International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Comprehensive Review of Error-Related Potentials in Brain–Computer Interfaces and Brain–AI Interaction: Trends, Challenges, and Future Directions

Sathvik Reddy[1], Jathin M[2], Srujan T M[3], Prashant P Patavardhan[4]

[1,2,3,4]Department of Electronics and Communication Engineering, RV Institute of Technology and Management, Bangalore–560076, Karnataka, India

Abstract

Error-related potentials (ErrPs) are event-related EEG re- sponses elicited when a person detects a mismatch between an intended and an observed outcome. Over the past two decades, ErrPs have emerged as a key neural marker for improving the reliability and usability of brain–computer interfaces (BCIs). This review summarizes recent ad- vances in ErrP-based BCIs, with a focus on multitask motor control, subject-independent classification, and emerging brain–AI interaction paradigms. We first outline the neurophysiological basis of ErrPs and classical applications for error correction in BCI spellers and motor- imagery systems. We then discuss recent work on multitasking sensori- motor control, robust classification methods, and generic classifiers that can generalize across users and recording conditions. A special emphasis is placed on frameworks that use ErrPs as implicit feedback to adapt arti- ficial intelligence (AI) agents, particularly large language models (LLMs), in closed loops. In these systems, ErrPs serve as feedback about agree- ment or disagreement with the AI's output, enabling the construction of subject-specific AI behaviour. Across the surveyed studies, we identify common challenges such as low single-trial accuracy, inter- and intra-subject variability, practical deployment constraints, and a strong depen- dence on attention and task context. We conclude by outlining promising directions for improving robustness, personalization, and scalability of ErrP-based BCIs and brain–AI systems.

Keywords: Error-related potentials · Brain–computer interface · EEG · Neural error monitoring · Subject-independent classification · Brain–AI interaction · Large language models

1 Introduction

One of the long-standing ambitions of neurotechnology is to enable seamless com- munication between the human brain and artificial systems. Brain–computer interfaces (BCIs) attempt to decode

patterns of neural activity and translate them into commands for external devices, ranging from cursors and wheelchairs to robotic arms and speech synthesizers. Classical BCIs have relied mainly on actively generated signals such as motor imagery, steady-state visual evoked po- tentials (SSVEPs), or P300 responses. Although these paradigms have enabled impressive demonstrations, they usually demand sustained attention, require extensive training, and can be fragile in the face of distraction, fatigue, or envi- ronmental disturbance.

Error-related potentials (ErrPs) offer a complementary approach. Rather than encoding a deliberate command, ErrPs reflect the brain's automatic re- action when an outcome is perceived as wrong, surprising, or inconsistent with expectations. This performance-monitoring mechanism is deeply embedded in human cognition and operates even when the person is not consciously trying to communicate. As a result, ErrPs can be exploited in BCIs as an implicit, passive feedback channel: the system monitors whether its own actions are judged as correct or erroneous by the user's brain.

Early work in this area showed that ErrPs could be detected during simu- lated BCI control and used to reject or correct erroneous decisions. Subsequent research broadened the range of tasks to include P300 spellers, motor-imagery BCIs, robotic manipulation, and human–robot supervision. In parallel, advances in machine learning and signal processing, such as deep learning and Riemannian geometry, have improved the decoding of ErrPs at the single-trial level. More recently, ErrP signals have been integrated into brain–AI interaction frameworks where they help adapt the behaviour of large language models or other AI agents to individual users.

Despite this progress, several issues hinder wide deployment. Single-trial de- coding often remains only moderately above chance, particularly in subject- independent scenarios or in naturalistic environments. ErrP amplitudes depend strongly on cognitive state, task engagement, and feedback timing, which leads to substantial variability. Practical BCIs must also contend with constraints on hardware, power consumption, setup time, and user comfort, and emerging brain–AI systems must address ethical and privacy concerns when neural data are used to personalize AI behaviour.

This review provides a detailed and structured synthesis of these develop- ments. It adopts an organization similar to comprehensive surveys in related do- mains, aiming to bridge conceptual background, methodological advances, and system-level considerations into a coherent narrative.

