Ziyuan Yin

Human communication often fails not because information is absent, but because meaning is inadequately represented. As I grew older, I began to notice the real consequences of these mismatches: when surface expressions diverge from the structures they are intended to convey, even well-intentioned exchanges can fracture, producing misunderstandings, distorted decisions, and widening cognitive gaps between individuals and groups. Encountering Wittgenstein later gave intellectual structure to what I had long sensed that language is fundamentally limited, and meaning does not always survive its translation into words. This recognition sparked my earliest, somewhat idealistic ambition: if we could understand how meaning is represented in the brain, perhaps we could build systems that reduce miscommunication and allow people to transmit their thoughts more faithfully. Over time, this ambition matured into the scientific question that has guided every step of my research since: how does the brain construct, stabilize, and transform meaning?

In developing this line of inquiry, I was first fortunate to work in a computational neuroscience laboratory at the City University of Hong Kong. Through modeling swallowing acoustics and clinical signals, I began to see that studying meaning ultimately requires studying representation, specifically how information is structured, encoded, and transformed within a system. Signals that initially appeared chaotic at the surface revealed stable internal structure when represented appropriately. Learning to extract information from noisy, nonstationary data was not merely technical training; it established my first systematic understanding of how complex systems encode latent structure.

My interest in representational structure deepened during a startup project on LLM-based generative agents, where I implemented memory modules and retrieval mechanisms to support long-range narrative consistency. By incorporating temporal weighting into memory embeddings, I enabled agents to retrieve past experiences based not only on semantic similarity but also on contextual relevance over time. While this approach improved retrieval behavior and temporal coherence, it also exposed a fundamental limitation: although these models exhibit strong contextual disambiguation, their semantic representations remain highly context-dependent and tend to drift across time, lacking the persistent abstraction characteristic of human conceptual organization. Confronting this gap shifted my focus from performance optimization to a deeper question of representation, prompting me to ask how conceptual meaning is structured computationally and why artificial systems struggle to maintain stable semantics across contexts.

These questions followed me to UCSF, where my work expanded into large-scale neurophysiology. While modeling long-term EEG from comatose patients and ECoG during spreading depolarizations, I encountered neural signals whose statistical properties vary substantially across time and physiological state, often obscuring the structure I aimed to characterize. To address this, I adopted representation-learning approaches that map highly heterogeneous EEG into lower-dimensional latent spaces, using autoencoder-based models to extract stable structure from nonstationary signals. By projecting long-term EEG into such latent manifolds, it became possible to characterize brain states and their temporal trajectories in a way that was both physiologically meaningful and robust to surface-level variability. Through this process, I learned that extracting interpretable structure from neural data depends critically on representational choices, and that appropriate low-dimensional embeddings are essential for capturing the dynamics by which neural systems organize, transform, and encode internal states. This methodological perspective later converged with my longstanding interest in meaning when I joined a project examining auditory and linguistic responses in the cortex of comatose patients. For the first time, I was able to approach semantic structure at the neural level as a representational problem grounded in measurable neural dynamics rather than an abstract theoretical construct.

Taken together, these experiences gradually sharpened my focus toward a single central question that I hope to pursue in my PhD: how the brain flexibly reshapes the representation of a concept across different contexts while preserving the stability of its hierarchical semantic structure. I aim to study how a concept is projected into different semantic subspaces depending on context, whether these projections follow predictable cortical dynamics, and whether they parallel the deeper semantic and reasoning dimensions revealed in large language models through residual disentanglement.

Addressing this question requires studying language representations across multiple temporal and representational scales, and explicitly linking neural activity to structured computational models. My goal is to design experimentally grounded frameworks that make contextual modulation of semantic representations explicitly testable, by projecting neural activity into low-dimensional representational spaces and quantitatively comparing context-dependent shifts with disentangled semantic embeddings from large language models across time and cortical regions. The Neural Acoustic Processing Lab provides a strong methodological foundation for this approach. In particular, Professor Mesgarani's work on hierarchical neural encoding offers concrete tools for tracing how representations evolve from acoustic input to higher-level linguistic structure across cortical regions and time scales. His recent use of deep neural networks and residual disentanglement further enables systematic comparisons between neural representations and model-derived features, which directly supports the type of analysis I aim to pursue.

If given the opportunity to pursue doctoral training in this environment, I hope to deepen my investigation of how semantic representations are structured, stabilized, and flexibly transformed in the brain by building on my prior work in representational learning across clinical neurophysiology, generative models, and language-related neural data. Across my training, I have repeatedly encountered the same constraint: questions about neural semantics only become tractable when neural measurements and computational representations are analyzed within a shared representational framework. An environment that treats this coupling not as an auxiliary technique but as a central object of inquiry is therefore essential for the kind of research I aim to pursue.