

Enhancing Genre Fidelity in Children’s Story Generation via Fine-Tuned LLMs

1st Sudipto Pramanik
Department of EEE
Daffodil International University

2nd Anik Chawdhury
Department of EEE
Daffodil International University

3rd Pallab Biswas
Department of EEE
Daffodil International University

Abstract —

The increasing demand for genre-specific creative writing, particularly in children’s storytelling, has revealed significant limitations in the controllability and coherence of large language models (LLMs) during story generation. While models such as LLaMA-2 are capable of generating fluent narratives, they often fail to consistently align them with specific genre prompts, reducing their thematic consistency.

This paper addresses the critical challenge of enhancing genre fidelity in children’s story generation by presenting a genre-conditioned story generation framework that employs a fine-tuned LLaMA-2 7B model, trained on a curated, multi-genre children’s story dataset. To evaluate the genre alignment, we utilize a RoBERTa-based classifier trained for multi-class genre classification across key genres in children’s literature such as Horror, Science Fiction, Humor & Comedy, and Mystery & Detective.

Comparative analysis of stories generated by both the pretrained and fine-tuned models demonstrates that the fine-tuned model significantly improves genre controllability, outperforming the pretrained model. By improving genre fidelity, this work enhances the ability of LLMs to generate more thematically consistent and engaging children’s stories, supporting the development of controlled, genre-aware LLMs for creative writing applications.

Introduction —

The advancement of large language models (LLMs) has revolutionized natural language generation, especially in open-domain storytelling. However, a persistent challenge remains: directing these models to produce outputs that adhere strictly to specific narrative genres. In domains like children's education, interactive media, and digital literature platforms, maintaining structural and thematic genre coherence is not merely stylistic—it's essential for cognitive engagement and pedagogical alignment.

Existing pretrained models such as LLaMA-2, though powerful, exhibit weak genre conditioning due to their generalized training objectives. This often results in narratives that diverge from intended genre-specific expectations, leading to diminished reader immersion and trust in generative systems. To bridge this gap, we introduce a supervised framework that adapts LLMs for genre-conditioned storytelling. The pipeline incorporates genre tags during training and reformats input into instruction-based prompts, aligning with fine-tuning strategies that emphasize narrative intent.

Additionally, to quantify genre adherence objectively, we develop an independent evaluation mechanism using a genre classification model trained on transformer architecture. This dual-model setup not only validates narrative alignment but also facilitates systematic benchmarking of genre fidelity.

The proposed system advances controllable language generation by fine-tuning LLMs into genre-aware storytellers. This enables genre-constrained creativity, personalized education, and consistent AI-driven narratives.

Methodology —

This section outlines a genre-aware story generation pipeline using fine-tuned LLMs, comprising five stages:

