

# CHAT-BOT

曾秋旺、林亦盛、黃俊愷、王俊迪

# OUTLINE

- 前言
- 資料簡介及前處理
- **Model**簡介
- 成果**demo**
- 未來工作

# 前言

- 李宏毅教授 機器學習 **Kaggle**比賽
- 聊天機器人:閱讀一段對話後，從六個選項中，選出最適答案
- **Ex：** **A:**醫生在幫你擦藥的時候 爸爸在外面聽到 你在裡面喊痛的聲音，  
選項：  
**A:**我的心好像刀割的一樣  
**A:**我想去六福村玩  
**A:**美國隊長最愛浩克  
**A:**我有蘋果和一支筆  
**A:**我覺得不行  
**A:**我買了哀鳳十

# 資料簡介

- 訓練資料:

從台灣電視節目中擷取的對話，經查詢發現，是民視的八點檔戲劇

- 測試資料:

1. 對話與選項用逗號隔開
2. 不同選項用 “\t” 分開
3. 不同題目用 “\n” 分開

# 資料前處理

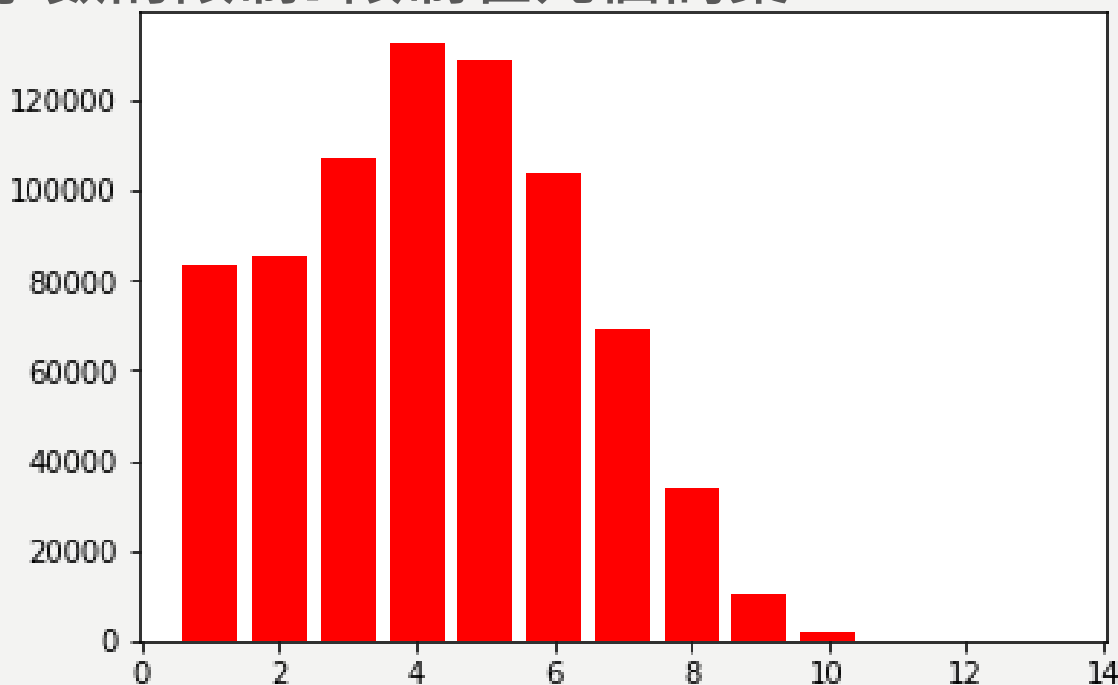
- 斷詞 : jieba
- Word2vec : genism
  - Size: 4096
  - 詞彙數量 : 12438

