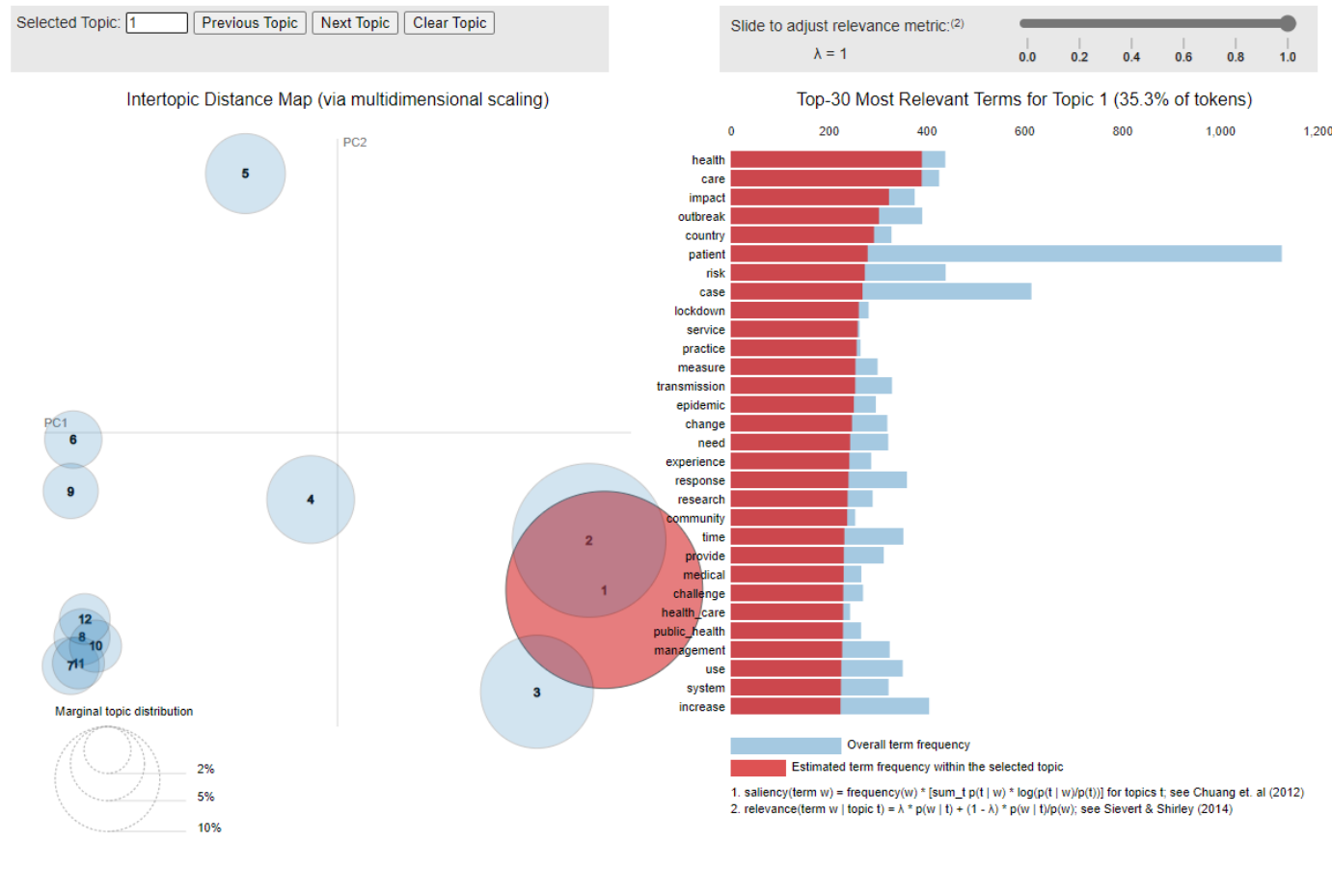




QUESTION ANSWERING MODEL FOR COVID 19 DATASET

