CUFE-CHS 2023

Cairo University - Faculty of Engineering - Credit Hour System

Speak News

Submitted By:

- Ahmed Ehab Farghal
- Abdallah Wael Marzouk
- Mohamed Amr Afifi
- Youssef Kadry Hashem

Under Supervision of :

Dr AbdelMoniem Bayoumi

Abstract

We present an interactive news summarizer system using avatar narration and text-to-speech conversion. Our solution revolutionizes news consumption by providing concise summaries for effortless listening or visual experience. Using advanced NLP and ML techniques, our system generates accurate and digestible news summaries, enhancing personalized and efficient information consumption.

Datasets

- LJ Speech for Tacotron Model.
- CNN Daily Mail for Text Summarizer.
- BBC LRS2 Lip syncing dataset for Avatar Generation.
- Twitter Sentiment dataset for text sentiment analysis.

Models Architecture

The Modified BERT Encoder To Fit Our Summarization Task

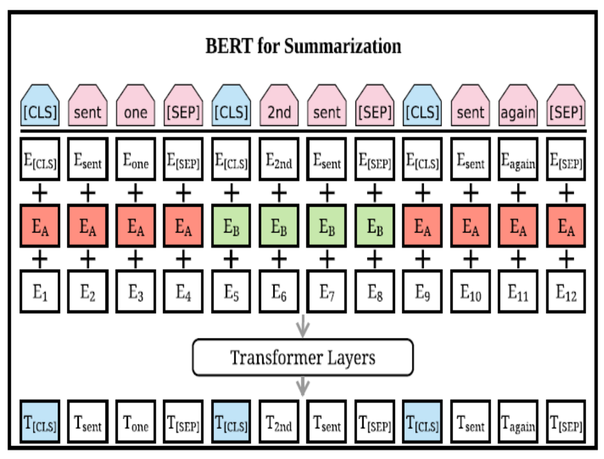


Figure 1. Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*.

The Transformer Encoder Decoder Model

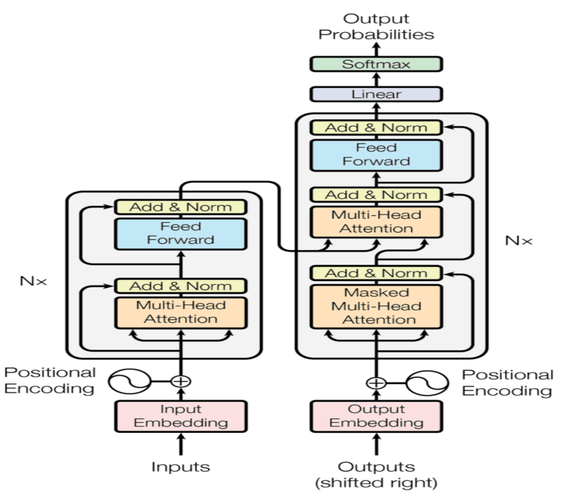
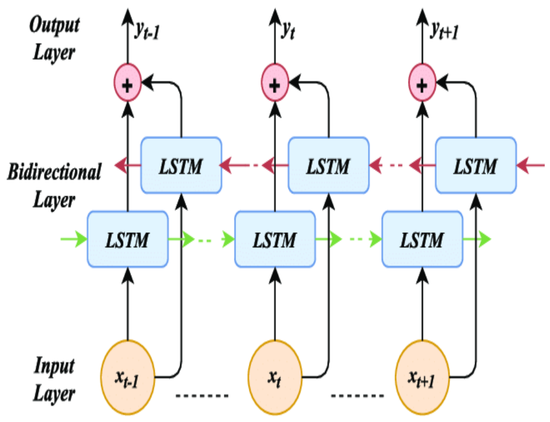


Figure 2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need.

