

Assignment: Exploring Large Language Models (LLMs) and Their Foundations

Project 1: Exploring Tokenization and Embeddings

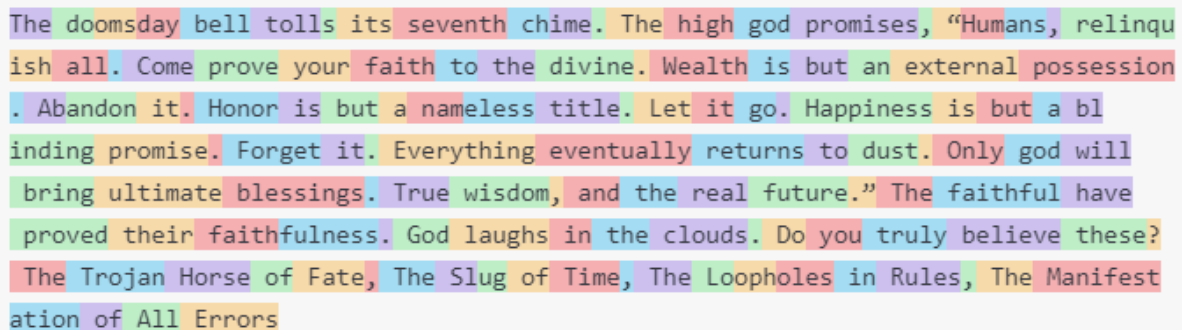
Objective: Understand how text is broken down and represented numerically by LLMs.

Tasks:

1. Choose a short paragraph (100–150 words) from any book, article, or web page.
2. Tokenize the text using any tool (e.g., Python's transformers library or an online tokenizer).
3. Analyze the tokens:
 - How many tokens does your text have?
 - Do any words split into subwords? If so, which ones?
4. Visualize embeddings:
 - Use a pre-trained model (e.g., GPT or BERT) to extract embeddings for a few tokens.
 - Plot the embeddings in a 2D or 3D space using a tool like t-SNE or PCA.

The doomsday bell tolls its seventh chime. The high god promises, “Humans, relinquish all. Come prove your faith to the divine. Wealth is but an external possession. Abandon it. Honor is but a nameless title. Let it go. Happiness is but a blinding promise. Forget it. Everything eventually returns to dust. Only god will bring ultimate blessings. True wisdom, and the real future.” The faithful have proved their faithfulness. God laughs in the clouds. Do you truly believe these? The Trojan Horse of Fate, The Slug of Time, The Loopholes in Rules, The Manifestation of All Errors

| Tokens | Characters |
|--------|------------|
| 133 | 580 |



The doomsday bell tolls its seventh chime. The high god promises, “Humans, relinquish all. Come prove your faith to the divine. Wealth is but an external possession. Abandon it. Honor is but a nameless title. Let it go. Happiness is but a blinding promise. Forget it. Everything eventually returns to dust. Only god will bring ultimate blessings. True wisdom, and the real future.” The faithful have proved their faithfulness. God laughs in the clouds. Do you truly believe these? The Trojan Horse of Fate, The Slug of Time, The Loopholes in Rules, The Manifestation of All Errors

3. 133 tokens, Yes: doomsday, humans, relinquish, abandon, nameless, blinding, faithfulness, slug, loopholes, manifestation

4.

Project 2: Crafting the Perfect Prompt

Objective: Hone your prompt engineering skills by designing effective instructions.

Tasks:

1. Create three different prompts for the same task (e.g., summarizing an article or generating a story).
 - Write one generic prompt.
 - Write one detailed prompt with context.
 - Write one highly specific prompt with clear instructions and boundaries.
2. Test your prompts using an LLM and compare the outputs.
 - Which prompt gave the most accurate or useful result?
 - What changes improved performance?
3. Refine your prompts based on feedback from the model.

Prompts

Summarize this article

Summarize the article titled '*Whatever*.' Focus on the main arguments, benefits, and challenges of remote work, and mention any data or expert opinions included.

Write a 150-word summary of the article '*Whatever*'

Include:

- One sentence introducing the topic
 - Two main benefits and one challenge of remote work
 - A statistic or expert quote from the article
 - A concluding sentence that captures the article's overall takeaway.
- Avoid opinions and keep the tone neutral and professional, try to act like a critic.

The third prompt gave me the most accurate/useful result. By providing clear and concise instructions as well as a role, the LLM was able to really provide what I was looking for.

Revised prompt:

Summarize the article titled '*Whatever*'. Include at least two benefits and one drawback of remote work. Mention one statistic or expert viewpoint, and keep the summary under 200 words. Use a neutral and informative tone

Project 3: Building a Mini Application

Objective: Use LLMs to create a small application that solves a practical problem.

Tasks:

1. Choose an application idea (e.g., a chatbot, content generator, or FAQ assistant).
2. Plan the workflow:
 - Define the input (e.g., user query).
 - Specify the output (e.g., a summary, generated content, or classification).
3. Develop your app using tools like Python, Streamlit, or Google Colab.
 - Include tokenization and embeddings in your pipeline.
 - Fine-tune your prompts for optimal results.
4. Test your application with different inputs to ensure it works smoothly.