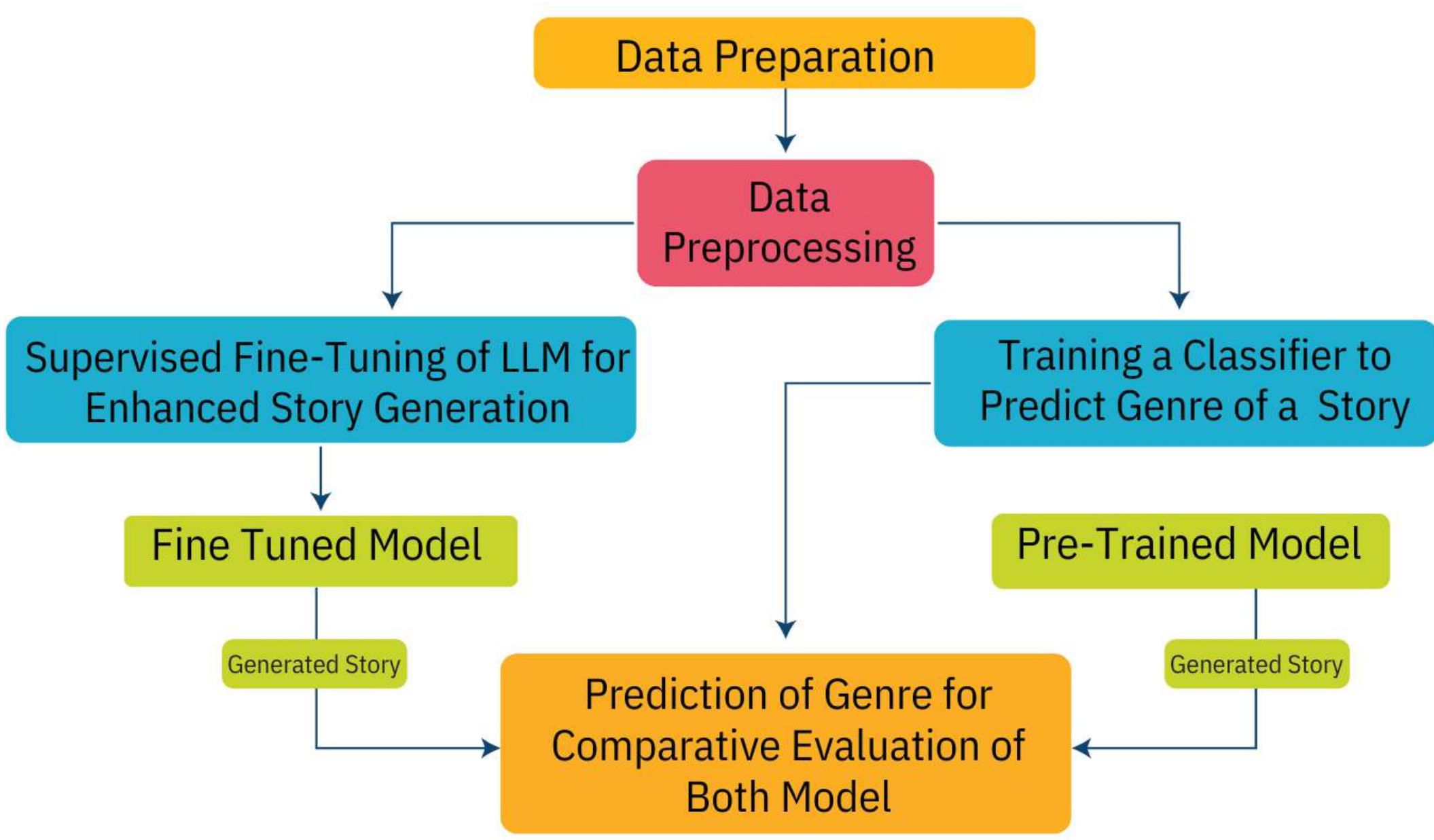


Fig. 1. Overall Workflow for Genre Consistent Story Generation

A. Dataset Preparation

A custom dataset of 200 stories was curated across four genres (Horror, Sci-Fi, Comedy, Mystery) due to the lack of genre-tagged corpora. Each entry includes a prompt, genre label, and a story output.

B. Data Preprocessing

Data was cleaned, shuffled, and truncated to fit token limits. Formatting was standardized and outliers removed to ensure quality inputs.

C. LLM Fine-Tuning

Using PEFT-LoRA, a pretrained LLM was fine-tuned on genre-labeled prompts converted into instruction format. The model was configured to accept genre-aware inputs, optimized with reduced training parameters, and uploaded for reproducibility.

