Date: July 18, 2025

Developer: Dua Mohyyuddin

Project: Robert Frost Poetry Generator

---

1. Today's Objectives

- [✔] Implement Byte Pair Encoding (BPE) tokenizer
- [✔] Upgrade model architecture for poetry generation
- [✔] Conduct performance benchmarking
- [✔] Team review and supervisor evaluation

---

2. Key Accomplishments

A. Tokenizer Implementation

Task: Replaced character-level tokenization with BPE

Technical Details:

- Built `BPETokenizer` class with merge operations
- Vocabulary size: 512 (256 bytes + 256 merges)
- Special handling for poetic structures (newlines, punctuation)

Validation:

```python
Input: "The woods are lovely, dark and deep,"
Encoded: [388, 119, 498, 262, 275, 256, 395, 118, 415]
Decoded: "The woods are lovely, dark and deep," (100% accurate)
```

Impact:

- 4.2x more efficient than character-level
- Better preserves poetic meter

B. Model Architecture Upgrade

Improvements:

Component	Before	After
Layers	4	8
Hidden Size	64	128
Attention Heads	4	8
Params	1.1M	5.7M

New Features:

- Gated Linear Units (GLU) in feedforward layers
- Dropout (p=0.1) for regularization
- Enhanced rhyme guidance during generation

C. Training Optimization

Key Changes:

1. OneCycleLR Scheduler:
    - Max LR: 3e-3 → Automatic warmup/cooldown
2. Gradient Clipping:
    - Threshold: 1.0 (prevents explosions)
3. Mixed Precision:
    - 2.8x faster training on GPU

Monitoring:

```
text
iter 500: loss 2.14, lr 2.7e-3, grad_norm 0.87
```

---

3. Evaluation Results

A. Generation Samples

Temperature Comparison:

- `temp=0.5`:
- text

```
The woods are lovely, dark and deep,
```

- But I have promises to keep...
- `temp=1.2`:
- text

```
A violet breeze through frozen trees,
```

- Whispering winter's lullabies...

Rhyme-Guided Output:

```
text
Prompt: "snow" → Rhymes: ["flow", "glow", "know"]
Generated:
"Where icy winds eternally blow,
And quiet blankets soften the snow."
```

B. Performance Metrics

Metric	Before	After

Training Speed	120it/s	340it/s
Val Loss	3.21	2.07
Rhyme Accuracy	62%	88%

---

4. Team Discussion Highlights

Supervisor Feedback:

- ✅ Excellent tokenizer efficiency
- ✅ Improved poetic coherence
- ⚠️ Suggestion: Add beam search for generation

Action Items:

1. Implement beam search (priority)
2. Expand rhyme dictionary
3. Test with larger dataset