Requires Specific Training Data

Significant Performance Improvement

Strong In-Context Learning

Negligible Performance Improvement

8B Model

405B Model

Annealing Impact on Model Performance