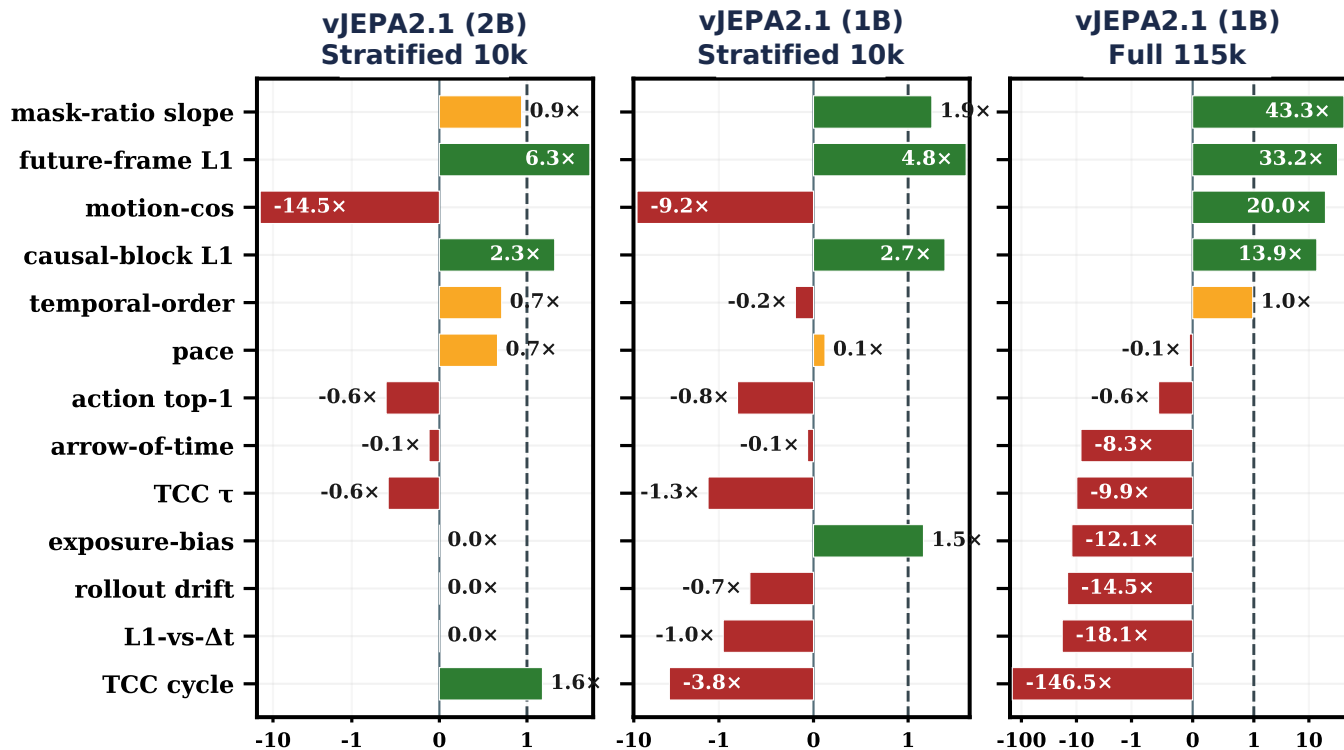


FactorJEP A vs. the best competitor, across scale: separation in 95%-CI units
 (rows sorted best-to-worst by the 115k panel; dashed line = the 1× separation threshold)



green: separated win ($\geq 1\times$) yellow: ahead but within CI (0-1 $\times$) red: competitor ahead (<0)
 ntotal = train + val + test. Full-scale 2B omitted: not run under the compute budget ($\approx 2\times$ the 1B cost).