FLATIRON CAPSTONE

By Allison Miller

BACKGROUND

In today’s world, messaging apps like Slack, Teams, and WhatsApp are used constantly across personal and professional settings.

But this leads to **information overload** — people miss important context, waste time reading long threads, or struggle to find decisions or action items buried in chats.

PROJECT GOALS

- Help save time, get the most important information, and improve communications by summarizing conversations
- Build an automated summarization model for dialogues.
- Train and evaluate it using the **SAMSum dataset**.
- Measure success using **ROUGE scores and qualitative comparison**.

METHODOLOGY

Used the **SAMSum dataset**, a collection of real, annotated conversations with human-written summaries

Removed null values, tokenized text using Hugging Face's tokenizers, and sample smaller subsets (100 training, 20 validation)

Fine-tuned facebook/bart-base using PyTorch and Hugging Face transformers.

METHODOLOGY

- Used small batches (batch size = 4) for GPU memory efficiency.
- Loss function: CrossEntropyLoss
- Optimizer: AdamW, LR = 5e-5
- Early stopping based on validation loss

METHODOLOGY

EVALUATION

We evaluated using **ROUGE-1**, **ROUGE-2**, and **ROUGE-L**:

- ROUGE-1: individual word overlap
- ROUGE-2: bigram overlap
- ROUGE-L: longest common subsequence (structure)
-

These metrics give a sense of **content coverage and fluency**.

RESULT

Our BART model achieved the following ROUGE scores:
ROUGE-1: 0.4104
ROUGE-2: 0.1746
ROUGE-L: 0.3394
ROUGE-Lsum: 0.3380

These are **solid results** given our small training set and short training time.
 ROUGE-1 above 0.4 means the model captured key words.
ROUGE-2 and L show it’s generating coherent, structured summaries.

ANALYSIS

- Handles short, task-based dialogues
- Preserves key decisions well
- Close to human-level summaries
- Struggles with long, emotional chats
- May drop small talk or context
- Limited by training data size

BUSINESS VALUE

- **Slack/Teams plugin** for auto-generated thread summaries

- **Customer service automation**
-
- **Meeting recap assistants**

CONCLUSION

- Built an effective dialogue summarizer
- Demonstrated strong ROUGE scores with limited data

Could improve with:

- More training data
- Larger model
- UI for user-friendly deployment

THANK YOU