**Objective:-**Design and develop a large language model (LLM) that generates human like, professional LinkedIn posts across various themes. The model should eventually work without any dependency on external APIs (example:- OpenAI, Anthropic).

# **Project Title:**

**LLM-Powered LinkedIn Post Generator - A Custom Fine-Tuned TinyLLaMA Model**

In a world where personal branding is key, writing engaging LinkedIn posts that feel professional yet human can be challenging especially across different themes like internships, promotions, certifications, and project completions.

So I decided to build a **custom Large Language Model (LLM)** that **generates human-like LinkedIn posts** tailored to different tones and themes **completely offline**, with **no reliance on OpenAI, Anthropic, or any external API.** The goal was to create a model that could generate professional posts with just a theme and tone as input and I’m proud to say the model now does exactly that!

## **Workflow**

### **Dataset Creation**

I started by curating a dataset of **340+ input-output pairs** in JSONL format.

Each data sample included:

* An instruction: the input prompt (e.g., “Write a humble LinkedIn post about internship”)
* A generated output: a well-written, human-style post relevant to the prompt
* Metadata like theme and tone to guide model learning

This dataset was completely handcrafted to ensure diversity and realism.

### **Model Selection**

I chose [**TinyLLaMA-1.1B-Chat**](https://huggingface.co/TinyLlama/TinyLlama-1.1B-Chat-v1.0) a small yet powerful open-source model with 1.1 billion parameters. It’s perfect for fine-tuning on personal machines without requiring GPUs or cloud APIs.

### **Fine-Tuning the Model**

* I fine-tuned the model using Hugging Face’s **Trainer** with **PEFT (Parameter-Efficient Fine-Tuning)** using LoRA.
* Training was done on my local machine using Apple Silicon’s **MPS** backend via PyTorch.
* The training ran for **3 epochs** over the dataset, optimizing the model to align with tone, theme, and style.

During training, loss decreased steadily, indicating that the model was learning well:

Initial loss: ~2.6

Final loss: ~0.2

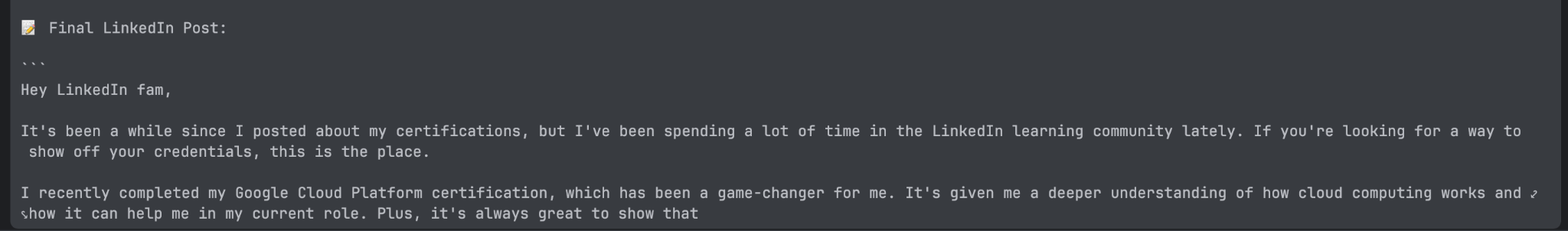
### **Inference and Generation**

Post fine-tuning, I created a generate\_custom.py script that takes in a custom theme and tone like:

{"theme": "certification", "tone": "confident", "post"

The model then **generates the rest of the post** in real-time — often producing fluent, context-aware, professional LinkedIn posts.

Here’s a output i received while training:



## **Achievements**

* Built an original dataset tailored to LinkedIn communication
* Successfully fine-tuned a Transformer-based model locally
* Eliminated all external API dependencies — **runs 100% offline**
* Generated human-like, structured LinkedIn posts across 20+ themes
* Compatible with MacBooks using Apple Silicon (MPS backend)

## **Tech Stack**

* **Model**: TinyLLaMA-1.1B-Chat
* **Libraries**: Hugging Face Transformers, Datasets, PEFT, BitsandBytes
* **Platform**: MacOS (M3)
* **Frameworks**: PyTorch, PEFT, Accelerate

## **What I Learned**

This project helped me dive deep into:

* Prompt engineering and dataset design
* Fine-tuning LLMs on resource-limited hardware
* Understanding LoRA-based PEFT
* Working with tokenizers and decoding strategies
* Designing ethical, locally deployable AI systems

## **Final Thoughts**

This project was more than a technical experiment — it was about combining **AI capabilities with real-world communication needs.** It was rewarding to see how a small model, properly fine-tuned, can generate posts that feel personal, professional, and ready to publish.

And the best part? It runs entirely on my laptop no cloud, no API bills. Just local, personalized, intelligent LinkedIn writing assistance.

I have attached all the program files and results in the drive link:

[LLM linkedin](https://drive.google.com/drive/folders/1_Z1jySCzths3D7O92ELXBXi1iajM4glW?usp=sharing)