NEURAL NETWORKS & DEEP LEARNING

EMAIL SPAM CLASSIFICATION USING NLP TECHNIQUES & LLM MODELS

ADONIA SEQUEIRA G31345330
GANESH KUMAR RAJASEKAR G37969806

AGENDA

- PROBLEM STATEMENT
- DATASET DESCRIPTION
- SYSTEM DESIGN
- RESULTS
- COMPARATIVE ANALYSIS

PROBLEM STATEMENT

Email spam continues to disrupt user productivity and compromise security. **Traditional keyword-based filters failing** to detect sophisticated and contextually nuanced spam messages.

This project addresses the need for an advanced spam classification system by leveraging **NLP models, BERT and Universal Sentence Encoder (USE)** to enhance the accuracy and reliability of email filtering.

EMAIL SPAM: A RISING CONCERN

Daily Email Usage: Between 2019 and 2024, the number of email users globally increased from 3.9 billion to 4.4 billion with projections reaching 4.8 billion by 2027.

Daily Spam Victims by Country: As of December 8, 2024, users in the United States and China were the most heavily targeted by spam, each receiving an estimated 7.8 billion spam emails daily.

Global Spam Volume: In December 2024, spam messages accounted for more than 46.8% of all email traffic worldwide.

DATASET DETAILS & PREPROCESSING STEPS

DATASET DESCRIPTION

```
        label                                              text
0           1  ounce feather bowl hummingbird opec moment ala...
1           1  wulvob get your medircations online qnb ikud v...
2           0   computer connection from cnn com wednesday es...
3           1  university degree obtain a prosperous future m...
4           0  thanks for all your answers guys i know i shou...
...       ...                                               ...
83443       0  hi given a date how do i get the last date of ...
83444       1  now you can order software on cd or download i...
83445       1  dear valued member canadianpharmacy provides a...
83446       0  subscribe change profile contact us long term ...
83447       1  get the most out of life ! viagra has helped m...

[83448 rows x 2 columns]
```

Size: [83448 rows x 2 columns] **Features:** Text & Label

Labels:
1 -> indicates that the email is classified as **spam**
0 -> indicates that the email is classified as **ham**

Text:
Contains the actual content of the email messages

Text Lowercasing:

Without lowercasing, the model might **see a particular word as two different words**, which can reduce its ability to correctly identify spam emails.

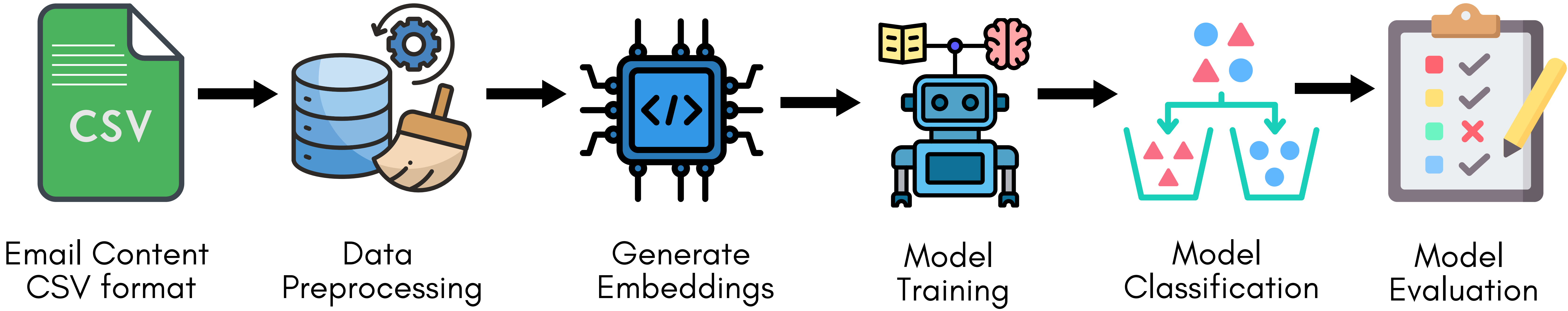
Special Character Removal:

Removing these characters **helps the model focus on the actual words and their meaning,** making it better at identifying spam content.

