**What Activation Topology Can Detect
(All effect sizes controlled for text length)**

- Truthfulness: **d = 0.05 DOES NOT WORK**
- Adversarial Inputs: **d = 0.80 WORKS**
- Computational Complexity: **d = 0.87 WORKS**
- Task Type (Grammar vs Reasoning): **d = 1.08 WORKS**

--- Large effect threshold (d=0.8)

Cohen's d (Effect Size, Length-Controlled)