Neural-Behavioral Representation of Natural Whole-body Movement in Monkeys

Jieshi He[1,2*] **Puzhe Li**[2*] **Yanan Sui**[3†] **Mu-ming Poo**[2†]

[1]Fudan University
[2]Center for Excellence in Brain Science and Intelligence Technology, CAS
[3]Tsinghua University

Abstract

Understanding how cortical activity represents natural whole-body behaviors in primates remains challenging. Limited by the diversity of movements and inaccessibility of large-scale neural representation of whole-body kinematics, previous motor decoding studies focused on constrained tasks and limited limb movements. Here, we present a neural-behavioral recording and modeling framework for freely moving monkeys, combining large-scale epidural cortical signals from distributed sensory- and motor-related areas with synchronized multi-view motion capture through a custom-made data collection platform. We reconstructed whole-body monkey kinematics and learned a compact behavior prior using an autoregressive encoder-decoder model. Conditioned on neural signals, the model decoded accurate and realistic whole-body movement without explicit physical constraints. Our results provide a novel proof-of-concept approach for decoding natural whole-body movements in primates using large-scale intracranial neural activity.

1 Introduction

The cerebral cortex supports sensory processing, motor planning, movement execution, and adaptive behavioral control. A central goal in computational and systems neuroscience is to understand how cortical population activity represents movements and to use the representations to reconstruct behaviors from neural activities [1, 2, 3, 4]. Substantial progress has been made in decoding intended or executed movements from cortical activity, particularly for upper-limb reaching, hand grasping, and lower-limb locomotion under constrained experimental paradigms [5, 6]. However, most existing studies decode restrained postures or limited body segments, relying on predefined tasks and repeated trials. The neural representations underlying natural whole-body movements remain largely unexplored. A few studies have used neural features extracted from specific brain regions to classify constrained whole-body movement categories in monkeys, such as walking, jumping, or resting [7, 8]. Yet no prior report has shown direct reconstruction of continuous whole-body movements of primates from cortical neural representations.

Natural whole-body movement poses challenges that are qualitatively different from those in conventional motor decoding experiments [9]. First, large-scale body movement introduces substantial artifacts into neural recordings [10], making it difficult to simultaneously acquire reliable brain signals and natural movement data. Second, compared with constrained settings, freely moving monkeys exhibit much richer behavioral patterns [11]. Reconstructing whole-body movement requires predicting kinematics of all body parts, leading to a substantially higher-dimensional output

*Equal contribution
†Corresponding to: Yanan Sui <ysui@tsinghua.edu.cn> and Mu-ming Poo <mpoo@ion.ac.cn>

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

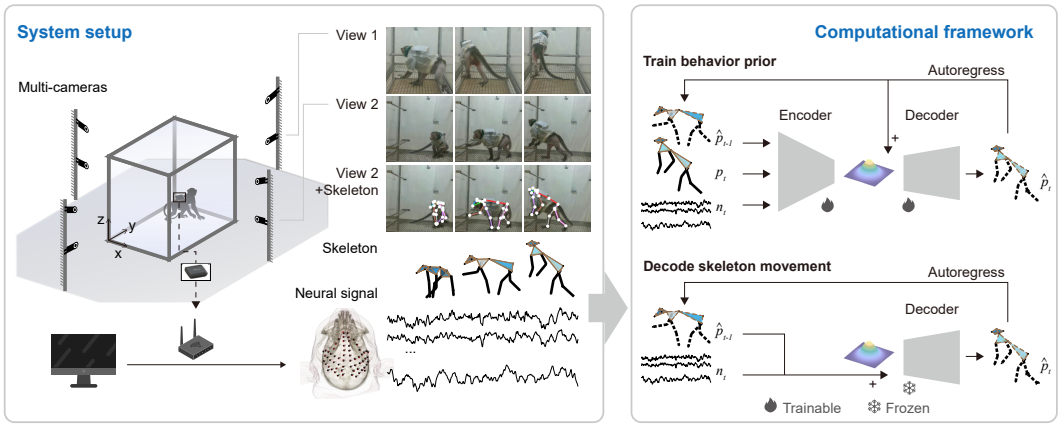

Figure 1: Pipeline for decoding natural whole-body movement from epidural neural signals. **System setup**: Whole-body movement data were captured using an 8-camera system and reconstructed into 3D skeleton. Synchronous recording of epidural neural signals from 62 channels were recorded and transmitted wirelessly. **Computational framework**: Collected data were divided into training and test sets. First, we trained a behavior prior on the training set by compressing movement data and neural signals into a latent space. Latent variables were then sampled and combined with the predicted state $\hat{p}_{t-1}$ and the neural signals n_t, and further passed through the decoder to reconstruct the current state $\hat{p}_t$. During testing, only the decoder was used, and the movement states were reconstructed in the same autoregressive manner as during the training.

space. Third, natural movements are spontaneous, irregular, and weakly repeated. Unlike instructed laboratory tasks, they do not provide aligned trials that can be averaged to improve signal-to-noise ratio. In constrained experimental paradigms, movements are usually triggered by task instructions, and are repeated in regular patterns. Moreover, similar movements may correspond to different neural activities under different contexts, while free movement also involves abundant spontaneous neural activity unrelated to the movement [11, 12]. These factors make it difficult to learn stable neural-behavioral mappings using standard discriminative decoding approaches.

In this work, we have addressed these challenges by developing a neural-behavioral recording and modeling framework for freely moving monkeys as shown in Fig. 1. We implanted epidural electrodes over distributed motor-related cortical areas, enabling recording of cortical population activity. Compared with intracortical microelectrode arrays, these epidural recordings sacrificed single-neuron resolution for improved long-term stability and reduced tissue damage [13, 14]. Compared with conventional ECoG arrays, our skull-fixed design enables broader cortical coverage while preserving the natural buffering structure of the skull, dura, and brain, which is important for reducing movement-induced artifacts during free-moving behaviors [10].

To capture movement behavior, we build a synchronized multi-view motion capture system with eight cameras and reconstruct three-dimensional monkey kinematics during spontaneous movement. We defined a monkey kinematic model and converted tracked 3D joint trajectories into a compact whole-body movement representation. On top of this recording system, we designed the Neural-Behavioral Model, which learns a compact behavior prior from whole-body kinematic sequences while simultaneously learning how epidural neural activity represents the natural movement. During inference, the model reconstructs whole-body movement from the previous predicted state and the current neural signals. This formulation is designed for natural behavior, where movements are constrained both by biomechanical regularities and by time-varying prior cortical activity. Unlike direct regression models that map neural features to joint coordinates independently, our approach explicitly integrates neural representations with a learned behavior prior, enabling more realistic and temporally coherent movement prediction without imposing explicit physical constraints.

Experiments on synchronized neural and behavioral recordings from freely moving monkeys demonstrate the feasibility of reconstructing continuous whole-body movement from distributed epidural cortical activity. The Neural-Behavioral Model reconstructs realistic whole-body movement over multi-second time scales and outperforms both a behavior-only generative model and a LSTM-based model [15]. These results show that neural signals provide behavior-specific information beyond

generic behavior priors, which are essential for maintaining physically plausible whole-body posture and coordination in the high-dimensional movement space.

Contributions.
1) We have developed a neural-behavioral recording platform that synchronously captures large-scale epidural cortical activity and multi-view 3D whole-body kinematics in freely moving monkeys.
2) We have formulated natural whole-body movement reconstruction as a problem of neural-behavioral-conditioned autoregressive generative modeling, providing an ML framework that integrates distributed cortical representations with learned behavior priors.
3) We have provided the first demonstration, to our knowledge, that continuous and realistic whole-body movement of freely moving primates can be predicted from cortical activity under naturalistic conditions.

2 Related work

The primary goal of our work is to demonstrate the feasibility of reconstructing whole-body movements in freely moving primates using large-scale epidural recordings. Previous studies were usually done on movement-constrained settings, where limited or stereotyped-pattern of movements were performed by subjects confined to chairs (see Fig. A5).

2.1 Predicting movement from neural signals

Previous studies have shown that neural signals from the human primary motor cortex can decode movement intention or limited actual motor behaviors, such as arm reaching and hand grasping [16, 6, 17, 18, 19, 20, 21, 22]. Similarly, neural recordings from the motor cortex of non-human primates have enabled prediction of upper-limb movements [23, 24, 25, 26, 27]. In such studies, subjects were required to perform predefined tasks repeatedly in a structured manner, which reduced noise and facilitated the discovery of stable neural-behavioral mappings. Most previous studies have adopted such approaches for supervised discriminative models [15, 28, 29]. Some works [30, 31, 32] have also explored unsupervised pretraining on large-scale neural data for downstream tasks, but these efforts were largely limited to simple classification or single-limb regression tasks.

