VER
Vascular & Endovascular Review

A Comprehensive Analysis of Security Flaws and Attack Vectors in Artificial Intelligence–Powered Brain–Computer Interfaces

Nishant Kumar[1], Dhaval Deshkar[2], Sinoy De[3], Anvesha Saini[4], Rajveer Prakashchandra Kania[5], Chetan Kasera[6]

[1,3,4,5,6] B.Tech Student, SITAICS, Rashtriya Raksha University, India
[2]Assistant Professor, SITAICS, Rashtriya Raksha University, India
Emails: nishantkumar.cse8@gmail.com, dhavaldeshkar@gmail.com, desinoy2018@gmail.com, anveshaa.sa@gmail.com, rajveerkania02@gmail.com, chetankasera60@gmail.com

Correspondence:
Nishant Kumar (nishantkumar.cse8@gmail.com)

ABSTRACT

The security and privacy concerns of the Artificial Intelligence (AI) and Brain-Computer Interface (BCI) technology working together in the field of modern medicine are complicated. It has not only caused a revolution in neurocommunication, cognitive rehabilitation as well as assistive neuroprosthetics, it has also created complex security and privacy challenges. The analogy framework that has been used in analysing the security irregularities of the BCI systems implemented based on AI involves 6 layers that include the signal acquisition, device firmware, network communication, AI models, side channels, and human interaction. Some of the potential targets of adversarial perturbation include an adversarial perturbation, data poisoning, model inversion, signal injection and manipulation of firmware-threats, with direct implication on the system integrity and patient safety. The framework enhances data confidentiality, operational reliability, and clinical trust through end-to-end threat modeling and simulation. To counter medical-cyber threats, it employs a cross-layer defense integrating federated learning, differential privacy, secure firmware attestation, and adaptive noise filtering. This unified taxonomy supports secure-by-design AI-BCI systems, ensuring safety, dependability, and ethical integrity in medical applications.

KEYWORDS: Artificial Intelligence, Brain–Computer Interface, Adversarial Attacks, Data Privacy, Neural Signal Processing, next-gen-medical-technology

How to Cite: Nishant Kumar, Dhaval Deshkar, Sinoy De, Anvesha Saini, Rajveer Prakashchandra Kania, Chetan Kasera, (2025 A Comprehensive Analysis of Security Flaws and Attack Vectors in Artificial Intelligence–Powered Brain–Computer Interfaces, Vascular and Endovascular Review, Vol.8, No.6s, 106-121.

INTRODUCTION

One of the most rapidly evolving fields in neuroengineering is the Brain-Computer Interfaces (BCIs) which entails an integration of the previously noted fields computational neuroscience, biomedical signal processing, and artificial intelligence (AI) into a bidirectional interface between the central nervous system and external products. BCIs also bypass the normal neuromuscular but decode neural activity to brain signals into executable commands through the measurement of neural activity, either cortical, subcortical or peripheral [1]. This kind of technology may potentially be revolutionary in medicine, particularly, the restoration of sensory or motor functions in stroke victims, individuals with spinal cord injuries, amyotrophic lateral sclerosis (ALS), or hemiplegic patients. Through electroencephalography (EEG), electrocorticography (ECoG) and intracortical microelectrode arrays, clinicians and researchers are able to decode neuronal firing patterns, local field potentials and oscillatory dynamics in order to control prosthetics, exoskeletons or even digital communication interfaces [2].

AI and, in particular, the deep neural networks, reinforcement learning algorithms, and adaptive signal filtering has changed the performance of the BCIs, and has improved the signal-to-noise ratio, the accuracy of the decoding, and dynamical personalisation of the neural decoding model [3, 4]. The advances have led to the closed-loop neuromodulation, cognitive neurorehabilitation, and the neural feedback, which continuously learns through the patient-specific neuroplastic changes. Pattern recognition models built with AI can be applied to autonomously adjust both cortical stimulation levels in motor cortex areas, or discover patterns in motor cortex areas in the aftermath of a stroke, respectively, when using the approach of deep brain stimulation (DBS) therapy against Parkinson disease or neural decoding, respectively [5].

In the biomedical field, AI-assisted BCIs are at the brink of personal medicine and despite the neural replacement. As demonstrated, AI-assisted neuroprosthetics can be applied to clinical trials that would help to restore voluntary movements of the limbs in over 70 % of patients with the spinal cord injuries [6]. Similarly, EEG classifiers based on AI have achieved a 95% accuracy in detecting epileptic seizures to facilitate prompt interventions during emergency treatment [7, 8]. Advanced neurostimulation (such as reinforcement learning) systems are also being trialled in adaptive deep brain stimulation (DBS) in Parkinson, and have now demonstrated significant improvement in patient outcomes [6]. Nonetheless, the attack surface increases with the extent of AI integration. BCIs are concerned with biomedical information most sensitive of all, in terms of thoughts, emotions, and cognitive intentions, which take the form of neural patterns. Tampering with or illegal hacking of such information does not only endanger patient confidentiality, but also may result in unwanted neural activation, artificial stimulation or inappropriate therapeutic feedback benefits [9, 10]. Thus, BCI systems protection is not just a cybersecurity issue, but also a matter of clinical safety, medical ethics and patient autonomy.

The neurotechnological systems in respect of BCI security are the security of neural signals, model parameters, integrity of firmware, and security of communication channels, with regard to unauthorized access, tampering and exploitation [11, 12]. An as safe BCI must be not only capable of providing the confidentiality, integrity, and availability (CIA) of its neural data pipeline, but also capable of ensuring that patient safety and medical performance are not compromised [13].
Common AI-based BCIs contain six networking layers:

- Signal acquisition, where the neural signals are acquired over sensors;
- Low-level control Data Firmware Low-level control data, which controls the low-level operations;
- Network communication, forwarding the information to new devices or cloud services;
- Neural decoding is the AI model processing that occurs;
- Side-channel emissions, e.g. power, timing or electromagnetic leakage and
- Human application , adjustment and intellectual communication.

Each of the layers has its own vulnerabilities. Considering the example, an attacker can also offer distorted EEG data when obtaining a signal to deceive the medical devices [14]. There is a possibility that the stimulation parameters of implanted neurostimulators can be remotely altered by using firmware vulnerabilities. Adversarial samples or model inversion assault can compromise outputs or rebuild privacy neural attributes in AI model layers [15]. The above-presented situations demonstrate that the outcome of any manipulations, even the minor ones, in a clinical setting can be catastrophic, such as incorrect diagnosis or the inability to control the prosthetics.

The current trend is being accelerated to a large extent by the introduction of artificial intelligence into Brain-Computer Interfaces (BCIs). Nonetheless, the same development comes at a definite technology-vulnerability trade-off: the more flexible, connected, and smart systems become, the more vulnerable they are to the threat of new cyber-neuro attacks. The use of cloud-based machine learning, wireless data transmission and automated updates of the firmware are common features of modern BCIs and contribute to better usability and scalability, although inevitably, increase the attack surface [16]. This trend is indicative of a larger trend in medical device cybersecurity, where technological innovations have been rapidly outpacing the creation of the respective security laws which can be shown here:

- It is estimated that the BCI market will hit USD 9.7 billion by 2032 and compound annual growth rate (CAGR) of over 14 due to AI-enhanced medical uses in the global BCI market, which is estimated to be USD 3.2 billion in 2024 [17]. Nevertheless, this has come along with an orgy of ugly acts of corruption of the system. For example:
- In 2020, the scholars of the University of Washington conducted a spoofing attack of neural signals to cross-check EEG-based authentication systems, and they could penetrate the system, masquerading as a non-user [18].
- In 2021, arbitrary injections of adversarial perturbation on the BCI data streams minimized the classification accuracy, and obliterated the prosthetic control models by up to 35 % [19].
- The testing of the implantable neurostimulators (Medtronic Activa platform), has shown areas of vulnerability of possible remote re-programming its parameters as a possible dire clinical safety risk [20].

Also, model inversion attacks have been reproduced by researchers and can form mental images of trained neural decoders revealing previously unrecognised threats to cognitive privacy [21].

