

Machine Learning Methods for Studying Latent Neural Activity Dynamics

Shufeng Kong^{1,2}, Fumei Deng, Xinyi Dong, Caihua Liu^{5,4,2*}, Weiwei Chen³, Yingheng Wang², Daniel Cao², Azahara Oliva³, Antonio Fernández-Ruiz³ and Carla P. Gomes²

¹ School of Software Engineering, Sun Yat-sen University, Zhuhai, China

² Department of Computer Science, Cornell University Ithaca, NY, USA

³ Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

⁴ Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

⁵ School of Computer Science and Artificial Intelligence, Foshan University, Foshan, China
{sk2299, cl2869, wc677, yw2349, dyc33, aog35, afr77}@cornell.edu, gomes@cs.cornell.edu

Abstract

Recent developments in brain recording are driving a demand for machine learning tools capable of decoding the latent structure of large populations of neurons. In this paper, we provide a comprehensive survey that outlines the trajectory of Latent Variable Models (LVMs) from early state-space models to more recent deep generative models. We organize the literature into three closely related domains: (1) Single-Region Latent Dynamics, which includes models such as linear dynamical systems to more complex dynamics represented by Recurrent Neural Networks (RNNs) and Neural Ordinary Differential Equations (ODEs); (2) Multi-Region Communication, which employs probabilistic as well as subspace methods to study how information is transferred across different brain areas considering synaptic propagation delays and network connectivity; and (3) Behavior-Aligned Modeling, which seeks to disentangle neural activity related to task performance from other internal states via supervised or contrastive learning. Finally, we conclude and discuss benchmarks, evaluation criteria, and open challenges, such as the ability to identify causal links or directionality of communication, to facilitate future research for bridging interpretable brain dynamics with reliable neural decoding.

1 Introduction

Recording the activity of hundreds of neurons across brain areas in behaving animals is becoming commonplace in neuroscience study thanks to the development of large-scale neural recording tools [Urai *et al.*, 2022]. This wealth of data has prompted a revolution on the way neural activity is analyzed, to a shift in focus from single neuron responses to understanding the dynamics of large populations. An important conceptual drive in the development of novel analytical and modeling approaches is the idea that complex, high-dimensional neural dynamics can be largely explained by a much reduced

set of factors, or ‘latent variables’ [Schneider *et al.*, 2023]. Therefore, the main goal for Artificial Intelligence (AI) in this domain is no longer just to analyze data, but to reverse-engineer the latent dynamical rules that govern how biological networks process information. To that end, there is a recent but growing appeals in the computational neuroscience community for Latent Variable Models (LVMs). Inspired by the neural manifold assumption [Gao *et al.*, 2017], which suggests that the activity of N neurons is driven by a latent state $\mathbf{x}_t \in \mathbb{R}^K$ evolving on a low-dimensional manifold, a recent trend is to develop LVMs that satisfy dynamical, distributed, and behavioral criteria [Perich *et al.*, 2025].

In light of the rapid developments and emerging challenges of AI in neuroscience, we present a comprehensive survey of this field to help the community keep track of its progress. Specifically, we introduce a taxonomy that categorizes existing works into three phases: (1) **Single-Region Latent Dynamics** move from static smoothing to dynamical modeling. The fundamental idea is to leverage Recurrent Neural Networks (RNNs) [Pandarinath *et al.*, 2018] and Neural Ordinary Differential Equations (ODEs) [Kim *et al.*, 2021] to infer dynamical features such as attractors; (2) **Multi-Region Communication** builds on the success of single-region models to capture distributed interactions. To move a step forward, sophisticated subspace identification methods [Gokcen *et al.*, 2022; Semedo *et al.*, 2022; Liu *et al.*, 2025] have been used to show how information is routed between brain areas; (3) **Behavior-Aligned Modeling** integrates specialized learning frameworks to enhance behavioral decoding. Specifically, it applies contrastive learning [Schneider *et al.*, 2023] and supervised disentanglement [Sani *et al.*, 2024; Sani *et al.*, 2021] to isolate latent factors that directly correlate with actions.

Lastly, the field currently stands at the precipice of a fourth paradigm: Neural Foundation Models. Inspired by the success of Large Language Models (LLMs), recent works [Azabou *et al.*, 2023; Zhang *et al.*, 2025] treat neural spikes as tokens, aiming to learn a universal neural grammar that generalizes across individuals. While existing reviews have focused on specific architectures, biological questions, or foundation models, we provide a unified perspective on latent variable methods for analyzing neural population dynamics, cross-

*Correspondence to Caihua Liu: cl2869@cornell.edu

regional interactions, and neuro-behavioral alignment within a common generative framework. We organize the literature into three complementary pillars: single-region latent dynamics, multi-region communication, and behavior-aligned modeling, offering a roadmap for ML researchers to engage with one of the most complex dynamical systems in nature.¹

2 Taxonomy and Preliminaries

2.1 Preliminaries and Problem Formulation

To compare different latent variable approaches, we characterize neural population analysis as a generative process, which yields the general framework defined below and lets us classify methods by how they infer the latent neural states.

We generalize over specific architectural choices by regarding a neural recording as a sequence $\mathbf{Y}$ of T observations. We can now formulate the generative process in terms of the governing equations as follows:

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_t, \epsilon_t), \quad (1)$$

$$\mathbf{y}_t \sim \mathcal{P}(g(\mathbf{x}_t, \mathbf{s}_t)), \quad (2)$$

where $\mathbf{u}_t$ represents external inputs, and ϵ_t denotes process noise. The observation is drawn from a probability distribution $\mathcal{P}$, parameterized by the mapping function $g(\cdot)$. The term $\mathbf{s}_t$ accounts for optional history-dependent covariates. Existing work on latent variable models uses many different types of objective functions; some allow for only inference (estimating the posterior probability distribution), while others combine them using learning objectives (maximizing the likelihood).

2.2 Taxonomy of Latent Variable Models

We introduce our taxonomy and classification for the surveyed papers (see Table 1) from the perspective of a model designer, focusing mainly on the interplay between the objective function and the neural dynamics.

Single-Region Latent Dynamics

The most basic level focuses on individual brain regions alone, characterizing the intrinsic rules and latent factors that govern intra-area dynamics. Although the absence of inter-regional coupling makes this seem simpler, the surveyed papers reveal subtle complexity that varies with the function f : Neural ODEs [Kim *et al.*, 2021] and RNNs [Pandarinath *et al.*, 2018] both model single-region dynamics, yet identify different computational primitives such as attractors or limit cycles.

Multi-Region Communication

Cognitive functions rely on distributed populations interacting across regions, so a primary goal is to understand how regions influence one another and route information—challenging because of the bidirectional loops pervading most areas. To separate local computation from information routing, we analyze the latent manifold extracted by g : shared latent components are classified as multi-region communication, while region-private components are local computation.

