

Fp-value of Dunn test for  $\delta$  (order 9)

|          |         |       |       |       |       |       |       |
|----------|---------|-------|-------|-------|-------|-------|-------|
| GPT-3.5  | 1.00    | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  |
| GPT-4    | 0.00    | 1.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  |
| Human    | 0.00    | 0.00  | 1.00  | 1.00  | 1.00  | 1.00  | 1.00  |
| LaMA-13B | 0.00    | 0.00  | 1.00  | 1.00  | 1.00  | 0.89  | 1.00  |
| LaMA-30B | 0.00    | 0.00  | 1.00  | 1.00  | 1.00  | 1.00  | 1.00  |
| LaMA-65B | 0.00    | 0.00  | 1.00  | 0.89  | 1.00  | 1.00  | 0.06  |
| LLaMA-7B | 0.00    | 0.00  | 1.00  | 1.00  | 1.00  | 0.06  | 1.00  |
|          | GPT-3.5 | GPT-4 | human | A-13B | A-30B | A-65B | MA-7B |