

# GPT2\_Chinese

The advancement of AI technology has made it possible to automatically generate many things with AI. The language model of the GPT series can already generate articles, answer questions, and translate. Here is a GPT-2 Chinese article generation APP for you, which can automatically generate various articles.

Version 20230223

# Applications

- Using the GPT2 algorithm can be applied to automatically generate lyrics, automatically generate ancient prose, automatically generate poems, automatically generate couplets, automatically generate articles, automatically generate novels, etc.

# How to use

- The main process is:

Prepare article samples ->

train -> generate articles

LEADERG - GPT2\_Chinese - Version 12

**Train**

GPU id

batch size

pretrained model Path

**Inference**

Select Generate Model :

GPU id

Number of samples

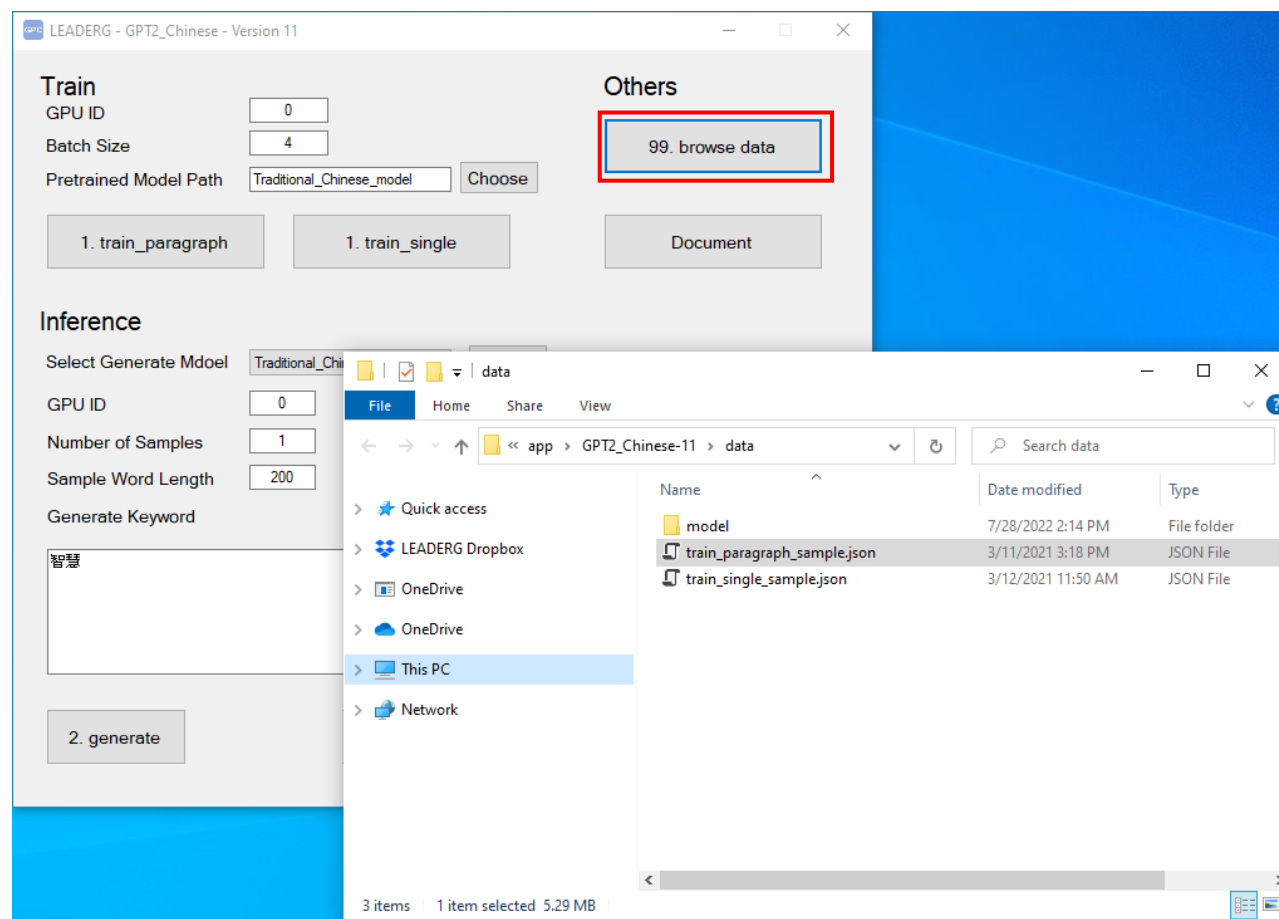
Sample word length

Generate keyword

**Others**

# Prepare the article

After pressing [99. browse data], open the data folder, you will see train\_paragraph\_sample.json and train\_single\_sample.json, one is a text format composed of many paragraph articles, and the other is a text format with a single article content.



# Prepare training samples of paragraph type

If the prepared sample is a paragraph type article, such as lyrics, ancient texts, news headlines, etc., the format of this book can refer to "train\_paragraph\_sample.json".

