

Individual Task Completion Time Vs. $F_1^{\mathcal{O}^*}$ Score Across Models

