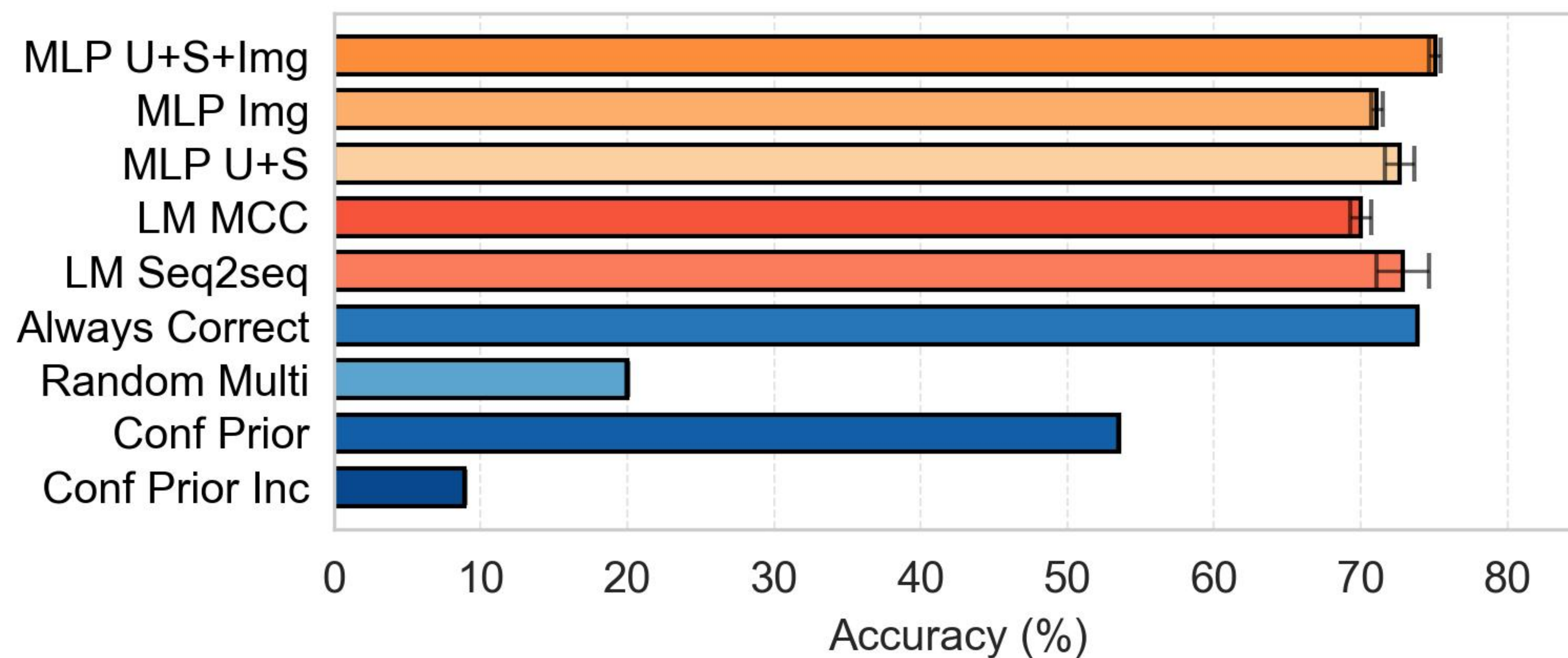
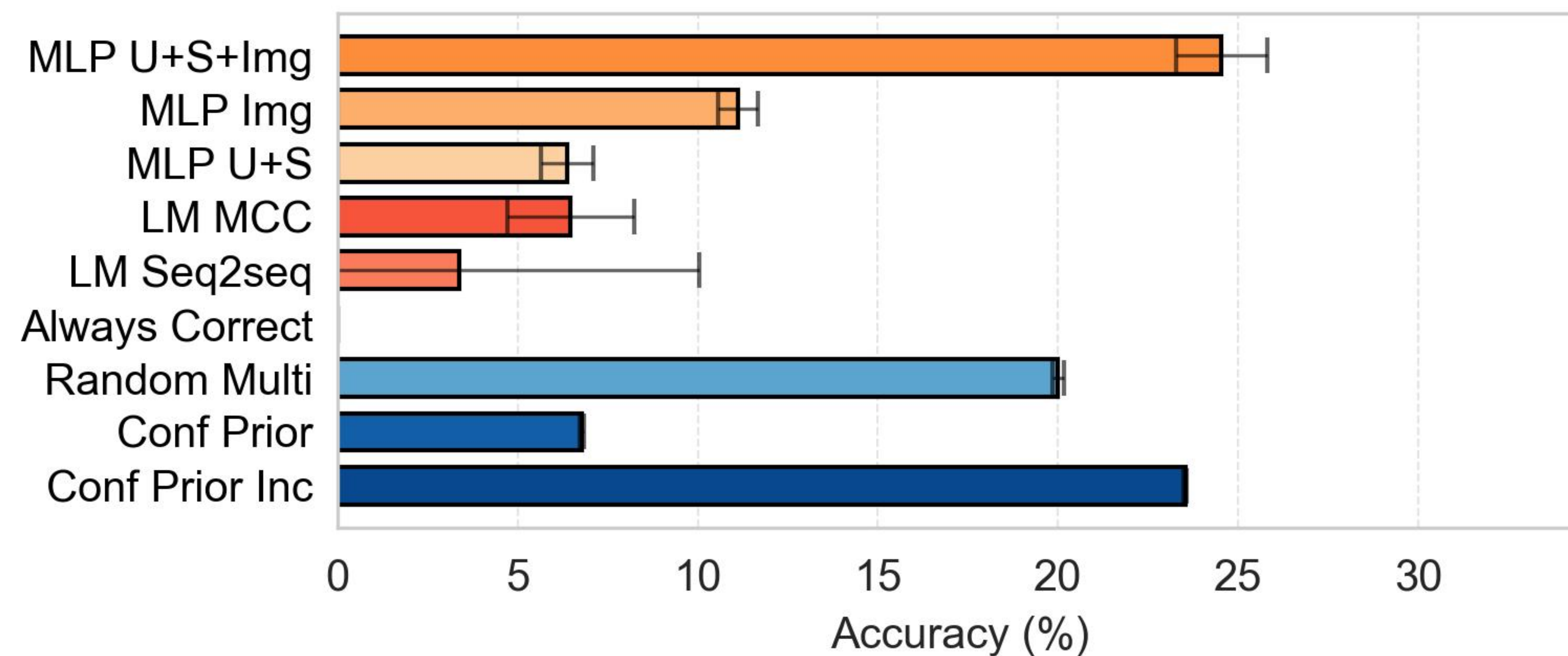


