Exploring Topological Properties in Artificial Neural Networks

Sophia Pi
Department of Computer Science, Northwestern University, Evanston, IL, USA

Background

- Neural systems often develop low-dimensional representations that capture the intrinsic geometry of the data.
- In biology, examples such as rodent head direction cells form circular manifolds, capturing the topology of spatial variables.
- In deep learning, internal activations tend to condense into lower-dimensional embeddings, although their topological nature is not yet fully understood.
- The goal is to explore whether ANNs trained on tasks with explicit topological structures (planar, spherical, cyclic) reveal similar geometric representations.
- Bridging insights from neuroscience and machine learning could improve model interpretability and inspire new architecture designs.

Prior Work

- Neuroscientific studies have identified manifold-like activity patterns in systems such as rodent head direction cells (Peyrache et al., 2015; Chaudhuri et al., 2019), motor cortex dynamics (Elsayed et al., 2016), and hippocampal representations (Bernardi et al., 2020).
- Previous deep learning research shows that representations in CNNs and transformers tend to become low-dimensional (Ansuini et al., 2019; Cohen et al., 2020), though these studies did not explicitly focus on aligning such representations with the intrinsic topology of the data.
- Some exploratory work in language models and molecular embeddings has hinted at structured internal representations.
- There remains a significant gap in testing whether ANNs can naturally develop embeddings that mirror the inherent topology of training data.

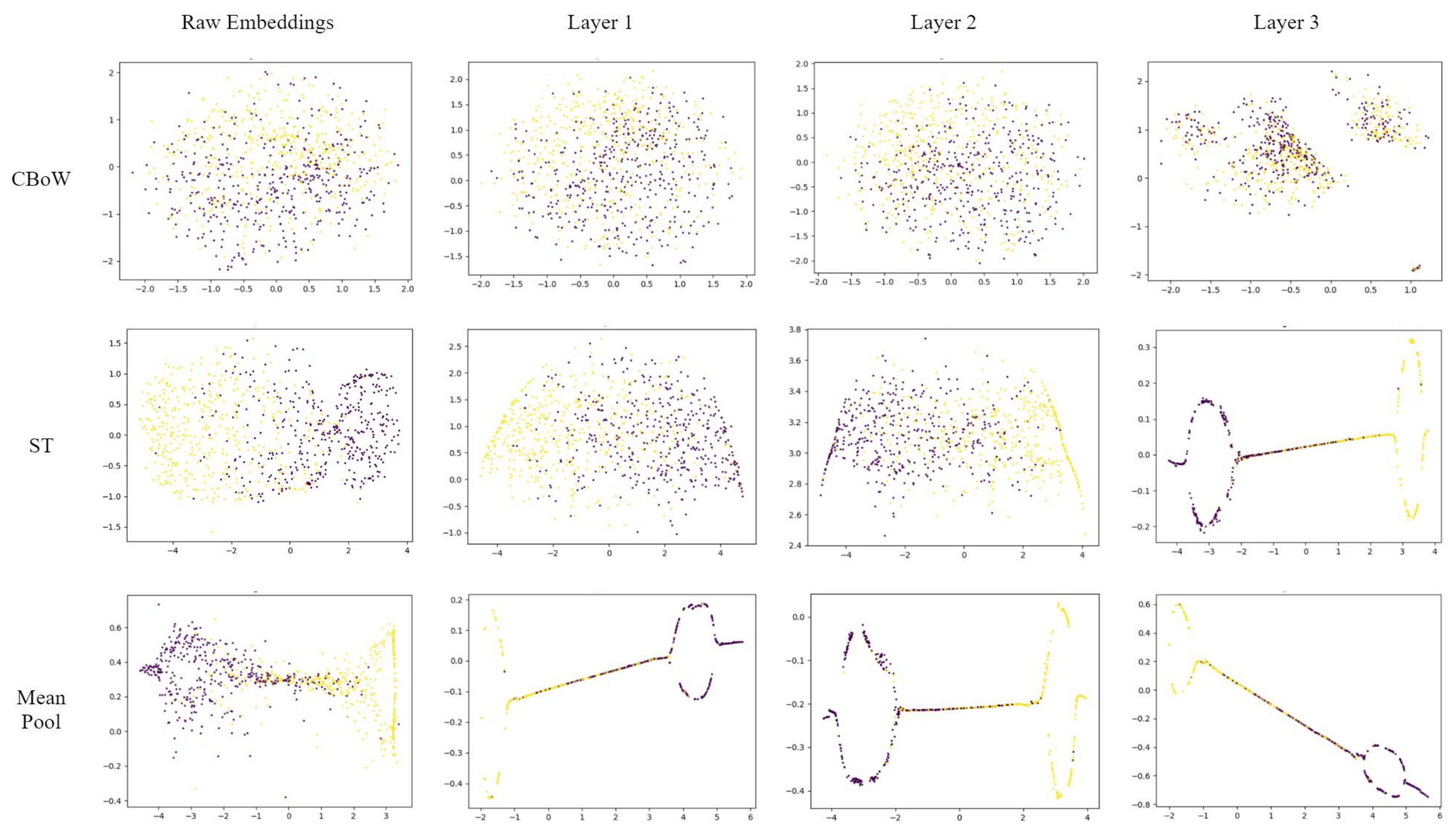
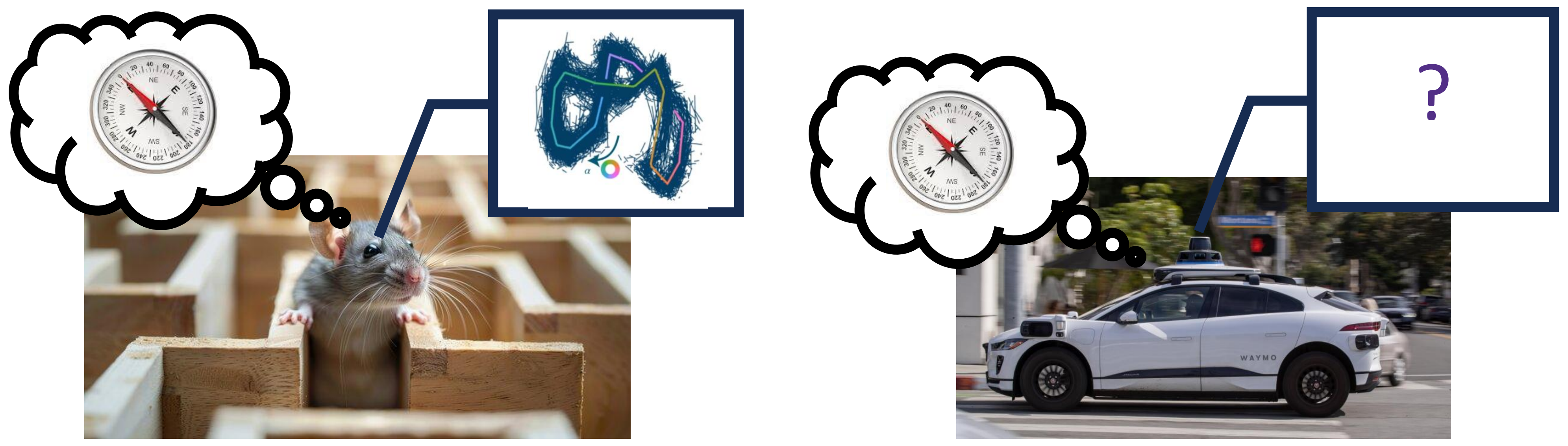


