RGCS REVISTA GREGORIANA DE CIENCIAS DE LA SALUD

ISSN 3028-8940

REVIEW ARTICLE

JULY-DECEMBER 2026;3(2):220-237
https://doi.org/10.36097/rgcs.v3i2.3254

Artificial intelligence-based clinical decision support systems in neurophysiology: conceptual framework, design criteria, and ethical considerations

Sistemas de apoyo a la decisión clínica basados en inteligencia artificial en neurofisiología: marco conceptual, criterios de diseño y consideraciones éticas

Duliet Hong* • Jaime E. Pérez • Jaime Azpiazu

Faculty of Engineering, International University of Cuanza, Cuito, Angola.

**Corresponding author*

Reception: 02-05-2026 *Acceptance:* 26-06-2026 *Publication:* 31-07-2026

ABSTRACT

Diagnosis in neurophysiology using visual evoked potentials (VEP) for multiple sclerosis and optic neuropathies remains specialist-dependent and subject to variability. Artificial intelligence (AI) offers transformative potential, but its responsible implementation is hindered by algorithmic metrics. This paper develops a conceptual framework for AI-based clinical decision support systems (CDSS) in neurophysiology, structured around three integrated dimensions: CDSS typology and foundations, responsible design criteria, and ethical-regulatory considerations. Six architectural principles are proposed: transparency, clinical anchoring in validated criteria (ISCEV 2023), preservation of physician judgment, false-alarm minimization, reproducibility, and interoperability, applicable to any system in the field. The framework is illustrated through an EEG-VEP system on the VEPCON dataset, integrating automated classification, clinical hierarchy over statistical thresholds, and a dashboard with an explicit ethical disclaimer. The responsible development of CDSS in neurophysiology demands a systemic approach that integrates clinical evidence, operational design, and medical device regulation, ensuring their role as support rather than as a substitute for clinical judgment.

Keywords: artificial intelligence, clinical decision support system, neurophysiology, visual evoked potentials, AI ethics, regulatory framework.

RESUMEN

El diagnóstico en neurofisiología mediante potenciales evocados visuales (VEP) para esclerosis múltiple y neuropatías ópticas sigue siendo dependiente del especialista y sujeto a variabilidad. La inteligencia artificial (IA) ofrece potencial transformador, pero su implementación responsable se obstaculiza por métricas algorítmicas. Este artículo desarrolla un marco conceptual para sistemas de apoyo a la decisión clínica (CDSS) basados en IA en neurofisiología, articulado en tres dimensiones integradas: tipología y fundamentos de los CDSS, criterios de diseño responsable y consideraciones ético-regulatorias. Se proponen seis principios arquitectónicos: transparencia, anclaje en criterios clínicos validados (ISCEV 2023), preservación del juicio médico, minimización de falsas alarmas, reproducibilidad e interoperabilidad, aplicables a cualquier sistema en el campo. El marco se ilustra mediante un sistema EEG-VEP sobre el *dataset* VEPCON, que integra clasificación automática, jerarquía clínica sobre umbrales estadísticos y un *dashboard* con advertencia ética explícita. El desarrollo responsable de CDSS en neurofisiología exige un enfoque sistémico que integre evidencia clínica, diseño operativo y regulación de dispositivos médicos, garantizando su rol como apoyo y no como sustituto del criterio clínico.

Palabras clave: inteligencia artificial, sistema de apoyo a la decisión clínica, neurofisiología, potenciales evocados visuales, ética de la IA, marco regulatorio.

Cite as: Hong, D., Pérez, J. E., & Azpiazu, J. (2026). Artificial intelligence-based clinical decision support systems in neurophysiology: conceptual framework, design criteria, and ethical considerations. *Revista Gregoriana de Ciencias de la Salud, 3*(2), 220-237. https://doi.org/10.36097/rgcs.v3i2.3254

INTRODUCTION

The integration of artificial intelligence (AI) into the diagnostic and therapeutic processes of contemporary medicine has generated unprecedented expectations within the scientific and clinical community. Machine learning algorithms and deep neural networks have demonstrated performance comparable to or exceeding that of human specialists in specific tasks related to diagnostic imaging, analysis of physiological signals, and clinical risk prediction (Topol, 2019; Shen et al., 2019). However, the enthusiasm generated by these results has not been matched by a rigorous reflection on the conceptual frameworks, design criteria, and ethical considerations that should guide the development of clinically responsible AI systems. The publication of international guidelines—such as those from the WHO (2021), the European Commission (2019), and the World Medical Association (2022)—demonstrates that the international community has become aware of this gap, but its translation into the concrete practice of developing diagnostic support systems in specific clinical specialties remains in its early stages.

