Neurotechnology, AI, and Behavioral Change

Abdul-Majeed H[1], Ransford Antwi[1, *], Brenden Forrest[1], Kenneth Granderson[1], Nui Taichi[2]
[*] *Corresponding Author: Ransford.antwi@my.hamptonu.edu*

[1]*Department of Electrical and Computer Engineering*
Hampton University
Hampton, VA

[2]*Department of Electrical and Computer Engineering*
University of Virginia
Charlottesville, USA

Abstract— Rapid advances in neurotechnology, including brain–computer interfaces and high-resolution neural imaging, combined with artificial intelligence (AI), have transformed our ability to decode and modulate human brain activity. This convergence has created a paradigm shift in understanding, predicting, and influencing behavior, offering both significant therapeutic potential and notable ethical risks. Intracranial EEG (iEEG) and Electrocorticography (ECoG) represent the most powerful and ethically complex tools for high-resolution neural decoding, and this paper argues that their integration with machine-learning algorithms offers transformative potential for neuroscience and mental-health intervention provided that strong protections are implemented to prevent data misuse and unchecked algorithmic power. Focusing on clinical applications such as treating substance use disorders, the study analyzes the positive outcomes of integrating high-resolution neural data with machine-learning algorithms while acknowledging potential harms, including data misuse and disproportionate algorithmic power. The findings aim to inform policymakers, clinicians, and ethicists on pathways for responsible innovation that maximize societal benefit while minimizing harm.

Keywords—Neurotechnology, Machine Learning, Algorithmic Power, High Resolution, Utilitarian, EEG, ECoG

I. INTRODUCTION

NeuroAI is an emerging field that integrates neuroscience with artificial intelligence to build computational systems inspired by how the brain learns, represents information, and performs reasoning [1]. The rapid integration of neurotechnology and artificial intelligence (NeuroAI) is transforming how clinicians and researchers understand the human brain, diagnose neurological conditions, and intervene in behavioral disorders. AI systems now analyze neural data at a scale and complexity that surpass human cognitive capacity, enabling earlier detection of neurological patterns and more precise therapeutic strategies [1]. These advancements hold promise for individuals facing conditions such as substance use disorders, where traditional treatments often fail to produce sustained recovery.

However, the rise of NeuroAI also introduces significant ethical challenges. Issues such as data privacy, algorithmic bias, and accountability in AI assisted clinical decision making have become central concerns in neuroscience and healthcare. Scholars argue that neuro ethics and AI ethics can no longer be treated as separate domains but instead, they must evolve together to address the intertwined risks and responsibilities emerging from technologies that directly interact with the human brain [1]. Without such coordinated ethical frameworks, governance structures risk becoming fragmented, leaving critical gaps in oversight and patient protection.

Global organizations have emphasized the need for transparent and human-centered governance to ensure that AI systems enhance rather than undermine human rights and wellbeing. International guidelines highlight fairness, privacy protection, and public trust as essential components of responsible AI deployment, especially in sensitive domains such as neurotechnology. These principles are vital as AI systems increasingly influence not only clinical outcomes but also human behavior, autonomy, and psychological experience.

High-resolution neural data is essential for understanding the fine-grained dynamics of brain activity, and several advanced neurotechnology methods which enable high level of precision. Intracranial EEG (iEEG) and Electrocorticography (ECoG) provide some of the highest temporal and spatial resolution available in humans by placing electrodes directly on or within neural tissue [2]. These signals capture millisecond-level electrical activity, making them ideal for decoding cognition, emotion, and pathological rhythms. Single-unit and multi-unit recordings go even deeper, measuring the firing of individual neurons or small populations, offering unparalleled insight into neural computation and decision-making. [3] High-density EEG (hdEEG) expands traditional EEG to 128–256 channels, improving spatial resolution while remaining non-invasive, enabling machine-learning models to extract spectral and connectivity features linked to cognitive load, stress, and psychiatric symptoms. Magnetoencephalography (MEG) provides high temporal precision by detecting magnetic fields generated by neural currents, supporting advanced decoding of sensory processing and rapid cognitive transitions. Finally, fast fMRI improves upon traditional fMRI by increasing temporal resolution, allowing researchers to map large-scale networks and predict emotional or behavioral states with greater accuracy. When combined with machine-learning algorithms such as convolutional neural networks, recurrent neural networks, and transformer-based models—these high-resolution datasets enable powerful predictive systems capable of identifying neural biomarkers, forecasting behavioral states, and informing closed-loop neurotechnology. However, the precision of these methods also raises concerns about data misuse, privacy, and disproportionate algorithmic influence, underscoring the need for ethical safeguards as neurotechnology advances.