Behavior-Aligned Modeling

This category operates within supervised or behavior-guided paradigms [Sani *et al.*, 2024; Schneider *et al.*, 2023], where the objective function acts as a functional constraint: the validity of a representation is judged by its alignment with behavior (e.g., “the latent space must disentangle motor output” or “the manifold must be invariant across subjects”). Such constraints effectively assign zero utility to any state $\mathbf{x}$ outside the set of behaviorally aligned states. We also examine large-scale pre-training for neurobehavioral modeling (e.g. Neuro-Foundation Models [Azabou *et al.*, 2023; Zhang *et al.*, 2025]) which target cross-subject generalization, zero-shot decoding, and multimodal integration.

3 Single-Region Latent Dynamics

3.1 Probabilistic and Linear State-Space Models

Gaussian Process-based Models

To handle the ill-posed nature of neural decoding, various state-space models have been developed. One of the first [Yu *et al.*, 2008] defines the latent state evolution not through explicit dynamics, but via a kernel function that enforces temporal smoothness. Specifically, *Gaussian-Process Factor Analysis* (GPFA) extracts neural trajectories by assuming the observed activity $\mathbf{y}_{:,t}$ is a linear projection of a latent state $\mathbf{x}_{:,t}$ driven by a Gaussian Process (GP) prior. The likelihood is given by:

$$\mathbf{y}_{:,t} \mid \mathbf{x}_{:,t} \sim \mathcal{N}(\mathbf{C}\mathbf{x}_{:,t} + \mathbf{d}, \mathbf{R}). \quad (3)$$

Much of the following work adds new inference strategies for non-Gaussian noise, since the standard GPFA assumption is ill-suited for discrete spike trains. The *Variational Latent Gaussian Process* (vLGP) [Zhao and Park, 2017] addresses this with a variational framework using a Poisson observation model, while *SNP-GPFA* [Keeley *et al.*, 2020] disentangles stimulus-locked signal from internal noise via Fourier-domain black-box variational inference. Finally, to overcome linear mappings, *P-GPLVM* [Wu *et al.*, 2017] employs GPs to govern both the temporal latent variable and non-linear tuning curves, discovering low-dimensional structure in complex manifolds such as hippocampal place-cell codes.

Linear Dynamical Systems

Unlike GPFA, *Linear Dynamical Systems* (LDS) explicitly model the generative process via a state transition matrix $\mathbf{A}$:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t, \quad (4)$$

where $\mathbf{u}_t$ represents external inputs. Then the Poisson Linear Dynamical System (PLDS) [Macke *et al.*, 2011] combines this linear backbone with a Poisson emission model:

$$\mathbb{E}[y_t^i \mid \mathbf{x}_t] = \exp([\mathbf{C}\mathbf{x}_t + \mathbf{d}]_i). \quad (5)$$

While PLDS captures temporal evolution better than GLMs, standard LDS ignores biological constraints. *CTDS* [Jha *et al.*, 2024] enforces Dale’s Law, which a neuron is purely excitatory or inhibitory, by restricting the connectivity columns to be non-negative, distinguishing excitatory from inhibitory populations and adding biophysical interpretability.

¹This paper’s appendix is available in the extended arXiv version. The project page is at <https://github.com/sk2299/neuro-ai-survey>.

Category	Representative methods
(1) Single-Region Latent Dynamics — intrinsic temporal evolution within one area	
Probabilistic & linear state-space	<i>Gaussian process</i> : GPFA [Yu <i>et al.</i> , 2008], vLGP [Zhao and Park, 2017], SNP-GPFA [Keeley <i>et al.</i> , 2020], P-GPLVM [Wu <i>et al.</i> , 2017]. <i>Linear dynamical</i> : PLDS [Macke <i>et al.</i> , 2011], CTDS [Jha <i>et al.</i> , 2024].
Deep sequential generative	<i>RNN VAEs</i> : LFADS [Pandarath et al., 2018], AutoLFADS [Keshtkaran <i>et al.</i> , 2022], DASC [Soo <i>et al.</i> , 2023], Stochastic Low-Rank RNN [Pals <i>et al.</i> , 2024]. <i>Transformer</i> : NDT [Ye and Pandarath, 2021].
Continuous & switching dynamics	<i>Neural ODE/SDE</i> : PLNDE [Kim <i>et al.</i> , 2021], FINDR [Kim <i>et al.</i> , 2023], ODIN [Versteeg <i>et al.</i> , 2023], LangevinFlow [Song <i>et al.</i> , 2026]. <i>Switching LDS</i> : rSLDS [Linderman <i>et al.</i> , 2017], gpSLDS [Hu <i>et al.</i> , 2024], DynCL [Gonzalez Laiz <i>et al.</i> , 2025].
(2) Multi-Region Communication — distributed information routing across areas	
Probabilistic generative	<i>Linear & GP</i> : DLAG [Gokcen <i>et al.</i> , 2022], mDLAG [Gokcen <i>et al.</i> , 2023], MRM-GP [Li <i>et al.</i> , 2024], ADM [Li <i>et al.</i> , 2025]. <i>Switching/nonlinear SSM</i> : MR-SDS [Karniol-Tambour <i>et al.</i> , 2024], mp-srSLDS [Glaser <i>et al.</i> , 2020], MRDS-IR [Dowling and Savin, 2025].
Deep representation learning	<i>Sequential/Transformer VAEs</i> : MR-LFADS [Liu <i>et al.</i> , 2025], GLM-Transformer [Xin and Kass, 2025], MtM [Zhang <i>et al.</i> , 2024]. <i>Network & geometric</i> : mRNNs [Perich and Rajan, 2020], FORCE/full-FORCE [DePasquale <i>et al.</i> , 2018], SPLICE [Hirayama <i>et al.</i> , 2017], MARBLE [Gosztolai <i>et al.</i> , 2025].
Subspace identification & alignment	<i>Communication subspace</i> : CCA [Semedo <i>et al.</i> , 2022], RRR-based [MacDowell <i>et al.</i> , 2025], Structured SNN+RRR [Gozel and Doiron, 2024]. <i>Multi-view</i> : CREIMBO [Mudrik <i>et al.</i> , 2025].
(3) Behavior-Aligned Modeling — disentangling behaviorally relevant latent factors	
Disentangled dynamic modeling	<i>Linear subspace</i> : PSID [Sani <i>et al.</i> , 2021], IPSID [Vahidi <i>et al.</i> , 2024]. <i>Nonlinear/RNN</i> : DPAD [Sani <i>et al.</i> , 2024], BRAID [Vahidi <i>et al.</i> , 2025], LINT [Valente <i>et al.</i> , 2022], SRNN [Zhang and Saxena, 2024].
Representation alignment & geometry	<i>Contrastive/geometric</i> : CEBRA [Schneider <i>et al.</i> , 2023], GLUE [Chou <i>et al.</i> , 2024]. <i>Cross-scenario consistency</i> : SABLE [Jude <i>et al.</i> , 2022], NLA [Cho <i>et al.</i> , 2023].
Generative & foundation models	<i>Deep generative</i> : LDNS [Kapoor <i>et al.</i> , 2024], BeNeDiff [Wang <i>et al.</i> , 2024], GNOCCHI [McCart <i>et al.</i> , 2024]. <i>Energy-based/autoregressive</i> : EAG [Ge <i>et al.</i> , 2026]. <i>Neuro-foundation</i> : POYO [Azabou <i>et al.</i> , 2023], POSSM [Ryoo <i>et al.</i> , 2026], NEDS [Zhang <i>et al.</i> , 2025].

