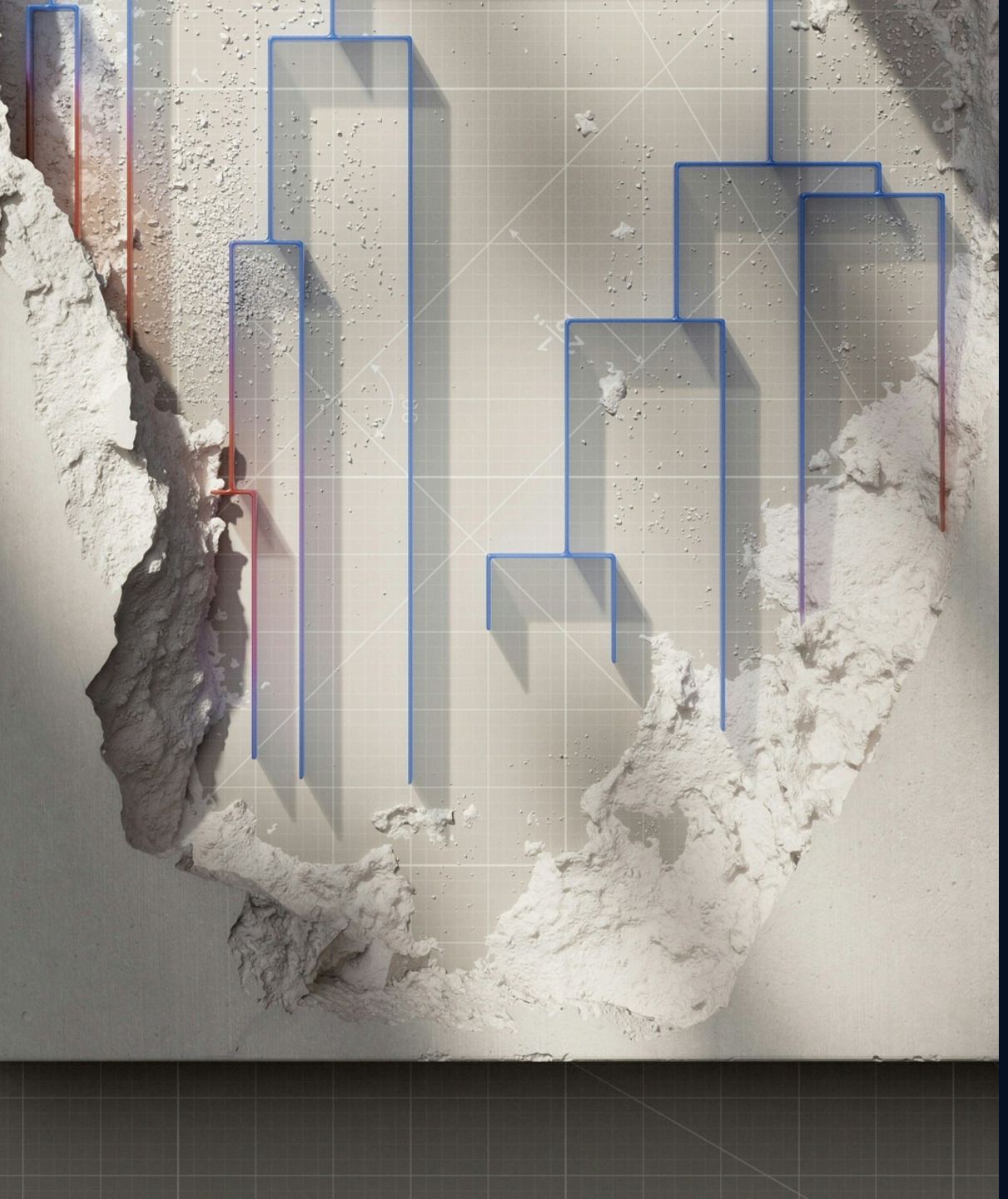
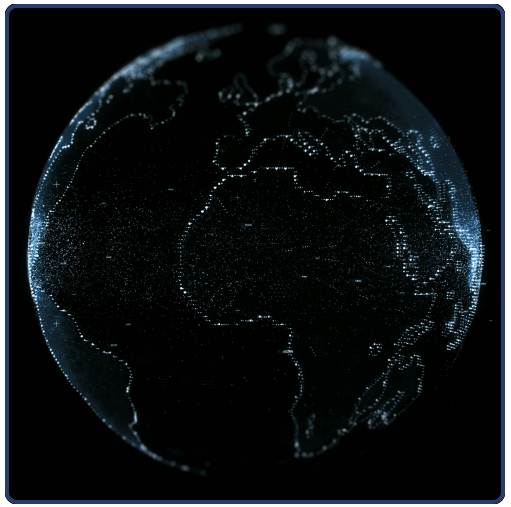