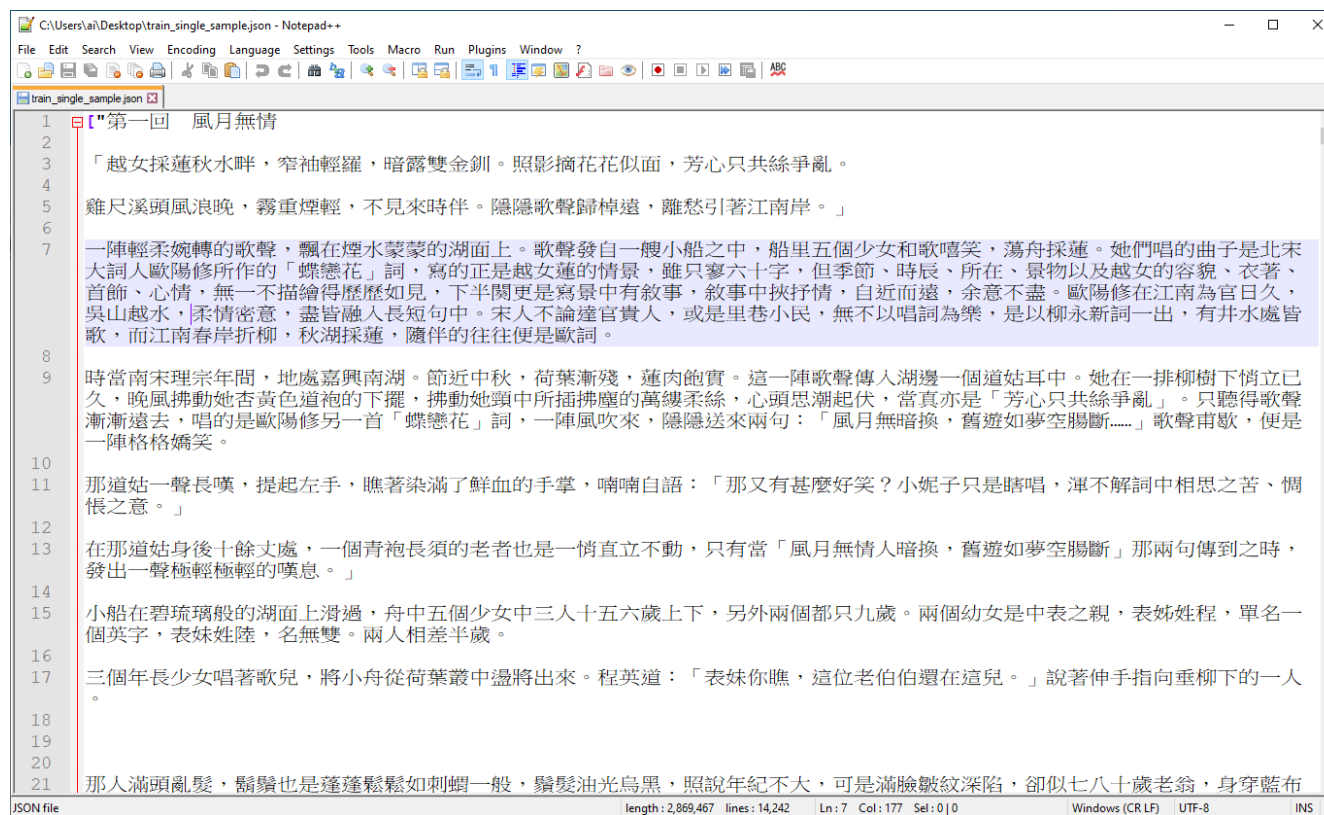
The format of the sample is ["paragraph 1", "paragraph 2", "...", "paragraph n"] , , taking the example of APP as an example, it is like [“哈通保卡山.....”, “帕哈林保卡山.....”, “夏侯獻.....”, ..... , “任天堂Labo ..... 截然不同。”] 。

```
train_paragraph_sample.json
```

```
1 [
  "哈通保卡山哈通保卡山是秘魯的山峰，位於該國中南部利馬大區，由米拉弗洛雷斯區和坦塔區負責管轄，屬於秘魯中部山脈的一部分，海拔高度5,420米。",
  "帕哈林保卡山帕哈林保卡山是秘魯的山峰，位於該國中南部利馬大區，由米拉弗洛雷斯區和坦塔區負責管轄，屬於秘魯中部山脈的一部分，海拔高度5,200米。",
