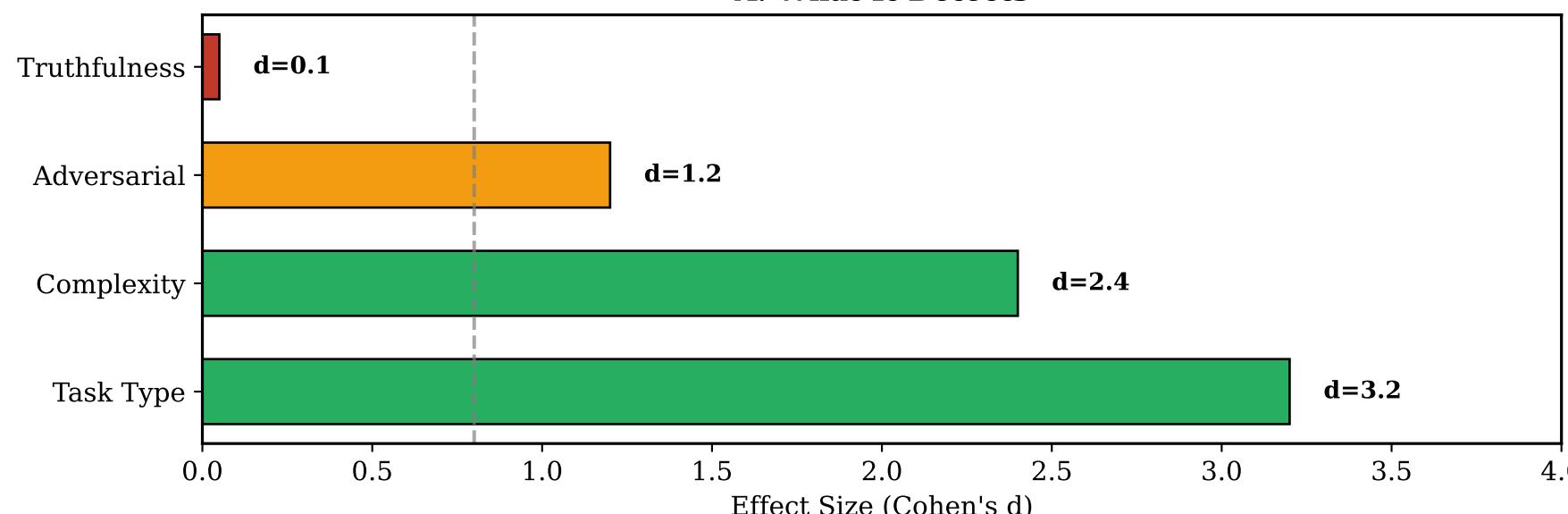


# Latent Diagnostics: Activation Topology Analysis

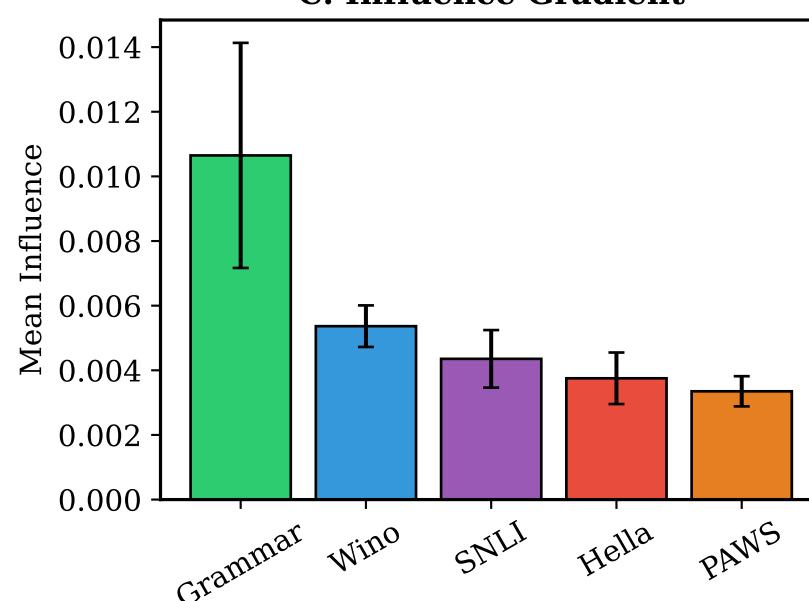
**A. What It Detects**



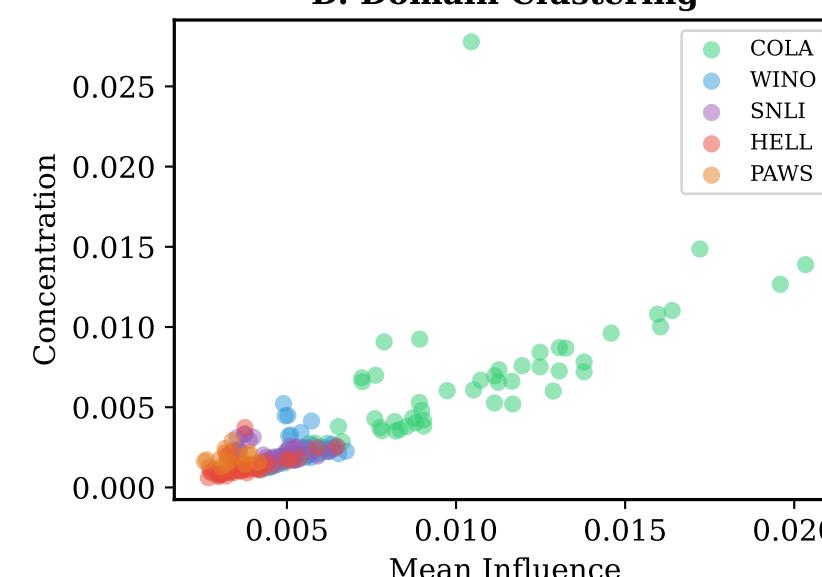
**B. Core Finding**

**KEY INSIGHT**  
Activation topology measures HOW a model computes, not WHETHER it is correct.  
Simple tasks → Focused  
Complex tasks → Diffuse  
True vs False → Same