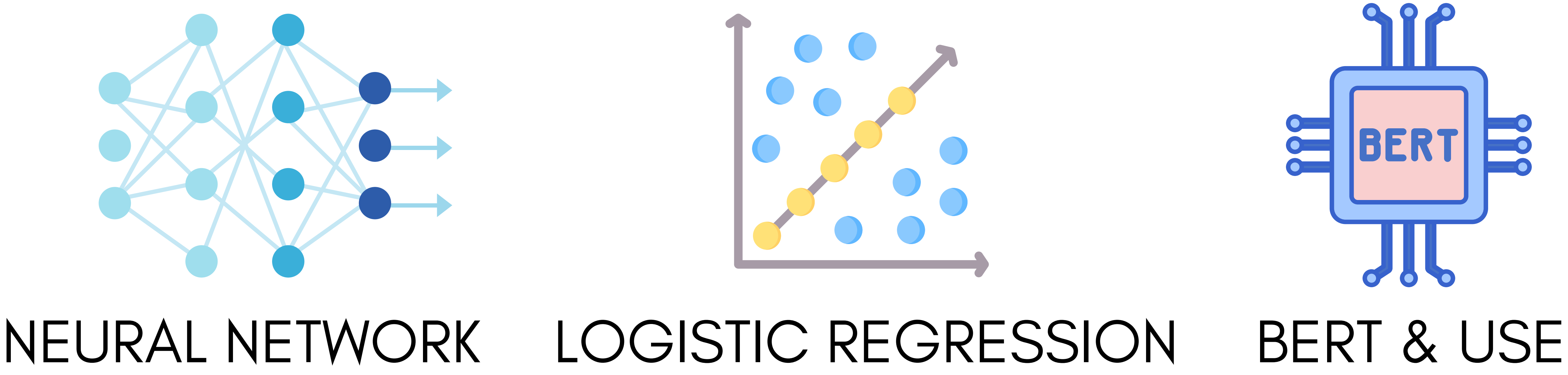
Whitespace Removal:

By removing unnecessary whitespaces, the text becomes cleaner and **easier for the algorithm to process** which overall improves performance.

SYSTEM DESIGN



MODEL



RESULTS

```
Enter the text (email/message) to classify as Spam or Ham: Thank you for your feedback.
Logistic Regression Prediction (USE): Ham
Logistic Regression Prediction (BERT): Ham
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 17ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 20ms/step
Neural Network Prediction (USE): Ham
Neural Network Prediction (BERT): Ham
```

Logistic Regression and Neural Networks, using USE and BERT embeddings, **accurately classified the message as Ham** by understanding its **meaning** and recognizing the **lack of common spam traits.**

```
Enter the text (email/message) to classify as Spam or Ham: Congratulations, you have successfully logged in.
Logistic Regression Prediction (USE): Spam
Logistic Regression Prediction (BERT): Ham
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 17ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 16ms/step
Neural Network Prediction (USE): Ham
Neural Network Prediction (BERT): Ham
```

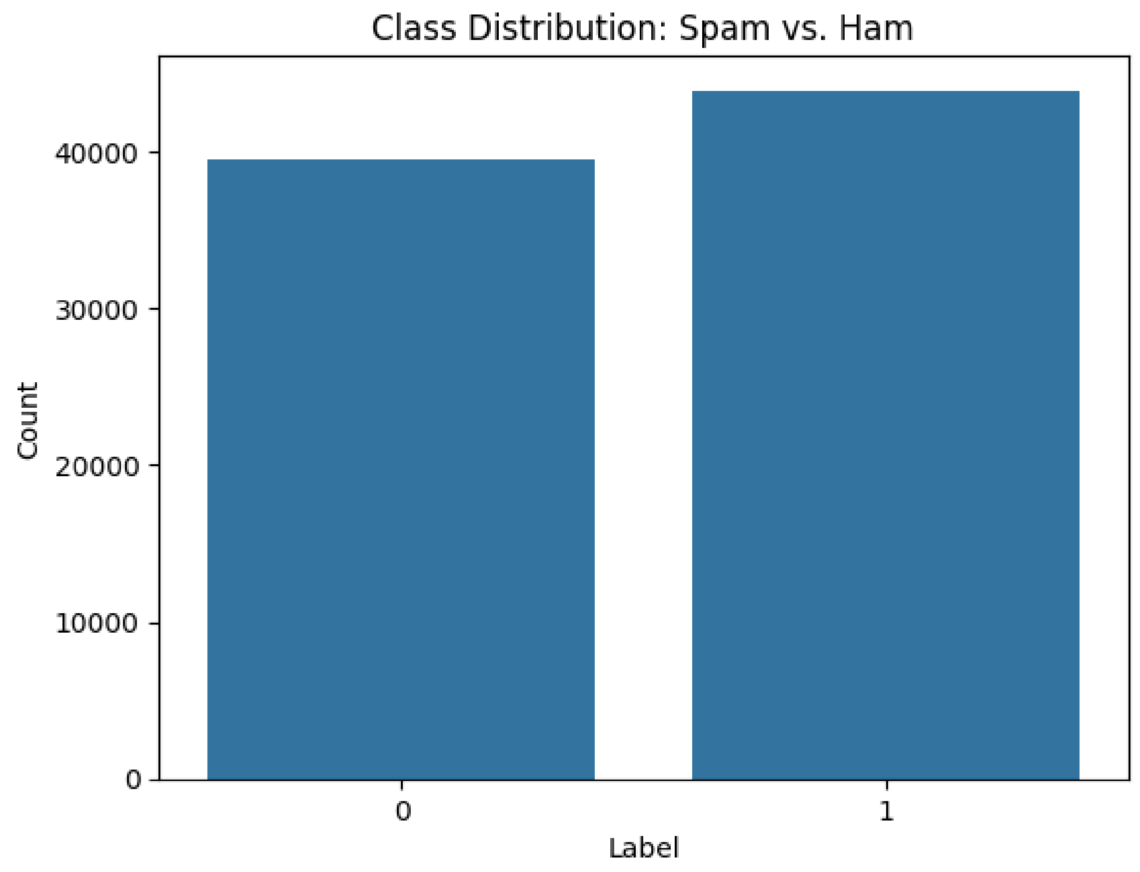
Logistic Regression (USE) misclassified "Congratulations" as spam due to keyword bias, while **BERT, with its deeper understanding of context, correctly identified it as "Ham."** **Neural Networks (USE/BERT)** performed well by **recognizing the message's genuine tone.**

