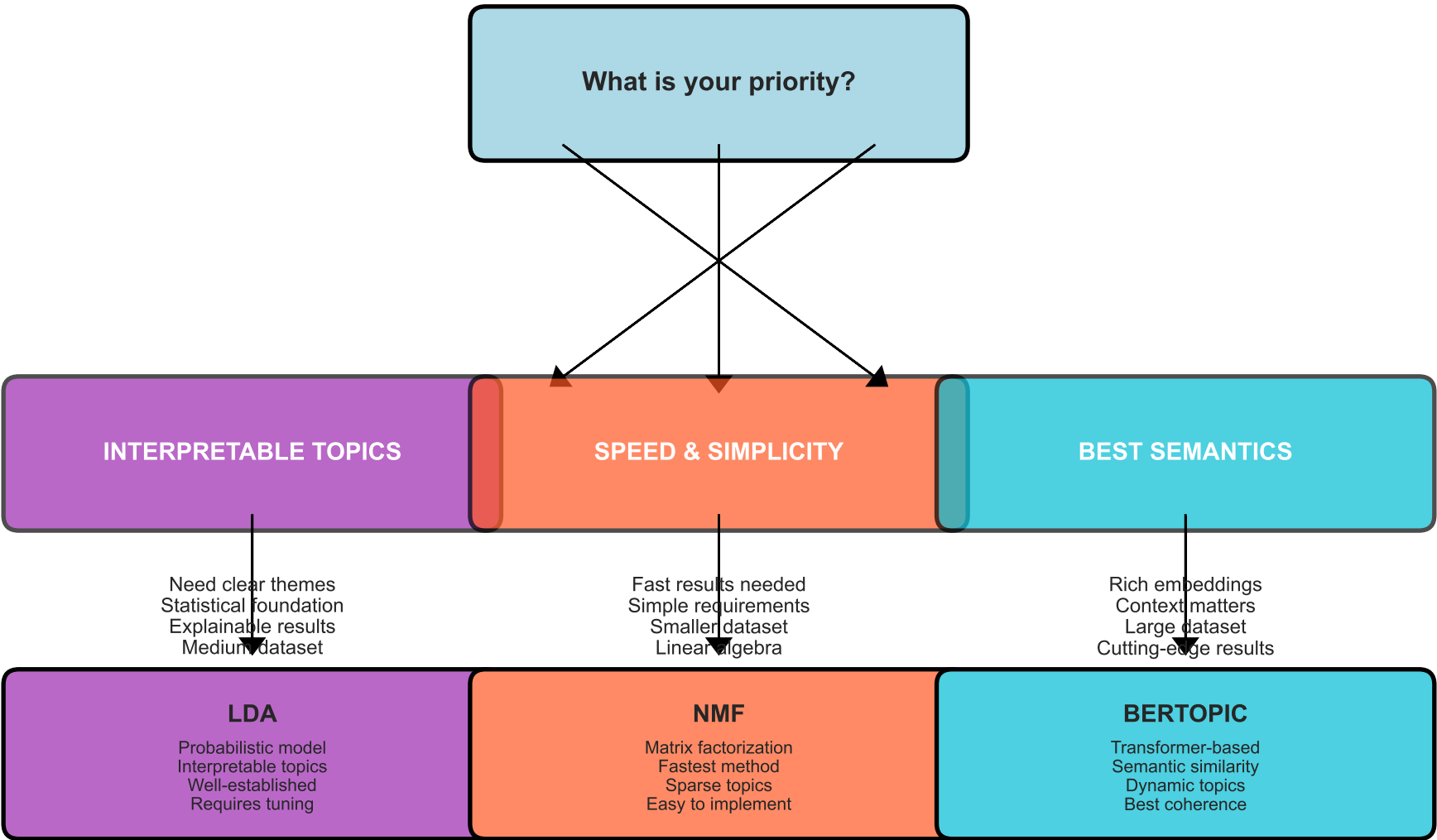


When to Use Which Topic Modeling Method: Decision Framework



Additional Considerations

Dataset Size: <1K docs → LSA/NMF simple; 1K-100K → LDA optimal; >100K → BERTopic scalable
Topic Count: Know K → LDA/NMF; Discover K → Hierarchical or BERTopic clustering
Languages: Multilingual → BERTopic with mBERT; Single language → Any method works
Real-time: Streaming topics → Online LDA; Batch analysis → Any method suitable
Coherence: Need coherent topics → BERTopic (best); LDA with tuning; NMF varies
Computation: Limited resources → NMF (fastest); GPU available → BERTopic; Medium → LDA

Principle: LDA for interpretable topics, NMF for speed, BERTopic for best coherence and modern semantics