“Myanmar Fake News Classification in Social Media Using Deep Learning”

1. Problem Statement

Fake news refers to false or misleading information presented as legitimate news. In Myanmar, the spread of fake news on social media platforms such as Facebook has increased significantly in recent years, particularly during political events, health crises, and natural disasters. This misinformation can cause public confusion, escalate social tensions, and harm trust in reliable information sources.

Detecting and filtering fake news in Myanmar is especially challenging because:

- Many fake stories closely mimic the style of real news.
- Burmese is a low-resource language with limited NLP tools.
- Current content moderation systems often fail to handle Myanmar text effectively.

This project aims to develop a deep learning model capable of classifying Myanmar-language news headlines and short articles as *Real* or *Fake*. By addressing the fake news problem, this project will contribute to a safer, more informed online environment.

2. Input and Output

Input: Myanmar-language news headlines or short articles in Unicode format.

Samples:

- “ရန်ကုန်မြို့တွင် မိုးအလွန်များ၍ မြစ်စီးသွားမည်ဟု အစိုးရကြေညာ”
- “ကမ္ဘာပေါ်တွင် ငြိမ်းချမ်းရေးအဆုံးသတ်ပြီး စစ်ပွဲအသစ် စတင်”

Output:

- Predicted Label **: Fake or Real**
- Along with a confidence score indicating the model's certainty.

Example Output Table:

Input Text	**Predicted Label**	**Confidence**
ရန်ကုန်မြို့တွင် မိုးအလွန်များ၍ မြစ်စီးသွားမည်ဟု အစိုးရကြေညာ	Fake	0.94
ကမ္ဘာပေါ်တွင် ငြိမ်းချမ်းရေးအဆုံးသတ်ပြီး စစ်ပွဲအသစ်စတင်"	Real	0.87

3. Dataset

Type: Binary Burmese text classification dataset (Fake vs. Real)

Sources:

- Real news:
  - Eleven Media, The Irrawaddy, Myanmar Now
- Fake news:
  - Myanmar Fact Check, AFP Fact Check Burmese
  - Archived Facebook posts flagged for misinformation

Planned Size: At least 1,000 labeled comments, balanced across both classes (500 Fake, 500 Real).

Labeling Process:

- A headline/article will be marked as *Fake* if verified by fact-checking sources or if it contains demonstrably false information.

- Label Studio: to define two labels such as Fake (1) vs. Real (0).

4. Expected Performance

With a clean and balanced dataset, and by fine-tuning a multilingual transformer such as XLM-RoBERTa, the aim is to achieve:

- Accuracy: 70% or higher
- Macro F1-Score: 70% or higher

Performance will be evaluated on a held-out test set to ensure he model works on unseen data. Additional evaluation, such as confusion matrix analysis, will be conducted to identify errors.

5. Motivation

The spread of fake news has serious social consequences in Myanmar. By developing a Myanmar fake news detection system, this project will:

- Support fact-checkers and content moderators in identifying false information
- Help reduce misinformation spread on social media
- Contribute to the development of Burmese NLP resources and models

This project is personally motivating as it addresses a critical real-world problem and uses deep learning to create a practical, socially impactful solution.