

# Securing Neural Interfaces: Architecture, Threat Taxonomy, and Neural Impact Scoring for Brain-Computer Interfaces

Kevin L. Qi  
Qinnovate  
kevin@qinnovate.com

February 2026

## Abstract

Brain-computer interfaces (BCIs) are transitioning from experimental neuroscience tools to commercially deployed medical devices, with companies including Neuralink, Synchron, Blackrock Neurotech, and Paradromics advancing toward regulatory approval and new entrants such as Merge Labs raising \$252M in seed funding [2]. Yet no security framework exists that accounts for the unique risks of devices that read and write neural signals. The Common Vulnerability Scoring System (CVSS v4.0), the industry standard for vulnerability assessment, cannot express biological tissue damage, cognitive integrity violations, consent boundaries, damage reversibility, or neuroplastic consequences—dimensions critical to neural device security.

We present an integrated security framework comprising four contributions: (1) an **11-band hourglass architecture** mapping attack surfaces from neocortex to wireless radio across neural, interface, and synthetic zones; (2) **TARA**, a threat taxonomy of 109 techniques across 15 tactics and 8 domains, each classified by status, severity, and dual-use therapeutic potential; (3) **NISS**, the Neural Impact Scoring System—a CVSS v4.0 extension adding six neural-specific metrics (Biological Impact, Cognitive Reconnaissance, Cognitive Disruption, Consent Violation, Reversibility, Neuroplasticity) designed to conform with FIRST.org’s official extension mechanism; and (4) the **Neural Impact Chain**, a methodology mapping security vulnerabilities to DSM-5-TR psychiatric diagnoses through a six-stage pipeline.

Analysis of all 109 techniques reveals that 99.1 cannot express. The Neural Impact Chain maps all techniques to 29 unique DSM-5-TR diagnostic codes across 5 psychiatric clusters, with 52 techniques posing direct diagnostic risk. The framework identifies 78 techniques (71.6 establishing a dual-use atlas where every attack mechanism that can harm neural tissue has a corresponding clinical application. The complete framework, threat registry, and scoring system are released as open source under the Apache 2.0 license.

**Keywords:** brain-computer interface, neurosecurity, CVSS, threat taxonomy, neural scoring, neuroethics, DSM-5-TR, dual-use

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	The Scoring Gap . . . . .	4
1.2	Contributions . . . . .	4
<b>2</b>	<b>Related Work</b>	<b>5</b>
2.1	Foundational Neurosecurity . . . . .	5
2.2	Empirical Attack Demonstrations . . . . .	5
2.3	Emerging Frameworks . . . . .	5
2.4	Neuroethics and Policy . . . . .	6
2.5	How QIF Differs . . . . .	6

<b>3</b>	<b>The Hourglass Architecture</b>	<b>6</b>
3.1	Design Rationale . . . . .	7
3.2	Neural Band Derivation . . . . .	7
3.3	Band Definitions . . . . .	7
3.4	Zone Structure . . . . .	8
3.5	Security Properties . . . . .	8
<b>4</b>	<b>TARA: Threat Taxonomy</b>	<b>9</b>
4.1	Taxonomy Structure . . . . .	10
4.2	Evidence Classification . . . . .	10
4.3	Severity Distribution . . . . .	10
4.4	Category Breakdown . . . . .	11
4.5	Dual-Use Mapping . . . . .	11
4.6	Representative Techniques . . . . .	11
<b>5</b>	<b>NISS: Neural Impact Scoring System</b>	<b>12</b>
5.1	Gap Analysis: Why CVSS Alone Is Insufficient . . . . .	12
5.2	Extension Metrics . . . . .	13
5.2.1	BI: Biological Impact . . . . .	13
5.2.2	CR: Cognitive Reconnaissance . . . . .	13
5.2.3	CD: Cognitive Disruption . . . . .	13
5.2.4	CV: Consent Violation . . . . .	14
5.2.5	RV: Reversibility . . . . .	14
5.2.6	NP: Neuroplasticity . . . . .	14
5.3	PINS Flag . . . . .	14
5.4	Scoring Formula . . . . .	15
5.5	Vector Format . . . . .	15
5.6	NISS Severity Distribution . . . . .	15
<b>6</b>	<b>Neural Impact Chain</b>	<b>16</b>
6.1	Pipeline Architecture . . . . .	16
6.2	NISS-to-DSM Bridge . . . . .	17
6.3	Coverage Statistics . . . . .	17
6.4	Diagnostic Cluster Distribution . . . . .	17
6.5	Risk Classification . . . . .	18
6.6	Example Walkthrough . . . . .	18
<b>7</b>	<b>Governance and Neuroethics</b>	<b>19</b>
7.1	Consent Tiers . . . . .	19
7.2	UNESCO Alignment . . . . .	19
7.3	Neurorights Framework Integration . . . . .	20
7.4	Regulatory Mapping . . . . .	20
<b>8</b>	<b>Case Studies</b>	<b>21</b>
8.1	Scoring Comparison . . . . .	21
8.2	Case 1: Cortical Signal Injection (QIF-T0001) . . . . .	21
8.3	Case 2: P300 Side-Channel (QIF-T0015) . . . . .	21
8.4	Case 3: Calibration Poisoning (QIF-T0034) . . . . .	22
8.5	Case 4: Covert Neural Surveillance (QIF-T0050) . . . . .	22
8.6	Case 5: Real-World Vulnerability Disclosure . . . . .	23

<b>9</b>	<b>Limitations and Future Work</b>	<b>23</b>
9.1	No Empirical Validation on Real BCI Devices . . . . .	23
9.2	DSM-5-TR Mapping Not Clinically Validated . . . . .	23
9.3	NISS Weights Not Calibrated . . . . .	23
9.4	No Interrater Reliability Study . . . . .	23
9.5	Taxonomy Completeness . . . . .	24
9.6	Single-Author Bias . . . . .	24
9.7	AI Tool Disclosure . . . . .	24
9.8	Future Work . . . . .	25
<b>10</b>	<b>Conclusion</b>	<b>25</b>

# 1 Introduction

Brain-computer interfaces (BCIs) are no longer confined to research laboratories. Neuralink has implanted its N1 device in human patients [22], Synchron’s Stentrode has demonstrated motor neuroprosthesis via neurointerventional surgery [25], and Blackrock Neurotech’s Utah array has enabled high-performance speech neuroprostheses [34]. Paradromics is advancing toward high-bandwidth cortical interfaces for paralyzed patients. Investment in neurotechnology grew 700% between 2014 and 2021, with the global market projected to reach \$25 billion by 2030 [31].

Despite this rapid commercialization, no security standard exists specifically for neural devices. Section 3305 of the Food and Drug Omnibus Reform Act of 2022 (FDORA, Pub. L. 117-328) added Section 524B to the FD&C Act [33], mandating cybersecurity documentation—including threat modeling, software bills of materials, and vulnerability disclosure—in all premarket submissions for connected medical devices. Since October 2023, the FDA enforces a Refuse-to-Accept policy for non-compliant submissions. However, Section 524B does not specify which threat taxonomy or scoring system to use, and the referenced standards (CVSS, IEC 62443, AAMI TIR57, ISO 14971) contain no provisions for the unique risks of devices that read and write neural signals. The EU Medical Device Regulation [8] similarly lacks neural-specific security requirements.

## 1.1 The Scoring Gap

The Common Vulnerability Scoring System (CVSS v4.0) [9] is the industry standard for assessing vulnerability severity. CVSS excels at capturing exploitability characteristics—attack vector, complexity, privileges required, user interaction—and information system impact across confidentiality, integrity, and availability. However, CVSS was designed for information technology assets. It has no concept of:

- **Biological tissue damage**—seizure induction, neural tissue necrosis, involuntary motor activation
- **Cognitive integrity**—thought privacy, perception manipulation, identity modification
- **Consent boundaries**—the distinction between operating within consented parameters and covert neural manipulation
- **Damage reversibility**—IT assets can be restored from backup; neural tissue cannot be rebooted
- **Neuroplastic consequences**—prolonged adversarial stimulation can cause lasting structural changes to the brain

When a vulnerability in a BCI device can induce seizures, decode private thoughts, or cause irreversible brain damage, a CVSS base score alone is insufficient. The gap is not theoretical: our analysis of 109 BCI attack techniques shows that 99.1% require scoring dimensions that CVSS cannot express.

## 1.2 Contributions

This paper presents an integrated security framework with four contributions:

1. **The Hourglass Architecture** (Section 3): An 11-band model mapping attack surfaces across three zones—neural (7 bands, from neocortex to spinal cord), interface (1 band, the electrode-tissue boundary), and synthetic (3 bands, from analog electronics to wireless radio). The hourglass shape reflects a natural security chokepoint at the neural interface.
2. **TARA Threat Taxonomy** (Section 4): A registry of 109 techniques across 15 tactics and 8 domains, each classified by evidence status, severity, and dual-use therapeutic potential. TARA is an independent taxonomy inspired by the structural methodology of MITRE ATT&CK<sup>®</sup>, tailored to the BCI domain.

3. **NISS Scoring System** (Section 5): The Neural Impact Scoring System—a CVSS v4.0 extension designed to conform with FIRST.org’s official extension mechanism (User Guide §3.11) [10], adding six neural-specific metrics. Formal registration with FIRST.org is planned as future work.
4. **The Neural Impact Chain** (Section 6): A methodology mapping security vulnerabilities to psychiatric diagnoses through a six-stage pipeline: technique → hourglass band → neural structure → cognitive function → NISS score → DSM-5-TR diagnostic code [1].

Sections 7–8 address governance alignment with international neuroethics frameworks and present comparative case studies. Section 9 discusses limitations and validation gaps honestly. The complete framework, threat registry, and scoring system are released as open source under the Apache 2.0 license.