  "夏侯獻夏侯獻，三國時人物。曹魏功臣夏侯氏裔身，世係不明。魏明帝曹叡時代，遼東公孫淵殺死孫吳使者，外有高句麗、濊貊為仇，局勢不穩。中領軍夏侯獻上奏，按公孫淵之父公孫康服從的先例，派遣使者陳說利害，打消公孫淵的叛意。後拜為領軍將軍。239年，曹叡臨終時，夏侯獻和燕王曹宇、曹休之子屯騎校尉曹肇、曹真之子武衛將軍曹爽、曹操側室杜夫人之子驍騎將軍秦朗受遺命輔佐皇太子曹芳。劉放、孫資建議以曹爽、司馬懿代曹宇等輔政。夏侯獻、曹肇、秦朗和曹宇一起被更迭，相對流淚迴傢。《三國誌注·劉放傳》引《世說新語》，記載之前夏侯獻、曹肇對劉放、孫資掌握政治中樞不滿。劉放、孫資身感危險，勸臨終的曹叡徵召徵伐遼東的司馬懿迴朝。夏侯獻先密通曹宇命令司馬懿往長安，劉放派的使者命令下達給司馬懿，司馬懿懷疑洛陽有事變，急忙迴到洛陽。",
  "內蒙古自治區長城列錶內蒙古自治區長城是位於中國內蒙古自治區的長城牆體及附屬設施，分布於76個縣、旗、市、區，建於戰國、秦、漢、北魏、隋、西夏、金、明等曆史時期，其中：因內蒙古自治區幅員遼闊，長城分布廣，分為三個子列錶：",
  "托羅約剋山托羅約剋山是秘魯的山峰，位於該國中南部利馬大區，由萬坦區和拉勞斯區負責管轄，屬於秘魯中部山脈的一部分，海拔高度5,400米。",
```