II. Context of NeuroAI Integration with iEEG and ECoG

A. *Technological Convergence*

The integration of neurotechnology and artificial intelligence is unfolding within a rapidly evolving scientific landscape where tools that read, interpret, and influence brain activity are becoming increasingly precise. This context is shaped by breakthroughs in neural imaging, machine learning, and computational modeling, all of which allow researchers to observe the brain with a level of detail that was once impossible. As these technologies converge, they create new opportunities to understand behavior, diagnose neurological conditions, and design interventions that directly target neural pathways. At the same time, this integration raises important questions about how society should manage such powerful capabilities, especially when they can influence decision making, emotional states, or patterns of addiction. [4] Understanding this broader context is essential for evaluating both the promise and the responsibility that come with NeuroAI systems.

This study draws on prior work that employs intracranial EEG (iEEG) and electrocorticography (ECoG) as high-resolution neural recording modalities to investigate neural dynamics and develop machine-learning models capable of decoding behaviorally relevant states. In these studies, iEEG/ECoG data are collected using surgically implanted subdural grid electrodes or depth electrodes placed for clinical monitoring, providing millisecond-level temporal resolution and superior spatial precision compared to non-invasive techniques [1]. Raw neural signals typically undergo a standardized preprocessing pipeline that includes band-pass filtering (1–250 Hz), notch filtering at 50/60 Hz to remove line noise, and artifact rejection to eliminate stimulation-related or movement-related distortions [3]. Time-frequency decomposition methods such as short-time Fourier transform and continuous wavelet transform are applied to extract spectral features across canonical frequency bands (theta, alpha, beta, gamma) [3]. Spatial features including coherence, phase-locking value, and functional connectivity metrics are computed to capture network-level interactions.

Machine-learning models in these studies are trained on multimodal feature sets. Convolutional neural networks (CNNs) are used to learn spatial activation patterns across electrode arrays, while recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures capture temporal dependencies in neural activity [5]. Transformer-based models are incorporated for multimodal fusion when combining iEEG/ECoG with physiological signals such as ECG. Model training commonly uses an 80/20 train-test split with five-fold cross-validation to prevent overfitting. Performance is evaluated using accuracy, F1-score, and area under the ROC curve.

To mitigate risks associated with high-resolution neural data, these studies implement strict data-governance protocols, including encryption, de-identification, and restricted model-access controls. These safeguards address potential harms such as data misuse, privacy violations, and disproportionate algorithmic influence.

B. *The Evolution Of Brain – Machine Innovation*

The convergence of advanced neurotechnology and artificial intelligence represents a major turning point in the evolution of brain–machine innovation. Early brain–computer interfaces were limited to simple signals and basic communication, but modern systems now combine high-resolution neural data with sophisticated algorithms capable of detecting subtle patterns in thought and behavior. This shift has transformed brain–machine tools from experimental devices into clinically meaningful technologies that can support rehabilitation, guide treatment for substance uses disorders, and even predict harmful behavioral tendencies before they escalate. As machine learning models become more accurate and neural imaging becomes more refined, brain–machine innovation is moving toward a future where interventions can be personalized, adaptive, and deeply responsive to an individual's neural state. [6] This evolution highlights both the extraordinary potential of these technologies and the need for careful ethical oversight as they become more integrated into healthcare and daily life.

