|  |  |
| --- | --- |
| Parameter | Effect |
| min\_cluster\_size=20 | Smaller value → more clusters, finer granularity |
| n\_neighbors=15 | UMAP setting; higher means smoother groups |
| n\_components=2 | Increase to 3 if using 3D visualization |
| min\_samples=5 | Higher → more conservative clustering (more noise) |

|  |  |
| --- | --- |
| Area | What You Can Fine-Tune |
| min\_cluster\_size, min\_samples in HDBSCAN | Get coarser or finer clusters |
| TF-IDF keyword extraction | Use full text instead of just intro |
| Cluster merging/splitting | Manually relabel or split noisy groups |
| Add metadata | Year, author country, citations |
| Visualization | Add keywords to legends or tooltips |
| Use BERTopic | Combine clustering + topic modeling for coherence |