In the field of clinical neurophysiology, this tension between technological potential and clinical responsibility is particularly evident in the analysis of visual evoked potentials (VEPs). VEPs are a well-established diagnostic tool for assessing the integrity of the visual pathway and are routinely used in the diagnosis of multiple sclerosis, optic neuritis, and optic neuropathies (Odom et al., 2023). However, their clinical interpretation remains anchored in manual processes that depend on the specialist's experience and lack automated normative comparison mechanisms (Manz et al., 2022). Developing an AI system that automates this interpretation without compromising patient safety or displacing the physician's clinical judgment requires more than just a high-performance classification algorithm: it requires a conceptual framework that coherently integrates clinical rationale, ethical principles, and regulatory requirements.

This article develops such a conceptual framework with specific reference to AI-based clinical decision support systems (CDSS) in neurophysiology. The framework is structured around three complementary axes: the typology and theoretical foundations of clinical decision support systems (CDSS) in healthcare, design criteria focused on transparency, explainability, and preserving the physician's role, and the ethical and regulatory considerations that condition the clinical implementation of these systems. Throughout the article, the framework is illustrated with reference to an EEG-VEP analysis system developed as a research project on the public dataset VEPCON (Pascucci et al., 2022), which serves as a concrete application case of the principles

discussed. The objective is not to report the results of this system, but rather to use it as an empirical reference that grounds the theoretical reflection in the reality of computational development applied to healthcare.

The relevance of this topic lies in a paradox frequently observed in the AI literature in medicine: the proliferation of systems with impressive performance metrics that, nevertheless, fail to achieve clinical implementation due to a lack of the necessary conceptual, ethical, and regulatory foundations for institutional trust (Topol, 2019; Cabitza et al., 2017). Understanding why this occurs—and what principles should guide the design to avoid it—is the central problem motivating this work.

METHODOLOGY

This paper develops a conceptual framework for AI-based clinical decision support systems in neurophysiology. As the objective is not to report primary experimental results but to propose a theoretical structure and operational criteria for responsible CDSS design, a qualitative methodology of conceptual synthesis and framework construction was adopted, structured in three sequential phases.

The first phase of the work consisted of the literature review and selection, for which a structured search was conducted in PubMed, IEEE Xplore, and Scopus databases using the following search strings: ("clinical decision support systems" OR "CDSS") AND ("artificial intelligence" OR "machine learning") AND ("neurophysiology" OR "visual evoked potentials" OR "EEG") AND ("ethics" OR "regulation" OR "explainability"). Publications in English and Spanish between 2017 and 2025 were included, prioritizing international clinical guidelines (ISCEV 2023, WHO 2021), regulatory frameworks (FDA SaMD, IMDRF, GDPR), and systematic reviews on CDSS and AI ethics. The selection was purposive, aimed at covering the three axes of the framework: technological foundations, design criteria, and ethical-regulatory considerations.

The second phase comprised the analysis and conceptual synthesis of the selected literature, from which an iterative thematic analysis was performed to identify recurrent categories in CDSS design. This process allowed the articulation of the framework into three integrated dimensions (typology, responsible design, and ethics/regulation) and the inductive derivation of the six proposed architectural criteria (transparency, clinical anchoring, preservation of physician judgment, false-alarm minimization, reproducibility, and interoperability). Each criterion was

cross-validated against its occurrence in the reviewed sources and its consistency with bioethical principles (Beauchamp and Childress) and trustworthy AI guidelines (European Commission, 2019).

The third phase was oriented towards illustrating the framework through an application case, for which the EEG-VEP system developed on the public VEPCON dataset (OpenNeuro ds003505) was used as an illustrative case study. This system integrates an SVM-RBF classifier, a normative profile based on the ISCEV 2023 standard, and an interactive clinical dashboard with an explicit ethical disclaimer. The selection of this case responds to the fact that the system was developed by the authors and allows the operational concretisation of the six design criteria in a real neurophysiological signal processing environment. It should be noted that the purpose of the case is not to clinically validate the system (which would require prospective studies with patients), but to demonstrate the technical and conceptual feasibility of the framework.

To conclude, it should be acknowledged that the main methodological limitation of this approach is its qualitative and propositional nature; the proposed criteria have not undergone empirical validation through clinical trials or usability studies with neurophysiologists. This limitation is explicitly recognized in the discussion and presented as a future research direction.

RESULTS AND DISCUSSION

Theoretical foundations of clinical decision support systems

Definition and historical evolution

Clinical decision support systems (CDSS) are defined as software tools that provide physicians, patients, or other healthcare professionals with intelligently filtered, timely, patient-specific knowledge and data to improve the quality of care (Sutton et al., 2020). This definition, which emphasizes the supportive nature and timeliness of the intervention, establishes a fundamental distinction from automated diagnostic systems: CDSS do not replace the physician but rather provide structured information so that the physician can make the decision.

