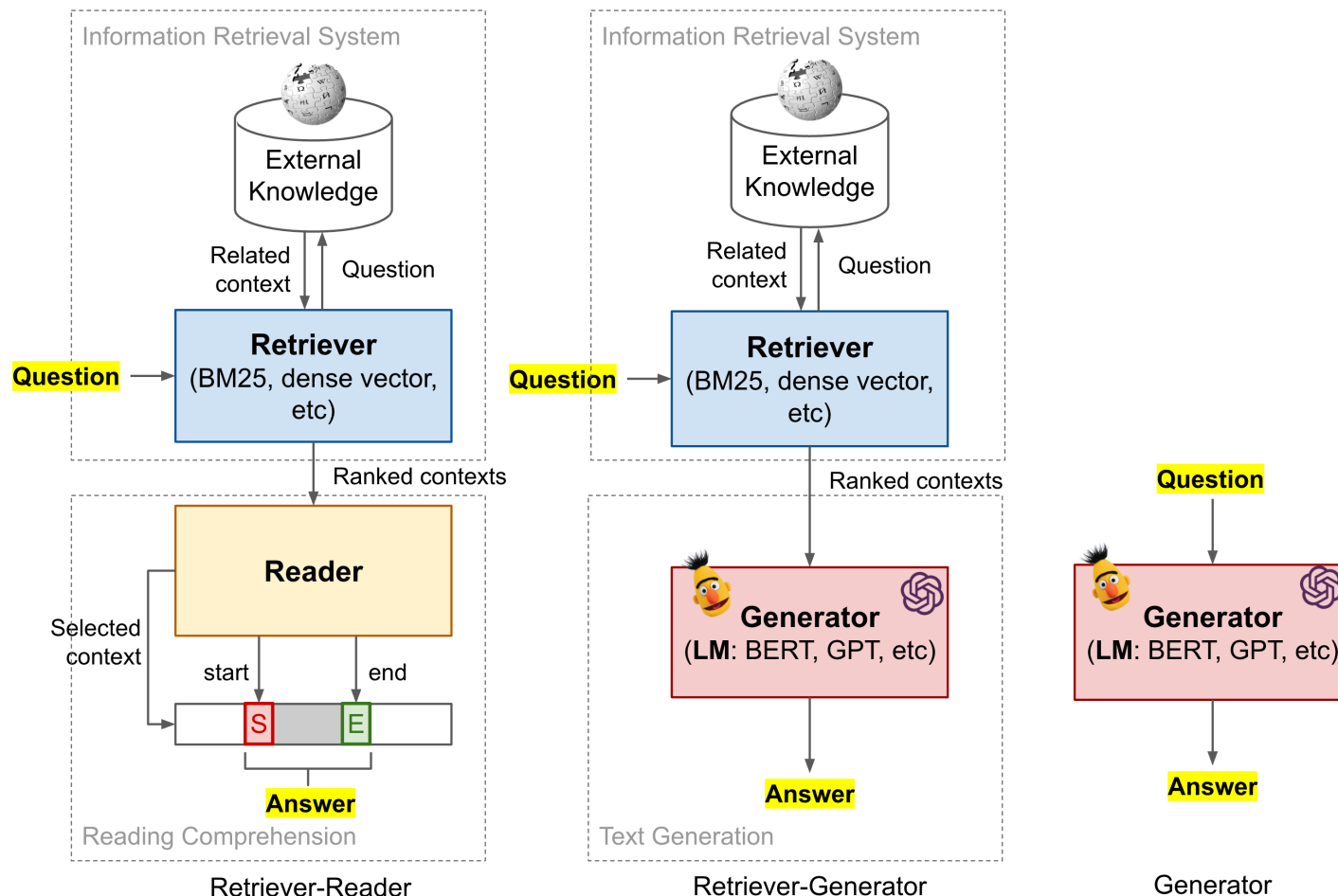# Readability Controlled Open-domain QA for COVID-19

