technical_documentation.md

markdown
CopyEdit
Technical Documentation
NewsBot Intelligence System 2.0
Author: Grisel Barrera
Version: 2.0
Date: August 2025

---

1. Overview
The **NewsBot Intelligence System 2.0** is an advanced Natural Language Processing (NLP) platform designed to process, classify, analyze, and generate news content. It builds on the Midterm NewsBot foundation, integrating additional features for **topic modeling, multilingual analysis, content generation, and conversational interaction**.

This production-ready system demonstrates mastery of:
- Advanced Topic Modeling (**LDA**, **NMF**)
- Pre-trained Language Models for text generation
- Cross-language analysis with translation workflows
- Conversational AI interface
- Scalable architecture for large-scale text processing

---

2. Objectives
- Automate classification and analysis of news articles.
- Discover emerging topics and trends.
- Provide multilingual capabilities for global news analysis.
- Generate human-like news summaries and responses.
- Offer a natural language query interface for interactive exploration.

---

3. System Architecture

```
+----------------------+
| Data Ingestion Layer |
| (BBC Dataset, APIs)  |
+----------+-----------+
           |
           v
+----------------------+
| Preprocessing Layer  |
| (Cleaning, Tokenizing)|
+----------+-----------+
           |
           v
+----------------------+
| Feature Extraction   |
```

```
| (TF-IDF, Word2Vec) |
+---------+-----------+
|
v
+--------------------+
| Model Layer |
| - LDA/NMF Topic Models|
| - Classification Model|
| - Transformer-based LM|
+---------+-----------+
|
v
+--------------------+
| Output Layer |
| (Summaries, Insights, |
| Interactive Queries) |
+--------------------+
```

```yaml
CopyEdit

---

## 4. Features
1. **Data Preprocessing** – Text cleaning, stop word removal, lemmatization.
2. **Feature Extraction** – TF-IDF and word embeddings.
3. **Classification** – Logistic Regression, SVM, Transformers.
4. **Topic Modeling** – LDA and NMF.
5. **Multilingual Analysis** – Translation integration.
6. **Conversational AI** – Q&A interface.
7. **Content Generation** – AI-generated summaries.

---

## 5. Installation
### Requirements
- Python 3.10+
- Virtual environment recommended

### Setup
```bash
git clone https://github.com/Grisel-Barrera-HCC-AI/NewsBot-Intelligence-
System-2.0.git
cd NewsBot-Intelligence-System-2.0
python -m venv venv
source venv/bin/activate  # Windows: venv\Scripts\activate
pip install -r requirements.txt
```

6. Usage

```
bash
CopyEdit
# Train the models
python train_model.py

# Launch the chatbot
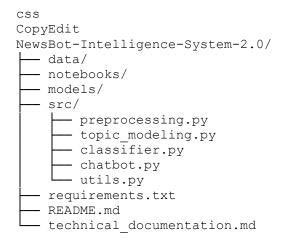python chatbot.py
```

---

7. Example

Input:
```
Tesla announces new battery technology that will improve range by 50%.
```

Output:

- **Category:** Technology
- **Topics:** renewable energy, EV innovation
- **Summary:** Tesla unveiled a breakthrough battery design offering a 50% range increase.

---

8. File Structure

```
css
CopyEdit
NewsBot-Intelligence-System-2.0/
├── data/
├── notebooks/
├── models/
├── src/
│   ├── preprocessing.py
│   ├── topic_modeling.py
│   ├── classifier.py
│   ├── chatbot.py
│   └── utils.py
├── requirements.txt
├── README.md
└── technical_documentation.md
```

---

9. Troubleshooting

Issue	Cause	Solution
`ModuleNotFoundError`	Missing dependency	`pip install -r requirements.txt`
Encoding error	Non-UTF8 dataset	Load with `encoding='utf-8'`
Slow training	Large dataset	Reduce dataset size

---

10. Future Improvements

- Real-time news scraping.
- More languages for multilingual analysis.
- Sentiment analysis.
- Flask/FastAPI deployment.

---

11. References

- Scikit-learn
- Gensim
- Hugging Face Transformers
- BBC News Dataset

```
yaml
CopyEdit

---

## **`README.md`**
```markdown
# NewsBot Intelligence System 2.0

## Overview
The **NewsBot Intelligence System 2.0** is an advanced NLP platform for
automated news classification, topic modeling, multilingual analysis, and AI-
generated summaries. It builds on the original NewsBot with enhanced
features, production-ready architecture, and a conversational AI interface.

---

## Features
- **Topic Modeling** with LDA & NMF
- **Classification** using TF-IDF, embeddings, and ML models
- **Multilingual Analysis** with translation workflows
- **Conversational AI** for Q&A interaction
- **Content Generation** using transformer-based models

---

## Installation
```bash
git clone https://github.com/Grisel-Barrera-HCC-AI/NewsBot-Intelligence-
System-2.0.git
cd NewsBot-Intelligence-System-2.0
python -m venv venv
source venv/bin/activate  # Windows: venv\Scripts\activate
pip install -r requirements.txt
```

Usage

```bash
CopyEdit
# Train the models
python train_model.py

# Launch the chatbot
python chatbot.py
```

Example Output

Input:
```
Tesla announces new battery technology that will improve range by 50%.
```

Output:

- **Category:** Technology
- **Topics:** renewable energy, EV innovation
- **Summary:** Tesla unveiled a breakthrough battery design offering a 50% range increase.

File Structure

```css
CopyEdit
NewsBot-Intelligence-System-2.0/
├── data/
├── notebooks/
├── models/
├── src/
│    ├── preprocessing.py
│    ├── topic_modeling.py
│    ├── classifier.py
│    ├── chatbot.py
│    └── utils.py
├── requirements.txt
├── README.md
└── technical_documentation.md
```

Author

Grisel Barrera – Applied AI & Robotics Program, Houston Community College

```yaml
CopyEdit
```