The history of CDSS dates back to the rule-based expert systems of the 1970s and 1980s, such as MYCIN—a bacterial infection diagnostic system developed at Stanford—and INTERNIST-I, a pioneer in the differential diagnosis of internal diseases (Miller, 1994). These systems encoded medical knowledge in the form of IF-THEN logical rules, with explicit and fully auditable inference logic. Their main limitation was scalability: updating medical knowledge

required the continuous participation of experts to rewrite the rules, a costly and time-consuming process. The advent of machine learning and, subsequently, deep learning, radically transformed the landscape: systems could now learn knowledge representations directly from large volumes of clinical data without requiring manual rule coding (Topol, 2019). The price of this gain in predictive capacity, however, was a proportional loss of interpretability: deep learning models operate as 'black boxes' whose internal logic is opaque to both clinicians and developers.

Types of CDSS according to the inference mechanism

The specialized literature distinguishes four main types of CDSS according to their inference mechanism, each with its own advantages, limitations, and characteristic clinical applications (Sutton et al., 2020; Shortliffe and Sepúlveda, 2018). Table 1 summarizes this typology with specific examples from the field of neurological health.

Table 1. Typology of clinical decision support systems according to inference mechanism

CDSS type	Inference mechanism	Examples of healthcare applications	Main advantages and limitations
Rule-based CDSS	Inference engines based on manually encoded clinical rules	Drug interaction alerts; medication dosage reminders	High transparency; easy to audit; limited generalization capability
Machine Learning (Data-driven) CDSS	Models trained on historical data (SVM, Random Forest, Artificial Neural Networks)	Radiological image classification; sepsis prediction; EEG analysis	High predictive performance; "black-box" risk; requires labeled datasets
Hybrid CDSS	A combination of clinical rules and machine learning models	Optic neuropathy diagnosis; VEP risk traffic-light systems	Balance between interpretability and predictive performance
Ontology-based CDSS	Representation of medical knowledge using controlled vocabularies and ontologies	Diagnostic coding systems; HL7/FHIR interoperability	High interoperability; costly development; complex maintenance and updating

Adapted from Sutton et al. (2020) and Shortliffe and Sepúlveda (2018).

Hybrid systems, which combine explicit clinical rules with machine learning models, represent the most promising approach for neurological diagnostic applications. Their fundamental advantage is that they allow the integration of established clinical knowledge—such as the ISCEV

standard criteria for VEP interpretation—with the ability of ML algorithms to detect complex statistical patterns in high-dimensional EEG signals. The EEG-VEP analysis system developed by the authors illustrates this approach: an SVM-RBF classifier operates on a multidomain vector of 121 features, but its output is integrated with a hierarchical risk traffic light algorithm that gives absolute priority to the ISCEV 2023 clinical criterion (P100 latency > 115 ms as an indicator of demyelination) over any statistical threshold (Odom et al., 2023). This architecture ensures that expert clinical knowledge is not subordinated to the statistical performance of the algorithm, a principle that Cabitza et al. (2017) call 'clinical primacy' in CDSS design.

The problem of technological fragmentation

A systematic review of 47 publications on computational EEG analysis systems by Manz et al. (2022) identified a recurring pattern in the literature: the proliferation of partial developments that address only one component of the analysis pipeline—preprocessing, classification, or visualization—without articulating the complete cycle from the raw signal to the clinical interface. Less than 15% of the reviewed systems provided any interactive visualization capabilities, and less than 10% included automated normative comparison. Fernández-Varela et al. (2022) pointed out that this fragmentation stems from an asymmetry of incentives in academia: scientific publication rewards algorithmic innovation, not clinical integration. The result is a literature rich in algorithmic refinements but poor in systems transferable to clinical practice. Overcoming this technological fragmentation is one of the central conceptual challenges of the field, and its solution requires adopting a systems design perspective that takes the needs of the clinical user, not the algorithm's performance metrics, as its starting point.

Design criteria for responsible CDSS in neurophysiology

The centrality of the clinical user

The most fundamental principle of responsible CDSS design is to focus on the clinical user as the starting point of the design process, not as a passive recipient of a pre-configured technological solution (Sutton et al., 2020). This principle, called 'user-centered design,' has concrete implications for the system's architecture: the neurophysiologist does not need to understand how an SVM-RBF classifier works, but does need to know which electrophysiological parameters are altered, by what magnitude they deviate from the normative range, and what clinical criterion establishes the alarm threshold. A system that reports 'AUC = 0.728' is not clinically useful; a system that reports 'RISK: P100 latency = 128.5 ms in three occipital channels, ISCEV

criterion >115 ms active' is.

