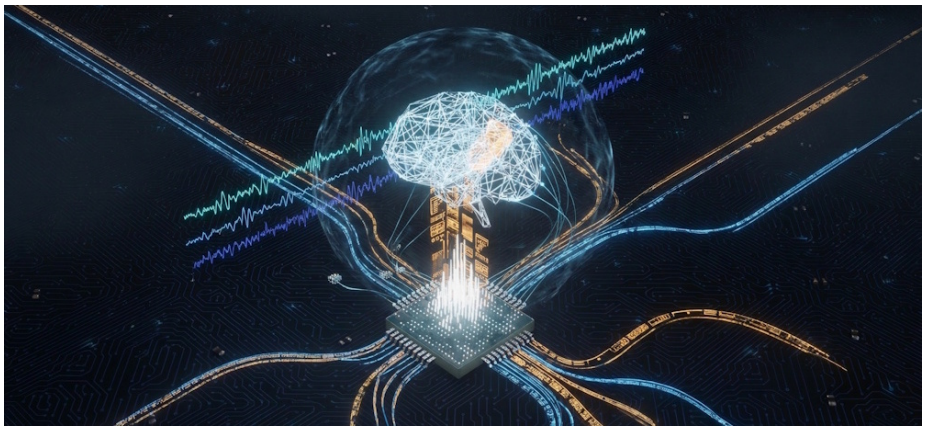
THOR-NeuroBench Challenge 2026

Challenge Website



1 Benchmarking Challenge

The National Science Foundation THOR: The Neuromorphic Commons, in partnership with the NeuroBench project, proposes a two-phase benchmarking challenge aimed at advancing scalable, efficient, and societally relevant neuromorphic algorithms.

As artificial intelligence pushes the boundaries of human-machine interaction, deploying continuous, real-time physiological decoding on edge devices remains severely constrained by power and latency. The NSF THOR-NeuroBench Challenge invites the global research community to bridge this gap by harnessing the benefits of neuromorphic systems and architectures. Designed to emulate the biological brain's event-driven dynamics, neuromorphic computing leverages sparsity and spiking neural networks to execute complex pattern recognition locally on highly constrained edge devices, untethering them from the cloud. To rigorously and objectively measure these advancements, teams will utilize the NeuroBench collaborative framework. Through NeuroBench's standardized metrics, participants will evaluate their models not merely on accuracy, but on memory footprint and compute efficiency. This will ensure that the winning

solutions represent true, deployable leaps in neuromorphic engineering.

To drive this innovation, participants will benchmark their architectures against the OpenBMI Dataset, specifically targeting the high-density 62-channel EEG Motor Imagery (MI) paradigm. In this task, neural activity is captured as subjects internally rehearse the movement of their hands without physical execution. EEG signals are notoriously complex, noisy, and non-stationary, making their spatio-temporal features an ideal match for neuromorphic hardware that excels at processing asynchronous time-series data. Unlike early datasets limited by small cohorts, OpenBMI spans 54 subjects across multiple sessions. This scale is critical, demanding algorithms capable of robust cross-subject generalization and session-to-session reliability. Crucially, the data natively captures **BCI illiteracy**, where certain users struggle to produce machine-readable EEG patterns, challenging engineers to design highly adaptive algorithms resilient to diverse neural profiles.

This challenge extends beyond algorithmic optimization into enhancing societal impact. Motor Imagery-based Brain-Computer Interfaces (BCIs) hold life-changing potential for individuals suffering from severe motor impairments, including stroke, ALS, and spinal cord injuries. Today, patients with locked-in syndrome are largely dependent on caregivers, and the deep learning models capable of decoding their neural intent require bulky, power-intensive hardware confined to clinical laboratories. By successfully mapping complex EEG decoding onto ultra-low-power neuromorphic systems, this challenge aims to liberate BCI technology. Our ultimate objective is to pave the way for real-time, wearable decoders that can be seamlessly integrated into motorized wheelchairs or robotic prosthetics, bringing us closer to a future where paralyzed individuals can autonomously interact with their world through thought alone.

1. Establish reproducible and standardized benchmarking practices using NeuroBench metrics.
2. Promote scalable spiking neural network (SNN) algorithm development under resource constraints.
3. Encourage solutions that address applications beneficial to society.
4. Bridge algorithmic innovation (Phase 1) with hardware realization (Phase 2 on hardware neuromorphic platforms hosted by NSF THOR).

Phase 1 focuses on algorithmic innovation under scale constraints, ensuring fairness, reproducibility, and meaningful comparison. Phase 2 will extend top-performing solutions to neuromorphic hardware platforms.

2 Timeline

The challenge follows a structured, milestone-based timeline designed to ensure transparency, fairness, and adequate preparation time for participants.

