



# 20 Large Language Model Project Ideas to Elevate Your Resume



Aqsazafar · [Follow](#)

7 min read · May 8, 2024



9



*20 Large Language Model Project  
Ideas to Elevate Your Resume*

When it comes to making your resume stand out in the competitive job market, showcasing your skills and expertise is crucial. As a budding data

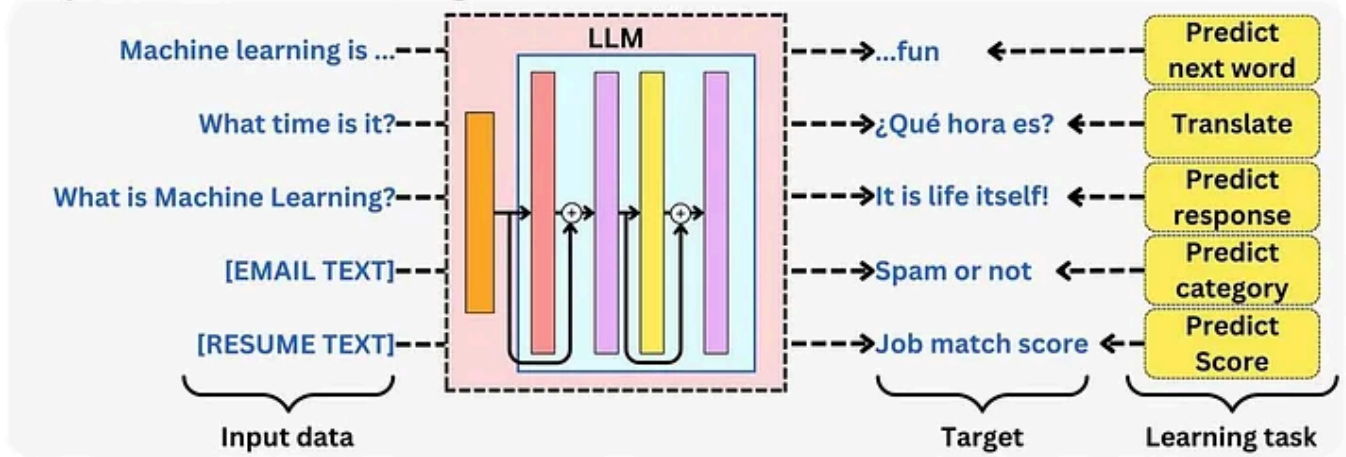
scientist, programmer, or AI enthusiast, one way to grab attention is by including large language model (LLM) projects on your resume.

These projects not only demonstrate your proficiency in natural language processing (NLP) but also highlight your ability to tackle complex problems and deliver innovative solutions. In this guide, we'll explore 20 LLM project ideas that are sure to impress potential employers and set you apart from the crowd.