The translation of computational output into clinical language is, therefore, a first-order design function, not a cosmetic addition. Abbas et al. (2025) identified in their review of CDSS in neurophysiology three elements of clinical communication that clinicians consider essential: (1) visualization of the original electrophysiological signal on which the recommendation is based; (2) explicit quantification of deviation from normative reference values; and (3) identification of the established clinical criterion (clinical practice guideline, international standard) that activates the alert. All three elements are present in the subject analysis module of the reference EEG-VEP system: the actual VEP waveform loaded from the signal file, the z-scores per feature relative to the normative profile of 19 healthy subjects, and the explicit indication of the ISCEV 2023 criterion when activated (Pascucci et al., 2022; Odom et al., 2023).

Explainability and interpretable artificial intelligence

The explainability of AI systems in medicine—usually referred to in the literature under the term Explainable AI (XAI)—has emerged as one of the most debated principles in the ethics of clinical AI. The underlying argument is that medical trust in a diagnostic support system requires understanding the logic of its recommendations: a doctor who cannot verify why the system is recommending an alert cannot assume responsibility for the clinical decision made based on it (Arrieta et al., 2020). This requirement for explainability has direct consequences on the choice of ML algorithms: an SVM-RBF classifier with feature importance derived from the analysis of separation hyperplane weights is more interpretable than a deep convolutional neural network, although the latter can achieve higher AUC under controlled conditions.

Arrieta et al. (2020) propose an operational distinction between inherently interpretable models—whose decision logic is accessible by construction (decision trees, logistic regression, LDA)—and post-hoc explainable models—whose logic is not directly accessible but can be approximated using techniques such as LIME (Local Interpretable Model-agnostic Explanations) or SHAP (SHapley Additive exPlanations). In the context of VEP analysis, interpretability has a specific additional dimension: the electrophysiological parameters of the feature vector (P1 latency, N170 amplitude, alpha power) are directly meaningful to the clinical neurophysiologist, making feature-based classifiers intrinsically more interpretable to the clinical user than deep learning models on raw signal, whose internal representations lack direct clinical correlates.

Six fundamental design criteria

Based on the review by Sutton et al. (2020), Abbas et al. (2025), Arrieta et al. (2020) and Cabitza et al. (2017), six fundamental criteria are proposed for the design of responsible CDSS in neurophysiology. Table 2 presents them with their conceptual justification and their concrete application in the reference EEG-VEP system.

Table 2. Fundamental criteria for the design of responsible CDSS in neurophysiology

Design principle	Description	Application in the EEG–VEP system
Transparency and explainability (XAI)	The system should communicate the factors underlying each recommendation in a manner that is understandable to clinicians, without requiring technical knowledge of the underlying algorithms.	VEP system: z-scores reported for each feature, explicit normative thresholds, and the ISCEV criterion identified as the highest-priority decision rule.
Grounding in validated clinical criteria	Decision rules should be based on current international clinical guidelines rather than solely on statistical patterns derived from data.	Integration of the ISCEV 2023 criterion (P100 > 115 ms) as the primary decision rule in the traffic-light risk algorithm.
Preservation of the physician's role	The system provides recommendations rather than diagnoses. The clinician retains full authority and responsibility for the final decision.	Explicit ethical notice displayed on the dashboard: *"This system is a decision-support tool and does not replace the clinical judgment of the specialist."*
Minimization of alert fatigue	Systems with high false-alarm rates contribute to operator fatigue and reduce clinicians' trust in the system.	100% specificity within the normative dataset: no healthy participant generates a false-risk alert.
Reproducibility and auditability	Results should be reproducible and traceable, and the decision-making process should be auditable at any time.	Open-source code (GitHub/Zenodo), fixed random seed (42), and fully documented library versions.
Interoperability	The system should be compatible with healthcare data standards and clinical information technology infrastructures.	Use of the BIDS/OpenNeuro format for input data; web-based architecture (Plotly Dash) compatible with any standard web browser.

The application examples correspond to the EEG-VEP analysis system described in Pascucci et al. (2022) and Odom et al. (2023).

These six criteria do not constitute an exhaustive checklist, nor are they independent of

each other. Transparency and explainability are necessary conditions for preserving the physician's role: a clinician cannot responsibly exercise their decision-making authority on a recommendation whose logic they cannot verify. Minimizing false alarms and reproducibility are necessary conditions for institutional trust: a system that generates unjustified alerts or whose results are not reproducible erodes the credibility of the system as a whole, regardless of the soundness of its algorithmic foundations (Sutton et al., 2020).