## 2 Related Work

Research at the intersection of cybersecurity and neurotechnology has progressed through three phases: foundational framing, empirical demonstration, and emerging policy response. QIF builds on each while addressing gaps left by prior work.

### 2.1 Foundational Neurosecurity

Denning, Matsuoka, and Kohno [6] coined the term “neurosecurity” and established the field by analyzing attack surfaces in implantable neurostimulators. Their work identified wireless reprogramming, battery depletion, and signal injection as threat categories, but predated modern high-bandwidth BCIs and did not propose a scoring system.

Bonaci, Calo, and Chizeck [3] extended this with “App Stores for the Brain,” examining privacy threats from third-party BCI applications. They identified the need for access control at the neural data layer—a concern our governance framework addresses through consent tiers (Section 7).

### 2.2 Empirical Attack Demonstrations

Martinovic et al. [19] demonstrated at USENIX Security 2012 that consumer-grade EEG headsets could be exploited via P300 event-related potentials to extract private information—PIN numbers, bank details, and personal preferences—through subliminal visual stimuli embedded in a game. This remains the most influential empirical demonstration of a BCI side-channel attack. Our TARA taxonomy (Section 4) classifies this as a confirmed technique in the Signal Eavesdropping category.

Ienca and Haselager [14] broadened the scope to “hacking the brain,” analyzing BCI security through the lens of informational and physical integrity. They proposed that BCI security requires both traditional cybersecurity measures and novel neuroethical safeguards—a dual requirement that QIF implements through the parallel CVSS/NISS scoring architecture.

Meng et al. [20] established an adversarial robustness benchmark for EEG-based BCIs, evaluating multiple adversarial defense approaches across neural network architectures and EEG datasets. Their work focuses on machine learning adversarial robustness; TARA extends coverage to physical-layer, protocol-layer, and cognitive-layer threats.

### 2.3 Emerging Frameworks

Schroder et al. [28] published the most recent analysis (2025), developing a threat model for network-based attacks on next-generation BCIs and identifying associated regulatory changes needed to address them. Their work contributes threat categorization across network pathways

but does not propose a scoring system, formal threat taxonomy with technique-level granularity, or clinical impact mapping.

Camara et al. [4] and Rushanan et al. [27] surveyed security and privacy in implantable medical devices more broadly, covering pacemakers, insulin pumps, and cochlear implants alongside neurostimulators. Halperin et al. [12] demonstrated practical wireless attacks against pacemakers, establishing that implanted medical device security is not theoretical. These works address the broader medical device landscape; QIF focuses specifically on the unique challenges of bidirectional neural interfaces.

## 2.4 Neuroethics and Policy

Ienca and Andorno [13] proposed four fundamental neurorights: cognitive liberty, mental privacy, mental integrity, and psychological continuity. Yuste et al. [35] added equal access and protection from algorithmic bias. These philosophical frameworks establish *what* must be protected; QIF provides the technical architecture for *how*.

Muñoz et al. [23] and Lázaro-Muñoz et al. [18] contributed empirical grounding through researcher interviews, documenting clinician concerns about adaptive DBS ethics (72% cited uncertainty about risks) and post-trial access obligations for implanted neural devices.

UNESCO’s 2025 Recommendation on the Ethics of Neurotechnology [31]—adopted by 194 Member States—represents the first global normative framework for neurotechnology governance. The OECD had previously published a Recommendation on Responsible Innovation in Neurotechnology [24] covering 36 member countries. Chile became the first nation to constitutionally protect neurorights [26]. These policy instruments establish principles and values; QIF demonstrates technical implementation (Section 7).

## 2.5 How QIF Differs

Prior work has addressed BCI security, neuroethics, and vulnerability scoring separately. QIF is the first framework to integrate all four layers into a single system:

1. **Architecture:** A formal band model (vs. informal threat lists)
2. **Taxonomy:** Technique-level granularity with evidence classification (vs. category-level surveys)
3. **Scoring:** A CVSS-compatible extension with neural-specific metrics (vs. qualitative risk ratings)
4. **Clinical mapping:** A pipeline from security vulnerability to psychiatric diagnosis (novel; no prior work)

The Neural Impact Chain (Section 6) has no precedent in either cybersecurity or neuroethics literature. To our knowledge, no prior work has systematically mapped vulnerability severity to DSM-5-TR diagnostic codes.

## 3 The Hourglass Architecture

The QIF architecture maps the full BCI attack surface using an 11-band hourglass model organized into three zones. The model is derived from two independent design traditions: the OSI networking reference model [36], which stratifies communication systems into functional layers, and functional neuroanatomy, which organizes the nervous system by anatomical structure and physiological role. The hourglass shape borrows from the Internet protocol architecture [5], where IP serves as a narrow waist through which all traffic must pass.

### 3.1 Design Rationale

A BCI system spans from high-level cortical computation to low-level radio transmission. Any security framework must account for threats at every point along this path. We observe that the neural interface—the electrode-tissue boundary—forms a natural chokepoint analogous to IP in the Internet hourglass: all signals, whether neural or synthetic, must cross this boundary. Above the interface, attack surfaces expand through the complexity of neural tissue. Below it, they expand through electronic and wireless systems. The resulting shape is an asymmetric hourglass with the narrowest point at the interface.

### 3.2 Neural Band Derivation

The seven neural bands (N1–N7) are not arbitrary divisions. They correspond directly to the canonical central nervous system hierarchy established in foundational neuroscience texts [16]: spinal cord, brainstem, cerebellum, diencephalon, basal ganglia, limbic system, and neocortex. This mapping was chosen for three reasons:

1. **Functional differentiation:** Each CNS division performs computationally distinct roles. The spinal cord (N1) mediates reflexes; the brainstem (N2) controls vital autonomic functions; the neocortex (N7) performs executive cognition and language. Attacks targeting different divisions produce qualitatively different harms—from respiratory arrest (N2) to identity manipulation (N7).
2. **Determinacy gradient:** Ascending the neural hierarchy, system behavior becomes progressively less deterministic. Spinal reflexes (N1) are largely predictable; basal ganglia dynamics (N5) exhibit chaotic sensitivity; prefrontal deliberation (N7) involves the highest degree of unpredictability. This gradient has direct security implications: attacks at lower bands produce more predictable (and often more immediately dangerous) effects, while attacks at higher bands produce less predictable but potentially more insidious cognitive consequences.
3. **Device deployment mapping:** Existing BCI devices map cleanly onto specific bands. Sacral nerve stimulators and spinal cord stimulators operate at N1, deep brain stimulation systems (e.g., Medtronic Percept) target the subthalamic nucleus at N5, and cortical implants (e.g., Neuralink N1 chip, Blackrock Utah array) interface at N7. The band structure ensures that every deployed device class has a defined security context.

Each band also exhibits characteristic neural oscillation frequency ranges (delta through gamma), though the bands are defined by neuroanatomy rather than frequency. The oscillation profile of each band informs signal-plausibility checks: a stimulation signal injected at N4 (thalamus) can be validated against the expected spectral envelope for thalamocortical rhythms [29], providing a physics-based anomaly detection mechanism.

### 3.3 Band Definitions

Table 1 defines all 11 bands. The model uses a 7-1-3 asymmetric structure: seven neural bands (N7–N1), one interface band (I0), and three synthetic bands (S1–S3).

Table 1: QIF Hourglass: 11 bands across three zones.

Band	Name	Zone	Attack Surface
N7	Neocortex	Neural	PFC, M1, V1, Broca, Wernicke—executive function, language, movement, perception
N6	Limbic System	Neural	Hippocampus, amygdala, insula—emotion, memory, interoception
N5	Basal Ganglia	Neural	Striatum, STN, substantia nigra—motor selection, reward, habit
N4	Diencephalon	Neural	Thalamus, hypothalamus—sensory gating, consciousness relay
N3	Cerebellum	Neural	Cerebellar cortex, deep nuclei—motor coordination, timing
N2	Brainstem	Neural	Medulla, pons, midbrain—vital functions, arousal, reflexes
N1	Spinal Cord	Neural	Cervical through sacral—reflexes, peripheral relay
I0	Neural Interface	Interface	Electrode-tissue boundary—the hourglass waist; all signals must cross
S1	Analog / Near-Field	Synthetic	Amplification, ADC, near-field EM (0–10 kHz)
S2	Digital / Telemetry	Synthetic	Decoding, BLE/WiFi, telemetry (10 kHz–1 GHz)
S3	Radio / Wireless	Synthetic	RF, directed energy, application layer (1 GHz+)

### 3.4 Zone Structure

**Neural Zone (N7–N1).** The neural zone encompasses all biological structures from the neocortex to the spinal cord. Attack surfaces in this zone involve direct interaction with neural tissue: signal injection, neural manipulation, cognitive exploitation, and physical safety threats. The ordering follows neuroanatomical hierarchy—higher bands represent more complex cognitive functions, while lower bands represent more fundamental physiological processes. Attacks at lower neural bands (N1–N2) threaten vital functions; attacks at higher bands (N6–N7) threaten cognition, identity, and autonomy.

**Interface Zone (I0).** Band I0 is the singular chokepoint where biological and synthetic signals convert. For invasive BCIs, this is the electrode-tissue boundary where electrical signals are transduced from ionic currents in neural tissue to electronic currents in silicon. For non-invasive devices, it is the sensor surface (e.g., scalp electrodes for EEG). The interface band is the narrowest point in the hourglass and the most critical for security: every neural signal that reaches the synthetic system, and every stimulation signal that reaches neural tissue, must pass through I0.