Post Announcement and Registration, the challenge is officially launched and teams may begin forming and registering. The Phase 1 Dataset Release marks the start of the technical preparation period, providing participants with the training/validation data, evaluation rules, scale constraints, and integration details for the NeuroBench benchmarking framework. When Phase 1 Submissions Open, teams may begin uploading containerized models for evaluation, with leaderboard feedback (if applicable) based on the public validation set. The Phase 1 Submissions Close date represents the firm deadline after which no further updates are accepted, ensuring a fair evaluation window. Finally, on the Phase 1 Results Announcement, final rankings (computed on the hidden held-out test set or newly generated samples) are released, and top-performing teams are formally invited to participate in Phase 2 hardware deployment.

Date	**Challenge Schedule**
23 Feb 2026	Challenge Announcement
24 March 2026	Phase 1 Dataset Release & Challenge Details
25 March 2026	Registration Opens
10 June 2026	Phase 1 Submissions Open
05 August 2026	Phase 1 Submission Deadline
16 Sept 2026	Phase 1 Results Announced
30 Sept 2026	Phase 2 Invitations

Table 1: Challenge Schedule

3 Participation

1. The challenge is open to US-based individuals, colleges/universities, scientific research institutions, labs, groups, and enterprises. The maximum number of team members is 3.

2. To ensure compliance with local regulations during the challenge, all participants should comply with the export control laws of their country. In case of any negative impact on the challenge due to violation of any export control laws, the team members reserve the right to disqualify the relevant contestants and take legal action.

3. Participants shall express their intent to participate in this challenge by filling in the registration form.

4 Dataset, Challenge and Code Harness

The core of the NSF THOR-NeuroBench Challenge evaluates algorithms on their ability to accurately and efficiently decode complex human intent. For this

purpose, the challenge utilizes the widely cited OpenBMI Dataset Publication, specifically isolating the Electroencephalography (EEG) Motor Imagery (MI) paradigm.

4.1 Dataset Structure and Specifications

The OpenBMI dataset provides an unprecedented scale of high-quality EEG recordings, distinguishing it from smaller, legacy datasets that often suffer from overfitting.

- **Cohort Size**: Data was collected from 54 healthy participants across two distinct sessions, separated by multiple days. This scale is vital for testing session-to-session algorithmic reliability.
- **Recording Hardware**: Signals were acquired using a 62-channel Ag/AgCl electrode setup following the international 10-05 system.
- **Temporal Resolution**: The raw continuous data was recorded at a high sampling rate of 1000 Hz.
- **Task Paradigm**: During the MI task, subjects were prompted with visual cues to imagine the kinetic movement of either their left hand or right hand (a two-class classification problem). Each trial consists of a 4-second active imagination phase, flanked by resting baselines. This yields 100 trials per class, per session, for each individual subject.

Download Raw Database : OpenBMI Dataset
Please download only the files titled with ***EEG_MI*** *labels*
e.g.- sess02_subj54_EEG_MI.mat

4.2 Encoding Mechanism

Standard Deep Neural Networks (DNNs) ingest continuous, floating-point analog signals. However, neuromorphic architectures operate on event-driven, binary logic. Therefore, the continuous EEG amplitudes must be converted into discrete, asynchronous spike trains before processing.

For this challenge, the provided data must be formatted as sparse, multi-dimensional binary tensors (representing 1 for a spike, 0 for no spike). Participants are encouraged to explore varying spike encoding schemes to optimize their models for Track 1, but have to adhere to a standardized (provided) spike format for Track 2. The baseline benchmarking harness will support two standard neuromorphic transformations:

Threshold-Based Delta Modulation: A spike is generated whenever the difference between consecutive voltage samples in a specific channel exceeds a predefined threshold. This method naturally exploits the temporal sparsity of the EEG signal, only registering dynamic changes in neural activity.

Direct Input Coding (Learnable Encoding): Participants may utilize the first layer of their Spiking Neural Network (SNN) as an encoding layer comprised

of Leaky Integrate-and-Fire (LIF) neurons. These neurons ingest the raw or downsampled EEG and output spikes based on trained synaptic weights and membrane thresholds.

The final input format fed into the NeuroBench algorithm track will take the shape of a discrete matrix: [Channels (62) $\times$ Time Steps (T)], where T is determined by the participant's chosen temporal binning resolution.

4.3 Train and Evaluation Split

To rigorously evaluate the real-world viability of the submitted neuromorphic models, the challenge implements a strict, multi-phase cross-validation strategy designed to prevent data leakage and test true generalization.

4.4 Evaluation Metrics

Scoring in the THOR-NeuroBench Challenge utilizes the standardized NeuroBench evaluation harness. Because this is a neuromorphic grand challenge, raw accuracy is the gating requirement, but hardware efficiency determines the ultimate victor.