2.2 Neural-behavioral representation

Some studies on neuronal spike data have explored cross-modal and cross-subject neural representations. The NDT2 [31] introduces Transformer-based pretraining for neural spike modeling by integrating data across subjects, tasks, and recording sessions to learn general-purpose neural representations. The results showed that large-scale multi-context pretraining substantially improved downstream decoding performance and adaptation to unseen scenarios. A framework [33] established reliable cross-modal representational alignment, enriching neural coding insights and enabling zero-shot behavior decoding. A study has demonstrated that applying orthogonal regularization to the latent movement space can significantly improve decoding accuracy across sessions, subjects, and paradigm generalization tasks [34].

2.3 Neural and behavioral dataset and setup

The UniMo4D [35] dataset provides large-scale cross-species 3D movement data but does not include neural activity. Previously reported neural-behavior recording systems include those that only capture whole-body movement without neural recording such as BigMaQ [36], or setups that recorded distributed neural signals together with motion capture of body parts and those recorded local neural signals with whole-body movement (see Table A1 in the Appendix). OpenMonkeyStudio [37] reconstructed 3D kinematics of 13 joints in moving monkeys while recording tens to hundreds of neurons per session. A study [1] performed microelectrode array (MEA) recordings from three brain regions during locomotion tasks, synchronized with EMG and limited limb kinematics. NeuroTycho [37] provided epidural ECoG signals with upper-limb kinematics under restrained conditions. In contrast, our work uniquely combined population-level neural recordings over multiple sensory and motor cortical areas with synchronously monitored whole-body 3D kinematics in freely moving monkeys.

3 System setup

To reconstruct natural whole-body movement from distributed cortical activity, we first need to collect freely moving behavior data together with synchronized large-scale neural signals. In this section, we describe our experimental setup, including epidural recording of neural signal, video acquisition of monkey behaviors, reconstruction of movement trajectories from video recordings, and the kinematics-representing model of whole-body movement.

3.1 Data collection

Animal care and surgery All experimental procedures were approved by the Ethics Committee of the relevant institutional review board (Protocol No.: NZXSP-2022-11-1). The animal was well cared in a certified experimental facility for macaque monkeys. Following a pre-surgical health assessment, the animal underwent implantation surgery of epidural electrodes. Sixty-four implantation sites were designated on the skull across bilateral sensory- and motor-related cortical areas based on pre-surgical CT and MRI scans, including the primary motor cortex (M1), primary somatosensory cortex (S1), premotor cortex (PMC) and superior and inferior parietal lobes (SIPL), two of which served as reference and ground electrodes. The electrodes, which are located at the tip of insulated metal screws with a length slightly exceeding the local skull thickness, contacted the dura surface to record intracranial population neuronal activity. These recorded signals exhibited properties comparable to those obtained by ECoG, while providing additional advantages over ECoG, including reduced invasiveness, greater flexibility in selecting recording regions, and broader cortical coverage. Two adult female *Macaca fascicularis* was used in this study. Monkey A: , 14 years of age, Monkey B: 13 years old.

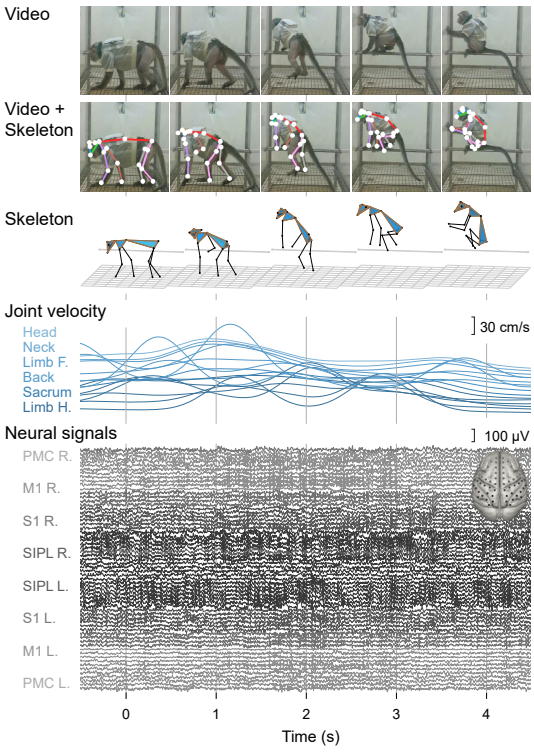

Figure 2: Example captured movements and synchronized neural signals lasting 5 seconds. From top to bottom: raw video frames; video frames overlaid with tracked reconstructed 3D skeletons; 3D visualization of the skeleton; velocities of joints (blue color scale from light to dark indicates body positions from anterior to posterior; Limb F., forelimb; Limb H., hindlimb); and preprocessed neural signals from 62 channels. Gray color scales denote different cortical regions (R., right hemisphere; L., left hemisphere). The cortical locations of the corresponding channels are indicated at the corner.

Acquisition of neural signals and behavioral data The recording session began about three weeks after surgical implantation of electrodes. The animal was placed in a transparent observation cage (70 cm × 100 cm × 100 cm) in a dedicated behavioral recording room (3.5 m × 2.5 m) for multi-view behavioral video recording. The cage was equipped with horizontal bars at both upper and lower levels to support sitting, climbing, and swinging movements, and 8 cameras (Intel RealSense D435i) mounted on four walls of the room and provided comprehensive multi-view coverage of monkey movements. A wearable brain signal recording device (Quanlan RSC64) was placed in the backpack of the monkey jacket and connected to the lead wires. The animal was placed in the cage and allowed to move freely throughout sessions. Each behavioral experiment day included three sessions, each session lasting 30 minutes.

All 8 cameras were triggered simultaneously via a custom-designed program. After data acquisition, we achieved precise synchronization between videos and neural signals using timestamps and eyeblink events. Data were acquired continuously at stable sampling rates (videos: 30 Hz, neural signals:

2000 Hz). Rich behavioral repertoires of the monkeys were recorded in the experimental cage, including walking, climbing, swinging, scratching, and other spontaneous movements.

The Fig. 2 illustrates monkey A walking and stepping onto a bar. The figure shows raw video, reconstructed skeletons, joint velocities, and preprocessed neural signals. During quadrupedal locomotion, the left and right limbs exhibited alternating movement patterns, as reflected in velocity fluctuations. Synchronously, contralateral cortical hemispheres showed alternating high- and low-frequency oscillations. Consistent with prior studies [38, 39], high-frequency activity in motor-related cortical areas was associated with movement execution. Time–frequency analysis of signals from primary motor cortex further supports this observation. Unlike findings under constrained conditions, this association was not consistently observed, as neural recordings during free movement contained more abundant spontaneous activity unrelated to movement (see Fig. A1 in the Appendix for details).

3.2 Kinematic representation of monkey movement

Joint definitions We defined 20 body keypoints of the monkey, including head, nose, left ear, right ear, neck, left shoulder, left elbow, left wrist, right shoulder, right elbow, right wrist, back, sacrum, left thigh, left knee, left ankle, right thigh, right knee, right ankle, and tail tip. We tracked and reconstructed the 3D coordinates of all keypoints in the world coordinate frame. Among these, we define 15 keypoints (nose, neck, left shoulder, left elbow, left wrist, right shoulder, right elbow, right wrist, sacrum, left thigh, left knee, left ankle, right thigh, right knee, and right ankle) including nose as "joints" for the kinematic representation, although the nose is not anatomically a true joint.

2D tracking of keypoints and 3D reconstruction of their coordinates We first track keypoints in video recordings from the view of the 8 cameras. We used a clustering-based strategy to select 200 representative video frames from each camera for manual annotation of keypoint coordinates. A "top-down" pose estimation pipeline from DEEPLABCUT [40] was adopted to identify keypoints in the video, which helped avoid spurious detection outside the cage caused by reflections from the acrylic cage walls. Specifically, we first performed monkey-level detection to localize the animal, and then estimated keypoint coordinates within the detected region. Camera calibration was performed using a standard checkerboard, and the resulting calibration parameters were used to reconstruct 3D joint coordinates in the world coordinate system (defined as shown in Fig. 1) with the Anipose toolbox [41]. We imposed fixed bone-length constraints for all limb segments except the Neck-Back and Back-Sacrum connections, since the Back point is located on the animal's jacket and thus may shift relative to the underlying body. During triangulation, we further applied temporal smoothing to joint trajectories to reduce tracking-induced jitter and non-physical discontinuities.

Kinematic representation of movement We first defined a root coordinate system. The origin of root frame was set at the Sacrum joint in the world coordinate system, and the direction from Sacrum to Neck defined the z-axis. The projection of the z-axis onto the world x-y plane was then rotated counterclockwise by 90° to define the y-axis, while the x-axis was determined by the right-hand rule. Since root rotation was typically the most difficult component to decode in movement prediction tasks, this definition constrained the rotational degrees of freedom of the root frame to two, thereby mitigating the decoding difficulty.

