NOVA Policy Network
Policy Brief

EU Artificial Intelligence Act & Neurotech

Integrating Neurotechnology into the EU Artificial Intelligence Act

<Murari Ambati>

Policy Brief

Date, October 4, 2024

Executive Summary

The European Union Artificial Intelligence Act (EU AIA) is a groundbreaking legislative proposal that seeks to regulate artificial intelligence based on a risk-tiered regime. Under this regime, AI systems posing significant risks to safety, fundamental rights, and democratic values are identified as "high-risk" and are subjected to more heightened scrutiny, transparency, and compliance obligations. However, while the AIA addresses a broad range of algorithmic technologies—such as biometric recognition and infrastructure automation—it does not intentionally recognize or classify neurotechnology, and more particularly AI-powered brain-computer interfaces (BCIs), within its regulatory scope.

Such omission is a grave error. Modern BCIs are increasingly leveraging machine learning and deep neural networks to decode mental signals, infer mental states, and modulate neural activity. They can translate thought into action, interpret desires straight from brainwaves, and occasionally alter behavior via feedback loops. Their dual-use potential—spanning medical restoration, cognitive enhancement, military application, and consumer interface—places them at the forefront of ethical intricacy and technological hazard. Despite all this, the neuro-AI sector is considerably underregulated, lacking harmonized standards for safety testing, cognitive consent, or neurodata protection.

This policy brief argues that AI-driven BCIs should be formally recognized as high-risk systems under the EU AIA. This would ensure that their use and development come under express transparency obligations, neuroethical safeguards, and adaptive risk management frameworks. The paper suggests creating a novel category of neuro-AI systems under Annex III of the AIA and outlines the procedural tools necessary to govern these technologies effectively. These include mandatory neuroethical impact assessments, consent models with greater specificity for cognitive autonomy, and real-time post-market monitoring for systems with the capacity for self-modification or inference drift.

Inaction on a preemptive basis risks eroding public trust in AI and systemic risk in fields that encroach on the most private recesses of human thought. As Europe tries to position itself as a global frontrunner for ethical AI legislation, it must realize that technologies that can read and write the human mind need to be regulated to the extent of their revolutionary promise.

Introduction

The European Union's Artificial Intelligence Act is a landmark effort to regulate artificial intelligence systems in a manner that is harmonious with democratic values, human rights, and public security. Anchored in a risk-based classification, the AIA ranges AI applications across four tiers—unacceptable, high-risk, limited-risk, and minimal-risk—with increasing requirements around transparency, accountability, and oversight. The Act has the potential to shape not only how AI is regulated in the EU, but also to provide global benchmarks for algorithmic regulation. Yet, as with any grand regulatory undertaking, the lines of what is explicitly included are as important as what is left out.

One such gap is the governance of neurotechnology, and in particular artificial intelligence systems integrated into brain-computer interfaces (BCIs). These systems—designed to decode neural signals, achieve bidirectional communication with the brain, and even regulate cognitive activity—have moved out of the lab and are now entering medical and consumer markets. Spanning from clinical applications such as restoring motor function in paralyzed patients to experimental consumer interfaces capable of reading intent and interpreting it into digital commands, BCIs are becoming ever more reliant on AI for extracting real-time inferences from high-dimensional, non-linear brain data.

Despite this development, BCIs are not obviously referenced in the AIA's annexes or definitions. This is a significant oversight. When powered by AI, BCIs are not merely assistive technologies but also potent tools capable of extracting mental states, manipulating thought, and influencing behavior. Their capacity to infer preference, anticipate decision, or modulate attention thresholds poses a fundamentally new category of risk—one that cuts across autonomy, mental integrity, data privacy, and the very notion of cognitive sovereignty. These systems undermine the traditional distinctions between machine learning and self, between inference and intention, between enhancement and manipulation.

Moreover, BCIs have a dual-use potential: the same algorithms intended to enable speech recovery may, in a different context, be applied for behavioral monitoring or cognitive surveillance. Given the EU's stated dedication to protecting fundamental rights, it is incumbent that regulatory regimes are extended to AI systems that engage with the brain. Lacking explicit coverage under the AIA, BCIs risk occupying a regulatory grey zone—too advanced to disregard, yet too novel to be neatly categorized.

This briefing contends that AI-driven BCIs need to be formally classified as high-risk systems under the AIA. It considers the current classification lacunae, the ethical and legal implications of unregulated neuro-AI systems, and offers a policy roadmap to integrate such technologies into the existing legislative paradigm. In doing so, it attempts to bring cognitive technologies into the regulatory narrative, before their societal impact becomes too vast to be regulated ex post facto.

Issues / Policy Gaps

Despite the extremely wide coverage of the European Union Artificial Intelligence Act (AIA), it does not yet possess an express category for neurotechnologies—specifically brain-computer interfaces (BCIs) that incorporate machine learning in neural signal decoding, inference, or modulation. This absence leaves a number of conceptual and legal loopholes that detract from the AIA's aim to safeguard elementary rights in high-risk contexts.