NEURAL NEWSROOM: AI-POWERED NEWS SUMMARIZATION

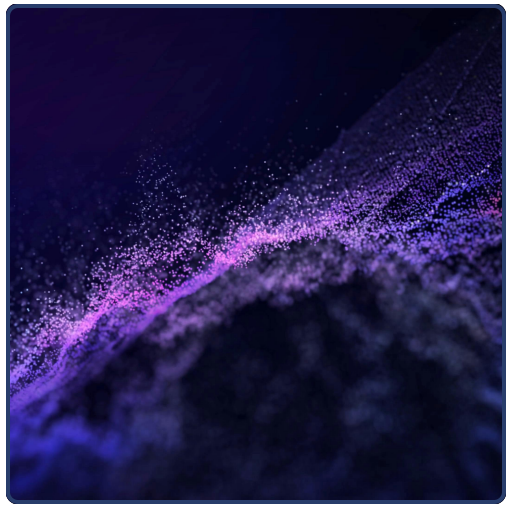
By La Shawn Sykes and Frank Hanan

APPROACH - TEAM ROLES



Frank Hanan

Machine Learning Engineer



La Shawn Sykes

Front-end Developer

Project Manager

EXECUTIVE SUMMARY

AI-Driven News Production

AI algorithms can automatically generate news articles, headlines, and even entire stories, reducing the workload on human journalists.

Personalized News Feeds

AI-powered recommendation systems can curate news content tailored to individual user preferences, improving user engagement.

Multilingual Translation

AI-powered translation tools can instantly convert news articles into multiple languages, enabling global distribution and accessibility.

The integration of AI technology into the news industry is revolutionizing how news is produced, distributed, and consumed, leading to increased efficiency, personalization, and reliability.

DATA COLLECTION

- ## Accessing the Guardian API

  The Guardian API provides a way to retrieve news articles and related data from the Guardian news organization. This API can be used to collect data for news summarization and analysis.

- ## Retrieving Article Data

  The Guardian API allows developers to search for and retrieve specific articles based on keywords, topics, or other criteria. This can be used to gather a dataset of news articles for processing and analysis.

- ## Handling API Requests

  Interacting with the Guardian API involves making HTTP requests to the appropriate API endpoints and handling the response data, which is typically returned in JSON format. This requires managing API authentication, rate limiting, and error handling.

- ## Storing and Preprocessing Data

  Once the article data is retrieved from the Guardian API, it needs to be stored and preprocessed for use in the neural newsroom system. This involve cleaning the text, extracting relevant metadata, and converting the data into a format suitable for machine learning tasks.

DATA CLEAN-UP AND EXPLORATION

Data Cleaning Approach	Insights Gained
Removed missing values and outliers using established statistical methods	Identified key variables that had a significant impact on the target variable
Performed feature engineering to create new variables that better captured the underlying relationships in the data	Uncovered previously unknown patterns and trends that provided valuable context for the neural network model

*Adapted from the analysis conducted by the data science team at the 'Neural Newsroom' project.

APPROACH - TECHNOLOGIES

- **Python:Programming Language**
  Chosen for its robust libraries and ease of use in data processing and machine learning
- **TensorFlow/Keras:Deep Learning**
  Utilized to build and train our CNN-LSTM hybrid model, offering flexibility and powerful neural network capabilities
- **The Guardian API:Data Source**
  Provided a rich, diverse dataset of news articles, allowing real-time access to current news
- **NLTK: Text Processing**
  Natural Language Toolkit, employed for tokenization, stop word removal, and other essential NLP tasks in our data preprocessing pipeline
- **Gradio:User Interface**
  Enabled the creation of an intuitive, interactive web interface for users to input queries and receive summarized news
- **Matplotlib: Data Visualization**
  Used to create training history plots, confusion matrices, and other visual representations for model evaluation and analysis

DATA IS A STORY

CONFUSION MATRIX

business	culture	environment	politics	science	sport	technology	world	
21	0	1	0	0	0	0	0	bus
0	5	0	0	0	0	1	0	cul
0	0	13	0	0	0	0	0	env
0	0	0	9	0	0	0	0	pol
0	0	0	0		0	0	1	sci
1	3	1	0	0	105	0	1	spo
1	0	0	0	2	0	0	0	tec
2	0	12	6	1	0	3	58	wor

CHALLENGES AND SUCCESSES

Handling Noisy Input Data

Overcome the challenge of processing user-generated news articles with inconsistent formatting, spelling errors, and irrelevant information.