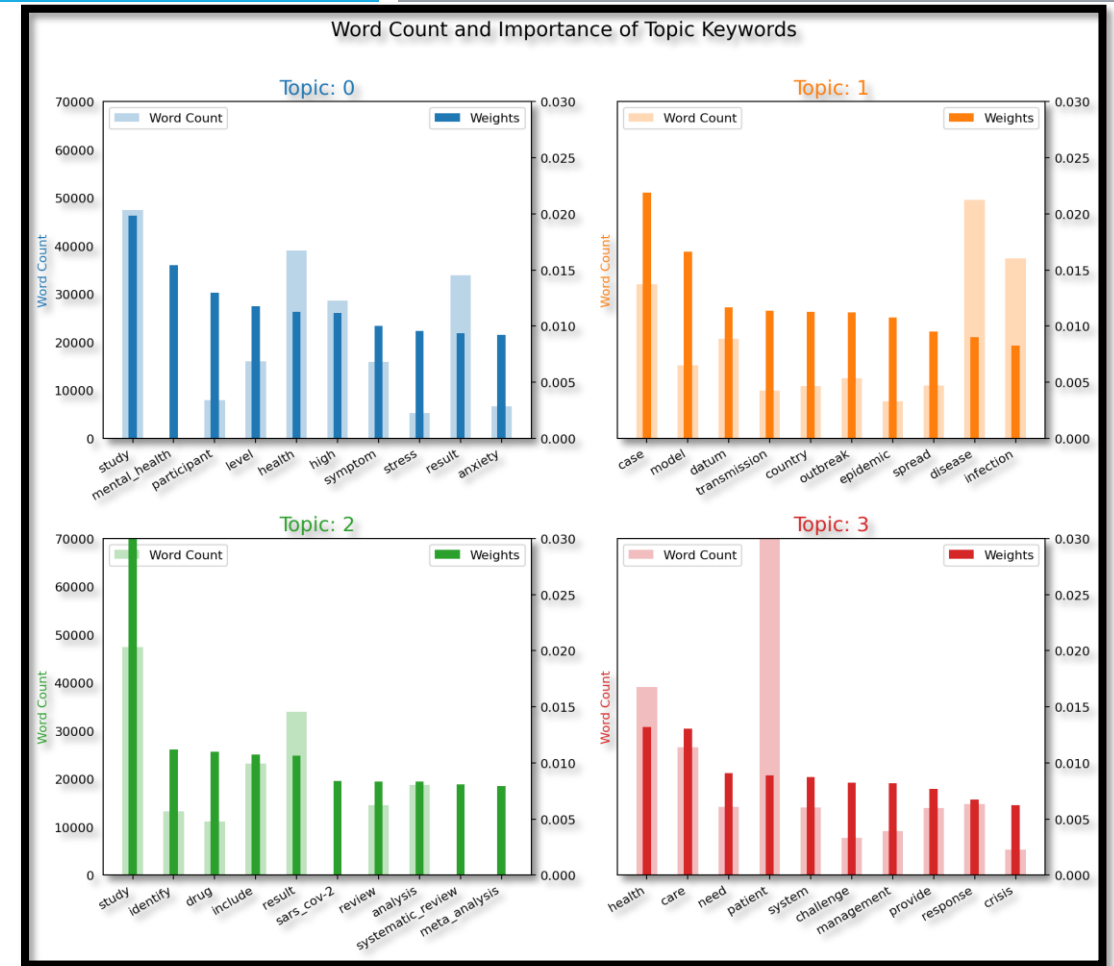
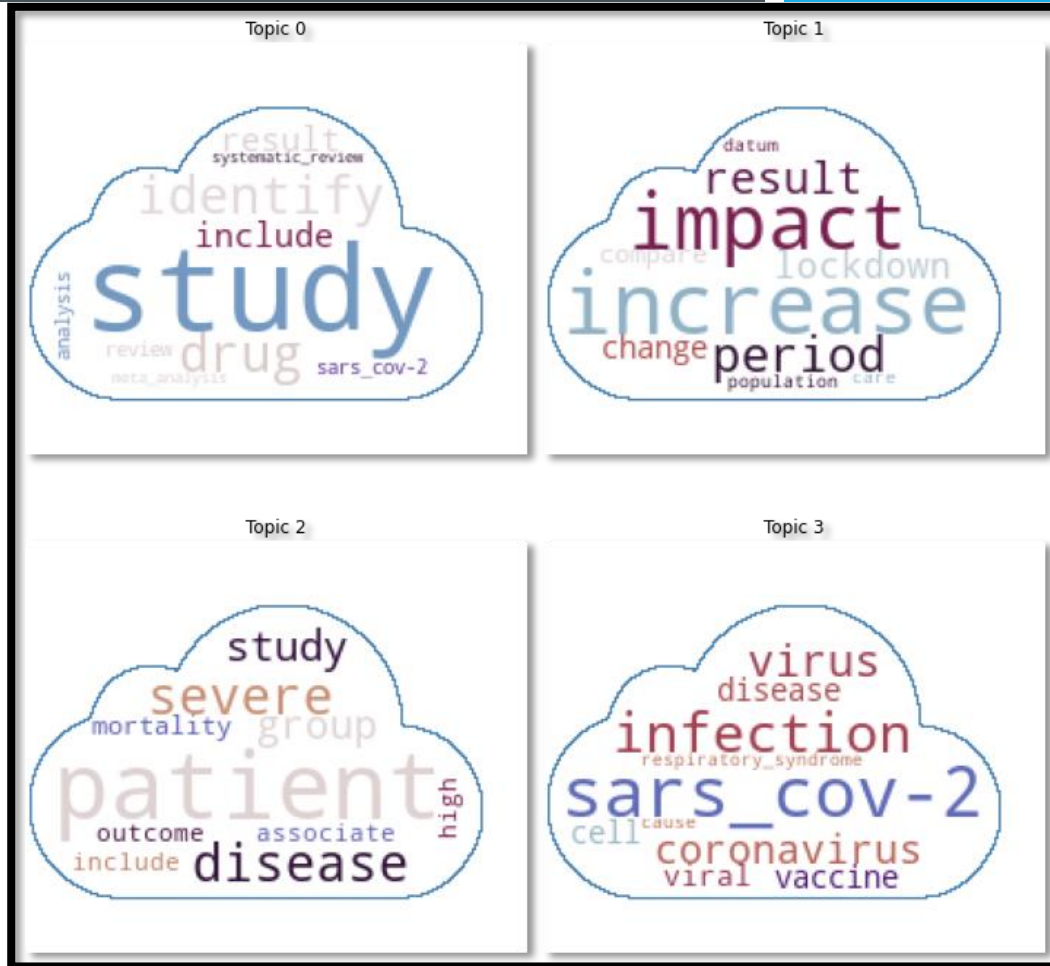
EXPLORATORY ANALYSIS OF COVID ARTICLES DATASET