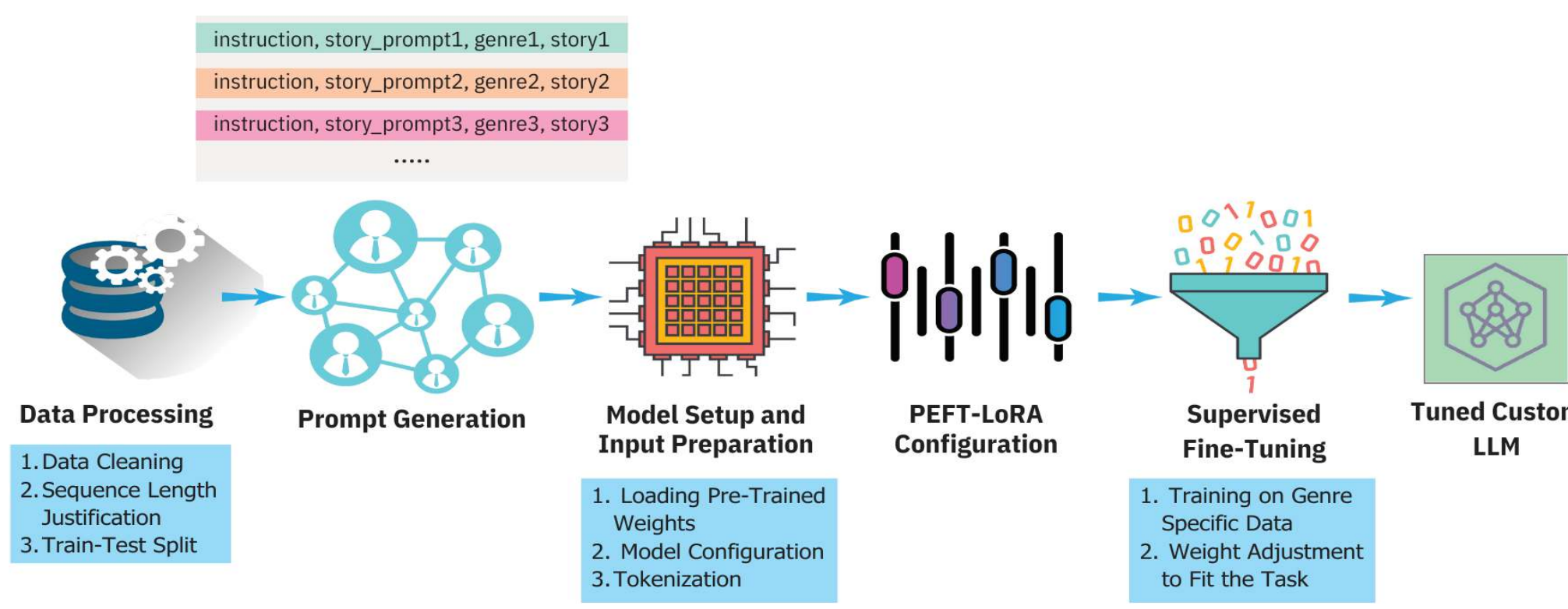


Fig. 2. Fine-Tuning of LLM to Enhance Generation

D. Genre Classifier Training

A genre classifier based on xlm-roberta-large was trained to evaluate genre fidelity. Prompts and generated stories were tokenized, encoded, and classified for training.