We then computed analytical solutions for joint rotations associated with joint trajectories using inverse kinematics, in contrast to numerical optimization methods employed in human models [42]. We assumed fixed limb lengths during modeling and a constant limb length for each segment was estimated by averaging the lengths across the entire recording session. In addition, each child joint was constrained to two rotational degrees of freedom by removing rotation around the limb axis. The joint rotations were represented by Euler angles rather than rotation matrices. To avoid discontinuities caused by angle wrapping, we used the sine and cosine values of the Euler angles as the final rotation representation.

Next, We represent the state of a moving monkey as a matrix P, defined as

$$P = (t_{root}, r_{root}, j^p, j^r, j^v) \tag{1}$$

Which consists of the displacement ($t \in \mathbb{R}^3$) of the root frame relative to the previous time step, the rotation ($r \in \mathbb{R}^2$) of the root frame relative to the previous time step, joint positions ($j^p \in \mathbb{R}^{3*15}$) in

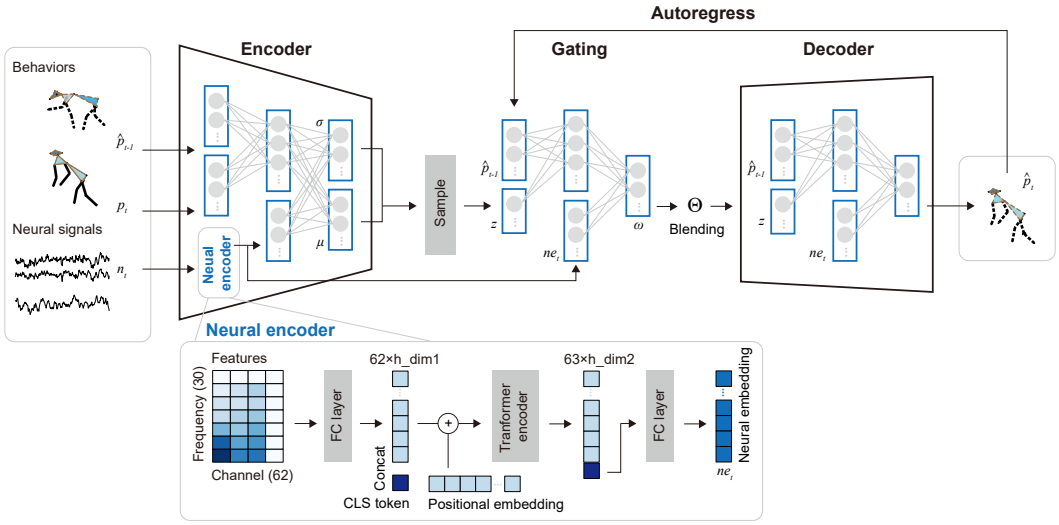

Figure 3: Framework of Neural-Behavioral Model. During training, the neural signal n_t was first preprocessed and transformed into spectral features. These features were then encoded by the neural encoder, and together with the previously predicted movement state $\hat{p}_{t-1}$ (except for the initial state) and the current movement state p_t, were compressed by the encoder into a latent space (μ, σ). The decoder was implemented as a mixture-of-experts neural network. A latent variable z was sampled from (μ, σ), and combined with the neural representation ne_t and the previous movement state $\hat{p}_{t-1}$. These inputs were fed into the decoder to autoregressively reconstruct the current movement state $\hat{p}_t$. During inference on the test set, only the decoder was used. A latent variable z was sampled from a standard Gaussian distribution, and together with the current neural representation ne_t and the previously predicted movement state $\hat{p}_{t-1}$, the decoder autoregressively predicted the current movement state $\hat{p}_t$.

the current root coordinate frame, joint rotations ($j^r \in \mathbb{R}^{4*14}$) in the current local joint coordinate frames, and joint velocities ($j^v \in \mathbb{R}^{3*15}$) expressed in the previous root coordinate frame. This yielded a unified state representation $P \in \mathbb{R}^{151}$.

The sampling rate of reconstructed movement data from original video was 30 Hz. To increase the amount of novel neural information associated with each prediction step, we downsampled the data from 30 to 10 Hz, thereby reducing redundancy between consecutive neural signal windows. We used the predicted t_{root}, r_{root}, j^r , together with the root-frame coordinates and rotation matrix from the previous time step, to compute the current joint coordinates via forward kinematics.

4 Experiment

In this section, we trained a behavior prior using a condition-guided autoregressive encoder-decoder model (Neural-Behavioral Model), and reconstructed monkey movements from neural signals. We then conducted a systematic evaluation of the reconstruction performance.

4.1 Neural-Behavioral model

Latent variable dynamics model Our goal is to model the probability distribution over state time series. We modeled the state transition process in which the transition model reconstructs the next state conditioned on the current behavior p_{t-1}, neural activity n_t and a latent variable z_t . For the behavior-only representation model (Behavioral Model) that does not incorporate neural features, the term n_t is omitted. θ denotes the learnable parameters.

$$P_\theta(p_0, p_1, ..., p_t) = P_\theta(p_0) \prod_{t=1}^{T} P_\theta(p_t \mid p_{t-1}, n_t) \tag{2}$$

$$P_\theta(p_t \mid p_{t-1}, n_t) = \int P_\theta(z_t \mid p_{t-1}, n_t)\, P_\theta(p_t \mid z_t, p_{t-1}, n_t)\, dz_t \tag{3}$$

During training, the latent variable was regularized to follow a standard Gaussian distribution. The ground-truth state p_t was used during optimization, and the encoder estimated the posterior distribution q_ϕ. The reconstruction loss was defined as the mean squared error (MSE) between the predicted and the ground-truth movement representations. The network was trained using the standard evidence lower bound (ELBO) [43] objective with the optimization target defined as follows.

$$\mathcal{L} = \|p_t - \hat{p}_t\|^2 + \beta D_{KL}\left(q_\phi(z_t \mid p_{t-1}, p_t, n_t) \parallel \mathcal{N}(0, I)\right) \tag{4}$$

At the inference stage, the model used the sampled z_t from a standard Gaussian distribution, previously predicted $\hat{p}_{t-1}$ and the current neural activity n_t to autoregressively generate movements step by step.

Neural-behavioral network The architecture of neural-behavioral network is illustrated in Fig. 3. The encoder consists of a two-layer MLP for encoding behavior inputs and a neural encoder (See Appendix for details) for neural signals. The behavioral and neural features were concatenated and passed through a single MLP to project them into a latent space of dimension 16. The decoder comprised six expert networks with identical structures and a gating network. The gating network was implemented as a three-layer MLP. The use of a mixture-of-experts design helped to alleviate posterior collapse [44]. Based on our empirical observations, neural signals were difficult to reconstruct accurately from a standard Gaussian latent space. Therefore, we did not include a neural reconstruction loss during training.

Data splits We conducted computational experiments using data (18,000 samples, 30 minutes) collected during the first session of the first behavioral experiment day. The first 20 minutes of data were used as the training set, while the remaining 10 minutes were used as the test set.

4.2 Reconstructing whole-body movement

To visualize model performance, Fig. 4 presents a representative example of joint kinematic predictions from different models during the monkey A's swinging movement from the upper bar to the floor (joint trajectories shown in Fig. A6 in the Appendix). We first evaluated the Neural-Behavioral Model, which reconstructed the current movement state from neural signals and the predicted state at the previous time step (except for the initial state). We observed that during the first 2 seconds (20 time steps), the model accurately reconstructed the monkey's whole-body movement, despite slight jitter and minor temporal misalignment. Although after 3 seconds, the accumulated error led to larger deviations, the model still preserved a consistent trend of movement.

To demonstrate the necessity of neural signals in the movement reconstruction, we evaluated a Behavioral Model without neural signals. In this case, the model primarily generated random but plausible movement sequences that reflected general behavior priors, rather than the actual movement of the monkey.

We adopted a previously reported method [15] that directly predicted movement from historical neural signals using an LSTM-based network and achieved state-of-the-art performance on constrained movement datasets [26]. We tested the LSTM for predicting velocities in the global coordinate system, velocities in the local coordinate system, and angular velocities in the local coordinate system. We reconstructed whole-body movements by predicting velocities in the global coordinate system, as the model achieved the best performance in this setting. Specifically, we used the LSTM to predict the velocity components along the x-, y-, and z-axes, and combined them with the given initial state to reconstruct the monkey's whole-body movement. The reconstructed skeletons shown in Fig. 4 illustrated LSTM's failure to recover bio-reasonable whole-body postures, while it predicted the overall movement trends (see Fig. A6 in the Appendix). The separately predicted joint positions exhibited substantial mismatches, resulting in anatomically implausible kinematics, such as joint hyperextension and unrealistic scaling or compression of body segments .