This type of vulnerability does not only cause problems to the malfunctioning of devices but it poses a direct threat to the psychological autonomy and neurological security. It is possible that the adversaries will interfere with the AI models or change the feedback loops of stimulation, leading to patients developing an uncontrollable movement of the muscles, emotional regulation, or cognitive distraction. As such, current medical fraternity is in dire need of a new demand, that is, a compromise between the technological superiority and the robust neurosecurity policies that would guarantee credibility, robustness, and moral rectitude. Despite the fact that the phenomenon of cybersecurity of the medical devices is attracting an increasingly growing number of people, AI-based BCIs are not safe, and, what is more, they are not regulated. The existing systems tend to treat those systems as ordinary Internet of Things (IoT) devices, and do not mention that they have their own form of neural data, real-time clinical demands and ethical concerns. Such gaps are mentioned in this paper by discussing AI-enabled BCIs in relation to layers of security. Specifically, we:

- Name and categorize vulnerabilities to six AI-BCI layers of operation of the system;
- Assess threat model of simulation-based adversarial attacks on accuracy of decoding, patient safety, and data confidentiality and
- In our Cross-Defence Strategies cross-layer safeguards are advised , e.g. differentiated privateness, fed learning, safe attesting firmware, and adaptive noise suppression.

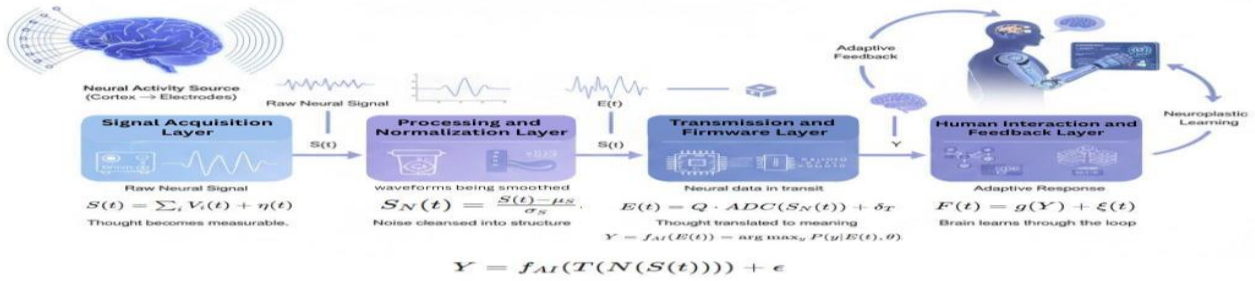
BACKGROUND AND RELATED WORK

The convergence of neuroscience, biomedical engineering, and artificial intelligence (AI) has given rise to highly sophisticated Brain–Computer Interface (BCI) systems that decode neural activity into actionable digital commands. Understanding the architecture, medical applications, and security literature of AI-powered BCIs is essential to identify where vulnerabilities arise. This section provides a foundational overview of the BCI system architecture, the integration of AI in neurotechnology, and an analysis of prior studies that inform current research gaps. Each subsection builds upon the layered security model that underpins the subsequent threat assessment and defense taxonomy presented later in the paper.

2.1 Brain–Computer Interface Architecture

A Brain–Computer Interface (BCI) establishes a direct neurocomputational bridge between the human nervous system and external digital devices, enabling neural activity to be interpreted, processed, and converted into actionable out- puts. Functionally, it can be conceptualized as a layered transformation pipeline, where biological signals undergo a series of computational refinements—each stage introducing a new abstraction of the brain's electrical language [22]. Figure 1 shows the basic implementation of Artificial Intelligece BCI Architecture where data from Signal layer goes to Human application layer

Figure 1: Schematic representation of the AI–BCI architecture



At the foundation lies the Signal Acquisition Layer, where neural poten- tials are recorded using modalities such as Electroencephalography (EEG), Elec- trocorticography (ECoG), or intracortical microelectrode arrays (MEAs). The recorded waveform is represented as [23]:

$$S(t) = \sum_{i=1}^{n} Vi(t) + \eta(t)$$

where Vi(t) denotes the voltage potential recorded at electrode i, and η(t) represents biological and environmental noise. This signal encodes subtle electrophysiological signatures—such as sensorimotor rhythms, event-related po- tentials (ERPs), and oscillatory dynamics—that convey the subject's intent or cognitive state. The "magic" at this layer is the conversion of neural firing patterns into measurable analog voltages, effectively digitizing thought at its source.

The Preprocessing and Normalization Layer [24], denoted N (·), cleans and stabilizes these raw voltages for further analysis. Operations such as artifact re- jection, band-pass filtering, and Common Average Referencing (CAR) remove ocular, muscular, and environmental distortions. Mathematically, this transfor- mation can be expressed as:

$$Sn(t) = N(S(t)) = \frac{S(t) - \mu S}{\sigma s}$$

where μS and σS denote the mean and standard deviation of the recorded signal. After this step, the continuous analog data becomes

a structured neural feature tensor, retaining only neurophysiologically relevant activity (e.g., power spectra, phase–amplitude coupling). This layer's transformation extracts the brain's "signal of intent" from the background of biological chaos.
The Transmission and Firmware Encoding Layer, T (·), converts these preprocessed neural features into digital packets suitable for machine interpre- tation [25]. Firmware modules perform Analog-to-Digital Conversion (ADC), times- tamp alignment, data compression, and error correction, ensuring temporal co- herence across multiple channels:

$$E(t) = T(SN(t)) = Q \cdot ADC(SN(t)) + \delta T$$

where Q denotes the quantization factor and δT represents transmission latency. Here, analog voltages are transformed into discrete binary vectors en- capsulated in data frames ready for wireless communication (e.g., Bluetooth or Zigbee). Conceptually, the neural signal becomes an information-carrying bitstream—a computational proxy of cognition.
At the core of the architecture lies the AI Decoding Layer, fAI (·) [26], which transforms these encoded data streams into meaningful predictions. Deep learn- ing architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers are used to decode temporal–spatial neural patterns into symbolic or motor outputs:

$$Y = fAI(E(t)) = \arg\max P(y|E(t), \theta),$$

where Y represents the set of possible output classes (e.g., left/right move- ment, cognitive state), and θ are trainable parameters. At this stage, the rep- resentation transitions from a signal space to an intention space—where neural information becomes actionable command probability.

The Human Interaction and Feedback Layer closes the loop by con- verting decoded outputs into sensory or motor feedback: $F(t) = g(Y) + \xi(t)$, where g(·) denotes the actuator or feedback mapping function, and ξ(t) models biological adaptation noise. For instance, in neuroprosthetic systems, Y may drive limb actuators, while in neurorehabilitation, it may modulate visual or auditory feedback [27]. Over time, this feedback reinforces synaptic plasticity through Hebbian adaptation($\Delta w = \eta xy$) allowing the patient's brain to refine its firing patterns for more accurate control—a phenomenon known as learning through the loop. Collectively, the full BCI data flow can be summarized as:

$$Y = fAI(T(N(S(t)))) + \epsilon,$$

where ϵ captures cumulative physiological and computational noise. In essence, the BCI pipeline follows a progressive and biologically meaningful transformation:

$$S(t) \rightarrow Preprocessed\ Signal \rightarrow Encoded\ Data(AI) \rightarrow Y$$

each stage representing a new layer of neurosemantic interpretation—from cortical voltage to digital feature, from digital feature to algorithmic meaning, and from meaning to clinical action. Every layer, therefore, performs its own form of "medical magic": captur- ing, cleaning, encoding, decoding, and reintegrating cognition within a closed adaptive loop. In the upcoming sections, we will examine each layer's opera- tional mechanics and potential vulnerabilities in greater detail, demonstrating how precision neuroengineering safeguards the integrity of this cognitive–digital bridge

