NLP Project Report: TrendStory

## 1. Introduction

TrendStory is an NLP project that extracts real-time trending topics from Google News and YouTube, and generates contextual stories using advanced language models. The project supports story generation in various tones such as tragic, hopeful, and poetic. It includes a RESTful API, a gRPC interface, a user-friendly Gradio UI, Docker containerization, and automated testing with Postman and Python.

## 2. Features Implemented

- Combined trend extraction using GNews and YouTube APIs
- Story generation using Microsoft Phi-2 model
- Tone customization: tragic, hopeful, poetic
- REST API via FastAPI
- gRPC interface using Protobuf and grpcio
- Docker containerization for deployment
- Gradio interface for demo
- Postman collection for endpoint testing
- Python test suite with performance evaluation

## 3. Implementation Steps

- Set up trend extraction using GNews and YouTube API keys
- Developed FastAPI app with `/generate_story` and `/trends` endpoints
- Created Dockerfile for containerized deployment
- Implemented gRPC server with matching client and compiled protobuf
- Built Gradio UI with dropdown and TTS support
- Generated Postman collections to test API under different cases

- Added `test_api.py` script to validate endpoints and benchmark performance

4. Testing & Evaluation

- Postman tested: `/generate_story` (valid, missing topic, invalid tone)

- Python tested: performance test (5 stories in sequence)

- Sample test output attached in `test_api.py`

- JSON responses validated

5. Model Used

Microsoft Phi-2 (https://huggingface.co/microsoft/phi-2)

- Transformer-based causal language model

- Used via Hugging Face Transformers library

- CPU-only fallback supported

6. Known Limitations

- Phi-2 model requires high RAM or GPU to run efficiently

- Trends may not load properly if APIs fail or internet is down

- No fine-tuning yet for Roman Urdu

7. Conclusion

TrendStory successfully integrates trend monitoring and AI-driven storytelling in a scalable, modular pipeline. It demonstrates key skills in LLM usage, API development, containerization, and interface design.