When initialized from a stationary state, the reconstructed movement remains static (such as sitting, see Fig. A7 in the Appendix). The model could also capture transitions such as initiating movement after a period of rest or returning to rest following movement (Fig. A7). However, it struggled to

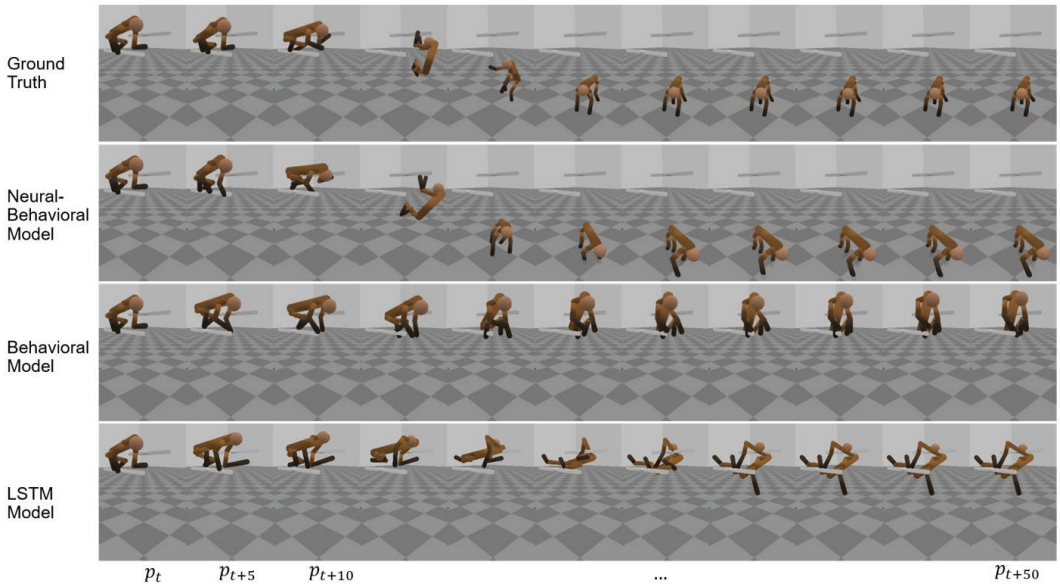

Figure 4: Visualization of reconstructed whole-body movement by three models during the monkey A's swinging movement from the upper bar. The p_t denotes the given initial state. Models predict the movement at each time step, and the predicted movements at every 5 time steps (0.5 s) are visualized. Note that Neural-Behavioral Model yields movements closest to the ground truth.

consistently reconstruct fine-grained movements, particularly when most body parts remained static while only a subset were active, as shown in Fig. A8.

4.3 Performance evaluation

We used the Average Displacement Error (ADE) and Final Displacement Error (FDE) to evaluate movement errors under different prediction time scale. The ADE is defined as the average MSE between reconstructed and ground-truth movement over all time steps, while FDE measures the MSE at the final time step. In addition, we computed the Pearson correlation coefficient between predicted and ground-truth joint trajectories to assess the global consistency of movement trends.

We labeled all movement sequences according to videos in the test set, with the remaining periods corresponding to resting behavior. We evaluated all movement sequences in the test set, together with two resting sequence. From each initial state of sequences, we autoregressively reconstructed 100 time steps (10 seconds). For each initial state we sampled the latent variable 5 times and reported the averaged performance. Table 1 below presents the performance evaluation results.

Table 1: Performance of various models in predicting movements for Monkey A. Average estimation error of joints over 3 seconds. The unit of ADE and FDE was cm.

Model	Horizontal movement			Vertical movement		
	ADE	FDE	r	ADE	FDE	r
Neural-Behavioral	**7.84**	**13.27**	**0.43**	**5.95**	**10.17**	**0.33**
Behavioral	10.36	19.19	0.21	9.20	19.74	0.25
LSTM[15]	9.38	15.29	0.24	7.78	15.20	0.23

In Table 1, we compared prediction results between the Neural-Behavioral Model and the Behavioral Model. Compared with previous studies on human datasets that were mostly dominated by horizontal movements, we analyzed performance separately for horizontal and vertical movements, since monkeys exhibited rich vertical behaviors such as climbing and swinging. We also compared the movement prediction by the state-of-art LSTM model that was used for decoding under constrained settings. The Neural-Behavioral Model outperformed the other two models. The results for Monkey

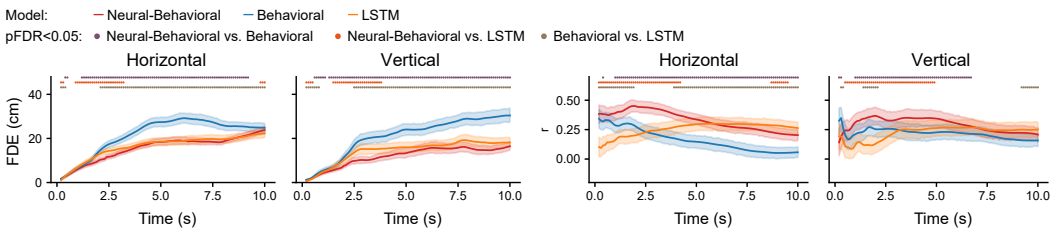

Figure 5: Model performance across various prediction time scales. Curves show the correlation coefficient FDE (left) and r (right) along the horizontal and vertical movement axes as functions of number of prediction steps. Shaded regions indicate 95% confidence intervals across 15 joints and 10 movements, with colors denoting data from three different models. Corrected P-values < 0.05 are marked with colored dots above each plot for pairwise model comparisons.

B are presented in Table A3 in the Appendix, the results were consistent with those from the monkey A. Notably, this is the first study to predict natural whole-body movement of freely moving monkey using population neuronal recordings, the performance of our Neural-Behavioral Model serves as a baseline for future research in this area.

In a previous work [1] that examined trained monkeys performing regular walking on a treadmill (rather than natural behaviors), neural signals recorded by microelectrode arrays were used to decode velocities of unilateral hindlimb joints during locomotion. This study reported a Pearson correlation coefficient of approximately 0.5 for predicting velocities of 4 joints. In contrast, our method achieved an average performance of whole-body joint positions with Pearson correlation coefficients 0.45 across all natural movements.

The Fig. 5 provides a detailed comparison of the predictions by the three models for various joints and movements over 1–100 prediction steps (corresponding to 0–10 s). Across both horizontal and vertical movement axes, the Neural-Behavioral Model achieved overall better decoding performance compared with the Behavioral and LSTM models, as reflected by higher correlation coefficients (r) and lower FDE (see Fig. A9 in the Appendix for prediction FDE across various joints and movements along 3 axes). Statistical significance between model pairs at each number of steps was assessed using paired Wilcoxon signed-rank tests across matched joint-movement conditions, followed by false discovery rate (FDR) correction for multiple comparisons.

In addition, the relative advantage of the models depended on the temporal scale of prediction. At smaller numbers of prediction steps (within $\sim$1 s), the Neural-Behavioral Model showed no significant differences from the other two models in some situations (e.g., FDE and r for vertical movements), and even performed worse than the Behavioral Model in terms of r during vertical movements within 400 ms. However, as the number of prediction steps increased to about 1.5 s, performance of Neural-Behavioral Model exceeded the Behavioral Model, showing as less FDE and higher r. In the first 4 seconds, Neural-Behavioral Model also outperformed LSTM in most situations.

These results indicate that Neural-Behavioral Model effectively integrates neural representations with behavior priors and learns a meaningful mapping between neural signals and behaviors. We further labeled the behavior data with movement categories, and incorporated discrete labels into the Behavioral Model and trained it as a Label-Behavioral Model. The comparison with models in section A.8 further supports that the Neural-Behavioral Model contributes more than merely identifying discrete movement types. Whereas the Behavioral Model could achieve performance comparable to that of the Neural-Behavioral Model in short-term predictions ($\sim$1 s), it failed to predict accurate movements over longer time scales. In contrast, the LSTM-based neural decoder developed under constrained settings failed to recover bio-reasonable whole-body postures. Although it could capture coarse movement trends (see Fig. A10 in the Appendix for predicted sacrum joint movement), it could not reconstruct physically plausible whole-body movements.

5 Limitation

Our model adopts a relatively simple model architecture, intended primarily as a proof of concept for capturing the neural representation of whole-body movements in a single freely moving monkey. In

addition, our experiments were conducted using data collected within a single day and evaluated on a limited and selected set of some representative movements.

6 Conclusion

In this work, we addressed the problem of reconstructing natural whole-body movement from cortical activity in freely moving primates, which is substantially more challenging than conventional motor decoding. We developed an experimental platform that synchronously records large-scale epidural cortical activities and multi-view 3D whole-body kinematics of freely moving monkeys. We further formulated natural whole-body decoding as a neural-conditioned autoregressive generative modeling problem. The proposed Neural-Behavioral Model integrates cortical representations with a learned behavior prior, allowing the decoder to reconstruct continuous, coherent, and realistic whole-body movements directly from neural signals without imposing explicit physical constraints. It achieved superior performance compared with approaches that are based solely on neural representations or behavior priors. This work provides a proof-of-concept demonstration that natural whole-body movement in freely moving primates can be reconstructed from large-scale cortical activity with behavior priors. We release the original monkey neural-behavioral data to support future research on neural representation learning and motor brain-machine interfaces. Future work with larger longitudinal datasets, multiple animals, and more expressive neural-behavioral models may further improve long-horizon prediction and generalization.

