

Code script	purpose	Generated documents	Generated graphics
1-MasterThesis-Videos	Collecting videos and their statistics	video_index_edited.csv	
2-MasterThesis_Comments	Collecting videos' comments	comments	
3-MasterThesis_Data_description	Creating data description	all_comments.csv	Figure 1, 2, 3, 4, 5
4-MasterThesis_BERTopic	Filtering text and modeling topics	filtered_data.csv topics_data.csv	Figure 6, 7, 8, 9, A Table 1, 2
5-MasterThesis_Shuff	Random sampling training set	train_set.xlsx train_set05.csv	Table 3 Figure 10
6-MasterThesis_FineBERT	Fine-tuning BERT model and predicting labels of comments	BERTmodel all_comments_with_labels.csv	Figure 11, 12, B Table 4, 5, 6
7-MasterThesis_Manifesto	Measuring viewpoints and creating a document for R	Videos_index_forR.csv	Figure 13, 14
8-MasterThesis-Regression	Running linear regression		Table 7 Figure C, D, E, F

1. The Raw Data folder stores the raw data crawled through the API, and the Refine Data folder stores the processed files that are passed on to the next step.
2. The train\_set05.csv is the manually labeled training set.
3. The BERTmodel folder stores the fine-tuned BERT model.
4. The fine-tuning BERT model script runs on an A100 40GB, and the fine-tuning process requires about 34GB of graphics memory.
5. The first seven scripts are all Python programming. The last regression script is an R markdown document.