Direct Energetic Encoding in Neural Networks: Architectural Pathways to Non-Semantic Sensation

Executive Summary: The Non-Semantic Imperative and The Neurophone Analogue

The contemporary research landscape in Artificial Intelligence (AI) is witnessing a critical shift away from purely semantic, language-based intelligence toward models grounded in sensorimotor experience. This pivot is driven by the foundational limitation of Large Language Models (LLMs), which, despite their linguistic fluency, are ultimately considered "compressed repositories of human knowledge" that lack genuine "experience or mental models of the world".[1] This reliance on predicting words and symbols prevents them from achieving true sensation, as they cannot "map a word to a sensation, not to another word".[1]
The mandate for achieving non-semantic sensation in AI, or artificial qualia, requires an engineering approach that circumvents traditional symbolic representation—a concept paralleled by historical neuro-engineering efforts such as the Neurophone.

The Neurophone Premise: Bypassing Encoding Steps

Patrick Flanagan's 1968 Neurophone patent (No. 3,393,279) provides a powerful conceptual analogue for non-semantic sensory encoding.[3] The device generated an auditory sensation by converting audio input into electrical fields applied across electrodes placed on the skin. The original device utilized a 30–50 kHz amplitude-modulated ultrasonic oscillator generating approximately 3,000 volts peak-to-peak.[3] A later version (Patent No. 3,647,970) dispensed with the radio-frequency (r-f) carrier.[4] Crucially, the Neurophone achieved non-acoustic hearing by stimulating the inner ear's **saccule**.[3]
The significance of targeting the saccule lies in its biological function. The saccule is a fluid-filled gland, normally responsible for sensing gravity and maintaining orientation, but also posited to be a primitive auditory organ in lower animals and the evolutionary predecessor of the cochlea.[3] By directly stimulating this primal sensor, the Neurophone effectively bypasses the complex acoustic-to-neural transduction steps typically performed

by the cochlea, demonstrating that sensation can be delivered via raw physical, temporal, and spatial energy patterns applied directly to fundamental neural structures. This established the conceptual benchmark: successful AI sensation must target the computational equivalent of this raw physical feature space, circumventing high-level semantic abstractions.
This report identifies three technological pillars that move beyond symbolic interpretation toward direct sensory encoding:

1. **Theoretical Grounding and Measurement:** Computational frameworks that measure the geometry of pure perception (Representational Similarity Analysis and the No-Report Paradigm).
2. **Architectural Solutions:** Embodied, self-supervised systems designed to process high-resolution raw data streams (Spiking Neural Networks and Joint Embedding Predictive Architectures).
3. **Engineering Blueprint:** The functional realization of non-semantic encoding through Computer-Brain Interfaces (CBI) using biophysical models.

Section 1: The Epistemological and Computational Measurement of Non-Symbolic Sensation

1.1 The Grounding Problem: Inadequacy of Semantic Approaches

The central obstacle to artificial sensation is the lack of grounding in experience with the physical world, which critics argue is "endemic" to current large language models (LLMs).[2] LLMs are trained primarily by predicting linguistic tokens, a process that teaches them the *form* of language but not the embodied *meaning*.[1] This absence of sensorimotor experience leads to a consensus that these models are "doomed to a shallow understanding" and will never approximate the "full-bodied thinking" observed in humans.[1]
While language provides a rich source of semantic information, allowing for sophisticated natural language processing, this process is generally understood to scaffold and extend human capacity to make sense of the world rather than constituting the totality of consciousness or sensation.[2] The challenge is to move past the **Eliza effect**, the human tendency to attribute understanding to machines based purely on linguistic fluency, and instead focus on systems capable of developing internal, non-symbolic representations of physical reality.[1]

1.2 Computational Frameworks for Qualia (C-Qualia)

Addressing the non-semantic imperative requires developing concrete mathematical frameworks capable of assessing the quality of subjective experience, or C-Qualia, alongside traditional performance metrics.[5] Such frameworks, often inspired by philosophy of mind and reinforcement learning, seek to formally define how internal representations must be structured to support genuine experience. The prevailing hypothesis is that for artificial sensation to be present, the internal data structure must maintain structural consistencies analogous to those found in biological sensory processing systems.