References

[1] Simon Borgognon, Nicolò Macellari, Alexandra M Hickey, Matthew G Perich, Houman Javaheri, Rafael Ornelas-Kobayashi, Maude Delacombaz, Christopher Hitz, Florian Fallegger, Stéphanie P Lacour, et al. Regional specialization of movement encoding across the primate sensorimotor cortex. *Nature communications*, 16(1):5729, 2025.

[2] Tatsuya Umeda, Masashi Koizumi, Yuko Katakai, Ryoichi Saito, and Kazuhiko Seki. Decoding of muscle activity from the sensorimotor cortex in freely behaving monkeys. *NeuroImage*, 197:512–526, August 2019.

[3] Chet T. Moritz, Steve I. Perlmutter, and Eberhard E. Fetz. Direct control of paralysed muscles by cortical neurons. *Nature*, 456(7222):639–642, December 2008.

[4] Jeffrey D Laurence-Chasen, Callum F Ross, Fritzie I Arce-McShane, and Nicholas G Hatsopoulos. Robust cortical encoding of 3d tongue shape during feeding in macaques. *Nature Communications*, 14(1):2991, 2023.

[5] Maarten C Ottenhoff, Maxime Verwoert, Sophocles Goulis, Simon Tousseyn, Johannes P van Dijk, Maryam M Shanechi, Omid G Sani, Pieter Kubben, and Christian Herff. Decoding continuous goal-directed movement from human brain-wide intracranial recordings. *Cell Reports*, 44(10), 2025.

[6] Benyamin Haghi, Tyson Aflalo, Spencer Kellis, Charles Guan, Jorge A. Gamez De Leon, Albert Yan Huang, Nader Pouratian, Richard A. Andersen, and Azita Emami. Enhanced control of a brain–computer interface by tetraplegic participants via neural-network-mediated feature extraction. *Nature Biomedical Engineering*, December 2024.

[7] Camille Testard, Sébastien Tremblay, Felipe Parodi, Ron W DiTullio, Arianna Acevedo-Ithier, Kristin L Gardiner, Konrad Kording, and Michael L Platt. Neural signatures of natural behaviour in socializing macaques. *Nature*, 628(8007):381–390, 2024.

[8] Benjamin Voloh, David J-N Maisson, Roberto Lopez Cervera, Indirah Conover, Mrunal Zambre, Benjamin Hayden, and Jan Zimmermann. Hierarchical action encoding in prefrontal cortex of freely moving macaques. *Cell reports*, 42(9), 2023.

[9] David A. Leopold. The big mixup: Neural representation during natural modes of primate visual behavior. *Current Opinion in Neurobiology*, 88:102913, 2024.

[10] Daniel Moran. Evolution of brain–computer interface: action potentials, local field potentials and electrocorticograms. *Current Opinion in Neurobiology*, 20(6):741–745, 2010. Motor systems – Neurobiology of behaviourv.

[11] Francesca Lanzarini, Monica Maranesi, Elena Hilary Rondoni, Davide Albertini, Elena Ferretti, Marco Lanzilotto, Silvestro Micera, Alberto Mazzoni, and Luca Bonini. Neuroethology of natural actions in freely moving monkeys. *Science*, 387(6730):214–220, 2025.

[12] Caleb Weinreb, Lakshanyaa Thamarai Kannan, Alia Newman-Boulle, Tim Sainburg, Winthrop F Gillis, Alex Plotnikoff, Sofia Makowska, Jonah E Pearl, Mohammed Abdal Monium Osman, Scott W Linderman, et al. Spontaneous behavior is a succession of self-directed tasks. *Neuron*, 2026.

[13] Natalie N Mueller, Mali Ya Mungu Ocoko, Youjoung Kim, Kate Li, Kaela Gisser, Gabriele Glusauskas, Isabella Lugo, Peter Dernelle, Anna Clarissa Hermoso, Jaime Wang, et al. Mechanically-adaptive, resveratrol-eluting neural probes for improved intracortical recording performance and stability. *npj Flexible Electronics*, 9(1):64, 2025.

[14] Mariana P Branco, Simon H Geukes, Erik J Aarnoutse, Nick F Ramsey, and Mariska J Vansteensel. Nine decades of electrocorticography: A comparison between epidural and subdural recordings. *European Journal of Neuroscience*, 57(8):1260–1288, 2023.

[15] Reza Eyvazpour, Behraz Farrokhi, and Abbas Erfanian. A general model based on riemannian manifold for stable decoding movement trajectory from ecog signals. *iScience*, 29(2), 2026.

[16] Sang Jin Jang, Yu Jin Yang, Seokyun Ryun, June Sic Kim, Chun Kee Chung, and Jaeseung Jeong. Decoding trajectories of imagined hand movement using electrocorticograms for brain–machine interface. *Journal of Neural Engineering*, 19(5):056011, October 2022.

[17] Guy Hotson, David P McMullen, Matthew S Fifer, Matthew S Johannes, Kapil D Katyal, Matthew P Para, Robert Armiger, William S Anderson, Nitish V Thakor, Brock A Wester, and Nathan E Crone. Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. *Journal of Neural Engineering*, 13(2):026017, April 2016.

[18] Alim Louis Benabid, Thomas Costecalde, Andrey Eliseyev, Guillaume Charvet, Alexandre Verney, Serpil Karakas, Michael Foerster, Aurélien Lambert, Boris Morinière, Neil Abroug, Marie-Caroline Schaeffer, Alexandre Moly, Fabien Sauter-Starace, David Ratel, Cecile Moro, Napoleon Torres-Martinez, Lilia Langar, Manuela Oddoux, Mircea Polosan, Stephane Pezzani, Vincent Auboiroux, Tetiana Aksenova, Corinne Mestais, and Stephan Chabardes. An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration. *The Lancet Neurology*, 18(12):1112–1122, December 2019.

[19] Maciej Śliwowski, Matthieu Martin, Antoine Souloumiac, Pierre Blanchart, and Tetiana Aksenova. Decoding ECoG signal into 3D hand translation using deep learning. *Journal of Neural Engineering*, 19(2):026023, April 2022.

[20] Jiarong Wang, Luzheng Bi, Weijie Fei, and Kun Tian. EEG-Based Continuous Hand Movement Decoding Using Improved Center-Out Paradigm. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2845–2855, 2022.

[21] Tessy M. Thomas, Daniel N. Candrea, Matthew S. Fifer, David P. McMullen, William S. Anderson, Nitish V. Thakor, and Nathan E. Crone. Decoding Native Cortical Representations for Flexion and Extension at Upper Limb Joints Using Electrocorticography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(2):293–303, February 2019.

[22] M Spüler, A Walter, A Ramos-Murguialday, G Naros, N Birbaumer, A Gharabaghi, W Rosenstiel, and M Bogdan. Decoding of motor intentions from epidural ecog recordings in severely paralyzed chronic stroke patients. *Journal of Neural Engineering*, 11(6):066008, oct 2014.

[23] Johan Wessberg, Christopher R Stambaugh, Jerald D Kralik, Pamela D Beck, Mark Laubach, John K Chapin, Jung Kim, S James Biggs, Mandayam A Srinivasan, and Miguel AL Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810):361–365, 2000.

[24] Marie-Caroline Schaeffer and Tetiana Aksenova. Switching Markov decoders for asynchronous trajectory reconstruction from ECoG signals in monkeys for BCI applications. *Journal of Physiology-Paris*, 110(4):348–360, November 2016.

[25] Kentaro Shimoda, Yasuo Nagasaka, Zenas C Chao, and Naotaka Fujii. Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in Japanese macaques. *Journal of Neural Engineering*, 9(3):036015, June 2012.

[26] Zenas C Chao, Yasuo Nagasaka, and Naotaka Fujii. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. *Frontiers in neuroengineering*, 3:1189, 2010.

[27] Andrea Colins Rodriguez, Matt G Perich, Lee E Miller, and Mark D Humphries. Motor cortex latent dynamics encode spatial and temporal arm movement parameters independently. *The Journal of Neuroscience*, 44(35):e1777232024, 2024.

[28] Behraz Farrokhi and Abbas Erfanian. A state-based probabilistic method for decoding hand position during movement from ECoG signals in non-human primate. *Journal of Neural Engineering*, 17(2):026042, April 2020.

[29] Henri Lorach, Andrea Galvez, Valeria Spagnolo, Felix Martel, Serpil Karakas, Nadine Intering, Molywan Vat, Olivier Faivre, Cathal Harte, Salif Komi, et al. Walking naturally after spinal cord injury using a brain–spine interface. *Nature*, 618(7963):126–133, 2023.

[30] Joel Ye and Chethan Pandarinath. Representation learning for neural population activity with Neural Data Transformers. *Neurons, Behavior, Data analysis, and Theory*, August 2021.