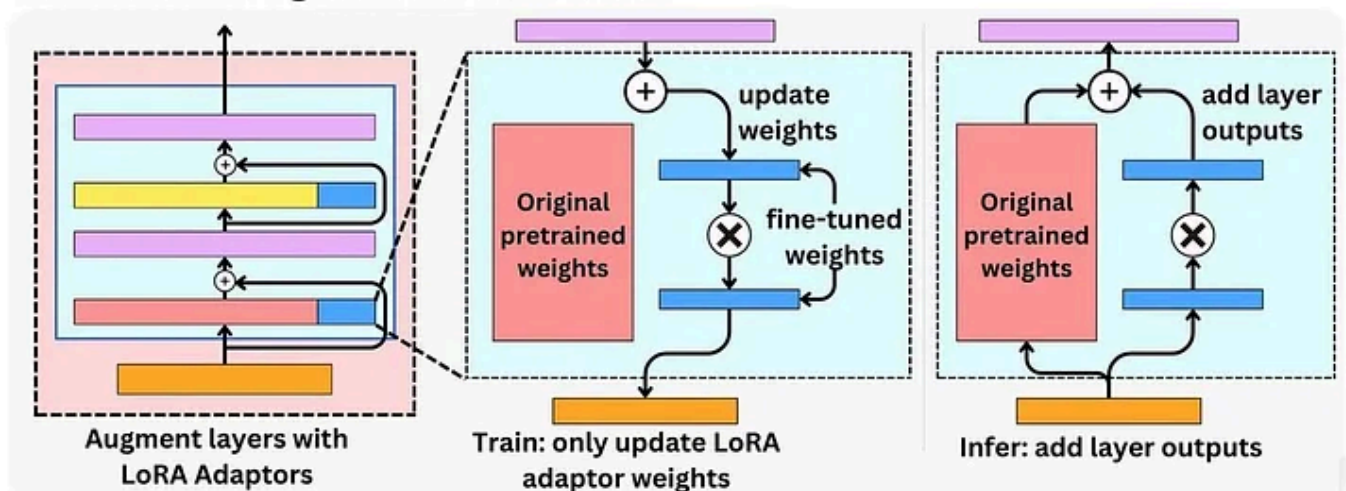
# How to Fine Tune LLMs

TheAiEdge.io

Step 1: Choose the learning task with the related data



Minimize training time and cost: LoRA



Multi fine-tuned tasks pipeline

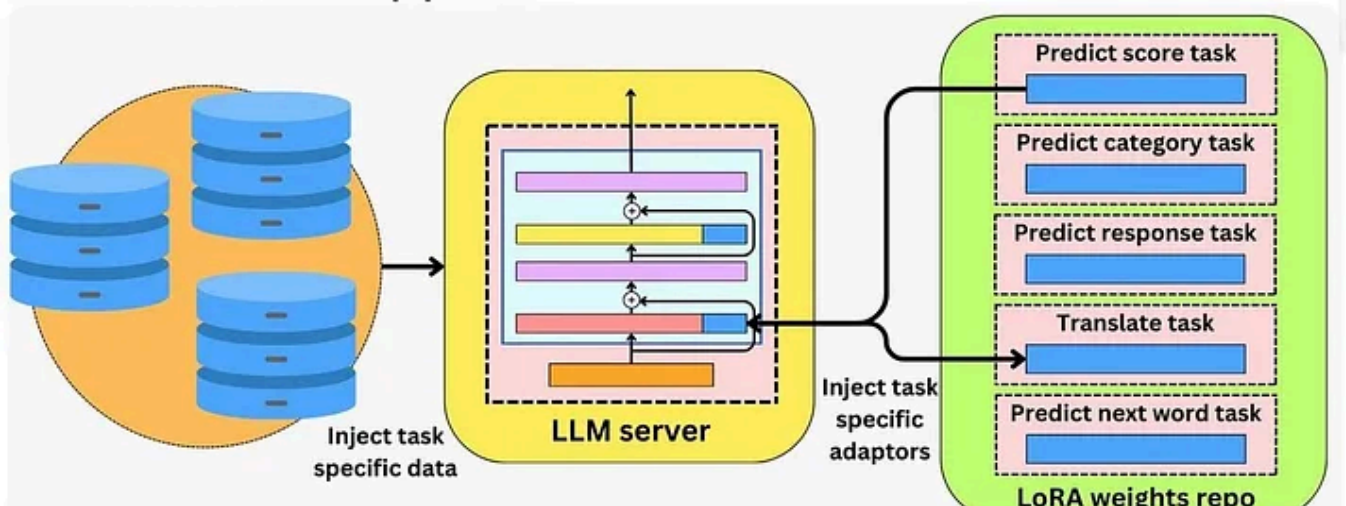


Image Credit- TheAiEdge.io

## 1. Text Summarization Tool

**Description:** Create a tool that can automatically summarize lengthy documents or articles into concise and coherent summaries. Utilize techniques such as extractive summarization, where important sentences are selected from the original text, or abstractive summarization, where a new summary is generated based on the content.

**Dataset:** You can use datasets like CNN/Daily Mail or PubMed for news articles and scientific papers, respectively. These datasets come with paired article-summary pairs, making them suitable for training and evaluation.

## 2. Sentiment Analysis on Social Media Data

**Description:** Develop a model that analyzes the sentiment of social media posts or comments. By identifying positive, negative, or neutral sentiments, you can extract valuable insights from large volumes of user-generated content.

