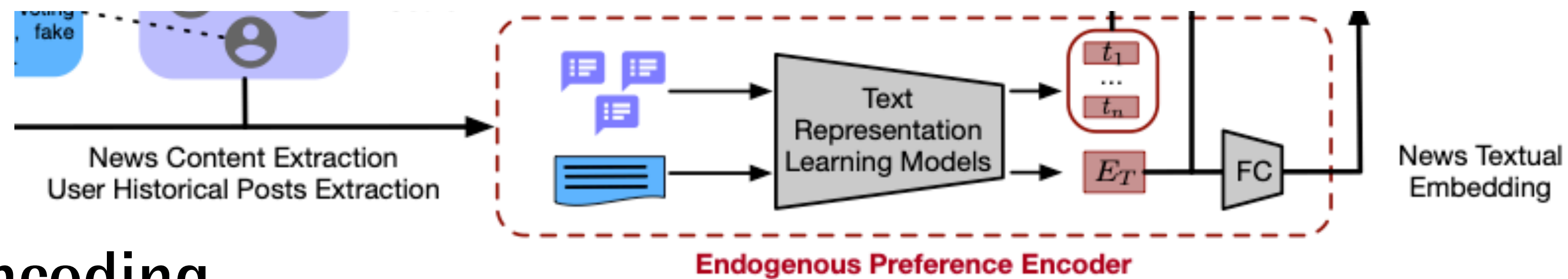


# Approach

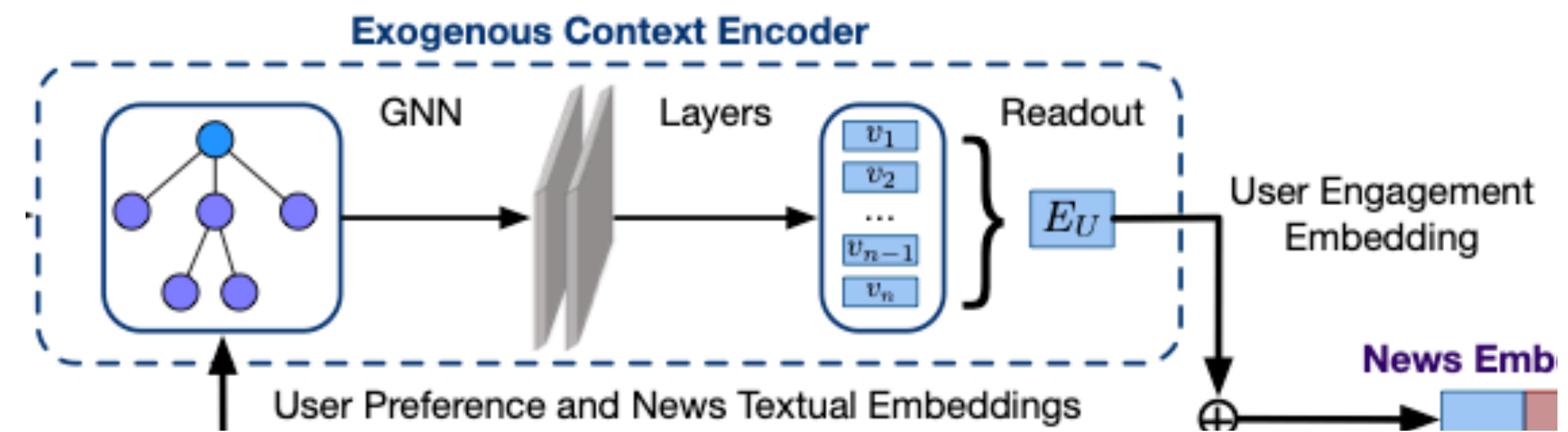
## Endogenous Preference Encoding



- word2vec (spaCy)
  - Average the vectors of existing words in combined recent 200 tweets to get user preference representation.
  - The news textual embedding is obtained similarly.
- BERT (BERT-large)
  - Due to BERT's input sequence length limitation (512 tokens), couldn't use BERT to encode 200 tweets as one sequence, so authors resort to encode each tweet separately and average them afterward to obtain a user's preference representation.
  - Generally, the tweet text is way shorter than the news text, authors empirically set the max input sequence length of BERT as 16 tokens to accelerate the tweets encoding time

# Approach

## Exogenous Context Extraction



- Given a news piece on social media, the user exogenous context is composed of all users that engaged with the news.
- Utilize the retweet information of news pieces to build a news propagation graph.
- The root node represents the news pieces, and other nodes represent users who share the root news.
- Define a new piece as  $v_1$ , and  $\{v_2, \dots, v_n\}$  as a list of users that retweeted  $v_1$  ordered by time.