# 資料前處理

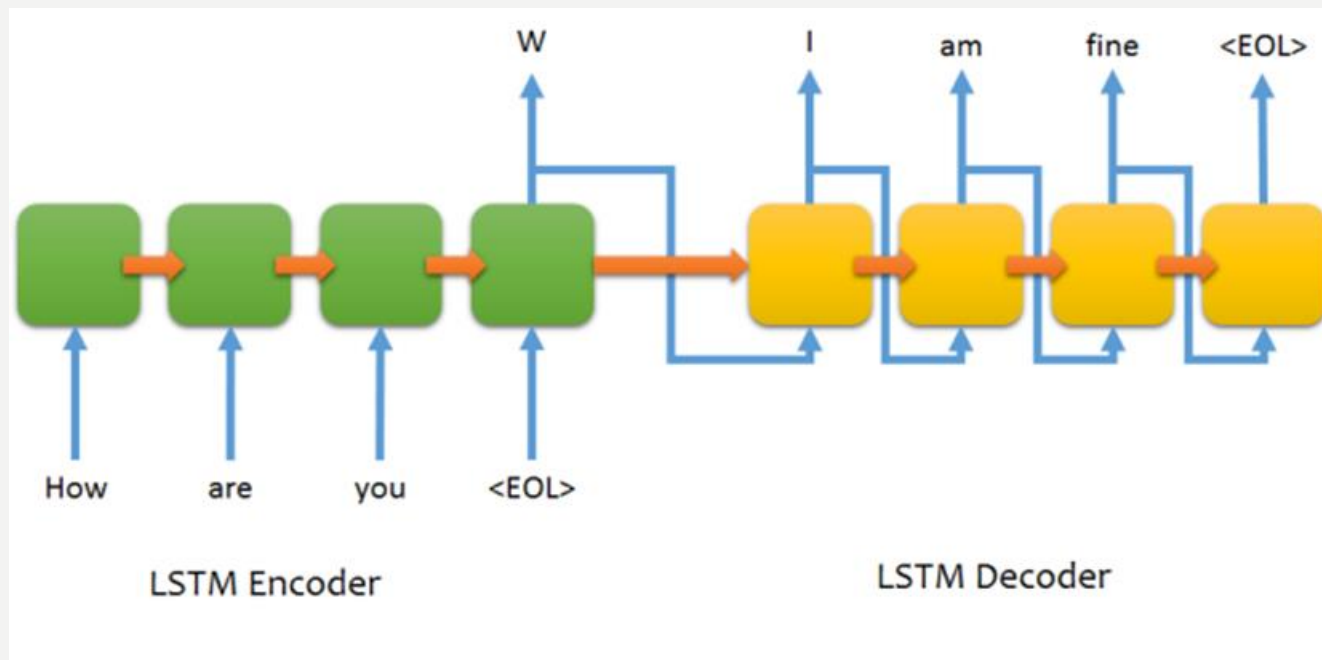
1. 將測試資料中的人名提出，加進詞庫裡。

**Ex:** 陳家輝、家輝、李有土

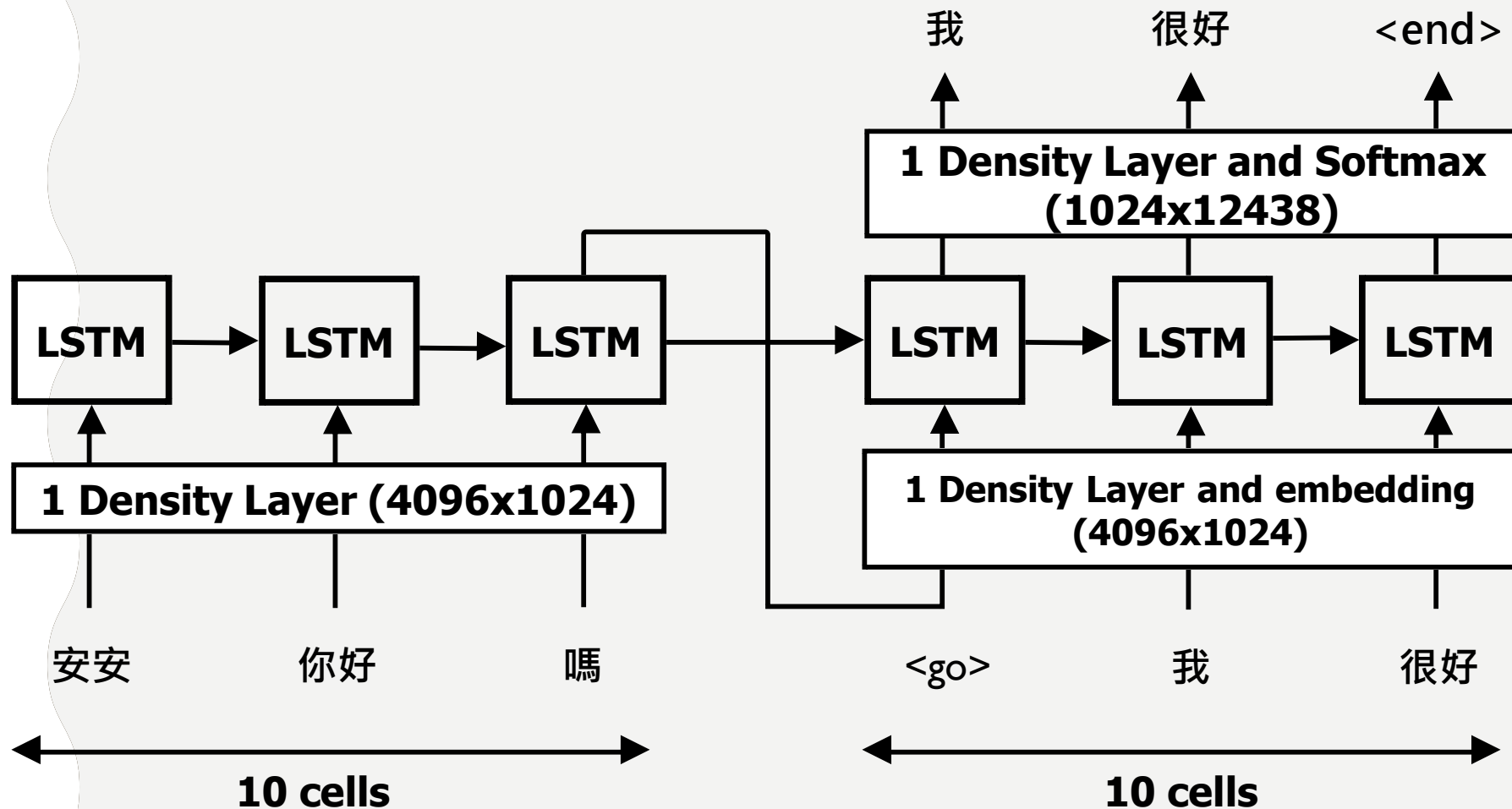