**Dataset:** Utilize datasets like the Sentiment140 dataset, which contains over 1.6 million tweets labeled with sentiment (positive or negative).

Alternatively, you can collect data from social media platforms using their APIs.

## 3. Chatbot for Customer Support

**Description:** Build a chatbot that can interact with customers to address their inquiries and provide assistance. This project showcases your ability to design conversational agents using advanced NLP algorithms.

**Dataset:** You can use datasets like the Cornell Movie Dialogues Corpus or the Ubuntu Dialogue Corpus for training your chatbot. Additionally, you may

need domain-specific data or customer support logs to fine-tune the chatbot for specific tasks.

## **4. Language Translation Service**

**Description:** Create a language translation service that can accurately translate text from one language to another. This project demonstrates your proficiency in multilingual NLP and language modeling techniques.

**Dataset:** Utilize parallel corpora such as the WMT (Workshop on Machine Translation) datasets, which contain aligned sentences in multiple languages. Additionally, you can use publicly available translation datasets for specific language pairs.

## **5. Named Entity Recognition (NER) System**

**Description:** Develop a system that can automatically identify and classify named entities (such as persons, organizations, and locations) in text documents. This project showcases your expertise in information extraction and entity recognition.

**Dataset:** Use datasets like the CoNLL-2003 dataset, which contains labeled entities in English text across various domains. You can also annotate your own dataset using tools like spaCy or NLTK.

## **6. Text Classification for News Articles**

**Description:** Build a model that can classify news articles into different categories (e.g., politics, sports, entertainment). This project demonstrates your ability to perform document classification tasks using supervised learning algorithms.

**Dataset:** Utilize datasets like the Reuters-21578 dataset or the BBC News dataset, which contain news articles labeled with predefined categories. These datasets are widely used for text classification tasks.

## **Best Large Language Models (LLMs) Courses**

*Introduction to Large Language Models– Coursera*

*Generative AI with Large Language Models– Coursera*

*Large Language Models (LLMs) Concepts– DataCamp*

*Prompt Engineering for ChatGPT– Vanderbilt University*

*Introduction to LLMs in Python– DataCamp*

*ChatGPT Teach-Out– University of Michigan*

*Large Language Models for Business– DataCamp*

*Introduction to Large Language Models with Google Cloud– Udacity FREE Course*

*Finetuning Large Language Models– Coursera*

*LangChain with Python Bootcamp– Udemy*

## **7. Email Spam Detection**

**Description:** Create a spam detection system that can accurately classify incoming emails as either spam or legitimate. This project showcases your

skills in text classification and anomaly detection.

**Dataset:** Use datasets like the Enron Email Dataset or the SpamAssassin Public Corpus, which contain labeled emails (spam or ham). You may also need to preprocess the emails to extract relevant features for classification.

## **8. Question Answering System**

**Description:** Develop a system that can answer questions posed in natural language based on a given context or knowledge base. This project demonstrates your understanding of machine comprehension and question-answering techniques.

**Dataset:** Utilize datasets like SQuAD (Stanford Question Answering Dataset) or MS MARCO (Microsoft Machine Reading Comprehension), which contain pairs of questions and answers derived from real-world sources like Wikipedia.

## **9. Text Generation with Recurrent Neural Networks (RNNs)**

**Description:** Train a recurrent neural network (RNN) to generate coherent and contextually relevant text. This project showcases your expertise in sequence modeling and language generation.

**Dataset:** Use text corpora such as the Gutenberg Project or the WikiText datasets, which contain large amounts of text in various genres. You can also use domain-specific datasets for specialized text generation tasks.

## **10. Paraphrase Detection**

**Description:** Build a model that can identify paraphrases (i.e., rephrased versions) of a given text. This project demonstrates your understanding of

semantic similarity and paraphrase identification techniques.

**Dataset:** Utilize datasets like the Quora Question Pairs dataset or the Microsoft Research Paraphrase Corpus, which contain pairs of sentences labeled as paraphrases or non-paraphrases.

## **11. Text-to-Speech (TTS) Synthesis**

**Description:** Develop a text-to-speech synthesis system that can convert written text into spoken speech. This project showcases your skills in speech synthesis and audio processing.