# Prepare a training sample for an entire article

If the prepared sample is of the type of an entire article, such as the content of a novel prepared by the APP, please refer to “train\_single\_sample.json” for the format of the sample. The format of the sample is [“an entire article content”], taking the example of the APP as an example, it is like [“第一回 風月無情 ..... 難為情。””]。



```
1 ["第一回 風月無情
2
3 「越女採蓮秋水畔，窄袖輕羅，暗露雙金釧。照影摘花花似面，芳心只共絲爭亂。
4
5 雞尺溪頭風浪晚，霧重煙輕，不見來時伴。隱隱歌聲歸棹遠，離愁引著江南岸。」
6
7 一陣輕柔婉轉的歌聲，飄在煙水蒙蒙的湖面上。歌聲發自一艘小船之中，船里五個少女和歌嬉笑，蕩舟採蓮。她們唱的曲子是北宋大詞人歐陽修所作的「蝶戀花」詞，寫的正是越女蓮的情景，雖只寥六十字，但季節、時辰、所在、景物以及越女的容貌、衣著、首飾、心情，無一不描繪得歷歷如見，下半闕更是寫景中有敘事，敘事中挾抒情，自近而遠，余意不盡。歐陽修在江南為官日久，吳山越水，柔情密意，盡皆融入長短句中。宋人不論達官貴人，或是里巷小民，無不以唱詞為樂，是以柳永新詞一出，有井水處皆歌，而江南春岸折柳，秋湖採蓮，隨伴的往往便是歐詞。
8
9 時當南宋理宗年間，地處嘉興南湖。節近中秋，荷葉漸殘，蓮肉飽實。這一陣歌聲傳入湖邊一個道姑耳中。她在一排柳樹下佇立已久，晚風拂動她杏黃色道袍的下擺，拂動她頸中所插拂塵的萬縷柔絲，心頭思潮起伏，當真亦是「芳心只共絲爭亂」。只聽得歌聲漸漸遠去，唱的是歐陽修另一首「蝶戀花」詞，一陣風吹來，隱隱送來兩句：「風月無暗換，舊遊如夢空腸斷.....」歌聲甫歇，便是一陣格格嬌笑。
10
11 那道姑一聲長嘆，提起左手，瞧著染滿了鮮血的手掌，喃喃自語：「那又有甚麼好笑？小妮子只是瞎唱，渾不解詞中相思之苦、惆悵之意。」
12
13 在那道姑身後十餘丈處，一個青袍長須的老者也是一竄直立不動，只有當「風月無情人暗換，舊遊如夢空腸斷」那兩句傳到之時，發出一聲極輕極輕的嘆息。」
14
15 小船在碧琉璃般的湖面上滑過，舟中五個少女中三人十五六歲上下，另外兩個都只九歲。兩個幼女是中表之親，表姊姓程，單名一個英字，表妹姓陸，名無雙。兩人相差半歲。
16
17 三個年長少女唱著歌兒，將小舟從荷葉叢中盪將出來。程英道：「表妹你瞧，這位老伯伯還在這兒。」說著伸手指向垂柳下的一人。
18
19
20
21 那人滿頭亂髮，鬚鬚也是蓬蓬鬆鬆如刺蝟一般，鬚髮油光烏黑，照說年紀不大，可是滿臉皺紋深陷，卻似七八十歲老翁，身穿藍布
```

JSON file length: 2,869,467 lines: 14,242 Ln: 7 Col: 177 Sel: 0|0 Windows (CR LF) UTF-8 INS

# Model introduction

The APP provides 8 models:










- `chinese_lyrics_model`: Simplified Chinese lyrics model.
- `classical_chinese_model`: Simplified Chinese classical Chinese model.
- `couplet_model`: Simplified Chinese couplet model.
- `general_chinese_model`: Simplified Chinese article model.
- `general_chinese_small_model`: Simplified Chinese short article model.
- `JinYong_model`: Traditional Chinese Jin Yong novel model.
- `poetry_model`: Simplified Chinese ancient poetry model.
- `Traditional_Chinese_model`: Traditional Chinese article model.

Remark:

- `final_model` is the resulting model location after training.

