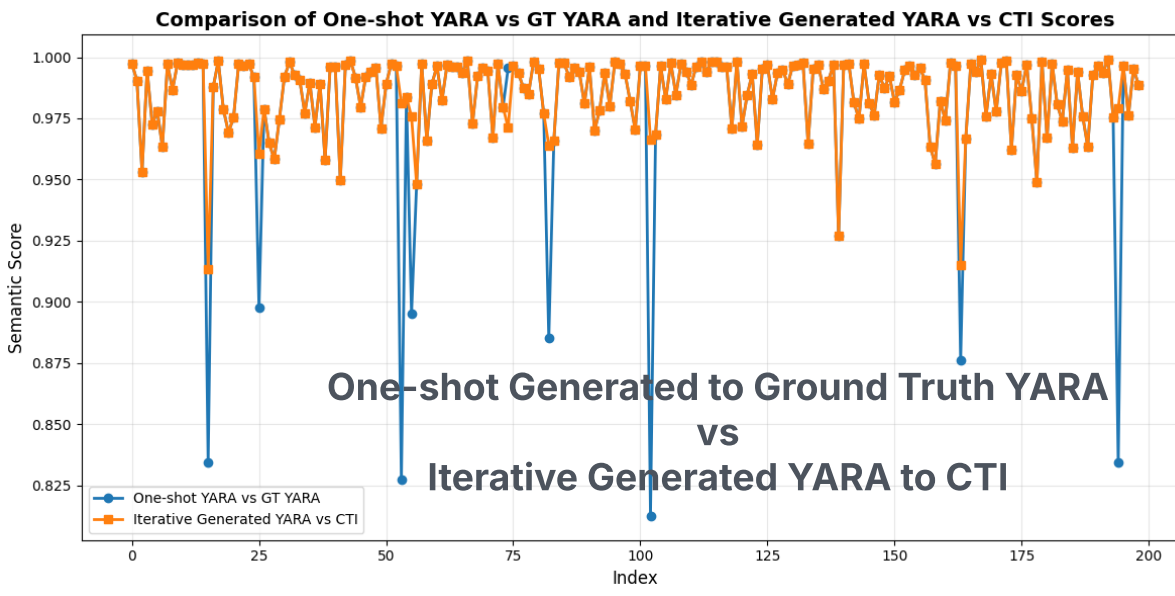
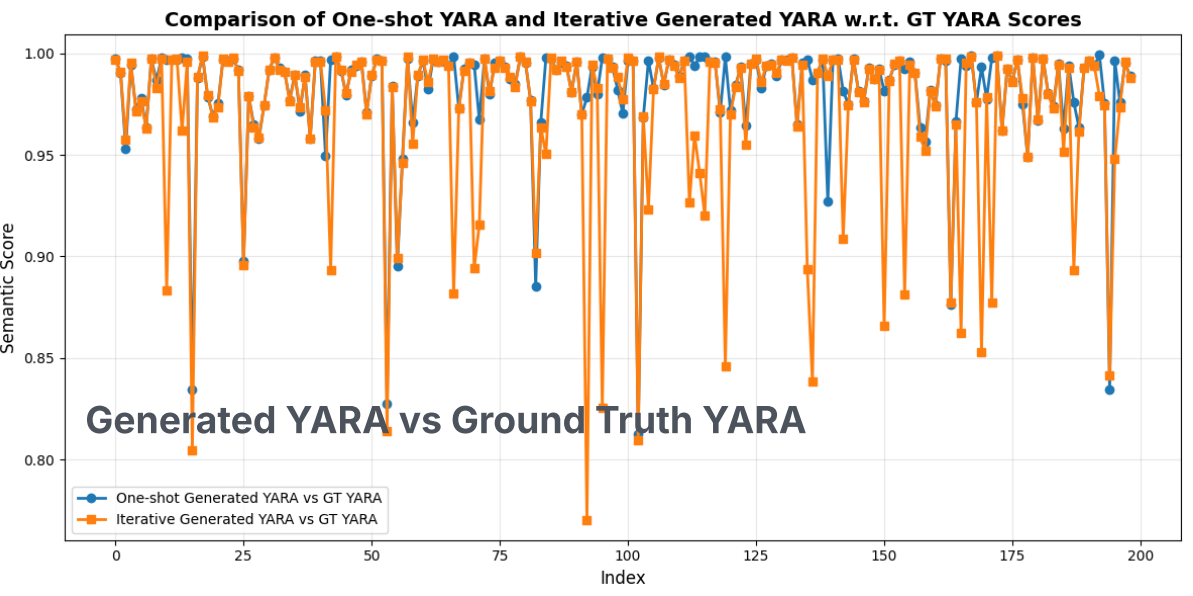
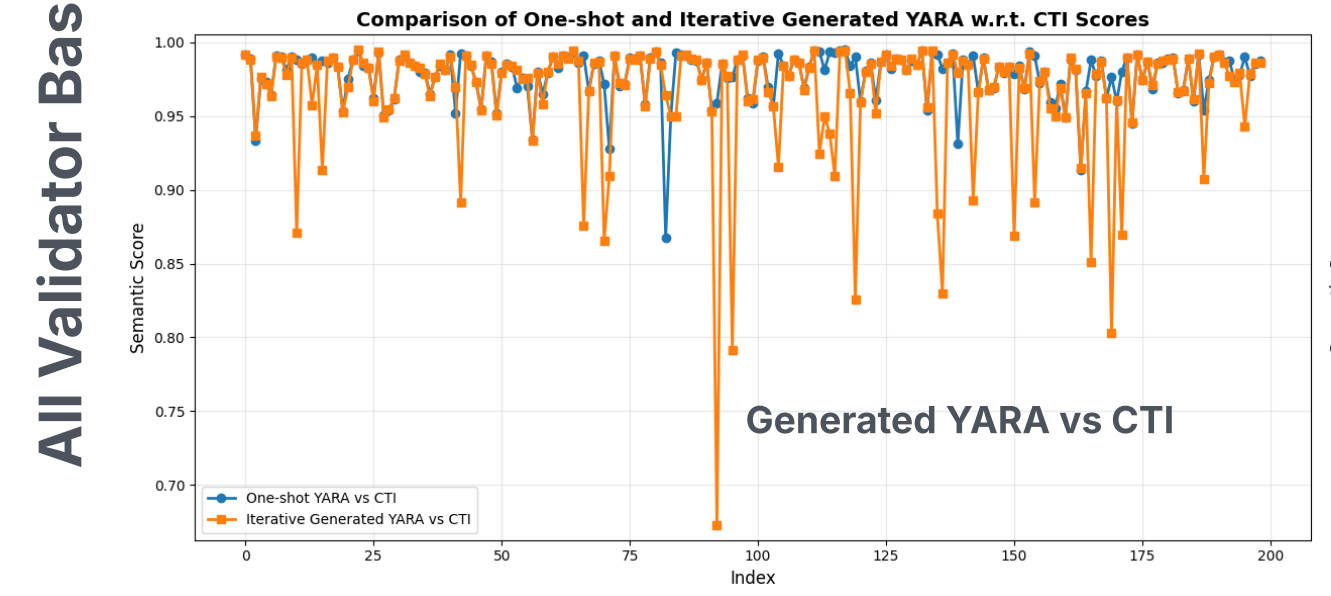
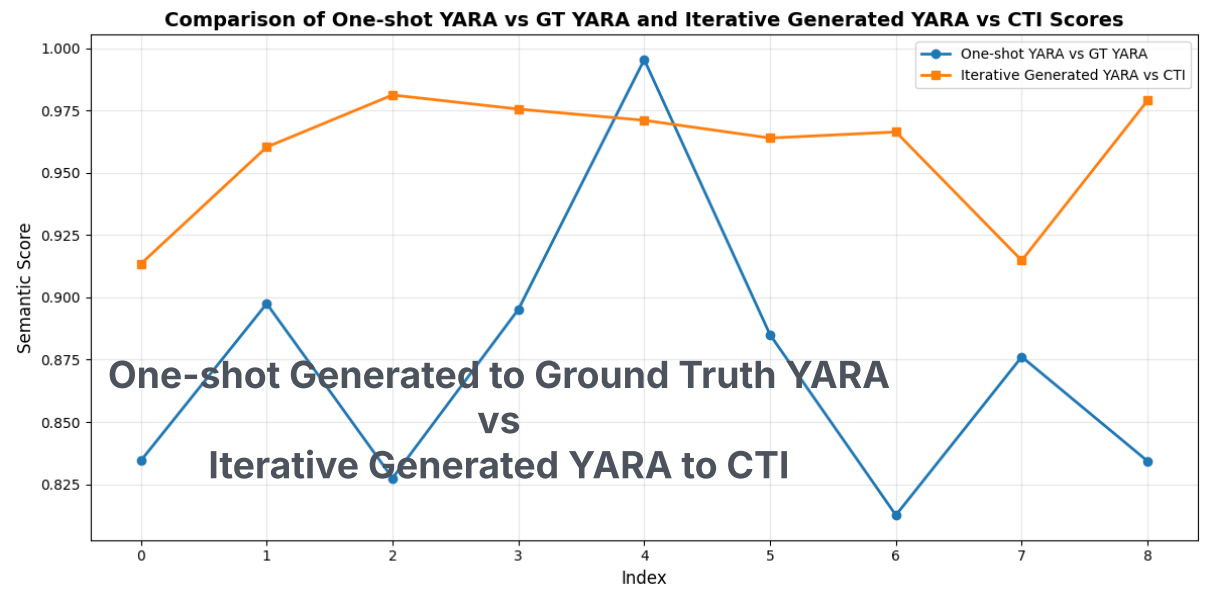
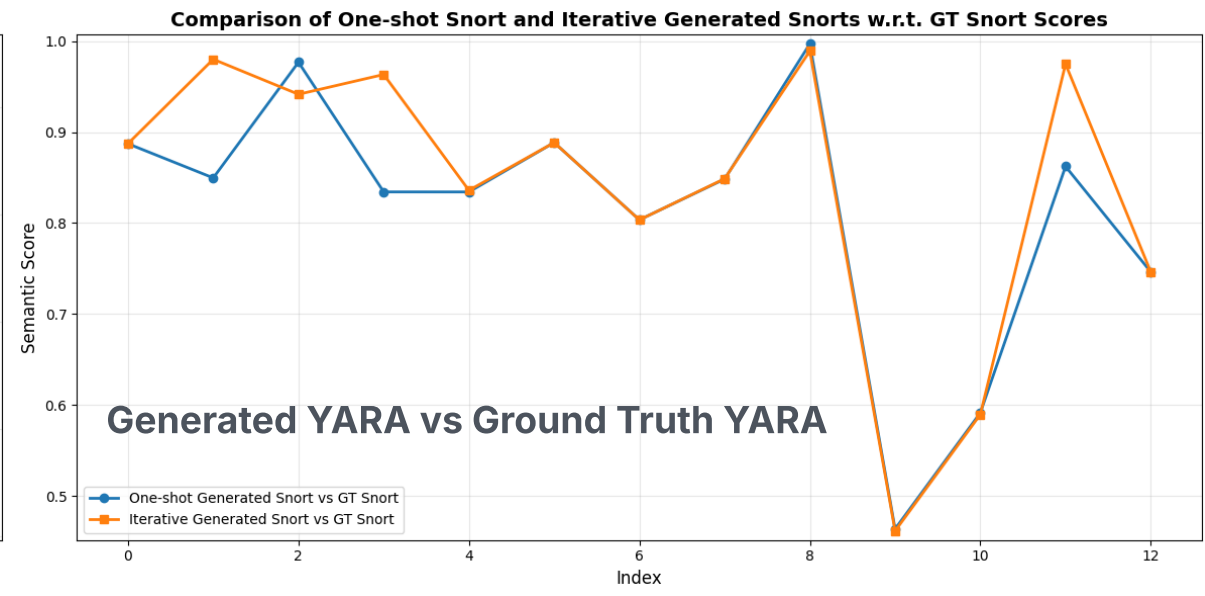
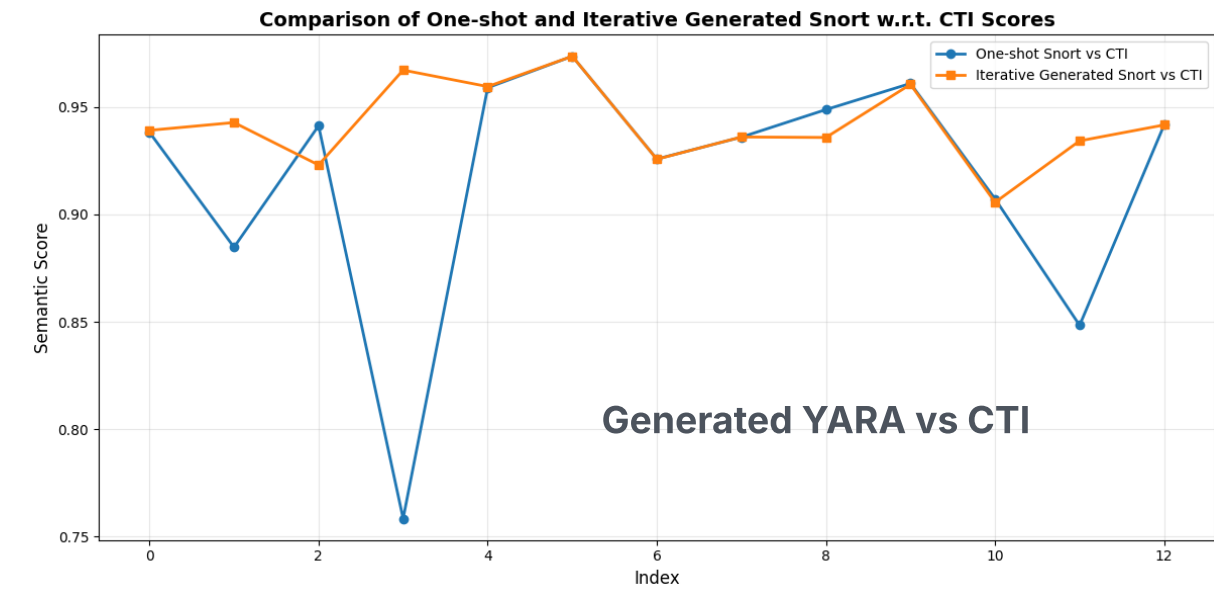
