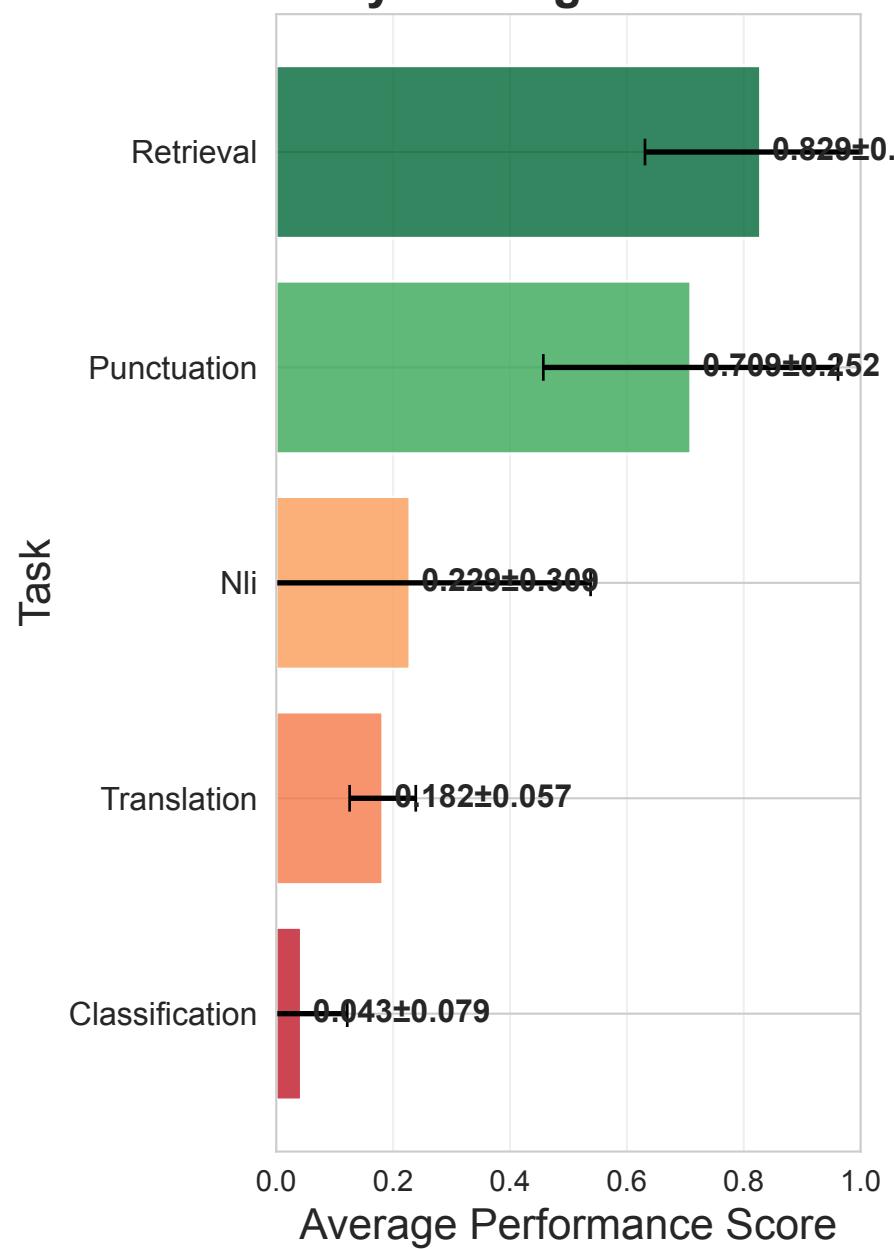


Task Difficulty: Average Model Performance



Task Difficulty: Performance Range Across Models