2 Organisation of the Paper

To offer a detailed and systematized account of ErrP research, this review is orga- nized into eight main sections. Section 1 introduces the motivation for studying ErrPs in the context of BCIs and brain–AI interaction and outlines why tra- ditional BCI paradigms are insufficient on their own. Section 2 describes the structure of the paper and clarifies how the subsequent sections build upon one another.

Section 3 provides the conceptual background and motivation. It explains the neurophysiological basis of ErrPs, discusses their characteristic components, introduces the main categories of ErrP paradigms, and positions ErrPs within broader neural recording and stimulation architectures. This section also high- lights why ErrPs are particularly attractive for adaptive and safety-critical sys- tems.

Section 4 presents a comprehensive literature review of ErrP-based systems. It summarises representative work on deep-learning models, classical machine- learning pipelines, hybrid and Riemannian approaches, subject-independent frame- works, and recent closed-loop brain–AI interaction systems.

Section 5 offers a comparative numerical analysis of the reviewed studies. It consolidates quantitative results on datasets, classification accuracy, latency, robustness, and hardware requirements, and discusses how system architectures balance performance with real-time and energy constraints.

Section 6 provides a critical discussion of the practical limitations and open problems, including generalization gaps, inter- and intra-subject variability, eco- logical validity, ethical issues, and privacy concerns. Section 7 outlines future research directions, focusing on multimodal fusion, low-power edge AI, improved sensing robustness, explainability, and privacy-preserving learning. Finally, Sec- tion 8 concludes the paper by summarizing the key insights and reflecting on the evolving role of ErrPs in human-centred neurotechnology.

A list of commonly used abbreviations, a declarations section, and the full set of references are included at the end of the paper.

3 Background and Motivation

Driver-state monitoring in vehicles and human-state monitoring in BCIs share a common principle: the human is an integral part of a larger cyber-physical control loop, and any deviation in human state can destabilize the system. In BCIs, the human brain forms the primary control source; errors in decoding or execution not only degrade performance but can also cause frustration, mistrust, or safety risks. ErrP-based monitoring aims to detect these mismatches as soon as they occur, thereby stabilizing the overall loop.

3.1 Neurophysiological Basis of ErrPs

ErrPs belong to the family of event-related potentials (ERPs), which are time- locked EEG responses to discrete events. The dominant components of ErrPs are typically observed over midline fronto-central electrodes, especially FCz and Cz. The earliest component is the error-related negativity (ERN or Ne), a negative deflection peaking around 50–150 ms after the moment the error is detected. It is thought to originate primarily from the anterior cingulate cortex and neighbour- ing medial frontal regions, reflecting rapid evaluation of conflict or mismatch.

Following the ERN/Ne, a positive component known as the error positivity (Pe) appears between roughly 200 and 400 ms. The Pe is associated with con- scious error awareness and the allocation of cognitive resources to process the mistake. Together, these components constitute a highly informative temporal pattern that can be exploited for decoding.

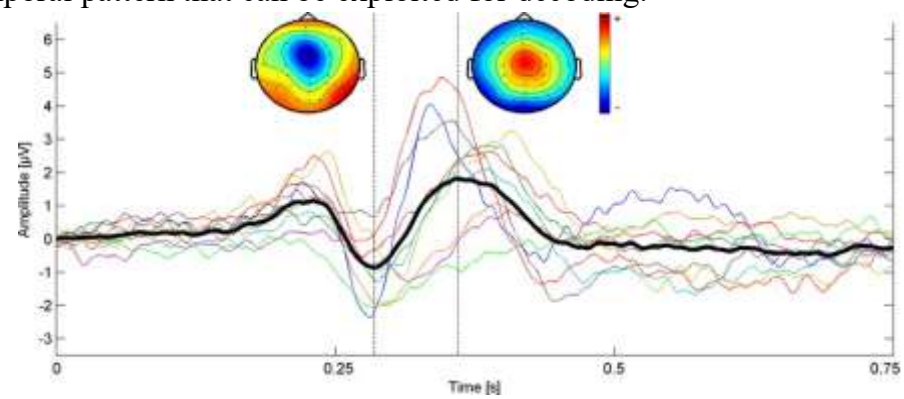

Fig. 1. Error-related potential (ErrP) waveform showing the typical early negativity (ERN/Ne) and late positivity (Pe) across multiple trials. Coloured lines denote indi- vidual trials, while the bold trace represents the average response. Scalp topographies illustrate the characteristic

fronto-central distribution of the components.

