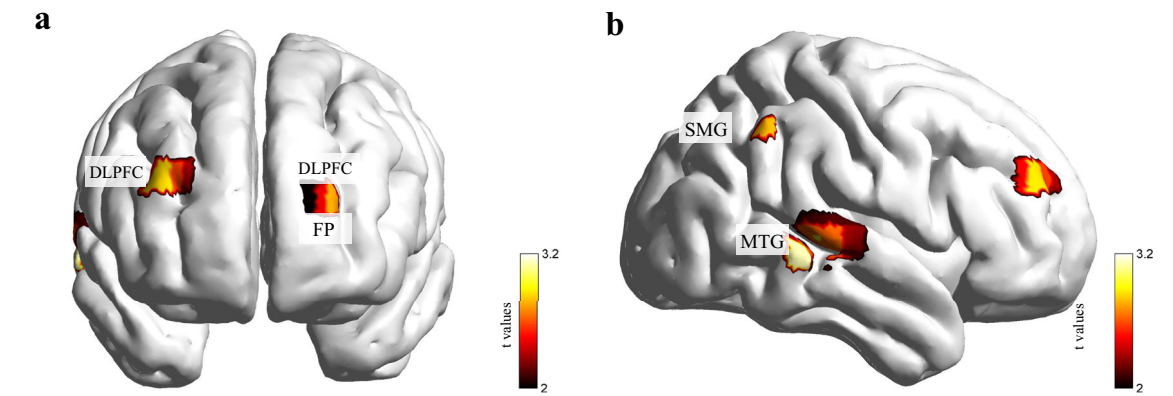
**Fig. 2 | Brain responses to metacognitive feedback.** The metacognitive feedback group showed significantly increased activation in (**a**) the dorsolateral prefrontal cortex (DLPFC), frontopolar area (FP), and (**b**) supramarginal gyrus (SMG) and middle temporal gyrus (MTG).
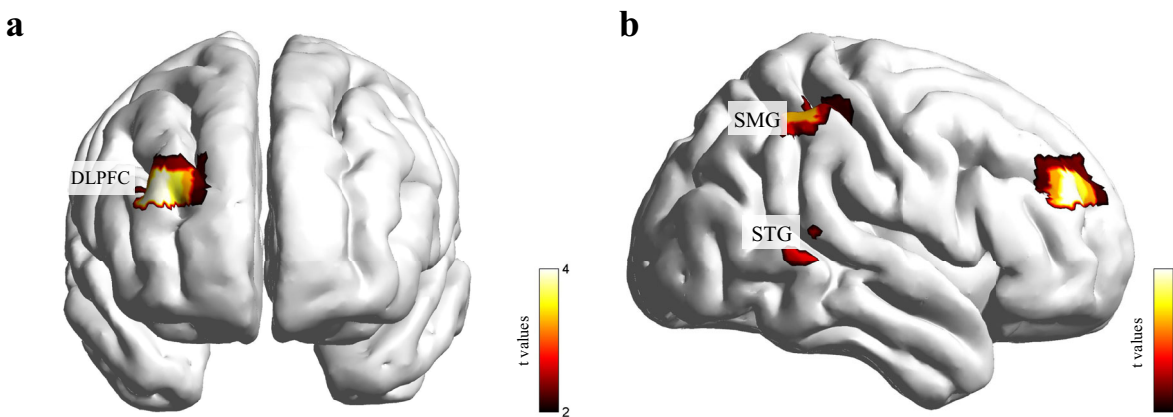


**Fig. 3 | Brain responses to affective feedback.** The affective feedback group showed significantly increased activation in (**a**) the dorsolateral prefrontal cortex (DLPFC), and (**b**) the supramarginal gyrus (SMG) and superior temporal gyrus (STG).

## Table 1 | Group differences in brain activation

| Regions-channels | Group | $M$ | $SD$ | $p$ | $\eta^2$ | Post hoc |
|---|---|---|---|---|---|---|
| DLPFC-CH 1 | MF | −0.0004 | 0.001 | < 0.001 | 0.21 | $M_{NF} > M_{MF}$ |
| | AF | −0.001 | 0.002 | | | $M_{NF} > M_{AF}$ |
| | NF | 0.002 | 0.005 | | | |
| DLPFC-CH 4 | MF | −0.0005 | 0.001 | < 0.001 | 0.19 | $M_{NF} > M_{MF}$ |
| | AF | −0.0008 | 0.002 | | | $M_{NF} > M_{AF}$ |
| | NF | 0.002 | 0.003 | | | |
| DLPFC-CH 7 | MF | 0.001 | 0.002 | < 0.001 | 0.21 | $M_{MF} > M_{NF}$ |
| | AF | 0.002 | 0.002 | | | $M_{AF} > M_{NF}$ |
| | NF | −0.0008 | 0.002 | | | |
| FP-CH 8 | MF | 0.001 | 0.002 | < 0.001 | 0.18 | $M_{MF} > M_{AF}$ |
| | AF | −0.0007 | 0.002 | | | $M_{MF} > M_{NF}$ |
| | NF | −0.0008 | 0.002 | | | |
| SMG-CH 11 | MF | −0.0003 | 0.002 | < 0.001 | 0.17 | $M_{AF} > M_{MF}$ |
| | AF | 0.001 | 0.002 | | | $M_{AF} > M_{NF}$ |
| | NF | −0.0005 | 0.002 | | | |
| SMG-CH 12 | MF | 0.0006 | 0.001 | 0.025 | 0.08 | $M_{MF} > M_{AF}$ |
| | AF | −0.0004 | 0.002 | | | |
| | NF | 0.0002 | 0.001 | | | |
| STG-CH 17 | MF | −0.001 | 0.001 | 0.007 | 0.11 | $M_{AF} > M_{NF}$ |
| | AF | 0.001 | 0.001 | | | |
| | NF | −0.003 | 0.001 | | | |
| MTG-CH 19 | MF | 0.001 | 0.002 | 0.005 | 0.12 | $M_{MF} > M_{AF}$ |
| | AF | −0.001 | 0.002 | | | $M_{MF} > M_{NF}$ |
| | NF | −0.0007 | 0.004 | | | |

transfer scores. Consequently, we employed machine learning analyses to take a more predictive approach to understanding transfer performance. This multivariate strategy allows us to examine how multiple factors, including brain activity features, contribute to the neural-behavioral relationship.

Second, for all groups, machine learning models using brain activity from the entire learning module predicted transfer scores more accurately than models based on individual phases of the learning process. This suggests that transfer performance is influenced by the interplay of the learning, assessment, and feedback phases, which justifies the use of complete learning modules in our models. The Extra Trees model, an ensemble learning method, provided the best prediction of transfer scores across all groups, enhancing performance through multiple decision trees. It automatically identifies key features within high-dimensional data, highlighting variables most relevant to prediction through feature importance analysis. This capability is valuable for complex neuroscience data, such as fNIRS signals, which often contain multiple dimensions and complex temporal dynamics. Compared to similar models such as Random Forests, Extra Trees introduces additional randomness during training by randomly selecting both features and segmentation thresholds. This added randomness helps reduce overfitting, particularly beneficial when dealing with limited sample sizes.

For the MF group, the results revealed that the mean $R^2$ achieved 0.88, the mean MAE achieved 0.72, and the mean MSE achieved 0.93. Figure 5 shows the Shapley additive explanation (SHAP) summary plot, which ranks features based on their importance in predicting transfer scores for the MF

group. The SHAP values for each activity are displayed along the x-axis, ranging from −1 to 1. A positive SHAP value increases the probability of high scores, while a negative SHAP value decreases this likelihood. Key brain regions that predicted transfer scores included the FP, SMG, DLPFC, and fusiform gyrus (FG). Among these, the features "CH-9-HbO-max", "CH-9-HbO-min", and "CH-9-HbR-min" did not demonstrate a simple linear relationship with the prediction outcomes. That is, the values of these features did not directly correlate with higher or lower scores. For the feature "CH-11-HbR-kurt", smaller values were associated with a higher likelihood of the model predicting higher scores, suggesting that lower HbR concentration in the supramarginal gyrus was linked to better performance. Conversely, for the feature "CH-7-HbO-max", larger values were associated with a higher likelihood of predicting higher scores, indicating that greater activity in the dorsolateral prefrontal cortex was related to higher student performance.