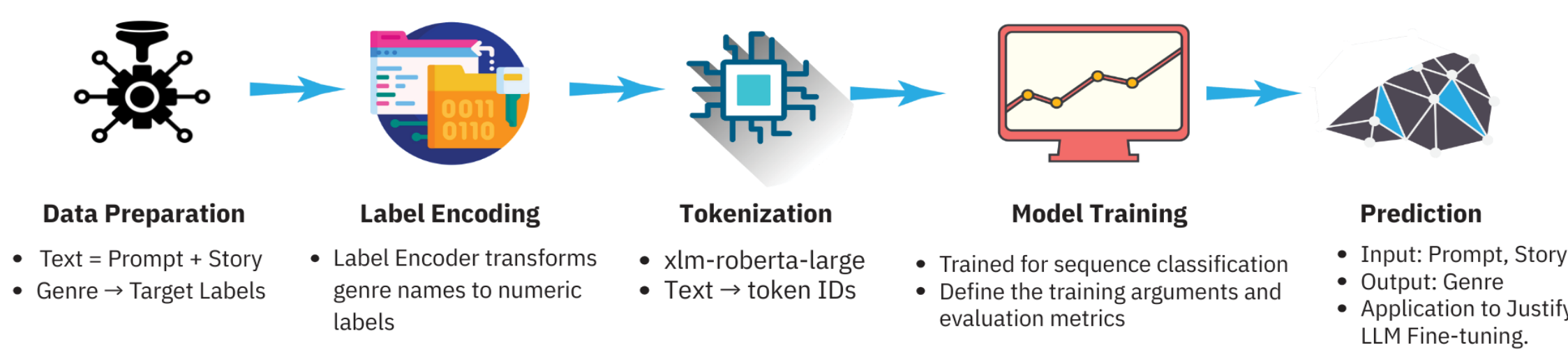


Fig. 3. Training of Genre Classifier Model

E. Model Comparison

The responses of the fine-tuned and pretrained models were evaluated both statistically and visually to assess the improvements achieved through fine-tuning.

Result —

By evaluating the responses of both models using the classifier, two distinct confusion matrices were generated. The derived statistical metrics clearly

demonstrate the superior performance of the fine-tuned model. Moreover, human evaluation of the generated responses further confirms the fine-tuned model’s enhanced relevance, coherence, and overall quality.