The Bi-Directional LSTM Model



Tacotron Model Architecture

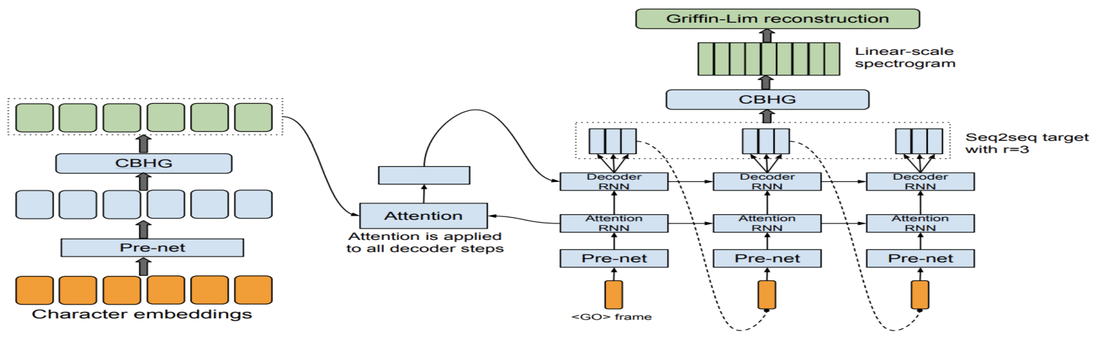


Figure 3. Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., ... & Saurous, R. A. (2017). Tacotron: Towards end-to-end speech synthesis.