Figure 1 shows an example of an ErrP waveform averaged over multiple trials, together with scalp topographies of the ERN/Ne and Pe. The pronounced peaks demonstrate why these components are attractive targets for event-related BCIs.

3.2 Neural Recording, Stimulation, and Control Loops

Modern neurotechnology increasingly relies on bidirectional communication be- tween the nervous system and external devices. BCIs provide a pathway for recording neural activity and translating it into commands, while neural stim- ulation systems deliver electrical or other forms of stimulation back to the ner- vous system to modulate activity. Many advanced systems involve prosthetic or robotic effectors that act upon the environment, sensors that provide feed- back, and higher-level controllers that determine how to combine neural and environmental information.

Figure 2 illustrates this bidirectional architecture. Neural recording modules capture brain activity, which is processed and used to control prosthetic limbs or other devices. Sensory feedback and artificial stimulation return information to the brain. ErrPs naturally arise in this loop when the effectors behave unex- pectedly or when stimulation does not match internal predictions. Consequently, ErrP decoding can serve as a safety and adaptation layer within broader neural control systems.

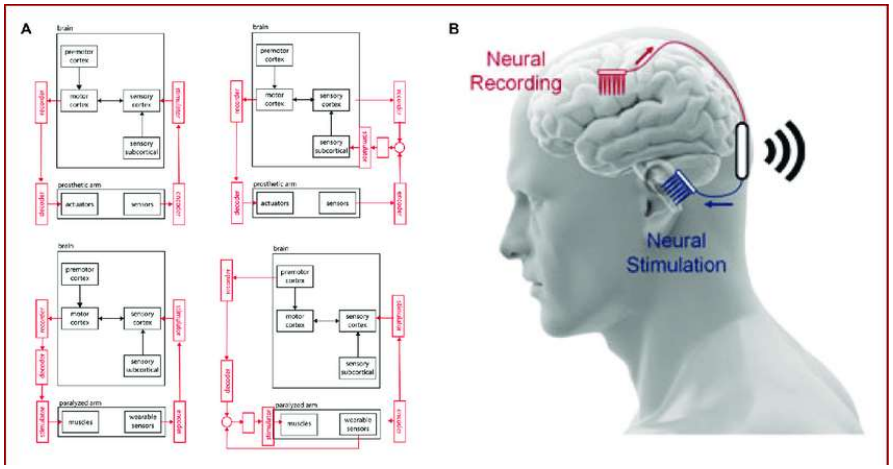

Fig. 2. Illustrative schematics of neural recording and stimulation pathways, and their integration with external effectors. Motor cortical activity can be decoded and used to control prosthetic devices, while sensory information and artificial feedback are con- veyed back to the brain. ErrP-based monitoring can be embedded in such loops to detect mismatches between intended and observed outcomes.

3.3 Taxonomy of ErrP Paradigms

ErrPs can be divided into several broad categories according to how errors are generated and perceived. Response-locked ErrPs occur when the participant per- sonally makes an incorrect response, for instance pressing the wrong key in a speeded reaction-time task. Feedback-related ErrPs arise when external feedback indicates that a decision or action was incorrect, even if the participant's motor response itself was correct. Observation-related ErrPs are elicited when a person observes another agent, such as a robot or another human, making a mistake. Finally, in BCI contexts, interaction ErrPs occur when the system misinterprets the user's intention, such as moving a cursor in the wrong direction or selecting an incorrect symbol.

These categories partly overlap at the neural level but differ in timing, am- plitude, and scalp

distribution. Feedback-based and observation-based ErrPs, in particular, are especially relevant for passive BCIs and brain–AI interaction, because they do not require explicit voluntary responses.

3.4 Motivation for ErrP-Based BCIs and Brain–AI Systems

The motivation for integrating ErrPs into BCIs is twofold. First, ErrPs can substantially improve the reliability and usability of existing BCIs by providing an automatic error-monitoring channel. Instead of relying solely on decoding primary control signals, a system can check whether its outputs are implicitly endorsed or rejected by the user's brain. This concept has been demonstrated in spellers, motor-control tasks and robot supervision scenarios, where ErrP detection enables error correction, undo actions and adaptive recalibration.

