Kaiwen Hu

https://kaotty.github.io kaiwen_hu@berkeley.edu

Education

School of EECS, UC Berkeley Berkeley, U.S.
PhD in EECS, advised by Prof. Somayeh Sojoudi *Aug. 2025 - May 2030 (Expected)*

School of EECS, Peking University Beijing, China
BS in Computer Science, advised by Prof. Yisen Wang *Sep. 2021 - July 2025*

Research Interests

- Multimodal Reasoning and Tool Use: Understanding how SFT shapes RL dynamics to improve reasoning and reliable tool use in multimodal language models.
- Reinforcement Learning: Developing scalable RL algorithms on long-horizon tasks.
- On-policy Distillation: Exploring how to enable decision-level knowledge transfer and selective supervision for long-horizon reasoning.

Publications and Preprints

* indicates equal contributions

- **PRIOR: Pre-Call Reasoning in SFT and Its Influence on Tool-Use Retention under RL**
Kaiwen Hu, Hyunin Lee, Joseph Suh, Serina Chang, Shuai Shao, Yong Zhang, Namyong Park, Xiaoqiang Luo, Luke Simon, Somayeh Sojoudi
Under review.
- **Value Gradient Flow: Behavior-Regularized RL without Regularization**
Haoran Xu*, **Kaiwen Hu***, Somayeh Sojoudi, Amy Zhang
The 14th International Conference on Learning Representations (**ICLR**), 2026.
- **Projection Head is Secretly an Information Bottleneck**
Zhuo Ouyang*, **Kaiwen Hu***, Qi Zhang, Yifei Wang, Yisen Wang
The 13th International Conference on Learning Representations (**ICLR**), 2025.
- **Understanding the Role of Equivariance in Self-Supervised Learning**
Yifei Wang*, **Kaiwen Hu***, Sharut Gupta, Ziyu Ye, Yisen Wang, Stefanie Jegelka
The 38th Annual Conference on Neural Information Processing Systems (**NeurIPS**), 2024.

Research Experiences

Equivariant Self-supervised Learning Theory (Accepted at NeurIPS 2024) Oct. 2023 - May 2024

- **Summary:** In this work, we seek to establish a theoretical explanation for the principle of equivariant self-supervised learning (E-SSL) using the information theory. We utilize the explaining-away effect to analyze the mutual information between augmentation and class information, which indicates that E-SSL can indeed encourage the model to extract class-relevant features from data and benefit downstream tasks. We also propose three principles for the design of E-SSL and explain how advanced E-SSL designs echo with our theory.

Contributions:

- Use a toy model to investigate how to maximize the mutual information between the augmentation information and the class information when given the encoder feature.
- Discover that a balanced mixture of augmentation information and class information makes a good feature and that a larger action space is preferable for learning a better feature.
- Verify that acquiring class information can help E-SSL by explicitly injecting and eliminating class information during pretraining.

- Verify that aggressive base transformations can effectively prune shortcuts like learning from colors in E-SSL, encouraging the model to rely more on extracting class information during pretraining.

Contrastive Learning Theory (Accepted at ICLR 2025) June 2024 - Oct. 2024

- **Summary:** In this work, we aim to investigate the role of the projection head in contrastive learning from an information theory perspective. We establish theoretical lower and upper bounds for the downstream performance, suggesting that the projection head should act as an information bottleneck. In addition, we optimize the design of the projection head by means of training and structural regularization, which improves downstream performance.

  Contributions:

- Establish the theoretical lower and upper bounds on downstream performance and propose that the key point is to control the mutual information between the encoder and projector features.
- Add the matrix mutual information between the encoder and projector features as a regularization term to the contrastive loss, which increases downstream task accuracy and aligns well with theory.
- Discover that discretizing the projector feature improves downstream performance across different datasets and propose a brief theoretical explanation for this method.

Behavior-regularized Reinforcement Learning (Accepted at ICLR 2026) Nov. 2024 - Sep. 2025

- **Summary:** In most RL settings, we need a KL divergence term to control the distance between the reference policy and the learned policy, but it inherently limits the support of the learned policy to that of the reference policy. In this work, we introduce Value Gradient Flow (VGF), a particle-based gradient flow method to learn the policy, which does not apply any regularization term. We theoretically prove that VGF has an impilcit regularization by controlling the VGF learning step, and is able to explore beyond the support of the reference policy. VGF demonstrates superior performance on RL benchmarks such as D4RL and OGBench.

  Contributions:

- Theoretically prove that VGF implicitly imposes regularization on the learned policy by showing that the Maximum Mean Discrepancy (MMD) has an upper bound that is linear with respect to the VGF learning rate and the VGF step.
- Theoretically prove that VGF can enable the learned policy to explore beyond the support of the reference policy by showing that the ϵ-support of the learned policy is almost surely not a subset of that of the reference policy in both discrete and continuous action spaces.
- Evaluate the performance of VGF on D4RL and OGBench.

Tool-Integrated Multimodal Reasoning (Under Review) May 2026 - Sep. 2026

- **Summary:** We study how SFT data shapes tool-use retention during RL in multimodal language models. We find that deeper tool-use demonstrations do not necessarily produce more persistent tool use. Through theoretical analysis and controlled experiments, we investigate selective tool retention and develop a data-filtering strategy that improves downstream reasoning and tool-enabled performance.

  Contributions:

- Conduct controlled experiments with matched SFT token budgets, identifying pre-call reasoning as a potential factor in tool-use calibration and retention during RL.
- Derive idealized outcome-only optimization dynamics showing that stronger SFT initializations will result in higher overall and tool conditional accuracy, instead of higher tool initiation rate.
- Develop SFT data filtering based on pre-call reasoning length, improving post-RL accuracy across five benchmarks.

Internship Experiences

Apple (Foundation Model Team) May 2026 - Aug. 2026

- Investigating how training data difficulty and mixture composition affect reinforcement learning performance through data filtering and mixing experiments.

Skills

- Programming Languages: C, C++, Python (Pytorch).
- Tools: VS code, Github, Latex, Notions, Markdown.