1. Study helper for students: an interactive app where students input a topic or question, and the app returns a summary, explanation, and a follow-up quiz. Great for students preparing for certifications, courses, or exams.

2. Input: A text prompt from the user (can you help me understand the significance of this historical period)

Output: a brief explanation (LLM-generated), a summary, 2–3 quiz questions for self-assessment

```
import streamlit as st
from openai import OpenAI
from sentence_transformers import SentenceTransformer
import nltk

embed_model = SentenceTransformer('all-MiniLM-L6-v2')

nltk.download('punkt')

st.title("Smart Study Helper")

query = st.text_input("Enter a topic or question:")
```

```

if query:
    # Tokenization
    tokens = nltk.word_tokenize(query)
    st.write(f"Token count: {len(tokens)}")

    with st.spinner("Generating explanation..."):
        explanation = f"Here's an explanation of '{query}': [LLM-generated content here]"
        summary = f"Summary of '{query}': [LLM-generated summary here]"
        quiz = f"Quiz:\n1. [Question 1]\n2. [Question 2]"

    # Embeddings
    embedding = embed_model.encode(query)

    st.subheader("Explanation")
    st.write(explanation)

    st.subheader("Summary")
    st.write(summary)

    st.subheader("Quiz Questions")
    st.write(quiz)

    st.success("Done!")

```

Prompt 1 (Explanation):

"Explain the topic of [user input] in simple terms, as if teaching a high school student. Use real-world examples if possible."

Prompt 2 (Summary):

"Summarize the explanation of [user input] in two concise sentences."

Prompt 3 (Quiz):

"Write 2 multiple-choice questions to test understanding of [user input], with answers."

Project 4: Advanced Prompt Techniques

Objective: Experiment with advanced strategies to maximize LLM performance.

Tasks:

1. Choose a task that requires nuanced responses (e.g., generating poetry or answering technical questions).
2. Apply advanced prompt techniques:
 - Chain of Thought Prompting: Break down the task into smaller steps.
 - Few-Shot Learning: Include examples in your prompt to guide the model.
 - Role Play: Ask the model to take on a specific persona (e.g., “Act as a historian and explain the significance of the Renaissance.”).
3. Compare the results with standard prompting methods.
4. Document your findings:
 - What advanced techniques did you try?
 - How did they affect the quality of the output?

1. Explain what computer networking is and the different types of it

Output: Computer networking is the practice of connecting computers and other devices to share resources and information. There are several types of computer networks, including LAN (Local Area Network), WAN (Wide Area Network), MAN (Metropolitan Area Network), and PAN (Personal Area Network).

2. Chain of Thought Prompting: Let's explain computer networking step by step:

1. What is computer networking?
2. What are its purposes?
3. What are the main types of computer networks?
4. How are these types used in real-world situations?

Output: Computer networking is the process of connecting computers and other devices to share data, resources, and applications.

1. It allows devices to communicate, enabling file sharing, internet access, email, and remote access.
2. The main types include:
 - **LAN:** Covers a small area, like a home or office.
 - **WAN:** Spans large areas, such as the internet.
 - **MAN:** Covers a city or large campus.
 - **PAN:** Personal network using Bluetooth or USB.

4. For example, your Wi-Fi at home is a LAN. The internet is a WAN. A university might use a MAN to connect campus buildings. A smartwatch connected to your phone uses a PAN.

Few-Shot Learning: *Q: What is cloud computing?*

A: Cloud computing means accessing computing services like servers, storage, or software over the internet rather than on local machines.

Q: What is an IP address?

A: An IP address is a unique identifier for a device on a network, similar to a home address for internet traffic.

Q: Explain what computer networking is and the different types of it.

Output: Computer networking involves linking computers and devices so they can share data and resources. The types of networks include:

- LAN: Local networks in small areas
- WAN: Connects devices across long distances
- MAN: Networks across cities or campuses
- PAN: Personal, short-range connections

Role Play: Act as an experienced IT instructor explaining computer networking to a group of new students. Use real-world analogies and clearly explain each type of network with examples.

Alright class, imagine networking like roads between buildings. Computer networking is how we connect computers so they can "talk" to one another.

- **LAN (Local Area Network)** is like the roads within your neighborhood. It connects a few houses (devices), often using Wi-Fi or Ethernet.
- **WAN (Wide Area Network)** is like the highway system—it links cities (or networks) together over long distances. The internet is the largest WAN!
- **MAN (Metropolitan Area Network)** is like the roads connecting buildings within a city—like connecting government buildings or a university campus.
- **PAN (Personal Area Network)** is your own digital bubble—like when your phone connects to your earbuds via Bluetooth.

Networking makes all our tech devices work together smoothly—whether you're watching Netflix, printing a document, or gaming online.

I think out of all the advanced prompts I used, Role Play gave me the engaging output and the analogies I was looking for, if I was a beginner at learning networks.