For the AF group, the results revealed that the Extra Trees model was the best in predicting transfer scores. The mean $R^2$ achieved 0.90, the mean MAE achieved 0.61, and the mean MSE achieved 0.83. Figure 6 shows the SHAP summary plot that orders features based on their importance in predicting transfer scores within the AF group. Key brain regions that predicted transfer scores included the SMG, DLPFC, MTG, FP, and. Among them, greater values of features, such as "CH-13-HbO-max" and "CH-19-HbR-max", were associated with higher predicted scores. Conversely, greater values of the feature "CH-3-HbO-max" were linked to lower predicted scores. This indicates that higher HbO concentration in the supramarginal gyrus and higher HbR concentration in the middle temporal gyrus were associated with higher transfer scores, while higher HbO concentration in the dorsolateral prefrontal cortex was linked to lower transfer scores. However, the influence of brain activity in the frontopolar area on prediction outcomes mirrored that of the metacognitive group, showing a nonlinear relationship.

For the NF group, the results revealed that the Extra Trees model was the best in predicting transfer scores. The mean $R^2$ achieved 0.87, the mean MAE achieved 0.92, and the mean MSE achieved 1.42. Figure 7 shows the SHAP summary plot that orders features based on their importance in predicting transfer scores in the NF group. Key brain regions that predicted transfer scores included the SMG, MTG, FP, and DLPFC. Among these, higher values of features, such as "CH-13-HbR-max", "CH-19-HbO-max", were associated with higher predicted scores. In contrast, smaller values of the feature "CH-19-HbR-min" were linked to lower predicted scores. This suggests that higher HbR concentration in the supramarginal gyrus or higher HbO concentration in the middle temporal gyrus corresponded to higher transfer scores.
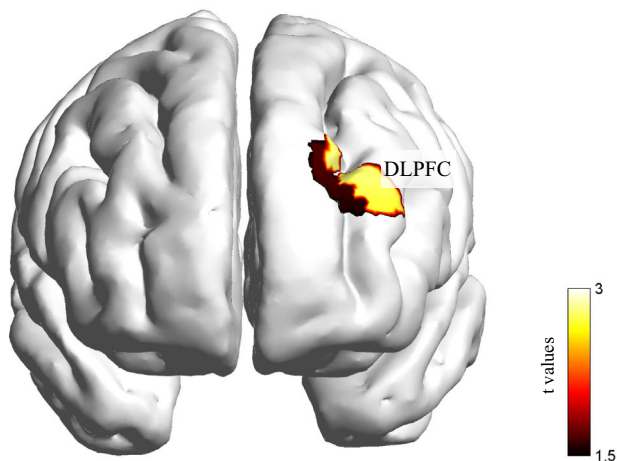
### fNIRS results across phases

**Comparison of the brain activation between phases.** For the MF group, the learning phase, compared to the rest phase, showed significantly greater activation in the DLPFC and MTG. Specifically, channel 7 (corrected $p = 0.0498$) and channel 19 (corrected $p = 0.022$) were significantly more active during learning than the rest. No significant differences were found between the assessment and rest phases across all channels. During the feedback phase, the DLPFC, FP, and MTG showed significantly greater activation compared to the rest, with channel 7 (corrected $p = 0.035$), channel 8 (corrected $p = 0.022$), and channel 19 (corrected $p = 0.039$) being more active.

For the AF group, channel 7 in the DLPFC showed significantly greater activity during the learning phase compared to the rest (corrected $p < 0.001$). The assessment and feedback phases also elicited significantly larger activation in the DLPFC relative to rest, with channel 7 again showing significant increases (corrected $p < 0.001$).



**Fig. 4 | Brain responses to neutral feedback.** The neutral feedback group showed significantly increased activation in dorsolateral prefrontal cortex (DLPFC).



**Fig. 5 | Identification of important features through SHAP in the metacognitive feedback group.** The left side of the ordinate lists the feature names, following the naming convention: channel - channel number - biomarker (HbO or HbR) - calculated feature value (average, maximum, minimum, peak, slope, or kurtosis). Darker red denotes higher activity values, while lighter blue represents lower activity values. In the MF group, key brain regions that predicted transfer scores included the frontopolar area (FP), supramarginal gyrus (SMG), dorsolateral prefrontal cortex (DLPFC), and fusiform gyrus (FG).
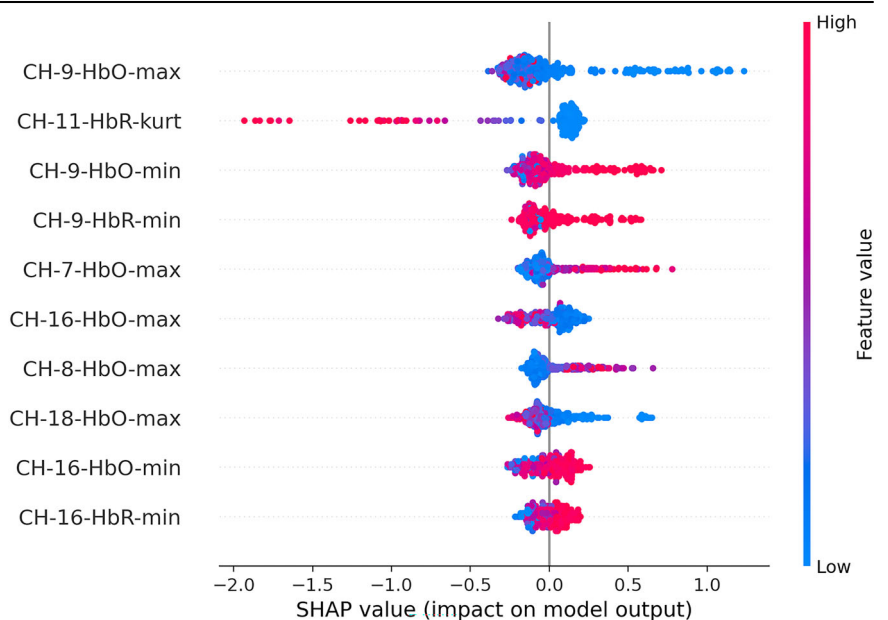
**Fig. 6 | Identification of important features through SHAP in the affective feedback group.** The left side of the ordinate lists the feature names, following and naming convention: channel—channel number - biomarker (HbO or HbR)—calculated feature value (average, maximum, minimum, peak, slope, and kurtosis). Darker red denotes higher activity values, while lighter blue represents lower activity values. In the AF group, key brain regions that predicted transfer scores included the supramarginal gyrus (SMG), dorsolateral prefrontal cortex (DLPFC), middle temporal gyrus (MTG), and frontopolar area (FP).
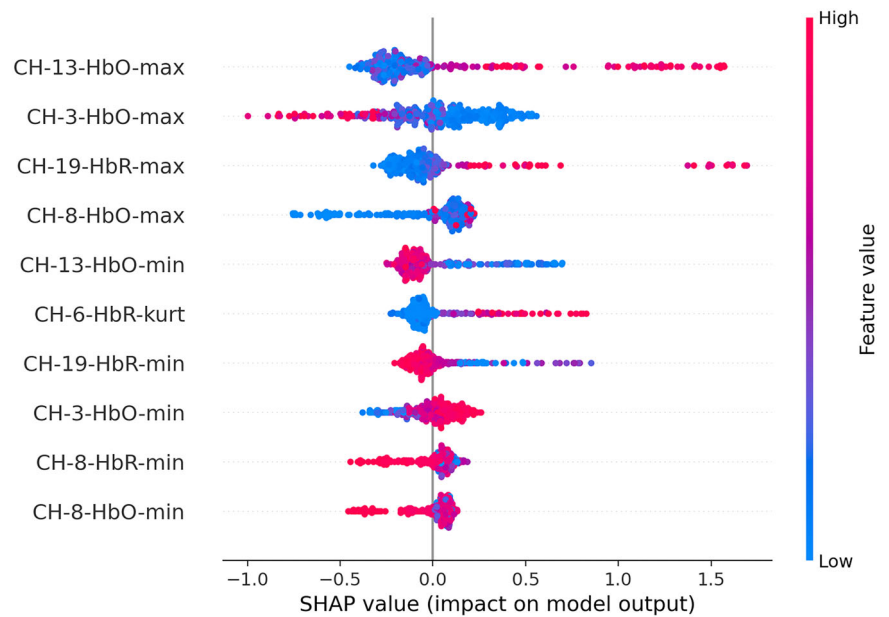