These six criteria do not constitute an exhaustive checklist, nor are they independent of each other. Ethical Considerations in the Development of AI-Based Clinical Diagnostic Services (CDSS)

Bioethical principles applied to clinical AI

The ethical analysis of AI systems in medicine has tended to be organized around the four classic principles of bioethics proposed by Beauchamp and Childress (1979): autonomy, beneficence, non-maleficence, and justice. Their application to AI-based CDSS takes on specific nuances that deserve detailed attention (Jobin et al., 2019; WHO, 2021).

Autonomy. In the context of CDSS, autonomy has a dual dimension: that of the patient, who must consent to the use of automated systems in their diagnostic process; and that of the physician, whose capacity to exercise independent clinical judgment should not be displaced by the inertia of algorithmic recommendations. This last risk—known as 'automation bias' or 'complacency bias'—means that the clinician tends to follow the system's recommendation without the critical scrutiny they would apply to their own judgment, especially when the system has a reputation for high accuracy (Goddard et al., 2012). Countering this bias requires that the system design actively emphasize the supportive—not substitutive—nature of the recommendation, through explicit ethical warnings in the interface and by presenting the model's uncertainty alongside its recommendation.

Beneficence and non-maleficence. The imperative of beneficence demands that the system effectively improve the patient's clinical outcomes. In the absence of formal clinical validation with real-world pathological data, no diagnostic AI system can claim to meet this principle: it may demonstrate statistical performance on historical data, but the impact on patient outcomes can only be determined through prospective clinical studies (Topol, 2019). The principle of non-maleficence adds a symmetrical requirement: the system must not cause harm, either through misclassifications leading to inappropriate treatments (false positives) or through diagnostic

omissions that delay necessary treatment (false negatives). The choice of the risk classification threshold—and, in particular, the hierarchical priority given to the ISCEV clinical criterion over the statistical thresholds in the reference EEG-VEP system—is an ethical decision as much as a technical one: prioritizing sensitivity (minimizing false negatives) implies accepting a higher false alarm rate, with its own clinical and psychological consequences for the patient.

Justice. The principle of justice raises questions about equity in access to the benefits of clinical practice safety systems (CPSS) and the absence of systematic bias in their recommendations. AI models trained on homogeneous populations—such as a normative corpus of healthy adults from a single European university—can generate normative profiles that do not represent the variability of other populations in terms of age, ethnicity, baseline health conditions, or acquisition equipment (Obermeyer et al., 2019). A variable-precision imaging (VPI) system developed on a normative corpus of young adults from Northern Europe can incorrectly classify older patients as at risk, even though their P100 latencies are physiologically longer due to normal neuronal aging. Explicitly documenting the profile of the reference population—and the resulting limitations on generalizability—is not only a methodological requirement but also an ethical imperative of justice.

Privacy, data protection, and sovereignty of clinical data

The development of AI-based clinical decision support systems (CDSS) in the healthcare field involves the processing of particularly sensitive data: electroencephalographic recordings contain information not only about the patient's neurological health, but potentially about their cognitive activity, emotional state, and, as neuroscience advances, aspects of their mental privacy (Ienca and Andorno, 2017). The European Union's General Data Protection Regulation (GDPR) (2016/679) establishes specific requirements for the processing of health data, including the legal basis for processing, the principle of data minimization, and the right not to be subject to automated decisions with significant legal effects (Martínez Martínez, 2018). In the Ibero-American context, equivalent laws—such as Mexico's Federal Law on the Protection of Personal Data Held by Private Parties or Argentina's Personal Data Protection Law—establish similar frameworks with variations in the rigor of the requirements.

The adoption of the FAIR (Findable, Accessible, Interoperable, Reusable) principles in the management of neurophysiological data (Wilkinson et al., 2016) and the use of standardized formats such as BIDS (Brain Imaging Data Structure; Gorgolewski et al., 2016) and open access

repositories such as OpenNeuro (Markiewicz et al., 2021) represent a partial response to these concerns: they facilitate scientific reproducibility and reduce dependence on proprietary datasets, but they do not resolve questions about the privacy of individual patient data in real-world clinical settings. Robust anonymization, informed consent management, and the definition of data governance protocols are necessary conditions for the transition from academic prototype to a deployed clinical system.

