My research interests:

1. Statistical learning theory:
    - Handling distribution shift (unsupervised domain adaptation, domain generalization, domain extrapolation)
        - Settings:
            - Covariate shift
            - Model shift
            - Label shift
        - Methods:
            - Label propagation
            - Learning invariant features (DRO / meta learning)
            - Importance reweighting
            - Optimal transport
    - Understanding representation learning
        - Self-supervised learning:
            - Architecture
            - Implicit bias
            - Feature learning
        - Representation learning for continual learning
        - Evaluation: how to evaluate pre-trained model without (labeled) target data
2. Optimization:
    - Designing algorithms for structured problems that guarantee (global) convergence
        - Min-max optimization
        - Optimizing over samples (inverse problems, adversarial training, recovering samples)
    - Model compression/pruning/knowledge distillation
        - NAS-based compression
        - Adaptive hyper-parameter choosing during pruning
    - Data distillation/pruning
        - Connections to domain adaptation
3. AI Safety
    - Reconstruction Attacks and Defenses
    - Watermark/copyright preservation
    - Alignment to human values
4. Miscellaneous topics and phenomena in deep learning
    - NN fitting random labels
    - Lottery ticket hypothesis
    - Evaluation and calibration for OOD generalization
    - Fine-grained feature learning vs neural collapse