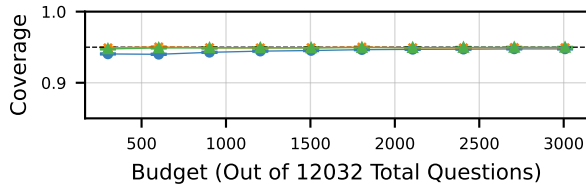
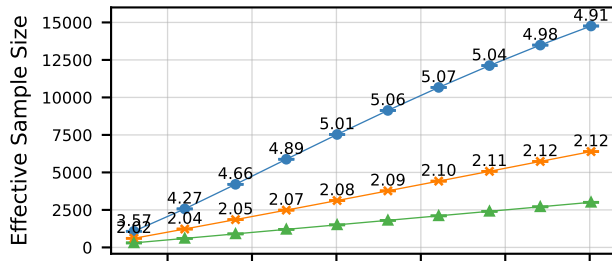
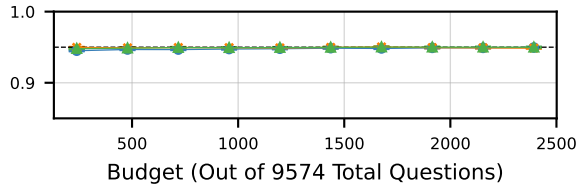
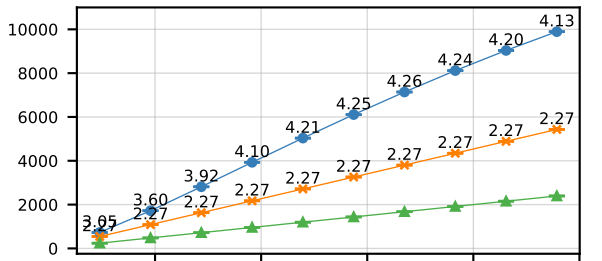


MMLU-Pro



BBH+GPQA+IFEval+MATH+MuSR



Legend: FAQ (blue line with circles), Best Baseline (Post-hoc Per Budget) (orange line with crosses), Uniform (green line with triangles), ---- 95% Coverage (Bottom Subplots)