Table 1: A taxonomy of latent neural modeling. Methods are organized into three pillars: (1) Single-Region Dynamics (e.g., GPFA, RNNs), modeling intrinsic temporal evolution; (2) Multi-Region Communication (e.g., CCA, DLAG), capturing distributed information routing; and (3) Behavior-Aligned Modeling (e.g., CEBRA, PSID), using disentanglement for functional identification. The framework culminates in Neuro-Foundation Models (e.g., POYO), which leverage large-scale pre-training.

3.2 Deep Sequential Generative Models

RNN-Based Variational Autoencoders

This class shows strong recurrence between time steps. *Latent Factor Analysis via Dynamical Systems* (LFADS) [Pandarath et al., 2018] is a key example: a Sequential Variational Autoencoder (SVAE) that infers the initial condition g_0 of a generator RNN. Then *AutoLFADS* [Keshtkaran *et al.*, 2022] eliminates the need for expert hyperparameter tuning by integrating Population-Based Training (PBT) and Coordinated Dropout, which enables the model to self-optimize its architecture. *DASC* [Soo *et al.*, 2023] further solves the gradient instability issues inherent in training leaky RNNs (which mimic membrane time constants) by introducing dynamically aligned skip connections. Finally, the *Stochastic Low-Rank RNN* (SLRNN) [Pals *et al.*, 2024] injects interpretability into the black box of RNNs by constraining the recurrent matrix to a low-rank structure, which allows for analytical derivation of fixed points and flow fields.

Transformer-Based Models

The *Neural Data Transformer* (NDT) [Ye and Pandarath, 2021] adapts the Transformer to neuroscience to capture more complex temporal interdependencies. Instead of recurrence,

NDT uses self-attention between temporal queries Q and keys K across the entire recording session. NDT utilizes a masked modeling objective (inspired by BERT), in which spikes are randomly masked and the model is trained to reconstruct them. This approach achieves faster inference than LFADS, demonstrating that explicit recurrence is not strictly necessary for recovering autonomous neural dynamics.

3.3 Continuous & Switching Dynamics

Continuous-Time Neural ODEs & SDEs

Continuous-time models prevent discretization errors. *PLNDE* [Kim *et al.*, 2021] and *ODIN* [Versteeg *et al.*, 2023] utilize Neural Ordinary Differential Equations (ODEs) to capture deterministic evolution. Formally, these models define the derivative as:

$$\dot{\mathbf{z}} = f_{\psi}(\mathbf{z}, t) + \sum_u \delta_{t_u}(\mathbf{g}_{\phi}^u + \eta_u). \quad (6)$$

To incorporate physical priors into continuous dynamics, LangevinFlow [Song *et al.*, 2026] further models the evolution of neural latents as an underdamped Langevin equation. Within a sequential VAE framework, this model unifies in-

trinsic deterministic dynamics with unobserved stochastic influences:

$$m \frac{\partial \mathbf{v}}{\partial t} = -\nabla_{\mathbf{z}} U(\mathbf{z}) - m\gamma \mathbf{v} + \sqrt{2m\gamma k_B \tau} \boldsymbol{\eta}(t) \quad (7)$$

To capture the oscillatory and flow-like characteristics inherent in neural activity, the potential function is innovatively parameterized as a locally coupled network of oscillators ($U(\mathbf{z}) = \mathbf{z}^T \frac{\mathbf{W}_z}{\|\mathbf{W}_z\|_2} \mathbf{z}$). Combined with a global Transformer decoder, the model can effectively disentangle and extract complex neural spatiotemporal dynamics.

However, neural dynamics often exhibit intrinsic variability. *FINDR* [Kim *et al.*, 2023] extends this paradigm to *Neural Stochastic Differential Equations* (SDEs). By parameterizing the drift with a Gated Neural Network and incorporating a diffusion term ($d\mathbf{z} = \mu(\mathbf{z}, \mathbf{u})dt + \sqrt{\Sigma}dW$), *FINDR* explicitly enables the visualization of latent flow fields and attractor structures, while simultaneously disentangling task-relevant dynamics from background noise.

Switching Linear Dynamical Systems

A *Switching Linear Dynamical System* (SLDS) models complex dynamics by preserving linearity within individual modes; strictly speaking, the system’s evolution remains linear as long as the underlying regime is maintained. An illustrative example is *Recurrent Switching Linear Dynamical System* (rSLDS), which approximates a global nonlinear landscape by dynamically switching between distinct linear modes. Formally, this class of methods considers functions such that:

$$\kappa_{ssl}(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^J (\kappa_{lin}^{(j)}(\mathbf{x}, \mathbf{x}') \cdot \kappa_{part}^{(j)}(\mathbf{x}, \mathbf{x}')). \quad (8)$$

In this framework, the switching kernel plays a pivotal role. As demonstrated in methods like *gpSLDS* [Hu *et al.*, 2024], the kernel enables the model to iteratively refine the mode allocation, searching for optimal structural decompositions. Finally, recent approaches leverage contrastive learning to enforce identifiability. *Dynamics Contrastive Learning* (DynCL) [Gonzalez Laiz *et al.*, 2025] introduces a framework that uncovers switching linear dynamics by optimizing an InfoNCE objective rather than a reconstruction loss. By differentiating through a bank of stable linear matrices, it identifies the underlying switching manifold without requiring the heavy computational overhead of generative RNNs.

4 Multi-Region Communication in Latent Space

4.1 Probabilistic Generative Models

These methods extend state-space models to capture non-stationary interactions and transmission delays between distinct neuronal populations.

Linear and Gaussian Process Models

A fundamental challenge in multi-region modeling is decoupling shared variance from local variance. The unified observation process is typically expressed as:

$$\mathbf{y}_t = \mathbf{C} \begin{bmatrix} \mathbf{x}_t^{shared} \\ \mathbf{x}_t^{private} \end{bmatrix} + \mathbf{d} + \boldsymbol{\epsilon}_t, \quad (9)$$

where latent states are partitioned into across-area variables $\mathbf{x}^{shared}$ and within-area variables $\mathbf{x}^{private}$.

DLAG [Gokcen *et al.*, 2022] and *mDLAG* [Gokcen *et al.*, 2023] integrated a multi-output squared-exponential kernel to account for the physical transmission time of neural signals. By explicitly parameterizing the delay τ between regions i and j in the covariance function, these models effectively learn directed functional connectivity alongside signal propagation latencies.

However, simple time shifts often fail to capture frequency-specific changes. In response to this, *MRM-GP* [Li *et al.*, 2024] employed complex-valued kernels to model phase delays. Furthermore, *ADM* [Li *et al.*, 2025] addresses non-stationary regimes by introducing a Markov Gaussian Process. Thanks to a time-varying transition matrix $\mathbf{A}_t$, the framework tracks how communication latencies evolve dynamically over the course of a single trial.

