# Amazon Sales Chatbot

## Phase 2: Progress Report

Presented by

- Minn Banya
- Wut Yee Aung
- Kyi Thin Nu
- Thamakorn Kiattikaikul
- Noppawee Teeraratchanon

Natural Language Processing

# Project Overview (summary)

- Our Amazon Sale Chatbot is a virtual assistant designed to enhance the shopping experience on the Amazon platform.
- Objectives
  - to increase user engagement, and satisfaction on the Amazon platform using NLP.
- Target audience
  - Amazon customers

# Our Data Source

Dataset information : to mention which dataset is used for which task???

- Amazon product dataset (extracted from meta data) - https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023

- Train dataset - https://github.com/amazonqa/amazonqa?tab=readme-ov-file

# Classification Model

- We want to classify the question into 17 classes (Toys_and_Games, Health_and_Personal_Care, Cell_Phones_and_Accessories, Home_and_Kitchen, Musical_Instruments, Baby, Sports_and_Outdoors, Patio_Lawn_and_Garden, Video_Games, Pet_Supplies, Tools_and_Home_Improvement, Beauty, Electronics ,Grocery_and_Gourmet_Food, Automotive, Office_Products, Clothing_Shoes_and_Jewelry)
- We have trained 2 models (biLSTM and CNN) on different parameter with small (500 sample questions per product category) and large dataset (1000 sample questions per product category). Moreover, we saved the model when the validation loss is improve. Therefore, we select the best model with specific parameter based on minimum validation loss

# biLSTM Model with small dataset

| Run Name | Created | Duration | min_val_loss |
|---|---|---|---|
| biLSTM500-5-epochs-128-hidden dim-2-num layers | 12 hours ago | 1.2min | 2.191 |
| biLSTM500-5-epochs-128-hidden dim-4-num layers | 12 hours ago | 1.9min | 2.543 |
| biLSTM500-5-epochs-128-hidden dim-6-num layers | 12 hours ago | 2.6min | 2.71 |
| biLSTM500-5-epochs-256-hidden dim-2-num layers | 11 hours ago | 1.7min | 2.835 |
| biLSTM500-5-epochs-256-hidden dim-4-num layers | 11 hours ago | 3.2min | 2.978 |
| biLSTM500-5-epochs-256-hidden dim-6-num layers | 11 hours ago | 4.8min | 3.262 |
| biLSTM500-5-epochs-512-hidden dim-2-num layers | 11 hours ago | 4.1min | 3.328 |
| biLSTM500-5-epochs-512-hidden dim-4-num layers | 11 hours ago | 8.4min | 3.626 |
| biLSTM500-5-epochs-512-hidden dim-6-num layers | 11 hours ago | 12.6min | 3.029 |
| biLSTM500-10-epochs-128-hidden dim-2-num layers | 11 hours ago | 1.8min | 3.255 |
| biLSTM500-10-epochs-128-hidden dim-4-num layers | 11 hours ago | 3.4min | 3.36 |
| biLSTM500-10-epochs-128-hidden dim-6-num layers | 11 hours ago | 4.8min | 3.165 |
| biLSTM500-10-epochs-256-hidden dim-2-num layers | 11 hours ago | 3.0min | 4.14 |
| biLSTM500-10-epochs-256-hidden dim-4-num layers | 11 hours ago | 5.9min | 3.917 |
| biLSTM500-10-epochs-256-hidden dim-6-num layers | 11 hours ago | 8.6min | 3.696 |
| biLSTM500-10-epochs-512-hidden dim-2-num layers | 10 hours ago | 7.8min | 4.771 |
| biLSTM500-10-epochs-512-hidden dim-4-num layers | 10 hours ago | 16.2min | 4.34 |
| biLSTM500-10-epochs-512-hidden dim-6-num layers | 10 hours ago | 25.2min | 3.966 |

For biLSTM, when we increase the epoch, hidden dimension, and number of layers, the model is worse.