**Dataset:** Use datasets like the LJSpeech dataset or the Mozilla Common Voice dataset, which contain audio recordings paired with corresponding text transcripts. You can also collect your own dataset using text-to-speech software.

## **12. Document Clustering**

**Description:** Create a system that can cluster similar documents together based on their content. This project demonstrates your proficiency in unsupervised learning and document clustering techniques.

**Dataset:** Utilize document collections like the 20 Newsgroups dataset or the Reuters Corpus, which contain documents grouped into predefined categories. You can also use web scraping techniques to collect documents from online sources.

## **13. Language Model Fine-Tuning**

**Description:** Fine-tune a pre-trained language model (such as GPT or BERT) on a specific domain or dataset to improve its performance on specialized



tasks. This project showcases your ability to adapt existing models to new domains.

**Dataset:** Use domain-specific text corpora or datasets relevant to your task for fine-tuning the language model. You can also combine multiple datasets to create a diverse training corpus.

## **14. Text Similarity Measurement**

**Description:** Develop a system that can measure the similarity between pairs of text documents or sentences. This project demonstrates your understanding of semantic similarity metrics and text comparison algorithms.

**Dataset:** Utilize datasets like the STS Benchmark or the SemEval Semantic Textual Similarity datasets, which contain pairs of sentences or documents annotated with similarity scores. You can also create your own dataset by collecting pairs of text data and manually annotating their similarity.

## **15. Multi-Turn Dialogue System**

**Description:** Build a dialogue system that can engage in multi-turn conversations with users on various topics. This project showcases your skills in dialogue management and conversational AI.

**Dataset:** Utilize datasets like the PersonaChat dataset or the DailyDialog dataset, which contain conversational data with multiple turns and diverse topics. You can also collect your own dataset using crowdsourcing or data scraping techniques.

## **16. Document Summarization with Attention Mechanisms**

**Description:** Implement a document summarization model using attention mechanisms to focus on important information. This project demonstrates your understanding of attention-based neural networks and summarization techniques.

**Dataset:** Use datasets like the CNN/Daily Mail dataset or the BigPatent dataset, which contain document-summary pairs suitable for training attention-based summarization models. You can also collect domain-specific datasets for specialized summarization tasks.

## **17. Text Generation with Transformers**

**Description:** Train a transformer-based language model (such as GPT-3) to generate high-quality text outputs. This project showcases your expertise in state-of-the-art language modeling techniques.

**Dataset:** Utilize large-scale text corpora like the BookCorpus or the Common Crawl dataset for pre-training transformer-based language models. You can also fine-tune these models on domain-specific datasets for specialized text generation tasks.

## **18. Named Entity Linking (NEL) System**

**Description:** Develop a system that can link named entities mentioned in text to their corresponding entities in a knowledge base or database. This project showcases your skills in entity linking and knowledge integration.

**Dataset:** Utilize datasets like the AIDA-CoNLL dataset or the WikiLinks dataset, which contain text documents annotated with links to entities in a knowledge base. You can also use publicly available knowledge bases like Wikipedia or DBpedia for entity linking.

## 19. Speech Emotion Recognition

**Description:** Create a model that can recognize the emotional content of spoken speech (e.g., happiness, sadness, anger). This project showcases your expertise in speech signal processing and emotion detection.

**Dataset:** Utilize datasets like the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) or the CREMA-D dataset, which contain audio recordings labeled with emotional categories. You can also collect your own dataset using audio recording devices and emotion annotation techniques.

## 20. Document-level Sentiment Analysis

**Description:** Build a model that can perform sentiment analysis at the document level, considering the overall sentiment expressed in a piece of text. This project demonstrates your ability to analyze and interpret sentiment in longer textual passages.

**Dataset:** Utilize datasets like the IMDb movie reviews dataset or the Yelp reviews dataset, which contain text documents labeled with sentiment polarity (positive or negative). You can also collect domain-specific datasets for specialized sentiment analysis tasks.

By including these diverse and innovative large language model projects on your resume, you not only demonstrate your proficiency in natural language processing and artificial intelligence but also showcase your creativity, problem-solving skills, and ability to deliver real-world solutions.