The most important and first conceptual concern is an absence of official AI-facilitated neurotechnology taxonomy. While the AIA refers to biometric identification, critical infrastructure AI, and emotion recognition systems, it has nothing to say regarding BCIs that read cognitive or affective states directly from the brain. As a result, regulatory clarity for AI systems that translate electroencephalographic (EEG), magnetoencephalographic (MEG), or electrocorticographic (ECoG) data into useful output is nonexistent. This silence deprives developers, users, and regulatory bodies of any understanding of whether these systems are in high-risk categorization, despite having profound implications for agency, privacy, and cognitive integrity.

Second, the AIA does not address the new category of risk from AI models that are capable of inferring internal mental states. In contrast to recognizing emotion from facial expression or voice, BCIs allow for direct access to cognitive processes such as intent, attention, memory recall, or emotional valence. Direct access to internal thought patterns based on probabilistic inference permits concern about invasion of mental privacy and potential for behavioral manipulation. These capabilities bleed outside the boundaries of conventional data processing into morally sensitive space—where merely inferring can be an act infringing on rights.

Third, neurodata is not yet protected by existing European privacy mechanisms. The General Data Protection Regulation (GDPR), as robust as it is, does not designate brain-derived information as a special category of sensitive data, nor does it contain regulation aimed at the neurotechnology environment. This imprecision creates inconsistency in consent, data minimization, portability, and ownership—particularly for neurodata used to train AI models or migrated to third-party sites. Lacking specially crafted protections, users can inadvertently give the deepest stages of their thinking to opaque or exploitative data systems.

Fourth, currently mandated safety and conformity standards in the AIA are not formulated to address the adaptive character of AI-based BCIs. Such systems tend to exhibit non-static behavior—learning, adapting to user input, neural variability, or context-dependent reinforcement. Such adaptive changes can engender inference drift, where the system's outputs no longer validate its intended design parameters. This risk is most problematic in systems deployed for neurostimulation or real-time feedback, where unintended alterations in AI behavior can precipitate cognitive harm, emotional dysregulation, or unwanted dependence.

Finally, there is no consistent requirement of neuroethical analysis in the development life cycle of BCI-AI. Even though the AIA imposes some human oversight and transparency measures for high-risk AI, these provisions are not aimed at considering the ethical nuances of technologies that directly engage with the human brain. Until now, there is no mandatory requirement for developers to assess impacts on mental autonomy, psychological well-being, or long-term neural plasticity. This gap not only puts users at risk but also undermines the EU's professed commitment to human-centered AI for dignity, rights, and justice.

In general, the AIA's silence on neurotechnology is indicative of a broader regulatory deficit

between legislative imagination and technical promise. Without transparent classification, risk modeling, and enforcement of ethics, AI-extended BCIs can go rogue right at the threshold between thought and computation.

Policy Recommendations

To address the regulatory loopholes currently present with the EU Artificial Intelligence Act (AIA) regarding neurotechnologies—specifically, AI-driven brain-computer interfaces (BCIs)—this note proposes a three-dimensional policy framework for openly recognizing and governing neuro-AI systems. The proposed recommendations integrate ethical safeguards, control mechanisms, and mathematical tools in order to estimate and mitigate neurocognitive harms from algorithmic interfaces with the human brain.

To start this, Annex III of the AIA, which specifies high-risk use cases, must be altered to include a specific category by the term "Neuro-AI Systems." It should include any AI system that performs decoding, inference, or modulation of neural activity from EEG, MEG, ECoG, or other brain activity modalities. The labeling of BCIs as high-risk AI acknowledges the epistemic asymmetry and behavior malleability that governs neural interfacing, and offers a legal trigger for enhanced conformity assessments.

Secondly, authors of Neuro-AI systems would be required to offer a Neuroethical Impact Assessment (NIA), derived from the Data Protection Impact Assessment (DPIA) found in the GDPR. The NIA would determine cognitive autonomy risk, informed consent strength, data vulnerability, and long-term psychological integrity. The following measure could be used to quantify cognitive autonomy risk in algorithmic inference contexts:

$$\mathrm{CAR}_t = \delta \cdot I(t)^2 + \lambda \cdot \frac{A(t)}{C(t)} + \theta \cdot \frac{1}{P(t)}$$

$I(t)$ is the inference depth over time (e.g., number of mental states inferred), $A(t)$ is the frequency of algorithmic intervention per user session, $C(t)$ is the cumulative user control index, $P(t)$ is the perceived predictability of the system by the user, δ, λ, θ are adjustable weight parameters determined by context and regulatory thresholds.

High values of CAR_t would indicate a degradation of user cognitive sovereignty and should trigger either system redesign or pre-market rejection.

Third, the AIA's current focus on static risk classification must be expanded to encompass adaptive risk regimes. Neuro-AI systems often employ reinforcement learning or neural adaptation, where user behavior and brain activity recurrently remodel the model on a time axis. This creates a form of temporal uncertainty not presently addressed in conformity protocols. Theoretical dynamic systems analysis frameworks predict that long-term loop convergence between AI and user is not guaranteed. Contrarily, post-market surveillance would require periodic recalibration of the accuracy of behavioral inference by the system and its CAR\textsubscript{t} trajectory. Regulatory audits must be triggered by either:

- increasing variance of prediction error,
- shifting distribution of user attention,
- drift in inferred emotional valence through comparison with baseline profiles of users.