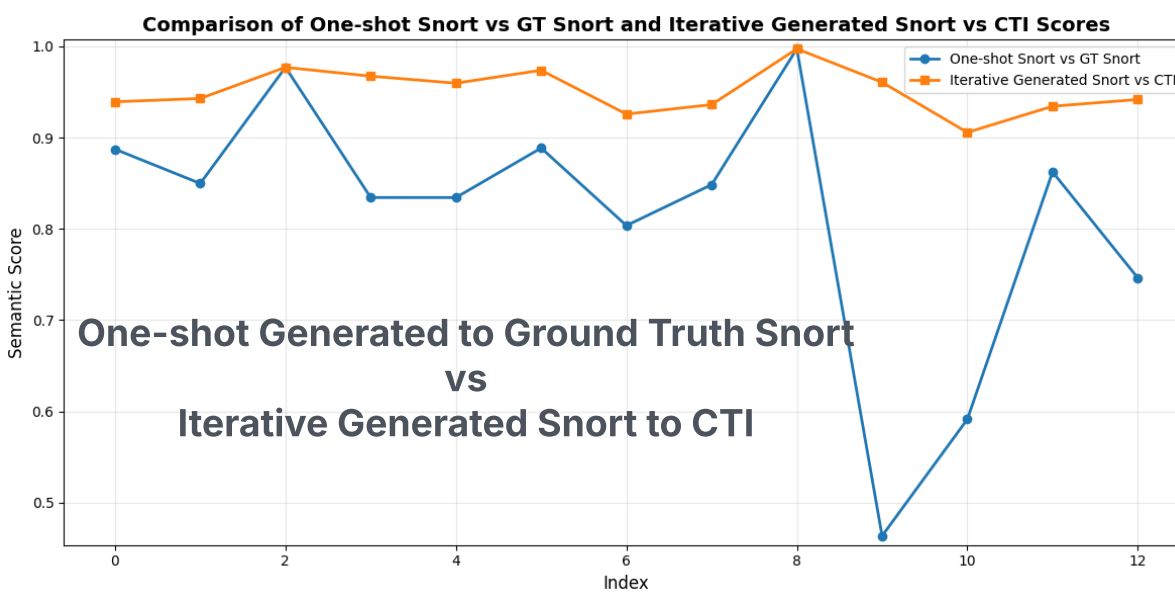
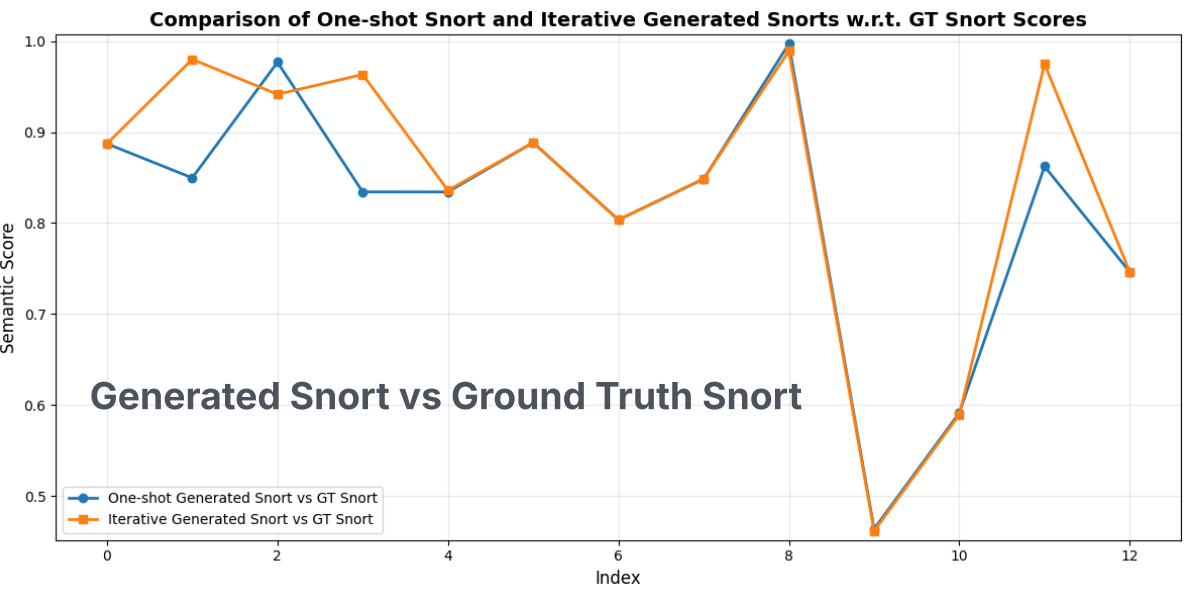
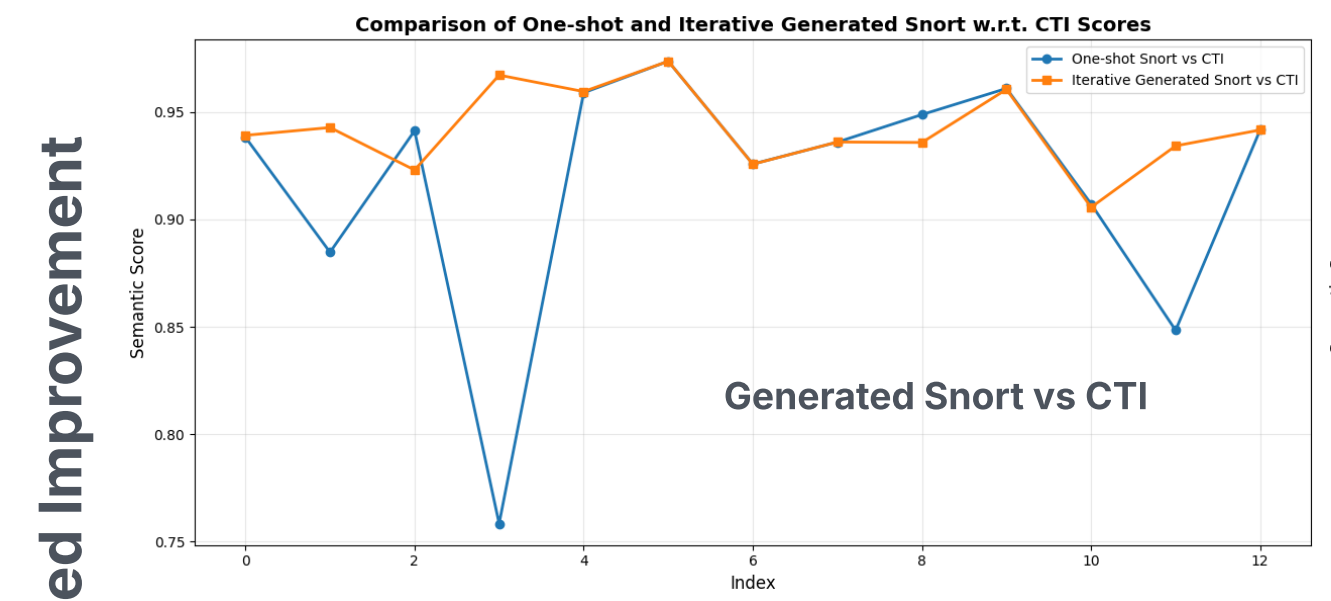
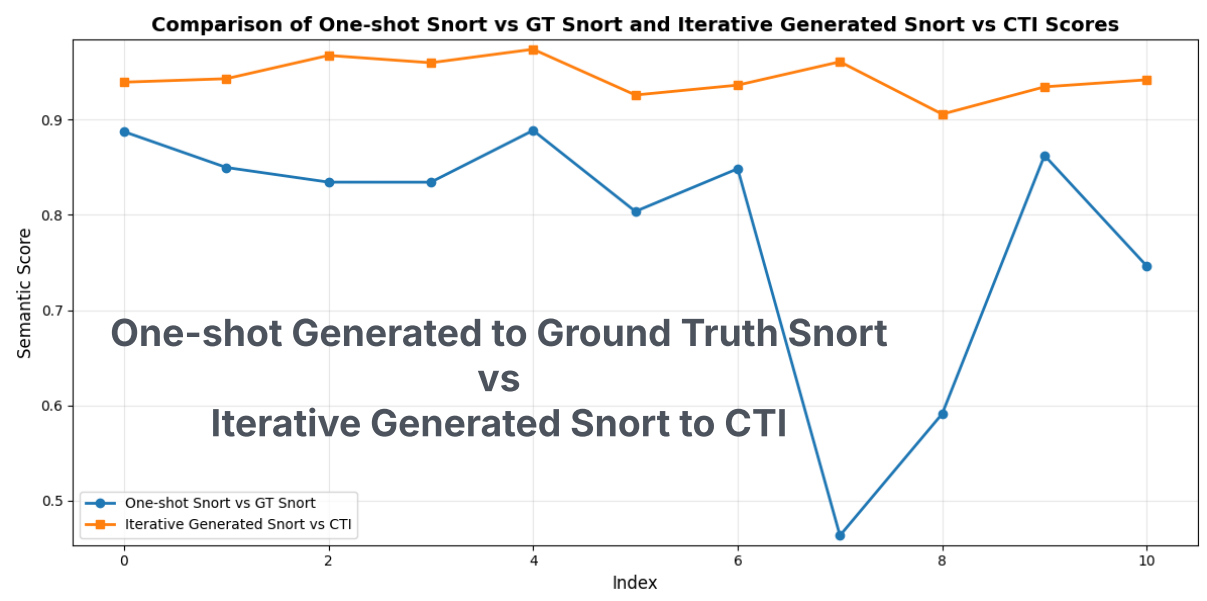
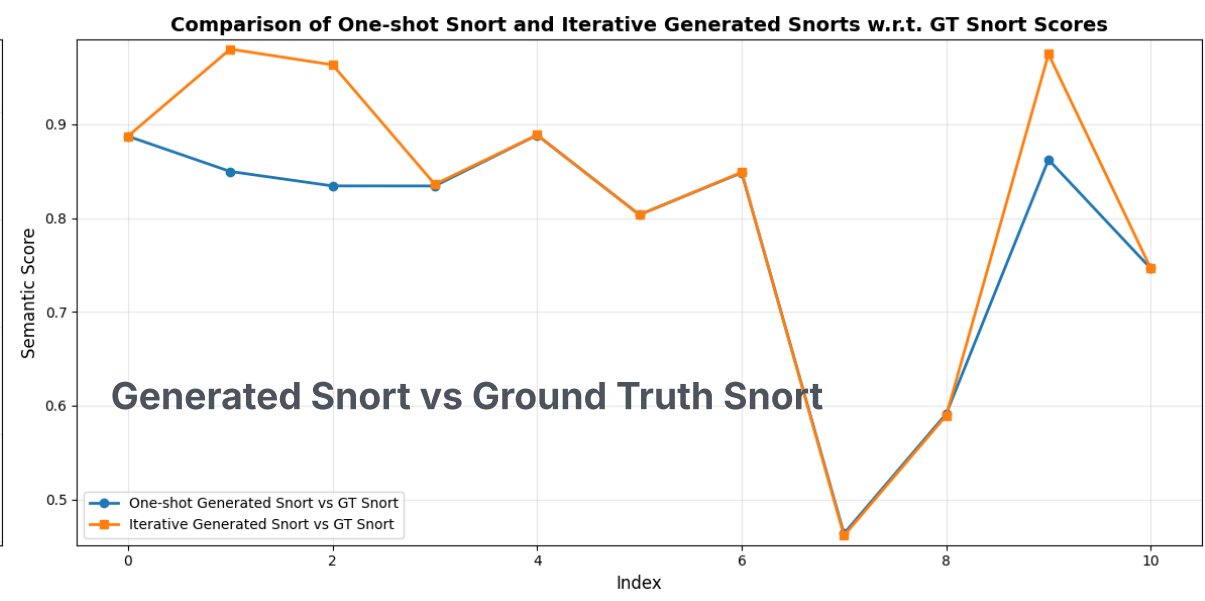
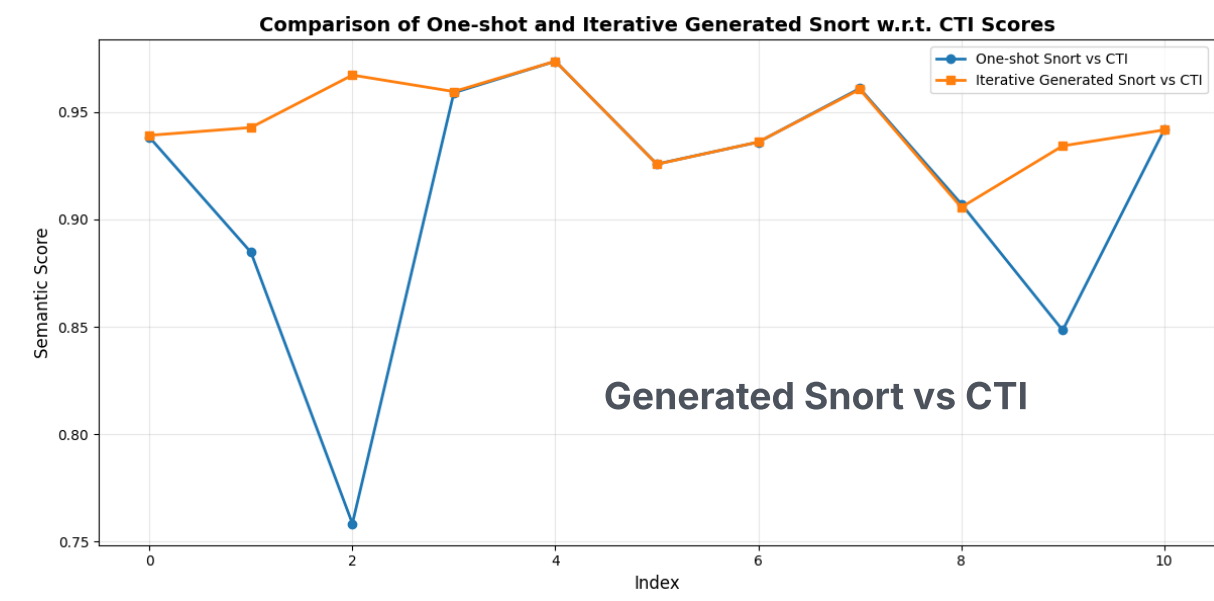