	topic1	topic2	topic3
0	sars_cov-2	copd	model
1	cell	brain	detection
2	protein	mirna	assay
3	virus	mouse	predict
4	drug	endothelial_cell	sample
5	vaccine	college_student	prediction
6	human	ace2_tmprss2	main_protease
7	ace2	expression_ace2	performance
8	viral	neurological_symptom	dataset
9	target	olfactory_dysfunction	2019-ncov

- Topic modeling to gain insight about data and the most important topics covered in this huge dataset
- Table of keywords for each topic will help us to find most relevant keywords for each topic
- For instance, topic1 is probably related to developed drug, and vaccine for COVID

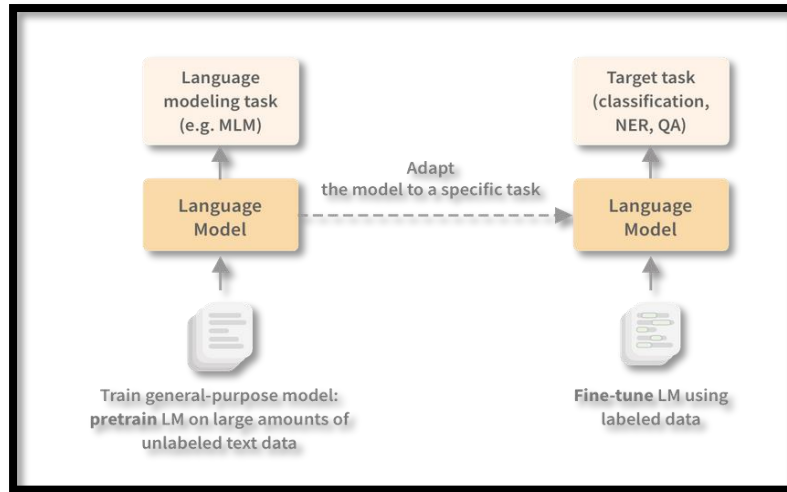
EXPLORATORY ANALYSIS OF COVID ARTICLES DATASET



- A word cloud with the size of the words proportional to the weight can give us a quick insight about data and keywords covered in each topic

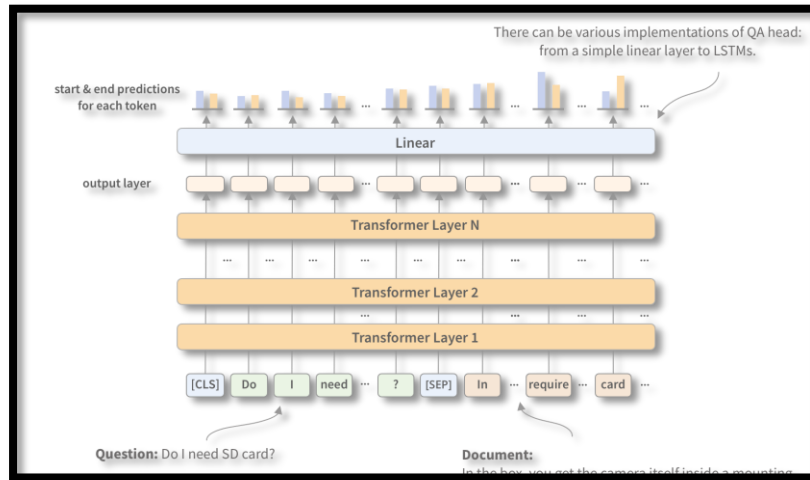
- When it comes to the keywords in the topics, the importance (weights) of the keywords matters.
- How frequently the words have appeared in the documents can give valuable insight to us.
- Words that occur in multiple topics and the ones with relative frequencies higher than the weight are often less important.