# biLSTM Model with large dataset

| Run Name | Created | Duration | min_val_loss |
|---|---|---|---|
| biLSTM1000-5-epochs-128-hidden dim-2-num layers | 8 hours ago | 2.6min | 1.977 |
| biLSTM1000-5-epochs-128-hidden dim-4-num layers | 8 hours ago | 4.4min | 2.257 |
| biLSTM1000-5-epochs-128-hidden dim-6-num layers | 7 hours ago | 6.1min | 2.55 |
| biLSTM1000-5-epochs-256-hidden dim-2-num layers | 7 hours ago | 4.0min | 2.748 |
| biLSTM1000-5-epochs-256-hidden dim-4-num layers | 7 hours ago | 7.4min | 2.739 |
| biLSTM1000-5-epochs-256-hidden dim-6-num layers | 7 hours ago | 11.7min | 2.887 |
| biLSTM1000-5-epochs-512-hidden dim-2-num layers | 7 hours ago | 10.6min | 3.19 |
| biLSTM1000-5-epochs-512-hidden dim-4-num layers | 7 hours ago | 21.9min | 3.326 |
| biLSTM1000-5-epochs-512-hidden dim-6-num layers | 6 hours ago | 35.8min | 2.977 |
| biLSTM1000-10-epochs-128-hidden dim-2-num layers | 6 hours ago | 4.4min | 3.062 |
| biLSTM1000-10-epochs-128-hidden dim-4-num layers | 6 hours ago | 8.0min | 3.212 |
| biLSTM1000-10-epochs-128-hidden dim-6-num layers | 6 hours ago | 11.3min | 2.975 |
| biLSTM1000-10-epochs-256-hidden dim-2-num layers | 5 hours ago | 7.5min | 3.818 |
| biLSTM1000-10-epochs-256-hidden dim-4-num layers | 5 hours ago | 14.4min | 3.736 |
| biLSTM1000-10-epochs-256-hidden dim-6-num layers | 5 hours ago | 21.6min | 3.11 |
| biLSTM1000-10-epochs-512-hidden dim-2-num layers | 5 hours ago | 21.2min | 4.36 |
| biLSTM1000-10-epochs-512-hidden dim-4-num layers | 4 hours ago | 41.1min | 3.891 |
| biLSTM1000-10-epochs-512-hidden dim-6-num layers | 4 hours ago | 1.0h | 3.758 |

The biLSTM model is better when the dataset is larger

# CNN Model

## Small dataset

| Run Name | Created | Duration | min_val_loss |
|---|---|---|---|
| CNN500-30-epochs-150-n_filters | 8 hours ago | 7.0min | 2.099 |
| CNN500-30-epochs-100-n_filters | 8 hours ago | 6.4min | 2.073 |
| CNN500-30-epochs-50-n_filters | 9 hours ago | 4.8min | 2.076 |
| CNN500-20-epochs-150-n_filters | 9 hours ago | 4.9min | 2.1 |
| CNN500-20-epochs-100-n_filters | 9 hours ago | 4.5min | 2.066 |
| CNN500-20-epochs-50-n_filters | 9 hours ago | 3.4min | 2.111 |
| CNN500-10-epochs-150-n_filters | 9 hours ago | 2.5min | 2.191 |
| CNN500-10-epochs-100-n_filters | 9 hours ago | 2.3min | 2.202 |
| CNN500-10-epochs-50-n_filters | 9 hours ago | 1.8min | 2.231 |

## Large dataset

| Run Name | Created | Duration | min_val_loss |
|---|---|---|---|
| CNN1000-30-epochs-150-n_filters | 7 hours ago | 13.3min | 1.871 |
| CNN1000-30-epochs-100-n_filters | 7 hours ago | 12.1min | 1.867 |
| CNN1000-30-epochs-50-n_filters | 8 hours ago | 8.9min | 1.898 |
| CNN1000-20-epochs-150-n_filters | 8 hours ago | 9.3min | 1.886 |
| CNN1000-20-epochs-100-n_filters | 8 hours ago | 8.3min | 1.897 |
| CNN1000-20-epochs-50-n_filters | 8 hours ago | 6.3min | 1.927 |
| CNN1000-10-epochs-150-n_filters | 8 hours ago | 4.7min | 2.007 |
| CNN1000-10-epochs-100-n_filters | 8 hours ago | 4.2min | 2.028 |
| CNN1000-10-epochs-50-n_filters | 8 hours ago | 3.3min | 2.042 |