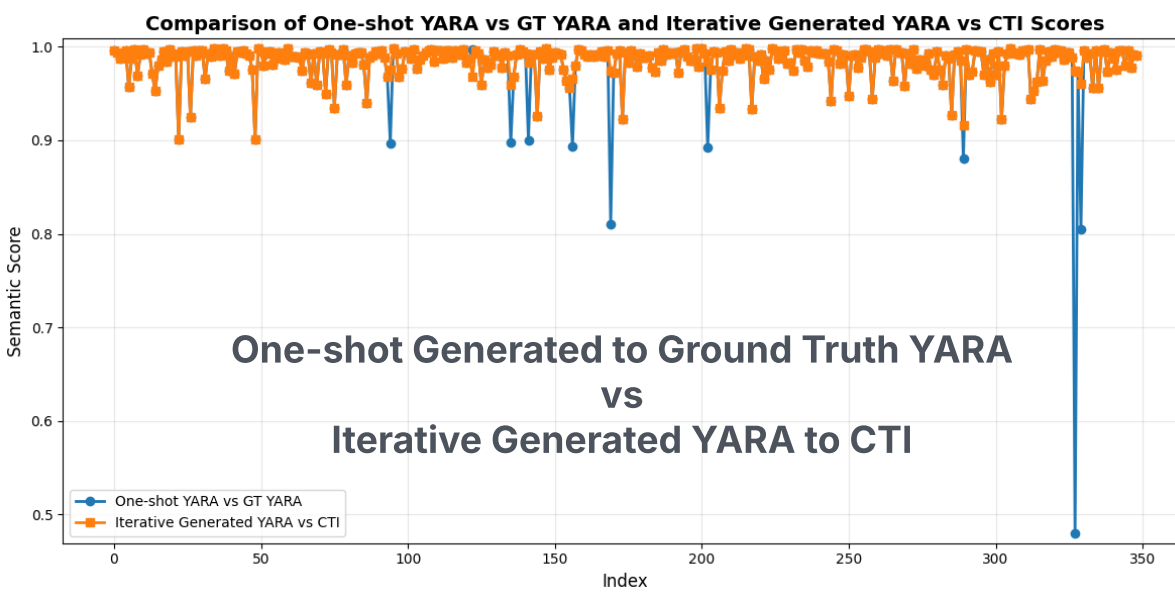
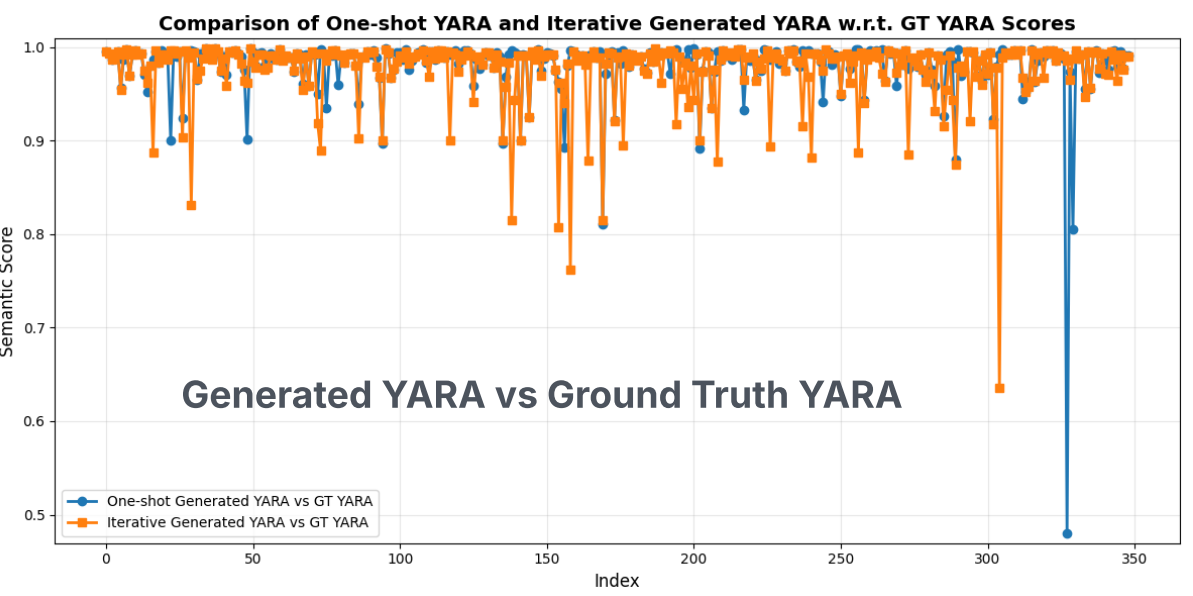
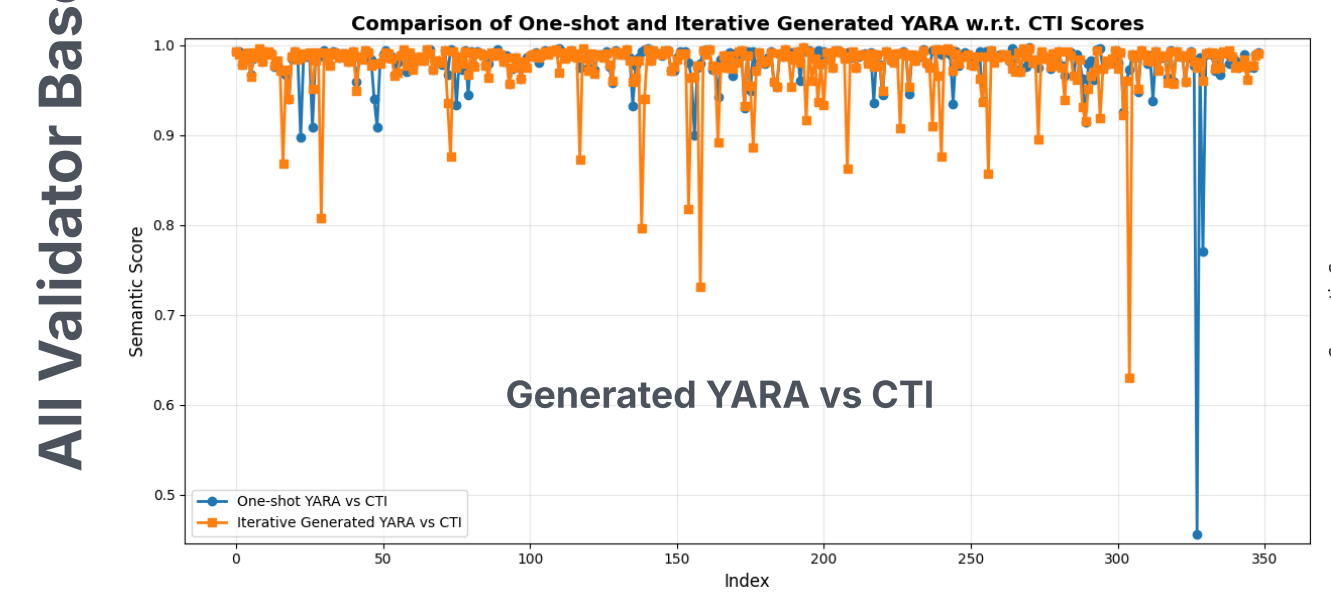
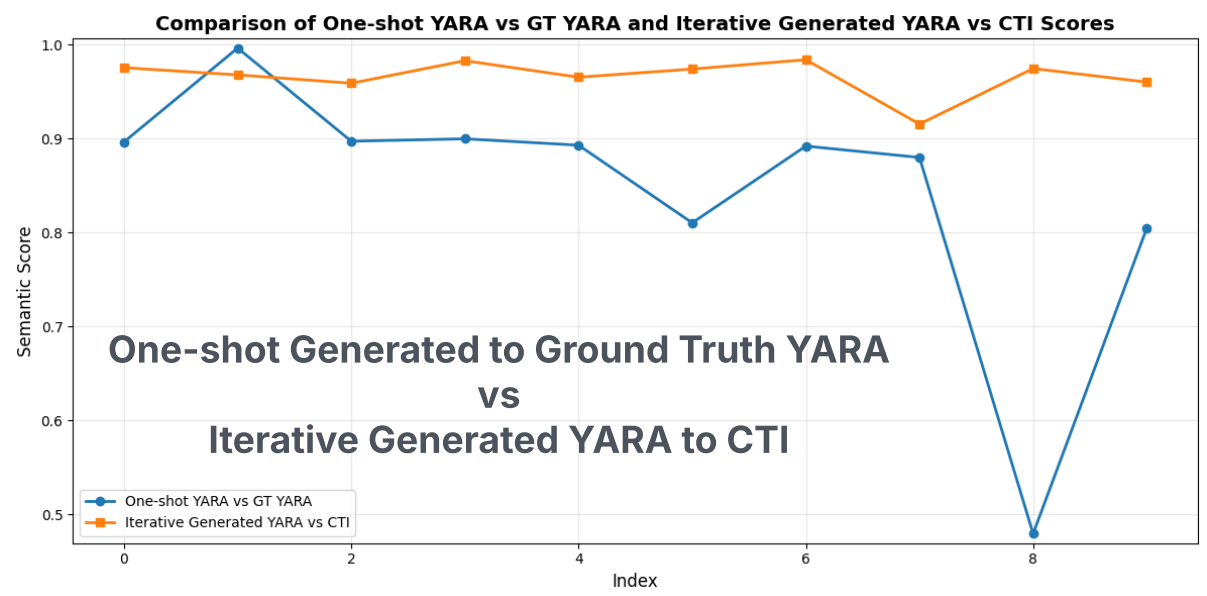
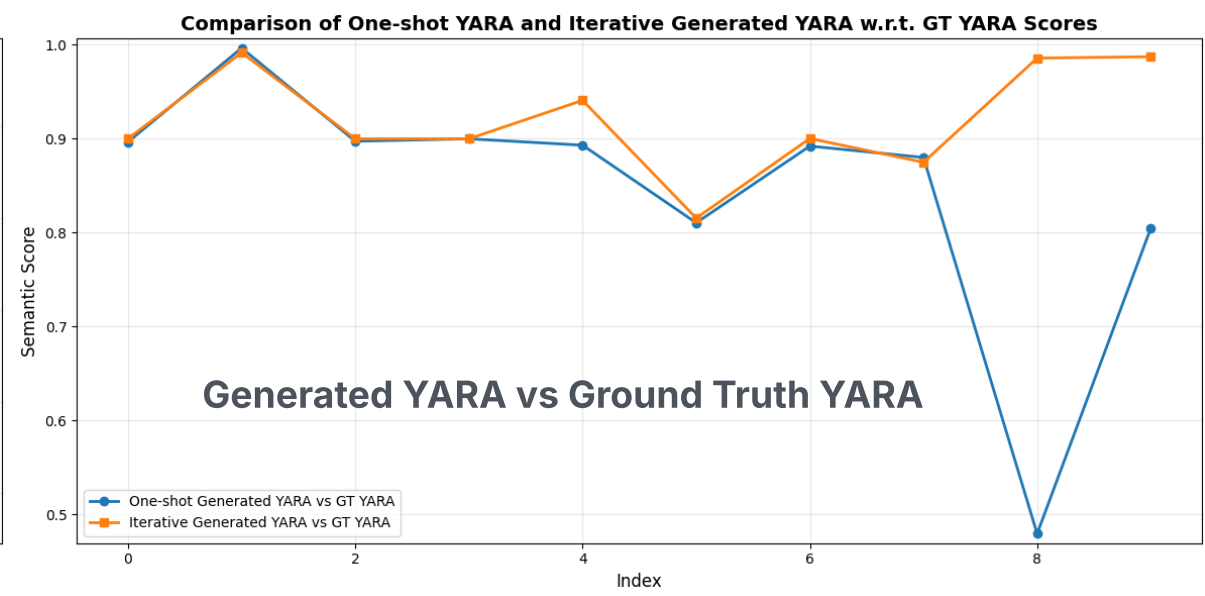
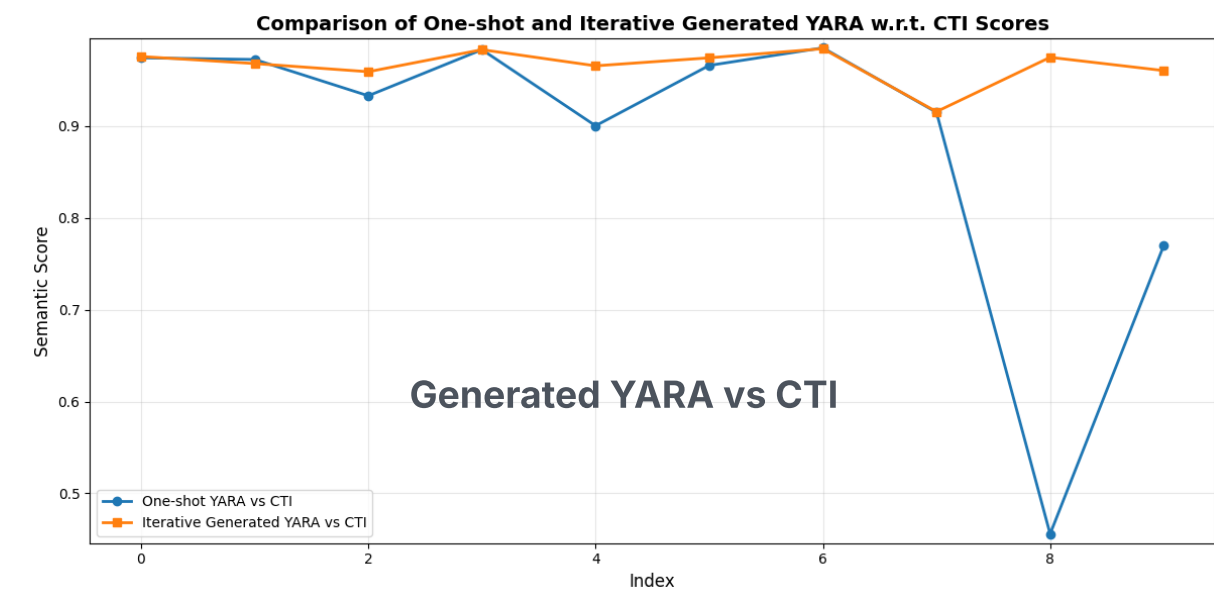