The risk of algorithmic overconfidence and distributed responsibility

One of the more subtle ethical challenges of deploying CDSS in clinical practice is the redistribution of moral and legal responsibility it generates. When a physician makes a diagnosis based solely on their clinical judgment, the responsibility lies with them. When that diagnosis is supported by the recommendation of an AI system, responsibility is distributed in a complex way among the physician, the system developer, the institution that deployed it, and potentially, the manufacturer of the acquisition equipment (Mittelstadt et al., 2016). This dispersion of responsibility—which some authors call the 'distributed responsibility problem' or responsibility gap—has no purely technical solution: it requires legal and regulatory frameworks that clearly define the responsibilities of each actor in the medical AI ecosystem. The classification of CDSS as software medical devices (SaMDs) within current regulatory frameworks (FDA, CE MDR) is a first step in this direction, but their effective implementation in academic and open-source systems like the one described in this work requires further reflection on the responsibilities of researchers developing tools potentially transferable to the clinical setting.

Regulatory framework for the development and implementation of CDSS

The concept of software as a medical device (SaMD)

The SaMD, coined by the IMDRF (International Medical Device Regulators Forum) working group in 2013, designates software that performs one or more medical functions without being part of a hardware medical device. A diagnostic support system that analyzes EEG-VEP signals and issues neurological risk classifications would fall into this category, with specific regulatory implications depending on the jurisdiction and the level of risk associated with its clinical function (IMDRF, 2013). Table 3 summarizes the main regulatory frameworks applicable in the most relevant jurisdictions for the Ibero-American context.

Table 3. Regulatory framework for CDSS as software medical devices (SaMDs) in major jurisdictions

Regulatory framework	Regulatory instrument	Key requirements	Implications for neurological CDSS
FDA (United States)	*Software as a Medical Device (SaMD) Action Plan* (2021)	Risk classification into four levels based on impact on clinical decision-making and the severity of the medical condition.	A CDSS reporting P100 > 115 ms as an indicator of demyelination requires validated clinical evidence before commercialization.
CE MDR (European Union)	*Regulation (EU) 2017/745*; *MDCG 2019-11 Guidance*	Classification into Classes I–III according to invasiveness and risk; CDSS typically falls under Class IIa or IIb.	The VEP system would likely be classified as Class IIa (non-invasive diagnostic support) and would require certification by a notified body.
ISO 13485 / IEC 62304	Medical software quality management	Requirements for software life-cycle processes, requirements traceability, and risk management (ISO 14971).	Applicable to the EEG–VEP analysis pipeline during the transition from research prototype to certified medical device.
PAHO/WHO (Latin America)	Regulation of medical software in Latin America: PAHO guidelines (2021)	Progressive harmonization with the FDA and CE frameworks; regulatory requirements vary among countries.	CDSS developed in Latin American academic settings should comply with the applicable regulatory framework before any clinical implementation.

FDA: Food and Drug Administration (USA); CE-MDR: European Conformity-Medical Device Regulation 2017/745; PAHO: Pan American Health Organization.

Regulatory implications for the development cycle

For researchers developing CDSS in academic settings, the regulation of SaMD has a fundamental implication that is frequently overlooked in the early stages of development: the path from a functional academic prototype to a clinically deployable system is long, costly, and conditioned by clinical evidence requirements that go far beyond computational performance metrics. The FDA's risk classification for SaMD establishes four levels based on the interaction between the patient's condition (critical, serious, non-serious) and the impact on clinical decision-making (inform, guide, treat/diagnose). A VEP analysis system that reports on the risk of demyelination in the context of suspected multiple sclerosis would operate at the highest risk level—serious condition, impact of treatment/diagnosis—requiring the most stringent level of

clinical evidence before approval (FDA, 2021).

This regulatory reality should not be interpreted as an obstacle to the development of clinical practice safety systems (CPSS) in academic settings, but rather as a clear horizon of responsibility that must be integrated into project planning from its initial phases. In practice, this means that systems developed in early stages must be honestly presented as research prototypes with a scope limited to the academic context, without claims of diagnostic equivalence with established clinical practice. The reference EEG-VEP system operates under this principle: the explicit ethical disclaimer on the dashboard and the comprehensive documentation of methodological limitations—particularly the lack of clinical validation with real pathological data—constitute the practical embodiment of this position of epistemic responsibility.

From metrics to clinical value: challenges in AI neurophysiology

The conceptual framework developed in this article allows us to identify a structural tension in the field of AI-based clinical diagnostic systems (CDSS) in neurophysiology: the asymmetry between the speed of technological advancement and the speed of clinical validation, regulation, and ethical reflection processes. EEG classification algorithms can improve their AUC by 5 points in weeks through hyperparameter adjustments or architectural changes; prospective clinical validation of a diagnostic system requires years, institutional ethical approval, and collaboration with leading neurological services. This asymmetry generates implicit pressure toward the publication of technical systems with impressive metrics, but without the necessary clinical support for their responsible implementation, perpetuating the paradox identified by Topol (2019): more AI, less AI useful in practice.