Wav2lip Model Architecture

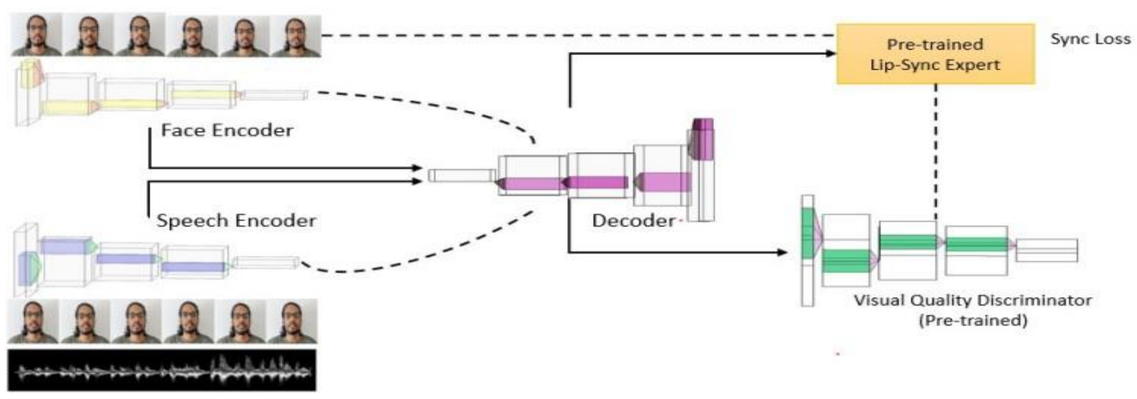
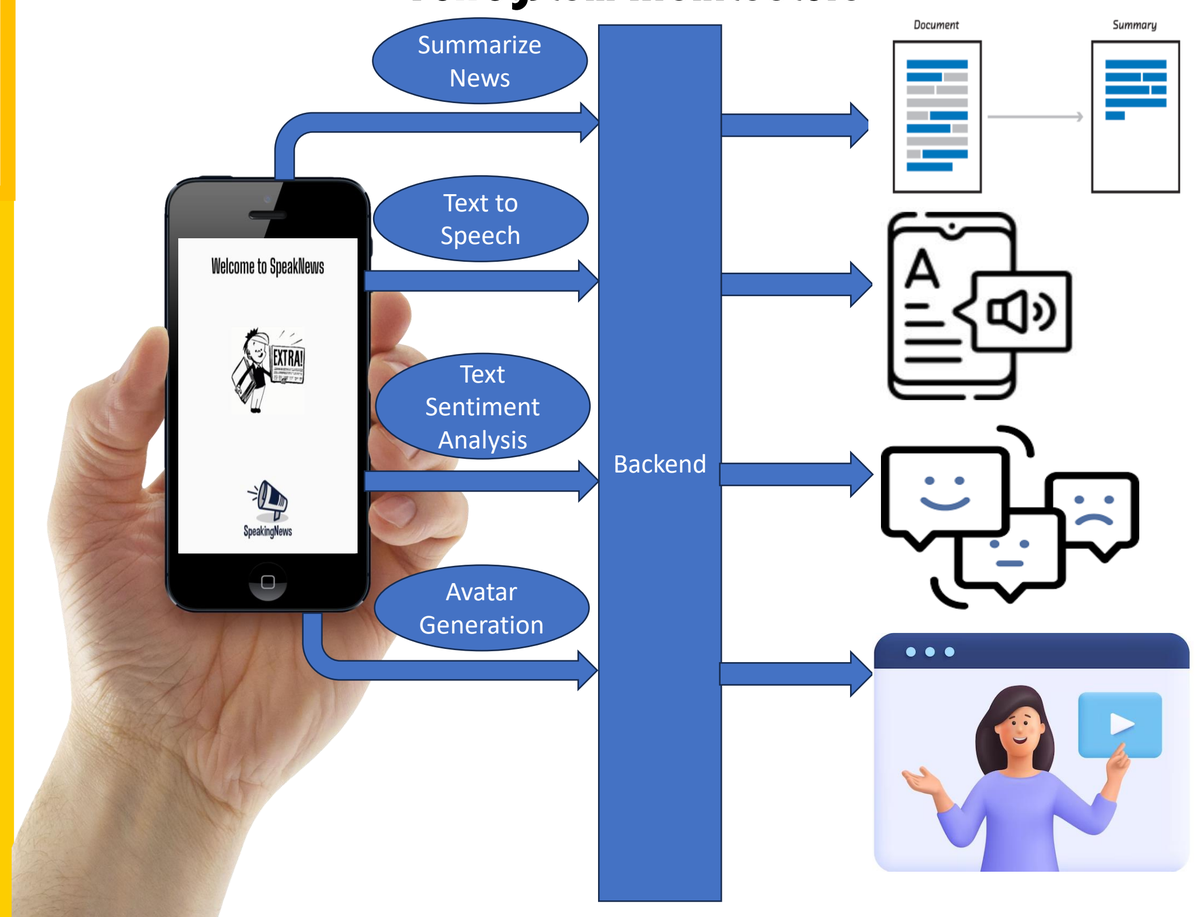
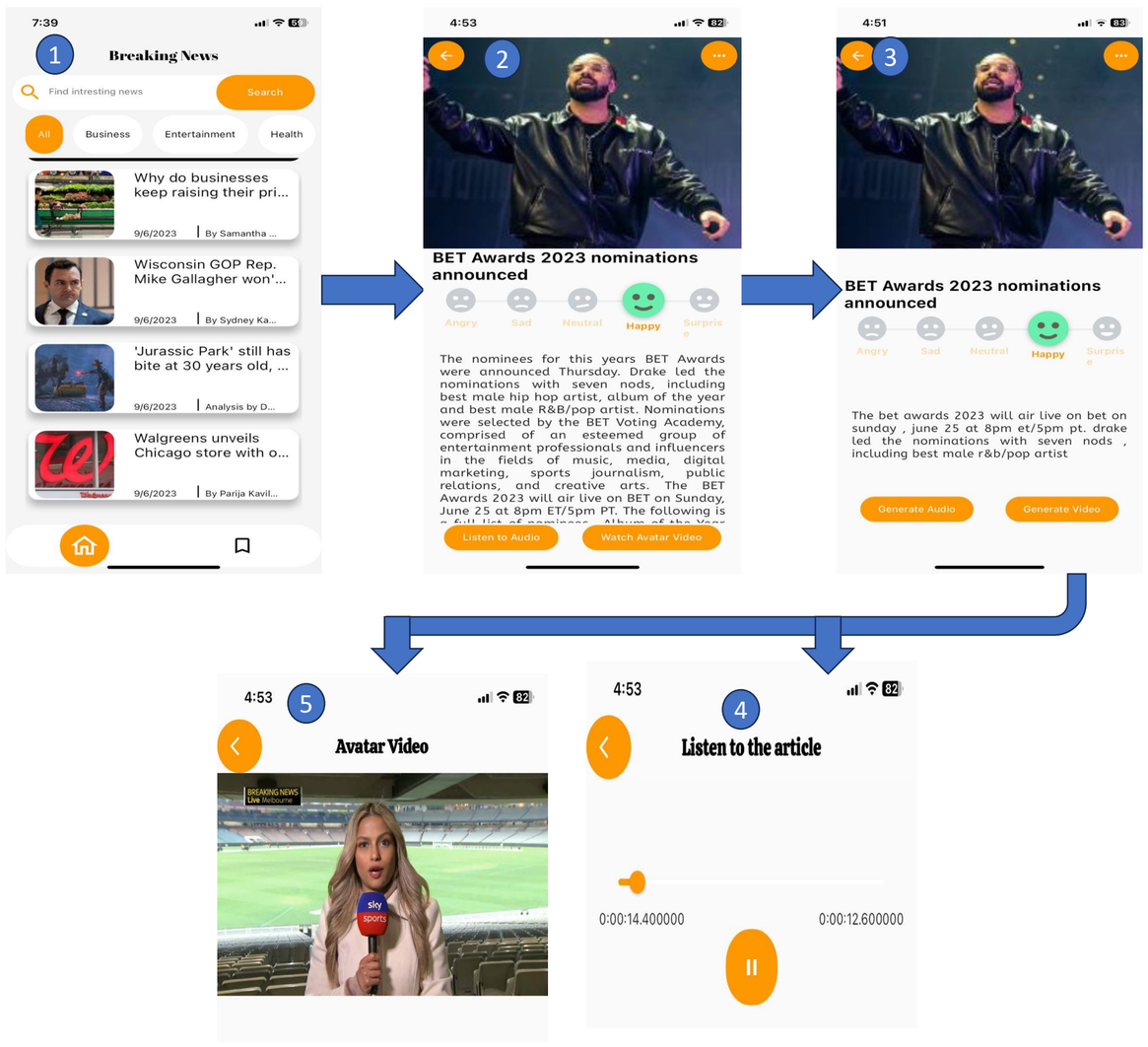