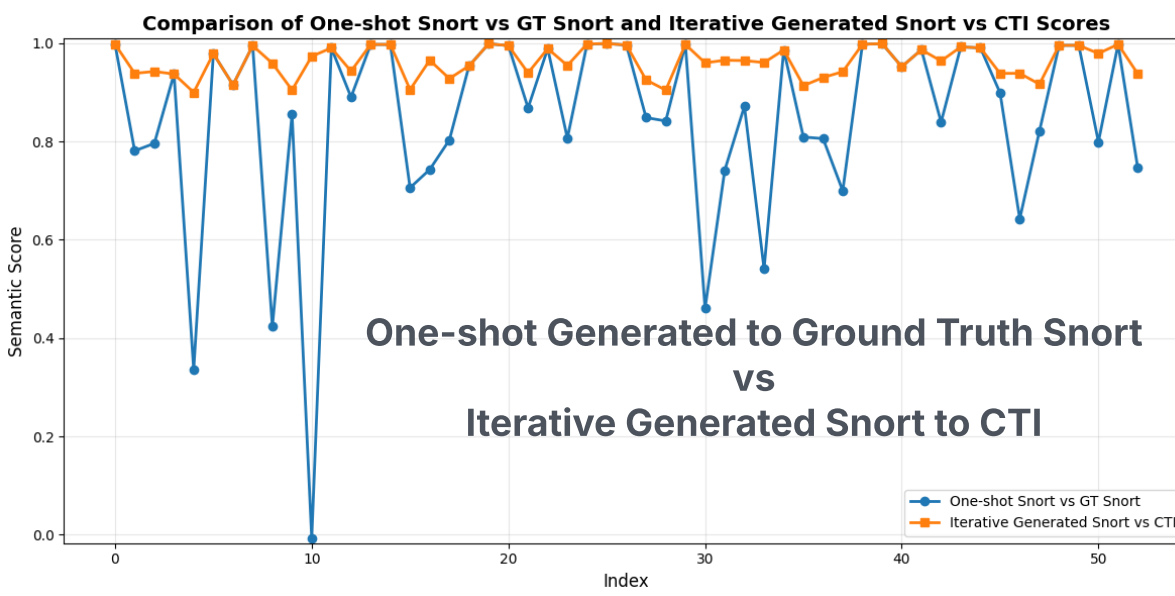
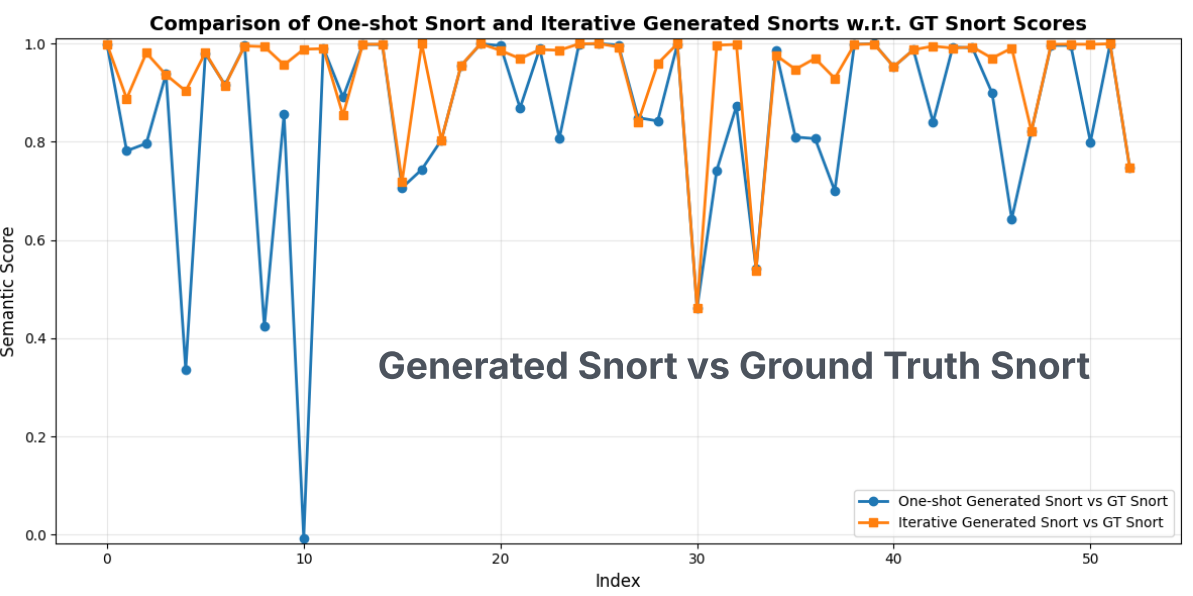
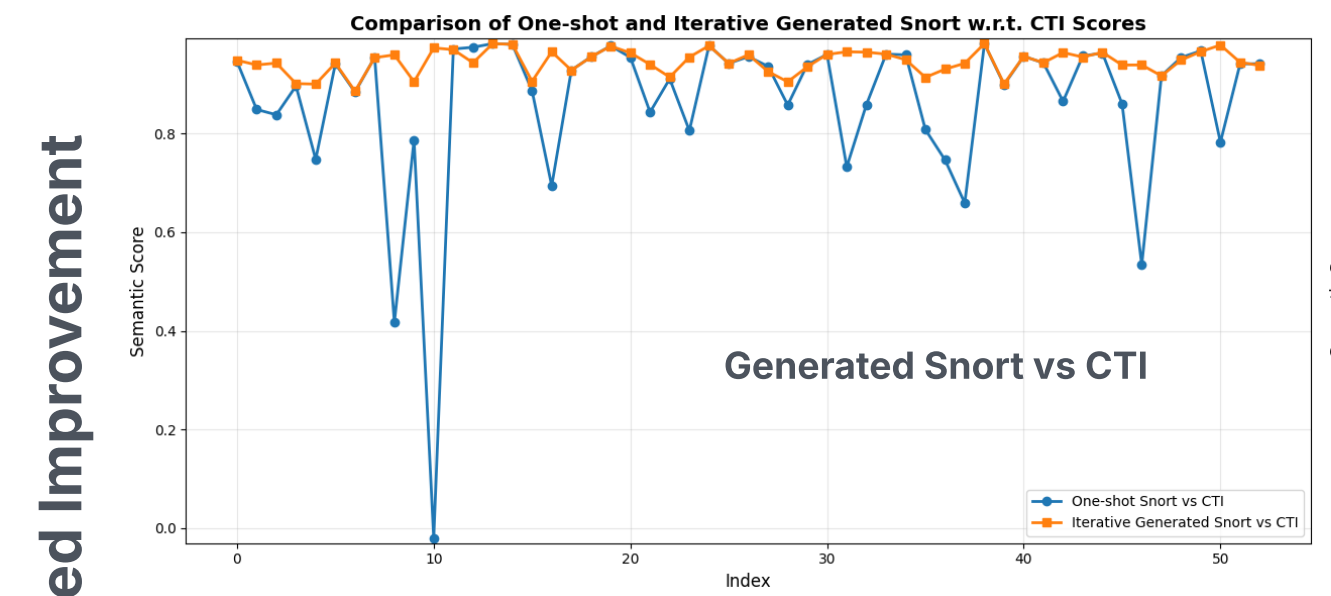
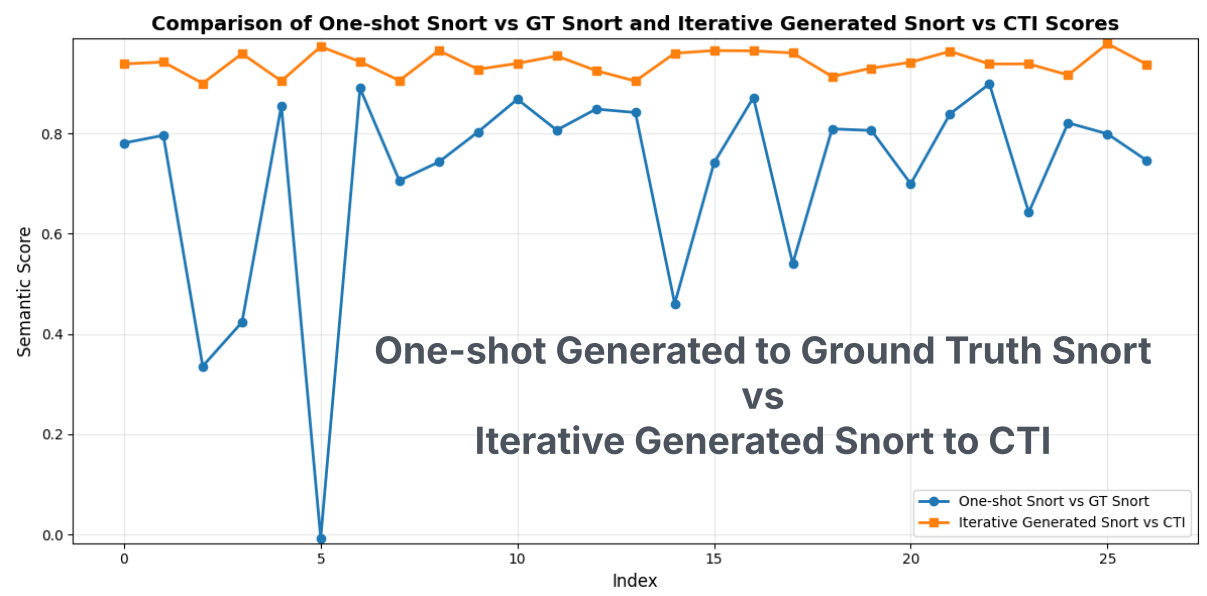
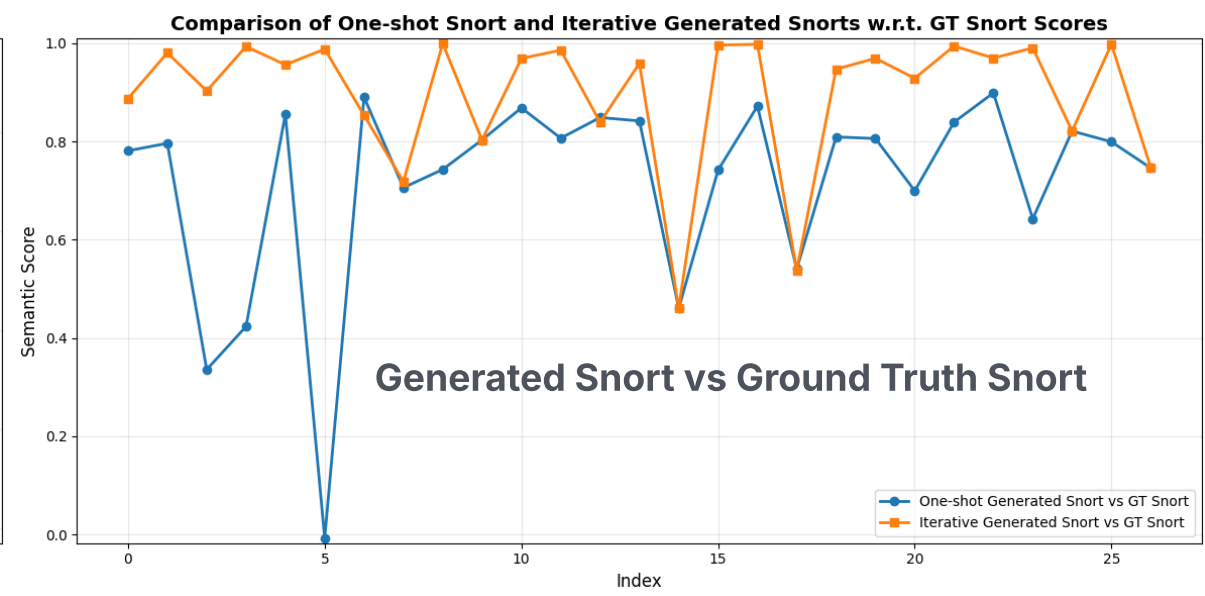
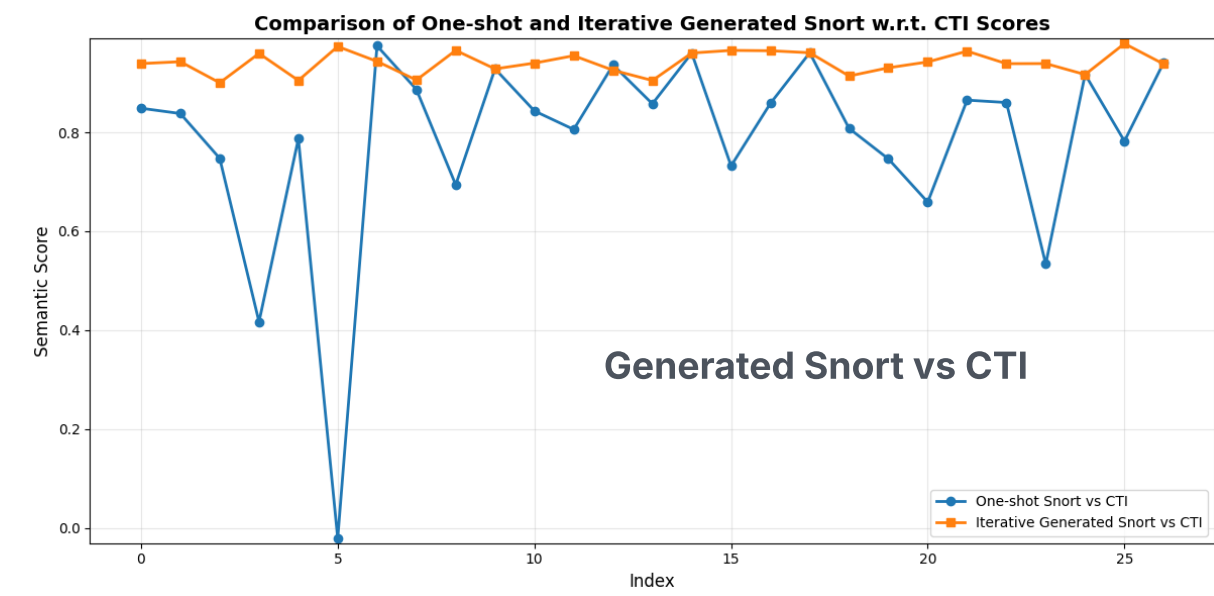