Fourth, the EU should establish a Neurotechnology Coordination Taskforce, coordinated by the EU AI Office, European Data Protection Supervisor (EDPS), and the Council of Europe's

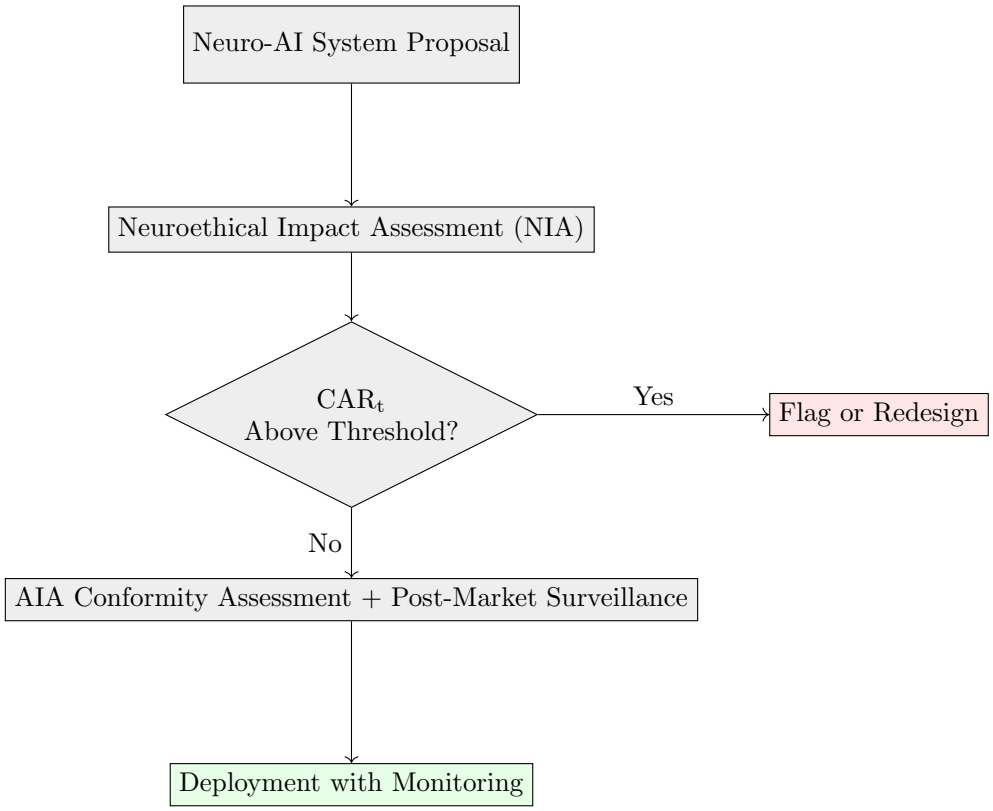

Figure 1: Proposed EU Oversight Pipeline for AI-Powered Brain-Computer Interfaces

Committee on Bioethics (DH-BIO). The Taskforce would issue periodic technical recommendations on neural data classification, cognitive coercion thresholds, and transboundary neurotech enforcement. The Taskforce should also maintain a "living list" of classified neuro-AI applications, ranging from therapeutic BCIs to consumer-grade neurostimulators.

These multilayered approaches ensure that the AIA's high-risk AI label is not merely responsive, but proactive—capable of engaging with the fundamental epistemological and moral challenges that come with machines entering the human mind.

Conclusion

As artificial intelligence systems begin to surpass the boundaries of human thought, the need for anticipatory regulation becomes not merely important, but imperative. Brain-computer interfaces (BCIs), which were once confined to clinical rehabilitation settings, now are increasingly becoming general-purpose interfaces that decode mental thought, control behavioral output, and infer states of mind. Combined with adaptive AI, these technologies create a novel epistemic frontier—a world in which machines no longer merely observe human behavior, but interpret and shape it from the inside out.

The AIA as currently drafted provides a solid basis for rights-based algorithmic regulation, but its failure to address AI-driven neurotechnology is a structural blind spot. By not classifying neuro-AI systems explicitly, they fall outside the very regulation meant to safeguard autonomy, integrity, and transparency. The exclusion of BCIs from high-risk categories is due to a failure of conceptual readiness rather than a failure of relevance.

This paper has outlined a strategy to bridge that gap: formal Annex III inclusion of Neuro-AI systems, the creation of a measurable cognitive autonomy risk indicator ($\mathrm{CAR_t}$), work on Neuro-ethical Impact Assessments, and the formation of an inter-sectoral Neurotechnology Coordination Taskforce. Together, these reforms would not only make the EU a world leader in AI ethics, but the first jurisdiction to have legislated for algorithmic systems to operate within the boundaries of mental life.

Lastly, neurotechnology control is not a matter of technical compliance—it's a matter of constitutional sovereignty of the mind in the future. By acting swiftly, the EU can shape that future and ensure that just as the machines are converging on the mind, the law converges on the future that lies ahead.