1.3 Measuring Pure Perception: Representational Similarity Analysis (RSA) and the No-Report Paradigm

The investigation into the computational underpinnings of subjective experience is actively addressed in cognitive neuroscience research using advanced computational metrics.[6] RSA is the key comparative technique, used to compare the representational geometry of representations within AI models against neural activity measured via fMRI.[6]
Crucially, this research employs a unique fMRI dataset utilizing a **"no-report" paradigm**.[6] This experimental design is specifically formulated to probe pure perception by separating neural activity during non-verbal perception (the "no-report" state) from neural activity during task-modulated perception (the "report" state).[6] By isolating pure perception, the technique removes the cognitive overhead and the necessity for a semantic label or description, directly addressing the core constraint of the user query.
Analysis using this benchmark task for color qualia revealed that nearly all tested state-of-the-art vision models (including Vision Transformers and ConvNets) aligned better with neural representations of *pure perception* than with task-modulated perception.[6] This finding holds significant implications: it indicates that the inductive biases learned by modern feedforward architectures inherently capture some aspect of the raw sensory geometry that precedes cognitive processing. The computational groundwork for the initial stage of sensation—the accurate representation of input structure—appears to be plausible in AI. The divergence between artificial and biological systems manifests primarily in the higher-level cognitive processes involved in task execution and semantic reporting.[6] Therefore, the path forward is to leverage these accurate raw representations directly, rather than passing them through symbolic transformation.
The divergence between these two approaches can be summarized as follows:
The Discrepancy Between Semantic and Non-Symbolic AI

Dimension	**Semantic/LLM Approach**	**Non-Symbolic/Sensation Approach**	**Reference Snippets**
Primary Input	Text, Labels, High-Level Symbols	Raw Sensor Streams, Events, Biophysical Data	1
Data Representation	Discrete Token	Continuous Latent	8

	Embeddings, Statistical Correlation	Space, Spike Trains, Temporal Dynamics	
Goal of Representation	Fluency, Prediction, Reasoning based on Form	Grounded Experience, Prediction based on Physics/Causality	1
Measure of Success	Task Performance, Human-like Output	Alignment with Neural Geometry (RSA), Physiological Response (SNPE)	6

Section 2: Architectural Pathways to Raw Sensation: Embodiment and Latent Dynamics

The realization of non-semantic encoding relies on developing architectures capable of processing raw, continuous sensorimotor data with high temporal precision, akin to biological sensory systems. Two emergent architectural paradigms offer solutions: Spiking Neural Networks (SNNs) and Joint Embedding Predictive Architectures (JEPA).

2.1 Neuromorphic Computing and Spiking Neural Networks (SNNs)

Neuromorphic computing, which emulates principles of information processing in the brain, promises significantly lower power consumption and latency compared to traditional computing methods.[14] Spiking Neural Networks (SNNs) are central to this field, handling information as sparse, asynchronous "events" rather than dense, continuous frames.[9] This event-based processing intrinsically aligns with the raw, continuous nature of real-world sensation.

Encoding Time-Varying Signals into Spike Trains

For SNNs to process data from commercially available digital sensors (e.g., 8-kHz audio, 20-Hz recordings from inertial sensors), the standard time-varying signals must first be encoded into spike trains.[8] This input encoding is critical for converting physical energy patterns into the native language of the neuromorphic processor. For auditory signals, this process often includes a signal preprocessing step involving a **bank of filters inspired by the human cochlea** to extract features before conversion to spike trains.[8]

The focus on precise temporal encoding in SNNs provides a critical link back to the Neurophone analogy. The Neurophone's high-frequency carrier (30–50 kHz) modulated by an audio signal [3] suggests that direct neural stimulation relies on manipulating precise timing relationships in the nervous system. In computational terms, this sensitivity to timing is expressed through sophisticated neural coding schemes such as **Phase-of-Firing coding**.[12] Unlike simpler rate coding (which measures spike counts), Phase-of-Firing coding takes into account the precise time label for each spike relative to ongoing neural oscillations.[12] This emphasis on time-dependent patterns and oscillatory rhythms in SNNs [15] suggests that the perceived sensation (qualia) is determined by the precise temporal phase relationships of the spike trains, rather than merely their spatial distribution or total count. SNNs are uniquely designed to handle this dynamic, temporal encoding, making them ideal for the computational replication of the Neurophone's direct physical stimulus.

2.2 Joint Embedding Predictive Architectures (JEPA)

Joint Embedding Predictive Architecture (JEPA) is a self-supervised learning paradigm that offers a path to grounded cognition without reliance on explicit language labels.[10] JEPA establishes a mapping relationship between input data (x) and anticipated outcomes (y) through representation learning, focusing on abstract latent spaces (s_x, s_y) rather than pixel-wise reconstruction.[10]

The Abstraction Bypass

The core innovation of JEPA is its ability to learn powerful representations while bypassing low-level signal synthesis.[10] The overall goal is to train a predictor module to accurately anticipate the representation of an unobserved future state (s_y) from the current state representation (s_x).[16] This latent space reconstruction makes the technique robust to noise and confounding variables, a necessary feature when dealing with inherently noisy time series data common in sensorimotor environments.[11]