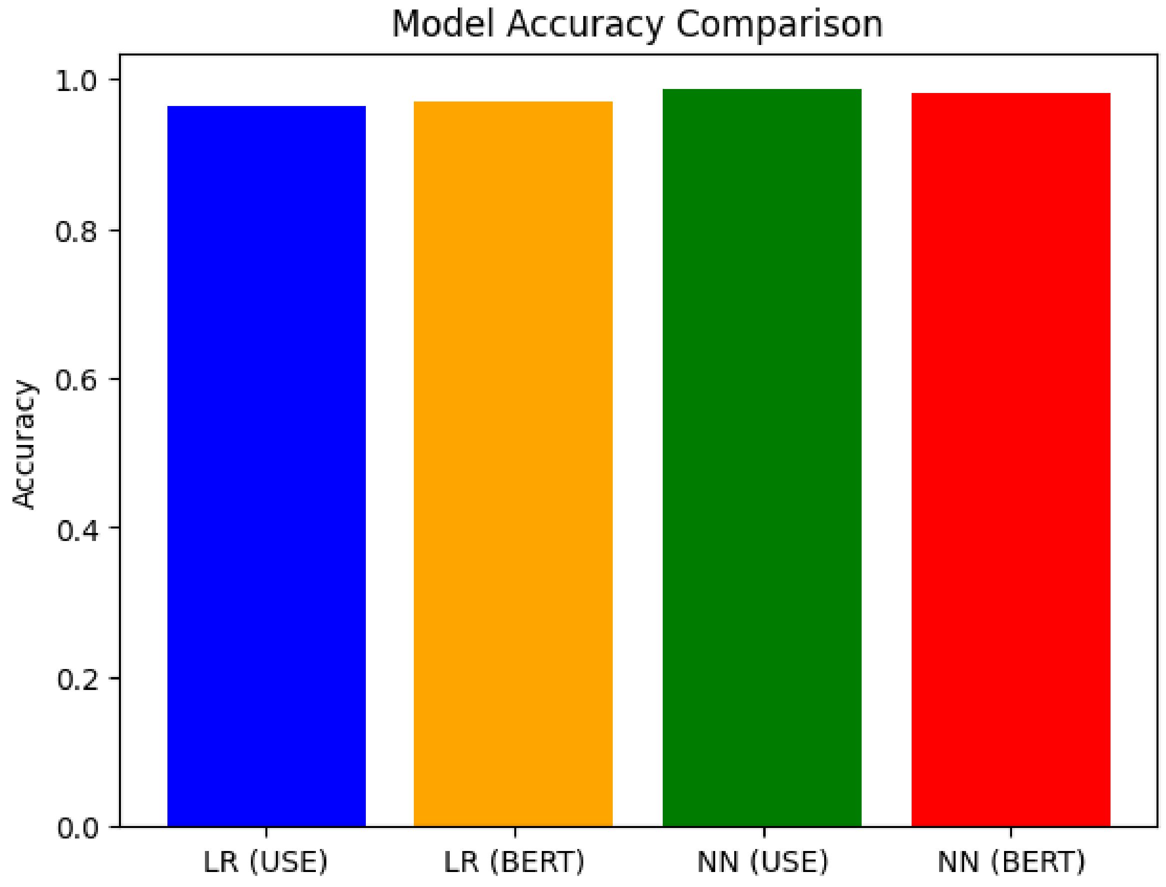
```
Enter the text (email/message) to classify as Spam or Ham: You have won 25000 dollars.
Logistic Regression Prediction (USE): Spam
Logistic Regression Prediction (BERT): Spam
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 27ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 45ms/step
Neural Network Prediction (USE): Spam
Neural Network Prediction (BERT): Spam
```