Figure 4. Prajwal, K. R., Mukhopadhyay, R., Namboodiri, V. P., & Jawahar, C. V. A lip sync expert is all you need for speech to lip generation in the wild.

Full System Architecture



SpeakNews in Action!



Results

Text Summarizer Results

Model	R1↑	R2↑	RL↑
Extractive Summarizers			
SUMO	41.00	18.40	37.20
TransformerEXT	40.90	18.02	37.17
BERTSUMEXT (Ours)	43.25	20.24	39.63
BERTSUMEXT w/o interval embeddings (Ours)	43.20	20.22	39.59
Abstractive Summarizers			
TransformerABS (Ours)	40.21	17.76	37.09
BERTABS	41.72	19.39	38.76
T5	45.62	25.58	36.53

Table 1: ROUGE F1 results on CNN/DailyMail test set (R1 and R2 are shorthands for unigram and bigram overlap; RL is the longest common subsequence).

Text to Speech Results

Model	Mean Opinion Score ↑
Tacotron (Ours)	3.82 ± 0.085
Parametric	3.69 ± 0.109
Concatenative	4.09 ± 0.119

Table 2: Mean Opinion Score results where the subjects were asked to rate the naturalness of the stimuli in a 5-point Likert scale score.

Text Sentiment Analysis Results

Method	Precision↑	Recall↑	F1-score↑	Accuracy↑
LSTM (Ours)	0.89	0.90	0.90	0.92
CNN	0.83	0.80	0.81	0.81
Bi-GRU	0.81	0.82	0.82	0.80
RNN	0.89	0.86	0.92	0.94

Table 3: The table showcases precision, recall, F1-score, and accuracy for each different approaches for Sentiment Analysis in different papers.

Wav2lip Results

Method	LSE-D↓	LSE-C ↑	FID↓
Without Lip-syncing	16.89	2.577	-
Speech2Vid	14.39	1.471	17.96
LipGAN	10.90	3.279	11.91
Wav2Lip (Ours)	9.53	6.41	13.65

Table 4: The table showcases Lip sync error difference, lip sync error confidence, and FrÃlchet Inception Distance.

Future Work

Improving the summarization architecture to handle a wider range of article types and domain-specific content. Moreover, Enhancing the speech synthesis quality, improving the prosody and intonation of the generated speech, and expanding language support. Finally ,we could the avatar animation to enhance realism and expressiveness.

Acknowledgements

We are grateful to Allah for providing us with the strength to overcome challenges. We express our sincere appreciation to Dr. AbdElMoniem Bayoumi for his invaluable support and guidance throughout our research and thesis writing. Lastly, we thank our families for their unwavering patience, encouragement, and support.

AEM AET **CCE** CEM CIE EEE HEM IEM MDE MEE MEM PPC SEE STE WEE