Inter Annotator Agreement Results

- < 0: No agreement (worse than chance)
- 0.01-0.20: Slight agreement
- 0.21-0.40: Fair agreement
- 0.41-0.60: Moderate agreement
- 0.61-0.80: Substantial agreement
- 0.81-1.00: Almost perfect agreement

We use Cohen's kappa for binary variables, and Weighted Cohen's kappa for ordinal variables.

Annotator1 & Annotator2

- **Twitter dataset**

Cohen's Kappa for emotional appeal: 0.354
Cohen's Kappa for audience adaptation: 100%
Weighted Kappa for clarity: 0.438
Weighted Kappa for evidence: 0.503
Weighted Kappa for rebuttal: 0.306
Weighted Kappa for fairness: 0.555

- **CONAN dataset**

Cohen's Kappa for emotional appeal: 0.520
Cohen's Kappa for audience adaptation: 100%
Weighted Kappa for clarity: 0.580
Weighted Kappa for evidence: 0.704
Weighted Kappa for rebuttal: 0.316
Weighted Kappa for fairness: 0.741

Annotator1 & Annotator3

- **Twitter dataset**

Cohen's Kappa for emotional appeal: 0.505
Cohen's Kappa for audience adaptation: 0.657 // 99%
Weighted Kappa for clarity: 0.557
Weighted Kappa for evidence: 0.193
Weighted Kappa for rebuttal: 0.847
Weighted Kappa for fairness: 0.723

- **CONAN dataset**

```
Cohen's Kappa for emotional appeal: 0.243
Cohen's Kappa for audience adaptation: 100%
Weighted Kappa for clarity: 0.063
Weighted Kappa for evidence: 0.219
Weighted Kappa for rebuttal: 0.117
Weighted Kappa for fairness: 0.572
```

- **Twitter dataset**

```
Cohen's Kappa for emotional appeal: 0.362
Cohen's Kappa for audience adaptation: 99%
Weighted Kappa for clarity: 0.347
Weighted Kappa for evidence: 0.213
Weighted Kappa for rebuttal: 0.281
Weighted Kappa for fairness: 0.401
```

- **CONAN dataset**

```
Cohen's Kappa for emotional appeal: 0.184
Cohen's Kappa for audience adaptation: 100%
Weighted Kappa for clarity: 0.301
Weighted Kappa for evidence: 0.117
Weighted Kappa for rebuttal: 0.343
Weighted Kappa for fairness: 0.484
```

Overall agreement with Weighted Kappa

- **Twitter dataset**

```
Average Weighted Kappa for emotional_appeal: 0.408
Average Weighted Kappa for audience_adaptation: 0.219 // 99%
Average Weighted Kappa for clarity: 0.448
Average Weighted Kappa for evidence: 0.304
Average Weighted Kappa for rebuttal: 0.478
Average Weighted Kappa for fairness: 0.560
```

- **CONAN dataset**

```
Average Weighted Kappa for emotional_appeal: 0.316
Average Weighted Kappa for audience_adaptation: 100%
Average Weighted Kappa for clarity: 0.315
Average Weighted Kappa for evidence: 0.347
Average Weighted Kappa for rebuttal: 0.259
Average Weighted Kappa for fairness: 0.599
```