Switching and Nonlinear State-Space Models

Communication between brain areas is not static; it is gated by internal cognitive states (e.g., attention). To capture this, researchers leverage *Switching State-Space Models* (SSMs). The theoretical foundation for this approach is rSLDS [Linderman *et al.*, 2017]. Although originally designed for general non-linear population dynamics, rSLDS introduced the critical mechanism of “recurrence,” where the discrete regime switch depends on the continuous latent history:

$$z_t \mid z_{t-1}, \mathbf{x}_{t-1} \sim \text{Cat}(\pi(z_{t-1}, \mathbf{x}_{t-1})) \quad (10)$$

$$\mathbf{x}_t \mid \mathbf{x}_{t-1}, z_t \sim \mathcal{N}(f_{z_t}(\mathbf{x}_{t-1}), \mathbf{Q}_{z_t}). \quad (11)$$

In the context of distributed circuits, this generative structure is adapted to gate information flow rather than just local dynamics. For instance, *MR-SDS* [Karniol-Tambour *et al.*, 2024] extends the framework to multi-region settings, using the discrete state z_t to toggle specific communication channels. Similarly, *mp-srSLDS* [Glaser *et al.*, 2020] enhances the transition function $\pi(\cdot)$ with Neural Networks to model state-dependent stickiness, effectively capturing how neural populations sustain stable modes of communication over time.

4.2 Deep Representation Learning Models

Sequential and Transformer VAEs

To capture the nonlinear nature of neural communication, methods leverage deep representation learning. *MR-LFADS* [Liu *et al.*, 2025] learns directed nonlinear influence via a Sender-Receiver architecture. Specifically, it infers a latent state from a source region that acts as a forcing input to a target region by assuming the instantaneous firing rate log-intensity $\log \lambda(t)$ is composed of additive functional components. The decomposition is given by:

$$\log \lambda(t) = \mathcal{D}(\mathbf{x}_{local}) + \mathcal{C}(\mathbf{x}_{coupling}) + \mathcal{H}(\mathbf{y}_{history}), \quad (12)$$

where $\mathcal{C}$ represents the population-coupling component.

Standard deep learning models often lack transparency regarding trial-specific factors. The *GLM-Transformer* [Xin and Kass, 2025] addresses this by introducing a self-attention mechanism fused with *Generalized Linear Model* (GLM) interpretability. On the other hand, communication dynamics often vary a lot across distinct recording environments. *MtM*

(Multi-task Masking) [Zhang *et al.*, 2024] deals with these interdependencies by learning universal spatiotemporal representations via masked modeling, enabling generalization across different sessions and subjects.

Network and Geometric Dynamics

Methods in this category focus on the functional connectivity and intrinsic topology of the neural population. *mRNNs* [Perich and Rajan, 2020] fits large-scale time-series to a unified dynamical system, extracting a directed interaction matrix by assuming the evolution follows $\tau\dot{\mathbf{x}} = \mathcal{F}(\mathbf{x}, \mathbf{u}, \Theta)$. Complementing this inference approach, *FORCE* and *full-FORCE* [DePasquale *et al.*, 2018] provide protocols for training the recurrent connectivity of Neural Networks to match target dynamics. While these are primarily tools for modeling the mechanisms of neural computation rather than inferring latent states, they offer a framework for understanding how complex dynamics emerge from recurrent architectures.

A distinct and emerging trend moves beyond temporal prediction to focus on the geometry of the neural manifold. *SPLICE* [Hirayama *et al.*, 2017] pioneers this by introducing a hierarchical model with subspace pooling, learning representations that are invariant to local transformations (such as phase shifts). Building on this geometric perspective, *MARBLE* [Gosztolai *et al.*, 2025] addresses the sensitivity of models to local recording bases. It introduces an unsupervised geometric deep learning framework that decomposes dynamics into Local Flow Fields (LFFs), using contrastive learning to enforce invariance to local rotations and ensuring the learned dynamics are intrinsic to the underlying manifold.

4.3 Subspace Identification and Alignment

Methods for modeling interactions across distributed brain circuits must account for specific structural constraints to accurately capture inter-regional communication. Among these, the most critical requirements include achieving a low-rank dimensionality to isolate shared signals, maintaining designated orthogonality for factor disentanglement, and satisfying specific predictive targets between populations.

Communication Subspace Analysis

Communication Subspace Analysis (CSA) methods quantify the relationship between source activity $\mathbf{X}$ and target activity $\mathbf{Y}$ via a rank-constrained linear mapping:

$$\hat{\mathbf{Y}} \approx \mathbf{X}\mathbf{B}_{\text{RRR}}, \quad \text{s.t. rank}(\mathbf{B}_{\text{RRR}}) = m, \quad (13)$$

where m represents the dimensionality of the communication channel. *Canonical Correlation Analysis* (CCA) [Semedo *et al.*, 2022] identifies dimensions where source and target are maximally correlated. *Reduced Rank Regression* (RRR) [MacDowell *et al.*, 2025] prioritizes predictive accuracy under rank constraints. *Structured SNN + RRR* [Gozel and Dörion, 2024] extends this framework by embedding the regression objective within a Spiking Neural Network training loop, enforcing biological plausibility (e.g., separate E/I populations) on the learned subspace.

Multi-View Alignment

Multi-view methods mitigate dependency on session-specific dynamics, which mainly follow the dictionary learning

paradigm. To capture the underlying dynamical process, *CREIMBO* [Mudrik *et al.*, 2025] employed multiple ensemble compositions to decompose the observed activity $\mathbf{Y}^d$ into session-specific factors. Based on it, the model formulated the latent trajectory $\mathbf{X}^d$ using a shared library of dynamics:

$$\mathbf{Y}^d \approx \mathbf{A}^d \mathbf{X}^d, \quad \mathbf{X}_{t+1}^d = \mathbf{F}_t^d \mathbf{X}_t^d, \quad (14)$$

$$\mathbf{F}_t^d = \sum_{k=1}^K c_{k,t}^d \mathbf{f}_k. \quad (15)$$

The extracted features $\{\mathbf{f}_k\}_{k=1}^K$ could then serve as a dictionary of latent sub-circuits invariant across sessions. However, the acquisition of shared motifs is not always straightforward, and the framework effectively captures how different brain regions recruit these universal dynamical motifs over time via the coefficients $c_{k,t}^d$.

5 Behavior-Aligned Latent Modeling

5.1 Disentangled Dynamic Modeling

The core hypothesis here is that the full latent state $\mathbf{x}_t$ can be factorized into orthogonal subspaces: $\mathbf{x}_t = [\mathbf{x}_t^{\text{beh}}; \mathbf{x}_t^{\text{null}}]$, where only $\mathbf{x}_t^{\text{beh}}$ drives the target behavior.