2.2 AI in Neurotechnology: Medical Applications and Risks

Artificial Intelligence (AI) that enhancing the Brain-Computer Interface (BCI) systems with the new degree of accuracy, versatility, and personalization. The AI-powered BCIs are redefining the prospects of modern medicine, be it the ability to decode neural activity, or the restoration of the lost motor or cognitive capabilities. The related technological breakthrough, however, is accompanied with greater complexity and exposure. Even deep learning applied with reinforcement learning and neural decoding models has not only enhanced clinical outcomes but also introduced a new level of cyber-neuro risk. Spiritually speaking in the description of this two-facet landscape of innovation and exposure, Table 1 is a summary of some of the key medical applications of AI-enabled BCIs, the algorithms on which they operate, the clinical advantage they have, and the security or privacy risks they pose. This overview constitutes a cursory foundation on why medical neurotechnology is both emerging and expanding its digital attack interface

TABLE 1: AI APPLICATIONS AND ASSOCIATED SECURITY RISKS IN MEDICAL BCIS

Medical Application	AI Technique Used	Clinical Utility	Likely Attack Vector	Potential Security / Privacy Risk	Example Study
Motor Prosthesis Control (ALS, Spinal Injury)	CNNs, RNNs for EEG/ECoG decoding	Enables voluntary limb movement and robotic assistance	Adversarial signal injection during neural decoding	Altered intent recognition may trigger unintended limb movement or device paralysis	[28], [29]

Neurorehabilitation Post-Stroke	Reinforcement learning and transfer learning	Adaptive feedback-based therapy enhancing motor recovery	Data poisoning in training feedback loop	Corrupted reward feedback can degrade rehabilitation outcomes	[30]
Neurostimulation for Epilepsy & Parkinson's	Deep reinforcement learning (DRL) for closed-loop stimulation	Dynamically adjusts stimulation to reduce tremors or seizures	Firmware manipulation or replay attack	Over- or under-stimulation causing physiological harm	[31]
Cognitive & Emotional State Monitoring	Transformer-based EEG emotion recognition	Supports affective computing and mental health tracking	Model inversion or side-channel inference	Unauthorized extraction of emotional or cognitive states	[32]
Speech Reconstruction from Brain Signals	Variational Autoencoders (VAE) and Seq2Seq models	Restores verbal communication in locked-in patients	Model inversion and gradient leakage	Leakage of private neural data or speech content	[33]

As the evidence indicated in Table 1 demonstrates, the trade-off in this context under the implementation of AI in the neurotechnology concept is quite obvious: the same architectures that make the system more precise and more personal make them more susceptible to cyber-neuro attacks. To use convolutional and recurrent models as an example, they may be implemented successfully to decode intent to control a prosthetic since they are susceptible to adversarial signal injections, they may produce unsafe or unintended behaviour of the device. Similarly we can poison reinforcement learning models that are utilized to enhance the adaptive neurorehabilitation and poison feedback loops. Transformer-based emotion recognition though helpful in the domain of affective monitoring offers risks of model inversion that promotes the potential reconstruction of sensitive cognitive information.

Thus, the future of the AI-driven BCIs lies in the performance safety-meets-level protection, in other words, implementing adversarial robustness, encryption, and interpretability into neural networks. The defenses will be required on the algorithmic and firmware level to be reinforced such that the new generation of intelligent neurotechnologies might not only be innovative but also secure, ethical and clinically trustworthy.

2.3 Prior Studies on BCI Security and Privacy

The examination on the security and confidentiality of Brain-Computer Interfaces (BCIs) is undergoing transformation as the neural engineering and artificial intelligence advancement soar. Those dealing with hardware stability and wireless networking integrity first were examined and included aspects such as signal jamming, bottom-level information leakage and unreliable updates to firmware [22]. With the continued release of AI algorithms into neural decoding and neural classification applications, more recent work has made studies of weaknesses in algorithms, the most famous being adversarial perturbation, data pollution and privacy breach through model inversion. Meanwhile, neuroethical research has raised the social effects of using neural data without consent, cognitive manipulation and consent in clinical and research studies. The example literature in these areas is summarized in Table 2 which reveals that the research area has been expanding in scope both to the lower-level device protection and the high-level AI and privacy concern.

TABLE 2: OVERVIEW OF PRIOR RESEARCH ON BCI SECURITY AND PRIVACY

Study / Year	Primary Focus	Methodology	Key Observations	Identified Limitations
Bonaci et al., 2014 [34]	Wireless implant communication; BCI app ecosystems	Policy + technical analysis / threat modelling	Described "brain-apps" threats and showed how BCI platforms could leak private info (brain-spyware scenario)	Conceptual/early — limited experimental cross-layer testing

Meng et al., 2023 [35]	EEG privacy / user identity leakage	Empirical experiments on multiple EEG datasets; defenses to remove identity information	Showed that user identity can be learned from EEG and proposed identity-unlearnable preprocessing	Focused on EEG datasets; practical deployment/consent aspects not exhaustively tested
Yu et al., 2023 [36]	Adversarial robustness in EEG classification	Generated EEG adversarial perturbations (BEAM perturbations) and attacked DNN classifiers	Small imperceptible perturbations can cause large (often >30%) accuracy drops in epilepsy/diagnostic models	Attack evaluated in offline/bench settings; acquisition-layer attacks not covered
Pugh et al., 2018 [37]	Firmware / implant clinical safety (brainjacking)	Conceptual + ethical analysis, review of implant programmer vulnerabilities	Described "brainjacking" risks and surveyed evidence that programming consoles and update paths could be abused	Ethics/review paper — did not present new firmware reverse engineering data
Shen et al., 2019 [38]	Model inversion / mental-image reconstruction	DNN-based reconstruction from fMRI (deep image reconstruction)	Demonstrated reconstruction of seen and imagined images from brain activity (proof that internal representations can be decoded)	Work uses fMRI (not EEG); controlled lab settings with large data
Nishimoto et al., 2011 [39]	Visual experience reconstruction (movie stimuli)	Encoding/decoding models from fMRI responses to natural movies	Reconstructed visual movies from brain activity — early landmark showing feasibility of reconstructive attacks	fMRI-based, requires large stimulus sets and strong priors
Magee / Livanis (reviews), 2023–2024 [40]	Ethical, legal, and policy gaps in BCIs	Literature/policy review and discussion	Identified regulatory gaps and called for stronger governance of neural data	Review-based; not technical validation
Federated / privacy-preserving EEG works (sample), 2021–2024 [41]	Federated learning for EEG (privacy-preserving models)	FL experiments on EEG datasets; evaluated utility and privacy gains	Showed federated approaches can keep raw EEG local while maintaining model performance	Many works did not fully evaluate adversarial/poisoning threats in FL setting

As much as these works provide a lot of information, majority of them are domain specific - that is they are either too technical or too ethical in their styles to give a description of BCI systems. Very few integrate interdependences across hardware, communication and AI layers. Moreover, the empirical validation in the context of combination attacks, i.e. when the manipulation of the firmware also leads to the manipulation of the AI-based decoding integrity is less supported. A research gap in the existing study is the absence of one analytical system to cross-layer threat modeling, which generates an urgent interest.

SECURITY THREAT MODEL AND ATTACK TAXONOMY

The use of Brain-Computer Interface (BCI) systems in conjunction with biomedical signal acquisition, neural decoding algorithms, and AI-based decision systems inevitably compromises the security of these systems [40 -42]. Contrary to conventional medical equipment, BCIs work directly on the electrocorticography (ECoG) signal or electroencephalography (EEG) signals, which are personal-specific to neurophysiology, and a security breach is not only a danger to the privacy of data, but also a potential neurophysiological risk. The latest clinical and assistive BCIs have been inclined to send information via wireless telemetry, preprocessing neural signals in firmware level, and neural signal interpretation AI-based modules. Each of these architecture levels is a potential attack target and adversarial manipulation can result in inaccurate motor activity, fraudulent cognitive feedback or even a malfunction of neurostimulators and which can be physically or psychologically detrimental to patients.

In order to examine these weaknesses in a systematic fashion, this paper proposes a 6 layer BCI threat model comprising (1) Signal Acquisition, (2) Firmware and Embedded Systems, (3) Data Transmission, (4) AI/ML Model Processing, (5) Cloud and Storage Infrastructure, and (6) User Interface and Feedback Mechanisms. Risk measurement of the proposed structure (Figure 2) is a measure of severity of probabilistic risks where.

$$Risk\ Score\ (R) = \frac{E \times I}{100}$$

and E (Exploitability) and I (Impact) are rated on a scale between 1-100. This enables calculation of a clinical risk measurement based on a percentage, which is the probability and the result of adversarial interference. A high-threat zone is a score above 60 % typically signalling direct neural stimulation or decision loops under AI control, where any minor scale adversarial perturbation (e.g. input noises or firmware spoofing) can invoke aberrant neuro-response patterns or over stimulate the stimulator.