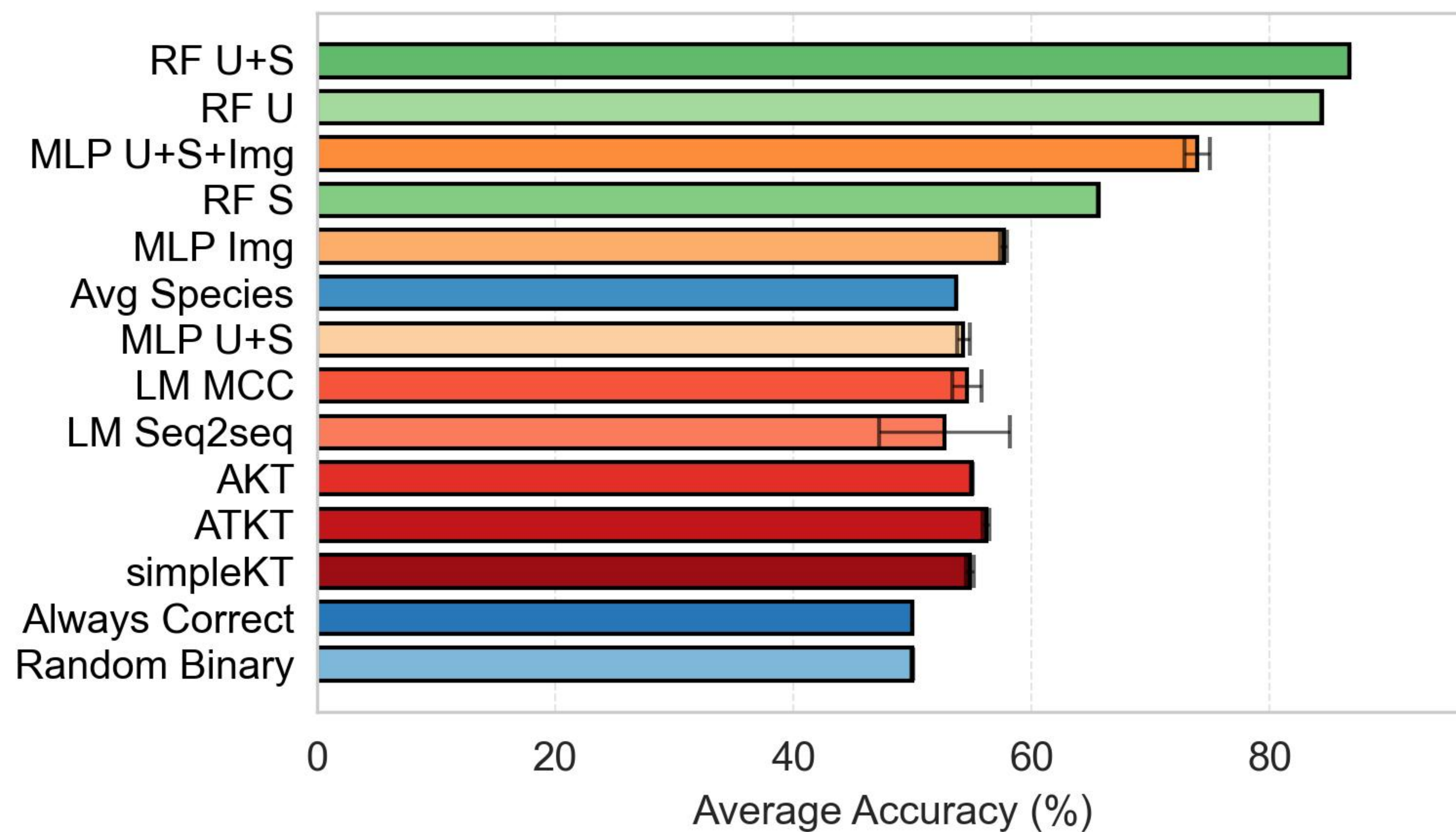
Multiple-Choice Accuracy



Multiple-Choice Incorrect Set Accuracy



Binary Macro Accuracy



Binary AP - Correctly Predicting User Mistakes