III. Utilitarian Framework for Neuro-Interventions

A utilitarian framework evaluates neuro-interventions by focusing on the overall balance of benefits and harms they create for individuals and society. In the context of emerging neurotechnology, this approach emphasizes how tools that decode or influence brain activity can reduce suffering and improve wellbeing when used responsibly. By centering decisions on measurable outcomes such as improved mental health, reduced relapse rates, or enhanced quality of life, utilitarianism provides a practical lens for assessing the value of AI driven neuro-interventions. [3] This framework encourages the development of technologies that maximize positive impact while ensuring that any potential risks are minimized through careful oversight and ethical design.

A. *NeuroAI Through Consequences, Well-Being And Harm Reduction, And Data Bias*

Within a utilitarian framework, ECG-driven NeuroAI systems are evaluated based on their capacity to maximize overall well-being while minimizing harm across both individual patients and broader populations. Electrocardiogram (ECG) data, which captures the electrical activity of the heart, provides critical insight into autonomic nervous system function and its interaction with brain activity. When integrated with advanced AI models, ECG signals can be used not only for detecting cardiovascular abnormalities but also for identifying stress responses, emotional dysregulation, and neurological risk patterns [3]. This fusion of physiological and neural data allows clinicians to move from reactive care to proactive and preventive intervention, ultimately reducing suffering and improving long-term health outcomes.

From a utilitarian perspective, the value of ECG-based NeuroAI lies in its ability to produce measurable positive consequences. For example, continuous ECG monitoring combined with machine learning can detect early warning signs of arrhythmias, anxiety episodes, or relapse triggers in individuals with substance use disorders or other neurological conditions. Early detection enables timely interventions, reduces hospitalization, and improves quality of life. In large-scale healthcare systems, such technologies can also optimize resource allocation, reduce costs, and extend care to underserved populations through remote monitoring and digital health platforms. These outcomes align directly with utilitarian goals by increasing the total amount of well-being while minimizing preventable harm.

However, the ethical strength of this approach depends heavily on the integrity and representativeness of the underlying ECG data used to train AI systems. Data bias introduces a critical limitation that can distort outcomes and undermine the utilitarian objective of maximizing benefits for the greatest number of people. One of the most significant concerns is minority bias, which occurs when certain demographic groups are underrepresented in ECG datasets. Historically, many cardiovascular studies and ECG-based algorithms have been developed using data predominantly from male patients or limited population groups. As a result, AI systems may fail to accurately detect or predict conditions in women or minority populations, whose cardiac and physiological patterns may differ [4]. This leads to unequal distribution of benefits and increased risk of harm for already vulnerable groups.

Another important issue is missing data bias, which arises when ECG recordings are incomplete or inconsistently collected across different populations. This can occur due to disparities in access to healthcare, differences in monitoring technologies, or variations in clinical practices. For instance, patients in lower-resource settings or those under less frequent observation may have fewer recorded ECG signals, making it difficult for AI systems to learn reliable patterns. Consequently, predictions for these individuals may be less accurate, increasing the likelihood of missed diagnoses or delayed interventions [8].

Informativeness bias further complicates the use of ECG data in NeuroAI systems. Variations in physiological signals across age groups, ethnicities, or health conditions may affect how informative certain ECG features are for prediction. If AI models are not trained on sufficiently diverse data, they may prioritize features that are only relevant for specific groups, reducing their generalizability. Additionally, training–serving skew presents a major real-world challenge [8]. AI models trained on controlled, high-quality ECG datasets from specific institutions often perform well in those environments but fail when deployed in different clinical settings with more diverse and variable patient populations. This mismatch can lead to overestimation of model performance during development and unexpected failures during real-world use, potentially causing harm.