: GPT2\_Chinese-11 > data > model

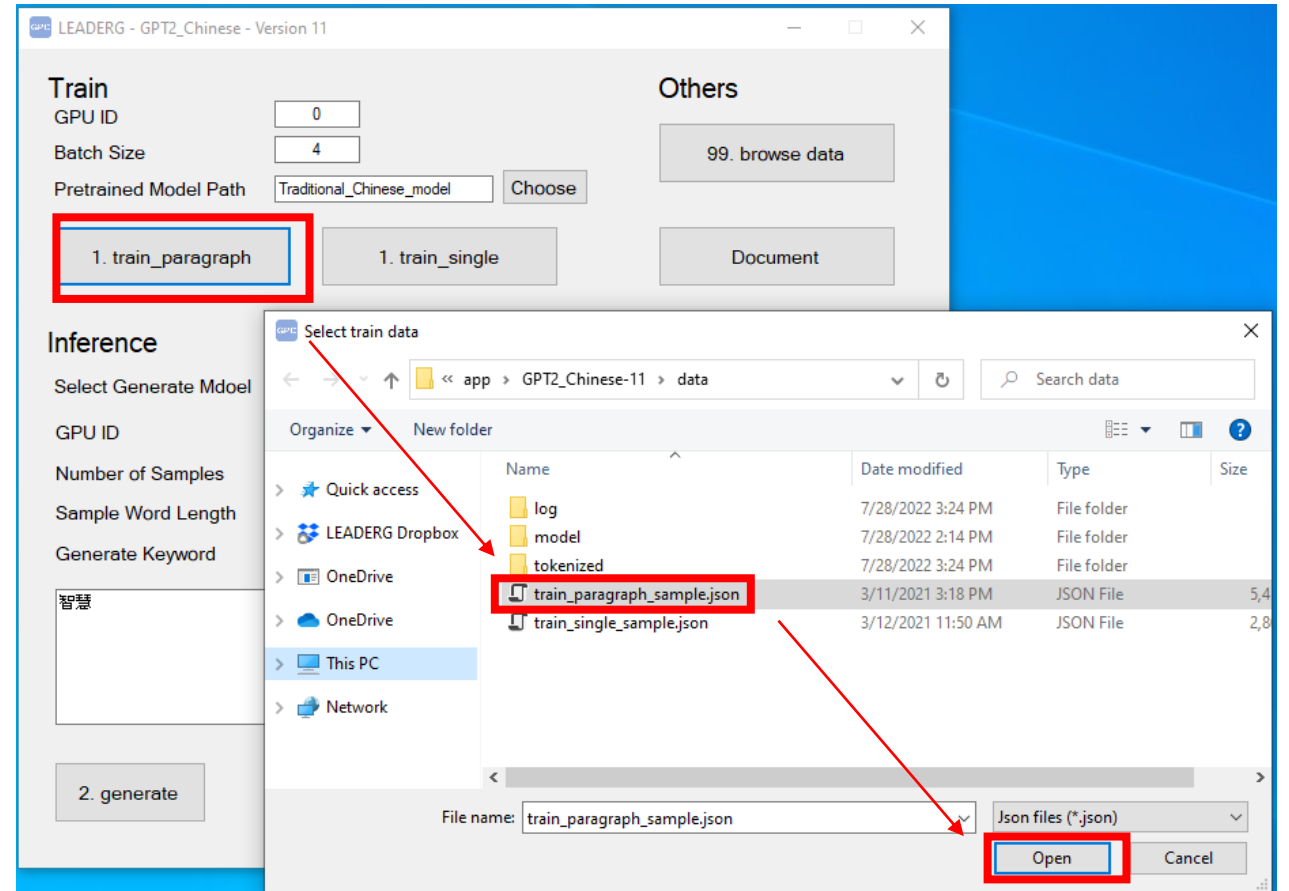
Name

-  `chinese_lyrics_model`
-  `classical_chinese_model`
-  `couplet_model`
-  `final_model`
-  `general_chinese_model`
-  `general_chinese_small_model`
-  `JinYong_model`
-  `poetry_model`
-  `Traditional_Chinese_model`

# 1. train\_paragraph

If the prepared text type is paragraph type, you can press [1. train\_paragraph] to select your own samples for training. The default Traditional\_Chinese\_model (Traditional Chinese model) is a pre-trained model.