**Fig. 7 | Identification of important features through SHAP in the neutral feedback group.** The left side of the ordinate lists the feature names, following the naming convention: channel - channel number - biomarker (HbO or HbR) - calculated feature value (average, maximum, minimum, peak, slope, or kurtosis). Darker red denotes higher activity values, while lighter blue represents lower activity values. In the NF group, key brain regions that predicted transfer scores included the supramarginal gyrus (SMG), middle temporal gyrus (MTG), frontopolar area (FP), dorsolateral prefrontal cortex (DLPFC).
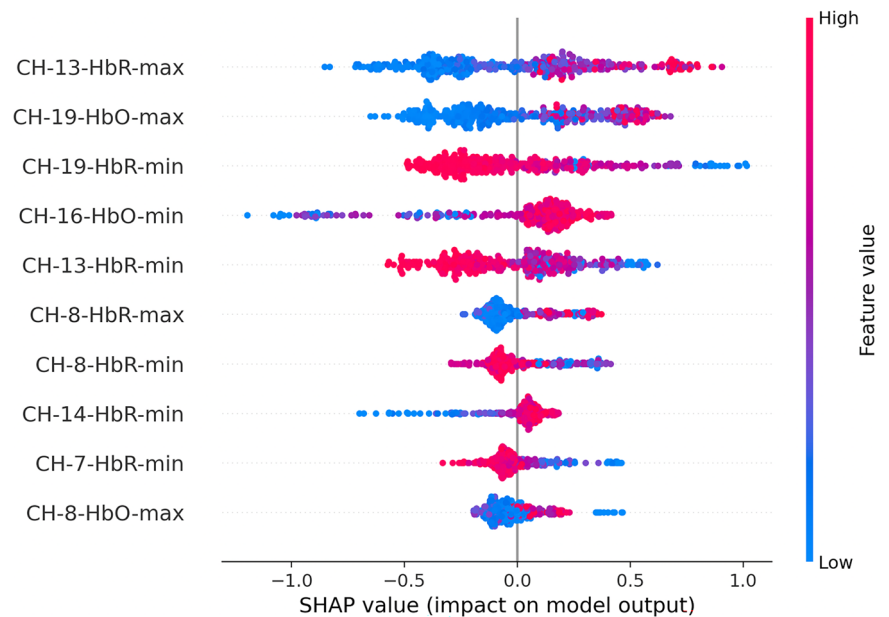


For the NF group, the learning phase also demonstrated higher activation compared to the rest, particularly in the DLPFC at channel 1 (*corrected* $p = 0.0498$). During the assessment phase, channel 1 (*corrected* $p = 0.031$) and channel 4 (*corrected* $p = 0.022$) in the DLPFC were more active than in the rest phase. Feedback compared to rest elicited significantly larger activation in the DLPFC at channel 1 (*corrected* $p = 0.035$).

**Neural-behavioral relationships.** Pearson correlation analysis indicated that channel 19 activation during the learning phase was positively correlated with retention scores in both the MF group ($r = 0.45$, $p = 0.015$) and the NF group ($r = 0.45$, $p = 0.013$), but not in the AF group. In the MF group, channel 8 activation during the feedback phase was positively correlated with retention score ($r = 0.44$, $p = 0.017$), an association not seen in the other groups. Furthermore, channel 8 activation during feedback was positively correlated with channel 19 activation

during learning. For the NF group, channel 1 activation during the assessment phase was negatively correlated with retention scores ($r = -0.379$, $p = 0.043$), an effect not observed in the other groups.

## Discussion
Feedback play an important role in the human-chatbot interaction. This study aims to elucidate the neurocognitive processes underpinning different feedback types (mainly metacognitive and affective feedback) during human-chatbot interactions and provide a neurophysiological explanation for why the certain feedback is useful. This study yielded three key findings concerning the processes involved in human-chatbot interactions: (1) it identified differences in learning outcomes based on the type of feedback provided, (2) it revealed the neural and psychological processes associated with different feedback types, and (3) it pinpointed key brain regions that predicted transfer scores across the various feedback groups.

Behavioral findings indicated that students who received metacognitive or affective feedback achieved better retention scores compared to those who received neutral feedback. Feedback that informs learners about their actions helps them reevaluate their abilities and to adapt their learning strategies, consequently increasing their chances of success[78]. As expected, metacognitive feedback resulted in better transfer scores, suggesting that it may encourage learners to engage more actively in the learning process, monitor their progress, and assess their understanding. This sense of autonomy and the ability to monitor their own learning enhances learners' capacity to grasp and apply the knowledge effectively. In biology education, students with higher self-awareness and stronger skills in monitoring, regulating, and controlling their learning process are more likely to develop a meaningful understanding of key biology concepts[79,80]. In addition, students in the MF group exhibited higher metacognitive sensitivity than the other groups. Metacognitive feedback enhances students' self-awareness and reflective thinking, helping them to reduce biases by guiding them to monitor and reflect on the learning process[81]. This enables students to quickly identify their mistakes and make corrective action more effectively. In contrast, affective and neutral feedback did not yield similar results, as they lacked the metacognitive cues needed for such reflection.

In terms of brain activation pattern, all groups showed activation in the dorsolateral prefrontal cortex (DLPFC), a region critical for working memory and attention[82–84]. Given the nature of the learning task, which involves conceptual knowledge requiring continuous memorization and comprehension, the DLPFC activation is likely due to these task demands. Furthermore, the MF group additionally activated the frontopolar area, supramarginal gyrus, and middle temporal gyrus. The frontopolar area is associated with metacognition[85,86], the supramarginal gyrus is related to theory of mind[87,88] and emotion recognition[89], and the middle temporal gyrus to semantic processing[90,91]. The greater activation of the frontopolar area and middle temporal gyrus in the MF group suggests that metacognitive feedback may foster deeper semantic processing and enhance understanding of complex concepts, which aligns with better learning outcomes. In contrast, the AF group showed activation in the supramarginal gyrus and superior temporal gyrus, with the former region linked to theory of mind[87,88] and emotion recognition[89], and the latter to social interaction and perception[43–45]. This pattern indicates that students in the AF group likely focused more on processing emotional information. However, due to equipment limitations (i.e., limited number of recording channels), it remains unclear whether these activations reflect emotional experiences or regulation, which could involve regions such as the anterior cingulate cortex and orbital cortex[92,93]. Future studies employing MRI could help address these questions. By contrast, we found that NF elicited greater DLPFC activation compared to the other feedback types, suggesting that students in the NF group may have been less influenced by external feedback or emotional states. While this finding was unexpected, it highlights the complex nature of brain responses to different feedback types. NF did not elicit the same activation patterns in other brain regions as seen in the MF or AF groups, which may suggest a different cognitive processing strategy, such as increased focus on task-related demands rather than emotional or metacognitive reflection.

In addition, in the MF group, higher metacognitive sensitivity was linked to higher activity in the frontopolar area, a relationship not observed in the other groups. This finding aligns with research suggesting that the frontopolar cortex plays a critical role in reflecting on what is known versus unknown and in monitoring and regulating cognitive processes. Our results further substantiate the role of the frontopolar cortex in metacognition, particularly in the context of explicit metacognitive judgments[63,94]. While previous studies have linked dorsolateral and medial PFC activation with self-awareness and metacognitive accuracy[95–97], our findings emphasize the frontopolar cortex in this process. This suggests that metacognitive feedback encourages students to reflect on, monitor, and regulate their learning, enhancing their ability to judge what they know and do not know, and thereby promoting metacognitive sensitivity. Furthermore, activation of the middle temporal gyrus in the MF group was significantly correlated with

retention scores, a relationship not observed in the other two groups. Although retention scores did not differ significantly between the MF and AF groups, the MF group showed higher mean scores, indicating a potential advantage for metacognitive feedback in promoting knowledge retention. Based on the association between brain activity and behavior, the effectiveness of metacognitive feedback may stem from its ability to enhance students' accuracy in judging their learning performance. This improvement allows students to adjust their learning strategies in real time, promoting better learning outcomes. As students continue to reflect, monitor, and regulate their learning process, their performance judgments become increasingly precise. Additionally, metacognitive feedback may encourage deeper processing of knowledge, which can help consolidate memory.

Furthermore, we observed that MTG activity was more intense during concept learning compared to resting in the MF group, and MTG activity during the learning phase was positively correlated with retention scores in both the MF and NF groups. This suggests that students in the MF group were not engaging in rote memorization but were instead working to understand the relationships among concepts, contributing to improved retention. Furthermore, FP activity was heightened when receiving metacognitive feedback than rest. Results showed that brain activation in the FP was specifically associated with the metacognitive feedback phase, as observed in the MF group, whereas neither the learning nor self-assessment significantly enhanced FP activity. Furthermore, FP activity during the feedback phase was positively correlated with both retention scores and MTG activity during the learning phase in the MF group. This suggests that students enhanced their comprehension monitoring while receiving metacognitive feedback and subsequently adjusted their learning strategies during the learning phase, thereby reinforcing memory. These findings imply that metacognitive feedback not only affects the feedback phase, but also influences the entire learning process.