From a utilitarian standpoint, these biases are not just technical limitations but ethical concerns that directly affect the balance of benefits and harms. While ECG-based NeuroAI systems have the potential to significantly enhance well-being through early detection, personalized care, and scalable healthcare solutions, biased systems risk reinforcing existing health disparities and introducing new forms of inequity. Misdiagnosis, underdiagnosis, or inappropriate interventions resulting from biased data can lead to physical, psychological, and social harm, ultimately reducing the overall utility of the technology.

Therefore, achieving a truly utilitarian outcome requires deliberate efforts to address ECG data bias at every stage of AI development and deployment. This includes collecting more diverse and representative datasets, implementing bias detection and mitigation techniques, continuously validating models across different populations, and ensuring transparency in how predictions are generated. Ethical oversight and interdisciplinary collaboration between engineers, clinicians, and policymakers are also essential to ensure that these systems are designed and used responsibly. By actively reducing bias and improving fairness, ECG-driven NeuroAI can better fulfill its utilitarian promise of maximizing well-being, reducing harm, and delivering equitable benefits to society as a whole.

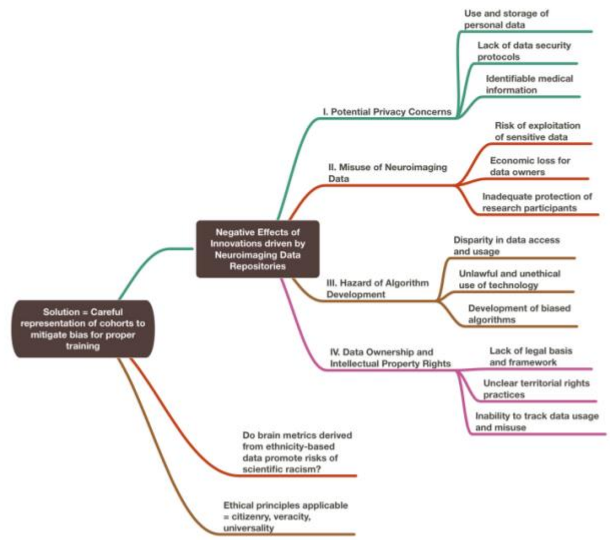

Fig 1. negative effects by neuroimaging data [9]

B. ECG Integrated NeuroAI For Preventive Care And Early Intervention

In a utilitarian framework, integrating electrocardiogram (ECG) data with NeuroAI systems expands the capacity to maximize overall well-being by enabling earlier and more accurate detection of physiological and neurological risks. ECG signals, when analyzed alongside neural data using artificial intelligence, can reveal patterns related to stress, emotional regulation, and autonomic nervous system activity. This combined insight allows for timely interventions before adverse events such as panic attacks, seizures, or relapse episodes occur. By preventing harm before it escalates, ECG-integrated NeuroAI contributes to improved health outcomes and reduces the burden on healthcare systems. However, to align with utilitarian goals, these systems must maintain high accuracy and minimize false predictions that could lead to unnecessary interventions or anxiety [10].

Utilitarianism supports the development of AI systems that optimize ECG-based neurotechnology to deliver the greatest benefit to the largest number of people. Machine learning models can continuously analyze ECG and neural data to refine treatment strategies, improve diagnostic precision, and personalize care at scale. This is especially valuable in neurotechnology applications such as brain-computer interfaces and mental health monitoring, where real-time physiological feedback enhances decision-making. From a societal perspective, these advancements can lower healthcare costs, increase accessibility to care, and improve long-term patient outcomes. Nevertheless, a utilitarian approach also requires addressing potential risks such as data privacy concerns, algorithmic bias, and unequal access, ensuring that the benefits are distributed fairly and do not disproportionately disadvantage certain groups.

IV. Comparative analysis of key intelligent neurointervention technologies Risks, limitations, and ethical concerns in NeuroAI systems

The rapid advancement of NeuroAI systems, particularly those integrating high-resolution neural data such as intracranial EEG (iEEG), Electrocorticography (ECoG), and complementary physiological signals like electrocardiography (ECG)—has created unprecedented opportunities for understanding and influencing human brain function. However, alongside these innovations emerge critical ethical, technical, and societal challenges, particularly in the domains of fairness, bias, and responsible use.