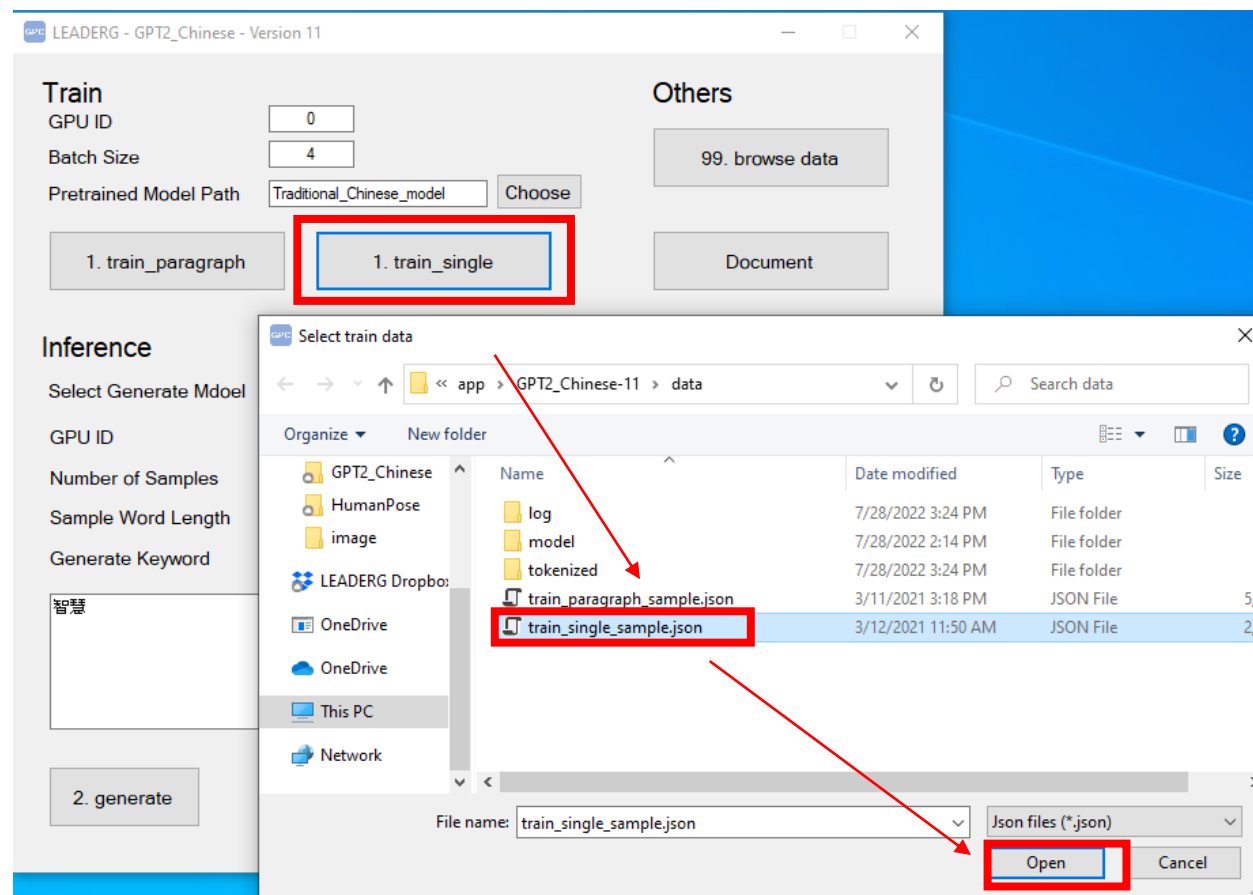
The trained model will exist in "final\_model".