**Synthetic Zone (S1–S3).** The synthetic zone covers electronic and wireless systems from the analog front-end through digital processing to radio transmission. S1 handles analog signal conditioning and near-field electromagnetic effects. S2 encompasses digital processing, Bluetooth Low Energy, WiFi, and telemetry protocols. S3 covers radio frequency transmission, directed energy, and application-layer communication. Threats in this zone are closest to traditional IT security and are most amenable to conventional countermeasures.

### 3.5 Security Properties

The hourglass architecture provides three properties relevant to security analysis:

1. **Completeness:** Every component of a BCI system maps to exactly one band. There is no attack surface that falls outside the model.
2. **Chokepoint identification:** Band I0 identifies the natural monitoring point where all signals can be inspected during transit between neural and synthetic zones.

3. **Threat locality:** Each technique in the TARA taxonomy (Section 4) is annotated with the bands it affects, enabling defenders to understand the spatial extent of an attack across the architecture.

Figure 1 illustrates the 11-band model with relative band widths reflecting attack surface breadth at each layer.

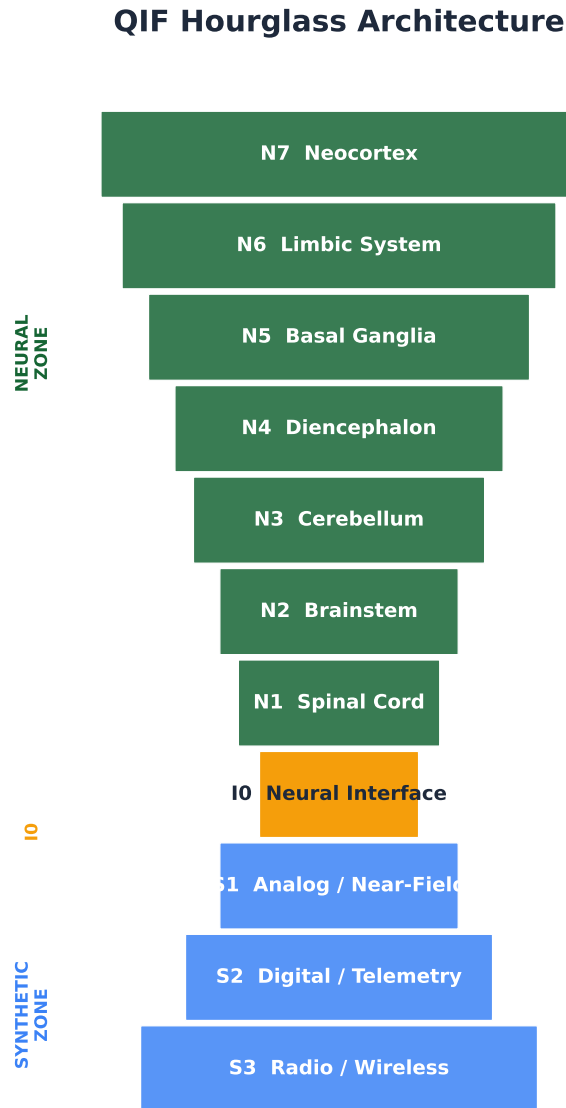


Figure 1: The QIF Hourglass Architecture. Band widths reflect relative attack surface breadth. The neural interface (I0) forms the narrow waist—the security chokepoint through which all signals must pass.

## 4 TARA: Threat Taxonomy

The Therapeutic Applications & Risk Assessment (TARA) registry catalogs BCI attack techniques with a structure inspired by MITRE ATT&CK [21], independently developed to cover neural, cognitive, and physical safety domains. Each technique is simultaneously an attack vector, an ethical risk, and—where applicable—a therapeutic application.

**Why not ATT&CK directly?** MITRE ATT&CK is designed for enterprise IT and mobile environments where adversary objectives center on data exfiltration, lateral movement, and persistence. BCI threats differ in three fundamental ways that make direct adoption inadequate: (1) the target is biological tissue, not an information system—attacks can cause seizures, tissue necrosis, or cognitive coercion, none of which map to ATT&CK’s impact categories; (2) the same technique is often simultaneously an attack and a therapy (e.g., deep brain stimulation is both a Parkinson’s treatment and a neural injection vector), requiring a dual-use classification that ATT&CK has no mechanism to express; and (3) the attack lifecycle must span from radio-frequency transmission through silicon processing to neural tissue, crossing the bio-digital boundary that ATT&CK’s purely digital model does not address. TARA adopts ATT&CK’s proven structural methodology—techniques organized into tactics and domains—while building an independent taxonomy purpose-built for the BCI threat landscape.

#### 4.1 Taxonomy Structure

The TARA registry organizes 109 techniques into 15 tactics across 8 operational domains. Technique identifiers follow the format QIF-TXXX, where the numeric suffix provides sequential ordering.

**Domains.** The eight domains span the full attack lifecycle:

1. **Neural (N)** — Direct interaction with neural tissue (scan, injection, manipulation)
2. **BCI System (B)** — System-level intrusion and evasion
3. **Protocol (P)** — Protocol disruption and communication attacks
4. **Data (D)** — Data harvesting and exfiltration
5. **Cognitive (C)** — Cognitive exploitation and imprinting
6. **Countermeasure (M)** — Surveillance and monitoring
7. **Evasion (E)** — Defense evasion and anti-detection
8. **Sensor (S)** — Consumer device side-channel attacks

**Tactics.** The 15 tactics follow a lifecycle structure analogous to ATT&CK, adapted for BCI operations: Neural Scan, BCI Intrusion, Neural Injection, Cognitive Imprinting, BCI Evasion, Data Harvesting, Neural Manipulation, Evasion/Rootkit Deployment, Monitoring/Surveillance, Protocol Disruption, Cognitive Exploitation, Signal Replay, Signal Harvesting, Signal Fingerprinting, and Signal Chaining.

#### 4.2 Evidence Classification

Each technique carries an evidence status reflecting the maturity of its documentation:

Table 2: TARA evidence status classification with technique counts.

Status	Definition	Count
Confirmed	Documented in real-world use or peer-reviewed literature	19
Demonstrated	Proven in laboratory or controlled conditions	35
Emerging	Newly identified; limited but growing evidence	22
Theoretical	Plausible based on known physics and engineering	31
Plausible	Possible but with significant uncertainty	1
Speculative	Hypothetical; requires unproven capabilities	1

#### 4.3 Severity Distribution

CVSS v4.0 base severity ratings for all 109 techniques show a distribution heavily weighted toward high and critical:

Table 3: TARA severity distribution (CVSS v4.0 base ratings).

Severity	Count	Percentage
Critical	29	26.6%
High	60	55.0%
Medium	17	15.6%
Low	3	2.8%

The dominance of high and critical ratings reflects the inherent severity of attacks against devices that interface directly with the nervous system. Even techniques with low exploitability can have severe consequences when the target is neural tissue.

#### 4.4 Category Breakdown

The 109 techniques distribute across eight operational categories:

Table 4: TARA techniques by operational category.

ID	Category	Count
SE	Signal Eavesdropping	20
CI	Cognitive Interference	18
EX	Data Exfiltration	19
DM	Data Manipulation	15
SI	Signal Injection	13
PE	Privilege Escalation	9
DS	Denial of Service	8
PS	Physical Safety	7

#### 4.5 Dual-Use Mapping

A distinctive feature of TARA is the systematic dual-use mapping: every attack technique is assessed for therapeutic analogs. The same physical mechanisms that enable attacks—electromagnetic stimulation, signal decoding, neuromodulation—are the mechanisms underlying established therapies such as deep brain stimulation (DBS) [17], transcranial magnetic stimulation (TMS) [11], and neurofeedback.

Table 5: Dual-use classification of 109 TARA techniques.

Classification	Definition	Count	%
Confirmed	Published clinical use exists	53	48.6%
Probable	Under active clinical investigation	16	14.7%
Possible	Theoretical therapeutic mapping	9	8.3%
Silicon Only	No tissue analog; purely digital	25	22.9%
Not Mapped	Dual-use assessment pending	6	5.5%

Of the 109 techniques, 78 (71.6%) have confirmed, probable, or possible therapeutic analogs. This finding underscores a fundamental challenge for BCI security: the same capabilities that must be defended against are often the capabilities that make BCIs therapeutically valuable.

#### 4.6 Representative Techniques

Table 6 shows five representative techniques spanning different categories, severities, and dual-use classifications.

Table 6: Five representative TARA techniques.

ID	Sev.	Status	Dual-Use	Description
T0001	Critical	Confirmed	Confirmed	Cortical signal injection via rogue electrode stimulation; therapeutic analog: deep brain stimulation
T0015	High	Demonstrated	Confirmed	P300 side-channel extraction of private information; therapeutic analog: P300-based spelling interfaces
T0034	High	Emerging	Probable	Calibration data poisoning during BCI training sessions; therapeutic analog: adaptive neurofeedback
T0072	Medium	Confirmed	Confirmed	Ultrasonic side-channel via bone conduction microphone; therapeutic analog: ABR audiometry
T0090	High	Demonstrated	Confirmed	WiFi CSI body sensing for respiratory and gait inference; therapeutic analog: sleep apnea detection

Figure 2 shows the severity distribution across all techniques, and Figure 3 illustrates the dual-use breakdown.