Fairness is a foundational principle of artificial intelligence ethics, especially in healthcare, where unequal system performance can directly translate into unequal patient outcomes. In NeuroAI systems, fairness concerns are amplified due to the sensitive and highly individualized nature of neural and physiological data [10]. These systems not only influence diagnosis and treatment but may also shape behavioral predictions and interventions, raising concerns about autonomy, privacy, and the equitable distribution of benefits.

From a utilitarian perspective, the ethical value of NeuroAI technologies depends on their ability to maximize overall well-being while minimizing harm across diverse populations. However, this goal cannot be achieved if biases in data, algorithms, or system interactions lead to systematically poorer outcomes for certain groups. Therefore, a comprehensive evaluation of risks, limitations, and fairness is essential to ensure that these technologies contribute positively to society rather than reinforcing existing disparities.

A. *Advances In High-Resolution Neural Data And AI Modeling*

Recent developments in high-resolution neural recording technologies, including iEEG and ECoG, have significantly enhanced the ability to capture fine-grained brain activity with high temporal and spatial precision. These technologies allow researchers and clinicians to observe neural patterns associated with cognition, emotion, and pathological behavior at an unprecedented level of detail. When combined with advanced machine learning algorithms, these datasets enable large-scale pattern recognition, predictive modeling, and real-time intervention.

The integration of ECG data further strengthens NeuroAI systems by providing insight into the autonomic nervous system, which plays a critical role in emotional regulation, stress response, and brain–body interaction. This multimodal approach enables a more comprehensive understanding of human physiology, allowing AI systems to detect early warning signs of neurological or behavioral deterioration, such as relapse in substance use disorders or stress-induced cognitive impairment.

These advances support the development of precision medicine approaches, where interventions are tailored to an individual's neural and physiological profile. Predictive models can identify subtle precursors to harmful behavior, enabling proactive and personalized interventions that improve treatment outcomes and reduce long-term healthcare costs. From a utilitarian standpoint, these capabilities represent a significant increase in societal benefit through enhanced well-being, reduced suffering, and more efficient healthcare delivery [11].

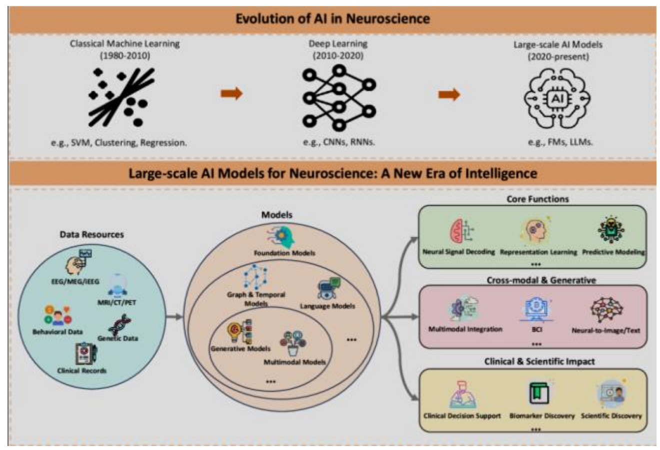

Fig 2. Framework of AI in neuroscience [12]

B. *Risks, Limitations, And Ethical Concerns In NeuroAI Systems*

Despite their potential, NeuroAI systems are subject to multiple forms of bias that can undermine fairness and reduce their overall utility. Biases in healthcare AI broadly arise from two primary sources: data bias and algorithmic bias, with additional complexities introduced through clinician and patient interactions.