Apple Siri Project
Zhehan, Lumingyuan and Xiao
Advisor: Gary Bécigneul
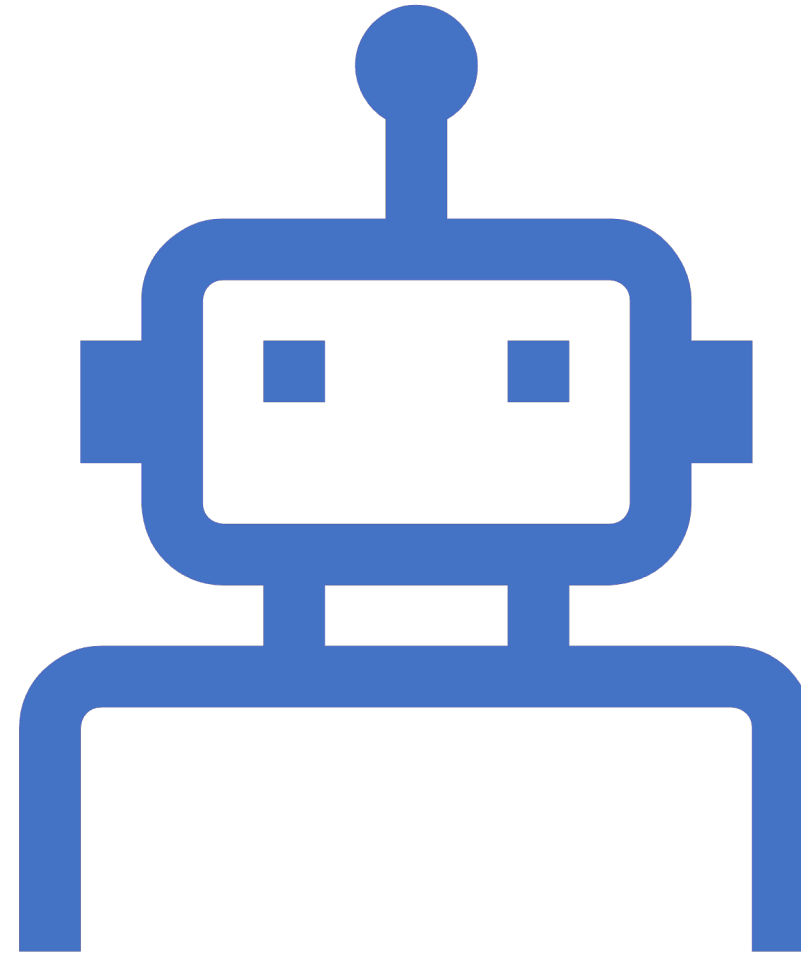
What is open domain QA?

- Open-domain QA is the task of answering factoid questions using a collection of general documents.

- Retrieve and comprehend one or more knowledge sources to get the correct answer.

- Regular method:
  - Retriever: TF-IDF, Bi-gram Hashing
  - Generator: BERT, GPT, etc.

Information Retrieval System

External Knowledge

Related context | Question

Retriever (BM25, dense vector, etc)

Question

Ranked contexts

Reader

Selected context

start | end

S | E

Answer

Reading Comprehension

Retriever-Reader

Information Retrieval System

External Knowledge

Related context | Question

Retriever (BM25, dense vector, etc)

Question

Ranked contexts

Generator (**LM**: BERT, GPT, etc)

Answer

Text Generation

Retriever-Generator

Question

Generator (**LM**: BERT, GPT, etc)