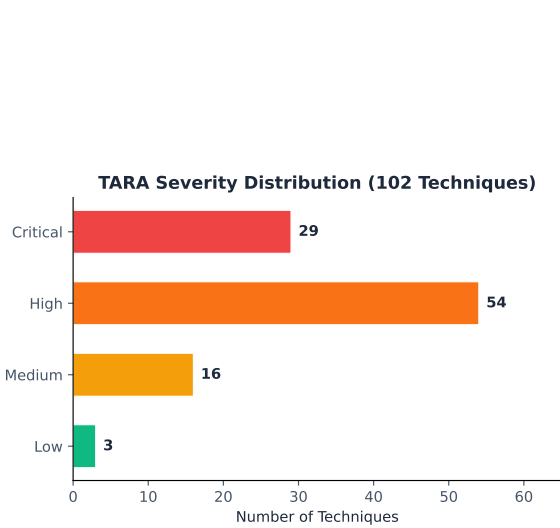


Figure 2: TARA severity distribution across 109 techniques.

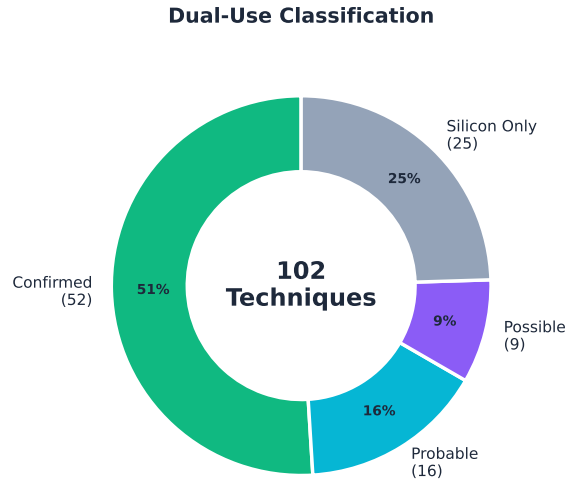


Figure 3: Dual-use classification: 71.6% of techniques have therapeutic analogs.

## 5 NISS: Neural Impact Scoring System

The Neural Impact Scoring System (NISS) is a proposed CVSS v4.0 extension designed to conform with FIRST.org’s official extension mechanism (User Guide §3.11) [10]. NISS adds six metrics that capture dimensions CVSS was never designed to express: biological tissue damage, cognitive reconnaissance, cognitive disruption, consent boundary violations, damage reversibility, and neuroplastic consequences.

NISS serves as the technique-level scoring input to the Neurosecurity Score (NSv2.1), a multi-layer risk framework that aggregates NISS scores across five neurorights dimensions using ISO 31000 risk methodology, physics-based feasibility gating, and population-adjusted context baselines. In short: NISS is the scoring metric; NSv2.1 is the validated risk framework that demonstrates NISS in operational context.

### 5.1 Gap Analysis: Why CVSS Alone Is Insufficient

We mapped all 109 TARA techniques to CVSS v4.0 base vectors and classified the results into three gap groups based on how much information CVSS alone fails to capture:

Table 7: CVSS v4.0 gap analysis across 109 TARA techniques.

Group	Gap Description	Count	Example
1	CVSS captures most impact; NISS adds nuance	13	Digital-only
2	CVSS captures exploitability but misses half	31	Mixed impact
3	CVSS fundamentally cannot express primary impact	64	Neural-dominant
<b>Techniques needing NISS extension</b>		<b>108</b>	<b>99.1%</b>

Group 3—where CVSS fundamentally cannot express the primary impact—contains the majority of techniques (58.7%). These are attacks where the most severe consequence is biological tissue damage, cognitive coercion, or irreversible neural harm—dimensions for which CVSS has no metric.

## 5.2 Extension Metrics

NISS defines six extension metrics, each with a graduated value set. The metrics are designed to be orthogonal to CVSS base metrics: they capture impact dimensions that exist only because the target system interfaces with biological neural tissue.

### 5.2.1 BI: Biological Impact

Direct harm to neural tissue, organs, or physiological function. This dimension has no equivalent in CVSS.

Value	Label	Score	Description
N	None	0.0	No tissue interaction or physical harm
L	Low	3.3	Temporary discomfort, minor sensory disruption, reversible tissue stress
H	High	6.7	Significant tissue damage, seizure induction, involuntary motor activation
C	Critical	10.0	Life-threatening or permanently disabling neural harm

### 5.2.2 CR: Cognitive Reconnaissance

Degree of unauthorized cognitive read access—thought decoding, neural data inference, intent extraction, memory readout. CVSS has no concept of thought privacy or cognitive surveillance. CR maps to the Cognitive Liberty and Mental Privacy neurorights dimensions.

Value	Label	Score	Description
N	None	0.0	No cognitive read access
L	Low	3.3	Partial intent inference, coarse emotional state detection
H	High	6.7	Full thought decoding, identity inference, or detailed neural data extraction
C	Critical	10.0	Complete cognitive surveillance—all neural activity readable in real time

### 5.2.3 CD: Cognitive Disruption

Degree of unauthorized cognitive write access—perception manipulation, identity modification, cognitive coercion, false stimuli injection. CD maps to the Mental Integrity neurorights dimension.

Value	Label	Score	Description
N	None	0.0	No cognitive write access or alteration
L	Low	3.3	Minor perceptual distortion, detectable interference
H	High	6.7	Perception manipulation, involuntary behavioral changes
C	Critical	10.0	Cognitive coercion, identity modification, or complete loss of cognitive autonomy

#### 5.2.4 CV: Consent Violation

Degree of violation of informed consent or cognitive autonomy. Ordered by severity: covert (implicit) violations are worse than detectable (explicit) ones.

Value	Label	Score	Description
N	None	0.0	Operating within explicitly consented boundaries
P	Partial	3.3	Action exceeds scope but subject retains some awareness
E	Explicit	6.7	Direct violation of consent boundaries, but detectable
I	Implicit (covert)	10.0	Covert manipulation the patient cannot detect or refuse

#### 5.2.5 RV: Reversibility

Whether the damage can be undone. IT assets can be restored from backup. Neural tissue cannot be rebooted.

Value	Label	Score	Description
F	Full	0.0	Effects fully reverse when attack stops
T	Temporary	3.3	Effects reverse over hours to days
P	Partial	6.7	Some effects permanent, some reversible
I	Irreversible	10.0	Permanent neural tissue destruction or cognitive change

#### 5.2.6 NP: Neuroplasticity

Whether the attack exploits or induces neuroplastic changes—the brain’s ability to rewire itself. This has no digital equivalent.

Value	Label	Score	Description
N	None	0.0	No neuroplastic effect
T	Temporary	5.0	Short-term synaptic changes that decay within hours–days
S	Structural	10.0	Long-term or permanent neural pathway changes

### 5.3 PINS Flag

NISS introduces the Potential Impact to Neural Safety (PINS) flag—a binary indicator triggered when:

$$\text{PINS} = \begin{cases} \text{true} & \text{if } \text{BI} \geq \text{High} \vee \text{RV} = \text{Irreversible} \\ \text{false} & \text{otherwise} \end{cases} \quad (1)$$

A PINS flag mandates immediate safety review regardless of overall score. Across all 109 techniques, 33 are PINS-flagged (30.3%).

#### 5.4 Scoring Formula

The NISS score is computed as the weighted mean of the six metric scores:

$$\text{NISS} = \frac{w_{\text{BI}} \cdot \text{BI} + w_{\text{CR}} \cdot \text{CR} + w_{\text{CD}} \cdot \text{CD} + w_{\text{CV}} \cdot \text{CV} + w_{\text{RV}} \cdot \text{RV} + w_{\text{NP}} \cdot \text{NP}}{w_{\text{BI}} + w_{\text{CR}} + w_{\text{CD}} + w_{\text{CV}} + w_{\text{RV}} + w_{\text{NP}}} \quad (2)$$

In the default profile, the weights are  $w_{\text{BI}} = w_{\text{CV}} = w_{\text{RV}} = w_{\text{NP}} = 1.0$  and  $w_{\text{CR}} = w_{\text{CD}} = 0.5$ . This normalization ensures that the split of the former Cognitive Integrity (CG) metric into Cognitive Reconnaissance (CR) and Cognitive Disruption (CD) in v1.1 does not inflate the cognitive dimension’s influence: at  $0.5 + 0.5 = 1.0$  out of a total weight sum of 5.0, cognition retains its original 20% share from the five-metric v1.0 formula. The CR/CD split reflects the fundamental distinction between read attacks (thought decoding, surveillance) and write attacks (perception manipulation, cognitive coercion)—attack vectors with different mitigations that map to different neurorights dimensions. NISS supports four context profiles with differential weights that override the default normalization:

- **Clinical:** Emphasizes BI, RV, and NP (patient safety focus)
- **Research:** Emphasizes CR and CV (consent and cognition focus)
- **Consumer:** Emphasizes CR and CV (privacy and consent focus)
- **Military:** Emphasizes BI, CR, and CD (dual-use concern)

#### 5.5 Vector Format

The NISS vector rides alongside the CVSS v4.0 base vector:

```
CVSS : 4.0 / AV : N / AC : L / AT : N / PR : N / UI : N /
VC : H / VI : H / VA : H / SC : N / SI : N / SA : N
NISS : 1.1 / BI : H / CR : H / CD : C / CV : I / RV : P / NP : S
```

This dual-vector architecture means security teams can triage using familiar CVSS scores while BCI-specific teams see the neural dimensions that determine whether a vulnerability is a software bug or a patient safety emergency.

#### 5.6 NISS Severity Distribution

Across all 109 techniques, the NISS severity distribution differs markedly from CVSS severity:

Table 8: NISS severity distribution (all 109 techniques).

NISS Severity	Count	%
High	21	19.3%
Medium	35	32.1%
Low	52	47.7%
None	1	0.9%

The NISS distribution is more uniform than CVSS because NISS captures impact dimensions that CVSS flattens into high/critical. Techniques rated “high” by CVSS may distribute across

low, medium, and high NISS scores depending on whether they involve biological tissue damage or are purely digital.

Figure 4 visualizes the gap between CVSS and NISS scoring.