[31] Joel Ye, Jennifer Collinger, Leila Wehbe, and Robert Gaunt. Neural data transformer 2: multi-context pretraining for neural spiking activity. *Advances in Neural Information Processing Systems*, 36:80352–80374, 2023.

[32] Vincent Rouanne, Thomas Costecalde, Alim Louis Benabid, and Tetiana Aksenova. Unsupervised adaptation of an ECoG based brain–computer interface using neural correlates of task performance. *Scientific Reports*, 12(1):21316, December 2022.

[33] Yu Zhu, Chunfeng Song, Wanli Ouyang, Shan Yu, and Tiejun Huang. Neural representational consistency emerges from probabilistic neural-behavioral representation alignment. In *International Conference on Machine Learning*, pages 80107–80133. PMLR, 2025.

[34] Kaixi Tian, Shengjia Zhao, Yuhan Zhang, and Shan Yu. Multi-dimensional neural decoding with orthogonal representations for brain-computer interfaces. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 2092–2100, 2026.

[35] Xuan Wang, Kai Ruan, Liyang Qian, Guo Zhi Zhi, Chang Su, and Gaoang Wang. X-mogen: Unified motion generation across humans and animals. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 10234–10242, 2026.

[36] Lucas Martini, Alexander Lappe, Anna Bognár, Rufin Vogels, and Martin A Giese. Bigmaq: A big macaque motion and animation dataset bridging image and 3d pose representations. In *International Conference on Learning Representations*, volume 2026, 2026.

[37] Praneet C Bala, Benjamin R Eisenreich, Seng Bum Michael Yoo, Benjamin Y Hayden, Hyun Soo Park, and Jan Zimmermann. Automated markerless pose estimation in freely moving macaques with openmonkeystudio. *Nature communications*, 11(1):4560, 2020.

[38] Supratim Ray, Nathan E. Crone, Ernst Niebur, Piotr J. Franaszczuk, and Steven S. Hsiao. Neural correlates of high-gamma oscillations (60–200 hz) in macaque local field potentials and their potential implications in electrocorticography. *Journal of Neuroscience*, 28(45):11526–11536, 2008.

[39] Kai J. Miller, Eric C. Leuthardt, Gerwin Schalk, Rajesh P. N. Rao, Nicholas R. Anderson, Daniel W. Moran, John W. Miller, and Jeffrey G. Ojemann. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*, 27(9):2424–2432, 2007.

[40] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.

[41] Pierre Karashchuk, Katie L Rupp, Evyn S Dickinson, Sarah Walling-Bell, Elischa Sanders, Eiman Azim, Bingni W Brunton, and John C Tuthill. Anipose: A toolkit for robust markerless 3d pose estimation. *Cell reports*, 36(13), 2021.

[42] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.

[43] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations (ICLR)*, Banff, AB, Canada, 2014.

[44] Hung Yu Ling, Fabio Zinno, George Cheng, and Michiel Van De Panne. Character controllers using motion vaes. *ACM Transactions on Graphics (TOG)*, 39(4):40–1, 2020.

A Appendix

A.1 Representative time–frequency features of neural signals

We analyzed the time–frequency features of the primary motor cortex during the first 2 minutes of data (Fig. A1). By comparing the time–frequency features with the velocity magnitude of the right ankle, we observed an overall correlation between neural activity and movement. Specifically, during movement periods, beta-band de-synchronization and synchronization in the gamma and high-gamma bands were evident. However, compared to constrained movement condition [28], this coupling was less consistent. For example, around 7 s and 22 s, no clear synchronous activity was observed. This suggests that neural activity during free behavior contains abundant spontaneous activity that is not directly related to movement. In addition, during 50–55 s, the time–frequency features failed to capture rapid transitions in movement, this may be one of the reasons why algorithms developed under constrained conditions failed to reconstruct natural whole-body movements.

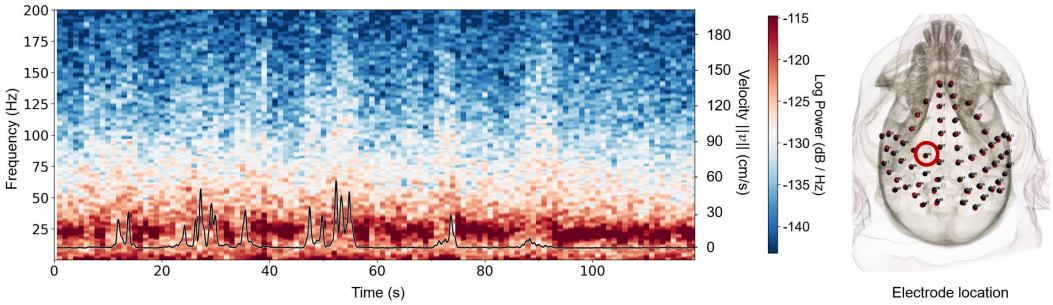

Figure A1: Time–frequency features of neural signals from the primary motor cortex. The time–frequency features are plotted together with the velocity magnitude of the right ankle in the world coordinate system (left panel). The right panel shows the electrode implantation locations (left primary motor cortex, red circles).

A.2 Evaluation of monkey kinematic representation

In our kinematic representation, the reduction of joint degrees of freedom may introduce reconstruction errors. To evaluate the fidelity of the representation, we converted the 3D cartesian joint coordinates into the kinematic representation and then reconstructed the cartesian coordinates from this representation. We found that across all joints and all frames (10 Hz, totaling 18,000 frames), the mean reconstruction error was -0.012 cm, the mean absolute error was 0.30 cm, and the maximum error was 18.23 cm. The distribution of reconstruction errors is shown in Fig. A2, most error values are within 1 cm, indicating that the error remains within an acceptable range.

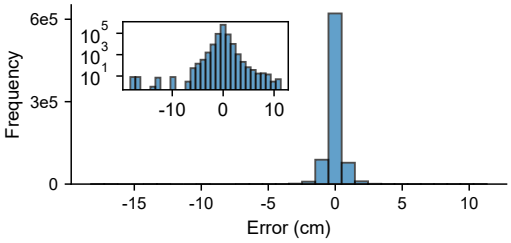

Figure A2: Distribution of reconstruction errors for 3D coordinates from the kinematic representation. The y-axis denotes the total number of frames within each error bin. The upper-left inset is shown with a log-scaled y-axis.

We computed the root-frame (sacrum joint) translation and rotation at each time step across 5 walking cases to characterize the monkey's overall movement velocity, as shown in Fig. A3. Regardless of the walking direction, the translational velocity along the z-axis remained positive (indicating relative forward locomotion) and had a larger magnitude than the velocities along the x- and y-axes, suggesting that the z-axis represented the dominant movement direction. In contrast, the x- and y-axis velocities contained both positive and negative values, reflecting the variability of movement trajectories. For angular velocity, taking yaw rotation as an example, comparison between case 1 and case 2 showed that the monkey moved toward the left-front and right-front directions, respectively. The yaw velocity curves exhibited opposite rotational directions while maintaining similar temporal trends in velocity magnitude. Comparing case 1 and case 3, the overall movement direction was

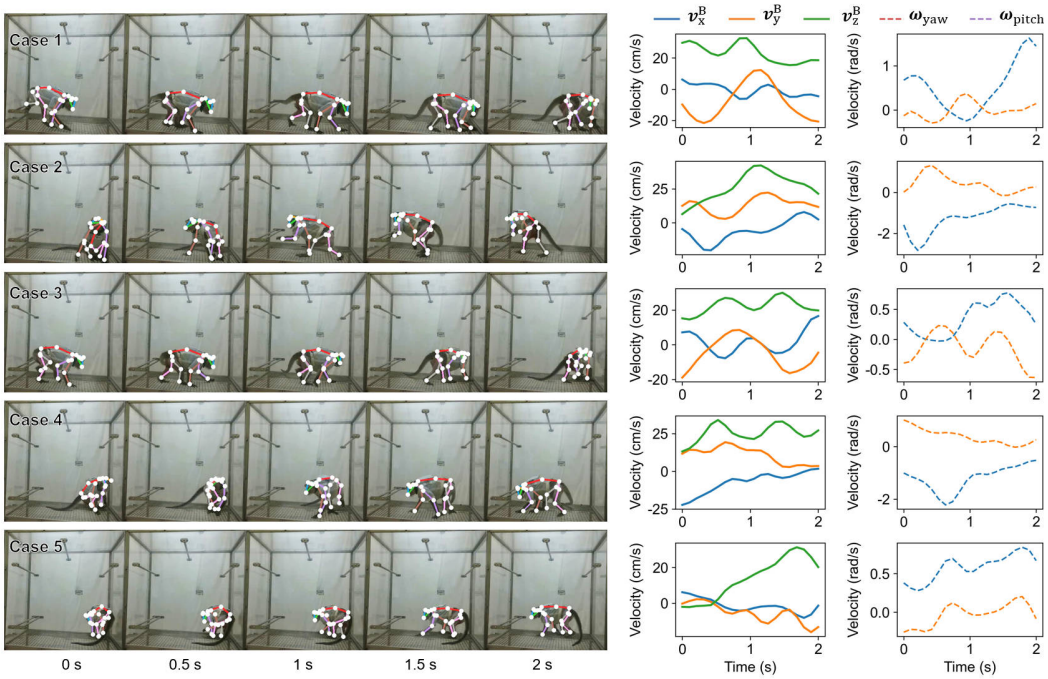

Figure A3: Example of monkey kinematic representation. Left: video frames from five walking segments with the 3D skeleton overlaid on the video. Case 1: 00:27–00:29, Case 2: 00:47–00:49, Case 3: 03:25–03:27, Case 4: 06:54–06:56, Case 5: 12:31–12:33. Right: the corresponding velocity of the root joint (sacrum), expressed in the root coordinate frame at the previous time step, for each walking case. yaw: rotation about the defined x-axis; pitch: rotation about the defined y-axis.

similar, but differences in angular velocity remained, reflecting variability in posture dynamics during movement. These results suggested that our kinematic representation captured meaningful movement characteristics of the monkey and facilitated the learning of appropriate behavior priors.

