

Model Efficiency: F1 Score vs Model Size  
(Marker size  $\square$  1/Latency)

