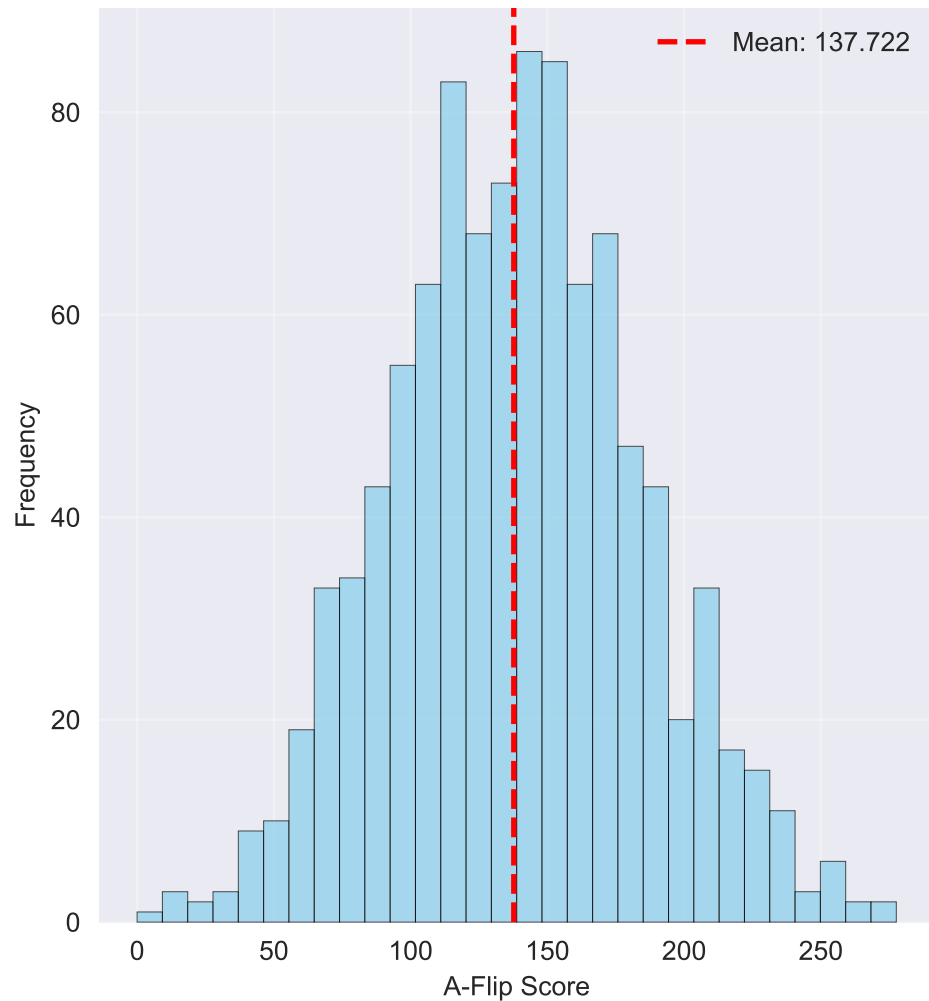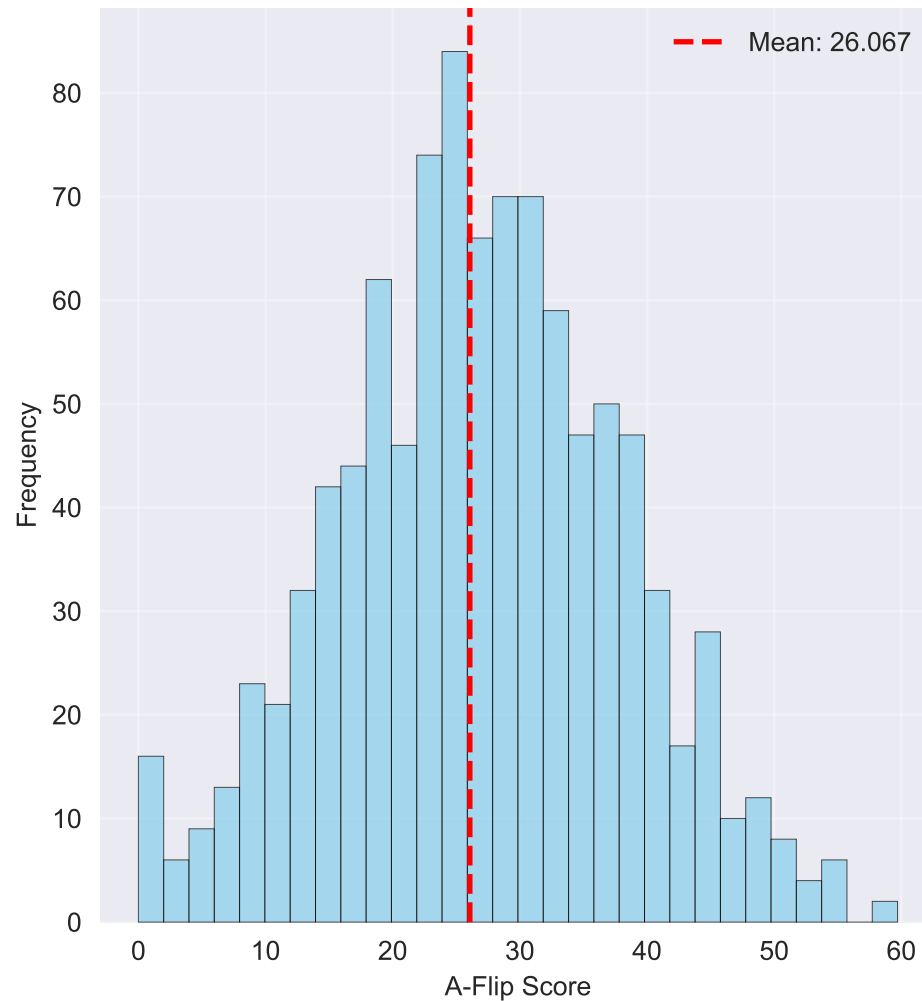A-Flip Score Distributions Across Your Model Implementations