

# Robustness Evaluation



## Noisy Inputs

Test with incomplete or corrupted data



## Paraphrasing

Rephrase queries in different ways



## Domain Shift

Test on out-of-distribution data



## Adversarial Cases

Challenge with edge cases

### Robustness Metrics



#### Performance degradation:

% accuracy drop under noise



#### Input sensitivity:

Response stability to variations



#### Failure mode analysis:

When and why it breaks



#### Recovery ability:

How well it handles errors

0%



10%

20%+

Target: <10% performance drop under typical real-world variations