- **Classification Accuracy (Track 1 Metric)**: The baseline capability of the model to correctly classify left-hand versus right-hand motor imagery. A baseline benchmark of average accuracy across the test splits will be provided to all participants after registration.
- **Compute Efficiency (Track 2 Constraint)**: Instead of counting dense Multiply-Accumulate (MAC) operations used in standard GPUs, the harness calculates the number of effective Synaptic Operations (SynOps) triggered by the sparse spike trains. Fewer spikes and sparser activations mean lower energy consumption on neuromorphic hardware. In addition, submissions in Track 2 will be evaluated based on their memory usage (Footprint).

4.5 Code Harness

The challenge will be implemented using the open-source NeuroBench benchmarking harness: `neurobench.readthedocs.io`. The harness provides dataset download scripts and automatic metric evaluation for torch-based neural network models.

Depending on the code implementation for the submission, the automatic evaluation in the NeuroBench harness may not be compatible. Each submission is responsible for ensuring that their method is properly measured, and the challenge review will also check all submissions.

A baseline example for this challenge, including training and evaluation with the NeuroBench harness, will be released soon. Prior examples using the NeuroBench harness, including for previous benchmark challenges, can be found in the NeuroBench github repo.

5 Challenge Tracks

To fully explore the potential of neuromorphic decoding while maintaining rigorous, real-world applicability, the NSF THOR-NeuroBench Challenge is divided into two distinct parallel tracks. This dual-track structure acknowledges a critical reality in current benchmarking: while the NeuroBench framework provides unparalleled precision in measuring the efficiency and cost of Spiking Neural Networks (SNNs), it does not currently measure the computational overhead of the data pre-processing stages prior to the network.

To prevent the offloading of heavy computation to an unmeasured pre-processing pipeline and to ensure a level playing field, we have designed the following tracks:

- **Track 1: Absolute Accuracy** This track is the proving ground for the absolute limits of neural decoding. The primary objective is to achieve the highest possible classification accuracy on the unseen test subjects. To foster maximum innovation, participants in Track 1 are completely unrestricted in how they condition the raw data. You are encouraged to explore novel spike encoding schemes, advanced spatial filters, or custom signal processing pipelines before feeding the data into your SNN. If you can extract a stronger signal through innovative data formatting, this is the track to showcase it.

- **Track 2: Compute Efficiency** This track represents the ultimate goal of the challenge: deploying highly efficient models onto constrained edge devices. In Track 2, participants are evaluated on a combined metric of accuracy and compute cost (energy measured as synaptic operations per sample). To ensure fair efficiency comparisons, all participants in this track must strictly utilize the standardized pre-processing pipeline provided in the challenge repository, which is formatted the raw 62-channel EEG data into continuous, dense floating-point tensors.

6 Awards

To recognize the engineering breakthroughs achieved during this event, we are proud to partner with the Edge AI Foundation, which has generously sponsored a $1,000 prize pool for the inaugural NSF THOR-NeuroBench Challenge.

Because the challenge specifically targets two distinct pillars of neuromorphic advancement, pushing the absolute limits of neural decoding and optimizing for deployable edge hardware, the prize pool is divided equally between the two tracks. The top-performing team in each category will be rewarded independently:

Track 1 Winning Team: $500

Track 2 Winning Team: $500

Beyond the monetary awards, the winning teams will have their methodologies highlighted on the NSF THOR Neuromorphic Commons website and will be granted priority access to THOR research infrastructure for their future research endeavors.

7 Organizers

THOR : The Neuromorphic Commons
THOR is a pioneering, NSF-funded national research hub designed to democratize access to large-scale neuromorphic computing. Led by and hosted at the University of Texas at San Antonio, THOR provides the broader scientific community with unprecedented, open access to state-of-the-art, brain-inspired hardware platforms, such as the massive SpiNNaker2 supercomputing architecture, that were previously restricted to select industry labs. By offering a collaborative ecosystem of cutting-edge hardware, open-source software tools, and standardized benchmarks, THOR aims to accelerate the transition of artificial intelligence from power-hungry data centers toward highly efficient, event-driven biological blueprints capable of transforming real-world applications. THOR is developed in partnership with University of Tennessee, Knoxville, University of California, San Diego, and Harvard University.

NeuroBench
NeuroBench is the premier collaborative framework for benchmarking neuromorphic computing algorithms and systems. Developed by a vast consortium of industry and academic leaders, NeuroBench provides standardized metrics, datasets, and methodologies to evaluate brain-inspired systems fairly and objectively. By participating in this challenge, teams will utilize the NeuroBench framework to measure not just the accuracy of their models, but their footprint, latency, and energy efficiency, ensuring that the winning solutions represent true advancements in deployable neuromorphic engineering.