Second, ErrPs provide a route to aligning AI systems with human prefer- ences and expectations at a deeper level. When an AI agent proposes actions or interpretations, users rarely provide explicit labels for every outcome. However, their brains continuously evaluate whether those outcomes match internal goals. ErrP-based systems exploit this implicit evaluation by decoding neural agree-ment or disagreement signals and feeding them back into AI learning processes. As AI systems scale in complexity and autonomy, such neural feedback may become an important component of human-centred AI alignment.

4 Literature Review

Research on ErrP-based systems has progressed through several methodological generations, analogous to the evolution observed in driver drowsiness detection and other safety-critical monitoring domains. Early work relied on handcrafted features and classical classifiers, followed by more sophisticated machine-learning pipelines, and most recently by deep-learning and hybrid approaches, including subject-independent frameworks and brain–AI interaction systems.

4.1 Foundational Studies and Classical Machine Learning

Foundational studies established that ErrPs could be reliably detected during simulated BCI feedback and used to correct classification errors. These experi- ments typically used controlled paradigms in which a cursor or symbol selection sometimes deviated from the user's intention, eliciting clear ErrPs. Linear clas- sifiers such as LDA and logistic regression were trained on features extracted from time-locked EEG segments. Despite their simplicity, these models demon-strated that single-trial ErrP decoding was feasible under favourable conditions and that incorporating ErrP-based correction could substantially increase overall BCI accuracy.

Subsequent work investigated the effects of task demands, feedback timing, and cognitive state on ErrP characteristics. For instance, arm-movement tasks with varying motor load showed that increased physical demands modulated the amplitude and latency of ErrPs, highlighting the importance of considering context when designing decoding algorithms. Studies on error awareness and cognitive control used ErrPs to explore how the brain monitors performance and adjusts behaviour, providing valuable insights for both neuroscience and BCI design.

4.2 Advanced Classification: Riemannian and Subject-Independent Approaches

As more data became available and computational methods advanced, researchers turned to more powerful classification frameworks. One influential line of work employed Riemannian geometry to classify covariance matrices derived from ErrP epochs. Instead of treating covariance matrices as vectors in Euclidean space, these methods respect the manifold structure of positive-definite matrices and use appropriate distance metrics. Riemannian classifiers have shown strong

performance for ERP-based BCIs, including ErrP decoding, and are relatively robust to moderate noise and non-stationarity.

Another important direction has been subject-independent classification. Col- lecting large amounts of calibration data from each individual user is inconve- nient and often impractical for real-world systems. To reduce calibration effort, several studies evaluated generic classifiers trained on pooled data from many participants and assessed their ability to generalize to unseen users. Ensemble methods such as Random Forests, domain-adaptation techniques, and transfer- learning strategies have been employed. While performance still typically lags behind subject-specific models, these approaches have made substantial progress toward plug-and-play ErrP-based BCIs.

4.3 Deep Learning and Hybrid Feature Pipelines

Deep-learning architectures have been widely adopted in EEG decoding due to their ability to learn spatial-temporal features directly from raw or minimally processed signals. For ErrP decoding, convolutional neural networks (CNNs) and temporal convolutional networks have been used to capture local waveform patterns, while recurrent models such as LSTMs model longer-range temporal dependencies. Some studies combine convolutional layers with attention mecha- nisms to focus on informative time windows or channels.

Hybrid pipelines often combine deep-learning components with handcrafted or Riemannian features. For example, covariance matrices extracted from EEG segments can be mapped to tangent space and fed into deep networks, or CNN layers can be combined with classical classifiers to balance interpretability and performance. These hybrid designs aim to exploit the strengths of both ap- proaches: the expressive power of deep networks and the robustness of well- understood statistical features.

4.4 ErrP-Driven Brain–AI Interaction

A particularly novel development is the integration of ErrP decoding into brain– AI interaction systems. Instead of using ErrPs solely to correct BCI outputs, these frameworks treat ErrP signals as feedback about the behaviour of an AI agent. One representative architecture couples an ErrP classifier with a large lan- guage model (LLM). In this setup, the user silently evaluates textual hypotheses generated by the LLM, such as guesses about an imagined object or proposed solutions to a task. Each hypothesis is presented on a screen, and the result- ing EEG segment is classified as reflecting agreement or disagreement based on ErrP-like patterns.