In this study, we focused on learning transfer, a core competency in modern education. In today's world, students are expected not only to memorize information but to apply and transfer knowledge across different contexts[98]. Transfer is considered an essential skill for success in the 21st century, as it demonstrates a student's ability to adapt and apply learned knowledge beyond immediate tasks[99]. Understanding the neural mechanisms behind transfer can inform how different types of feedback can be tailored to support this critical learning outcome. Although our initial correlation analyses did not reveal significant relationships between brain activity and transfer scores, our machine learning models demonstrated that brain activity features (e.g., average, maximum, minimum, peak, slope, and kurtosis) effectively predicted transfer scores across different feedback groups. These results highlight the complementary role of machine learning as a multivariate approach, enabling the identification of complex patterns in brain data that traditional analysis methods, such as general linear models, may not capture. First, key brain regions identified as predictors varied by feedback group and included the frontopolar area (associated with metacognition), the supramarginal gyrus area (related to theory of mind and emotion recognition), and the dorsolateral prefrontal (associated with memory and attention). Second, group-specific predictors were also observed. In the MF group, the fusiform gyrus known for its role in reading efficiency, emerged as a significant predictor. For the AF and NF groups, the middle temporal gyrus, associated with semantic processing, was highlighted as a key region. These findings of prediction models highlight distinct brain regions of interest associated with different types of feedback. Specifically, for the MF group, the key brain regions identified through prediction were consistent with those revealed by the activation analysis. Notably, while the frontopolar area was not significantly activated in the other two groups, its activity still emerged as a reliable predictor of transfer scores across all feedback types. This finding suggests that the frontopolar area plays a critical role in knowledge transfer regardless of the feedback type. In the future, monitoring the activity of these key brain regions in real time could enable adaptive feedback strategies to optimize learning and facilitate knowledge transfer more effectively. In addition, the association between neural activity and transfer performance suggests that

metacognitive feedback may enhance transfer performance by facilitating the integration of cognition, emotion, and metacognition.

In the current study, several limitations deserve noting. First, some external factors that may affect the effectiveness of the feedback types, such as students' motivation or learning style, were not included. Future studies should carefully address participant selection in this regard. Second, the feedback provided was not personalized, which could influence the effectiveness of different feedback types. Notwithstanding, we standardized the delivery of feedback to control for potential variations: each feedback type was given at the same frequency (15 times), consistently delivered after the self-assessment phase, and contained a uniform number of Chinese characters. Third, due to the limited number of channels and detection depth of fNIRS, our optode probe set only covered the frontal cortex and right temporoparietal regions, leaving other regions, particularly deeper brain areas, unexplored. Future studies should consider using whole-brain coverage or integrating other imaging technologies, such as fMRI, to consolidate these findings. Finally, the limited sample size and the gender composition of students can lead to bias, thereby affecting the generalizability of our conclusions. Future studies should include larger and more diverse samples to enhance the applicability of these findings.

Furthermore, this study focuses solely on comparing brain activity elicited by different types of feedback during human-chatbot interactions, without examining the feedback effects in human-human interactions. Previous research on human-human interactions has shown that cognitive feedback (e.g., yes-no verification or correct answers) engages frontoparietal brain regions, including the anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex (DLPFC), and parietal lobules[47,49,100]. The ACC is involved in error detection and expectation violation[48], while the DLPFC and superior parietal lobule support more complex processes such as error correction and performance adjustment[47,49,50]. In our study, metacognitive feedback during human-chatbot interaction also activated the DLPFC, consistent with findings from human-human feedback studies. However, unlike human-human interactions, metacognitive feedback in this context did not activate the superior parietal lobule but instead engaged the frontopolar area. This suggests that metacognitive feedback in human-computer interactions can, under certain conditions, substitute for teacher-provided feedback, particularly in promoting cognitive control. Future research should compare the effects of metacognitive and affective feedback on brain activity across human-human and human-chatbot interactions. Such comparisons could inform the design of human-machine interactions better suited for real-world educational applications.

This study advances our understanding of the effectiveness and neural basis of different types of feedback in human-chatbot interactions. In the chatbot-based learning environments, educators can tailor feedback to achieve various instructional goals. For example, metacognitive feedback can be developed to promote learning outcomes and enhance metacognitive sensitivity. Educators should consider integrating metacognitive feedback into human-chatbot interaction to guide students in continuously monitoring and reflecting on their learning processes, thereby promoting metacognitive sensitivity. In addition, metacognitive feedback may facilitate deeper semantic understanding during the learning phase, aiding in the internalization and retention of material. Furthermore, our results suggest that affective feedback may increase emotional processing, which in turn could promote positive emotions and memory retention. Positive emotions are essential components of learning motivation[101] and are closely related to academic performance, influencing subsequent learning behaviors[102]. Previous studies have consistently demonstrated that teacher-provided positive feedback, such as praise and encouragement, fosters positive emotions[102]. However, it remains unclear whether chatbot-based affective feedback elicits similar emotional benefits. Additionally, emotional response is not obvious in the single stages of the learning module, future research could enhance the anthropomorphic qualities of chatbots and diversify the forms of affective feedback to reduce social distance in human-computer interactions, thereby increasing the perceived authenticity of emotional responses. Finally, real-time monitoring of students' brain activity using functional neuroimaging

technologies during human-chatbot interactions could provide insights for adaptive feedback strategies, especially for promoting advanced skills like knowledge transfer. Machine learning analysis has shown that knowledge transfer is influenced by cognition, emotion, and metacognition. Future studies could focus on brain regions such as the frontopolar area, supramarginal gyrus area, and dorsolateral prefrontal cortex to predict transfer scores dynamically and deliver tailored feedback to optimize learning outcomes. In sum, this study underscores the importance of integrating diverse feedback types, including metacognitive and emotional feedback, into chatbot systems to foster deeper learning and improve student outcomes. While we acknowledge that translating neuroscience into educational practice is still a long-term goal, we believe these fNIRS findings are valuable for informing future research directions.

## Methods

### Participants

Ninety-three college students (63 females) with no history of neurological disorders were recruited from a university in Shanghai, China. All participants had normal or corrected-to-normal vision and were right-handed. None of them majored in biology science. Six participants were excluded from the data analysis because of the unsuccessful recording of their fNIRS data and missing screen recording files. Consequently, the behavioral and fNIRS results were based on the data from 87 participants. Each group consisted of 29 students. Our experimental procedures were conducted according to the Declaration of Helsinki and the research protocol was approved by the University Committee on Human Research Protection, East China Normal University (HR2-0068-2023). Written informed consent was obtained from all participants prior to the study.

### Materials

The learning materials focused on the human cardiovascular system, specifically the functions and structures of the human heart and blood, which were adapted from previous studies[103,104]. The content covered 15 key concepts, comprising approximately 1846 Chinese characters. To create a more engaging and realistic interactive experience, we adopted personalization design principles. First, the learning materials were presented using a combination of images, texts, and interactive buttons to capture students' interest. Second, the content was adapted to a conversational style[105], including the following changes: (1) the word "the" was replaced with "your" in each statement. For example, "Your blood flows through the veins throughout your body"; (2) Interactive leading sentences were added before introducing each concept, such as "Now, let's look at the structure of blood"; (3) The students' names were used to create a more personalized experience, for example, "Yin, do you understand heart tissue now?".

### Experimental procedure

**Pre-test**. All students first completed a knowledge test to determine their initial understanding of the material. The test included 16 multiple-choice questions, adapted from a previous study[104], with two questions modified to increase difficulty. Each question offered four options, with only one correct answer. Students received one point for each correct response.

**Human-chatbot interaction**. In the interaction session, the chatbot began by greeting the students, introducing itself, and asking for their names. A screenshot of this greeting phase is shown in Fig. 8. Then, students studied 15 learning modules. Each module, or trial, consists of three phases: learning, self-assessment, and feedback. At the end of each trial, students entered the rest phase and took a break. The flow for each trial is as follows: (i) *Learning phase*. The chatbot presented concepts in dialog boxes, allowing students to learn at their own pace. Once they have completed a concept, they could click the "go on learning" button to proceed. (ii) *Self-assessment phase*. After studying each concept, students assessed their understanding by selecting either the "understand" or "don't understand" button. (iii) *Feedback phase*. Based on the group to which students are assigned, the chatbot provides feedback tailored to
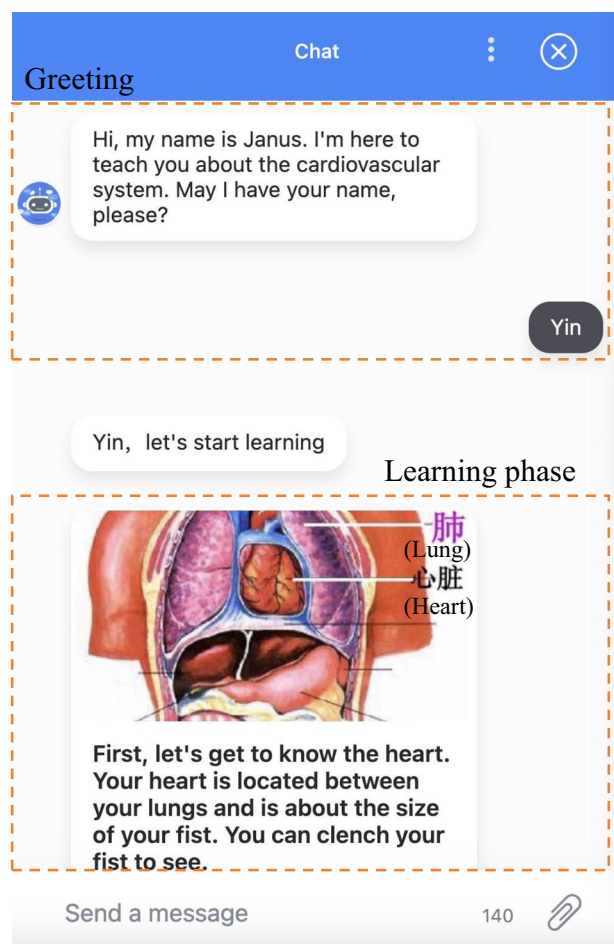
**Fig. 8 | Screenshot of the greeting phase.** The chatbot introduced itself and requested the student's name. Once the student enters their name, the chatbot remembers it and begins the learning phase.