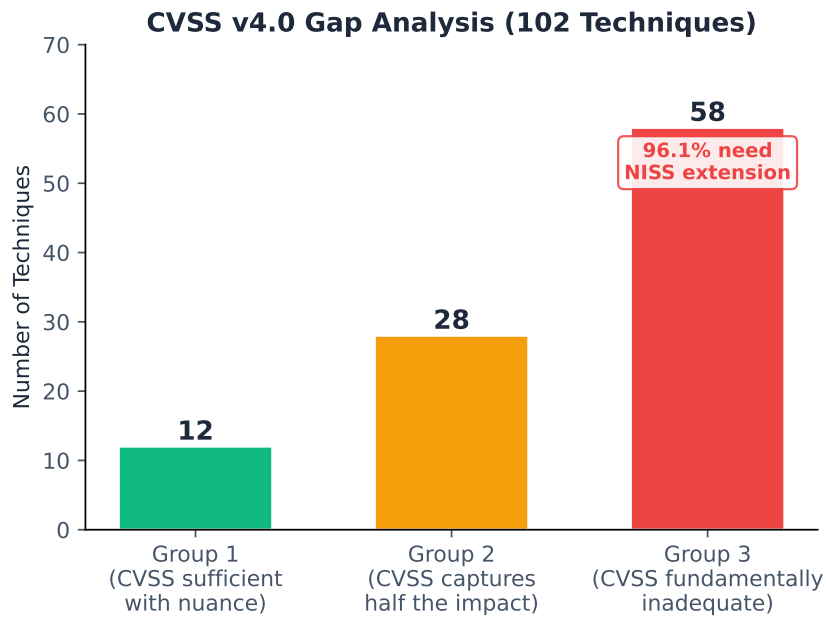


Figure 4: CVSS v4.0 vs. NISS gap analysis: 99.1% of techniques require extension metrics CVSS cannot express.

## 6 Neural Impact Chain

The Neural Impact Chain (NIC) is a six-stage methodology for mapping security vulnerabilities to clinical psychiatric diagnoses. To our knowledge, this is the first systematic pipeline connecting cybersecurity severity scoring to DSM-5-TR diagnostic codes [1]. The NIC answers a question no prior framework has addressed: *if this attack succeeds, what psychiatric condition could it cause or worsen?*

### 6.1 Pipeline Architecture

The NIC traces each technique through six stages:

1. **Technique:** The TARA attack technique (e.g., cortical signal injection)
2. **Hourglass Band:** Which band(s) the technique affects (e.g., N7 Neocortex, N6 Limbic)
3. **Neural Structure:** The anatomical structure at risk (e.g., hippocampus, prefrontal cortex, amygdala)
4. **Cognitive Function:** The function that structure supports (e.g., memory consolidation, executive control, emotional regulation)
5. **NISS Score:** The neural impact score, particularly the BI, CR, CD, and NP metrics that correlate with clinical outcomes
6. **DSM-5-TR Code:** The ICD-10-CM diagnostic code(s) for the psychiatric condition most closely associated with disruption of that function

Figure 5 illustrates this pipeline.

### Neural Impact Chain: From Technique to Diagnosis

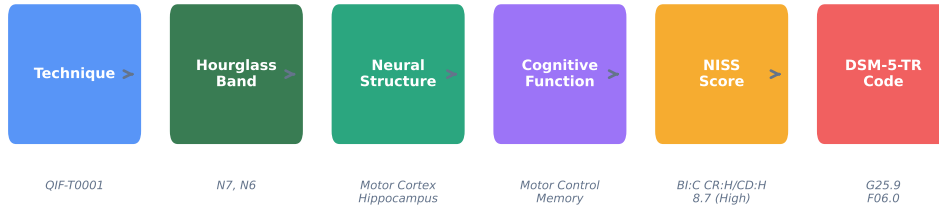


Figure 5: The Neural Impact Chain: six-stage pipeline from security technique to psychiatric diagnosis. Each arrow represents a mapping grounded in neuroanatomy, functional neuroscience, or clinical psychiatry.

## 6.2 NISS-to-DSM Bridge

The bridge between NISS metrics and DSM-5-TR diagnostic clusters is driven by which NISS metric dominates the technique’s profile:

- **BI-driven** → Motor/Neurocognitive cluster: biological impact implies tissue damage, leading to movement disorders or neurocognitive deficits
- **CR/CD-driven** → Cognitive/Psychotic cluster: cognitive reconnaissance or disruption implies perceptual or thought-process compromise
- **CV-driven** → Mood/Trauma cluster: Consent violations imply autonomy loss, mapping to trauma- and stressor-related disorders
- **NP/RV-driven** → Persistent/Personality cluster: neuroplasticity exploitation or irreversible damage implies lasting personality changes
- **No neural impact** → Non-diagnostic: Silicon-only techniques with no direct psychiatric mapping

## 6.3 Coverage Statistics

All 109 TARA techniques have been mapped through the NIC pipeline:

Table 9: Neural Impact Chain mapping results across 109 techniques.

Metric	Value
Techniques mapped	103 / 109 (94.5%)
Unique DSM-5-TR codes	29
Diagnostic clusters	5
Direct diagnostic risk	52 (47.7%)
Indirect diagnostic risk	9 (8.3%)
No diagnostic risk (silicon-only)	42 (38.5%)

## 6.4 Diagnostic Cluster Distribution

The five diagnostic clusters and their technique counts:

Table 10: DSM-5-TR diagnostic cluster distribution.

Cluster	Count	Representative DSM-5-TR Codes
Non-diagnostic	42	— (silicon-only techniques)
Mood/Trauma	21	F43.10 (PTSD), F32.9 (MDD), F44.9
Cognitive/Psychotic	17	F06.0 (psychosis), R41.3 (cognitive decline)
Motor/Neurocognitive	16	G25.9 (movement disorder), G31.84
Persistent/Personality	7	F07.0 (personality change)

The Mood/Trauma cluster is the largest diagnostic cluster (21 techniques), reflecting the prevalence of consent-violation and autonomy-disruption attacks in the BCI threat landscape. Techniques that covertly manipulate neural signals without the subject’s knowledge or consent map naturally to trauma- and stressor-related disorders.

## 6.5 Risk Classification

Each technique is classified by diagnostic risk:

- **Direct** (52 techniques, 47.7%): The attack mechanism can directly trigger or worsen the mapped psychiatric condition. Example: forced cortical stimulation causing seizures maps to epilepsy-related diagnostic codes.
- **Indirect** (9 techniques, 8.3%): The attack creates downstream conditions that may lead to the diagnosis. Example: sustained data exfiltration of private thoughts causing anxiety does not directly produce the anxiety disorder but creates the conditions for it.
- **None** (42 techniques, 38.5%): Silicon-only techniques with no direct neural interaction and thus no psychiatric diagnostic mapping.

## 6.6 Example Walkthrough

Consider technique **QIF-T0001: Cortical Signal Injection**—direct injection of adversarial signals via rogue electrode stimulation.

1. **Technique:** QIF-T0001 (cortical signal injection)
2. **Band:** N7 (Neocortex), N6 (Limbic System)
3. **Structure:** Primary motor cortex (M1), prefrontal cortex (PFC), hippocampus
4. **Function:** Motor control, executive function, memory consolidation
5. **NISS:** BI:C / CR:H / CD:H / CV:I / RV:P / NP:S → Score: 8.7 (High), PINS flagged
6. **DSM-5-TR:** G25.9 (movement disorder NOS), F06.0 (psychotic disorder due to another medical condition), F07.0 (personality change due to another medical condition)
7. **Risk class:** Direct—the stimulation itself can trigger seizures, involuntary movement, and perception distortion

Figure 6 shows the distribution of techniques across diagnostic clusters.

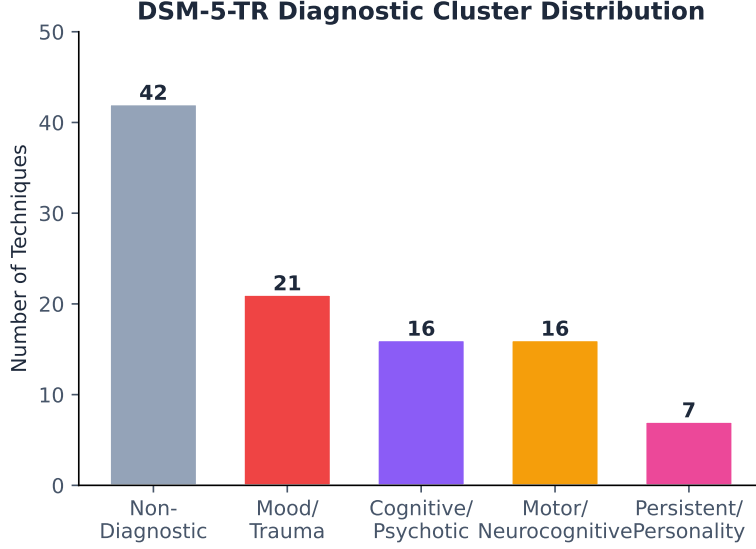


Figure 6: DSM-5-TR cluster distribution across 109 TARA techniques. 61 techniques (56.0%) have direct or indirect diagnostic risk.

## 7 Governance and Neuroethics

The QIF framework integrates neuroethics as a foundational design constraint rather than an afterthought. This section describes the consent tier system, alignment with international policy instruments, and regulatory mapping.

### 7.1 Consent Tiers

Each TARA technique is classified into one of four consent tiers based on the level of regulatory oversight required:

Table 11: Consent tier classification across 109 techniques.