2. 訓練字數的限制: 限制在九個詞彙



# MODEL簡介

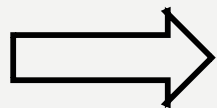
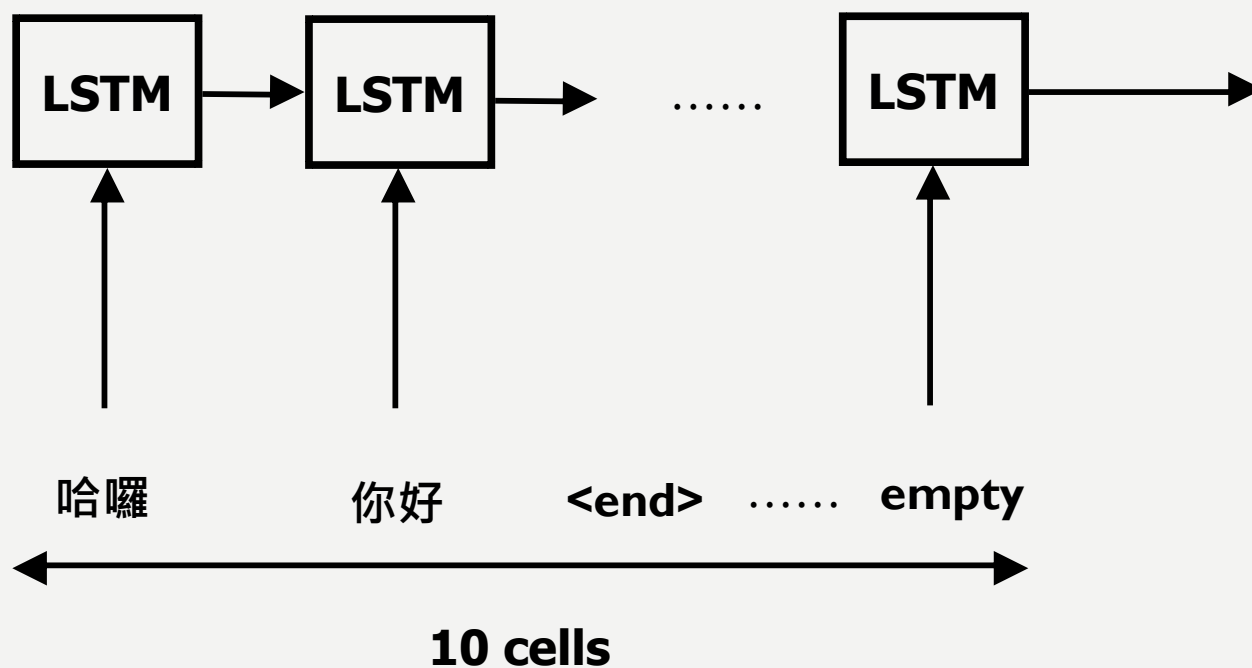


