# Steering Evaluation Aware Models to Act Like They Are Deployed

Tim Hua*, Andrew Qin*, Samuel Marks^, Neel Nanda^

## (Full poster coming this afternoon!)

As Seen In
ANTHROPIC
System Card:
Claude Opus 4.5