Decoded feedback is converted into textual cues describing whether the AI's guess is "too far," "closer," or "correct," and these cues are appended to the LLM's prompt. Over successive rounds, the LLM refines its hypotheses according to this implicit neural feedback. Even when the ErrP classifier achieves only modest single-trial accuracy, the iterative loop can converge to the correct answer by accumulating evidence. Figure 3 depicts such a closed-loop architecture.

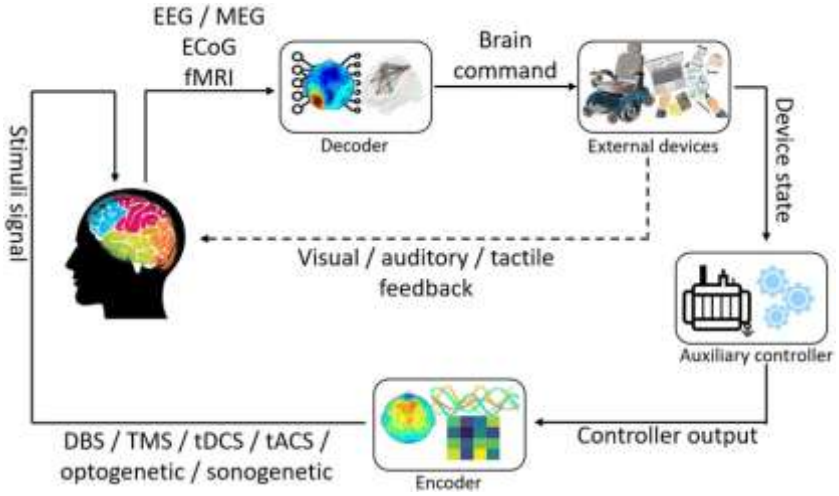

Fig. 3. Closed-loop brain–AI interaction framework in which the user evaluates AI-generated outputs, ErrP-based decoding estimates neural agreement or disagreement, and the resulting feedback is fed back into the AI model, enabling iterative subject- specific adaptation.

This type of system illustrates how ErrPs can serve as a bridge between human internal representations and high-dimensional AI models, potentially en- abling subject-specific semantic spaces and personalized AI behaviour without requiring explicit labelling from the user.

4.5 Summary of Reviewed Studies

Table **??** summarises a representative subset of ErrP studies that span the methodological spectrum described above. The entries illustrate how different research efforts emphasise various aspects such as multitasking control, subject independence, hybrid processing, and AI adaptation.

5 Comparative Numerical Analysis and System Architectures

Quantitative analysis provides a rigorous basis for comparing ErrP-based sys- tems that differ in experimental design, classification methods, and hardware im- plementation. Although the underlying datasets and evaluation protocols vary,

Table 1. Characteristics of major components of Error-Related Potentials (ErrPs).

Component	Typical Latency Range	Functional Interpretation
ERN / Ne	50–150 ms post-error	Rapid, automatic mismatch detection reflecting conflict monitoring; strongest over FCz/Cz.
Pe	200–400 ms post-error	Conscious error awareness, attentional allocation, and evaluation of error significance.
Feedback-Related Negativity (FRN)	200–300 ms after feedback	Processing of unexpected or negative feedback; commonly used in feedback-related ErrP paradigms.
Late Positivity (LP)	300–600 ms post-error	Higher-level appraisal of error context, decision evaluation, and behavioural adjustment.

several general trends emerge when examining performance metrics such as ac- curacy, latency, robustness, and energy consumption.

5.1 Dataset Characteristics and Performance

ErrP datasets differ in the number of subjects, number of trials per condi- tion, recording

hardware, and richness of task contexts. Studies using controlled laboratory paradigms with well-defined events and minimal artefacts often re- port single-trial classification accuracies above 80% with subject-specific models. When datasets contain more naturalistic tasks or higher noise levels, accuracies are typically lower but still above chance. Subject-independent models tend to show accuracies in the 65–80% range, depending on the diversity and quality of the training data.