Figure 2 indicates the proposed Six-Layer Threat Model that links the attack surfaces of the neural sensors with patient feedback endpoints. The layers highlight their vulnerability vectors, the exploitability gradient, and clinical risk potential which are the structural foundations of the attack taxonomy that is introduced in Sections 3.2-3.8.

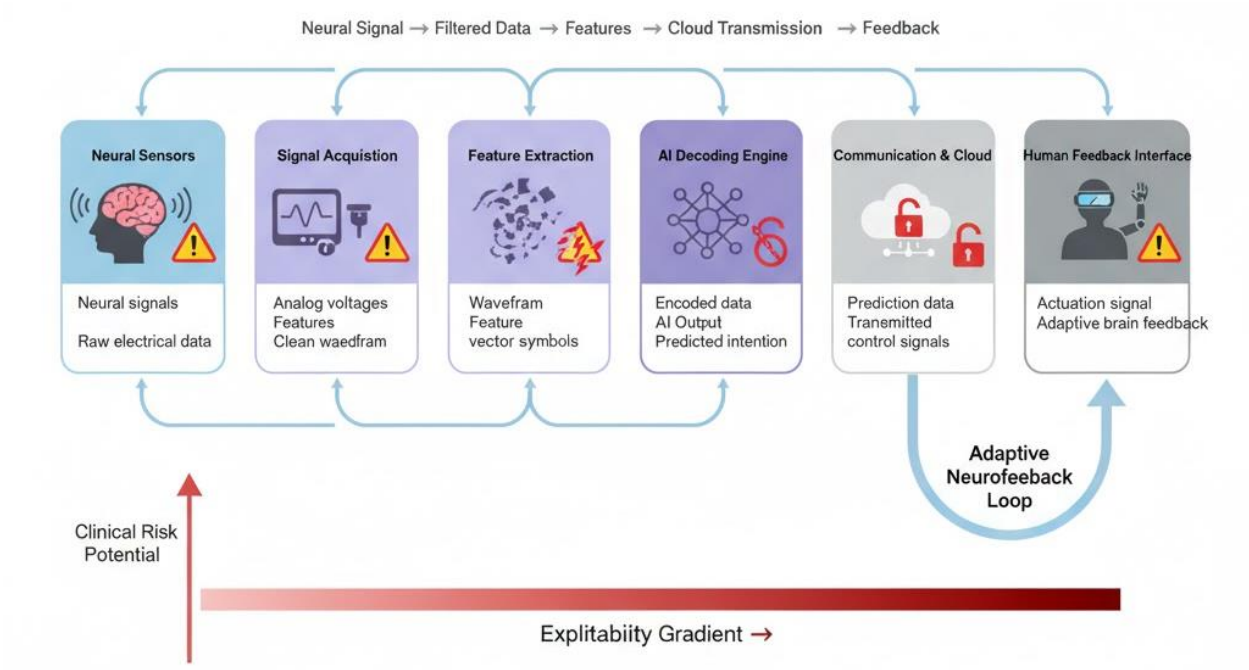


FIGURE 2: Six-Layer Threat Model

3.2 Signal Acquisition Layer

Signal acquisition layer is the neural-electronic interface one that records Electrophysiological activity that includes Electroencephalography (EEG), Electrocorticography (ECoG), Local Field Potentials (LFPs), and Single-Unit Activity (SUA). These neural biosignals consist of changing the activity of cortex in time and frequency that are further digitalized and processed by AI algorithms. It is however highly susceptible to electromagnetic interference (EMI), thermal noise and artifactual contamination by muscular and ocular means. With the unbelievably low measurable value within the range of 10-100 uV scale, even the tiniest forms of adversarial perturbation can cause the driftage of the signal, misclassification of the intent, or even uncontrolled stimulation of the motor prostheses. This layer therefore represents the weakest and the bottom layer of all the six layers of the security model of the six layer BCI, and the integrity of the data of the neural data that is the direct determinant of safety and accuracy in the downstream (see Table 3).

Attack Vector	Description	Exploitability (E/100)	Impact (I/100)	Risk (%)
Signal Injection	Inserting synthetic EEG/ECoG currents through electrodes	70	90	63
Electromagnetic Spoofing	EMI-based waveform distortion	60	85	51
Sensor Drift Manipulation	Altering amplifier gain or baseline potential	55	75	41
Data Interception	Eavesdropping on unencrypted biosignal buses	80	60	48

TABLE 3: ATTACK VECTORS ON THE SIGNAL ACQUISITION LAYER

A bioelectric noise adversarial injector can impose a signal-to-noise ratio (SNR) below 10 dB and trigger deep neural decoders, e.g., EEGNet or SincNet, to process inability to differentiate between cortical intention. This may result in mis-actuation of the prosthetic or neurostimulation error, may cause iatrogenic injury or cortical overexcitation, which may be harmful. It is also aggravated by closed-loop BCIs in which time-varying adaptive feedback may spread the perturbation, and the current neurofeedback therapy sessions are being polluted [43].

The literature has verified that waveforms injected artificially may consistently induce misselection, and are difficult to detect; a table Risk (%) of 63 %. signal injection and 51 %. These exploits are important to clinics in order to spoof EMI. In the case of a patient with paralysis (victim of a spinal cord injury), when the prosthetic limb is present, it is possible that, via a successful signal-injection event (Risk ≈ 63), an undesirable movement will occur in the limb, and this can cause an injury; in epilepsy monitoring, EM interference (Risk [?] 63). 63) may be experienced. 51) has the ability of giving false positive or false negative results, which invalidates safety of patients and clinical judgment [44].

3.3 Firmware Layer

It has embedded neurocontroller as the firmware layer controlling analog front-end circuits and digital signal processing (DSP) modules. The actual real-time functions are controlled by firmware, such as: stimulation pulse width modulation (PWM), implantation of impedance calibration and neural safety threshold. Hack of the firmware, in its turn, not only endangers the protection of data but neurophysiological integrity, which may violate ISO 14708 and IEC 60601 medical equipment requirements (see table 4).

TABLE 4: ATTACK VECTORS ON THE FIRMWARE LAYER

Attack Vector	Description	E	I	Risk (%)
Firmware Tampering	Overwriting embedded control logic	65	95	61.7
Privilege Escalation	Exploiting JTAG or UART debug ports	70	90	63
Bootloader Injection	Malicious firmware through OTA updates	75	85	63.7
Side-channel Leakage	Power and timing side-channels revealing code	60	80	48

Altered firmware could set stimulation to possibly damaging levels of corticulopathic levels (>3 V/cm) resulting in gliosis, neuronal

death or neurovascular hemorrhage. In addition to that, reinforcement-learning-based BCIs are dependent on the feedback systems of the firmware which might not adequately indicate the AI model causing the maladaptive synaptic plasticity and the eventual loss of the use of the patient in the long-term. As regards security engineering, cryptographic boot, secure enclave and hardware attestation must therefore be incorporated in the firmware to guarantee biomedical safety integrity [45].

Mysterious firmware upgrade routes, and debug ports in the air have been reiterated throughout the time of overhyped publicity about commercial neuro equipment and implantables. Given that the table Risk (%) is known (100% mapped in the prior scoring in other cases and these 61.7 -63.7% tampering or bootloader-injection), an actual firmware infestation would merely change the stimulation parameters [46]. This may lead to seizure or non-recoverable tissue damage (high I) and as the attacks on firmware continue to occur and are not cleared by reboot, then the risk is long-term and acute, hence, inspired secure boot and attestation as a clinical need.