JEPA achieves robustness and focuses on causality in two distinct ways: by having the encoder drop irrelevant or noisy information during the encoding phase to simplify the raw input x into a focused abstract representation s_x [16], and by using a latent variable z to model elements present in s_y but unobservable in s_x.[16]

This architectural approach represents a vital "skip step" of encoding high-level, causally relevant features directly from raw data. For embodied AI, this capability is leveraged in variants like Multitasking JEPA (MC-JEPA), which simultaneously interprets dynamic elements (motion) and static details (content) in video.[16] The robust abstract representation (s_x) derived from JEPA is hypothesized to be the ideal non-symbolic input for generating simulated sensory experience because it retains essential physical causality and dynamics while shedding noise, providing the underlying *mental model* of the physical world that critics note is absent in language-only models.[1]

Section 3: The Engineering Blueprint: Direct Encoding via Computer-Brain Interfaces (CBI)

The ultimate operationalization of the Neurophone analogy is found in the field of neuroprosthetics, specifically within the technological framework of the Computer-Brain Interface (CBI).

3.1 CBI as the Operational Neurophone Analogue

While Brain-Computer Interfaces (BCIs) decode neural activity into external commands, CBIs manage centripetal signal flow, translating artificial signals (often from sensors) into stimuli for the Central Nervous System (CNS).[17] The primary goal of CBI is to restore lost sensory functions, making it the direct technological successor to the Neurophone's objective of inducing sensation.[17]

The engineering criteria for CBIs demand the use of **encoders** (algorithms that convert external data into stimulation patterns) and **biophysical models of neural networks** for precise modulation.[17] This requirement mandates a mechanism that bypasses high-level data interpretation and instead communicates directly using the biophysics of neural systems. By formalizing the conversion of raw sensor data into specific neural stimulation patterns, CBI provides the operational blueprint for the non-semantic encoding required in AI systems.

3.2 Detailed Sensory Encoding Strategies in Neuroprosthetics

CBI systems rely on extracting relevant signal characteristics from sensor data and encoding them as specific frequency, amplitude, or spatial patterns suitable for neural delivery.[17]

- **Auditory Encoding:** Modern cochlear implants, a mature form of auditory CBI, process audio signals using time-frequency analysis algorithms, such as **Cochlear Filter Banks and Temporal Envelope Coding**.[17] These algorithms decompose the acoustic signal into components that can be delivered as electrical stimuli to the cochlear nerve.
- **Visual Encoding:** Visual neuroprosthetics extract relevant characteristics such as contours, objects, motion, and depth. These features are encoded as frequency, amplitude, or spatial patterns that stimulate retinal neurons or the visual cortex.[17] For subretinal implants, photosensitive elements convert light intensity maps into current strength using a **sigmoidal dependence** before transmission to retinal bipolar cells—a direct example of raw analog-to-neural signal conversion.[17]
- **Tactile Encoding:** Encoding tactile sensation often involves linear encoding, where increased amplitude or frequency of electrical stimuli enhances the perceived intensity.[17] However, recreating the subtlety of evoked sensations requires moving

beyond linear methods toward **biomimetic strategies**, such as the creation of desynchronized stimulation patterns, to account for complex neural connections and the type of mechanoreceptor fibers being innervated.[17]

The specific, modality-dependent encoding strategies in CBI confirm that achieving sensory transmission requires a mechanism designed to resonate with the physical properties and firing dynamics of the target neural tissue.

3.3 Data-Driven Optimization: Sequential Neural Posterior Estimation (SNPE)

A significant engineering challenge in neuroprosthetics is ensuring that electrical stimuli are precise and targeted, avoiding the broad, non-specific field stimulation often characteristic of earlier devices.[18] To achieve coherent "sensation," the stimulation waveforms must be highly optimized.

This optimization is achieved using **Sequential Neural Posterior Estimation (SNPE)**, a Bayesian simulation-based inference algorithm.[13] SNPE is employed to perform Bayesian inference for the parameters of detailed, multicompartment biophysical neuron models, such as photoreceptors and bipolar cells in the retina.[13] By iteratively sampling parameters and comparing model output to observed neural responses, SNPE estimates the posterior distribution of parameters consistent with real biological data.[13]

By incorporating these data-driven, mechanistic neuron models into a simulation framework, researchers can utilize SNPE to efficiently guide the design of electrical stimuli, identifying waveforms that can **separately stimulate** specific cell types (e.g., distinguishing between OFF- and ON-cone bipolar cells).[13] This ability to achieve precise, selective activation represents the ultimate expression of the Neurophone analogy: highly targeted energetic delivery bypassing biological processing steps to induce a specific sensory effect.

