

Pair	Total	Zeros	Light	Moderate	High	Mean	Median
Gemini ↔ GPT-4o	120	35	61	16	8	1.50	1.0
Gemini ↔ Grok	120	66	44	4	6	0.91	0.0
GPT-4o ↔ Grok	120	52	50	9	9	1.27	1.0

Table S2. Pairwise comparison of maximum indicator scores across all speeches. “Zeros” indicate perfect agreement (0-point difference), “Light” small discrepancies (1–2 points), “Moderate” mid-range divergences (3–4 points), and “High” substantial discrepancies (≥5 points). Mean and median summarize overall divergence per model pair. Results show that Gemini and Grok align most closely (mean ≈ 0.9), while GPT-4o exhibits slightly higher but still light-range differences when compared with both models (mean ≈ 1.3–1.5).