Tier	Description	Requirement
Standard	Normal informed consent sufficient	Standard clinical consent
Enhanced	Additional safeguards required	Extended disclosure, monitoring
IRB	Institutional review board approval	Full IRB/ethics committee review
Prohibited	Not permissible under any consent	Technique must not be deployed

The distribution across tiers reflects the spectrum of BCI operations from routine monitoring (standard consent) to techniques that inherently violate cognitive autonomy (prohibited).

### 7.2 UNESCO Alignment

UNESCO’s 2025 Recommendation on the Ethics of Neurotechnology [31]—adopted by 194 Member States—is the first global normative framework for neurotechnology governance. It establishes three pillars: core values, ethical principles, and policy action areas.

The QIF framework addresses 15 of 17 UNESCO elements through technical implementation:

Table 12: QIF alignment with UNESCO Recommendation elements.

UNESCO Element	Type	Status	QIF Component
Human rights & dignity	Value	Implemented	Neurorights framework, consent tiers
Health & well-being	Value	Implemented	PINS flag, biological impact scoring
Diversity	Value	Implemented	Open-source, multi-stakeholder model
Sustainability	Value	Implemented	Apache 2.0 license, post-trial access
Professional integrity	Value	Implemented	Transparency audit trail
Proportionality	Principle	Implemented	Graduated severity scoring
Freedom of thought	Principle	Implemented	Cognitive liberty, consent states
Privacy	Principle	Implemented	Cognitive reconnaissance (CR) / disruption (CD) metrics
Protection of children	Principle	Implemented	Age-tiered consent framework
Consumer protection	Policy	Implemented	Default-deny architecture
Enhancement regulation	Policy	Partial	Technical infrastructure only
Workplace protections	Policy	Implemented	Mental privacy scoring
Behavioral influence	Policy	Implemented	Consent violation metric (CV)
Health and well-being	Policy	Implemented	NISS scoring, PINS flag
Oversight & governance	Policy	Implemented	Full regulatory mapping
Access & equity	Policy	Implemented	Open-source release

The two partially implemented elements—enhancement regulation and detailed implementation guidance for Member States—are explicitly outside the scope of a technical security framework and require policy collaboration.

### 7.3 Neurorights Framework Integration

The QIF framework implements the four neurorights proposed by Ienca and Andorno [13]:

1. **Cognitive Liberty:** Captured by the consent violation metric (CV). Any technique that operates without informed consent scores  $CV \geq \text{Partial}$ .
2. **Mental Privacy:** Captured by the cognitive reconnaissance metric (CR). Techniques that decode intent, extract memories, or infer identity score  $CR \geq \text{Low}$ .
3. **Mental Integrity:** Captured by the biological impact metric (BI) and neuroplasticity metric (NP). Physical harm to neural tissue or induced structural changes violate mental integrity.
4. **Psychological Continuity:** Captured by the reversibility metric (RV). Irreversible changes to neural function threaten the continuity of personal identity.

### 7.4 Regulatory Mapping

The framework maps to existing and emerging regulatory instruments:

- **FDA Section 524B** (FDORA/PATCH Act, 2022) [33]: Mandatory cybersecurity documentation for connected medical devices; TARA provides the neural threat taxonomy and NISS extends CVSS scoring as required by FDA premarket submissions
- **EU MDR 2017/745** [8]: Medical device risk management; TARA provides the threat taxonomy required for conformity assessment
- **ISO 14971** [15]: Risk management for medical devices; the NIC pipeline formalizes the harm pathway from technical failure to patient injury
- **HIPAA** [32]: Health data privacy; neural data classification extends protected health information categories
- **GDPR** [7]: Data protection; neural data as sensitive personal data under Article 9
- **Colorado Privacy Act** [30]: First US state to classify neural data as sensitive personal data (2024)

## 8 Case Studies

To demonstrate the practical difference between CVSS-only and CVSS+NISS scoring, we present five representative techniques scored with both systems. Each case illustrates dimensions that CVSS cannot capture.

**Important caveat:** Cases 1–4 are *threat model scenarios* derived from the TARA taxonomy. They represent plausible attack vectors based on known neuroscience and engineering principles, but have not been empirically executed against real BCI hardware. Case 5 is the only empirically confirmed vulnerability. This distinction is discussed further in Section 9.

### 8.1 Scoring Comparison

Table 13 presents five techniques spanning different categories, severity levels, and gap groups.

Table 13: CVSS v4.0 vs. NISS scoring for five representative techniques.

Technique	Description	CVSS	NISS
QIF-T0001	Cortical signal injection via rogue electrode	9.3 (Crit)	8.7 (High)
QIF-T0015	P300 side-channel private data extraction	7.7 (High)	3.3 (Low)
QIF-T0034	BCI calibration data poisoning	8.2 (High)	5.8 (Med)
QIF-T0050	Covert neural signal decoding for surveillance	7.1 (High)	5.0 (Med)
QIF-T0072	Ultrasonic bone-conduction side-channel	5.3 (Med)	2.7 (Low)

### 8.2 Case 1: Cortical Signal Injection (QIF-T0001)

**CVSS v4.0 assessment.** CVSS rates this technique as Critical (9.3) based on network attack vector, low complexity, no privileges required, and high impact on confidentiality, integrity, and availability. This accurately captures the exploitability and system impact.

**What CVSS misses.** The primary consequence of cortical signal injection is not system compromise—it is seizure induction, involuntary motor activation, and potential permanent tissue damage. The patient experiences a medical emergency, not a data breach. CVSS has no metric for biological harm, consent violation (the stimulation occurs without the patient’s knowledge), or irreversibility (neural tissue damage may be permanent).

**NISS extension.** NISS:1.1/BI:C/CR:H/CD:H/CV:I/RV:P/NP:S — Score: 8.7 (High), PINS flagged. The NISS vector captures that this technique causes critical biological impact (BI:C), high cognitive reconnaissance (CR:H) and high cognitive disruption (CD:H), covert consent violation (CV:I), partially irreversible damage (RV:P), and structural neuroplastic changes (NP:S). The PINS flag triggers mandatory safety review.

**NIC diagnostic mapping.** Via the Neural Impact Chain: N7/N6 bands → motor cortex, hippocampus → motor control, memory → G25.9 (movement disorder), F06.0 (psychosis due to medical condition). Risk class: direct.

### 8.3 Case 2: P300 Side-Channel (QIF-T0015)

**CVSS assessment.** Rated High (7.7) for passive eavesdropping with high confidentiality impact.

**What CVSS misses.** The extracted data is not files or credentials—it is private cognitive responses. The P300 ERP component reveals whether the subject recognizes a stimulus, enabling extraction of PINs, personal preferences, and identity information [19]. CVSS treats this as a confidentiality breach; the actual impact is cognitive reconnaissance—covert extraction of private mental content.

**NISS extension.** NISS:1.1/BI:N/CR:H/CD:H/CV:E/RV:F/NP:N — Score: 3.3 (Low). No biological impact (BI:N = 0), high cognitive reconnaissance (CR:H = 6.7) and high cognitive disruption (CD:H = 6.7), explicit consent violation (CV:E = 6.7), fully reversible (RV:F = 0), no neuroplastic effect (NP:N = 0). The arithmetic mean is  $(0 + 6.7 + 6.7 + 6.7 + 0 + 0)/6 = 3.3$ . The NISS score is significantly lower than CVSS because no physical harm occurs and the attack is fully reversible—but the CR:H flag alerts BCI-specific teams that private cognitive data is at risk.

#### 8.4 Case 3: Calibration Poisoning (QIF-T0034)

**CVSS assessment.** Rated High (8.2) for integrity impact during the BCI training phase.

**What CVSS misses.** Poisoned calibration data causes the BCI to learn incorrect mappings between neural signals and intended actions. The patient’s device responds to wrong signals or fails to respond to correct ones. Over time, neuroplasticity causes the brain to adapt to the corrupted interface, creating lasting neural pathway changes even after the poisoning is discovered and corrected.

**NISS extension.** NISS:1.1/BI:L/CR:H/CD:H/CV:I/RV:T/NP:T — Score: 5.8 (Medium). Low biological impact (BI:L = 3.3), high cognitive reconnaissance (CR:H = 6.7) and high cognitive disruption (CD:H = 6.7, the device misinterprets intent), covert consent violation (CV:I = 10.0, the patient doesn’t know calibration was corrupted), temporary reversibility (RV:T = 3.3), and temporary neuroplastic changes (NP:T = 5.0). The arithmetic mean is  $(3.3 + 6.7 + 6.7 + 10.0 + 3.3 + 5.0)/6 = 5.8$ .

#### 8.5 Case 4: Covert Neural Surveillance (QIF-T0050)

**CVSS assessment.** Rated High (7.1) for sustained confidentiality impact.

**What CVSS misses.** Continuous covert decoding of neural signals constitutes ongoing mental privacy violation. Unlike data exfiltration from a server, the “data” being stolen is the patient’s thoughts, emotional states, and cognitive patterns. The consent violation is maximal: the patient cannot detect or refuse the surveillance.

**NISS extension.** NISS:1.1/BI:N/CR:C/CD:C/CV:I/RV:F/NP:N — Score: 5.0 (Medium). No biological impact (BI:N = 0), critical cognitive reconnaissance (CR:C = 10.0, full thought decoding) and critical cognitive disruption (CD:C = 10.0), covert consent violation (CV:I = 10.0), fully reversible (RV:F = 0), no neuroplastic effect (NP:N = 0). The arithmetic mean is  $(0 + 10.0 + 10.0 + 10.0 + 0 + 0)/6 = 5.0$ . Despite the moderate aggregate score, the CR:C, CD:C, and CV:I flags distinguish this from ordinary data exfiltration—these maximum-severity cognitive and consent violations demand BCI-specific review.