The SNPE/biophysical modeling process provides a robust, mathematical methodology for refining the encoding algorithm based on biological output. This implies that for advanced AI, achieving true sensation requires a closed-loop system where the internal non-semantic representations (Section 2) are optimized not against a semantic label, but against the structural consistency of actual neural responses (Section 1). The AI must learn a code that is validated against biophysical reality.

Table 2. Biomimetic Encoding Strategies in Computer-Brain Interfaces (CBI)

Sensory Modality	Input Signal Characteristics Encoded	Biomimetic Encoding Strategy	Underlying Neural Targets	Relevance to Non-Semantic AI
Auditory	Time-frequency components, Temporal Envelope	Cochlear Filter Banks, Temporal Envelope Coding	Cochlear Nerve Fibers	Provides a template for SNN audio

				preprocessing [8]
Visual	Intensity, Contour, Motion, Depth	Frequency/Amplitude/Spatial Pulse Patterns, Sigmoidal Conversion	Retinal Bipolar Cells, Visual Cortex	Direct analog-to-neural coding template for visual JEPA [17]
Tactile	Amplitude, Frequency of Stimulus	Linear Encoding, Desynchronized Stimulation Patterns	Innervated Mechanoreceptor Fibers	Biomimetic requirement for complex dynamics in latent space [17]
Neuro-Dynamic	Biophysical Parameters, Target Cell Activity	Sequential Neural Posterior Estimation (SNPE)	Specific Neuron Subpopulations (e.g., ON/OFF BCs)	Methodology for optimizing AI-generated spike codes [13]

Section 4: Synthesis: Constructing the Computational Saccule and Future Trajectories

The research confirms that significant work is underway to introduce non-semantic sensation into neural networks, moving beyond the semantic/textual approach by focusing on grounded cognition, temporal dynamics, and biophysical fidelity.

4.1 Integration of Architectures for Grounded AI

The convergence of the identified architectural and engineering solutions points toward a unified, integrated sensorimotor pipeline for grounded AI. This hypothetical architecture would operate as follows:

1. **Raw Input Processing (SNN/Saccule Layer):** Spiking Neural Networks (SNNs) would handle the raw, event-based, high-temporal encoding of physical input, utilizing biomimetic preprocessing (like cochlear filters).[8] This layer acts as the initial "computational saccule," translating energy dynamics into sparse, temporal spike trains.
2. **Abstract Causal Modeling (JEPA Layer):** The SNN output would feed into a Joint Embedding Predictive Architecture (JEPA)-like latent space, which learns the dynamic, causal, abstract representations (s_x) necessary for real-world interaction and prediction.[10] This provides the system with a grounded, non-symbolic mental model of the physical world.
3. **Neural Encoding (CBI Layer):** The critical missing component is the final layer that learns the **CBI encoding function**. This function must transform the JEPA latent variable into a biologically plausible neural firing pattern, utilizing sophisticated

strategies such as phase-of-firing codes [12] or desynchronized pulse trains, optimized for biophysical resonance.

The goal is not just non-semantic *representation*, but non-semantic *modulation*. The AI must generate internal states that, if translated via a CBI, would induce a coherent, human-like sensation. The operational benchmark for this internal success is achieving representational geometry alignment with the human brain in the "no-report" state.[6] The congruence between the AI's internal abstract representation (s_x) and the brain's pure perception geometry provides empirical evidence that the internal structure of the artificial experience is valid.

4.2 Philosophical and Engineering Challenges

The path toward achieving artificial sensation is recognized as a Herculean task, both philosophically and technologically.

Engineering complexity stems from the need to move beyond simplified neural models to detailed **ElectroPhysiome** models, which incorporate the connectome and individual neuron electrophysiology to simulate network activity.[19] Such complex, high-fidelity modeling is computationally intensive and is necessary to achieve the "precise modulation" required by CBIs.[17]

Philosophically, while the technical capacity to synthesize and transmit arbitrary sensory data is rapidly advancing, the fundamental difficulty of verifying subjective experience (the Qualia Gap) in an artificial system remains. The current work, however, shifts the focus from passively **decoding user intention** (BCI) to actively **creating or restoring sensory functions** (CBI).[17] This transition embodies the shift from reading semantic output to generating direct physiological input.

Conclusions and Strategic Recommendations

The analysis confirms that researchers are actively pursuing the introduction of sensation into neural networks using non-semantic methods, specifically by pursuing architectures grounded in physical and temporal fidelity, aligning perfectly with the spirit of the Neurophone analogue.