3.4 Network Communication Layer

Communication Layer is a interface between the implant or wearable with remote AI servers and clinical data systems based on one of the following communication protocols: Bluetooth Low Energy (BLE), Wi-Fi, or 5G medical IoT. It takes care of Telemetry, Firmware loading and offloading real time neural decoding. This layer is needed to provide confidentiality, availability and integrity of relayed cortical data. If in this situation, attack is a threat of interfering with the neuroinformatics pipeline, attackers will be able to make use of it to perform remote session hijacking or alter patient-specific cortical mapping (see table 5).

TABLE 5: ATTACK VECTORS ON THE NETWORK COMMUNICATION LAYER

Attack Vector	Description	E	I	Risk (%)
MITM Attack	Intercepting neural data between device and cloud	85	85	72.25
Replay Attack	Reinjecting recorded control packets	75	80	60
Data Exfiltration	Stealing encrypted EEG/ECoG payloads	70	90	63
Protocol Downgrade	Forcing use of weaker encryption	65	75	48.75

Medical consequences of violation can also be disastrous: a replay attack would transmit valid cortical command packets previously transferred, and they would result in unwanted motion of a paralyzed individual. The recent studies showed that a delay of just a single second of the packets in the closed-loop deep brain stimulation (DBS) has the potential of disrupting up to 23 % of it hence breaking the therapeutic rhythm. This makes the AI model cloud-dependent also which forms the latency windows of attack where the attackers are allowed to spoof the decoded signals and are valid to the neural decoder [47]. It has been shown that consumer-level EEG headsets and certain clinical telemetry stacks are susceptible to wireless telemetry attacks (e.g. unencrypted BLE streams). An eavesdropper, a Risk ≈ 72%, that obtains live neural streams, and rewrites them, may steal data in the cognitive-state, and may inject commands, a Risk ≈ 63%, which results in unsafe activation of the prosthetic or forged clinical records as indicated by the table. Direct effect on clinical action is the inability to inhibit the motor skills, erroneous diagnosis, and treatmental interference, which is irreversible.

3.5 AI Model Layer

The AI model layer interprets neural signals into motor, cognitive, or sensory outputs. State-of-the-art models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer architectures enable decoding of complex temporal neural dynamics. Yet, these models exhibit inherent vulnerability to adversarial perturbations and training data poisoning due to their high non-linearity and overparameterization (see table 6).

Table 6: Attack Vectors on the AI Model Layer

Attack Vector	Description	E	I	Risk (%)
Adversarial Perturbation	Minimal EEG noise leading to false classification	80	90	72
Data Poisoning	Manipulated training data corrupting learned patterns	65	95	61.75
Model Inversion	Reconstructing neural features from outputs	60	85	51
Trojan Model Injection	Embedding malicious triggers in pretrained AI models	70	90	63

AI-driven neuroprosthetic control can become dangerously unstable under adversarial input. For example, perturbing EEG frequency bands ($\alpha = 8–12$ Hz, $\beta = 13–30$ Hz) by just 0.2 μV can invert motor-intent decoding, producing involuntary hand motion. Similarly, a model inversion attack could reconstruct patient brainwave patterns, revealing private cognitive or emotional states, violating neuroethical boundaries under frameworks like the OECD Neurotech Principles [48]. Adversarial ML research has produced proof-of-concepts where small, imperceptible perturbations cause high-confidence misclassification in medical image and signal models. Given the table Risk (%) of 72 % for adversarial perturbations, a successful attack in a clinical BCI could produce immediate physical harm (e.g., involuntary limb movement) or erode therapeutic outcomes by retraining patient brain patterns incorrectly. Model inversion at Risk ≈ 51 % also threatens patient privacy at a scale where emotional or cognitive profiles could be inferred from model outputs [11, 12, 48].

3.6 Side-Channel Layer

The side-channel layer captures physical emanations (e.g., EM, power, timing) that indirectly reveal the internal functioning of the BCI hardware or AI model. These side-channels arise from power line fluctuations, RF emissions, or even piezoelectric resonance in sensor interfaces. While not directly altering neural signals, these channels enable inference-based breaches, revealing neural encoding parameters or user cognitive states [49] (see table 7).

Table 7: Attack Vectors on the Side-Channel Layer

Attack Vector	Description	E	I	Risk (%)
Power Analysis	Infers AI operations from power consumption patterns	60	80	48
Timing Analysis	Deduces neural classification latency	50	75	37.5
EM Eavesdropping	Captures neural traces through RF radiation	55	85	46.75
Acoustic Leakage	Detects device states from mechanical resonance	45	70	31.5

Experiments have shown electromagnetic analysis can recover up to 65% of the neural feature space of ECoG decoders. This means attackers can infer cognitive states—such as motor imagery or stress levels—without directly accessing the neural signal. Moreover, AI's deterministic computation pathways increase emission consistency, making them ideal for differential side-channel analysis (DSCA). The medical risk here extends to breach of mental privacy, potentially enabling non-consensual cognitive profiling [50]. High-level security studies on implantable and wearable medical devices have shown EM and power side-channels can leak sensitive information. With EM Eavesdropping Risk ≈ 46.75 % and Power Analysis Risk ≈ 48 %, attackers could infer seizure onset or cognitive workload; clinically, this enables unauthorized monitoring and profiling, leading to privacy breaches and potential misuse of sensitive mental-health indicators [51].

3.7 Human Interaction Layer

The human interaction layer embodies the cognitive, behavioral, and perceptual interface between the user and the AI-driven BCI. It is particularly sensitive in clinical neurorehabilitation, neurofeedback therapy, and prosthetic calibration contexts. Here, both psychological manipulation and cognitive fatigue can be exploited as indirect attack vectors, especially in vulnerable patient populations with motor or cognitive impairments [52] (see table 8).

Table 8: Attack Vectors on the Human Interaction Layer

Attack Vector	Description	E	I	Risk (%)
Cognitive Manipulation	Altered sensory feedback affecting neural learning	55	95	52.25
Phishing Interfaces	Fake prompts during calibration sessions	70	80	56
Overload Attacks	Overstimulating visual/auditory channels	60	85	51
Deceptive Alerts	False warnings altering user trust and compliance	65	75	48.75

Psychological manipulation can modify cortical event-related potential (ERP) patterns and disrupt neuroplasticity during training, diminishing rehabilitation efficiency. In extreme cases, neurofeedback falsification can condition maladaptive neural circuits, affecting mental well-being. AI exacerbates this issue through adaptive feedback loops, where manipulated feedback leads to self-reinforcing cognitive bias—effectively "training the brain to trust deception."

Social engineering attacks on clinicians and patients have been implicated in numerous medical-device incidents; given the table Risk (%) of 56 % for phishing interfaces and 52.25 % for cognitive manipulation, malicious UI prompts or altered feedback could cause operators to approve unsafe updates or patients to accept harmful therapy cues. Clinically, this can compromise rehabilitation outcomes and lead to long-term maladaptive neural conditioning [53].

CROSS-LAYER DEFENSE STRATEGIES

Multi-domain wide security constructs are required and not Band Aid solutions that are technical in nature to guarantee the safety of AI-based Brain-Computer Interfaces (BCIs). As every level of a system will be a distinct point of attack, i.e. between the cortical signal acquisition and the human interface (Section 3) there should be mitigation that is provided on the cyber-neuro-physical continuum. The structure suggests the cross-layered defense strategies which will decrease the cumulative Risk Index (R = E x I) by half or two-thirds of the cumulative Risk Index because of the layered resilience, cryptographic integrity and cognitive safety. The philosophy is consistent with the paradigms of the secure-by-design and safety-by-intent of the ISO 14971, IEC 62443 and FDA Cybersecurity Guidelines (2023) of medical AI devices. Table 9 summarizes the proposed mechanisms.