A recurring observation is that performance saturates as model complexity increases: moving from linear models to deep networks yields improvements, but the gains diminish once accuracies approach the mid-90% range, especially when the training data are limited or not sufficiently diverse. This suggests that collecting richer datasets, rather than simply increasing model complexity, is essential for further advances.

5.2 Classification Methods and Trade-offs

The variety of classification approaches used in ErrP studies can be compared along dimensions of accuracy, interpretability, computational cost, and suitabil- ity for real-time deployment. Table 2 summarizes the main strengths and limi- tations of several commonly used methods.

In practice, the choice of method depends on the target application. For embedded systems with tight latency and power budgets, lightweight linear or ensemble models may be preferable. In contrast, offline analyses or cloud-based systems may benefit from the additional capacity of deep networks.

Table 2. Comparison of classification methods employed in ErrP decoding.

Method	Strengths	Limitations
LDA / Regularized LDA	Fast and simple; well suited for online BCIs; performs reasonably with small datasets	Assumes Gaussian distributions and linear separability; limited capacity for complex non-linear patterns
SVM (Linear/RBF)	Strong performance in high-dimensional spaces; flexible kernels; robust margins	Sensitive to hyperparameters; ker-nel methods can be computation-ally demanding for real-time use
Random Forest / En-sembles	Handles noisy features; captures non-linear relationships; useful for subject-independent decoding	Requires sufficient training data; model size and inference time grow with number of trees
Deep Learning (CNN/RNN)	Learns spatial-temporal features end-to-end; can capture complex dynamics and cross-channel interactions	Needs large labeled datasets; prone to overfitting; high computational and energy cost
Riemannian Geometry Classifiers	Exploit covariance structure of EEG; strong performance for ERP decoding; robust to moderate noise	Require careful preprocessing; com-putationally more demanding; less intuitive for non-experts

5.3 EEG Acquisition and Paradigm Design

The design of EEG acquisition setups strongly influences both the quality of ErrP signals and the feasibility of real-time decoding. Table 3 provides a high- level summary of typical configurations across different application domains.

Table 3. Typical EEG setups and paradigms used in ErrP research. Values are repre-sentative rather than prescriptive.

Study Type	EEG Setup	Task / Paradigm
ErrP-based BCI spellers	32–64 channels, 256–512 Hz, midline focus	P300 or ERP spellers with occasional incorrect symbol selections
Multitasking motor control	64 channels, 500–1 000 Hz, dense frontal coverage	Multi-degree-of-freedom cursor or robot control with deliberate error feedback
Subject-independent analysis	Multi-subject datasets, 16–64 channels	Offline cross-subject decoding across multiple sessions and tasks
Robot action validation	32 channels, standard 10–20 montage	User observes robot performing correct and incorrect actions
Brain–AI ErrP feedback	4–8 channels (e.g., Fz, FC1, FC2, Cz, CP1, CP2)	LLM-generated guesses evaluated by user; agreement/disagreement encoded in ErrPs

There is a clear trend toward using fewer electrodes for practical systems, especially in brain–AI interaction where lightweight headsets and rapid setup are desirable. However, reducing channel count can degrade signal quality and decoding performance, which motivates research into optimized channel selection and spatial filtering.

5.4 Latency, Real-Time Performance, and Energy Use

For many applications, particularly safety-critical ones, latency is as important as accuracy. Real-time ErrP-based BCIs must process EEG segments shortly after feedback events and decide whether an error has occurred before the next system action is generated. Latencies below 100–150 ms are generally considered acceptable for closed-loop interaction.

Deep-learning models running on powerful GPUs can achieve high accuracy but may incur substantial latency and energy consumption, limiting their use on portable devices. Lightweight models deployed on embedded processors or neuromorphic hardware achieve lower latency and power usage but may sacrifice some accuracy. Balancing these trade-offs remains an active area of research, especially as edge-computing hardware becomes more capable.