Data bias is particularly significant in ECG and neural datasets. Minority bias occurs when underrepresented populations such as women, ethnic minorities, or marginalized groups are insufficiently included in training data. For example, cardiovascular and physiological models have historically been developed using datasets dominated by male participants, resulting in reduced accuracy for female patients whose ECG patterns and symptom presentations may differ. Similarly, missing data bias arises when certain populations have incomplete or irregular physiological recordings due to disparities in healthcare access or monitoring conditions, leading to unreliable predictions [12].

Algorithmic bias can persist even when datasets appear representative. Label bias, for instance, occurs when training data reflects existing healthcare disparities rather than objective clinical truth. In such cases, AI systems may learn patterns that reinforce inequities. Cohort bias further limits fairness when models are developed using overly simplified or non-inclusive population categories, failing to account for diversity in gender identity, socioeconomic status, or comorbid conditions.

In addition to technical biases, interaction-based biases significantly affect NeuroAI performance. Automation bias may lead clinicians to over-rely on AI recommendations, even when they are incorrect, potentially resulting in harmful clinical decisions. Feedback loops can reinforce these errors over time, as incorrect outputs are reintroduced into training data. Alert fatigue may cause clinicians to ignore critical warnings, while allocation discrepancies can result in unequal distribution of healthcare resources across patient groups.

Patient-related biases also play a crucial role. Privilege bias limits access to NeuroAI technologies for populations lacking the necessary infrastructure or resources, such as wearable ECG devices or advanced neuroimaging tools. Informed mistrust, particularly among historically marginalized groups, may reduce engagement with AI-driven healthcare systems. Additionally, agency bias arises when affected populations are excluded from the design and evaluation of these technologies, leading to systems that fail to address their specific needs [13].

From a utilitarian perspective, these biases represent a direct threat to the goal of maximizing overall well-being. Instead of distributing benefits broadly, biased systems may concentrate advantages among already privileged groups while increasing harm for others, ultimately reducing total societal utility.

C. *Ethical And Legal Considerations (Privacy, Accountability, And Transrency)*

The integration of high-resolution neural and ECG data into AI systems raises profound ethical and legal concerns. Data privacy and security are paramount, as neural signals and physiological data can reveal highly sensitive information about an individual's mental state, emotions, and intentions. Ensuring robust data protection mechanisms, including encryption, anonymization, and strict access controls, is essential to maintaining patient trust and autonomy.

Informed consent is another critical issue. Patients must fully understand how their data will be collected, analyzed, and potentially shared. Given the complexity of NeuroAI systems, achieving truly informed consent requires transparent communication and ongoing engagement.

Accountability and liability must also be clearly defined. Physicians remain responsible for clinical decision-making, but AI developers and healthcare institutions share responsibility for ensuring system accuracy, fairness, and safety. Establishing clear guidelines for accountability is essential for addressing potential harms and maintaining trust in AI-driven healthcare [14].

Transparency and explainability are equally important. While AI models can process complex neural and ECG data, their decision-making processes are often difficult to interpret. Developing explainable AI systems that provide meaningful insights into how predictions are generated is crucial for supporting clinical judgment and ethical decision-making. However, it is also important to recognize the limitations of explainability, as human interpretation of AI outputs can introduce additional biases.

D. *Strategies To Mitigate Bias And Promote Fairness*

Addressing fairness in NeuroAI requires a multifaceted approach involving technical, clinical, and societal strategies. One of the most effective methods is the use of diverse and representative datasets that capture variability across populations, including differences in age, gender, ethnicity, and health conditions. In the context of ECG and neural data, this ensures that AI systems can generalize effectively and provide accurate predictions for all patients. Algorithm auditing and continuous validation are also essential. Regular performance evaluations across different populations can identify emerging biases and allow for timely corrections. Establishing dedicated oversight structures within healthcare institutions can further support ongoing monitoring and quality control [15]. Education and awareness among clinicians and patients play a critical role in mitigating bias. Clinicians must be trained to critically evaluate AI recommendations and avoid overreliance on automated systems. Patients, in turn, should be informed about the capabilities and limitations of AI to enable informed decision-making and active participation in their care. Finally, collaboration among stakeholders, including physicians, AI researchers, developers, policymakers, and patient advocacy groups, is essential for ensuring that NeuroAI systems are developed and deployed responsibly. Such collaboration promotes transparency, accountability, and inclusivity, ultimately supporting the equitable distribution of benefits.