All models- Logistic Regression and Neural Networks using USE and BERT embeddings **accurately classified the message as Spam** due to the presence of **common spam indicators like monetary winnings and exaggerated claims.**

COMPARATIVE ANALYSIS

Class Distribution & Model Comparison



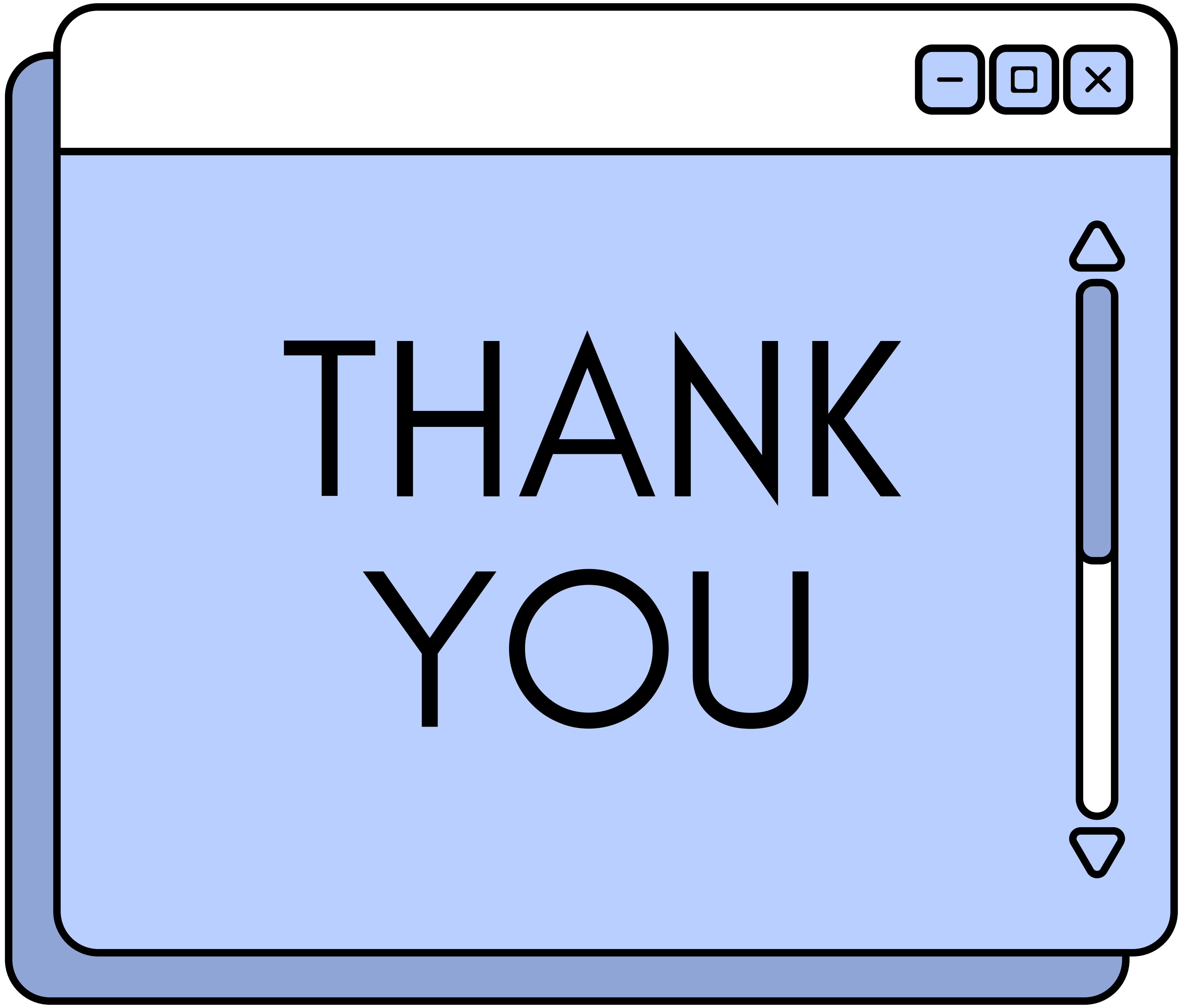


LR (USE) = 0.96

LR (BERT) = 0.97

NN (USE) = 0.98

NN (BERT) = 0.98


THANK YOU