Semantic Validator Based Improvement



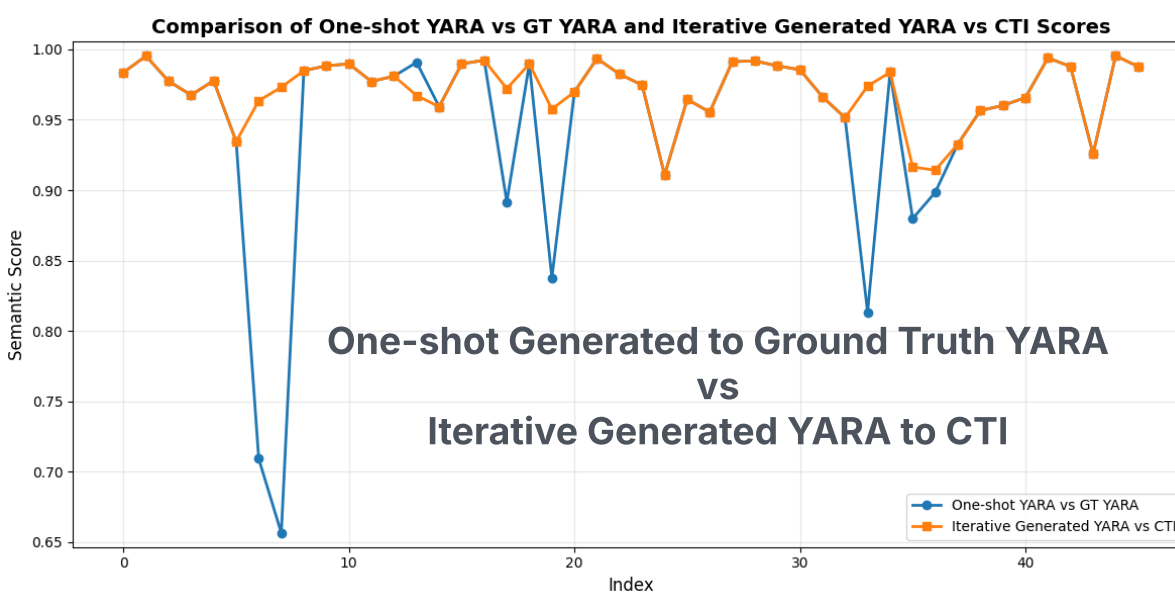
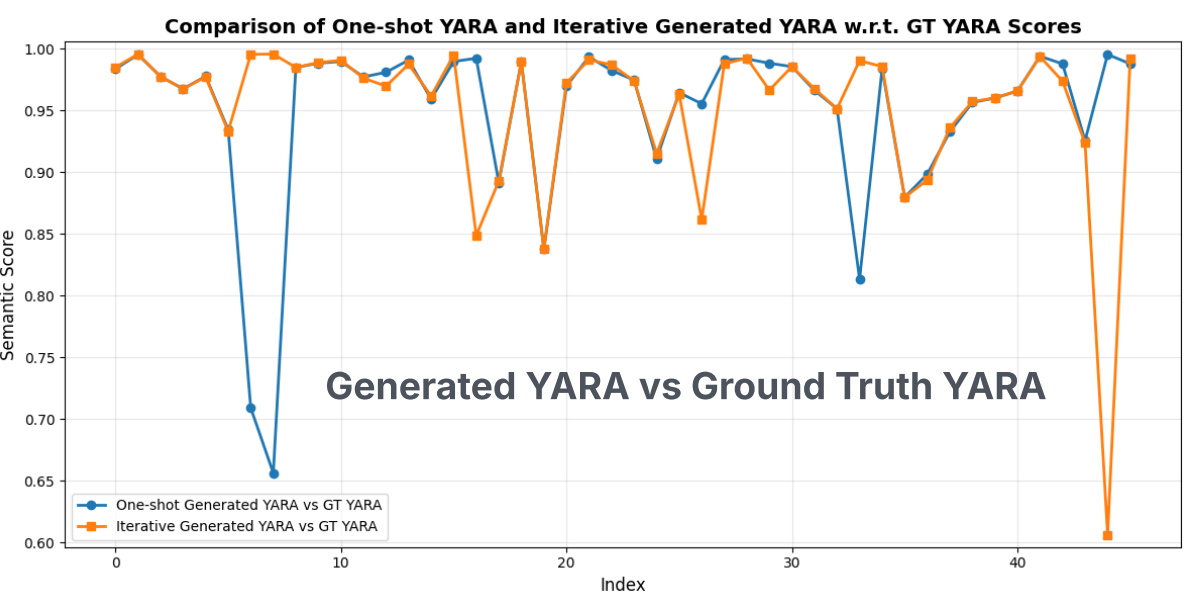
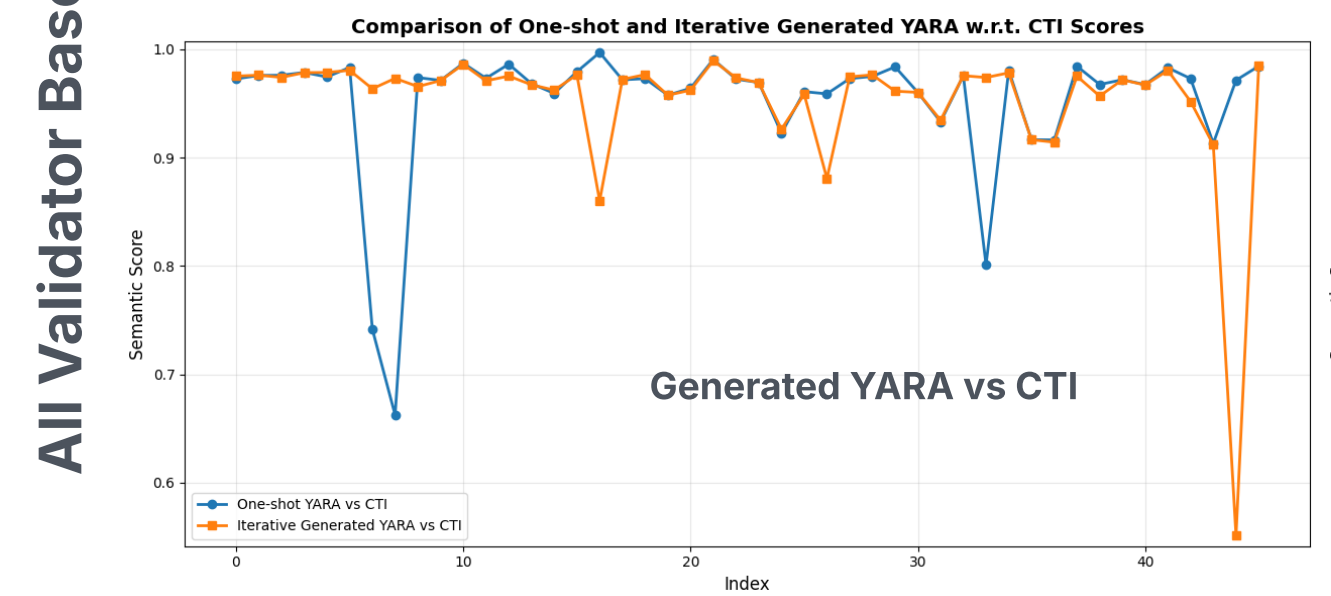
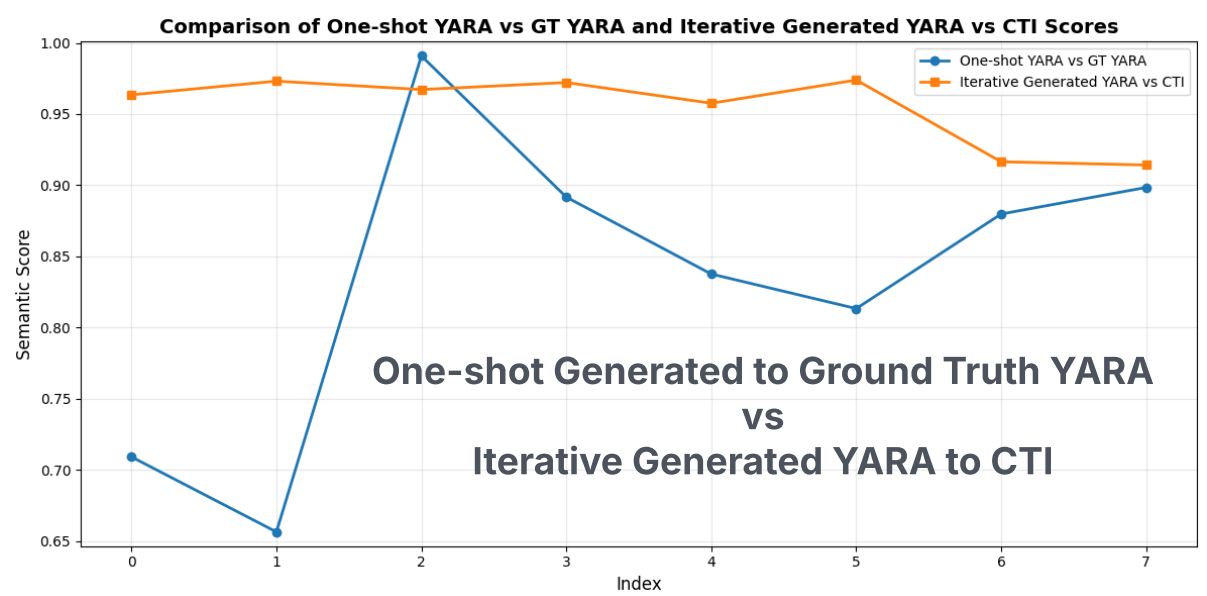
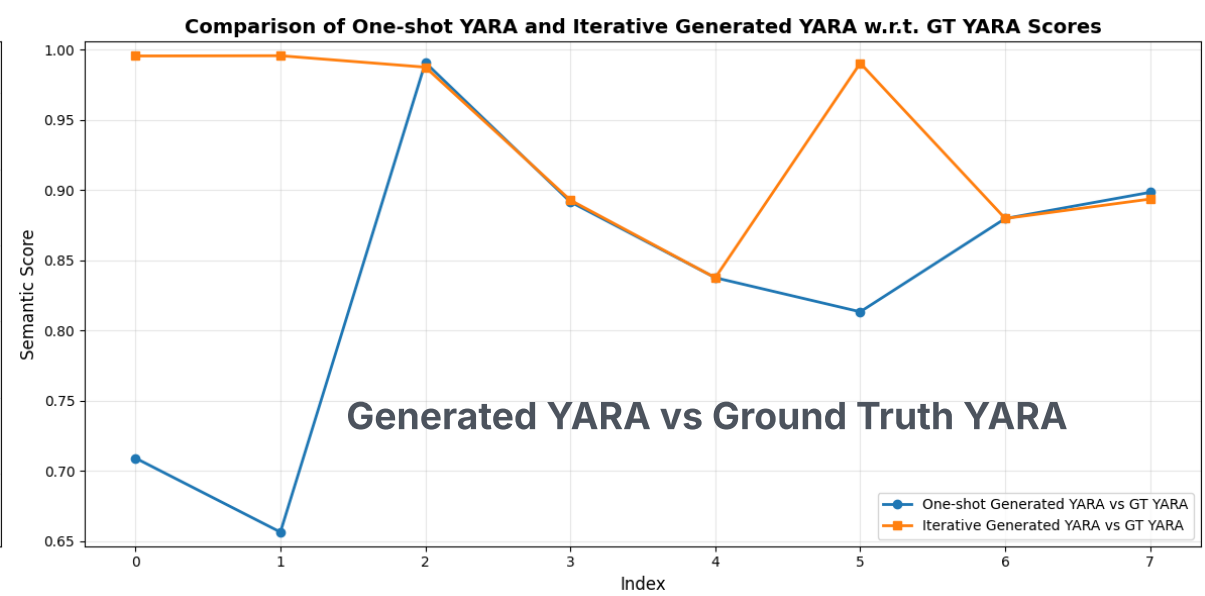
