

Kevin Yang

Microsoft Research

Title: Multimodal deep learning for protein engineering

Abstract:

Engineered proteins play increasingly essential roles in industries and applications spanning pharmaceuticals, agriculture, specialty chemicals, and fuel. Machine learning could enable an unprecedented level of control in protein engineering for therapeutic and industrial applications. Large self-supervised models pretrained on millions of protein sequences have recently gained popularity in generating embeddings of protein sequences for protein property prediction. However, protein datasets contain information in addition to sequence that can improve model performance. This talk will cover pretrained models that use both sequences, structures, and annotations to predict protein function or to generate functional protein sequences.