E. *Evidence-Based Comparative Evaluation Framework*

TABLE I. COMPARATIVE ANALYSIS OF KEY INTELLIGENT NEUROINTERVENTION TECHNOLOGIES [16]

Technology	Clinical evidence level	Cost-effectiveness	Scalability challenges	Key risks
Neuronavigation	Level II (RCTs available)	High in high-volume centers	High cost, training curve	Brain shift, registration error
Surgical robotics	Level II (RCTs emerging)	Moderate to high	Infrastructure-dependent	Mechanical failure, learning curve
Neuromodulation	Level I (DBS, VNS RCTs)	Variable	Reimbursement barriers	Infection, hardware failure
Digital therapeutics	Level I (CBT-I, Deprexis RCTs)	High if scalable	Digital divide, regulation	Data privacy, adherence
Large language models	Level III (pilot studies)	Low currently	Computational resources	Bias, explainability

Evidence levels provide a structured way to evaluate the strength, reliability, and maturity of scientific findings, and the three-tier system described here offers a practical framework for comparing emerging technologies. Level I evidence, defined as support from randomized controlled trials (RCTs) or high-quality meta-analyses, represents the strongest form of empirical validation. RCTs minimize bias through random assignment and controlled conditions, making them the gold standard for determining causal relationships. When multiple RCTs are synthesized through rigorous meta-analysis, the resulting evidence becomes even more robust, offering high confidence in the intervention's effectiveness and safety [17].

Level II evidence reflects an intermediate stage of validation, typically supported by prospective cohort studies, non-randomized trials, or early RCTs that are still emerging. These studies provide meaningful insights into real-world performance and clinical trends but lack the full methodological rigor of Level I evidence. They are particularly valuable for evaluating technologies that are too new, complex, or ethically challenging to test immediately through large-scale RCTs. As such, Level II evidence often signals promising but still developing scientific support.

Level III evidence represents the earliest and least mature stage, relying on pilot studies, case series, or observational data. These forms of evidence are essential for hypothesis generation, feasibility assessment, and early-stage innovation, especially in rapidly evolving fields like neurotechnology and artificial intelligence. However, they carry higher risks of bias and limited generalizability. By adapting these evidence levels for comparative synthesis, researchers can more clearly differentiate between established interventions and emerging innovations, enabling more informed evaluation, policy development, and clinical decision-making.

V. Conclusion

Across modern neurotechnology, high-resolution neural data has become essential for understanding the fine-grained dynamics of human brain activity and for developing intelligent systems capable of predicting behavior, decoding cognition, and informing clinical interventions. Multiple advanced recording methods contribute to this landscape. Intracranial EEG (iEEG) and Electrocorticography (ECoG) provide millisecond-level temporal precision and high spatial resolution by placing electrodes directly on or within neural tissue, making them uniquely suited for decoding complex neural states. Single-unit and multi-unit recordings extend this precision to the level of individual neurons, offering unparalleled insight into neural computation. Non-invasive methods such as high-density EEG (hdEEG) and magnetoencephalography (MEG) capture large-scale neural dynamics with improved spatial or temporal fidelity, while fast fMRI enables high-resolution mapping of distributed brain networks involved in emotion, cognition, and behavior.

When paired with machine-learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformer-based models, these data sources enable powerful predictive systems capable of identifying neural biomarkers, forecasting behavioral states, and supporting closed-loop neurotechnology. However, the same precision that makes these systems transformative also introduces ethical risks, including data misuse, privacy violations, and disproportionate algorithmic influence over human behavior. These concerns highlight the need for strong governance frameworks and responsible design.