Linear Subspace Disentanglement

PSID (Preferential Subspace Identification) [Sani *et al.*, 2021] pioneered this approach by redefining the state-space objective. Instead of maximizing variance explained in neural activity $\mathbf{y}$, it identifies a subspace that maximizes the mutual information between past neural activity and future behavior $\mathbf{z}$. This is achieved via a two-stage projection:

$$\hat{\mathbf{z}}_f = \mathbf{z}_f \mathbf{y}_p^T (\mathbf{y}_p \mathbf{y}_p^T)^{-1} \mathbf{y}_p, \quad (16)$$

where $\mathbf{y}_p$ is past activity and $\mathbf{z}_f$ is future behavior. *IPSID* [Vahidi *et al.*, 2024] extends this to distinguishing internal dynamics from external sensorimotor inputs $\mathbf{u}_t$, preventing the model from misinterpreting stimulus-driven activity as intrinsic computation.

Non-linear & RNN-based Disentanglement

To capture complex non-linear dynamics, *DPAD* [Sani *et al.*, 2024] employs a dual-objective RNN architecture. It learns a non-linear mapping by explicitly partitioning the latent space into behavior-aligned and behaviorally irrelevant dimensions, penalizing the latter to ensure they do not encode action-related information. *BRAID* [Vahidi *et al.*, 2025] advances this framework by integrating a forward predictive objective to resolve causal structure. Its generative model explicitly accounts for external control signals:

$$\mathbf{x}_{t+1}^{\text{beh}} = \mathbf{A}_\psi(\mathbf{x}_t^{\text{beh}}) + \mathbf{K}_\phi(\mathbf{u}_t) + \mathbf{w}_t, \quad (17)$$

thereby forcing the model to distinguish how inputs $\mathbf{u}_t$ causally drive the behavioral state, separate from internal residual dynamics. Furthermore, *LINT* [Valente *et al.*, 2022] enhances interpretability by imposing low-rank constraints on the RNN connectivity matrix. In contrast, *SRNN* [Zhang and Saxena, 2024] focuses on temporal disentanglement, utilizing switching dynamics to identify discrete cognitive states (e.g., “planning” vs. “execution”), effectively isolating distinct modes of neural computation.

5.2 Representation Alignment and Geometry

Contrastive Learning & Geometric Structure

CEBRA [Schneider *et al.*, 2023] represents a paradigm shift from generative to discriminative modeling. It employs contrastive learning to shape the latent space, pulling together neural time-points that share the same behavioral label (positive pairs) while pushing apart others. The InfoNCE-based objective is:

$$\mathcal{L} = \mathbb{E} \left[-\psi(\mathbf{x}_i, \mathbf{x}_i^+) + \log \sum_{j=1}^K e^{\psi(\mathbf{x}_i, \mathbf{x}_j)} \right]. \quad (18)$$

This allows CEBRA to align manifolds across different animals performing the same task, even if the underlying neurons are completely different. Complementing this, *GLUE* [Chou *et al.*, 2024] introduces a geometric metric (i.e., Manifold Capacity) to quantify the efficiency of this encoding:

$$\alpha_{\text{glue}} = \frac{\Psi_{\text{glue}} \cdot (1 + R_{\text{glue}}^{-2})}{D_{\text{glue}}}, \quad (19)$$

providing a standardized benchmark for comparing how distinct neural populations encode identical information.

Cross-Scenario Consistency Alignment

Experimental conditions often drift over time. *NLA* [Cho *et al.*, 2023] tackles temporal misalignment (jitter) using a differentiable time-warping function integrated into a contrastive objective. *SABLE* [Jude *et al.*, 2022] addresses distribution shift using unsupervised Domain Adaptation. It employs a gradient reversal layer to penalize the encoder if it retains session-specific information (e.g., recording artifacts), effectively learning a domain-invariant latent space robust to non-stationarity.

5.3 Generative and Foundation Models

Deep Generative Models

The first category focuses on deep generative priors, with a particular emphasis on *Diffusion-Based Models*. *LDNS* [Kapoor *et al.*, 2024] combines structured state-space models with Latent Diffusion; it first compresses spikes into a continuous latent space, then trains a diffusion model to generate these trajectories. *BeNeDiff* [Wang *et al.*, 2024] extends this to video diffusion, serving as an interpretability tool. It visualizes the behavioral meaning of a latent neuron by generating video frames conditioned on specific latent values using classifier-free guidance:

$$\hat{\epsilon}_\theta(\mathbf{z}_t, \mathbf{c}) = (1 + w)\epsilon_\theta(\mathbf{z}_t, \mathbf{c}) - w\epsilon_\theta(\mathbf{z}_t, \emptyset), \quad (20)$$

where $\mathbf{c}$ represents the behavioral condition and w scales the guidance strength.

GNOCCI [McCart *et al.*, 2024] extends the InfoDiffusion framework to neural time-series data, aiming to infer disentangled latent behavioral variables from recorded neural activity in an unsupervised manner. The method utilizes an auxiliary variable encoder to extract raw neural activity into a latent code $\mathbf{c}$ regularized via Maximum Mean Discrepancy (MMD); subsequently, a noise prediction network performs conditional diffusion generation conditioned on this code. To

guarantee high-fidelity generation while preserving the underlying behavioral structure, the model jointly optimizes the score-matching loss $\mathcal{L}_{\text{score}}$ and the direct data reconstruction loss $\mathcal{L}_{\text{recon}}$:

$$\begin{aligned} \mathcal{L}_{\text{score}} &= \mathbb{E} \left[\|\epsilon_i - \tilde{\epsilon}_i\|_2^2 \right] \\ \mathcal{L}_{\text{recon}} &= \mathbb{E} \left[\|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|_2^2 \right] \end{aligned} \quad (21)$$

where $\tilde{\epsilon}_i$ is the predicted noise and $\hat{\mathbf{x}}_0$ represents the reconstructed neural signal sample derived from the predicted noise. Once trained, researchers can conditionally generate high-quality neural activity for unseen behavioral conditions simply by performing a linear traversal within the disentangled latent code space.

Beyond diffusion, other generative frameworks enforce structural constraints. *Masked VAE* [Schulz *et al.*, 2025] unifies dimensionality reduction with imputation by training on randomly masked spike bins, forcing the model to learn robust joint statistics. Similarly, the *Multiscale Dynamical Model* [Abbaspourazad *et al.*, 2021] integrates disparate data types—discrete spikes (millisecond scale) and continuous Local Field Potentials (LFP, second scale)—within a hierarchical EM framework to identify shared latent states that bridge these temporal resolutions.

Energy-Based & Autoregressive Models

To eliminate the prohibitive iterative inference latency of diffusion models while maintaining high-fidelity neural spiking statistics, *EAG* [Ge *et al.*, 2026] introduces an Energy-based Autoregressive Generation framework to model neural population dynamics in continuous latent spaces without explicit likelihood computation. To preserve the stochastic nature of trial-to-trial neural variability, EAG optimizes an energy-based Transformer via the Energy Score, a strictly proper scoring rule defined as:

$$S(p_\theta, \mathbf{z}_{\text{data}}) = \mathbb{E}[\|\mathbf{z}_1 - \mathbf{z}_2\|^\alpha] - 2\mathbb{E}[\|\mathbf{z} - \mathbf{z}_{\text{data}}\|^\alpha], \quad (22)$$

where $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z} \sim p_\theta$ are independent samples generated by the model, and $\alpha \in (0, 2)$ ensures strict propriety to accurately capture single-trial neural statistics.