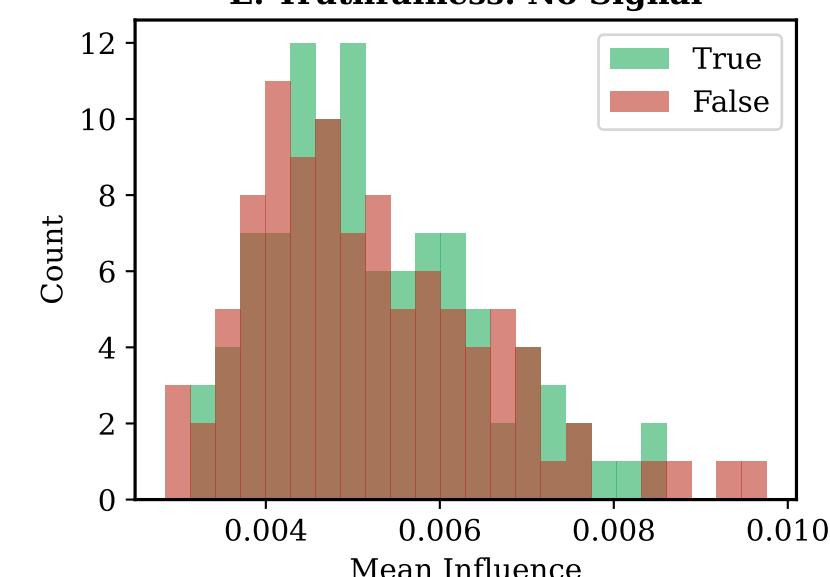
**C. Influence Gradient**



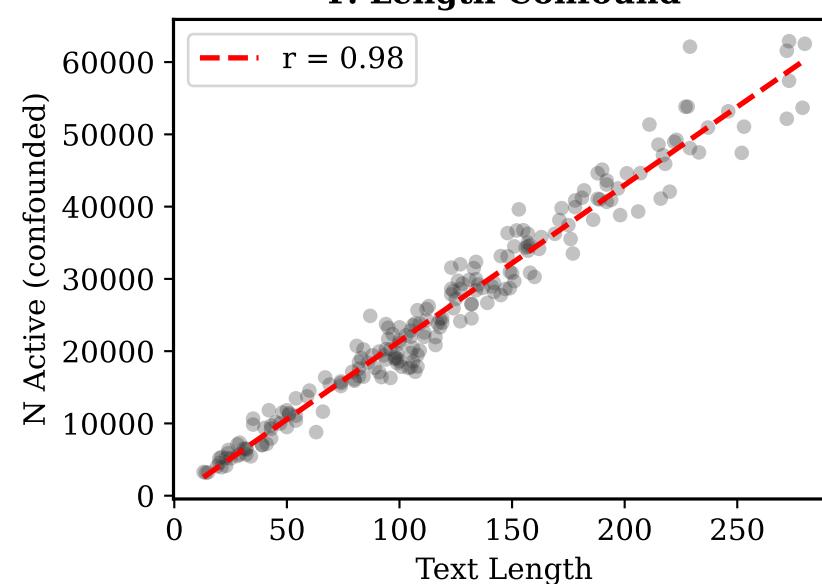
**D. Domain Clustering**



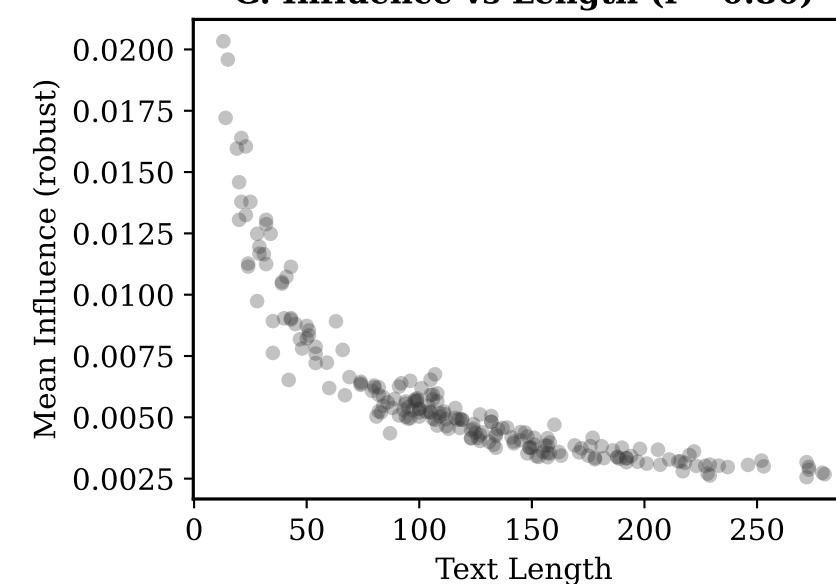
**E. Truthfulness: No Signal**



**F. Length Confound**



**G. Influence vs Length ( $r=-0.80$ )**



**H. Effect Sizes**