5.5 Composite Evaluation Metrics

To compare heterogeneous systems fairly, some researchers have proposed com- posite metrics that combine accuracy, latency, scalability and energy consump- tion into a single score. Although the exact weighting schemes vary, the general conclusion is that models with moderate complexity—such as optimized CNNs or well-designed hybrid pipelines—often provide the best balance between per- formance and deployability. Extremely complex models may deliver marginal accuracy gains at disproportionate computational cost, while overly simplistic models may be fast but insufficiently robust.

6 Discussion and Critical Analysis

The last decade of ErrP research has demonstrated considerable progress, yet several limitations prevent these systems from reaching their full potential in real-world applications. Perhaps the most fundamental challenge is variability. ErrPs differ substantially across individuals and across sessions, influenced by anatomical differences, cognitive strategies, fatigue, stress, motivation, and task engagement. Even within a single session, fluctuations in attention can attenu- ate or distort the ERN/Ne and Pe components, directly harming classification accuracy.

This variability complicates the creation of subject-independent models. While ensemble methods, domain adaptation, and transfer learning help, there remains a gap between the performance of subject-specific and generic classifiers. Closing this gap will require both improved algorithms and richer datasets that better capture the broad spectrum of human variability.

A second issue is ecological validity. Many studies employ simplified tasks with discrete events and limited sets of stimuli, conducted in controlled labora- tory environments. In contrast, real-world applications such as assistive robotics, smart home interfaces, or continuous brain–AI interaction operate in dynamic, noisy and unpredictable environments. Motion artefacts, electromagnetic inter- ference, and overlapping cognitive processes can all obscure ErrP signatures.

Designing paradigms and hardware that maintain robust signal quality in such conditions is non-trivial.

Third, there are important human-factor considerations. ErrP-based correc- tion mechanisms should support, not frustrate, users. If the system repeatedly misinterprets or overreacts to ErrPs, users may lose trust or experience cognitive overload. User-centred design is therefore critical: interfaces must provide clear and intuitive feedback, allow users to understand why corrections occur, and avoid excessive false alarms.

Finally, as brain–AI interaction frameworks mature, ethical and privacy is- sues come to the forefront. Neural data can implicitly reveal preferences, in- tentions, and cognitive states that users may not wish to share. ErrP-based personalization of AI agents means that aspects of an individual's internal se- mantic landscape may be encoded in model parameters. Ensuring secure storage, informed consent, and transparent control over such personalization is therefore essential. Privacy-preserving techniques such as on-device learning, federated learning, and differential privacy will likely play an important role in responsible deployment.

7 Future Research Directions

The next generation of ErrP-based systems must satisfy intersecting require- ments of robustness, efficiency, and ethical alignment. Several research avenues appear particularly promising.

Multimodal fusion is likely to become increasingly important. Combining ErrPs with other neural or physiological signals—such as P300, motor imagery, eye tracking, or heart-rate variability—can provide complementary information and improve reliability. Adaptive weighting strategies that emphasize the most informative modalities under current conditions could further enhance perfor- mance. For example, when visual attention is high, ErrP-based monitoring may dominate; when visual input is degraded, other modalities could take precedence. Another major direction is low-power edge AI. TinyML techniques, quan- tized neural networks, and custom hardware accelerators can bring sophisticated decoding models to wearable and portable devices without excessive energy con- sumption. Achieving real-time ErrP decoding on head-mounted systems with power budgets of a few watts or less will make BCIs more practical in everyday life.

Improving sensing robustness is equally critical. Dry electrodes, flexible elec- trode arrays, and integrated sensor headsets are being developed to reduce setup time and increase user comfort, but they often record noisier signals than high- quality gel electrodes. Advanced signal processing and adaptive filtering methods will be needed to compensate for this loss in signal quality. In addition, datasets should increasingly incorporate challenging scenarios such as movement, multi- tasking, and varied environments to foster algorithms that generalize beyond the lab.

Explainability and transparency must be integrated into the design of ErrP- based systems, especially those interacting with AI agents. Users and stakehold- ers should be able to inspect how

neural signals influence system behaviour. In- terpretable models, visualization tools, and user-friendly explanations can help build trust and support certification efforts in safety-critical domains such as medical devices and assistive robotics.

Finally, privacy-preserving learning strategies should become standard prac- tice rather than an afterthought. Federated learning can allow multiple users or devices to contribute to a shared model without exchanging raw EEG data. Combined with secure aggregation, encryption, and robust consent mechanisms, such approaches can enable large-scale improvement of ErrP-based systems while respecting individual autonomy.

