

Confusion Matrices - llama2-70b (Meta)  
Rows: Human verdict | Cols: Judge verdict |  $\Phi$  = FaviScore