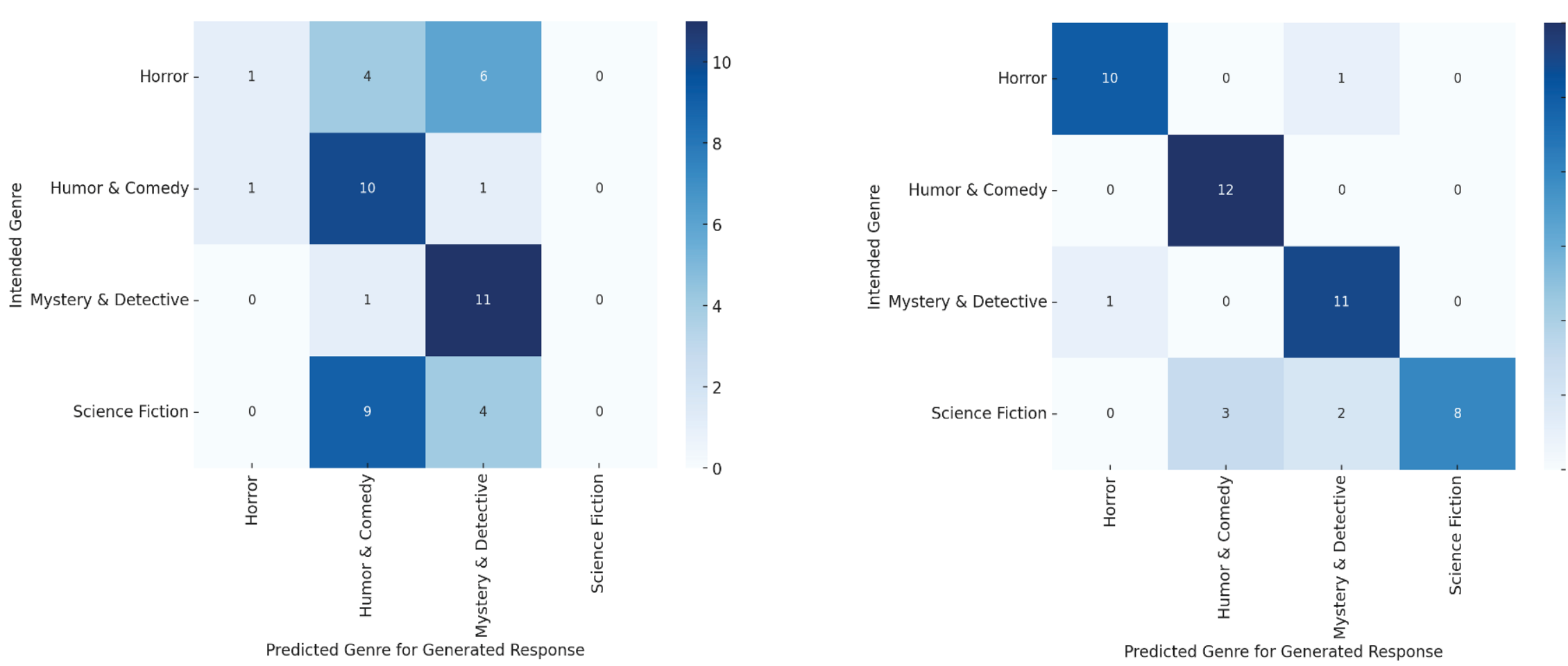


Fig. 4. Confusion Matrix for Pretrained and Fine-tuned Models

Metric	Pretrained LLaMa	Fine-tuned LLaMa
Accuracy	0.4583	0.8542
Precision	0.3542	0.8737
Recall	0.4602	0.8603
F1 Score	0.3391	0.8515

Fig. 5. Statistical Comparison of the Models’ Performance

Story Prompt

Write a story on “A curious boy finds a strange device that takes him to a new world.”

Genre

Science Fiction

Generate with Pretrained LLaMA-2

Generated Story

The boy was walking through the woods when he came across a strange looking rock. He picked it up and noticed that it was warm to the touch. As he was examining it, he heard a voice from inside the rock say, “Please help me, I’m trapped in here.”

Generate with Fine-tuned LLaMA-2

Generated Story

The World Within

One sunny afternoon, Ten-year-old Max was exploring the old attic when he found a small, rectangular device covered in glowing symbols. When he pressed a button, the world around him shifted, and before him appeared a lush forest.

Max stood still, his heart racing with excitement. He had no idea where he was, but the air felt light, as if he could fly. The trees seemed to whisper secrets, and the sun shimmered like a diamond caught in the wind.

Fig. 6. Visual Comparison of Generated Response 1

Story Prompt

A student checks out a book that keeps rewriting itself.

Genre

Mystery & Detective

Generate with Pretrained LLaMA-2

Generated Story

The story is about the missing book. The book is a library. It is the mystery book of the library and it is missing. There are many secretes in the book and the people are searching for the secreates. They are trying to find the secret of life. But they are not able to do so. So they have to wait for a long time.

Generate with Fine-tuned LLaMA-2

Generated Story

The Book That Rewrote Itself

Lena, a curious – and sometimes clumsy – student, was always fascinated by the library's old book of fairy tales. One afternoon, feeling brave, she pulled it off the shelf. As she flipped through the pages, something strange happened. The words began to shift and twist. Page after page rewrote itself until it told a completely new story.

"Once upon a time, there was a girl who could change fate. As she held a magical pen, the pen could write what she dreamed. But one dream was too powerful..."

Lena gasped. She knew this wasn't part of the fairytale. But as she turned the page, the words changed again:

"The girl used her dream to write a new world. She created a world where kindness ruled and wars ended. She wrote the future in her hands. But what if it wasn't a fairy story?"

Lena's mouth watered. "Is this really happening?" she whispered.

Fig. 7. Visual Comparison of Generated Response 2

Reference —

- [1]Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022). Word-craft: Story writing with large language models. Proceedings of the 27th International Conference on Intelligent User Inter-faces, 841-852.
- [2] Kasneci, E., Seßler, K., K¨uchemann, S., Bannert, M., De-mentieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

Implementation —

- Children’s E-book Generators
- Educational Storytelling Platforms
- Language Learning Platforms for Children
- Automated Story Compilation for Publishing