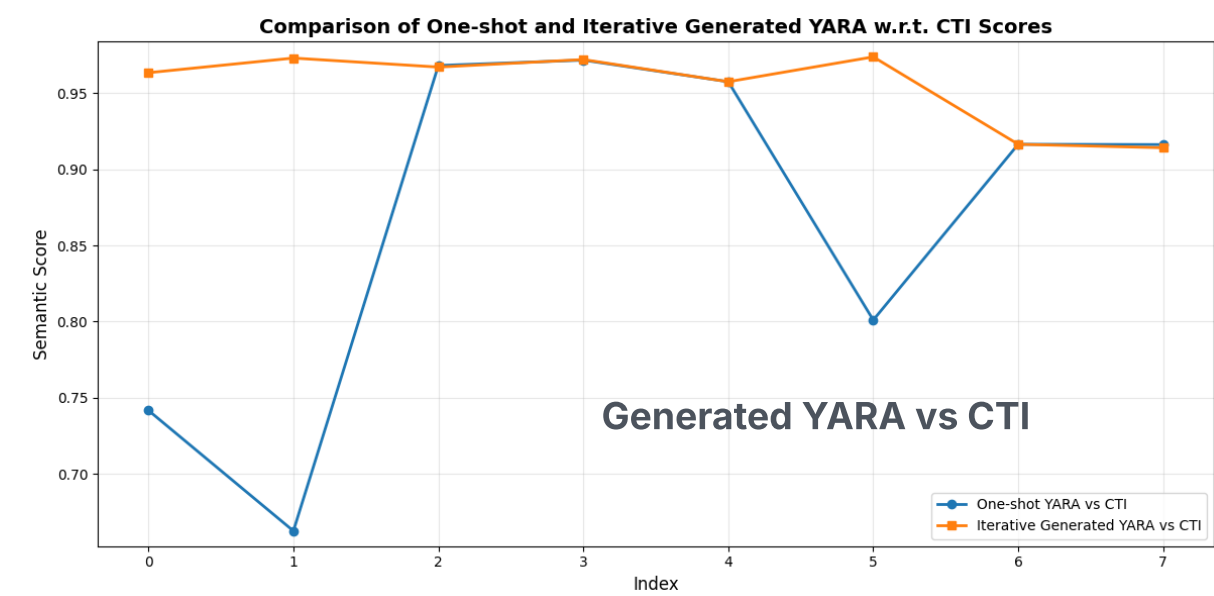
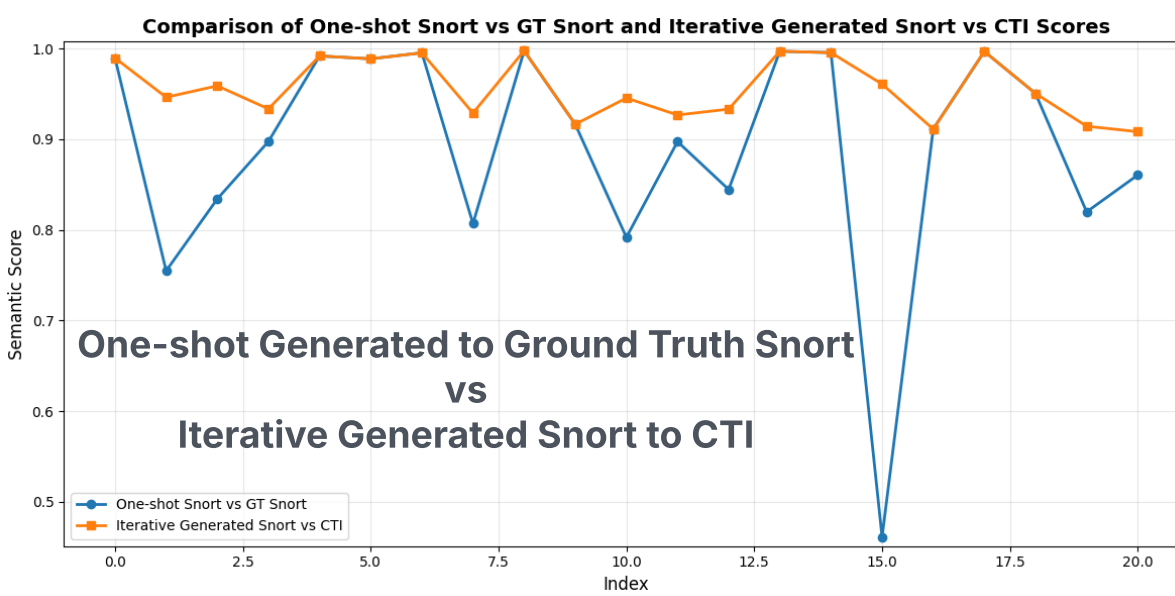
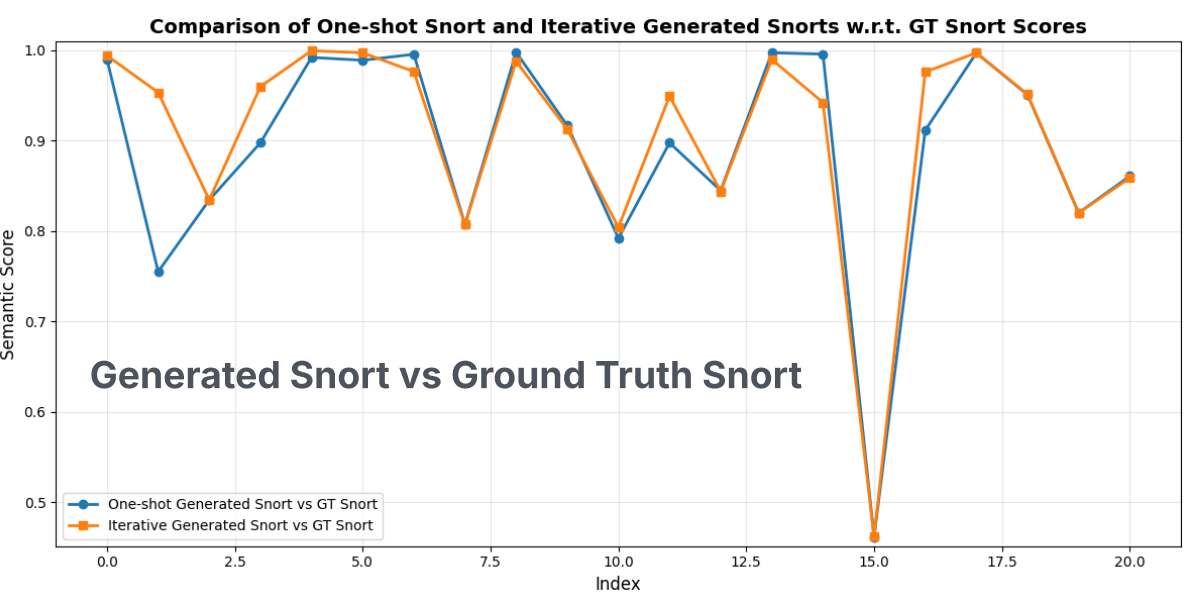
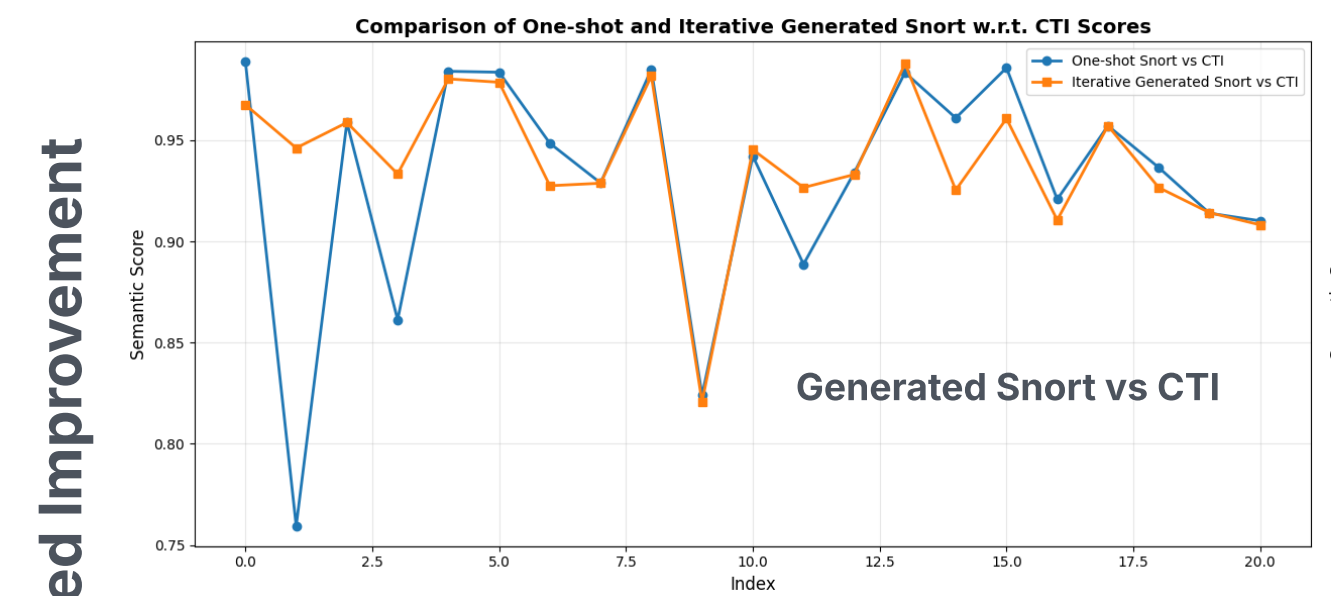
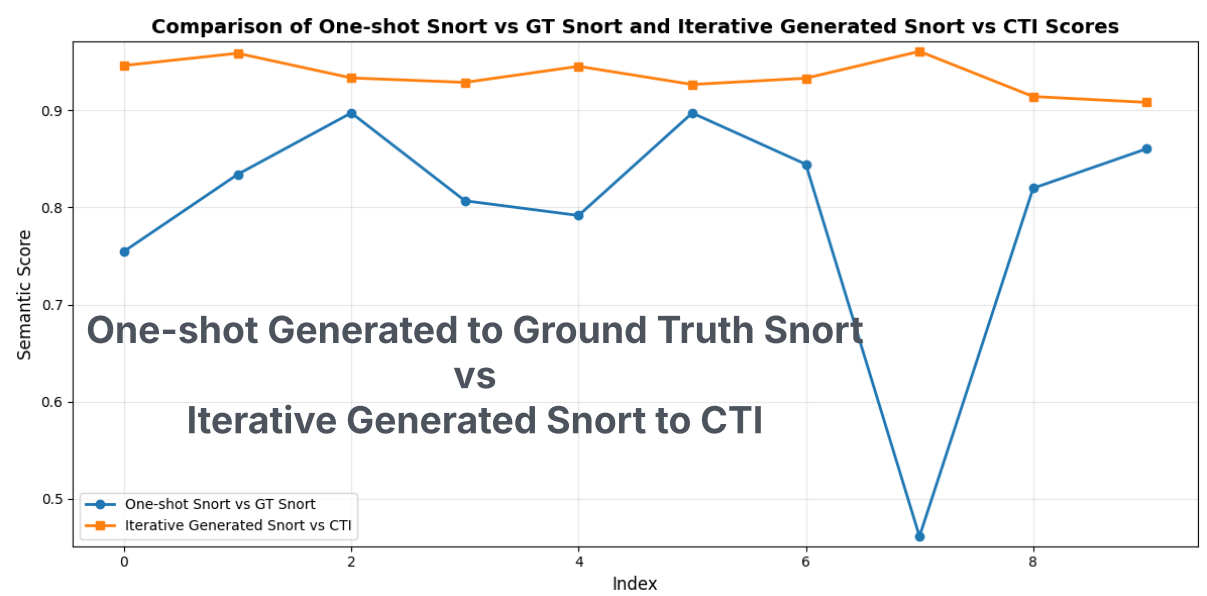
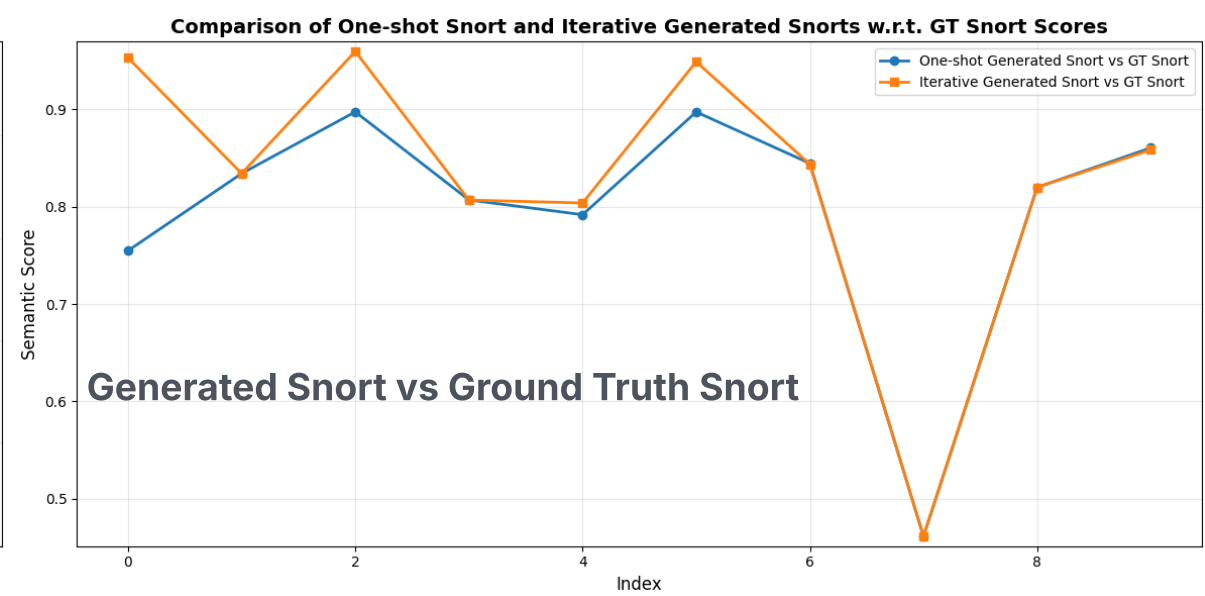