their self-assessment response. Specifically, in the MF group, the chatbot would first present a metacognitive question. The questions were adapted from previous studies[106–108]. If students pressed the "understand" button, the chatbot would ask questions to instruct students to evaluate, reflect, monitor, or regulate the learning process and plan learning strategies. The questions adopted were as follows: (1) are you satisfied with your current learning performance? (2) are you confident in mastering the above content? (3) can you explain the above concepts to others? (4) how do you better complete subsequent learning? Conversely, if students pressed the "don't understand" button, the chatbot would ask some questions to prompt students to monitor the learning process from knowledge understanding, effort, and concentration. Additionally, it would guide students in evaluating and reflecting on their learning progress to identify any comprehension challenges promptly. The questions adopted were as follows: (1) Are there any concepts you don't understand? (2) Are you making an effort to understand the content? (3) How focused are you on learning? (4) What is the most challenging part? Each learning module was paired with a metacognitive question. To prevent fatigue from repeated questioning, the four metacognitive questions are presented in rotation, with each question appearing up to 4 times. Afterward, students paused for 10 seconds and without input. Then, the chatbot provided a sentence about the students' possible metacognitive performance (e.g., I am satisfied with my current learning performance), and students responded on a five-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). This step further encouraged students to engage thoughtfully with the metacognitive questions.

In contrast to metacognitive feedback, the chatbots in the affective feedback group provided students with encouragement and praise. For example, when students pressed the "understand" button, the chatbot would offer positive reinforcement with statements such as: (1) "Great! I believe you will do better in the future learning." (2) "Great! It looks like you've grasped this part of the content well." (3) "That's great. You did a great job in the learning process." (4) "You're doing great! Keep up the good work with the upcoming learning materials." On the other hand, when students pressed the "don't understand" button, the chatbot would provide motivational encouragement with statements like: (1) "Don't be discouraged; it's normal to have difficulty understanding when you're just starting to learn." (2) "Don't worry. It's not easy for others to understand, either." (3) "Don't worry. This is a challenging concept for others as well." (4) "Don't be discouraged; you can try to understand this part of the content again." After receiving the feedback, students paused for 10 seconds without input, and then the chatbots prompted students to rate their emotional response (e.g., I am in a happy mood right now) on a five-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree), based on how well they felt the feedback matched their emotional state. In the NF group, regardless of the students' self-assessment, the chatbot simply instructed, "Let's take a break first and then continue with our learning." After a 10-second break, students were then asked to randomly click one of the numbers between 1 and 5. The screenshots of the feedback phase for all three groups are shown in Fig. 9.

**Post-test**. Students completed the retention and transfer tests, which took approximately 10 min. The retention test was identical to the initial knowledge test, but the questions were presented in a different order. The transfer test consisted of four questions designed to assess how well students applied the basic concepts and principles they had just learned to solve novel problems. Students' textual responses were scored by comparing them to standard answers, with a maximum total score of 15 points for the four questions. After each multiple-choice question, students were asked to rate their confidence in their answers. This confidence rating was used to measure students' metacognitive sensitivity by calculating the alignment between their accuracy and their self-assessed confidence level[21]. Metacognitive sensitivity reflects the ability to introspect and accurately assess one's own knowledge and performance.

### fNIRS data acquisition and preprocessing
A portable continuous-wave near-infrared optical imaging system was used (LIGHTNIRS, Shimadzu, Kyoto, Japan), with a sampling rate of 13 Hz and wavelengths of 730 and 850 nm. The system included eight sources and eight detectors, generating 20 different measurement channels. The distance between a source and a detector is 3 cm. The probes were strategically placed to cover each student's prefrontal cortex (see Fig. 10a) and right temporoparietal areas (see Fig. 10b). The center of the probe matrix was placed at the FPZ position according to the international 10-20 system. A 3D digitizer was used to determine spatial registration of fNIRS channels to MNI space[109]. Detailed information about coordinates and anatomical probabilities is provided in the supplementary material (see Supplementary Table 1).

The raw fNIRS data were processed using Matlab 2013b software. First, we registered the MNI coordinates using the NIRS_SPM toolbox. Second, we applied a wavelet-based method to remove global physiological noise[110]. Third, we used a general linear model (GLM) to analyze the pre-processed HbO data. Due to variations in the learning pace among students, their "onset" and "duration" differed. These values were manually extracted from the screen recording files for each student. In addition, the data were processed with a low-pass filter based on hemodynamic response function (HRF) and wavelet-MDL detrending method. Finally, we estimated Beta values for each of the twenty channels for each student in each condition, which reflected the degree of channel activation. The beta values were then exported to SPSS software for further analysis.
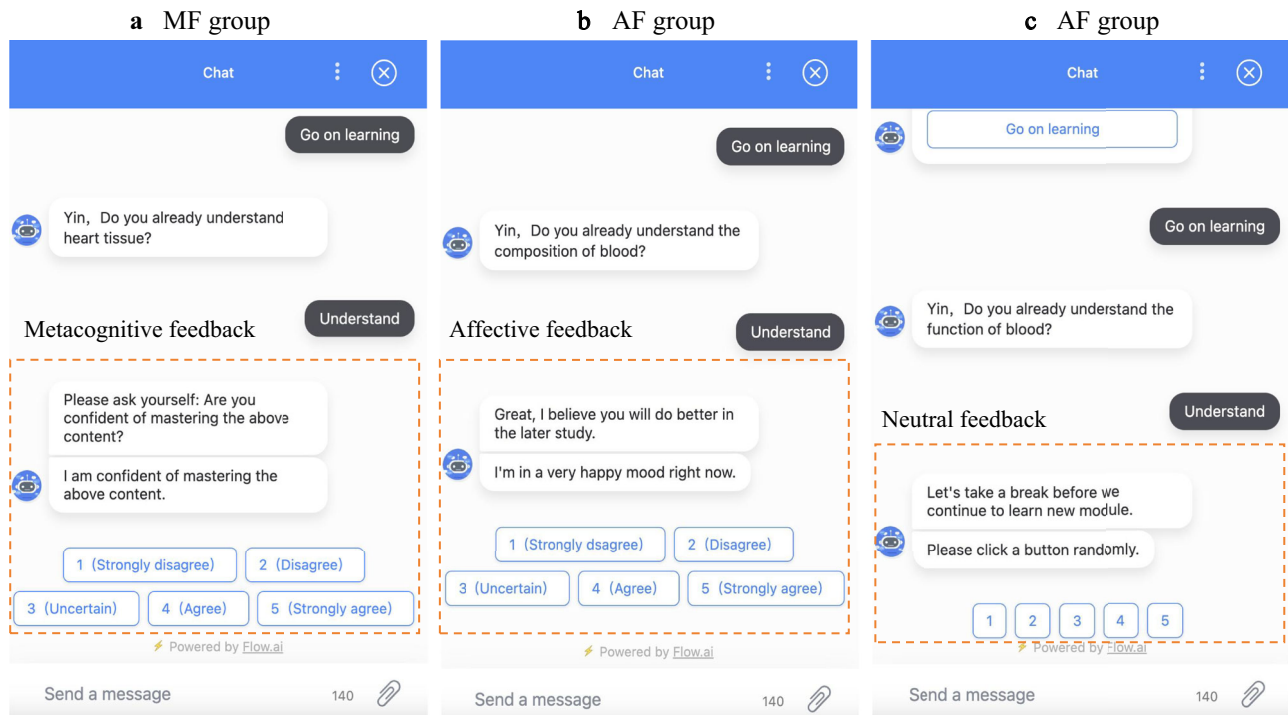
**Fig. 9 | Screenshots of the feedback phase for all three groups.** Following students' self-assessments, each group received different feedback content. **a** The metacognitive feedback group (MF group) received a metacognitive question, and then rated possible metacognitive performance. **b** The affective feedback group (AF group) received encouragement or praise, and then rated emotional response. **c** The neutral feedback group (NF group) received a break prompt, and then click one button randomly.