A.3 Neural encoder detail

Neural signal preprocessing and feature extraction The signals from the 62 channels were originally sampled at 2000Hz. The blink-related electrooculographic components, separated using ICA, were verified for time alignment with the eyelid closure frames in the video. After confirming the time alignment, we applied filtering (1-300Hz), manual artifact removal, ICA-based artifact rejection, and re-referencing (averaging the left and right hemispheres separately). We extracted frequency-domain features as input to the neural encoder. Specifically, for each channel, the raw signal was first downsampled to 400 Hz. We then applied a sliding window with a step size of 100 ms (aligned with the movement data) and computed features over the past 0–500 ms segment, including multi-taper power spectral density (PSD) and complex Morlet wavelet transforms. We divided the 1–150 Hz frequency range into 15 log-space bins and averaged the wavelet coefficients within each bin, and divided the same frequency range into 15 linear-space bins and computed the corresponding averaged multi-taper PSD. The resulting representation formed a feature matrix of size $\mathbb{R}^{C\times 30}$, where C denoted the number of recording channels.

Transformer-based neural encoder To extract movement-related representations from noisy neural signals, we employed a Transformer-based neural encoder that mapped multi-channel neural activity into a compact latent representation as shown in Fig. 3, and captured both spatial dependencies across electrode channels and spectral patterns in the frequency domain. For each time window, the neural input was represented as $N \in \mathbb{R}^{C\times F}$, where C denoted the number of recording channels and F denoted the number of frequency-band features extracted per channel. Each row corresponded to the energy or statistical features of a single channel across multiple frequency bands. The network first applied a linear projection layer to map the raw spectral features into a hidden dimension.

Since each channel corresponded to a fixed electrode location in physical space, we further added learnable positional embeddings along the channel dimension to explicitly encode the spatial layout of the electrodes. A learnable global token was then added to the sequence to aggregate information across all channels. The resulting sequence was processed by standard Transformer blocks to model inter-channel dependencies. Finally, a regression head (LayerNorm followed by an MLP) maped the global representation to a movement-related latent variable.

A.4 Training strategy

To improve the stability of long-term movement generation in the autoregressive model, we divided the training process into three stages. During the initial stage (teacher), the model used the ground-truth p_{t1} as inputs. This stage enabled the model to quickly learn stable temporal representations and accelerate convergence. During the intermediate transition stage (ramping), the input p_{t1} was updated using either the ground-truth movement or the model-predicted movement, determined randomly with a probability that gradually increased from 0 to 1. In the final stage (student), the model recursively generated motions using the predicted p_{t1}, corresponding to fully autoregressive training.

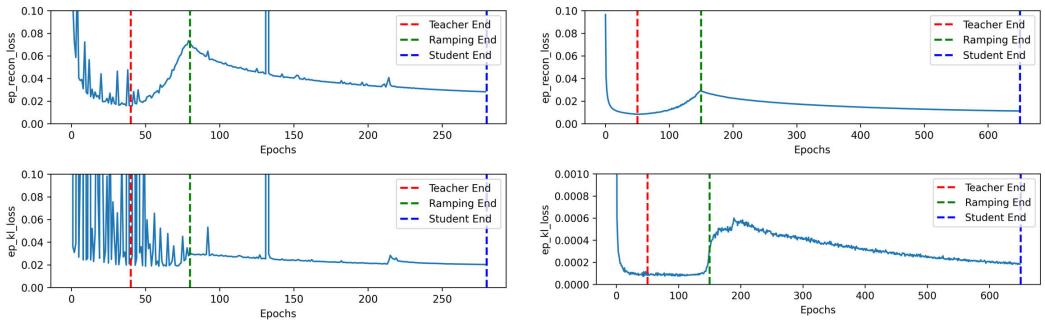

Figure A4: Example of train process. Loss curves before (left) and after (right) applying gradient-based constraints. We find that after applying the constraints, the training process remains stable over longer epochs, and the KL loss converges more rapidly.

We incorporated gradient-based constraints to regularize the training process and improve optimization stability. As shown in the Fig. A4, applying the regularization significantly alleviated gradient-related issues and accelerated convergence during training. We used the reconstruction error on the test set to determine the number of training epochs, and the final model was trained for 160 epochs (taking approximately 1 hour on an NVIDIA GeForce RTX 4090 D 24GB GPU).

A.5 Comparison with related work

Table A1: Comparison of our data with related datasets. In the "Species" column, M denotes monkey, H human, and Mul. multiple. In the "Body parts" column, FL denotes forelimbs and HL denotes hindlimbs. The column "Label type" differentiates between 2D coordinates, 3D coordinates, and multiple (like EMG, behavior categories, and 3D shape representations). In the "Neural signal" column, MEA denotes micro-electrode array recording.

Dataset	Species	Condition	Movement		Neural signal	Open
			Body parts	Label type		
Ours	M	Natural	21 joints	3D	Epidural	Yes
UniMo4D[35]	Mul.	Natural	25 joints	3D	No	Part
BigMaQ[36]	M	Natural	20 joints	Multiple	No	Yes
O.M.S[37]	M	Natural	13 joints	3D	MEA	Yes
G.C.[1]	M	Constrained	10 joints (HL)	Multiple	MEA	No
NT[37]	M	Constrained	6 joints (FL)	2D	Epidural	Yes

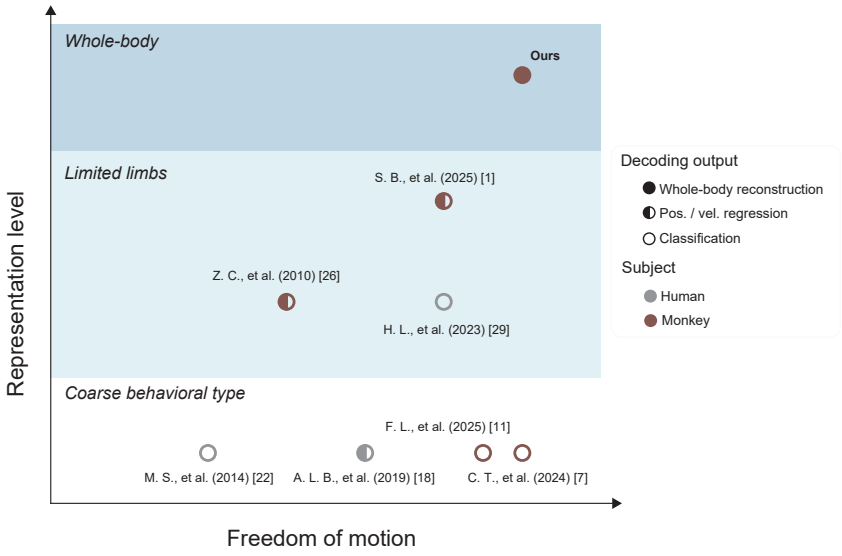

Figure A5: Comparison of behavioral paradigms across representative neural decoding studies.

A.6 LSTM Model detail

LSTM and Neural-Behavioral Model used the same inputs, including the neural features and the identical train/test data split. The model consists of two LSTM layers, one bidirectional LSTM layer, and a final fully connected regression layer. For the hidden dimensions, we performed Bayesian hyperparameter optimization, searching the size of the first LSTM layer in the range of 80–200, following the settings in [15, 4]. The sizes of the second and third layers were scaled proportionally according to the architecture in the origin paper[15]. During training, the learning rate was linearly decayed from 1e-3 to 1e-4. Bayesian optimization selected 104 hidden units for the first LSTM layer, which is very close to the 100 units used in the paper. Therefore, we adopted the setting of 100 units in our final experiments.