# MODEL簡介



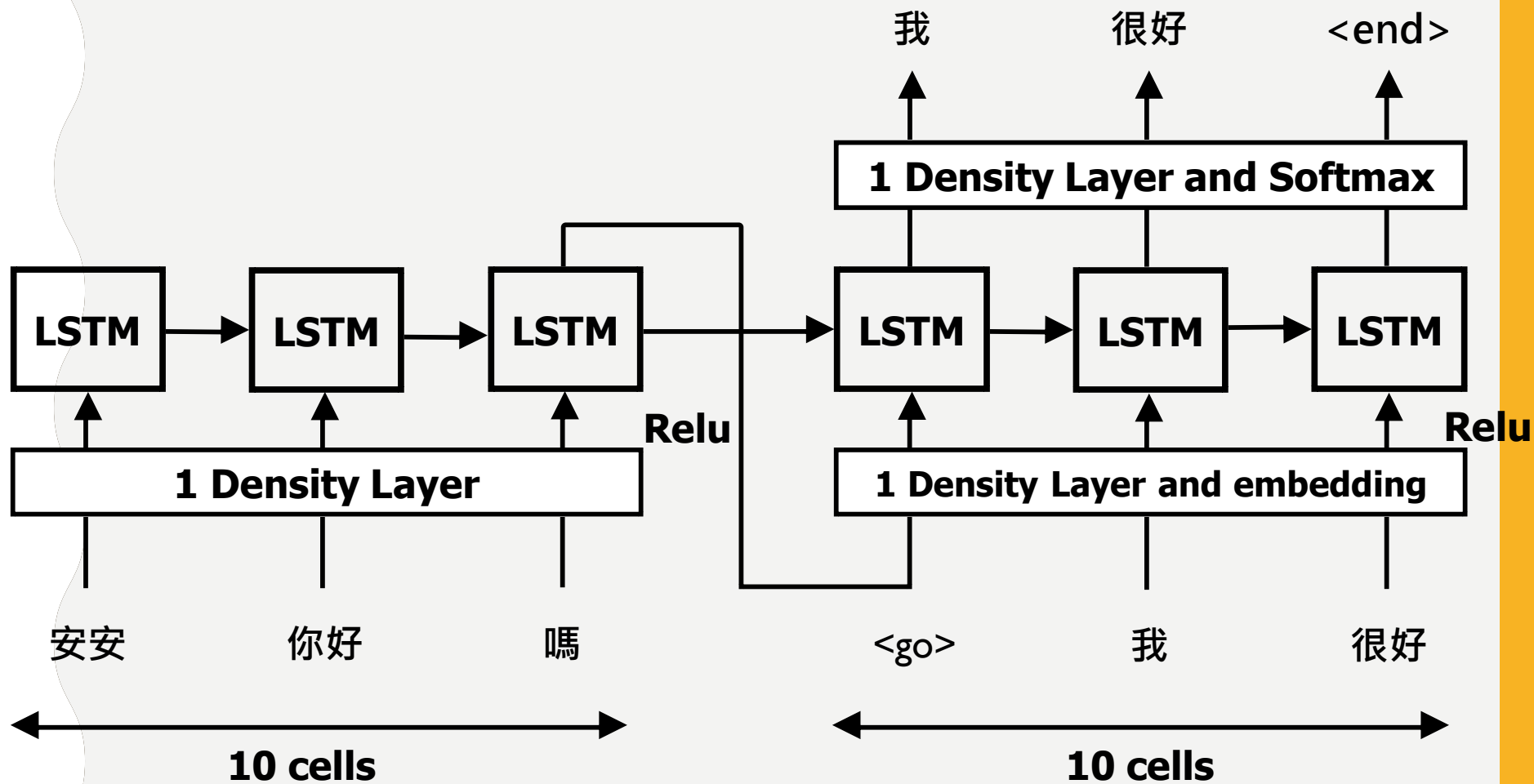


# MODEL簡介



只訓練前3個cell(包含<enc>)