For CNN, we can conclude that when the number of filters, epoch, amount of data increase, the model is better

# Classification Model

Based on minimum validation loss, CNN with 30 epochs, 150 number of filters is our best model and we trained the model with a larger dataset (2000 dataset per category)

```
test_loss, test_acc = evaluate(model, test_loader, criterion, test_loader_length)

print(f'Test Loss: {test_loss:.3f} | Test Acc: {test_acc*100:.2f}%')
```

✓ 1.1s

```
Test Loss: 1.880 | Test Acc: 44.22%
```

We have tested model with test set and the model got around 44% accuracy

# Language Model

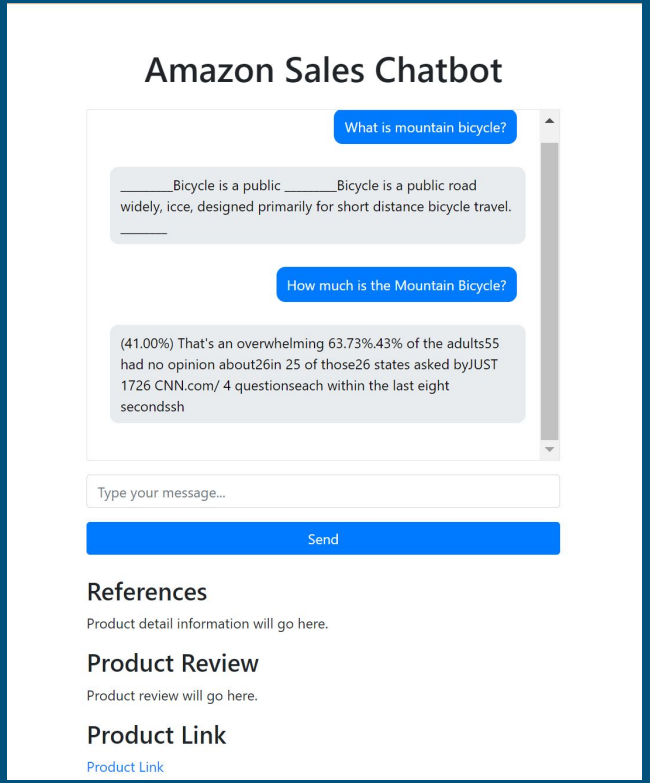
The product information from the meta datafiles are extracted and split into sample sizes of 100, 500 and 1000 products per category for experimentation.

The data samples are then stored as separate vector stores, and the application will determine which vector store the langchain should retrieve from based on the result of the classification model

# Design test cases

- Basic functionality testing
  - Verify that the chatbot can understand and respond to basic user queries.
- Product information retrieval testing
  - Ensure that the chatbot can provide accurate information of products from different categories.
- Recommendations testing
  - ensure that the chatbot can provide the recommendation based on different categories of products
- Error handling testing
  - To test that chatbot can handle error and unexpected question

# Web Design




Amazon Sales Chatbot

Chat Interface:

- Functionality:
  - Allows users to type messages (Type your message...), send them, and receive responses from the chatbot.
  - Messages are displayed in a visually distinct format, with user messages on the right and bot responses on the left.

Additional Sections:

- References:
  - Contains detailed information about the product.
- Product Review:
  - Displays product review content for user reference.
- Product Link:
  - Provides a link for users to access more detailed product information.

# Our Project GitHub Repository

https://github.com/minnbanya/NLP-Project

# Thank you
# For your attention