Whether you're applying for a data scientist position, a machine learning engineer role, or a research position in academia, these projects will

undoubtedly make a strong impression on potential employers and help you advance your career in the exciting field of artificial intelligence and NLP.

## You May Also Be Interested In

[Retrieval Augmented Generation Vs Fine Tuning LLM: Easy Guide](#)

[10 Best Large Language Models Courses and Training \(LLMs\)- 2024](#)

[How to Learn Large Language Models \(LLMs\)? \[Step-by-Step\]](#)

[What is Retrieval Augmented Generation \(RAG\) in AI? Full Guide](#)

Large Language Models

Llms In Healthcare

Retrieval Augmented

Generative Ai Tools

Generative Ai Use Cases



**Written by Aqsazafar**

965 Followers · 8 Following

Follow

Hi, I am Aqsa Zafar, a Ph.D. scholar in Data Mining. My research topic is “Depression Detection from Social Media via Data Mining”.

No responses yet





Write a response

What are your thoughts?

## More from Aqsazafar



Aqsazafar

### 7 Best Laptops for Data Analysis: A Comprehensive Guide

Data analysis has become an indispensable part of many professions, from finance and...

Sep 9, 2023



206



In Analytics Vidhya by Kia Eisinga

### How to create a Python library

Ever wanted to create a Python library, albeit for your team at work or for some open...

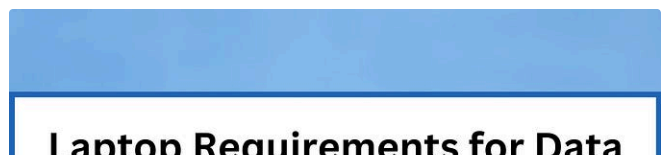
Jan 26, 2020



2.8K



33





In Analytics Vidhya by Harikrishnan N B

## Confusion Matrix, Accuracy, Precision, Recall, F1 Score

Binary Classification Metric

Dec 10, 2019



1.2K



6



Aqsazafar

## 8 Laptop Requirements for Data Scientists [Complete Guide]

In today's data-driven world, data scientists play a crucial role in analyzing information...

Jul 13, 2023

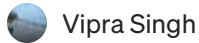


123



See all from Aqsazafar

## Recommended from Medium



Vipra Singh

## LLM Architectures Explained: NLP Fundamentals (Part 1)

Deep Dive into the architecture & building of real-world applications leveraging NLP...

★ Aug 15, 2024 🖱️ 2.4K 💬 21 📌



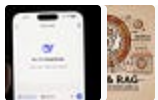
Raghu Teja Manchala

## How I Cracked My IBM AI Engineer Interview: Part 1—Technical

In this article, I share my detailed, step-by-step experience of how I performed and...

★ Jan 13 🖱️ 119 💬 2 📌

### Lists



#### Natural Language Processing

1962 stories · 1606 saves

#### ChatGPT prompts

51 stories · 2605 saves

#### AI Regulation

6 stories · 702 saves

#### Generative AI Recommended Reading

52 stories · 1670 saves

# GenAI with Python: Build Agents from Scratch (Complete Tutorial)

with Ollama, LangChain, LangGraph (No GPU, No APIKEY)

★ Sep 29, 2024 🖱️ 2.4K 💬 38 📌<sup>+</sup>

# Fine Tune Large Language Model (LLM) on a Custom Dataset with...

The field of natural language processing has been revolutionized by large language...

Jan 24, 2024 🖱️ 2.1K 💬 21 📌<sup>+</sup>

☐ In Data Science Collective by Hussein Jundi

## Automate Your Job Search with AI Agents

A step-by-step guide on building a team of AI agents that automate and refine the search...

★ Feb 17 🖱️ 507 💬 19 📌<sup>+</sup>

☐ In Artificial Intelligence in Plain En... by Ritesh Gu...

## Data Science All Algorithm Cheatsheet 2025

Stories, strategies, and secrets to choosing the perfect algorithm.

★ Jan 5 🖱️ 1.3K 💬 28 📌<sup>+</sup>

See more recommendations