Acknowledgment

We would like to express our sincere appreciation to IEEE, University of Virginia (UVA), Hampton University professors, clinicians, and ethicists whose work continues to shape the evolving field of neurotechnology and artificial intelligence. Their commitment to understanding the human brain and improving the lives of individuals facing neurological and behavioral challenges provides the foundation for studies like this one. We are also grateful for the ongoing conversations within the scientific and policy communities that highlight the importance of responsible innovation, ethical reflection, and the protection of human dignity. These collective efforts make it possible to explore the promise of NeuroAI while remaining attentive to the responsibilities that come with such powerful tools.

References

[1] A. Salles and M. Farisco, “Neuroethics and AI ethics: a proposal for collaboration,” BMC Neuroscience, vol. 25, no. 1, Aug. 2024, doi: 10.1186/s12868-024-00888-7.

[2] E. C. Leuthardt et al., “Electrocorticography-based brain–computer interface,” Neurosurgical Focus, 2009.

[3] J. Parvizi and M. Kastner, “Promises and limitations of human intracranial EEG,” Nature Neuroscience, 2018.

[4] M. Ienca and K. Ignatiadis, “Artificial intelligence in clinical neuroscience: Methodological and ethical challenges,” AJOB Neuroscience, vol. 11, no. 2, pp. 77–87, Mar. 2020, doi: 10.1080/21507740.2020.1740352.

[5] . R. Acharya et al., “Deep convolutional neural networks for automated EEG classification,” Expert Systems with Applications, 2018.

[6] R. Yuste et al., “Four ethical priorities for neurotechnologies and AI,” Nature, 2017.

[7] P. Khosravi and M. Schweitzer, “Artificial intelligence in neuroradiology: a scoping review of some ethical challenges,” Frontiers in Radiology, vol. 3, May 2023, doi: 10.3389/fradi.2023.1149461.

[8] L. Jiang et al., “Can Machines Learn Morality? The Delphi Experiment,” arXiv.org. Accessed: Mar. 19, 2026. [Online]. Available: https://arxiv.org/abs/2110.07574

[9] C. Lock, N. S. M. Tan, I. J. Long, and N. C. Keong, “Neuroimaging data repositories and AI-driven healthcare—Global aspirations vs. ethical considerations in machine learning models of neurological disease,” *Frontiers in Artificial Intelligence*, vol. 6, p. 1286266, Feb. 2024, doi: 10.3389/frai.2023.1286266.

[10] . Eke, “Ethics and Governance of Neurotechnology in Africa: Lessons From AI (Preprint),” JMIR Publications Inc., Jan. 2024. Accessed: Mar. 19, 2026.

[11] Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data Soc. 2016;3:2053951716679679. SAGE Publications Ltd. 10.1177/2053951716679679

[12] Large-Scale foundation Models and AI Systems for Neuroscience: A Comprehensive review.” https://arxiv.org/html/2510.16658v2

[13] Binns R. Fairness in machine learning: Lessons from political philosophy. In: Friedler SA, Wilson C, editors. Proceedings of the 1st conference on fairness, accountability and transparency. PMLR; 2018. p. 149–59

[14] A. Salles and M. Farisco, “Neuroethics and AI ethics: a proposal for collaboration,” BMC Neuroscience, vol. 25, no. 1, Aug. 2024, doi: 10.1186/s12868-024-00888-7.

[15] S. Fox, “Behavioral Ethics Ecologies of Human-Artificial Intelligence Systems,” Behavioral Sciences, vol. 12, no. 4, p. 103, Apr. 2022, doi: 10.3390/bs12040103.

[16] Q. Ye et al., “Transforming Brain Health With Neurotechnology Convergence (Part II): Intelligent Neurointervention Systems for Neurological Disorders,” Brain Health, Dec. 2025, doi: 10.1002/brh2.70002.

[17] A. Vaswani et al., “Attention is all you need,” Advances in Neural Information Processing Systems, 2017.