**Fig. 10 | Layout of optode emitters (red), detectors (blue), and channels (white). a** The left panel represents the prefrontal region, including the dorsolateral prefrontal cortex (CH 1, 2, 3, 4, 5, 6, 7), and frontopolar area (CH 8, 9, 10). **b** The right panel represents the right temporoparietal areas, including the supramarginal gyrus (CH11, 12, 13, 14), primary somatosensory cortex (CH16), superior temporal gyrus (CH 15, 17), fusiform gyrus (CH 18), middle temporal gyrus (CH19), and inferior temporal gyrus (CH20).
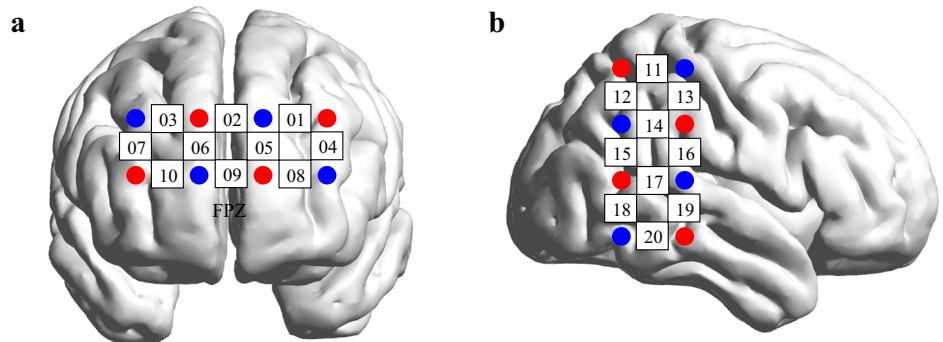
## Data analysis

For the behavioral data, we first conducted ANCOVAs to compare retention and transfer scores across the three groups. Next, we used a type 2 ROC function in MATLAB software to calculate the metacognitive sensitivity of the students[21].

For the fNIRS data, we first examined brain activation patterns associated with the entire learning module. In the analysis, "onset" refers to the time point corresponding to the start of the learning phase, while "duration" refers to the time from the beginning of the learning phase to the end of the feedback phase. We conducted a one-sample t-test on the beta values of all students in each group to investigate task-related brain activation regions resulting from different feedback types. We then performed an ANOVA to investigate group differences in brain activation.

To gain further insights into brain-behavior relationships during learning, assessment, and feedback, we then compared brain activation across phases using the rest phase as a baseline. This phase comparison analysis allowed us to examine how activation levels varied between different stages of the learning process and how these variations related to

learning outcomes. By focusing on the timing and role of brain activity during each phase, we were able to identify which phases might have a more significant impact on learning outcomes. Each learning module was further divided into three phases (i.e., learning, self-assessment, and feedback). The GLM method was used to compute beta values for each of the 20 channels in each phase (learning, self-assessment, feedback, and rest). In this context, "onset" refers to the time point marking the start of each phase, and "duration" refers to the length of each phase. A paired t-test on the beta values was conducted to assess the mean differences in the levels of activity for each channel across phases. The Benjamini-Hochberg procedure was applied to control for false discovery rate (FDR), with corrected significance levels presented in the results section.

To investigate the relationship between enhanced brain activity and behavior outcomes, Pearson correlations were calculated between neural activity and scores for retention, transfer, and metacognitive sensitivity.

Furthermore, to keep important signal features and more accurately capture task-related changes in brain activity, we developed different machine learning-based transfer score prediction models for the three

groups using python programming language in PyCharm. First, we performed data preprocessing, which included data augmentation, features extraction, and dataset generation. Each student underwent 15 trials (learning, assessment, feedback). For data augmentation, we divided each student's fNIRS data into 15 independent samples based on the total trial duration or the duration of individual phases[111]. In each sample, we extracted six statistical features (e.g., average, maximum, minimum, peak, slope, and kurtosis) from two biomarkers: oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) concentrations in each channel[38]. These features can comprehensively capture individual signal characteristics and have been applied in various predictive model studies[112]. Specifically, *Average* reflects the overall blood oxygen level throughout the task, indicating changes in brain activation. *Minimum and Maximum* reflect the fluctuation range of blood oxygen levels, representing the highest and lowest active states at different time points, and may be associated with peak or trough cognitive performance. *Slope* indicates the rate of change in blood oxygen concentration, revealing the brain's response speed and trend, which reflects the timeliness of brain responses to cognitive demands. *Kurtosis* describes the sharpness of the signal distribution, helping identify abnormal brain activity at certain moments and uncovering more precise brain functional responses. Then, we generated four datasets for each group: one for the entire learning module, and separate datasets for the learning phase, assessment phase, and feedback phase. Using 5-fold cross-validation, we split the data into training and testing datasets for each of these four sets.

Finally, each dataset comprises 435 (i.e., 29 participants, with 15 trials or phases per participant: $29 \times 15 = 435$) and 240 features per sample (i.e., 20 channels with six features each for HbO and HbR signals). We selected several machine learning models for comparison, including Random Forest, XGBoost, Gradient Boosting, Extra Trees, Adaboost Regressor, and Bagging Regressor. To further optimize model parameters and evaluate generalization performance, we used 5-fold cross-validation, which minimizes the risk of overfitting to the training set. Model effectiveness was assessed using $R$-squared ($R^2$), Mean Absolute Error (MAE), and Mean Squared Error (MSE). $R^2$ evaluates how much variance in the dependent variable is explained by the independent variables. MAE represents the average absolute difference between predicted and actual values, and MSE calculates the average of the squared differences between predicted and actual values.

Then, we selected the best model and employed the SHAP model to identify the most important predictors of transfer scores[113] for each group. The results are visualized using summary plots, which list the features according to their significance, where the darker color red denotes higher activity value, and a lighter color blue denotes lower activity values.

## Data availability
The data supporting this study's findings are available from the corresponding author on reasonable request.

## Code availability
The data were performed in MATLAB with standard functions and toolboxes and PyCharm. Custom codes are available from the corresponding author upon reasonable request.

## References

1. Hwang, G. J. & Chang, C. Y. A review of opportunities and challenges of chatbots in education. *Interact. Learn. Environ.* **31**, 4099–4112 (2023).
2. Fidan, M. & Gencel, N. Supporting the instructional videos with chatbot and peer feedback mechanisms in online learning: The effects on learning performance and intrinsic motivation. *J. Educ. Comput. Res.* **60**, 1716–1741 (2022).
3. Kuhail, M. A., Alturki, N., Alramlawi, S. & Alhejori, K. Interacting with educational chatbots: A systematic review. *Educ. Inf. Technol.* **28**, 973–1018 (2023).
4. Timmers, C. F., Walraven, A. & Veldkamp, B. P. The effect of regulation feedback in a computer-based formative assessment on information problem solving. *Comput. Educ.* **87**, 1–9 (2015).
5. Hattie, J. & Timperley, H. The power of feedback. *Rev. Educ. Res.* **77**, 81–112 (2007).
6. Narciss, S. et al. Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Comput. Educ.* **71**, 56–76 (2014).
7. Chou, C. Y. & Zou, N. B. An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models. *Int. J. Educ. Technol. High. Educ.* **17**, 1–27 (2020).
8. Villegas-Ch, W., Arias-Navarrete, A. & Palacios-Pacheco, X. Proposal of an architecture for the integration of a chatbot with artificial intelligence in a smart campus for the improvement of learning. *Sustainability* **12**, 1500 (2020).
9. Jeon, J. Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Comput. Assist. Lang. Learn.* **37**, 1–26 (2024).
10. Li, Y. et al. Using Chatbots to teach languages. In: *Proceedings of the Ninth ACM Conference on Learning, New York, NY, USA* General Chair: René F. Kizilcec 451–455 (Association for Computing Machinery, 2022).
11. Yin, J., Goh, T.-T. & Hu, Y. Using a Chatbot to provide formative feedback: a longitudinal study of intrinsic motivation, cognitive load, and learning performance. *IEEE Trans. Learn. Technol.* **17**, 1404–1415 (2024).
12. Karaoglan Yilmaz, F. G. & Keser, H. The impact of reflective thinking activities in e-learning: A critical review of the empirical research. *Comput. Educ.* **95**, 163–173 (2016).
13. Yilmaz, R. & Keser, H. The impact of interactive environment and metacognitive support on academic achievement and transactional distance in online learning. *J. Educ. Comput. Res.* **55**, 95–122 (2017).
14. Zhu, Y. et al. Instructor–learner neural synchronization during elaborated feedback predicts learning transfer. *J. Educ. Psychol.* **114**, 1427 (2022).
15. Thurlings, M., Vermeulen, M., Bastiaens, T. & Stijnen, S. Understanding feedback: A learning theory perspective. *Educ. Res. Rev.* **9**, 1–15 (2013).
16. Roll, I., Aleven, V., McLaren, B. M. & Koedinger, K. R. (2007). Designing for metacognition e applying cognitive tutor principles to the tutoring of help seeking. *Metacogn. Learn.* **2**, 125–140 (2007).
17. Rickey, D. & Stacy, A. M. The role of metacognition in learning chemistry. *J. Chem. Educ.* **77**, 915 (2000).
18. Teng, M. F., Qin, C. & Wang, C. Validation of metacognitive academic writing strategies and the predictive effects on academic writing performance in a foreign language context. *Metacogn. Learn.* **17**, 167–190 (2022).
19. Yeung, N. & Summerfield, C. Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. B: Biol. Sci.* **367**, 1310–1321 (2012).
20. Maniscalco, B. & Lau, H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**, 422–430 (2012).
21. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).
22. Ackerman, R. & Thompson, V. A. Meta-reasoning: Monitoring and control of thinking and reasoning. *TRENDS Cogn. Sci.* **21**, 607–617 (2017).
23. Ackan, S. & Tatar, S. An investigation of the nature of feedback given to pre-service English teachers during their practice teaching experience. *Teach. Dev.* **14**, 153–172 (2010).
24. Hyland, F. Providing effective support: Investigating feedback to distance language learners. *Open Learn.* **16**, 233–247 (2001).
25. Kluger, A. N. & DeNisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* **119**, 254–284 (1996).