The digital transformation of healthcare systems requires not only the development of algorithms with high predictive capacity, but also the training of professionals capable of understanding, using, and critically evaluating these technologies. In this context, emerging technologies based on artificial intelligence have demonstrated growing potential to improve teaching, research, and clinical practice processes, although their effective implementation depends on appropriate pedagogical, ethical, and organizational conditions. Rodríguez et al. (2026) emphasize that incorporating emerging technologies into higher education requires implementation models that integrate technological innovation with pedagogical strategies and evaluation of their effectiveness. Complementarily, the application of artificial intelligence in medicine presents disruptive opportunities in areas such as diagnostic support, personalized

medicine, and clinical decision-making, but also demands consideration of aspects related to governance, security, and professional responsibility (Pinargote et al., 2025). Likewise, artificial intelligence literacy in medical training is fundamental to preparing future professionals for an increasingly digitalized healthcare landscape, where these tools can foster autonomous learning, knowledge management, and the acquisition of clinical skills, provided they are used ethically and with adequate supervision (Salas et al., 2025).

The response to this tension cannot be either regulatory acceleration or a slowdown in technological advancement, but rather the early integration of responsible design criteria into the systems development cycle. The six criteria proposed in Table 2—transparency, clinical anchoring, preservation of the physician's role, minimization of alert noise, reproducibility, and interoperability—are not requirements added in the final stages of development, but rather architectural principles that should guide design decisions from the outset. Choosing an interpretable SVM-RBF classifier over a deep CNN, integrating the ISCEV criterion as an absolute priority rule, or establishing a random seed to guarantee reproducibility are examples of technical decisions that express ethical commitments.

The principle of justice deserves particular attention in the Ibero-American context, where health systems exhibit high heterogeneity in technological resources, availability of specialists, and demographic diversity of the populations served. A CDSS developed on a normative corpus of healthy adults from a European university institution—as is the case with the VEPCON dataset—can generate normative profiles that do not represent the variability of Latin American populations. This limitation does not invalidate the system, but it requires clear documentation, and expanding the normative corpus to include more diverse populations should be an explicit priority for future work.

The SaMD regulatory framework, although designed primarily for industrial and commercial settings, provides valuable concepts for academic projects: the notion of 'risk level' based on the clinical impact of the system's function offers a rational criterion for calibrating the rigor of the validation required before any claim of diagnostic utility. A system that 'reports' on statistical deviations from a normative profile has less stringent validation requirements than one that 'diagnoses' a specific disease; this distinction should be explicit in the system's academic and clinical communication.

CONCLUSIONS

Responsible development of AI-based clinical decision support systems in neurophysiology demands a systemic design approach integrating clinical, ethical, and regulatory foundations from the outset, moving beyond algorithmic performance metrics alone. The proposed framework aligns three integrated dimensions—CDSS typology, responsible design criteria, and ethical-regulatory considerations—yielding six architectural principles (transparency, clinical anchoring, physician role preservation, false-alarm minimization, reproducibility, and interoperability) applicable to any neurophysiological CDSS. Operationalized in the reference EEG-VEP system—through ISCEV 2023 hierarchical priority over statistics, 100% internal specificity, and an explicit ethical disclaimer—this confirms that ethical design is constitutive of, not adjunct to, technical rigor. Yet the critical gap remains clinical: the absence of prospective validation with real-patient data. Pending such rigorous validation, these systems must be honestly framed as research-support instruments, not substitutes for established practice—an epistemic responsibility that embodies, rather than undermines, the researcher's ethical commitment to patient welfare.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Jaime E. Pérez and Duliet Hong. **Data curation:** Duliet Hong. **Formal analysis:** Duliet Hong. **Funding acquisition:** Jaime E. Pérez. **Investigation:** Duliet Hong and Jaime E. Pérez. **Methodology:** Jaime E. Pérez and Duliet Hong. **Project administration:** Jaime E. Pérez. **Resources:** Jaime E. Pérez. **Software:** Duliet Hong. **Supervision:** Jaime E. Pérez. **Validation:** Jaime E. Pérez and Duliet Hong. **Visualization:** Duliet Hong. **Writing – original draft:** Duliet Hong. **Writing – review & editing:** Jaime E. Pérez and Duliet Hong.

REFERENCES

Abbas, M., et al. (2025). Explainable artificial intelligence in clinical neurophysiology: Current state and future directions. *Artificial Intelligence in Medicine, 152*, 102895. https://doi.org/10.1016/j.artmed.2024.102895

Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Beauchamp, T. L., & Childress, J. F. (1979). *Principles of biomedical ethics*. Oxford University Press.