8 Conclusion

This paper has presented a comprehensive review of error-related potentials in the context of brain–computer interfaces and emerging brain–AI interac- tion paradigms. Starting from the neurophysiological foundations of ErrPs, we outlined how these signals arise from the brain's performance-monitoring sys- tem and how they can be measured in non-invasive EEG recordings. We then surveyed key methodological developments, including classical machine-learning approaches, Riemannian geometry-based classifiers, deep-learning and hybrid pipelines, subject-independent frameworks, and novel closed-loop systems in which ErrPs guide the adaptation of large language models and other AI agents.

Comparative analysis across studies revealed common trade-offs between accuracy, latency, model complexity, and energy consumption. While several approaches achieve high performance in controlled experimental settings, chal- lenges remain in terms of generalization to diverse users and environments, ro- bustness in naturalistic conditions, and practical deployment on wearable or em-bedded hardware. Moreover, ethical and privacy concerns become increasingly salient as neural data are used not just for control but also for AI personalization.

Despite these obstacles, the trajectory of ErrP research is encouraging. The integration of ErrP-based monitoring into BCIs and AI systems promises to make these technologies more reliable, adaptive, and aligned with human users. As multimodal sensing, low-power AI, and privacy-preserving learning continue to advance, ErrP-driven systems are poised to play a central role in future hu- man–machine symbioses.

List of Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
BCI	Brain–Computer Interface
CNN	Convolutional Neural Network
EEG	Electroencephalography
ERN / Ne	Error-Related Negativity
ERP	Event-Related Potential
ErrP	Error-Related Potential
GPU	Graphics Processing Unit
LDA	Linear Discriminant Analysis
LLM	Large Language Model
LSTM	Long Short-Term Memory Network

Pe	Error Positivity
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SSVEP	Steady-State Visual Evoked Potential

Declarations

Availability of Data and Materials

This article is a review and does not involve the generation of new datasets by the authors. All datasets referenced in this manuscript originate from previously published studies, and details are available in the cited references.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank their institution and colleagues for providing the necessary facilities and academic environment that supported this work.

References

1. M. G. H. Coles, M. K. Scheffers, and L. Fournier, "Where did I go wrong? Errors, partial errors, and the nature of human information-processing," *Acta Psycholog- ica*, vol. 90, pp. 129–144, 1995.
2. M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biological Psychology*, vol. 51, no. 2–3, pp. 87–107, 2000.
3. A. Cruz, G. Pires, and U. J. Nunes, "Double ErrP detection for automatic error correction in an ERP-based BCI speller," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 26–36, 2017.
4. P. W. Ferrez and J. del R. Mill´an, "Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy," in *Proc. 4th Int. BCI Workshop and Training Course*, pp. 197–202, 2008.
5. F. Iwane, K. Okimura, T. Seki, and I. Nambu, "EEG error-related potentials en- code magnitude of errors and individual perceptual thresholds," *iScience*, vol. 26, no. 9, 2023.
6. A. Berkmush-Antipova *et al.*, "Yes or no? A study of ErrPs in the 'guess what I am thinking' paradigm with stimuli of different visual content," *Frontiers in Psychology*, vol. 15, p. 1394496, 2024.
7. M. Ullsperger, C. Danielmeier, and G. Jocham, "Neural mechanisms and temporal dynamics of performance monitoring," *Trends in Cognitive Sciences*, vol. 18, no. 5, pp. 259–267, 2014.
8. A. Farabbi and L. Mainardi, "Assessing the impact of stimulation environment and error probability on ErrP EEG response, detection and subject attention: an explorative

study," *Frontiers in Virtual Reality*, vol. 5, p. 1433082, 2024.
9. C. Kothe *et al.*, "The Lab Streaming Layer for synchronized multimodal record- ing," *bioRxiv*, 2024.
10. L. Fasnacht, "Referenced in JoVE paper: metadata unavailable," 2018.
11. A. Berkmush-Antipova *et al.*, "Do people dream about subject-specific AI? BCI based on ErrPs," *Journal of Visualized Experiments*, 2025 (in review).