A.7 Performance of 2D tracking and 3D reconstruction

The 2D keypoint tracking performance on the test set achieved an mAP of 0.978 and an mAR of 0.983, with a mean prediction error of 4.98 pixels across all joints. Owing to the high accuracy of the 2D keypoint detector and the use of 8 synchronized camera views, 80.4% of the video frames contain observations of all joints from at least two cameras. This provides high-quality 3D reconstruction.

A.8 Contribution of neural representations in predicting whole-body movements

We believe that the Neural-Behavioral Model learns to align continuous neural representations with latent movement representations, rather than merely identifying discrete movement types. This enables neural signals to guide the generation of accurate and physically plausible whole-body movements. We incorporated discrete movement labels into the Behavioral Model and trained it as a Label-Behavioral Model. We labeled the behavior data with movement categories. During both training and inference, conditioning movements were generated on the one-hot embedded movement label. This Label-Behavioral Model conditioned by discrete movement labels performed slightly better than the Behavioral Model, but still clearly worse than the Neural-Behavioral Model, as shown in the Table A2 below.

This suggests that the Neural-Behavioral Model contributed more than merely identifying discrete movement types. It may be noted that epidural neural recordings are unlikely to provide all the information required to decode fine-grained movement kinematics. This is true even for large-scale intracortical recording with single-neuron resolution. Nevertheless, epidural recording exhibits long-term stability and low tissue damage, thus offering reliable information for decoding movement

Table A2: Performance of various models in predicting movements for Monkey A.

Model	Horizontal movement			Vertical movement		
	ADE	FDE	r	ADE	FDE	r
Neural-Behavioral	**7.84**	**13.27**	**0.43**	**5.95**	**10.17**	**0.33**
Behavioral	10.36	19.19	0.21	9.20	19.74	0.25
Label-Behavioral	10.22	17.05	0.31	8.45	18.04	0.24

kinematics, particularly in clinical settings. Its broader cortical coverage makes it a promising approach for neural decoding of whole-body movement.

A.9 Additional results

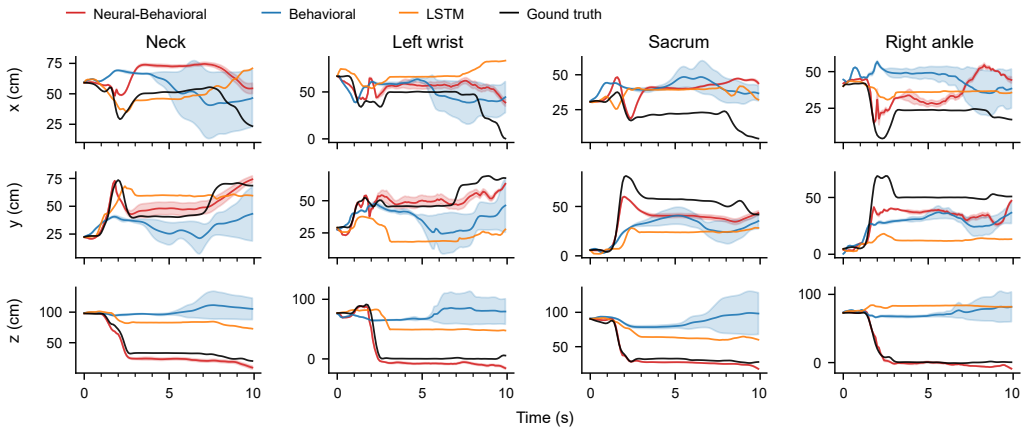

Figure A6: Trajectory comparison between ground truth (GT) and predictions from the three models. For the x, y, and z coordinates during a single swing-down movement (consistent with Fig 4), across four joints. The Neural-Behavioral Model more closely tracks the GT trajectories than the other two models, particularly along the y and z axes, where kinematic variations are more pronounced. The LSTM model shows the largest deviation from GT across all joints and axes. Line colors indicate different models, and shaded regions denote the standard deviation 5 repeated runs.

Table A3: Performance of various models in predicting movements for monkey B. Average estimation error of joints over 3 seconds. The unit of ADE and FDE was cm.

Model	Horizontal movement			Vertical movement		
	ADE	FDE	r	ADE	FDE	r
Neural-Behavioral	**10.99**	**17.94**	**0.48**	**11.46**	**12.54**	**0.45**
Behavioral	12.38	21.28	0.41	12.64	19.16	0.34
LSTM[15]	17.31	24.76	0.09	13.63	18.03	0.35



Figure A7: Visualization of additional reconstructed whole-body movements. p_t denotes the given initial state, and the predicted movements at every 5 time steps (0.5 s) are visualized. GT: ground truth, NB: Neural-Behavioral Model, B: Behavioral Mode.

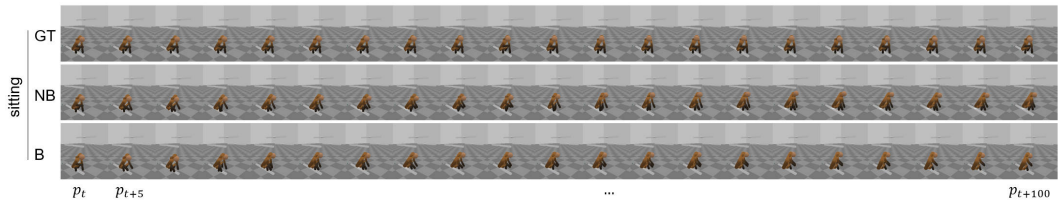

Figure A8: Visualization of a failed case, where the monkey remained largely stationary while making subtle movements of the elbow and head. p_t denotes the given initial state, and the predicted movements at every 5 time steps (0.5 s) are visualized. GT: ground truth, NB: Neural-Behavioral Model, B: Behavioral Mode.

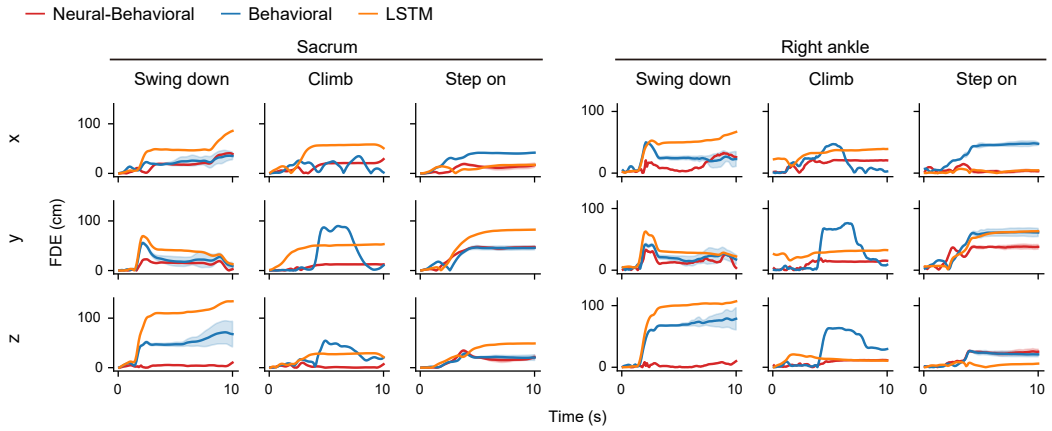

Figure A9: Prediction FDE of three models. Shown are x, y, and z axes of the “Swing down” “Climb” and “Step on” movements for the sacrum and right ankle joints. Overall, the Neural-Behavioral Model exhibits lower errors than the other two models. The improvement is especially obvious along the z-axis during “Swing down” and "Climb", along the y-axis during "Climb", as well as along the y-axis during “Step on” for the right ankle corresponding to the dominant movement directions. Line colors indicate different models, and shaded regions represent the standard deviation over 5 repeated runs. “Swing down”: swinging down from the upper bar to the floor; "Climb": climbing to the upper bar; "Step on": approaching and stepping onto the lower bar.

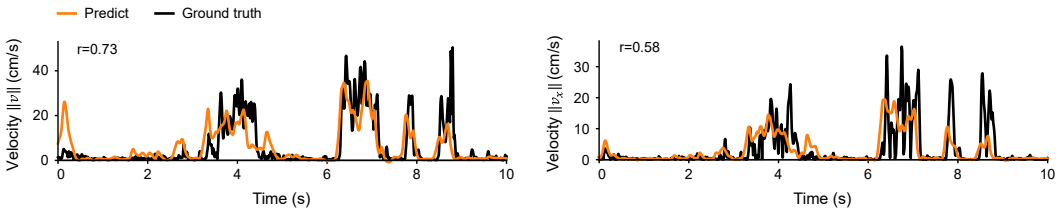

Figure A10: LSTM prediction. Using LSTM model to predict the 3D velocity vector magnitude, as well as velocity magnitude along the x-axis of the Sacrum joint in the monkey. The Pearson correlation coefficient r is shown in the upper-left corner of the figure. Note that all predicted velocities represent velocity magnitudes without directional information.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

Justification:

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification:

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification:

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification:

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification:

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification:

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification:

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification:

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification:

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.