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA*, *318*(6), 517–518. https://doi.org/10.1001/jama.2017.7797

Comisión Europea. (2019). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

FDA. (2021). *Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan*. U.S. Food and Drug Administration. https://www.fda.gov/medical-devices/software-medical-device-samd

Fernández-Varela, I., et al. (2022). A unified deep learning framework for real-time clinical EEG analysis. *Expert Systems with Applications*, *187*, 115895. https://doi.org/10.1016/j.eswa.2021.115895

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Gorgolewski, K. J., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044. https://doi.org/10.1038/sdata.2016.44

Ienca, M., & Andorno, R. (2017). Towards new human rights in the neurotechnology age. *Life Sciences, Society and Policy*, *13*(1), 5. https://doi.org/10.1186/s40504-017-0050-1

IMDRF. (2013). *Software as a Medical Device (SaMD): Key definitions*. International Medical Device Regulators Forum. https://www.imdrf.org/documents/software-medical-device-samd-key-definitions

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Manz, P., et al. (2022). Clinical decision support systems for EEG-based neurological diagnosis: A systematic review of implementation gaps. *Journal of Biomedical Informatics*, *129*, 104052. https://doi.org/10.1016/j.jbi.2022.104052

Markiewicz, C. J., et al. (2021). The OpenNeuro resource for sharing of neuroscience data. *eLife, 10*, e71774. https://doi.org/10.7554/eLife.71774

Martínez, R. (2018). Privacidad, datos de salud y tratamiento de datos sanitarios en el RGPD. *Revista Española de Protección de Datos, 3*, 1–32. https://doi.org/10.36151/reei.2018.09

Miller, R. A. (1994). Medical diagnostic decision support systems—past, present, and future. *Journal of the American Medical Informatics Association, 1*(1), 8–27. https://doi.org/10.1136/jamia.1994.95236141

Mittelstadt, B. D., et al. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Obermeyer, Z., et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Odom, J. V., et al. (2023). ISCEV Standard for clinical visual evoked potentials: 2023 update. *Documenta Ophthalmologica, 146*(3), 169–186. https://doi.org/10.1007/s10633-023-09926-9

Organización Mundial de la Salud. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization. https://www.who.int/publications/i/item/9789240029200

Pascucci, D., et al. (2022). VEPCON: Source imaging of high-density visual evoked potentials with multi-subject brain and connectivity data. *Scientific Data, 9*(1), 9. https://doi.org/10.1038/s41597-021-01116-1

Pinargote, C. A., Mero, J. G., Macías, V. D., Rodríguez, D., & García, M. A. (2025). Medicine of the future: Artificial intelligence, gene editing, and planetary health. *eVitroKhem, 4*, 213. https://evk.ageditor.ar/index.php/evk/article/view/213

Rodríguez, D., Chumbay, J. V., García, M. A., & Montero, Y. (2026). Tecnologías emergentes y aprendizaje universitario: revisión sistemática de su efectividad pedagógica y condiciones de implementación. *Revista San Gregorio, 1*(66), 120–136. https://doi.org/10.36097/rsan.v1i66.4039

Salas, L. E., Caamaño, S. R., Benalcazar, R. A., Turibbi, L. V., Quiroz, D. Z., & García, M. A. (2025). Artificial intelligence tools in the daily study of medical students. *EthAIca, 4*, 380. https://ai.ageditor.ar/index.php/ai/article/view/380

Shen, D., Wu, G., & Suk, H. I. (2019). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering, 19,* 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA, 320*(21), 2199–2200. https://doi.org/10.1001/jama.2018.17163

Sutton, R. T., et al. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *npj Digital Medicine, 3*(1), 17. https://doi.org/10.1038/s41746-020-0221-y

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Wilkinson, M. D., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3,* 160018. https://doi.org/10.1038/sdata.2016.18

Statement on the use of artificial intelligence: The authors acknowledge the use of generative artificial intelligence (AI) and AI-assisted technologies solely to improve the readability, language quality, and clarity of this manuscript. AI tools may also have been used to create or edit non-scientific graphical content, such as graphical abstracts, icons, illustrations, diagrams, or educational figures, provided that such use is transparently disclosed and does not misrepresent or fabricate research findings. AI tools were not used to generate, modify, enhance, interpret, or analyze scientific data, medical images (e.g., radiographs, computed tomography scans, magnetic resonance imaging, ultrasound images, histopathological images, microscopy images, or other diagnostic images), or any figures that constitute primary research data. The authors reviewed and edited all AI-assisted content as necessary and assume full responsibility for the accuracy, integrity, and originality of the final published work.

Disclaimer / Editor's Note: All publications' statements, opinions, and data are solely those of the individual authors and contributors, not *Revista Gregoriana de Ciencias de la Salud* or the editors. *Revista Gregoriana de Ciencias de la Salud* and/or the editors disclaim all responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.