

Confusion Matrices - gpt-4o-mini (OpenAI)
Rows: Human verdict | Cols: Judge verdict | Φ = FaviScore

