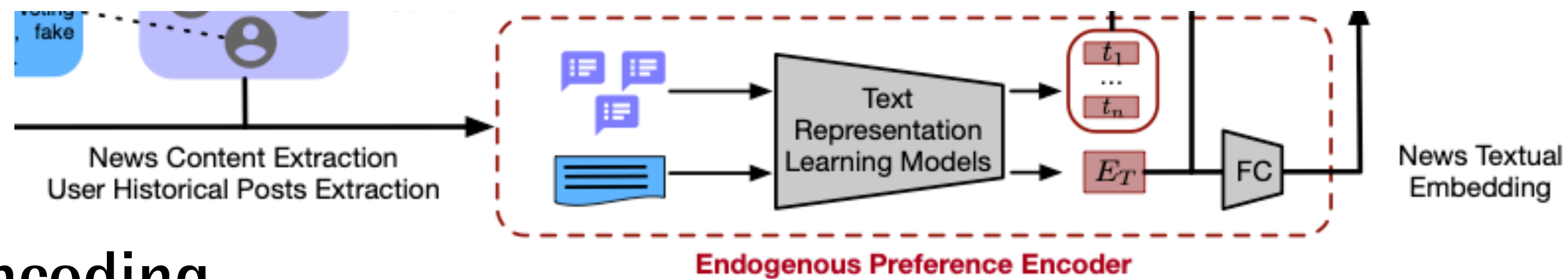


Approach

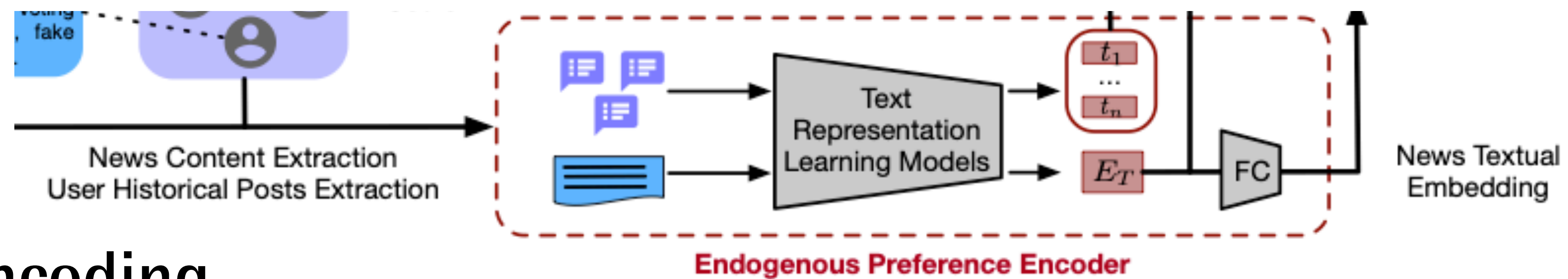
Endogenous Preference Encoding



- For inaccessible users (suspended or deleted account), use randomly sampled tweets from accessible users engaging the same news as its corresponding historical posts.
- Because deleting the inaccessible user will break the intact news propagation cascading and result in a less effective exogenous context encoder.
- Also remove the special characters (e.g. @) and urls, before applying text representation learning methods.

Approach

Endogenous Preference Encoding



- To encode the news textual information and user preferences, employ two types of text representation learning approaches based on language pertaining.
 - word2vec: choose the 680k 300 dimensional vectors pretrained by spaCy
 - BERT: employ pretrained embeddings (BERT-large) using bert-as-a-service
- Instead of training on the local corpus, the word embedding pretrained on large corpus are supposed to encode more semantic similarities between different words and sentences.