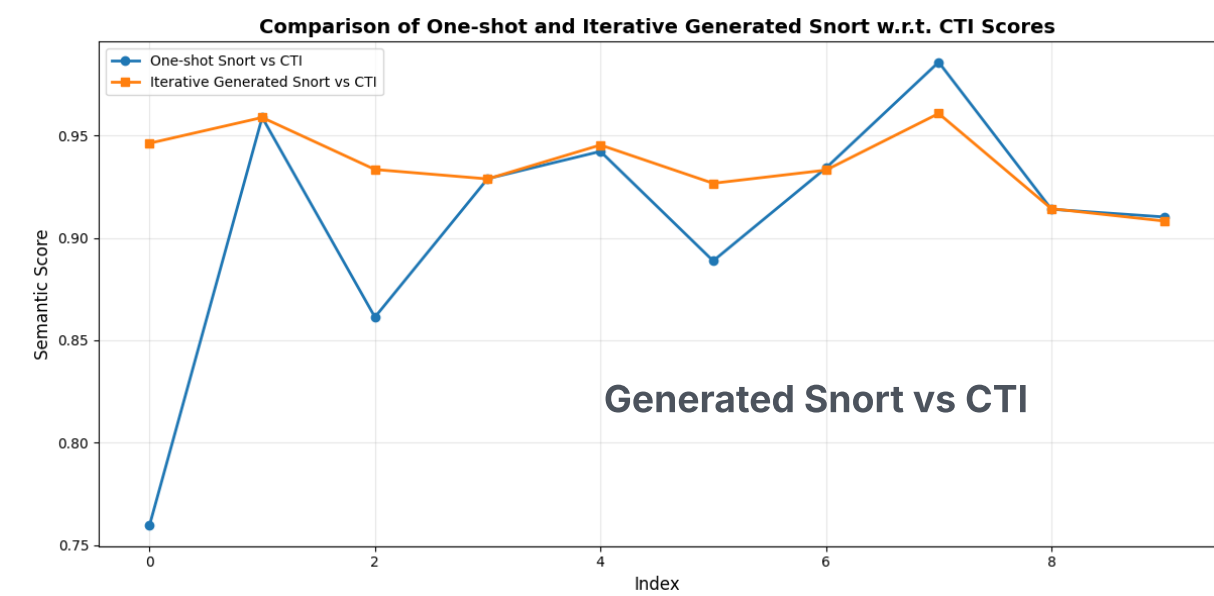
All Validator Based Improvement

Semantic Validator Based Improvement



All Validator Based Improvement

Semantic Validator Based Improvement



All Validator Based Improvement

Granite-3.3-8B-Instruct | One-Shot (Blue) vs Iterative with Validator Feedback (Orange)

Mistral-24B-Instruct-2501 | One-Shot (Blue) vs Iterative with Validator Feedback (Orange)

Granite-3.3-8B-Instruct | One-Shot (Blue) vs Iterative with Validator Feedback (Orange)