DSCI 510 Final Project – APPENDIX
Variable list and explanation

Variable Name in py files	**Variable definition**	**Variable source**
mid	Unique numbe of each post	Raw data without cleaning
uid	User ID	Raw data without cleaning
Comments_count/likes/reposts	Number of comments/likes/reposts	Raw data without cleaning
Text	Post text content	Translated from original posts with Google translate api
textlength	Number of words in a text	Counted based on the text splitting result of Spacy
Moral_score	A measurement considering bot moral and immoral judgement words into accounr	Computed through constructing a semantic graph of all tokens in a post. Moral_score = centrality of the token in a graph * the moral label of that word in the moral judgement dictionary (1 for moral, -1 for immoral, 0 for other)
Moral_wc	Number of moral judgement word	Counted based on dictionary developed by SoCo Lab https://aclanthology.org/2020.ccl-1.50/
Immoral_wc	Number of immoral judgement word	Counted based on dictionary developed by SoCo Lab https://aclanthology.org/2020.ccl-1.50/
Positive_prob	Sentiment score measuring the probability of a text expressing positive emotions	Calculated by calling SnowNLP methods
Average_commenter_follow	Average number of following accounts of all commenters below a source post	= Sum(commenter following number)/comment count
Average_commenter_follower	Average number of followers of all commenters below a source post	= Sum(commenter follower number)/comment count
Average_moral_word	Average number of moral judgement words of all comments below a source post	= Sum(moral word count of each comment)/comment count
Average_immoral_word	Average number of immoral judgement words of all comments below a source post	=Sum(immoral word count of each comment)/comment count

Commenter_gender_ratio	Proportion of male commenters among all below a source post	= count(male commenter)/count(female commenter)
Positive_negative_ratio	Ratio of positive comments and negative ones below a source post	Define positive probability > 50% as a positive comment, and below 50% as negative comment. Divide positive comment number by the negative comment number.
Average_text_lenth	Average number of Chinese words below a source post	Counted based on Spacy word splitting.
Average_moral_score	Average moral score of all comments below a source post	Computed with same method with moral score in source post for each comment and get the mean value.
Keyword_code	The keywords used to get every source post, indicating the general topic of posts and comments	Categorical variable with 7 levels. a) Openai b) AI risk c) Post deletion d) Sexual assault e) Homeless dogs hurt f) Cat abusing g) Du Meizhu scandal