26. Brinko, K. T. The practice of giving feedback to improve teaching: what is effective?. *J. High. Educ.* **64**, 574–593 (1993).

27. Ayedoun, E., Hayashi, Y. & Seta, K. Communication strategies and affective backchannels for conversational agents to enhance learners' willingness to communicate in a second language. In: *Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China* (eds André, E., Baker, R., Hu, X., Rodrigo, M. M. T. & du Boulay, B) 459–462 (2017).

28. Belland, B., Kim, C. M. & Hannafin, M. A Framework for designing scaffolds that improve motivation and cognition. *Educ. Psychol.* **48**, 243–270 (2013).

29. Moridis, C. N. & Economides, A. A. Affective learning: empathetic agents with emotional facial and tone of voice expressions. *IEEE Trans. Affect. Comput.* **3**, 260–272 (2012).

30. Wigfield, A., Klauda, S. L. & Cambria, J. in *Handbook of Self-regulation of Learning and Performance* (eds Zimmerman, B. J. & Schunk, D. H.) (33–48) (Routledge, 2011) https://doi.org/10.4324/9780203839010.ch3.

31. Yin, J., Goh, T. T. & Hu, Y. Interactions with educational chatbots: the impact of induced emotions and students' learning motivation. *Int. J. Educ. Technol. High. Educ.* **21**, 47 (2024).

32. Rajendran, R., Iyer, S. & Murthy, S. Personalized affective feedback to address students' frustration in ITS. *IEEE Trans. Learn. Technol.* **12**, 87–97 (2019).

33. Liu, C. C, Liao, M. G., Chang, C. H. & Lin, H. M. An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Comput. Educ.* **189**, 104576 (2022).

34. Arguedas, M., Daradoumis, T. & Xhafa, F. Analyzing how emotion awareness influences students' motivation, engagement, self-regulation and learning outcome. *Educ. Technol. Soc.* **19**, 87–103 (2016).

35. Villringer, A., Planck, J., Hock, C., Schleinkofer, L. & Dirnagl, U. Near infrared spectroscopy (nirs): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* **154**, 101–104 (1993).

36. Lei, X. & Rau, P. L. P. Emotional responses to performance feedback in an educational game during cooperation and competition with a robot: Evidence from fNIRS. *Comput. Hum. Behav.* **138**, 107496 (2023).

37. Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A. & Boas, D. A. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* **29**, 368–382 (2006).

38. Naseer, N. & Hong, K.-S. fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* **9**, 3 (2015).

39. Rajmohan, V. & Mohandas, E. Mirror neuron system. *Indian J. Psychiatry* **49**, 66–69 (2007).

40. Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192 (2004).

41. van Gog, T., Paas, F., Marcus, N., Ayres, P. & Sweller, J. The mirror neuron system and observational learning: implications for the effectiveness of dynamic visualizations. *Educ. Psychol. Rev.* **21**, 21–30 (2009).

42. Iacoboni, M. & Dapretto, M. The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* **7**, 942–951 (2006).

43. Allison, T., Puce, A. & McCarthy, G. Social perception from visual cues: Role of the STS region. *TRENDS Cogn. Sci.* **4**, 267–278 (2000).

44. Bigler, E. D. et al. Superior temporal gyrus, language function, and autism. *Dev. Neuropsychol.* **31**, 217–238 (2007).

45. Redcay, E. et al. Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage* **50**, 1639–1647 (2010).

46. Cavanagh, J. F., Figueroa, C. M., Cohen, M. X. & Frank, M. J. Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cereb. Cortex.* **22**, 2575–2586 (2011).

47. Crone, E. A., Zanolie, K., Van Leijenhorst, L., Westenberg, P. M. & Rombouts, S. A. Neural mechanisms supporting flexible performance adjustment during development. *Cogn., Cogn. Affect. Behav. Neurosci.* **8**, 165–177 (2008).

48. Luft, C. D. B., Nolte, G. & Bhattacharya, J. High-learners present larger mid-frontal theta power and connectivity in response to incorrect performance feedback. *J. Neurosci.* **33**, 2029–2038 (2013).

49. van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E. & Crone, E. A. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *J. Neurosci.* **28**, 9495–9503 (2008).

50. Zanolie, K., Van Leijenhorst, L., Rombouts, S. A. & Crone, E. A. Separable neural mechanisms contribute to feedback processing in a rule-learning task. *Neuropsychologia* **46**, 117–126 (2008).

51. Fleming, S. M. & Dolan, R. J. The neural basis of metacognitive ability. *Philos. Trans. R. Soc. B: Biol. Sci.* **367**, 1338–1349 (2012).

52. Vaccaro, A. G. & Fleming, S. M. Thinking about thinking: a coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain Neurosci. Adv.* **2**, 2398212818810591 (2018).

53. Fleming, S. M., Huijgen, J. & Dolan, R. J. Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.* **32**, 6117–6125 (2012).

54. Shimamura, A. P. Toward a cognitive neuroscience of metacognition. *Conscious. Cogn.* **9**, 313–323 (2000).

55. McCurdy, L. Y. et al. Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J. Neurosci.* **33**, 1897–1906 (2013).

56. Baird, B., Smallwood, J., Gorgolewski, K. J. & Margulies, D. S. Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *J. Neurosci.* **33**, 16657–16665 (2013).

57. Fleming, S. M., Ryu, J., Golfinos, J. G. & Blackmon, K. E. Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* **137**, 2811–2822 (2014).

58. Bellon, E., Fias, W., Ansari, D. & De Smedt, B. The neural basis of metacognitive monitoring during arithmetic in the developing brain. *Hum. Brain Mapp.* **41**, 4562–4573 (2020).

59. Takeuchi, N., Mori, T., Suzukamo, Y. & Izumi, S. I. Activity of prefrontal cortex in teachers and students during teaching of an insight problem. *Mind Brain Educ.* **13**, 167–175 (2019).

60. Carlen, M. "What constitutes the prefrontal cortex?". *Science* **358**, 478–482 (2017).

61. Pannu, J. K. & Kaszniak, A. W. Metamemory experiments in neurological populations: A review. *Neuropsychol. Rev.* **15**, 105–130 (2005).

62. Morales, J., Lau, H. & Fleming, S. M. Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* **38**, 3534–3546 (2018).

63. Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J. & Rees, G. Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541–1543 (2010).

64. Sabatinelli, D. et al. Emotional perception: meta-analyses of face and natural scene processing. *Neuroimage* **54**, 2524–2533 (2011).

65. Rauchbauer, B., Majdandžić, J., Hummer, A., Windischberger, C. & Lamm, C. Distinct neural processes are engaged in the modulation of mimicry by social group-membership and emotional expressions. *Cortex* **70**, 49–67 (2015).

66. Craig, A. D. How do you feel—now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* **10**, 59–70 (2009).

67. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).

68. Gunther Moor, B., van Leijenhorst, L., Rombouts, S. A. R. B., Crone, E. A. & Van der Molen, M. W. Do you like me? Neural correlates of social evaluation and developmental trajectories. *Soc. Neurosci.* **5**, 461–482 (2010).