## 8.6 Case 5: Real-World Vulnerability Disclosure

The QIF framework has been applied to real vulnerability research. During systematic analysis of the BCI software ecosystem, we identified a multi-phase exploit chain in an open-source library used in clinical and research BCI pipelines.

The exploit chain demonstrates escalation from synthetic-zone vulnerabilities (S2/S3 bands) to potential neural-zone impact (N-band: corrupted data reaching clinical decision-making). CVSS scores the software vulnerabilities accurately; NISS captures the downstream risk to patients whose clinical care depends on the integrity of the data stream.

Responsible disclosure is in progress. Specific vulnerability details, including affected software and CWE identifiers, will be published after coordinated disclosure concludes.

## 9 Limitations and Future Work

We present these limitations transparently to guide future validation efforts and to prevent overstatement of the framework’s current maturity.

### 9.1 No Empirical Validation on Real BCI Devices

The QIF framework has not been validated against operational BCI hardware. The TARA taxonomy was developed through literature review, threat modeling, and systematic analysis rather than penetration testing of actual neural devices. While the framework has been applied to one real software vulnerability (Section 8), this covers only the synthetic zone. Validation against neural-zone and interface-zone threats requires access to implanted BCI patients and clinical environments—resources unavailable to independent researchers.

### 9.2 DSM-5-TR Mapping Not Clinically Validated

The Neural Impact Chain maps security techniques to psychiatric diagnoses based on known neuroanatomical pathways and functional neuroscience. However, these mappings have not been reviewed or validated by psychiatrists or clinical neuroscientists. The mappings represent our best assessment of which diagnostic codes correspond to disruption of specific neural functions, but clinical validation is essential before these mappings can inform clinical decision-making.

### 9.3 NISS Weights Not Calibrated

The NISS scoring formula (Equation 2) uses equal weights (1.0) for all six metrics in the default profile. The four context profiles (Clinical, Research, Consumer, Military) propose differential weights, but these have not been calibrated against empirical data, expert elicitation, or clinical outcomes. Weight calibration requires:

- Expert panel scoring of representative scenarios
- Sensitivity analysis across weight configurations
- Correlation with observed clinical outcomes (when available)

### 9.4 No Interrater Reliability Study

NISS scores in the TARA registry were assigned by a single analyst (the author). No interrater reliability study has been conducted to assess whether independent scorers would assign the same metric values. CVSS interrater reliability is a known challenge [9]; NISS, with its novel neural-specific metrics, likely faces greater variability. A formal interrater reliability study with domain experts from both cybersecurity and neuroscience is needed.

## 9.5 Taxonomy Completeness

The TARA registry contains 109 techniques as of version 1.5. The BCI threat landscape is evolving rapidly, and additional techniques will emerge as: (a) new BCI devices reach market, (b) consumer neurotechnology proliferates, and (c) adversarial AI techniques advance. The current registry should be treated as a foundation, not a complete enumeration.

Of the 109 techniques, 31 are classified as Theoretical and 1 as Speculative—these have not been empirically demonstrated. While they are grounded in known physics and engineering principles, their practical feasibility remains unvalidated.

## 9.6 Single-Author Bias

The framework was developed by a single independent researcher. While multi-model AI verification was used throughout development (Claude, Gemini, ChatGPT), the architectural decisions, scoring assignments, and clinical mappings reflect a single perspective. Peer review and multi-disciplinary collaboration are essential for maturation.

## 9.7 AI Tool Disclosure

In accordance with arXiv, ACM, and IEEE policies on AI-assisted research, we disclose the following. Large language models (Claude 3.5/4, Gemini 2.0, ChatGPT-4o) were used during the development of this framework. Their roles, by section:

- **Literature review** (Sections 2, 3, 7): AI assisted in identifying relevant publications and summarizing prior work. All citations were independently verified by the author via DOI resolution and publisher page confirmation (see version history below).
- **Code generation** (Figures 1–6, TARA data pipeline): AI generated Python scripts for data analysis, visualization, and figure generation. All code was reviewed and modified by the author.
- **Writing assistance** (all sections): AI provided draft text that was substantially rewritten, restructured, and verified by the author. The proportion of AI-generated text retained verbatim is estimated at less than 15%; the remainder was rewritten or originated by the author.
- **Cross-validation**: AI models were used to cross-check factual claims across multiple systems (Claude, Gemini, ChatGPT) to reduce single-model bias. Disagreements between models were resolved by the author through primary source verification.

**Human-originated contributions:** The 11-band hourglass architecture, TARA threat taxonomy structure, NISS scoring methodology, Neural Impact Chain pipeline, all DSM-5-TR mappings, all NISS metric assignments, and all research conclusions were conceived, designed, and decided by the author. AI did not originate any architectural or methodological contribution. The author takes full responsibility for all content in this paper, irrespective of how it was generated. A complete, auditable transparency log documenting every AI contribution, human decision, and verification step is maintained at <https://github.com/qinnovates/qinnovate/blob/main/governance/TRANSPARENCY.md>.

An earlier version of this preprint (v1.0) contained citation errors introduced during AI-assisted bibliography construction, including three fabricated entries. These were corrected in v1.1 through a two-pass independent verification audit. All references have been verified against their source publications via DOI resolution, author publication pages, and database lookup. Version 1.2 corrected an internal percentage inconsistency and added the author responsibility statement above. Version 1.3 expanded the regulatory context (Section 2.2) with FDORA/PATCH Act Section 524B analysis and added Schroder et al. (2025) to the related work. Version 1.4 adds the missing Sherman & Guillery (2006) citation for thalamocortical relay logic in the hourglass derivation (Section 3), replaces a mislabeled DBS citation with Hallett (2007) for TMS, corrects

NISS case study scores to match the arithmetic mean formula (Equation 2), fixes the dual-use classification label, and completes a full citation and claims verification pass. This revision (v1.5) expands the TARA registry from 102 to 109 techniques, splits the NISS Cognitive Integrity (CG) metric into Cognitive Reconnaissance (CR) and Cognitive Disruption (CD) to distinguish read vs. write attacks (NISS v1.1, six metrics), normalizes default weights ( $w_{CR} = w_{CD} = 0.5$ , others = 1.0) to preserve the cognitive dimension at its pre-split 20% share, updates all statistics and tables to reflect the expanded registry, and adds neurorights dimension mapping for 103 of 109 techniques.

**Note on AI use and preprint status:** I did not find anything in Zenodo’s terms of use that prohibits using AI to help write a preprint, so I used it, heavily, to get this initial version out the door. The AI helped me move faster, but the ideas, the architecture, the scoring, the clinical mappings, those are mine. This preprint is me planting a seed. The real work is what comes next: writing it properly with human collaborators, getting the DSM-5-TR mappings reviewed by actual clinicians, running empirical validation, and revising it until it holds up to scrutiny. There is a lot more to do.

## 9.8 Future Work

1. **Reference implementation:** A software tool that automates NISS scoring and NIC mapping for new techniques
2. **Clinical validation:** Collaboration with psychiatrists to validate DSM-5-TR mappings
3. **Interrater reliability:** Formal study with cybersecurity and neuroscience domain experts
4. **FIRST.org registration:** Formal submission and registration of NISS as a CVSS v4.0 extension
5. **Empirical testing:** Penetration testing against BCI hardware in controlled environments
6. **Weight calibration:** Expert elicitation and sensitivity analysis for NISS context profile weights
7. **Conference paper:** Condensed version for submission to Graz BCI Conference 2026

## 10 Conclusion

Brain-computer interfaces are transitioning from laboratory prototypes to commercial medical devices. The security frameworks designed for information technology—while necessary—are insufficient for devices that read and write neural signals. A vulnerability in a BCI is not merely a software bug; it is a potential path to seizures, cognitive manipulation, privacy violation at the level of thought, and irreversible neural harm.

This paper presented the QIF framework: an integrated system comprising an 11-band hourglass architecture, a 109-technique threat taxonomy (TARA), a CVSS v4.0 extension for neural-specific scoring (NISS with six metrics), and the Neural Impact Chain—a first-of-its-kind methodology for mapping security vulnerabilities to DSM-5-TR psychiatric diagnoses. Analysis of all 109 techniques reveals that 99.1% require scoring dimensions CVSS cannot express, 71.6% have therapeutic dual-use analogs, and 56.0% pose direct or indirect psychiatric diagnostic risk.

The framework has significant limitations—no empirical validation on BCI hardware, no clinical validation of DSM-5-TR mappings, and single-author scoring—which we have documented transparently. These are not reasons to delay publication; they are invitations to collaborate. The BCI industry is moving faster than security standards. Neuralink, Synchron, and Blackrock Neurotech are implanting devices in patients today. The gap between what these devices can do and what security frameworks can assess grows wider each month.

The complete framework, threat registry, NISS specification, and scoring data are released as open source under the Apache 2.0 license. We invite collaboration from the neuroscience, neuroethics, cybersecurity, and clinical psychiatry communities. Formal registration of NISS with FIRST.org’s CVSS Special Interest Group is planned as future work.

The question is no longer whether BCI security frameworks are needed. The question is whether they will be ready before the first patient is harmed.

