HSE-bench Local Model Performance by Task Type