Neuro-Foundation Models

The third category comprises Neuro-Foundation Models, which prioritize cross-subject generalization. *POYO* [Azabou *et al.*, 2023] treats neural recordings as a sequence modeling task, utilizing a unit embedding tokenization scheme:

$$\mathbf{h}_i = \text{UnitEmbed}(\text{neuron ID}) + \text{RoPE}(t_i), \quad (23)$$

which allows the model to process arbitrary sets of neurons from different animals. By leveraging Cross-Attention Transformer blocks, POYO learns a universal mapping between neural tokens and behavioral variables. To alleviate the high computational complexity and latency constraints of such pure Transformer architectures in real-time, closed-loop settings, *POSSM* [Ryoo *et al.*, 2026] introduces a hybrid sequence framework. POSSM retains POYO’s single-spike tokenization and cross-attention mechanism to flexibly align neuron identities across diverse sessions, but replaces the heavy Transformer backbone with a modern recurrent

state-space model. This modification achieves up to a $9\times$ inference speedup while preserving powerful cross-subject generalization, scaling models for clinical deployment.

NEDS [Zhang *et al.*, 2025] unifies encoding (behavior $\rightarrow$ brain) and decoding (brain $\rightarrow$ behavior) into a single Transformer. It employs a multi-task masking strategy (i.e., randomly masking neural tokens, behavioral tokens, or cross-modal segments) to force the model to learn a joint, generalized representation. This pre-training paradigm allows NEDS to fine-tune on a new animal with limited data, effectively overcoming the data scarcity bottleneck inherent in individual experiments.

6 Datasets and Benchmarks

6.1 Standardized Benchmarks

To overcome the inconsistencies of ad-hoc evaluations, the field has adopted common benchmarking platforms. A foundational effort is the *Neural Latents Benchmark* (NLB) [Pei *et al.*, 2021], which evaluates models on standardized, held-out test sets rather than custom splits, providing rigorous and comparable metrics. Subsequent initiatives have expanded this framework to specialized task domains. While the commonly used monkey hand reaching task serves as a standard baseline, it is often ill-suited for isolating specific sensory feedback mechanisms. The Area2-Bump benchmark addresses this limitation by challenging models to explicitly decouple somatosensory perturbations. Similarly, addressing internal cognitive processes where timing evolves dynamically over a single trial, DMFC-RSG evaluates the capacity to capture non-motor dynamics and temporal interdependencies. Finally, Jenkins-Maze focuses on algorithmic efficiency, testing the scalability of strategies to meet the requirements of large-scale neural recordings.

6.2 Large-Scale Public Datasets

Foundation-Scale Corpora

The first category, Foundation-Scale Corpora, focuses on massive, multi-region coverage. The *International Brain Laboratory* (IBL) [Angelaki *et al.*, 2025] provides a massive corpus containing Neuropixels recordings from 279 brain regions across 139 mice. With over 600,000 neurons, it serves as the primary substrate for training Neuro-Foundation Models like POYO [Azabou *et al.*, 2023] and NEDS [Zhang *et al.*, 2025]. Similarly, the *Allen Brain Observatory* offers extensive datasets combining large-scale Neuropixels and calcium imaging recordings from multiple visual areas in head-fixed behaving mice. While their *Neuropixels* dataset is the standard for benchmarking multi-region communication [Siegle *et al.*, 2021], the *2-Photon* dataset facilitates geometric alignment testing due to its massive population count of over 63,000 neurons [de Vries *et al.*, 2020].

Task-Specific Repositories

The second category is Task-Specific Repositories. These platforms, such as the *CRCNS* repository, play a crucial role by aggregating datasets contributed by individual laboratories using different recording modalities, animal models, and behavioral tasks. Unlike foundation corpora, these datasets often focus on specific dynamical questions. For instance, the

M1/PMd Reach (pmd-1) dataset captures neural activity from the motor and premotor cortices during multi-target reaching tasks, providing a robust testbed for continuous dynamical systems modeling (e.g., LFADS). Taking a different perspective, the Hippocampus (hc-11) dataset records activity during spatial maze navigation; its distinct place-cell activity patterns make it ideal for testing discrete state-switching models (e.g., rSLDS) that track rapid transitions in cognitive representation. These datasets also have limitations: the prevalence of oversimplified tasks in head-fixed animals likely fails to capture the complexity of neural dynamics during naturalistic, freely moving behavior.

7 Discussion and Open Challenges

Below, we synthesize fundamental gaps and promising avenues for future development.

Identifiability and Interpretability. A central challenge is *identifiability*: guaranteeing that learned parameters match true biological mechanisms rather than merely fitting the data. Deep models often suffer from rotational ambiguity, where latent axes fail to map onto distinct cognitive processes. Constrained approaches such as Identifiable VAEs (iVAEs) [Khemakhem *et al.*, 2020] use auxiliary variables to make learned factors biologically unique.

From Correlation to Causal Circuits. Current methods are fundamentally observational, struggling to separate causal influence from unobserved “common drive” or to predict responses to perturbations. The field must advance toward Causal Representation Learning via in-silico “Digital Twins” [Kapoor *et al.*, 2024] that test counterfactual hypotheses—e.g., the outcome of optogenetic interventions—before physical experimentation.

Generalization and the “Universal Neural Grammar”. Neuro-Foundation Models face idiosyncratic neural codes that lack a standardized cross-subject vocabulary, preventing zero-shot transfer. The solution lies in In-Context Learning and geometric alignment [Schneider *et al.*, 2023]: treating neural dynamics as a relative manifold rather than absolute firing rates, so models calibrate to new subjects dynamically.

Computational Efficiency. Real-time brain-computer interfaces are hindered by the cost of Transformer and probabilistic models. Knowledge Distillation into lightweight “Student” networks, e.g. Spiking Neural Networks, enables efficient inference on low-power hardware.

8 Conclusion

In this paper, we have presented a comprehensive taxonomy of LVMs in neuroscience based on the interdependencies between three pillars: Single-Region Latent Dynamics, Multi-Region Communication, and Behavior-Aligned Modeling. The unified view we adopt in this survey relates hitherto unconnected strands of research, tracing the evolution from static smoothing to nonlinear dynamical systems and distributed information routing. Finally, we also outline open challenges regarding causality, interpretability, and geometry that must be addressed to successfully reverse-engineer the algorithms of biological intelligence.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (Grant No. 62506090); the National Key R&D Program of China (Grant No. 2025YFF0523900); an AI2050 Senior Fellowship, a Schmidt Sciences program; the National Science Foundation (NSF); the National Institute of Food and Agriculture (USDA/NIFA; 2023-67021-39829); and the Air Force Office of Scientific Research (AFOSR; FA9550-23-1-0322).

Contribution Statement

Fumen Deng and Xinyi Dong contributed equally to this work.

