

Inverse Scaling: Larger Models Resist Steering