# CHAT-BOT 1



# CHAT-BOT 1

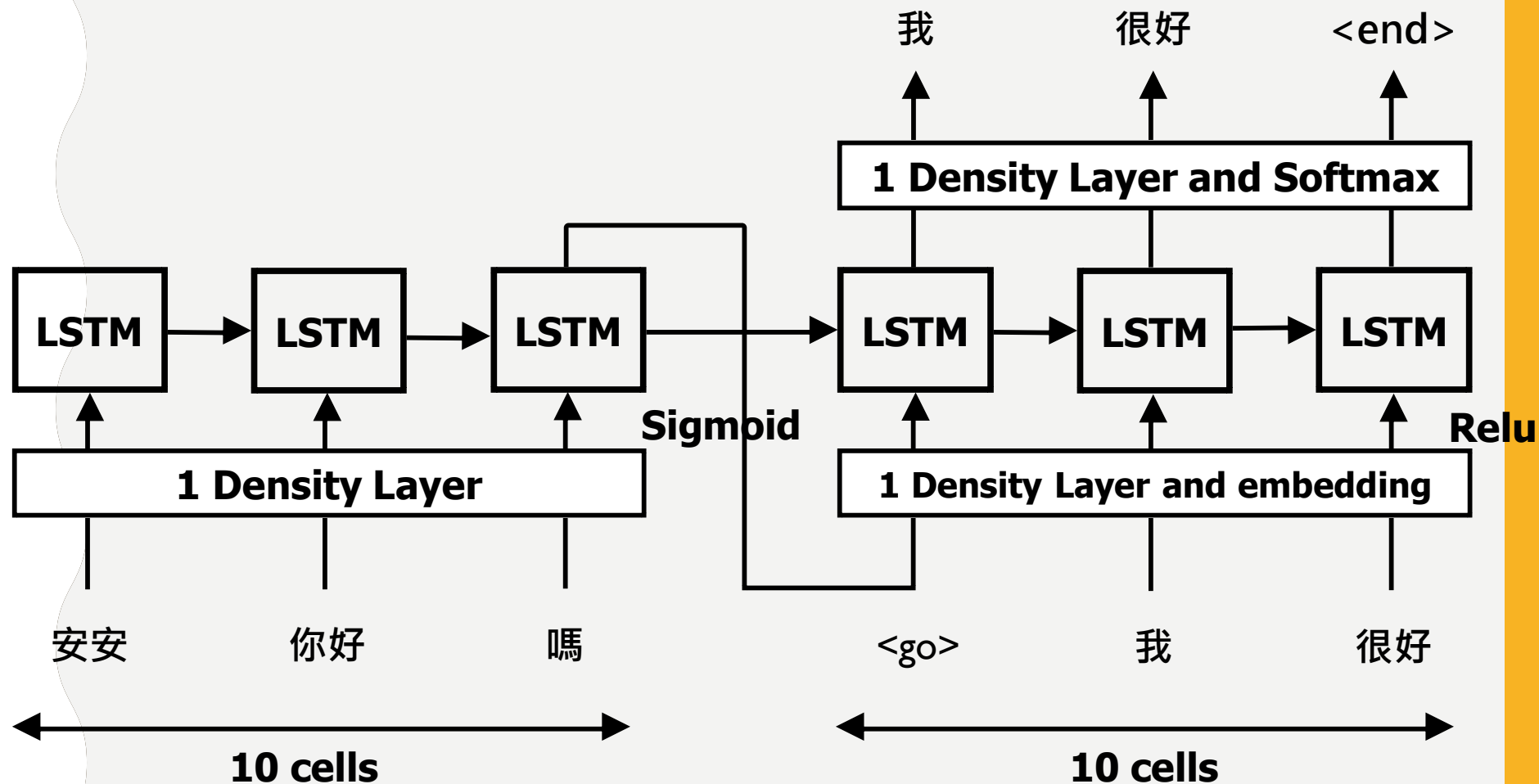
- **1 epoch**

- **Input:** 你還記得我之前有跟你說過
- **Output 1:** 妳我要要
- **Output 2:** 好

- **15 epoch**

- **Input:** 你還記得我之前有跟你說過
- **Output 1:** 我是真心想彌補你
- **Output 2:** 我怎麼會走

# MODEL簡介



# CHAT-BOT 2

- **1 epoch**

- **Input:** 你才剛生完小孩體力都還沒復原
- **Output 1:** 你
- **Output 2:** 我跟你說的

- **15 epoch**

- **Input:** 你才剛生完小孩體力都還沒復原
- **Output 1:** 我怎麼會趕你出去
- **Output 2:** 你怎麼可以這麼自私呢

# 成果DEMO

- [https://github.com/NTU-CSX-Project/R05631018/tree/master/Final\\_Chat-Bot](https://github.com/NTU-CSX-Project/R05631018/tree/master/Final_Chat-Bot)

# 未來工作

- 訓練句子與句子之間的語意判斷
- 考慮有沒有對話的因素
- 將**word2vec**也納入模型裡做訓練