Answer

Generator

Covid-19 and its influence

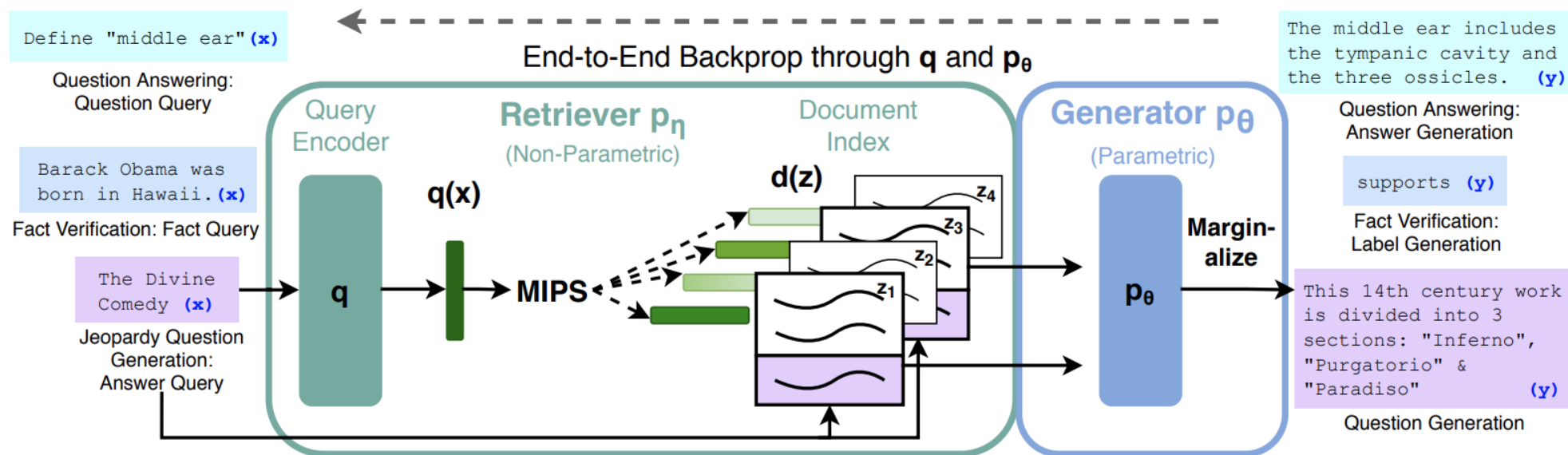WE MUST WORK TOGETHER TO DEFEND

† 2021-04-22

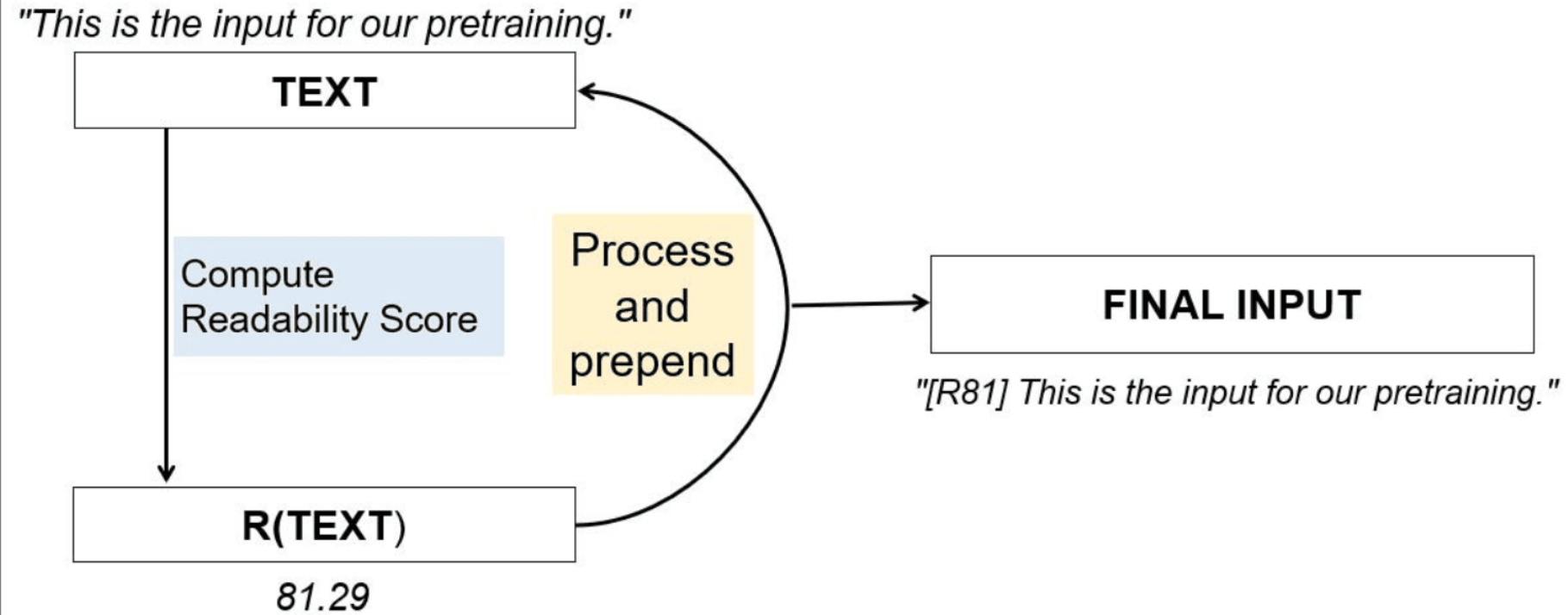# The need for a readability controlled QA model

A QA model which has ability to generate answers at different reading levels is required.

