

Confusion Matrices - gpt-3.5-turbo (OpenAI)
Rows: Human verdict | Cols: Judge verdict | Φ = FaviScore