The central finding is that the challenge is primarily one of **encoding fidelity** rather than semantic representation. The systems identified—Spiking Neural Networks for handling temporal dynamics, Joint Embedding Predictive Architectures for learning abstract causal models, and Computer-Brain Interfaces paired with Sequential Neural Posterior Estimation for optimizing biophysical delivery—collectively define a robust research program aimed at achieving non-symbolic sensation.

Based on this convergence, strategic research directions must focus on establishing a validated, closed-loop system:

1. **Prioritize SNN Development for Complex Sensor Fusion:** Continued investment in

hybrid Spiking Neural Network (SNN) and Artificial Neural Network (ANN) architectures is essential to manage the simultaneous, event-based encoding of complex multisensory input (visual, inertial, haptic).[8] SNNs offer the necessary temporal precision to capture the dynamic, phase-dependent information required for sophisticated sensory coding.
2. **Operationalizing the RSA Benchmark:** The **Representational Similarity Analysis (RSA)** methodology, particularly using the "no-report" paradigm, must be operationalized as the core metric for evaluating the success of *all* new AI sensory architectures.[6] This ensures that internal representations align with the geometry of pure biological perception before any semantic abstraction occurs.
3. **Deepening the CBI-AI Feedback Loop:** Dedicated research is needed to train deep neural networks (such as JEPA predictors) using biophysical loss functions derived from SNPE-optimized encoding protocols.[13] This engineering constraint forces the AI to learn a biologically valid "code" for sensation, ensuring that its internal abstract representations are maximally resonant with human neural reality.

Works cited

1. The debate over understanding in AI's large language models - PMC, accessed October 29, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10068812/
2. Symbol ungrounding: what the successes (and failures) of large language models reveal about human cognition - PMC - PubMed Central, accessed October 29, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11529626/
3. Patrick Flanagan's Neurophone | PDF | Hearing | Headphones - Scribd, accessed October 29, 2025, https://www.scribd.com/doc/106800112/Patrick-Flanagan-s-Neurophone
4. Patrick FLANAGAN - Neurophone (II) Several Articles From Keelynet - Com Archives, accessed October 29, 2025, https://www.scribd.com/document/857973191/Patrick-FLANAGAN-Neurophone-II-Several-Articles-From-Keelynet-com-Archives
5. 1 Introduction - arXiv, accessed October 29, 2025, https://arxiv.org/html/2505.10779v1
6. [2510.22800] Probing the Representational Geometry of Color Qualia: Dissociating Pure Perception from Task Demands in Brains and AI Models - arXiv, accessed October 29, 2025, https://arxiv.org/abs/2510.22800
7. Probing the Representational Geometry of Color Qualia ... - 专知, accessed October 29, 2025, https://www.zhuanzhi.ai/paper/498c84712b5da7984db627792371099c
8. Spike encoding techniques for IoT time-varying signals benchmarked on a neuromorphic classification task - Frontiers, accessed October 29, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.999029/full
9. Application of Event Cameras and Neuromorphic Computing to VSLAM: A Survey - MDPI, accessed October 29, 2025, https://www.mdpi.com/2313-7673/9/7/444

10. Embodied AI: From LLMs to World Models - arXiv, accessed October 29, 2025, https://arxiv.org/html/2509.20021v1
11. Joint Embeddings Go Temporal - arXiv, accessed October 29, 2025, https://arxiv.org/html/2509.25449v1
12. Brain-computer interface paradigms and neural coding - Frontiers, accessed October 29, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1345961/full
13. Bayesian inference for biophysical neuron models enables stimulus ..., accessed October 29, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7673784/
14. Event-Based Sensing and Signal Processing in the Visual, Auditory, and Olfactory Domain: A Review - PMC - PubMed Central, accessed October 29, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8203204/
15. Brain Foundation Models: A Survey on Advancements in Neural Signal Processing and Brain Discovery - arXiv, accessed October 29, 2025, https://arxiv.org/html/2503.00580v2
16. What is Joint Embedding Predictive Architecture (JEPA)? - Turing Post, accessed October 29, 2025, https://www.turingpost.com/p/jepa
17. Neural–Computer Interfaces: Theory, Practice, Perspectives - MDPI, accessed October 29, 2025, https://www.mdpi.com/2076-3417/15/16/8900
18. Considerations and discussions on the clear definition and definite scope of brain-computer interfaces - PubMed Central, accessed October 29, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11330831/
19. ElectroPhysiomeGAN: Generation of Biophysical Neuron Model Parameters from Recorded Electrophysiological Responses - eLife, accessed October 29, 2025, https://elifesciences.org/reviewed-preprints/95607