

ToM Performance vs Model Size (3-shot)

