Internship Final Report

Advanced Level

Student Name: Mahasweta Talik

University: Kalinga Institute of Industrial Technology

Major: Computer science

Internship Duration: August 1st, 2025 - August 31st, 2025

Internship Company: ShadowFox

Domain: AI/ML

Mentor: Kalai Maha

Coordinator: Mr. Aakash

Objectives

The primary goals I set for this advanced-level project were:

1. To explore and implement a modern Language Model (LM) using Python.
2. To analyze the LM's performance across various NLP tasks such as text generation and contextual understanding.
3. To evaluate the LM's strengths and limitations through structured experimentation and visualization.

Tasks and Responsibilities

This project involved a deep dive into LM technology, implemented entirely within a Jupyter Notebook. My key responsibilities included:

- **Language Model Selection:**
  Chose **BERT (Bidirectional Encoder Representations from Transformers)** for its robust contextual understanding and wide applicability in NLP tasks.
- **Model Implementation:**
  Set up the Hugging Face Transformers library and implemented BERT using Python. Loaded pretrained weights and tokenizers, and configured the model for inference.
- **Text Input Exploration:**
  Tested BERT on various input scenarios including masked word prediction, sentence classification, and semantic similarity. Documented model behavior and edge cases.
- **Performance Analysis:**
  Evaluated BERT's contextual accuracy, token-level predictions, and adaptability to domain-specific inputs. Compared results with simpler models like n-gram baselines.
- **Research Questions Formulation:**
  Investigated questions such as:
  - How well does BERT capture long-range dependencies in text?
  - Can BERT adapt to informal or domain-specific language?
  - What are its limitations in creative text generation?
- **Visualization of Results:**

  Used Matplotlib and Seaborn to visualize attention weights, token embeddings, and

  prediction confidence. Created comparative plots to highlight BERT's strengths.

- **Ethical and Technical Alignment:**
  Ensured responsible use of pretrained models, acknowledged biases, and aligned the project with current best practices in NLP research.
- **Documentation and Reporting:**
  Maintained detailed records of implementation steps, experimental results, and visualizations. Compiled a comprehensive notebook and final report

Learning Outcomes

This project significantly enhanced my understanding of advanced NLP systems. Key outcomes include:

- **Technical Mastery:**
  Gained hands-on experience with transformer-based architectures, tokenization, and attention mechanisms.
- **Analytical Depth:**
  Learned to design experiments that probe model behavior and interpret results meaningfully.
- **Visualization Skills:**
  Developed techniques to represent abstract model internals in accessible formats.
- **Research Orientation:**
  Strengthened my ability to ask critical questions and explore them through structured analysis.

Challenges and Solutions

1. **Model Complexity:**
   BERT's architecture and tokenization were initially overwhelming. I overcame this by studying official documentation and using simplified examples for debugging.
2. **Resource Constraints:**
   Faced memory limitations during inference. I optimized batch sizes and used CPU-based inference for lightweight testing.
3. **Interpretability:**
   Understanding attention weights was challenging. I used visualization libraries and tutorials to decode attention maps and token importance.

Conclusion

This project was a transformative experience in my journey as an aspiring AI/ML professional. Implementing and analyzing BERT deepened my appreciation for modern NLP techniques and equipped me with the skills to work with large-scale language models. I now feel confident in applying these models to real-world problems and contributing to the evolving field of AI.

Acknowledgments

I extend heartfelt thanks to my academic coordinator Mr. Aakash for his guidance, and to the open-source community and AI tools that supported my exploration. This project has reinforced my commitment to continuous learning and ethical AI development.