- Some important but simple knowledge such as the necessary of wearing medical masks must be known by general public.

- To some doctors or research scientists, they have requirement to get specific medical knowledge from the QA system.
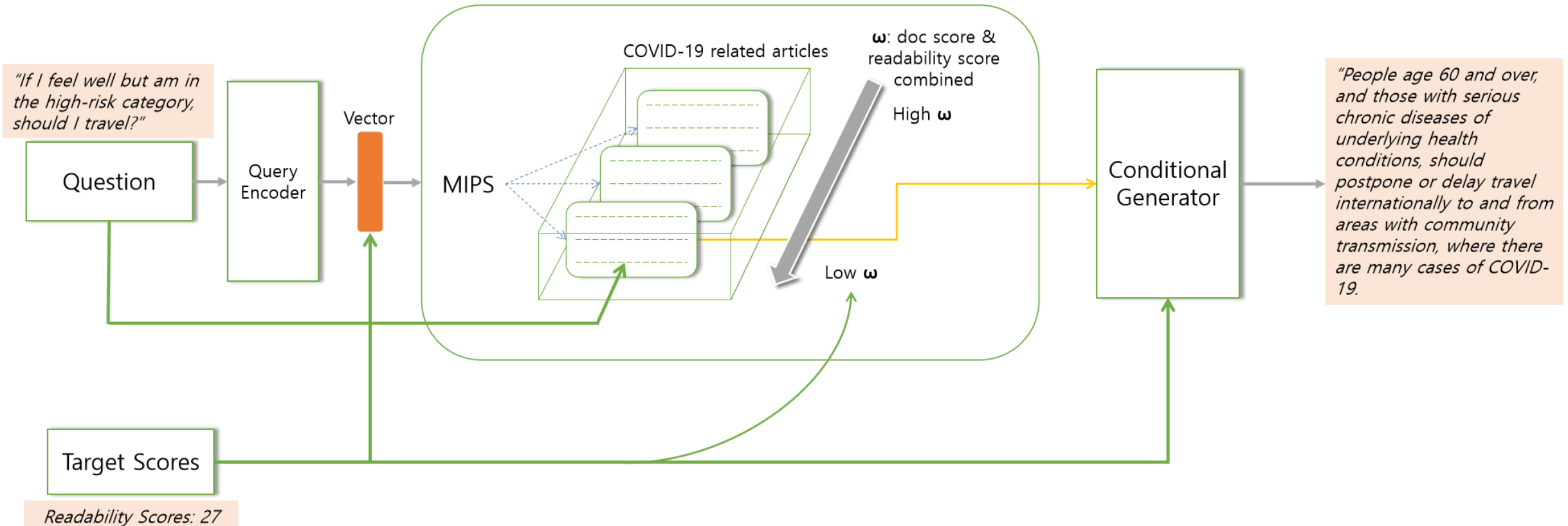
The introduction of RAG and CTRL

"This is the input for our pretraining."

TEXT

Compute
Readability Score

Process
and
prepend

FINAL INPUT

"[R81] This is the input for our pretraining."

R(TEXT)

81.29

The construction
of our model

Conditional BART

The construction of our model: RC-RAG

The building of RC-Conv Dataset

- In this dataset we extract the articles related with COVID-19 from the CORD-19 and Wikipedia dump to pretrain the Conditinal BART.

- Later we fine tune RC-RAG with the deepset and WHO QA pairs.

Training and evaluating

Compared with the origin RAG we have great progress in our benchmark.

We evaluate the model using the readability metrics

Conclusion

- The paper introduce the RC-RAG to make a conditional answer by giving the readability scores.

- We made the dataset ourselves from the articles and the papers from CORD-19 and WHO QA pairs

- We train the RC-RAG in this dataset and have a great results.

Thanks, Question?

Github Repo: github.com/Caplimbo/covid_QA