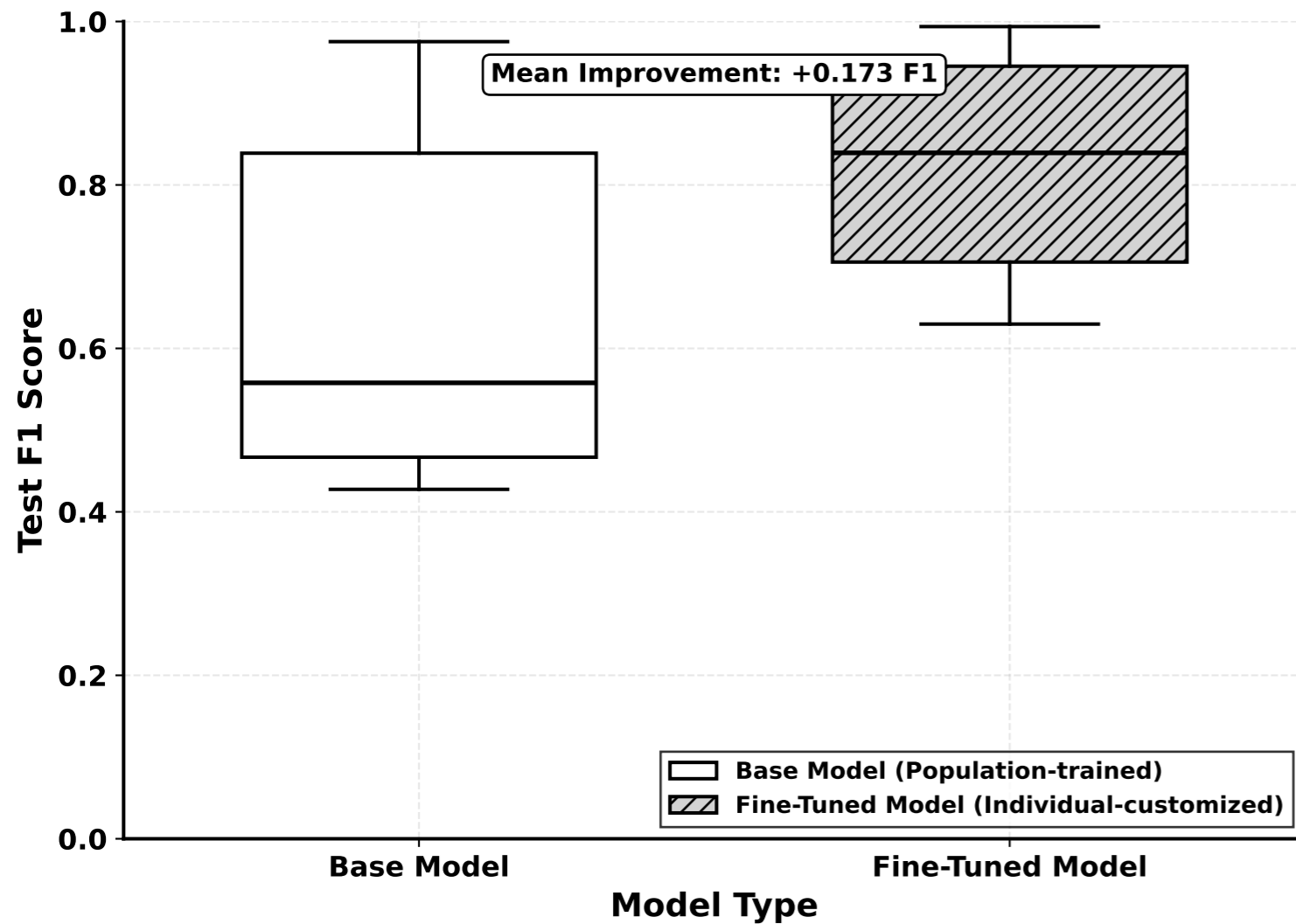


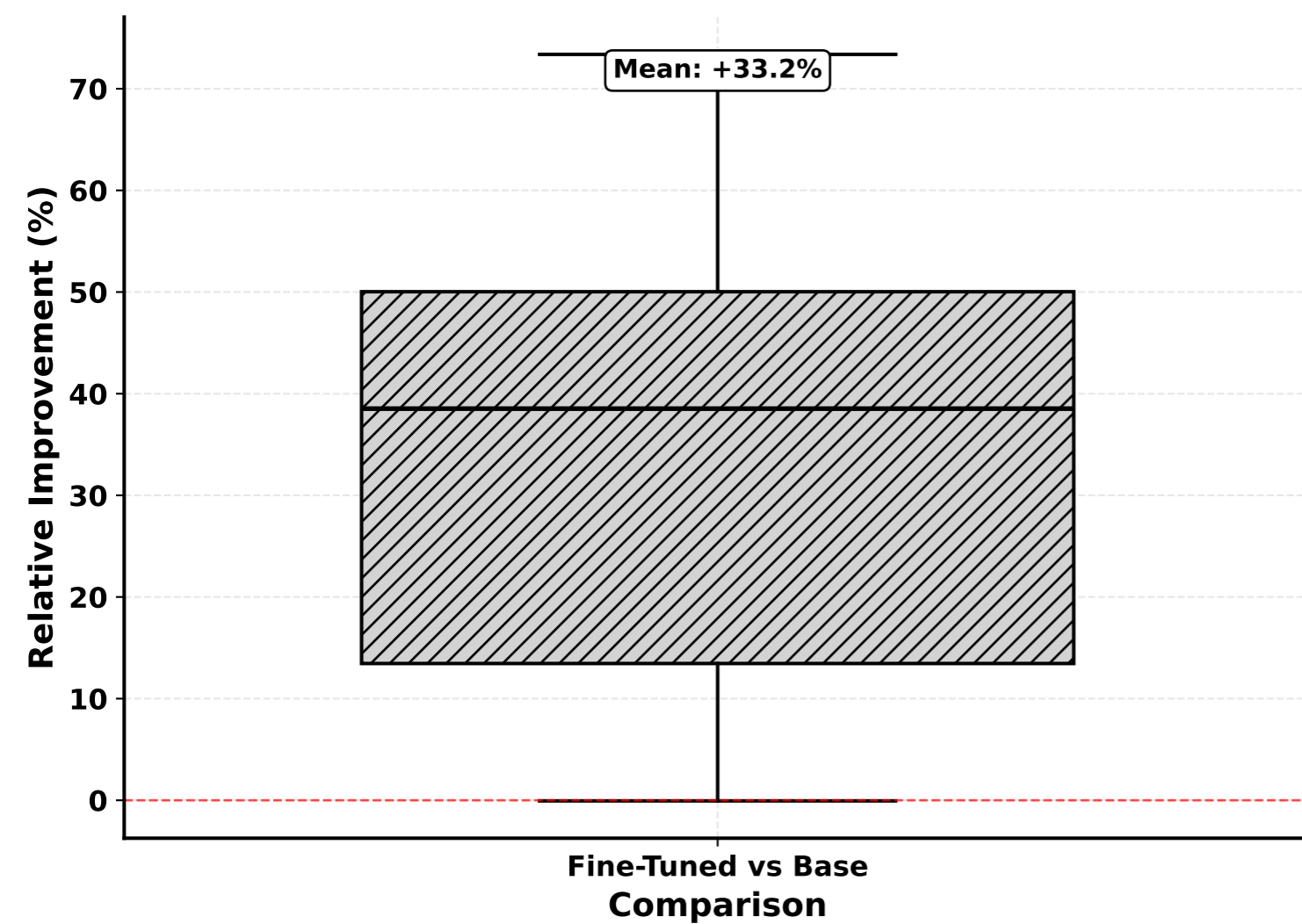
(A) F1 Score Comparison



(B) Absolute Improvement Distribution



(C) Relative Improvement Distribution



(D) Relative to Room for Improvement

