

1.

A.

BERT uses the encoder part of Transformer, which uses self-attention. Therefore, in an extractive question answering task, all the words in the **paragraph** and the **question** are considered input of the encoder, and the encoder encodes and output 2 number representing the start and the end of the location of the **answer** in the **paragraph**.

B.

GPT uses the decoder part of Transformer, which uses auto-regressive. Therefore, in an extractive question answering task, the words in the **paragraph** and the **question** are considered input of the decoder, and the decoder decodes and output every character one by one, but every time a character is outputted it will be treated as a decoder input for all future characters that are going to be generated. Assume the **answer** is X
X_i's generation will be decided by the decoder with the **paragraph**, the **question** and **X₀ to X_{i-1}** as input. Like a game of word chain(文字接龍).

2.

The following are 3 pairs of prompts comparison, each with corresponding screenshots and observations.

第一對比較：

「用文章裡面的文字回答問題」嘗試限制 model 回答直接在文章裡抓答案出來，「回答問題」則試圖讓 model 自由發揮，自行判斷要 output 什麼。最終結過 ID5「用文章裡面的文字回答問題」的回答錯誤、ID8、ID9「用文章裡面的文字回答問題」的回答與 Ground-Truth 較為相近。從這個結果可以觀察到，嘗試限制 model 從文章裡抓答案可以達到的效果在於，如果原本 model 自由發揮也就是「回答問題」無法得到正確答案，直接從文章裡抓比較有機會接近 Ground-Truth。但是如果原本 model 自由發揮就能得到答案的話，限制 model 從文章裡抓答案反而可能會造成像 ID5 的現象，答非所問。

用文章裡面的文字回答問題	回答問題
ID,Ground-Truth,Prediction 0,2007年1月16日,2007年 1,200公里,80 2,廣州,廣州 3,2015,1990年 4,文化大革命,文化大革命 5,廣州,廣州文章:廣州是中國最重要的開放城市之一。廣州是中國最重要的開 6,廣州,廣州 7,武昌起義,1911 8,香港,廣州和香港 9,1921,1913 10,7月,2010 11,2008年,1992 12,從化市,廣州的人口密度是廣州的一百倍分之一以下。 13,80,80% 14,非洲黑人,阿拉伯國家 15,普通話,廣東話	ID,Ground-Truth,Prediction 0,2007年1月16日,2007年 1,200公里,80 2,廣州,廣州 3,2015,1990年 4,文化大革命,文化大革命 5,廣州,廣州 6,廣州,廣州 7,武昌起義,1911 8,香港,1989 9,1921,廣州 10,7月,2010 11,2008年,2008年 12,從化市,廣州的人口密度是100倍分之一以下。 13,80,80% 14,非洲黑人,阿拉伯國家 15,普通話,廣東話

第二對比較：

「這是一個文字接龍遊戲，完成文字接龍」主要是因為教授上課時的說法得到的啟發，讓 model 自行決定想幹嘛甚至不是回答問題也可以，「你獲得了資訊，利用資訊產生問題的答案」則是預設給 model 情境，使它在這個情境下產生 output，並且希望它是一個問題的答案。最終結果 ID12、ID15 產生結果不同、並且都與 Ground-Truth 不同，不過可以觀察出在有預設情境的情況下，也就是「你獲得了資訊，利用資訊產生問題的答案」，model 可以至少產生出一個答案（廣州的中心區、廣東話），相對於完全不預設情境玩文字接龍，model 有時候會亂接（廣州的人口密度是 100 被分之一下。、廣州本地人與外國人交流），也許是因為「遊戲」這兩個字使 model 目的性無法足夠明確。

這是一個文字接龍遊戲，完成文字接龍	你獲得了資訊，利用資訊產生問題的答案
ID,Ground-Truth,Prediction 0,2007年1月16日,2007年 1,200公里,80 2,廣州,廣州 3,2015,1990年 4,文化大革命,文化大革命 5,廣州,廣州 6,廣州,廣州 7,武昌起義,1911 8,香港,香港和深圳 9,1921,廣州 10,7月,2010 11,2008年,1992 12,從化市,廣州的人口密度是100倍分之一以下。 13,80,80% 14,非洲黑人,阿拉伯國家 15,普通話,廣州本地人與外國人交流	ID,Ground-Truth,Prediction 0,2007年1月16日,2007年 1,200公里,80 2,廣州,廣州 3,2015,1990年 4,文化大革命,文化大革命 5,廣州,廣州 6,廣州,廣州 7,武昌起義,1911 8,香港,香港和深圳 9,1921,廣州 10,7月,2010 11,2008年,1992 12,從化市,廣州的中心區 13,80,80% 14,非洲黑人,阿拉伯國家 15,普通話,廣東話

第三對比較：

「答案是在文章裡與問題有高度相關性的詞」為純中文，「答案 is in the 文章 and is highly related to the 問題」，會做此嘗試是因為，猜測 model 對英文的理解度比較高，但是把重複出現關鍵字換成中文也許會有意想不到的好的結果。但是結果顯示，中英夾雜的情況下，反而會讓 model 在某些情況直接失去回答問題的能力，例如 ID5（唐）、ID8（台北市台北市台北市）、ID14（空白）。

答案是在文章裡與問題有高度相關性的詞	答案 is in the 文章 and is highly related to the 問題
ID,Ground-Truth,Prediction 0,2007年1月16日,2007年 1,200公里,80 2,廣州,廣州 3,2015,1990年 4,文化大革命,文化大革命 5,廣州,廣州的對外貿易發展到明清时期,廣州的對外貿易 6,廣州,廣州 7,武昌起義,1911 8,香港,1989 9,1921,廣州 10,7月,2010 11,2008年,1992 12,從化市,廣州的人口密度是100倍分之一以下。 13,80,80% 14,非洲黑人,阿拉伯國家 15,普通話,廣東話	ID,Ground-Truth,Prediction 0,2007年1月16日,2007年 1,200公里,200 2,廣州,廣州 3,2015,1990年 4,文化大革命,文化大革命 5,廣州,唐 6,廣州,廣州 7,武昌起義,1911 8,香港,香港,深圳,台北,台北市,台北市,台北市,台北市,台北市 9,1921,清朝末期,廣州被稱為廣州被稱為廣州被稱為 10,7月,2010 11,2008年,1992 12,從化市,廣州的人口密度是100倍分之一以下。 13,80,80% 14,非洲黑人, 15,普通話,廣東話