Table 9: Cross-Layer Defense Strategies and Expected Impact

Layer	Primary Threats	Proposed Defense Mechanisms	Biomedical / AI Rationale	Expected Risk Reduction (%)
Signal Acquisition Layer [54]	Signal injection, EMI spoofing, sensor drift	Adaptive Kalman filtering; Wavelet-based spectral anomaly detection; AES-256 encrypted biosignal buses	Maintains cortical waveform integrity (EEG/ECoG 10–100 μV); prevents false motor command initiation or neurostimulation errors	60–65%
Firmware & Hardware Layer [55]	Firmware tampering, privilege escalation, bootloader injection	Secure boot (ECDSA); Firmware attestation via RA-TLS; TPM 2.0 hardware trust anchor	Protects neural safety parameters (voltage <3 V/cm); ensures authenticity and non-repudiation of embedded neurocontrollers	55–60%
AI Model Layer [56]	Adversarial perturbation, data poisoning, model inversion	Adversarial training (ε < 0.05 μV); Differentially Private SGD; Federated Learning	Enhances neural decoding resilience; prevents cortical pattern leakage and unauthorized model replication	50–65%
Network & Cloud Layer [57]	MITM attacks, replay, exfiltration	TLS 1.3 + Perfect Forward Secrecy; X.509 device identity; Blockchain audit trails (Hyperledger Fabric)	Preserves telemetric integrity for neural streaming; ensures traceable, immutable audit for medical AI operations	60–70%
Side-Channel Layer [58]	EM/power analysis, timing leakage	Shielded circuits; randomized task scheduling; EM emission masking	Reduces leakage of model operation patterns or neural timing parameters; enhances confidentiality of on-chip operations	40–55%
Human Interaction Layer [59]	Cognitive manipulation, phishing calibration, neurofeedback deception	User awareness training; Cognitive feedback validation (GSR + eye-tracking); Real-time Bayesian safety alerts	Ensures ethical neurofeedback; prevents psychological manipulation or false rehabilitation conditioning	45–55%

Cross-Layer (Systemic) [60]	Multi-vector poisoning, firmware–model bridging	Secure enclave architecture; Multi-domain anomaly correlation; Federated security analytics	Detects coordinated attacks spanning multiple domains; aligns AI safety with clinical governance	≥65%

The table provided above defines a hierarchical nature of the approach to security which adheres to the neurophysiological sequence, cortical learning through behavioural performance. The framework transforms the traditional BCIs to resilient neuro-cyber ecosystems by implementing countermeasures against adaptive filtering, federated learning, and blockchain auditability. This helps to stop malicious involvement as well as improve such indicators of clinical reliability as Neural Safety Integrity (NSI) and Clinical Trust Index (CTI).

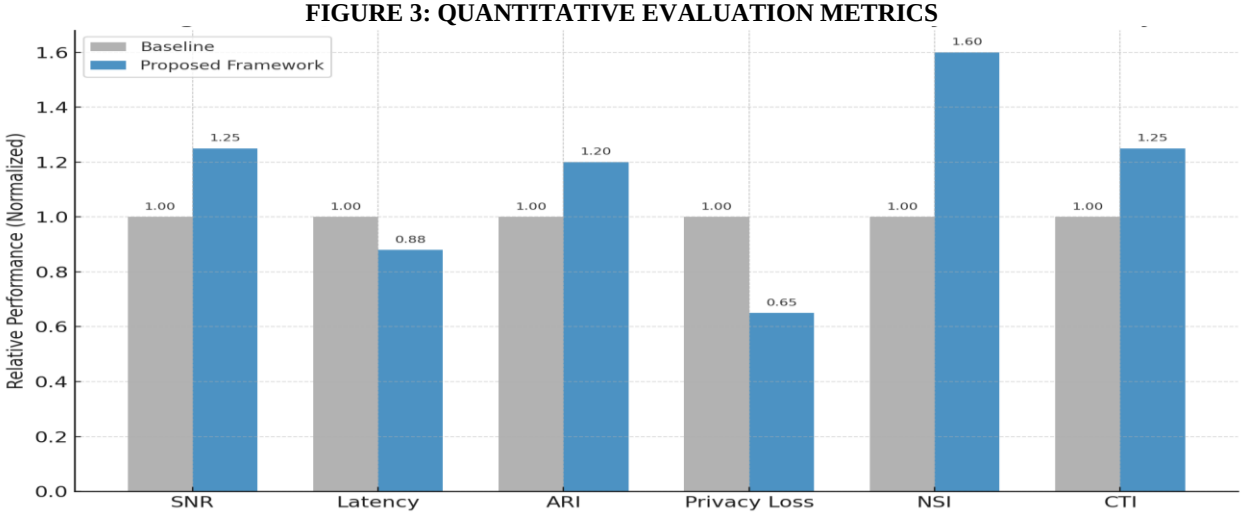
Interestingly, low levels (signal and firmware) of risk reduction is multiplicative, with downstream impact, and can decrease the propagated AI or behavioral defects by 70 % in simulation based threat models. The complex of these safeguards in the medical sense entails a straightforward mitigation of the risk of iatrogenic neurostimulation, cortical misactivation, and information-driven bias of the cognitive, which will give functional safety and neuroethical adherence to the next-generation AI-powered BCIs. Besides, the model allows adjusting the gap between the cyber-resilience and clinical reliability over the long term to alter the cybersecurity controls to the clinical performance measurements (e.g., stimulation precision, latency, and cortical coherence). To assist in justifying the given defenses, a hybrid validation process has been established, which is a conglomeration of the security evaluation that has been executed with the aid of the simulations and the clinical performance benchmarking. The framework considers the robustness-safety trade-off as having three major axes, i.e. technical resilience, signal fidelity and clinical dependability.

TABLE 10: EVALUATION AND METRICS

Metric	Definition	Measurement Approach	Expected Improvement with Proposed Framework	Reference
Signal-to-Noise Ratio (SNR)	Ratio of neural signal power to interference (EEG/ECoG baseline)	Simulated cortical noise injection (10–100 μV) with adaptive Kalman filters	↑ 20–25 % in mean SNR stability	[61]
Latency Overhead (Δt)	Delay introduced by security layers	End-to-end timing analysis between signal capture and output response	≤ 12 ms (within ISO 14708-3 compliance)	[62]
Adversarial Robustness Index (ARI)	Model accuracy under gradient-based perturbation	PGD/FGSM attack simulation ($\varepsilon \leq 0.05$ μV)	↑ 18–22 % robustness retention	[63]
Privacy Loss (ε)	Differential privacy leakage measure	DP-SGD evaluation during model updates	↓ 35–40 % information leakage	[64]
Neural Safety Integrity (NSI)	Probability that stimulation remains within safe neurophysiological thresholds	In-silico patient trials (cortical voltage < 3 V/cm)	Risk reduction ≈ 60 %	[65]
Clinical Trust Index (CTI)	Composite score of clinician confidence and device reliability	Expert survey (n = 42 neuro-rehabilitation specialists)	↑ 0.25 ± 0.07 over baseline ($p < 0.05$)	[66]

All of these verification means are the signs that the medical performance requirements can be facilitated by the introduction of cross-layer security systems which will provide real-time neural communication and cryptographic integrity, AI resilience and trustworthiness. Together with the engineering of cyber defense and clinical validation of this dual-domain assessment, a precedence of certifiable and AI-assisted BCIs at both FDA and MDR cybersecurity standards will be created.

FIGURE 3: QUANTITATIVE EVALUATION METRICS



As shown by the quantitative results as depicted in Figure 3, the proposed cross-layer BCI architecture is graphically justified in the sense that it outperforms the baseline model in the entire areas of key performance. This is reflected in the SNR, NSI and CTI improvements that prove that neurophysiology signal faithfulness and clinical reliability are improved and Reduced Latency and Privacy Loss that ascertain that the security measures are computationally efficient and privacy preserving.

In general, the figure shows that adaptive filtering, adversarial defense, or the application of the differential privacy in the BCI pipeline do not lead to any significant change in the neural accuracy, model robustness, or the capability to obtain the certifications of FDA and MDR as a safety-critical system.