References

- [Abbaspourazad *et al.*, 2021] Hamidreza Abbaspourazad, Mahdi Choudhury, Yan T Wong, et al. Multiscale low-dimensional motor cortical state dynamics predict naturalistic reach-and-grasp behavior. *Nat. Commun.*, 12(1):607, 2021.
- [Angelaki *et al.*, 2025] Dora Angelaki, Brandon Benson, Julius Benson, Daniel Birman, et al. A brain-wide map of neural activity during complex behaviour. *Nature*, 645(8079):177–191, 2025.
- [Azabou *et al.*, 2023] Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, et al. A unified, scalable framework for neural population decoding. *NeurIPS*, 36:44937–44956, 2023.
- [Cho *et al.*, 2023] Cheol Jun Cho, Edward Chang, and Gopala Anumanchipalli. Neural latent aligner: cross-trial alignment for learning representations of complex, naturalistic neural data. In *ICML*, pages 5661–5676, 2023.
- [Chou *et al.*, 2024] Chi-Ning Chou, Royoung Kim, Luke A Arend, Yao-Yuan Yang, et al. Geometry linked to untangling efficiency reveals structure and computation in neural populations. *bioRxiv*, 2024.
- [de Vries *et al.*, 2020] Saskia EJ de Vries, Jerome A Lecoq, Michael A Buice, Peter A Groblewski, Gabriel K Ocker, Michael Oliver, David Feng, Nicholas Cain, Peter Ledochowitsch, Daniel Millman, et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature neuroscience*, 23(1):138–151, 2020.
- [DePasquale *et al.*, 2018] Brian DePasquale, Christopher J Cueva, Kanaka Rajan, G Sean Escola, et al. full-force: A target-based method for training recurrent networks. *PloS one*, 13(2):e0191527, 2018.
- [Dowling and Savin, 2025] Matthew Dowling and Cristina Savin. Nonlinear multiregion neural dynamics with parametric impulse response communication channels. In *ICLR*, 2025.
- [Gao *et al.*, 2017] Peiran Gao, Eric Trautmann, Byron Yu, et al. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv*, page 214262, 2017.
- [Ge *et al.*, 2026] Ningling Ge, Sicheng Dai, Yu Zhu, and Shan Yu. Energy-based autoregressive generation for neural population dynamics. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(1):309–317, Mar. 2026.
- [Glaser *et al.*, 2020] Joshua Glaser, Matthew Whiteway, John P Cunningham, Liam Paninski, et al. Recurrent switching dynamical systems models for multiple interacting neural populations. *NeurIPS*, 33:14867–14878, 2020.
- [Gokcen *et al.*, 2022] Evren Gokcen, Anna I Jasper, João D Smedo, Amin Zandvakili, et al. Disentangling the flow of signals between populations of neurons. *Nat. Comput. Sci.*, 2(8):512–525, 2022.
- [Gokcen *et al.*, 2023] Evren Gokcen, Anna Jasper, Alison Xu, Adam Kohn, et al. Uncovering motifs of concurrent signaling across multiple neuronal populations. *NeurIPS*, 36:34711–34722, 2023.
- [Gonzalez Laiz *et al.*, 2025] Rodrigo Gonzalez Laiz, Tobias Schmidt, and Steffen Schneider. Self-supervised contrastive learning performs non-linear system identification. In *International Conference on Learning Representations*, volume 2025, pages 62224–62261, 2025.
- [Gosztolai *et al.*, 2025] Adam Gosztolai, Robert L Peach, Alexis Arnaudon, Mauricio Barahona, et al. Marble: interpretable representations of neural population dynamics using geometric deep learning. *Nat. Methods*, pages 1–9, 2025.
- [Gozel and Doiron, 2024] Olivia Gozel and Brent Doiron. Between-area communication through the lens of within-area neuronal dynamics. *Sci. Adv.*, 10(42):eadl6120, 2024.
- [Hirayama *et al.*, 2017] Jun-ichiro Hirayama, Aapo Hyvärinen, and Motoaki Kawanabe. Splice: fully tractable hierarchical extension of ica with pooling. In *International Conference on Machine Learning*, pages 1491–1500. PMLR, 2017.
- [Hu *et al.*, 2024] Amber Hu, David Zoltowski, Aditya Nair, David Anderson, et al. Modeling latent neural dynamics with gaussian process switching linear dynamical systems. *NeurIPS*, 37:33805–33835, 2024.
- [Jha *et al.*, 2024] Aditi Jha, Diksha Gupta, Carlos Brody, and Jonathan W Pillow. Disentangling the roles of distinct cell classes with cell-type dynamical systems. *NeurIPS*, 37:33668–33690, 2024.
- [Jude *et al.*, 2022] Justin Jude, Matthew G Perich, Lee E Miller, and Matthias H Hennig. Robust alignment of cross-session recordings of neural population activity by behaviour via unsupervised domain adaptation. *arXiv*, 2022.
- [Kapoor *et al.*, 2024] Jaivardhan Kapoor, Auguste Schulz, Julius Vetter, Felix Pei, et al. Latent diffusion for neural spiking data. *NeurIPS*, 37:118119–118154, 2024.
- [Karniol-Tambour *et al.*, 2024] Orren Karniol-Tambour, David M Zoltowski, E Mika Diamanti, Lucas Pinto, et al. Modeling state-dependent communication between brain regions with switching nonlinear dynamical systems. In *ICLR*, 2024.
- [Keeley *et al.*, 2020] Stephen Keeley, Mikio Aoi, Yiyi Yu, Spencer Smith, et al. Identifying signal and noise structure in neural population activity with gaussian process factor models. *NeurIPS*, 33:13795–13805, 2020.
- [Keshtkaran *et al.*, 2022] Mohammad Reza Keshtkaran, Andrew R Sedler, Raed H Chowdhury, Raghav Tandon, et al. A large-scale neural network training framework for generalized estimation of single-trial population dynamics. *Nat. Methods*, 19(12):1572–1577, 2022.
- [Khemakhem *et al.*, 2020] Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvärinen. Variational autoencoders and nonlinear ica: A unifying framework. In *AISTATS*, pages 2207–2217, 2020.
- [Kim *et al.*, 2021] Timothy D Kim, Thomas Z Luo, Jonathan W Pillow, and Carlos D Brody. Inferring latent dynamics underlying neural population activity via neural differential equations. In *ICML*, pages 5551–5561, 2021.
- [Kim *et al.*, 2023] Timothy Doyeon Kim, Thomas Zhihao Luo, Tankut Can, Kamesh Krishnamurthy, et al. Flow-field inference from neural data using deep recurrent networks. *bioRxiv*, 2023.
- [Li *et al.*, 2024] Weihai Li, Chengrui Li, Yule Wang, and Anqi Wu. Multi-region markovian gaussian process: An efficient method to discover directional communications across multiple brain regions. *PMLR*, 235:28112, 2024.
- [Li *et al.*, 2025] Weihai Li, Yule Wang, Chengrui Li, and Anqi Wu. Learning time-varying multi-region brain communications via scalable markovian gaussian processes. In *ICML*, 2025.

