

p-value of Dunn test for γ (order 3)

| | GPT-3.5 | GPT-4 | Human | LaMA-13B | LaMA-30B | LaMA-65B | LLaMA-7B |
|----------|---------|-------|-------|----------|----------|----------|----------|
| 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 | 0.85 | 0.01 | 1.00 |
| LaMA-13B | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.65 | 1.00 |
| LaMA-30B | 0.00 | 0.00 | 0.85 | 1.00 | 1.00 | 1.00 | 1.00 |
| LaMA-65B | 0.00 | 0.00 | 0.01 | 0.65 | 1.00 | 1.00 | 0.08 |
| LLaMA-7B | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.08 | 1.00 |
| | GPT-3.5 | GPT-4 | Human | LaMA-13B | LaMA-30B | LaMA-65B | LLaMA-7B |