Research Statement

Yingchen Xu | yingchenxu20@gmail.com | ycxuyingchen.github.io

Research Vision

My research seeks to understand how learning systems can acquire *internal models of the world* that support reasoning, planning, and generalization from limited experience. Humans do this naturally, whereas many modern learning systems—despite impressive gains from large-scale optimization—remain fragile when observations are partial, interaction is expensive, or task objectives change. These internal models, often referred to as *world models*, are therefore a natural focus for understanding and closing the gap between current learning systems and robust decision-making. My completed work responds to this challenge by focusing on the practical bottlenecks that arise when world models are used in real-world decision-making, including data collection without rewards, long-horizon reasoning, and high-dimensional control.

Many of my contributions are grounded in embodied control, where information is inherently limited, costly to acquire, and tightly coupled to action. These conditions expose not only practical challenges, but also suggest that the way world models are learned under *information constraints* plays a central role in determining their ability to support robust reasoning and generalization—an insight that motivates the next phase of my research.

Completed Research

My completed research addresses a sequence of concrete bottlenecks that arise when applying world models to planning and control in realistic settings.

1) Scalable data collection for reward-free world modeling. World models only generalize to the extent that their training experience is diverse. In *CASCADE* [1], I addressed the bottleneck of collecting broad experience without engineered rewards by training a *population* of explorers guided by information-theoretic objectives. Complementary behaviors maintain coverage and prevent collapse into narrow datasets, providing a scalable recipe for reward-free trajectory collection that supports downstream transfer.

2) Temporal abstraction for long-horizon planning. Long-horizon tasks require reasoning over multiple time scales; planning directly over primitive actions is brittle and computationally expensive. In *IQL-TD-MPC* [2], I developed an offline hierarchical model-based framework in which a high-level manager plans in a temporally abstract latent space, while a low-level policy executes short-horizon control. This decomposition makes planning tractable by pushing search into abstractions learned for decision-making rather than mere compression.

3) Generative priors for high-dimensional control. High-dimensional control, such as humanoid locomotion, requires plans that are both physically feasible and rapidly adaptable. In *H-GAP* [3], I bridged generative modeling and control by training a trajectory-diffusion model to serve as a motion prior within model predictive control, enabling zero-shot adaptation by sampling candidate futures that satisfy new objectives. In follow-on work [4, 5], we generalized this paradigm into a reusable control substrate that enables robust performance under perturbations without retraining.

4) Latent dynamics for reasoning beyond control. More recently, I investigated whether similar principles apply beyond control. In *TokenAssorted* [6], we hypothesized that reasoning in language models corresponds to trajectories through a latent semantic state space, analogous to planning in control. We found that standard transformers often collapse these intermediate dynamics, limiting generalization. Introducing explicit latent variables improved both performance and interpretability, reinforcing a broader lesson from my control work: representations optimized solely for prediction are insufficient for robust inference without appropriate structural constraints.

Future Research: World Models Under Information Constraints

My future research asks how world models should be learned when information is limited, costly, or biased, as is typical in realistic decision-making settings. While my completed work shows that diverse data, temporal abstraction, and generative priors can make world models usable in practice, these systems often rely on training conditions that provide more information than is available at test time. As a result, models may perform well by exploiting perceptual or task-specific shortcuts that do not transfer when objectives change or observations are partial.

I hypothesize that learning under *information constraints* is a key driver of generalization and sample efficiency. When models cannot rely on full observability or exhaustive reconstruction, they are forced to learn compact representations that capture task-relevant structure rather than surface detail. This perspective is complementary to scaling: rather than replacing large models, information constraints shape what large models are pressured to represent as capacity increases.

A central direction of my future work studies **active world models**, in which perception is treated as a decision rather than a passive input. Instead of training models on full sensory streams, I will investigate settings where agents must decide what information to acquire to support downstream reasoning and planning. While vision provides a natural testbed, the underlying question is general: how selective observation influences the representations learned by a world model.

Learning under information constraints naturally leads to the problem of **belief state learning in continual settings**. When observations are partial and future queries are unknown, effective decision making requires maintaining a persistent belief that integrates evidence over time and tracks uncertainty over unobserved variables. I plan to study world models whose internal states explicitly represent such uncertainty, and to evaluate whether these belief states support more stable planning, adaptation to new objectives, and principled exploration.

A final direction connects these ideas to **decision making under off-policy and biased data**. Realistic learning regimes often involve logged interaction, human demonstrations, or mixtures of policies. Building on my prior work in reward-free exploration and offline planning, I aim to study how world models learned under information constraints can better reuse such data for planning and control, measuring success by robustness and reuse rather than exhaustive prediction.

Overall, this agenda does not aim to outperform large generative models at full reconstruction. Instead, it asks what representations are required to support counterfactual, task-conditioned reasoning under realistic constraints. By treating observation as an active process and belief as a first-class object, my goal is to develop world models that generalize by learning transferable structure, rather than by consuming ever larger amounts of data.

References

[1] Yingchen Xu, Jack Parker-Holder, Aldo Pacchiano, Philip J. Ball, Oleh Rybkin, Stephen J. Roberts, Tim Rocktäschel, and Edward Grefenstette. *Learning General World Models in a Handful of Reward-Free Deployments*. 2022. arXiv: 2210.12719 [cs.LG]. URL: https://arxiv.org/abs/2210.12719.

[2] Yingchen Xu, Rohan Chitnis, Bobak Hashemi, Lucas Lehnert, Urun Dogan, Zheqing Zhu, and Olivier Delalleau. *IQL-TD-MPC: Implicit Q-Learning for Hierarchical Model Predictive Control*. 2023. arXiv: 2306.00867 [cs.LG]. URL: https://arxiv.org/abs/2306.00867.

[3] Zhengyao Jiang, Yingchen Xu, Nolan Wagener, Yicheng Luo, Michael Janner, Edward Grefenstette, Tim Rocktäschel, and Yuandong Tian. *H-GAP: Humanoid Control with a Generalist Planner*. 2023. arXiv: 2312.02682 [cs.LG]. URL: https://arxiv.org/abs/2312.02682.

[4] Andrea Tirinzoni, Ahmed Touati, Jesse Farebrother, Mateusz Guzek, Anssi Kanervisto, Yingchen Xu, Alessandro Lazaric, and Matteo Pirotta. *Zero-Shot Whole-Body Humanoid Control via Behavioral Foundation Models*. 2025. arXiv: 2504.11054 [cs.LG]. URL: https://arxiv.org/abs/2504.11054.

[5] Harshit Sikchi, Andrea Tirinzoni, Ahmed Touati, Yingchen Xu, Anssi Kanervisto, Scott Niekum, Amy Zhang, Alessandro Lazaric, and Matteo Pirotta. *Fast Adaptation with Behavioral Foundation Models*. 2025. arXiv: 2504.07896 [cs.LG]. URL: https://arxiv.org/abs/2504.07896.

[6] DiJia Su, Hanlin Zhu, Yingchen Xu, Jiantao Jiao, Yuandong Tian, and Qinqing Zheng. *Token Assorted: Mixing Latent and Text Tokens for Improved Language Model Reasoning*. 2025. arXiv: 2502.03275 [cs.CL]. URL: https://arxiv.org/abs/2502.03275.