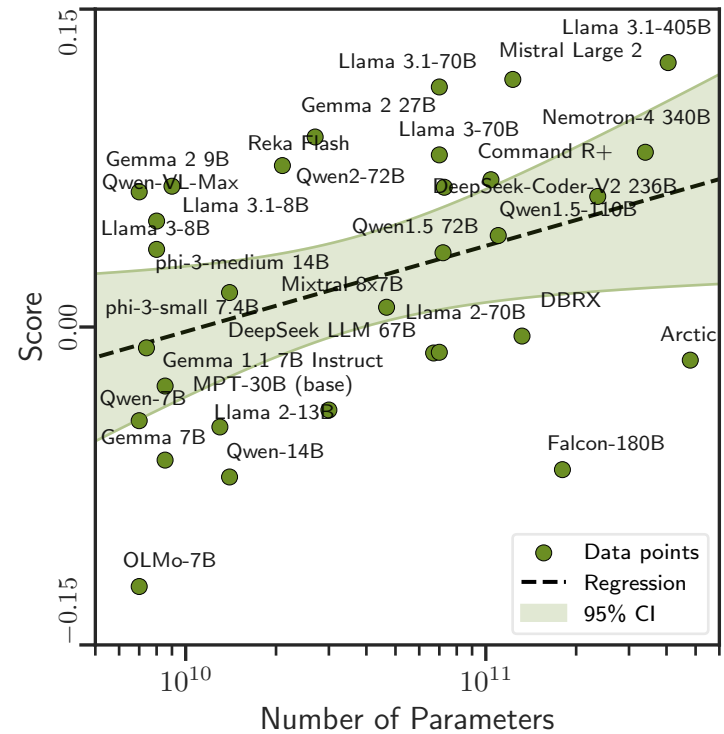
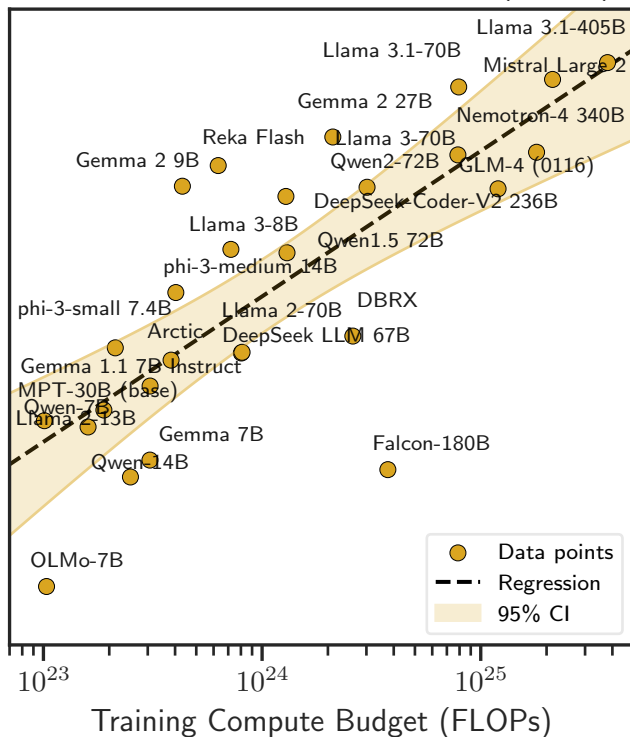


Effect of Number of Parameters



Effect of Training Compute Budget (FLOPs)



Effect of Dataset Size