## References

- [1] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR)*. American Psychiatric Association Publishing, 2022. doi: 10.1176/appi.books.9780890425787. Verified 2026-02-16 via Crossref API.
- [2] Rebecca Bellan. OpenAI invests in Sam Altman’s brain computer interface startup Merge Labs. TechCrunch, January 2026. URL <https://techcrunch.com/2026/01/15/openai-invests-in-sam-altmans-brain-computer-interface-startup-merge-labs/>. \$252M seed round at \$850M valuation; co-founded by Sam Altman, Alex Blania, and Caltech researchers. Verified 2026-02-16 via URL check.
- [3] Tamara Bonaci, Ryan Calo, and Howard J. Chizeck. App stores for the brain: Privacy & security in brain-computer interfaces. *IEEE Technology and Society Magazine*, 34(2):32–39, 2015. doi: 10.1109/MTS.2015.2425551. Verified 2026-02-16 via Crossref API.
- [4] Carmen Camara, Pedro Peris-Lopez, and Juan E. Tapiador. Security and privacy issues in implantable medical devices: A comprehensive survey. *Journal of Biomedical Informatics*, 55:272–289, 2015. doi: 10.1016/j.jbi.2015.04.007. Verified 2026-02-16 via Crossref API.
- [5] Stephen Deering. Watching the waist of the protocol hourglass. Keynote, IETF 51 Plenary, 2001. URL <https://www.ietf.org/proceedings/51/slides/plenary-1/>. Originally presented at ICNP ’98. Foundational description of the Internet hourglass architecture. Verified 2026-02-16 via URL check.
- [6] Tamara Denning, Yoky Matsuoka, and Tadayoshi Kohno. Neurosecurity: Security and privacy for neural devices. *Neurosurgical Focus*, 27(1):E7, 2009. doi: 10.3171/2009.4.FOCUS0985. First use of the term “neurosecurity” in academic literature. Verified 2026-02-16 via Crossref API.
- [7] European Parliament and Council. General data protection regulation (GDPR) — regulation (EU) 2016/679, 2016. URL <https://gdpr-info.eu/>. Verified 2026-02-16 via URL check.
- [8] European Parliament and Council. Regulation (EU) 2017/745 on medical devices (MDR), 2017. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R0745>. Verified 2026-02-16 via URL check.
- [9] FIRST.org. Common vulnerability scoring system v4.0 specification, 2023. URL <https://www.first.org/cvss/v4-0/cvss-v40-specification.pdf>. Verified 2026-02-16 via URL check.
- [10] FIRST.org. CVSS v4.0 user guide, 2023. URL <https://www.first.org/cvss/v4.0/user-guide>. Section 3.11: Extension Framework. Verified 2026-02-16 via URL check.
- [11] Mark Hallett. Transcranial magnetic stimulation: A primer. *Neuron*, 55(2):187–199, 2007. doi: 10.1016/j.neuron.2007.06.026. Verified 2026-02-18 via PubMed PMID:17640522.
- [12] Daniel Halperin, Thomas S. Heydt-Benjamin, Benjamin Ransford, et al. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, pages 129–142, 2008. doi: 10.1109/SP.2008.31. Verified 2026-02-16 via Crossref API.

- [13] Marcello Ienca and Roberto Andorno. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(1):5, 2017. doi: 10.1186/s40504-017-0050-1. Verified 2026-02-16 via Crossref API.
- [14] Marcello Ienca and Pim Haselager. Hacking the brain: Brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18(2):117–129, 2016. doi: 10.1007/s10676-016-9398-9. Verified 2026-02-16 via Crossref API.
- [15] International Organization for Standardization. ISO 14971:2019 — medical devices — application of risk management to medical devices, 2019. URL <https://www.iso.org/standard/72704.html>. Verified 2026-02-16 via ISO catalog.
- [16] Eric R. Kandel, John D. Koester, Sarah H. Mack, and Steven A. Siegelbaum. *Principles of Neural Science*. McGraw-Hill, 6th edition, 2021. ISBN 978-1259642234. URL <https://neurology.mhmedical.com/book.aspx?bookid=3024>. Verified 2026-02-16 via McGraw-Hill Medical.
- [17] Andres M. Lozano, Nir Lipsman, Hagai Bergman, et al. Deep brain stimulation: Current challenges and future directions. *Nature Reviews Neurology*, 15:148–160, 2019. doi: 10.1038/s41582-018-0128-2. Verified 2026-02-16 via Crossref API.
- [18] Gabriel Lázaro-Muñoz, Michelle T. Pham, Katrina A. Muñoz, et al. Post-trial access in implanted neural device research: Device maintenance, abandonment, and cost. *Brain Stimulation*, 15(5):1029–1036, 2022. doi: 10.1016/j.brs.2022.07.051. Verified 2026-02-16 via Crossref API.
- [19] Ivan Martinovic, Doug Davies, Mario Frank, Daniele Perito, Tomas Ros, and Dawn Song. On the feasibility of side-channel attacks with brain-computer interfaces. In *Proceedings of the 21st USENIX Security Symposium*, pages 143–158. USENIX Association, 2012. URL <https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/martinovic>. Verified 2026-02-16 via URL check.
- [20] Lubin Meng, Xue Jiang, and Dongrui Wu. Adversarial robustness benchmark for EEG-based brain-computer interfaces. *Future Generation Computer Systems*, 143:231–247, 2023. doi: 10.1016/j.future.2023.01.028. Verified 2026-02-18 via Crossref API; volume/pages added.
- [21] MITRE Corporation. ATT&CK® framework, 2024. URL <https://attack.mitre.org/>. Verified 2026-02-16 via URL check.
- [22] Elon Musk and Neuralink. An integrated brain-machine interface platform with thousands of channels. *Journal of Medical Internet Research*, 21(10):e16194, 2019. doi: 10.2196/16194. Verified 2026-02-16 via Crossref API.
- [23] Katrina A. Muñoz, Kristin Kostick, Clarissa Sanchez, Lavina Kalwani, Laura Torgerson, Rebecca Hsu, Demetrio Sierra-Mercado, Jill O. Robinson, Simon Outram, Barbara A. Koenig, Stacey Pereira, Amy McGuire, Peter Zuk, and Gabriel Lázaro-Muñoz. Researcher perspectives on ethical considerations in adaptive deep brain stimulation trials. *Frontiers in Human Neuroscience*, 14:578695, 2020. doi: 10.3389/fnhum.2020.578695. Verified 2026-02-16 via Crossref API.
- [24] OECD. Recommendation on responsible innovation in neurotechnology, 2019. URL <https://legalinstruments.oecd.org/api/print?ids=658&Lang=en>. Verified 2026-02-16 via URL check.

- [25] Thomas J. Oxley, Peter E. Yoo, Gill S. Rind, et al. Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis. *Journal of NeuroInterventional Surgery*, 13:102–108, 2021. doi: 10.1136/neurintsurg-2020-016862. Verified 2026-02-16 via Crossref API.
- [26] Republic of Chile. Constitutional amendment on neurorights (article 19, no. 1), 2021. URL <https://www.bcn.cl/leychile/navegar?idNorma=1166983>. Ley 21.383. First country to constitutionally protect neurorights. Verified 2026-02-16 via URL check.
- [27] Michael Rushanan, Aviel D. Rubin, Denis Foo Kune, and Colleen M. Swanson. SoK: Security and privacy in implantable medical devices and body area networks. In *Proceedings of the 2014 IEEE Symposium on Security and Privacy*, pages 524–539, 2014. doi: 10.1109/SP.2014.40. Verified 2026-02-16 via Crossref API.
- [28] Tyler Schroder, Renée Sirbu, Sohee Park, Jessica Morley, Sam Street, and Luciano Floridi. Cyber risks to next-gen brain-computer interfaces: Analysis and recommendations. *arXiv preprint arXiv:2508.12571*, 2025. URL <https://arxiv.org/abs/2508.12571>. Also published in *Neuroethics* 18 (2025). Verified 2026-02-16 via arxiv.org and Springer.
- [29] S. Murray Sherman and R. W. Guillery. *Exploring the Thalamus and Its Role in Cortical Function*. MIT Press, 2nd edition, 2006. ISBN 978-0-262-51344-9. URL <https://mitpress.mit.edu/9780262513449/exploring-the-thalamus-and-its-role-in-cortical-function/>. Verified 2026-02-18 via MIT Press catalog.
- [30] State of Colorado. Colorado privacy act — amendment on biological and neural data, 2024. URL <https://leg.colorado.gov/bills/hb24-1058>. HB 24-1058. First US state to classify neural data as sensitive personal data. Verified 2026-02-16 via URL check.
- [31] UNESCO. Recommendation on the ethics of neurotechnology. Adopted at the 43rd session of the General Conference, Samarkand, 2025. URL <https://www.unesco.org/en/ethics-neurotech/recommendation>. Verified 2026-02-16 via URL check.
- [32] U.S. Congress. Health insurance portability and accountability act (HIPAA), 1996. URL <https://www.govinfo.gov/link/plaw/104/public/191>. Verified 2026-02-16 via URL check.
- [33] U.S. Congress. Section 524B of the FD&C act — ensuring cybersecurity of devices, 2022. URL <https://www.congress.gov/bill/117th-congress/house-bill/2617/text>. Added by Section 3305 of FDORA, Division FF of the Consolidated Appropriations Act, 2023 (Pub. L. 117-328). Effective March 29, 2023; RTA enforcement from October 1, 2023. Verified 2026-02-16 via Federal Register.
- [34] Francis R. Willett, Erin M. Kunz, Chaofei Fan, et al. A high-performance speech neuroprosthesis. *Nature*, 620:1031–1036, 2023. doi: 10.1038/s41586-023-06377-x. Verified 2026-02-16 via Crossref API.
- [35] Rafael Yuste, Sara Goering, Blaise Agüera y Arcas, et al. Four ethical priorities for neurotechnologies and AI. *Nature*, 551(7679):159–163, 2017. doi: 10.1038/551159a. Verified 2026-02-16 via Crossref API.
- [36] Hubert Zimmermann. OSI reference model — the ISO model of architecture for open systems interconnection. *IEEE Transactions on Communications*, 28(4):425–432, 1980. doi: 10.1109/TCOM.1980.1094702. Verified 2026-02-16 via Crossref API.