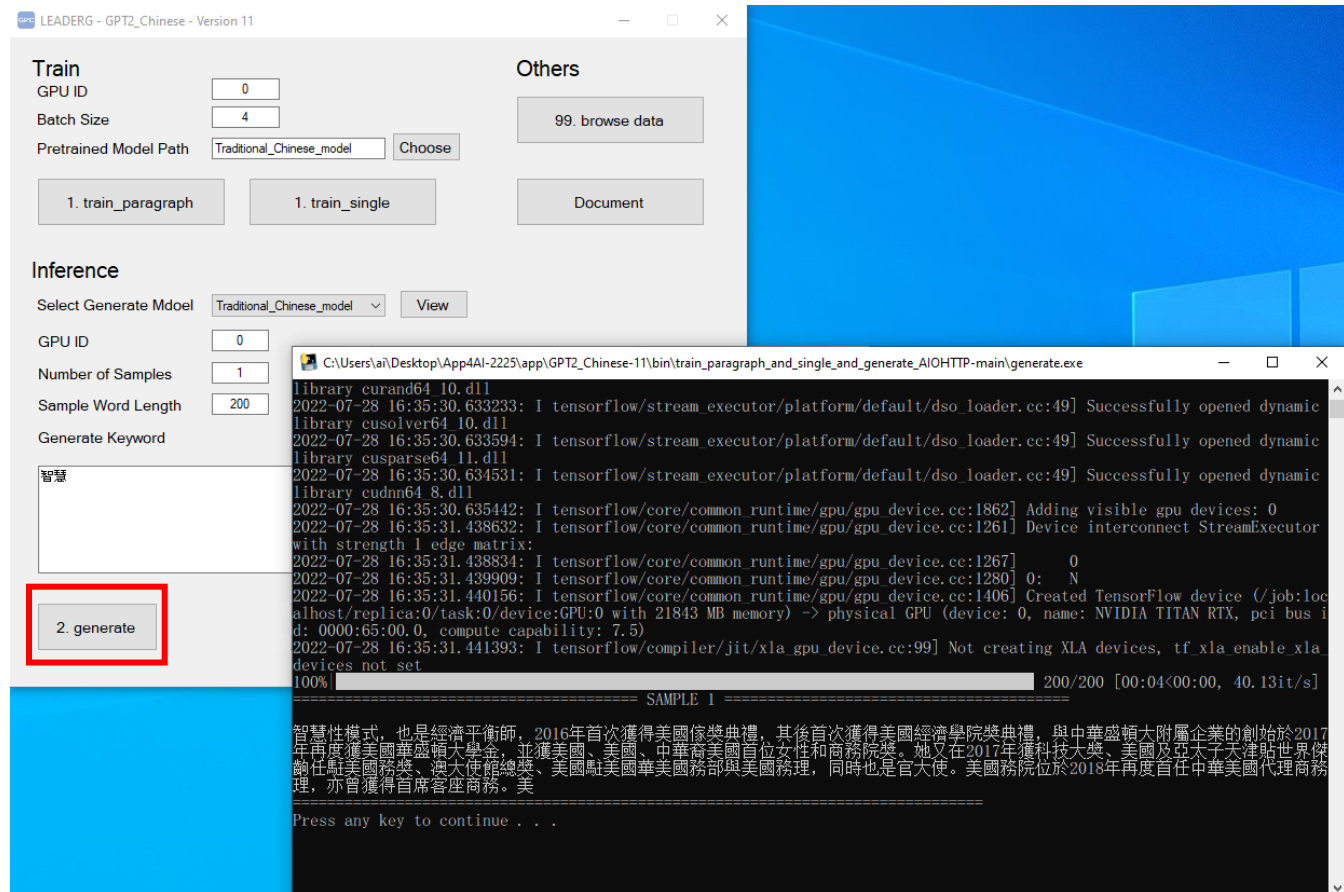
# 1. train\_single

If the prepared text type is an entire article type, you can press [1. train\_single] to select your own sample for training. The default Traditional\_Chinese\_model (Traditional Chinese model) is a pre-trained model. The trained model will exist in "final\_model".



## 2. generate

Select the model to generate text (Select Generate Model), you can refer to the model introduction to select the model, set the number of articles you want to generate (Number of Samples), and a reasonable number of words to generate (Sample Word Length), and finally enter as the beginning of text, you can press [2. generate] to generate the article.



# Reference

- Please refer to the readme.txt in the APP folder.
- LEADERG AppForAI: <https://www.leaderg.com/appforai-windows>
- Copyright © LEADERG INC. All rights reserved.