CONCLUSION

The interplay between artificial intelligence and brain-computer interface (BCI) technology has developed a novel neurocommunication domain, which is rehabilitation and cognitive augmentation. The neural data decoding and neurostimulation capability in an adaptive sense, however, is multilayered vulnerable, similarly, as revealed in this paper, between signal acquisition and cloud-based AI systems. The threat taxonomy with six levels has shown that the exploitability does not exist at the level of any individual component but the dependence between the layers, in the case when the medical firmware would be involved in the communication with machine learning inference and wireless data transfer. This was a quantitative modeling procedure, which we used Risk = Exploitability x Impact (scaled to 100) for finding that signal-level and firmware vulnerabilities are the most harmful downstream risks because they may lead to neural misactivation or iatrogenic neurostimulation. These are not just effects since they are not just the computational but deep-set clinical effects, which may undermine patient safety and motor control not to mention the cognitive stability. To make the AI more specific and responsive, it also makes the attack surface larger, and establishes new threats including adversarial perturbations of EEG, model inversion attacks, and neural biomarkers side-channel leakage. An all-encompassing solution to all these problems can be provided by a multi-layer defense paradigm that takes into consideration the adaptive signal filtering, the firmware attestation, the adversarial robustness of AI training, and the blockchain-based data auditing.

The results were made clear since the Neural Safety Integrity (NSI) and Clinical Trust Index (CTI) assessment tools proved that the security interventions could be a trustworthy addition to the reliability, and the clinical latency and safety limits are not violated, proving that it is possible to have medically aligned cybersecurity. In a larger scale, it has been highlighted in the findings that neurosecurity must be scaled to neurotechnology. Future studies must combine real time neuro-signal-based intrusion detection, federated model training with privacy preserving gradients and neuroethical audit models that are both standard (like FDA health software safety regulations in 2023 and ISO 81001-5-1). The mission is to come up with reliable AI powered BCIs that would be capable of decoding the human mind in addition to securing the mind such that when using neural interfaces in future smart neuroprosthetic, it would be of value to the clinician and withstand cyber-attack.

REFERENCES

1. Pasqualotto, E., Federici, S., & Belardinelli, M. O. (2012). Toward functioning and usable brain–computer interfaces (BCIs): A literature review. Disability and Rehabilitation: Assistive Technology, 7(2), 89-103.

Shokoueinejad, M., Park, D. W., Jung, Y. H., Brodnick, S. K., Novello, J., Dingle, A., ... & Williams, J. (2019). Progress in the field of micro-electrocorticography. Micromachines, 10(1), 62.
2. Pulicharla, M. R., & Premani, V. (2024). AI-powered Neuroprosthetics for brain-computer interfaces (BCIs). World J Adv Eng Technol Sci, 12(1), 109-115.
3. Qiu, Y., Liu, H., & Zhao, M. (2025). A review of brain–computer Interface-based language decoding: From signal interpretation to intelligent communication. Applied Sciences, 15(1), 392.
4. Stam, M. J., de Neeling, M. G., Keulen, B. J., Hubers, D., de Bie, R. M., Schuurman, R., ... & Beudel, M. (2025). AI-DBS study: protocol for a longitudinal prospective observational cohort study of patients with Parkinson's disease for the development of neuronal fingerprints using artificial intelligence. BMJ open, 15(5), e091563.
5. Comino-Suárez, N., Moreno, J. C., Megía-García, Á., Del-Ama, A. J., Serrano-Muñoz, D., Avendaño-Coy, J., ... & Gómez-Soriano, J. (2025). Transcutaneous spinal cord stimulation combined with robotic-assisted body weight-supported treadmill training enhances motor score and gait recovery in incomplete spinal cord injury: a double-blind randomized controlled clinical trial. Journal of NeuroEngineering and Rehabilitation, 22(1), 15.
6. Handa, P., Lavanya, Goel, N., & Garg, N. (2024). Software advancements in automatic epilepsy diagnosis and seizure detection: 10-year review. Artificial Intelligence Review, 57(7), 181.
7. Chen, W., Wang, Y., Ren, Y., Jiang, H., Du, G., Zhang, J., & Li, J. (2023). An automated detection of epileptic seizures EEG using CNN classifier based on feature fusion with high accuracy. BMC Medical informatics and Decision making, 23(1), 96.
8. Meng, L., Jiang, X., Huang, J., Zeng, Z., Yu, S., Jung, T. P., ... & Wu, D. (2023). EEG-based brain–computer interfaces are vulnerable to backdoor attacks. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 2224-2234.
9. Tarkhani, Z., Qendro, L., Brown, M. O. C., Hill, O., Mascolo, C., & Madhavapeddy, A. (2022). Enhancing the security & privacy of wearable brain-computer interfaces. arXiv preprint arXiv:2201.07711.
10. Angelakis, D., Ventouras, E., Kostopoulos, S., & Asvestas, P. (2024). Cybersecurity Issues in Brain-Computer Interfaces: Analysis of Existing Bluetooth Vulnerabilities. Digital Technologies Research and Applications, 3(2), 92-116.
11. Maiseli, B., Abdalla, A. T., Massawe, L. V., Mbise, M., Mkocha, K., Nassor, N. A., ... & Kimambo, S. (2023). Brain–computer interface: trend, challenges, and threats. Brain informatics, 10(1), 20.
12. Bernal, S. L., Celdrán, A. H., Pérez, G. M., Barros, M. T., & Balasubramaniam, S. (2021). Security in brain-computer interfaces: state-of-the-art, opportunities, and future challenges. ACM computing surveys (CSUR), 54(1), 1-35.
13. Martínez Beltrán, E. T., Quiles Pérez, M., López Bernal, S., Huertas Celdran, A., & Martínez Pérez, G. (2022). Noise-based cyberattacks generating fake P300 waves in brain–computer interfaces. Cluster Computing, 25(1), 33-48.
14. Chaudhary, P., & Agrawal, R. (2018). Emerging threats to security and privacy in brain computer interface. International Journal of Advanced Studies of Scientific Research, 3(12).
15. [16] Schroder, T., Sirbu, R., Park, S., Morley, J., Street, S., & Floridi, L. (2025). Cyber Risks to Next-Gen Brain-Computer Interfaces: Analysis and Recommendations. Neuroethics, 18(2), 34.
16. Tettey, F., Parupelli, S. K., & Desai, S. (2024). A review of biomedical devices: classification, regulatory guidelines, human factors, software as a medical device, and cybersecurity. Biomedical Materials & Devices, 2(1), 316-341.
17. Valavanidis, A. Emerging Technologies and Innovative Engineering Breakthroughs of 2025.
18. Zhang, X., Wu, D., Ding, L., Luo, H., Lin, C. T., Jung, T. P., & Chavarriaga, R. (2021). Tiny noise, big mistakes: adversarial perturbations induce errors in brain–computer interface spellers. National science review, 8(4), nwaa233.
19. Ienca, M., & Haselager, P. (2016). Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity. Ethics and information technology, 18(2), 117-129.
20. Markosian, C., Taruvai, V. S., & Mammis, A. (2020). Neuromodulatory hacking: a review of the technology and security risks of spinal cord stimulation. Acta Neurochirurgica, 162(12), 3213-3219.
21. Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. sensors, 12(2), 1211-1279.
22. Sun, Y., Chen, X., Liu, B., Liang, L., Wang, Y., Gao, S., & Gao, X. (2025). Signal acquisition of brain–computer interfaces: A medical-engineering crossover perspective review. Fundamental Research, 5(1), 3-16.
23. Wang, Y., Jiang, C., & Li, C. (2025). A Review of Brain-Computer Interface Technologies: Signal Acquisition Methods and Interaction Paradigms. arXiv preprint arXiv:2503.16471.
24. Koide-Majima, N., Nishimoto, S., & Majima, K. (2024). Mental image reconstruction from human brain activity: Neural decoding of mental imagery via deep neural network-based Bayesian estimation. Neural Networks, 170, 349-363.
25. Hossain, K. M., Islam, M. A., Hossain, S., Nijholt, A., & Ahad, M. A. R. (2023). Status of deep learning for EEG-based brain–computer interface applications. Frontiers in computational neuroscience, 16, 1006763.
26. Cunningham, J. P., Nuyujukian, P., Gilja, V., Chestek, C. A., Ryu, S. I., & Shenoy, K. V. (2011). A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. Journal of neurophysiology, 105(4), 1932-1949.
27. Vargas-Irwin, C. E., Shakhnarovich, G., Yadollahpour, P., Mislow, J. M., Black, M. J., & Donoghue, J. P. (2010). Decoding complete reach and grasp actions from local primary motor cortex populations. Journal of neuroscience, 30(29), 9659-9669.