69. Mayberg, H. S. et al. Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am. J. Psychiat.* **156**, 675–682 (1999).

70. Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T. & D'Esposito, M. Neural activity during social signal perception correlates with self-reported empathy. *Brain Res.* **1308**, 100–113 (2010).

71. Jackson, P. L., Rainville, P. & Decety, J. To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain* **125**, 5–9 (2006).

72. Rojiani, R., Zhang, X., Noah, A. & Hirsch, J. Communication of emotion via drumming: dual-brain imaging with functional near-infrared spectroscopy. *Soc. Cogn. Affect. Neurosci.* **13**, 1047–1057 (2018).

73. Fund, Z. Effects of communities of reflecting peers on student–teacher development—Including in-depth case studies. *Teach Teach* **16**, 679–701 (2010).

74. Kretlow, A. G. & Bartholomew, C. C. Using coaching to improve the fidelity of evidence based practices: a review of studies. *Teach. Educ. Spec. Educ.* **33**, 279–299 (2010).

75. Smutny, P. & Schreiberova, P. Chatbots for learning: A review of educational chatbots for the Facebook messenger. *Comput. Educ.* **151**, 103862 (2020).

76. Chang, C. Y., Hwang, G. J. & Gau, M. L. Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *Br. J. Educ. Technol.* **53**, 171–188 (2022).

77. Bibauw, S., François, T. & Desmet, P. Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Comput. Assist. Lang. Learn.* **32**, 827–877 (2019).

78. Zacharopoulos, G., Hertz, U., Kanai, R. & Bahrami, B. The effect of feedback valence and source on perception and metacognition: An fMRI investigation. *Cogn. Neurosci.* **13**, 38–46 (2022).

79. Eilam, B. & Reiter, S. Long-term self-regulation of biology learning using standard junior high school science curriculum. *Sci. Educ.* **98**, 705–737 (2014).

80. Martin, B. L., Mintzes, J. J. & Clavijo, I. E. Restructuring knowledge in biology: Cognitive processes and metacognitive reflections. *Int. J. Sci. Educ.* **22**, 303–323 (2000).

81. Simonovic, B., Vione, K., Stupple, E. & Doherty, A. It is not what you think it is how you think: a critical thinking intervention enhances argumentation, analytic thinking and metacognitive sensitivity. *Think. Skills Creat.* **49**, 101362 (2023).

82. Funahashi, S. Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited. *Front. Syst. Neurosci.* **9**, 1–14 (2015).

83. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).

84. Rossi, A. F., Bichot, N. P., Desimone, R. & Ungerleider, L. G. Top–down attentional deficits in macaques with lesions of lateral prefrontal cortex. *J. Neurosci.* **27**, 11306–11314 (2007).

85. Qiu, L. et al. The neural system of metacognition accompanying decision-making in the prefrontal cortex. *PLoS. Biol.* **16**, 1–27 (2018).

86. Miyamoto, K., Setsuie, R., Osada, T. & Miyashita, Y. Reversible silencing of the frontopolar cortex selectively impairs metacognitive judgment on non-experience in primates. *Neuron* **97**, 980–989 (2018).

87. Paul, S. et al. Autistic traits and individual brain differences: functional network efficiency reflects attentional and social impairments, structural nodal efficiencies index systemising and theory-of-mind skills. *Mol. Autism* **12**, 1–18 (2021).

88. Schurz, M., Tholen, M. G., Perner, J., Mars, R. B. & Sallet, J. Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: a review using probabilistic atlases from different imaging modalities. *Hum. Brain Mapp.* **38**, 4788–4805 (2017).

89. Wada, S. et al. Volume of the right supramarginal gyrus is associated with a maintenance of emotion recognition ability. *PLoS ONE* **16**, 1–12 (2021).

90. Raposo, A., Moss, H. E., Stamatakis, E. A. & Tyler, L. K. Repetition suppression and semantic enhancement: an investigation of the neural correlates of priming. *Neuropsychologia* **44**, 2284–2295 (2006).

91. Gold, B. T. et al. Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *J. Neurosci.* **26**, 6523–6532 (2006).

92. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* **35**, 121–143 (2012).

93. Dalgleish, T. The emotional brain. *Nat. Rev. Neurosci.* **5**, 583–589 (2004).

94. Fleur, D. S., Bredeweg, B. & van den Bos, W. Metacognition: ideas and insights from neuro-and educational sciences. *NPJ Sci. Learn.* **6**, 1–13 (2021).

95. Fleming, S. M. *Know Thyself: The Science of Self-Awareness* (Hachette, 2021).

96. Mansouri, F. A., Koechlin, E., Rosa, M. G. & Buckley, M. J. Managing competing goals—a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* **18**, 645–657 (2017).

97. Molenberghs, P., Trautwein, F. M., Böckler, A., Singer, T. & Kanske, P. Neural correlates of metacognitive ability and of feeling confident: a large-scale fMRI study. *Soc. Cogn. Affect. Neurosci.* **11**, 1942–1951 (2016).

98. Wambsganss, T., Kueng, T., Soellner, M. & Leimeister, J. M. ArgueTutor: An adaptive dialog-based learning system for argumentation skills. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan* (2021).

99. van Zyl, S. & Mentz, E. Deeper Self-Directed Learning for the 21st Century and Beyond. In *Self-Directed Learning and the Academic Evolution From Pedagogy to Andragogy* (eds Patrick, C. H. & Jillian, Y.) 50–77 https://doi.org/10.4018/978-1-7998-7661-8.ch004 (IGI Global, 2022).

100. Cavanagh, J. F., Figueroa, C. M., Cohen, M. X. & Frank, M. J. Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cereb. Cortex.* **22**, 2575–2586 (2012).

101. Pekrun, R., Goetz, T., Titz, W. & Perry, R. P. Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research. *Educ. Psychol.* **37**, 91–105 (2002).

102. Li, L., Gow, A. D. I. & Zhou, J. The role of positive emotions in education: a neuroscience perspective. *Mind Brain Educ.* **14**, 220–234 (2020).

103. Lin, L. et al. The effect of learner-generated drawing and imagination in comprehending a science text. *J. Exp. Educ.* **85**, 142–154 (2017).

104. Lin, L., Ginns, P., Wang, T. & Zhang, P. Using a pedagogical agent to deliver conversational style instruction: what benefits can you obtain. *Comput. Educ.* **143**, 103658 (2020).

105. Mayer, R. E., Fennell, S., Farmer, L. & Campbell, J. A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *J. Educ. Psychol.* **96**, 389–395 (2004).

106. Yılmaz, F. G. K., Olpak, Y. Z. & Yılmaz, R. The effect of the metacognitive support via pedagogical agent on self-regulation skills. *J. Educ. Comput. Res.* **56**, 159–180 (2018).

107. Cabales, V. Muse: Scaffolding metacognitive reflection in design-based research. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* General Chairs: Stephen, B. & Geraldine, F. 1–6 (Association for Computing Machinery, 2019).

108. Zheng, L., Li, X., Zhang, X. & Sun, W. The effects of group metacognitive scaffolding on group metacognitive behaviors, group performance, and cognitive load in computer-supported collaborative learning. *Internet High. Educ.* **42**, 13–24 (2019).

109. Pan, Y., Cheng, X., Zhang, Z., Li, X. & Hu, Y. Cooperation in lovers: an f NIRS-based hyperscanning study. *Hum. Brain Mapp.* **38**, 831–841 (2017).
110. Duan, L. et al. Wavelet-based method for removing global physiological noise in functional near-infrared spectroscopy. *Biomed. Opt. Express* **9**, 3805–3820 (2018).
111. Huang, M. et al. Joint-channel-connectivity-based feature selection and classification on fNIRS for stress detection in decision-making. *IEEE Trans. Neural Syst. Rehabil. Eng.* **30**, 1858–1869 (2022).
112. Gao, Y. et al. Functional brain imaging reliably predicts bimanual motor skill performance in a standardized surgical task. *IEEE Trans. Biomed. Eng.* **68**, 2058–2066 (2020).
113. Lundberg, S. M. & Lee S. I. A unified approach to interpreting model predictions. In: *31st Conference on Neural Information Processing Systems, NIPS 2017, Long Beach, CA, USA*, (eds Guyon I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. & Garnett, R) Vol. 30 (Curran Associates, Inc, 2017).

## Author contributions
All authors contributed to the conceptual design, research methodology, and manuscript preparation for this study. Significant roles beyond these contributions are noted below. JY led overall research operations, including software development, study design, and manuscript preparation. H.X. was particularly involved in model development, validation, and review. Y.H. and Y.P. contributed to conceptualization, resources, and leading project goals. All authors read and approved the final manuscript.

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41539-025-00311-8.

**Correspondence** and requests for materials should be addressed to Yafeng Pan or Yi Hu.

**Reprints and permissions information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.