Many extensive research studies have focused on long text TM. The traditional long text TM models such as Latent Dirichlet Allocation (LDA) (Blei et al. [2003](https://link.springer.com/article/10.1007/s10462-022-10254-w#ref-CR23)), Latent Semantic Analysis (LSA) (Deerwester et al. [1990](https://link.springer.com/article/10.1007/s10462-022-10254-w#ref-CR47)), Probabilistic Latent Semantic Analysis (PLSA) (Hofmann [1999](https://link.springer.com/article/10.1007/s10462-022-10254-w#ref-CR84)), and Non-negative Matrix Factorization (NMF) (Lee and Seung [2001](https://link.springer.com/article/10.1007/s10462-022-10254-w#ref-CR108)) are popular for discovering latent semantic topic structures in long texts. They do not require prior annotation or labelling processes. These traditional topic models visualize each text and document as a mixture of different themes (topics) which are distributed over the text. They employ statistical approaches such as variational techniques and Gibbs sampling to infer the trending topics of each document through complex order word co-occurrence patterns (Ostrowski [2015](https://link.springer.com/article/10.1007/s10462-022-10254-w#ref-CR175)).