Achieving High Summarization Accuracy

Developed advanced natural language processing techniques to generate concise and informative summaries that capture the key points of the original news articles.

Ensuring Timely Delivery

Implemented efficient algorithms and optimized system architecture to provide news summaries to users in near-real-time, meeting their demand for up-to-date information.

Enhancing Multilingual Support

Extended the system to handle a wide range of languages, enabling users from diverse linguistic backgrounds to access personalized news summaries.

Improving User Experience

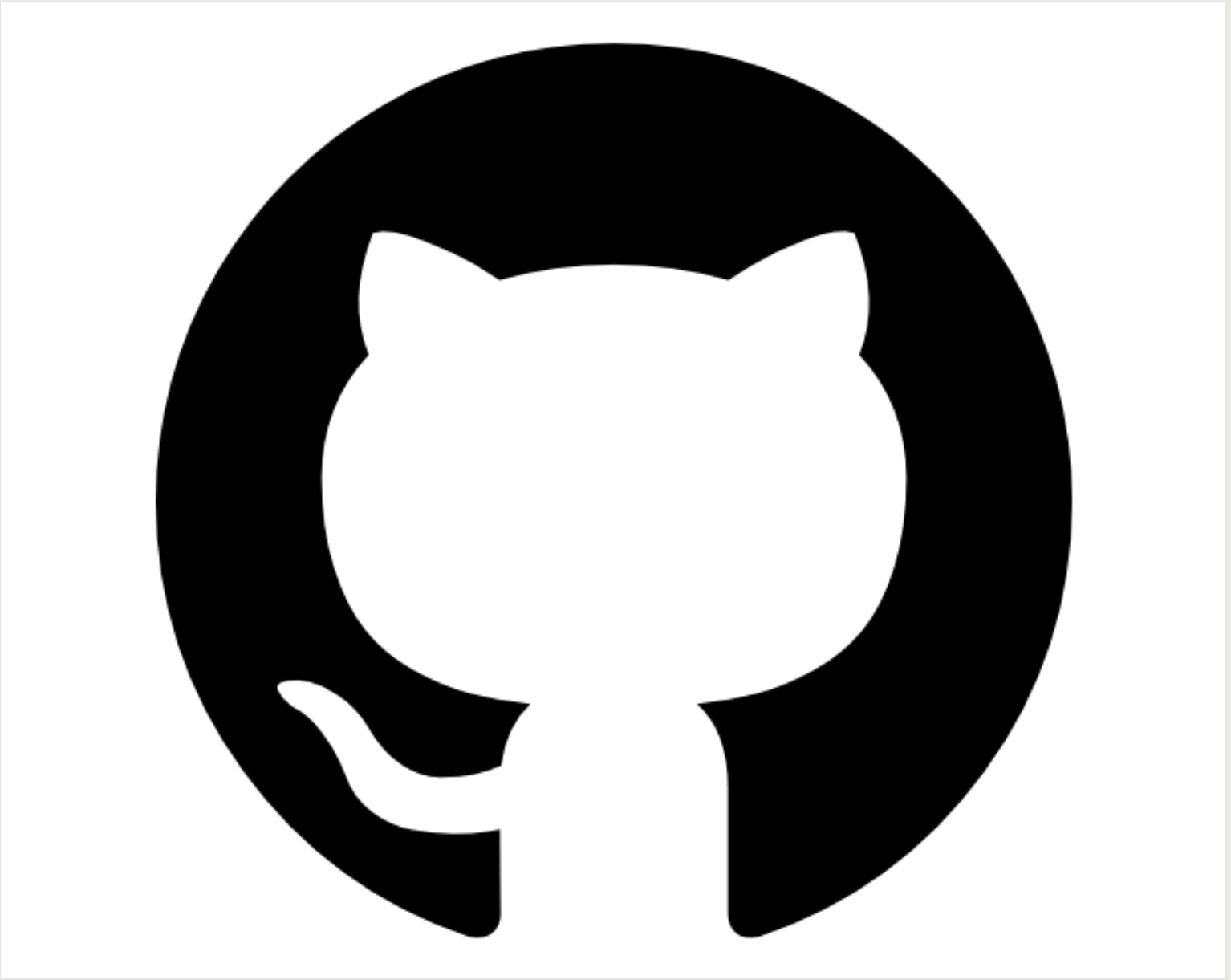
Iteratively refined the user interface and interaction design to create an intuitive and engaging experience for users, encouraging regular usage of the news summarization service.

NEXT STEPS

- Expand NLP Models
- Enhance Summarization Accuracy
- Integrate Multi-Modal Processing
- Improve User Interface Design

GITHUB

Project Data Repository



DEMO