Fig. 1: Preliminary analyses show that simple language models can progressively learn semantic properties of textual queries

Do artificial neural networks emulate the **latent topological structures** found in biological neural networks?



[image taken from Chaudhuri et. Al. 2019]

Proposed Methodology

Datasets:
Tasks with clearly defined topologies (e.g., planar/spherical coordinates in Apolloscape, toroidal temporal variables, graph-structured road networks).

Model Architectures:
Graph neural networks (GNNs) for road network graphs.
3D models (or point-based models) for LIDAR point clouds.
Use recurrent or transformer-based sequence models for cyclic and/or temporal data.

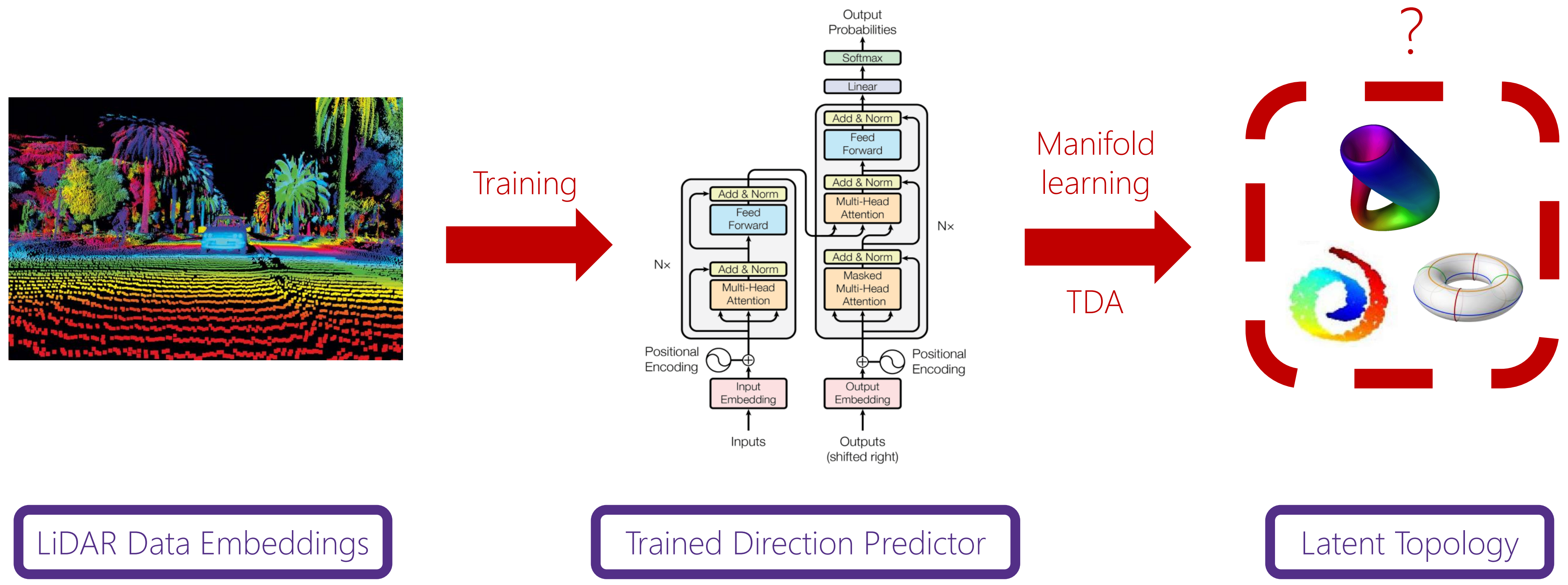
Analysis Techniques:
Extract intermediate activations and weights from key layers (point cloud encoder, temporal module, prediction head).
Apply dimensionality reduction techniques (PCA, t-SNE, UMAP) to visualize the structure of these embeddings.
Compute persistent homology, Betti numbers, etc. to quantify the topological features of the representations.

Architecture Plan for LiDAR Data (BEV Model):

Point Cloud Encoder: Use PointNet++ or PointPillars-style networks to extract spatial features from raw LiDAR data.

Temporal Module: Integrate an LSTM/GRU or Transformer layer to track movement and develop a sense of direction over time.

Direction Prediction Head: A classification/regression module to predict the agent's current heading.



LiDAR Data Embeddings

Trained Direction Predictor

Latent Topology

Expected Results

Positive Case:
- Internal representations align with the task's topology (e.g., circular manifolds for cyclic tasks or spherical surfaces for spatial tasks).
- Visualization methods reveal clear low-dimensional structures—such as rings or surfaces—that correspond to the expected topology.
- Topological metrics (e.g., persistent homology, Betti numbers) show significant features matching theoretical predictions.

Negative Case:
- The network fails to develop distinct topological structures; embeddings appear diffuse or lack clear geometric patterns.
- Dimensionality reduction methods do not reveal meaningful low-dimensional manifolds that correspond to the inherent topology of the data.
- TDA metrics do not show statistically significant topological invariants, suggesting that the current training regimes may not encourage such representations.

Discussion

Implications for Neuroscience & AI:
- If ANNs exhibit topologically meaningful internal representations, it supports the notion of shared computational principles between biological and artificial systems.
- Positive results could enhance model interpretability by linking network activations to well-understood topological constructs.

Challenges and Considerations:
- A lack of clear topological structure might indicate that standard training paradigms are insufficient for inducing such representations.
- Results could motivate the exploration of new training strategies (e.g., topology-based regularizers) to guide the learning process.

Findings may offer insights into why certain architectures are particularly effective for tasks involving spatial or cyclic data. Could suggest design principles in neuroscience-inspired models and more explainable AI systems.

References

Ansuini, A.; Laio, A.; Macke, J. H.; Zoccolan, D.; and Stella, L. 2019. Intrinsic dimension of data representations in deep neural networks. Advances in Neural Information Processing Systems, 32: 6109–6119.
Bernardi, S.; Benna, M. K.; Rigotti, M.; Munuera, J.; Fusi,S.; and Salzman, C. D. 2020. The geometry of abstraction in the hippocampus and prefrontal cortex. Cell, 183(4): 954–967.e21.
Chaudhuri, R.; Gerc¸ek, B.; Pandey, B.; Peyrache, A.; and Fi-ete, I. 2019. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nature Neuroscience, 22(9): 1512–1520.
Elsayed, G. F.; Lara, A. H.; Kaufman, M. T.; Churchland,M. M.; and Cunningham, J. P. 2016. Reorganization be-tween preparatory and movement population responses in motor cortex. Nature Communications, 7: 13239.
Peyrache, A.; Lacroix, M. M.; Petersen, P. C.; and Buzs´aki,G. 2015. Internally organized mechanisms of the head di-rection sense. Nature Neuroscience, 18(4): 569–575.

I would like to express my sincere gratitude to my men-tor, Jeshwanth Challagundla, for his invaluable guidance and support throughout this project, and to Dr. Jason Grant, Chair of the AAAI-UC, for his leadership and dedication to fostering undergraduate research.

sophiapi2026.1@u.northwestern.edu