MODEL IMPLEMENTATION



Adapting the model to your domain

- New language models are capable of learning new context given new training texts



Question Answering Neural Network Architecture

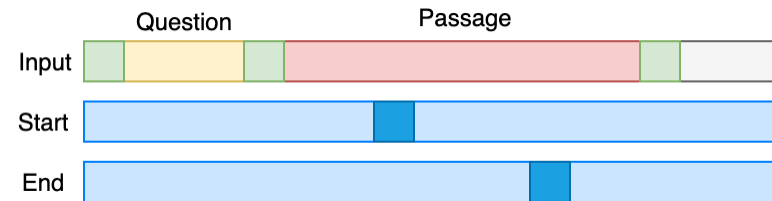
- Tokens are converted into word vectors by a language model and these vectors are used to generate logits

Question: Is smoking a risk factor

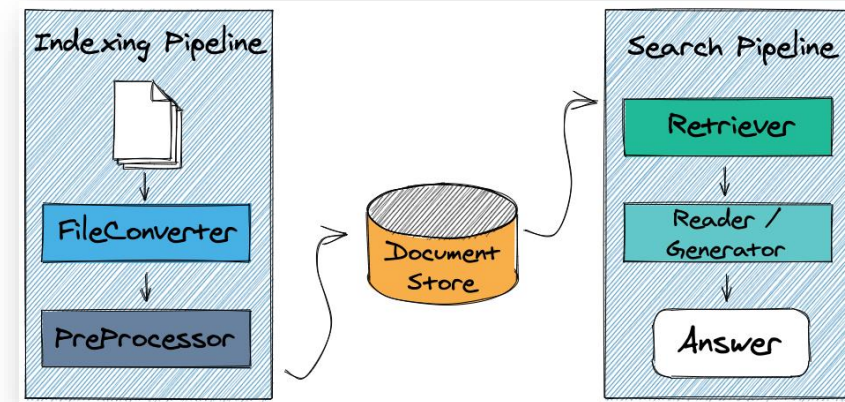
Answer: cigarette smoking has been proven to be a risk factor in the development of many diseases Research

Paper: smoking is correlated with the prognosis of coronavirus disease 2019 (covid-19) patients: an observational study

Context: background: cigarette smoking has been proven to be a risk factor in the development of many diseases. however, it remains controversial with respect score:14.78



- The positions with high logit values (dark blue) will likely be chosen as the start and end of the answer span.



Retrieval in a typical QA system

INSIGHTS AND CONCLUSION

Question: What are the public health mitigation measures that could be effective for control?

#1 Answer: social distancing and mandated face covering Research score:10.49

#2 Answer: wearing of face masks Research score:8.46

#3 Answer: viral testing together with contacts tracing, strict implementation of lockdown, and public cooperation Research score:8.34

#4 Answer: number of trips per day and the percentage of people wearing face masks Research score:8.00

#5 Answer: mobility restrictions, physical distancing, hygienic measures, socio-economic restrictions, communication and international support mechanisms score:5.68

Question: Is a pre-existing pulmonary disease a risk factor ?

#1 Answer: patients with preexisting pulmonary vascular diseases also appear at increased risk of morbidity and mortality score:9.93

#2 Answer: increased likelihood of mortality score:3.89

#3 Answer: it is thus unknown whether pulmonary arterial hypertension belongs to the risk factors of severe covid-19 disease. score:3.32

#4 Answer: there are no reports score:2.04

#5 Answer: there is an increased risk of having a pulmonary embolism during the sars-cov-2 pandemic, which affects patients with a different clinical profile and more often causes distal pulmonary embolisms. score:1.31

Question: What type of people are more likely to get infected?

#1 Answer: critical-type patients Research score:9.71

#2 Answer: critical-type score:9.09

#3 Answer: those who either themselves or a member of their household belonged to a vulnerable group score:3.98

#4 Answer: older score:2.60

#5 Answer: respiratory infectious score:2.13

Question: What are efforts targeted at a coronavirus vaccine ?

#1 Answer: development of effective vaccines score:12.84

#2 Answer: re-establishing public trust in government agencies score:9.73

#3 Answer: to support people and give them correct information about vaccines score:9.72

#4 Answer: vaccine hesitancy and foster vaccine confidence score:9.17

#5 Answer: psychological and behavioral interventions to boost vaccine efficacy and reduce side effects score:7.79