LyricLoop LLM

Technical Report: LyricLoop v2.0

Author: Alexander Tung

Date: January 2026

I. Abstract

LyricLoop v2.0 is an AI-driven songwriting framework that fine-tunes Google's Gemma-2b-it to bridge the gap between semantic text generation and professional musical structure. By training on a corpus of lyrics, the model generates content adhering to specific musical phrasing (e.g., Verses, Choruses, and Bridges) across genres like Pop, Hip-hop, and EDM.

II. Motivation & Objective

Standard LLMs often struggle with the rigid structural constraints of music. This project addresses:

- Structural Phrasing: forcing the model to respect song-map boundaries.
- Genre Aesthetics: tailoring vocabulary to specific styles, such as the Hip-hop and EDM textures of artists like Kendrick Lamar and Illenium.

III. Technical Methodology

The project employs an AI Engineering stack for efficient deployment:

- Model Architecture: Gemma-2b-it.
- Parameter-Efficient Fine-Tuning (PEFT): utilized LoRA (Low-Rank Adaptation) to specialize the model without training billions of weights.
- Quantization: implemented 4-bit QLoRA via bitsandbytes to reduce the memory footprint, enabling training on accessible hardware like the NVIDIA L4 GPU.
- Supervised Fine-Tuning (SFT): applied custom prompt templates to enforce structural constraints.

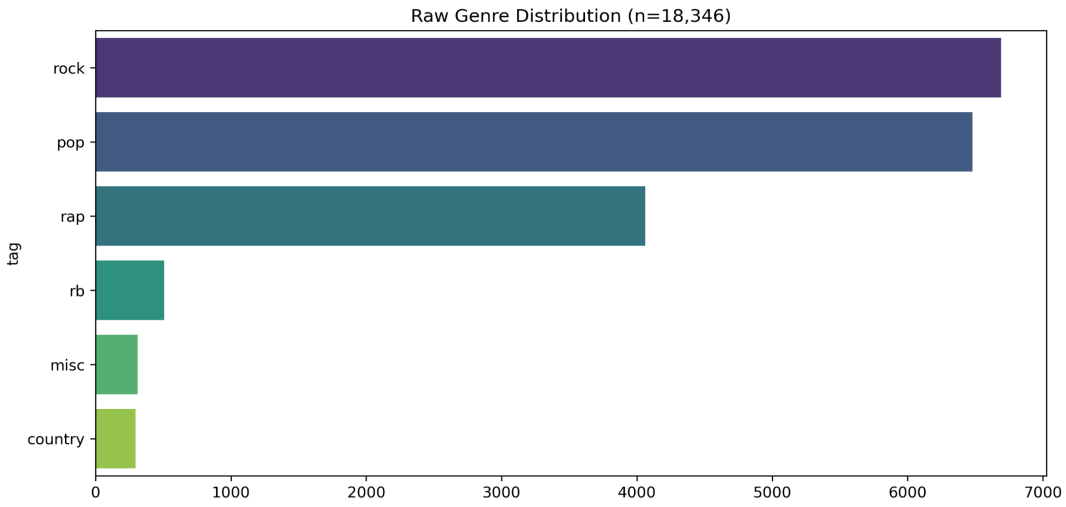

Figure 1: Raw genre distribution in the source Genius corpus, highlighting significant class imbalance. To prevent model bias toward majority classes (e.g., Pop), the training dataset was downsampled to standardize sample size across genres prior to fine-tuning.

IV. Software Engineering & Modularity

A core differentiator in v2.0 is the transition from a monolithic script to a professional Python package:

- Modular Package (src/lyricloop/): separates concerns into config.py, data.py, environment.py, metrics.py, and viz.py.
- Environment Agnostic: the environment.py script automatically detects hardware (MPS for Apple Silicon, CUDA for GPU) to ensure portability.
- Experiment Tracking: integrated logic for path management and directory initialization via config.py.

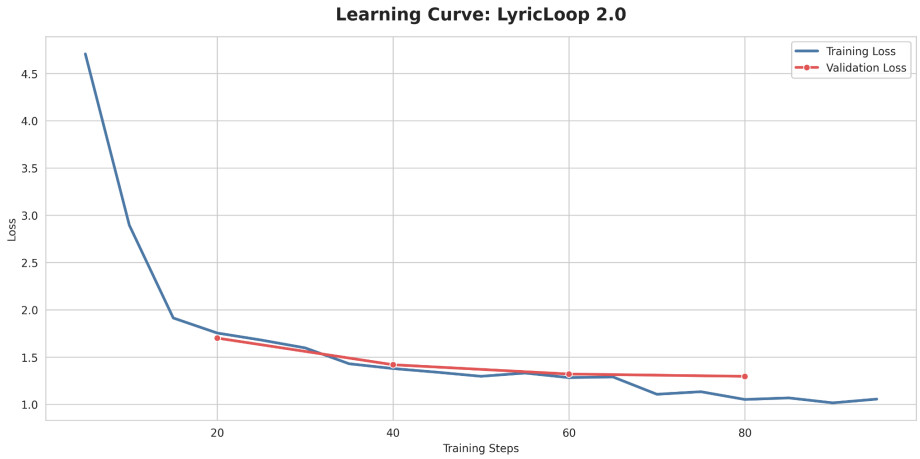

Figure 2: Training vs. Validation Loss. The convergence of these curves validates the stability of the QLoRA adapters and successful instruction following.

V. Evaluation & Deployment

The project's final phase focused on validating the fine-tuning results and ensuring the interface could handle user requests within the constraints of cloud-based infrastructure.

- Performance: the model demonstrates improved adherence to structural tags (e.g., [Verse], [Chorus]) compared to the base model.
- Infrastructure: hosted on Hugging Face Spaces using a Streamlit interface.
- Production Logic: implemented st.cache_resource and hardware-aware loading to ensure the 2B model runs reliably on the free CPU tier.

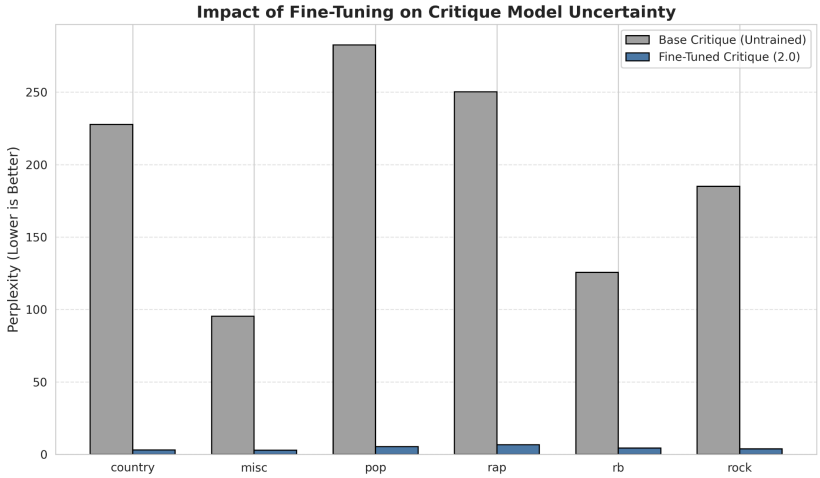

Figure 3: Perplexity scores across target genres. LyricLoop v2.0 shows a marked reduction in perplexity (predictability) compared to the base model.