- [Linderman *et al.*, 2017] Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, et al. Bayesian learning and inference in recurrent switching linear dynamical systems. In *AISTATS*, pages 914–922, 2017.
- [Liu *et al.*, 2025] Belle Liu, Jacob Sacks, and Matthew D. Golub. Accurate identification of communication between multiple interacting neural populations. In *ICML*, 2025.
- [MacDowell *et al.*, 2025] Camden J MacDowell, Alexandra Libby, Caroline I Jahn, Sina Tafazoli, et al. Multiplexed subspaces route neural activity across brain-wide networks. *Nat. Commun.*, 16(1):3359, 2025.
- [Macke *et al.*, 2011] Jakob H Macke, Lars Buesing, John P Cunningham, Byron M Yu, et al. Empirical models of spiking in neural populations. In *NeurIPS*, volume 24, 2011.
- [McCart *et al.*, 2024] Jonathan D McCart, Andrew R Sedler, Christopher Versteeg, Domenick Mifsud, Mattia Rigotti-Thompson, and Chethan Pandarinath. Diffusion-based generation of neural activity from disentangled latent codes. *ArXiv*, 2024.
- [Mudrik *et al.*, 2025] Noga Mudrik, Ryan Ly, Oliver Ruebel, and Adam Shabti Charles. CREIMBO: Cross-regional ensemble interactions in multi-view brain observations. In *ICLR*, 2025.
- [Pals *et al.*, 2024] Matthijs Pals, A Erdem Sağtekin, Felix Pei, Manuel Gloeckler, et al. Inferring stochastic low-rank recurrent neural networks from neural data. *NeurIPS*, 37:18225–18264, 2024.
- [Pandarinath *et al.*, 2018] Chethan Pandarinath, Daniel J O’Shea, Jasmine Collins, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods*, 15(10):805–815, 2018.
- [Pei *et al.*, 2021] Felix Pei, Joel Ye, David Zoltowski, Anqi Wu, et al. Neural latents benchmark’21: evaluating latent variable models of neural population activity. *arXiv*, 2021.
- [Perich and Rajan, 2020] Matthew G Perich and Kanaka Rajan. Rethinking brain-wide interactions through multi-region ‘network of networks’ models. *Curr. Opin. Neurobiol.*, 65:146–151, 2020.
- [Perich *et al.*, 2025] Matthew G Perich, Devika Narain, and Juan A Gallego. A neural manifold view of the brain. *Nat. Neurosci.*, 28(8):1582–1597, 2025.
- [Ryoo *et al.*, 2026] Avery Hee-Woon Ryoo, Nanda H Krishna, Ximeng Mao, Mehdi Azabou, Eva L Dyer, Matthew G Perich, and Guillaume Lajoie. Generalizable, real-time neural decoding with hybrid state-space models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.
- [Sani *et al.*, 2021] Omid G Sani, Hamidreza Abbaspourazad, Yan T Wong, Bijan Pesaran, et al. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nat. Neurosci.*, 24(1):140–149, 2021.
- [Sani *et al.*, 2024] Omid G Sani, Bijan Pesaran, and Maryam M Shanechi. Dissociative and prioritized modeling of behaviorally relevant neural dynamics using recurrent neural networks. *Nat. Neurosci.*, 27(10):2033–2045, 2024.
- [Schneider *et al.*, 2023] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, 2023.
- [Schulz *et al.*, 2025] Auguste Schulz, Julius Vetter, Richard Gao, Daniel Morales, et al. Modeling conditional distributions of neural and behavioral data with masked variational autoencoders. *Cell Rep.*, 44(3), 2025.
- [Semedo *et al.*, 2022] João D Semedo, Anna I Jasper, Amin Zandvakili, Aravind Krishna, et al. Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat. Commun.*, 13(1):1099, 2022.
- [Siegle *et al.*, 2021] Joshua H Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, 2021.
- [Song *et al.*, 2026] Yue Song, T. Anderson Keller, Yisong Yue, Pietro Perona, and Max Welling. Langevin flows for modeling neural latent dynamics, 2026.
- [Soo *et al.*, 2023] Wayne Soo, Vishwa Goudar, and Xiao-Jing Wang. Training biologically plausible recurrent neural networks on cognitive tasks with long-term dependencies. *NeurIPS*, 36:32061–32074, 2023.
- [Urai *et al.*, 2022] Anne E Urai, Brent Doiron, Andrew M Leifer, and Anne K Churchland. Large-scale neural recordings call for new insights to link brain and behavior. *Nat. Neurosci.*, 25(1):11–19, 2022.
- [Vahidi *et al.*, 2024] Parsa Vahidi, Omid G Sani, and Maryam M Shanechi. Modeling and dissociation of intrinsic and input-driven neural population dynamics underlying behavior. *PNAS*, 121(7):e2212887121, 2024.
- [Vahidi *et al.*, 2025] Parsa Vahidi, Omid G Sani, and Maryam M Shanechi. Braid: Input-driven nonlinear dynamical modeling of neural-behavioral data. *arXiv*, 2025.
- [Valente *et al.*, 2022] Adrian Valente, Jonathan W Pillow, and Srdjan Ostojic. Extracting computational mechanisms from neural data using low-rank rnns. *NeurIPS*, 35:24072–24086, 2022.
- [Versteeg *et al.*, 2023] Christopher Versteeg, Andrew R Sedler, Jonathan D McCart, and Chethan Pandarinath. Expressive dynamics models with nonlinear injective readouts enable reliable recovery of latent features from neural activity. *arXiv*, 2023.
- [Wang *et al.*, 2024] Yule Wang, Chengrui Li, Weihai Li, and Anqi Wu. Exploring behavior-relevant and disentangled neural dynamics with generative diffusion models. *NeurIPS*, 37:34712–34736, 2024.
- [Wu *et al.*, 2017] Anqi Wu, Nicholas A Roy, Stephen Keeley, and Jonathan W Pillow. Gaussian process based nonlinear latent structure discovery in multivariate spike train data. *Advances in neural information processing systems*, 30, 2017.
- [Xin and Kass, 2025] Qi Xin and Robert E Kass. Identifying interactions across brain areas while accounting for individual-neuron dynamics with a transformer-based variational autoencoder. *arXiv*, 2025.
- [Ye and Pandarinath, 2021] Joel Ye and Chethan Pandarinath. Representation learning for neural population activity with neural data transformers. *arXiv*, 2021.
- [Yu *et al.*, 2008] Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen Ryu, et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *NeurIPS*, volume 21, 2008.
- [Zhang and Saxena, 2024] Yongxu Zhang and Shreya Saxena. Inference of neural dynamics using switching recurrent neural networks. *NeurIPS*, 37:131456–131481, 2024.
- [Zhang *et al.*, 2024] Yizi Zhang, Yanchen Wang, Donato Jiménez-Benetó, Zixuan Wang, et al. Towards a “universal translator” for neural dynamics at single-cell, single-spike resolution. *NeurIPS*, 37:80495–80521, 2024.
- [Zhang *et al.*, 2025] Yizi Zhang, Yanchen Wang, Mehdi Azabou, Alexandre Andre, et al. Neural encoding and decoding at scale. *arXiv*, 2025.
- [Zhao and Park, 2017] Yuan Zhao and Il Memming Park. Variational latent gaussian process for recovering single-trial dynamics from population spike trains. *Neural Comput.*, 29(5):1293–1316, 2017.