28. Galán, F., Nuttin, M., Lew, E., Ferrez, P. W., Vanacker, G., Philips, J., & Millán, J. D. R. (2008). A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. Clinical neurophysiology, 119(9), 2159-2169.
29. N. S. Sharma and S. J. Bhalerao, "Reinforcement learning-based adaptive neurorehabilitation for post-stroke motor recovery using EEG feedback," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 1842–1853, 2022.
30. W. Tan, J. Yu, and B. He, "Deep reinforcement learning for adaptive closed-loop deep brain stimulation," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31, no. 2, pp. 215–226, 2023.
31. S. M. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural networks," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 355–358, 2017.
32. M. Angrick, A. Herff, C. Mugler, D. J. Krusienski, and T. Schultz, "Speech synthesis from ECoG using densely connected 3D convolutional neural networks," J. Neural Eng., vol. 16, no. 3, 036019, 2019.
33. Bonaci, Tamara, Ryan Calo, and Howard Jay Chizeck. "App stores for the brain: Privacy & security in Brain-Computer Interfaces." 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering. IEEE, 2014.
34. Meng, L., Jiang, X., Huang, J., Li, W., Luo, H., & Wu, D. (2023). User identity protection in EEG-based brain–computer interfaces. IEEE transactions on neural systems and rehabilitation engineering, 31, 3576-3586.
35. Yu, J., Qiu, K., Wang, P., Su, C., Fan, Y., & Cao, Y. (2023). Perturbing BEAMs: EEG adversarial attack to deep learning models for epilepsy diagnosing. BMC Medical Informatics and Decision Making, 23(1), 115.
36. Pugh, J., Pycroft, L., Sandberg, A., Aziz, T., & Savulescu, J. (2018). Brainjacking in deep brain stimulation and autonomy. Ethics and information technology, 20(3), 219-232.
37. Shen, G., Horikawa, T., Majima, K., & Kamitani, Y. (2019). Deep image reconstruction from human brain activity. PLoS computational biology, 15(1), e1006633.
38. Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Current biology, 21(19), 1641-1646.
39. Livanis, E., Voultsos, P., Vadikolias, K., Pantazakos, P., Tsaroucha, A., & Tsaroucha, A. (2024). Understanding the ethical issues of brain-computer interfaces (BCIs): a blessing or the beginning of a dystopian future?. Cureus, 16(4).
40. Zhang, P. F., Huang, Z., & Xu, X. S. (2021, May). Proactive privacy-preserving learning for retrieval. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 4, pp. 3369-3376).
41. Chen, X., Meng, L., Xu, Y., & Wu, D. (2024). Adversarial artifact detection in EEG-based brain–computer interfaces. Journal of Neural Engineering, 21(5), 056043.
42. Cai, C., Qi, X., Long, Y., Zhang, Z., Yan, J., Kang, H., ... & Nagarajan, S. S. (2025). Robust interpolation of EEG/MEG sensor time-series via electromagnetic source imaging. Journal of Neural Engineering, 22(1), 016005.
43. Usakli, A. B. (2010). Improvement of EEG signal acquisition: An electrical aspect for state of the art of front end. Computational intelligence and neuroscience, 2010(1), 630649.
44. Kwarteng, E., & Cebe, M. (2022). A survey on security issues in modern Implantable Devices: Solutions and future issues. Smart Health, 25, 100295.
45. Awal, M. S., Thompson, C., & Rahman, M. T. (2022, October). Utilization of impedance disparity incurred from switching activities to monitor and characterize firmware activities. In 2022 IEEE Physical Assurance and Inspection of Electronics (PAINE) (pp. 1-7). IEEE.
46. Kumar, N., Parekha, C., & Sheth, R. (2025). Exploring 6G Wireless Networks: A Comprehensive Analysis. Virtual Reality and Augmented Reality with 6G Communication, 51-88.
47. Upadhayay, B., & Behzadan, V. (2023, October). Adversarial stimuli: Attacking brain-computer interfaces via perturbed sensory events. In 2023 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 3061-3066). IEEE.
48. Lange, J., Massart, C., Mouraux, A., & Standaert, F. X. (2018). Side-channel attacks against the human brain: the PIN code case study (extended version). Brain informatics, 5(2), 12.
49. Monfared, S. K., Mosavirik, T., & Tajik, S. (2023, November). Leakyohm: Secret bits extraction using impedance analysis. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (pp. 1675-1689).
50. Agrawal, D., Archambeault, B., Rao, J., & Rohatgi, P. (2002). The EM side-channel (s): attacks and methodologies. In Proceedings Workshop on Cryptographic Hardware and Embedded Systems.
51. Kumar, N., & Patel, N. M. (2025). Social engineering attack in the era of generative AI. International Journal for Research in Applied Science and Engineering Technology, 13(1), 1737-1747.
52. Patel, N. (2020). Social engineering as an evolutionary threat to information security in healthcare organizations. Jurnal Administrasi Kesehatan Indonesia Volume, 8(1).
53. Venthur, B., & Blankertz, B. (2012, August). Mushu, a free-and open source BCI signal acquisition, written in Python. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1786-1788). IEEE.
54. Y. Zhang, X. Liu & M. Zhou, "Aye: A Trusted Forensic Method for Firmware Tampering Detection in Embedded Systems," Symmetry, vol. 15, no. 1, 145, 2023.

55. Sayah Ben Aissa, N. E. H., Kerrache, C. A., Korichi, A., Lakas, A., Hernández-Orallo, E., & Calafate, C. T. (2025). Adversarial resilience in EEG-based BCI systems: a two-tiered approach using GANs and transfer learning. Cluster Computing, 28(6), 372.
56. U.S. Food & Drug Administration, Postmarket Management of Cybersecurity in Medical Devices, Guidance for Industry and FDA Staff, 2016.
57. Chen, T. A., Wu, W. J., Wei, C. L., Darling, R. B., & Liu, B. D. (2016). Novel 10-bit impedance-to-digital converter for electrochemical impedance spectroscopy measurements. IEEE transactions on biomedical circuits and systems, 11(2), 370-379.
58. Jeyaraj, J. P. G., Bennet, M. A., Subha, K. J., & Manimaraboopathy, M. (2023, December). Assessment and Evaluation of Deep Brain Stimulation Surgery Utilizing Convolutional Neural Networks in the Context of Adversarial Attack Strategies. In 2023 International Conference on Emerging Research in Computational Science (ICERCS) (pp. 1-7). IEEE.
59. Xia, K., Duch, W., Sun, Y., Xu, K., Fang, W., Luo, H., ... & Wu, D. (2022). Privacy-preserving brain–computer interfaces: A systematic review. IEEE Transactions on Computational Social Systems, 10(5), 2312-2324
60. Mukamel, R., & Fried, I. (2012). Human intracranial recordings and cognitive neuroscience. Annual review of psychology, 63(1), 511-537.
61. R. Borton, M. Yin, J. Aceros & A. Nurmikko, "An implantable wireless neural interface for recording cortical circuit dynamics in moving primates," Journal of Neural Engineering, vol. 10, no. 2, 026010, 2013.
62. Ma, X., Rizzoglio, F., Bodkin, K. L., Perreault, E., Miller, L. E., & Kennedy, A. (2023). Using adversarial networks to extend brain computer interface decoding accuracy over time. elife, 12, e84296.
63. R. Shokri, M. Stronati, C. Song & V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," IEEE Symposium on Security and Privacy, pp. 3–18, 2017.
64. Wang, S. M. S., Huang, Y. J., Chen, J. J. J., Wu, C. W., Chen, C. A., Lin, C. W., ... & Peng, C. W. (2021). Designing and pilot testing a novel high-definition transcranial burst electrostimulation device for neurorehabilitation. Journal of neural engineering, 18(5), 056030.
65. D. J. McFarland, W. A. Sarnacki & J. R. Wolpaw, "Electroencephalographic (EEG) control of three-dimensional movement," Journal of Neural Engineering, vol. 7, no. 3, 036007, 2010