

**Supplementary Table 2.** Model Diagnostics for Latent Dirichlet Allocation Models across 10 Random Seeds

| No. of topics (k) | Mean perplexity | Mean log-likelihood |
|-------------------|-----------------|---------------------|
| 3                 | 516             | -179,377            |
| 4                 | 489             | -172,389            |
| 5                 | 467             | -167,469            |
| 6                 | 448             | -163,585            |
| 7                 | 434             | -160,404            |
| 8                 | 420             | -157,985            |
| 9                 | 408             | -155,653            |
| 10                | 399             | -153,656            |

Values are means across 10 random seeds. Mean semantic coherence for k=3–10 was -.404, -.423, -.427, -.425, -.432, -.434, -.434, and -.438, respectively; mean cross-seed stability was .488, .440, .445, .464, .405, .387, .399, and .345. The number of topics were selected by considering these diagnostics together with topic distinctiveness, adjacent-solution merging/fragmentation, and representative-document interpretability.