**Project Title:** Poetry Style Generative AI: Fine-tuning a Language Model for Interactive Poetry Style Generation

**Project Description:** This project aims to create an interactive web application where users can engage with a generative AI model that emulates a distinct poetic style. The core of the system will involve fine-tuning a pre-trained Large Language Model (LLM) on curated datasets of poetry to capture and replicate the unique stylistic characteristics of chosen poems. Users will be able to provide prompts or topics, and the application will generate new poetry that embodies the selected style, offering a unique method to explore literary stylings with artificial intelligence.

**Final Topic Area:** For this project, we will be working in NLP. Given the intent of the project to generate text in the poetic style of selected poems representing a defined time-period and style, working in NLP is the appropriate application versus Vision.

**Dataset Description:** The selected dataset for this project is the *merve/poetry* dataset sourced from Hugging Face. This dataset is specifically chosen for its focus on diverse poetic styles and time periods, which is crucial for fine-tuning our generative AI model to emulate specific poetic genres.

The dataset consists of 573 poems authored by 67 unique poets. Its total file size is less than 1KB, making it efficient for training and deployment. Each sample is presented as structured text, with clear delineations for different attributes of the poem.

The *merve/poetry* dataset is monolingual (English) and is well-suited for our project due to its structured nature and rich content.

It includes key features for each poem:

* **Content:** The full text of the poem, which will serve as the primary input for the LLM fine-tuning process. This is directly relevant to learning poetic structures, vocabulary, and stylistic nuances.
* **Author:** The name of the poet, enabling us to train and differentiate models based on individual authorial styles.
* **Poem name:** The title of the poem.
* **Age:** The period to which the poem belongs (e.g., Renaissance, Modern), which can be utilized to guide the model towards specific historical or genre-based poetic characteristics.
* **Type:** The subject or genre of the poem (e.g., Love, Nature, Mythology & Folklore), allowing for genre-specific model fine-tuning and content generation.

**Model Selection:** For this project, we will employ a pre-trained GPT-2 model. GPT-2 (Generative Pre-trained Transformer 2) is a Transformer-based decoder-only architecture. Developed by OpenAI, it utilizes a stack of Transformer decoder blocks. This architecture is autoregressive, meaning it predicts the next token in a sequence based on all the preceding tokens, making it highly effective for generative tasks. GPT-2 was notable for its training on a massive and diverse corpus of internet text, allowing it to learn extensive linguistic patterns, grammar, syntax, and semantics.

Leveraging a model already trained on a vast corpus of text significantly reduces the computational resources and time required for training from scratch. The pre-trained model has already acquired a broad understanding of language, which we will then specialize through fine-tuning on our specific poetry dataset. This approach is highly recommended for achieving high-quality generative outputs efficiently within the scope of this project.

GPT-2 is an appropriate choice for our project due to several key reasons directly aligning with our research questions and project goals:

* **Strong Generative Capabilities:** As a generative pre-trained transformer, GPT-2 excels at producing coherent, contextually relevant, and human-like text. This is fundamental to our goal of generating new poetry that embodies specific styles. Its autoregressive nature allows for the sequential generation of text, which is natural for poetic forms.
* **Fine-tuning Potential for Stylistic Emulation:** GPT-2's architecture and pre-training enable effective fine-tuning on smaller, domain-specific datasets. This is crucial for our project, as we aim to fine-tune the model on curated datasets of poetry (like *merve/poetry*) to capture and replicate the unique stylistic characteristics of chosen poets or poetic genres. The model's learned general linguistic knowledge serves as a powerful starting point, which can then be specialized for the intricate patterns of poetic language.
* **Adaptability to Prompts:** GPT-2 can generate conditional samples, meaning it can generate text based on a given prompt or topic. This directly supports the user interaction feature of our web application, where users provide prompts, and the model generates poetry in response.
* **Open-Source Availability and Community Support:** While larger, more recent models exist, GPT-2 offers a balance of strong performance and accessibility. Its open-source nature, coupled with robust support from libraries like Hugging Face Transformers, makes it practical for research and development, allowing for easier implementation, fine-tuning, and deployment within a web application.

**Research Questions:**

1. **How does the fine-tuning process on the *merve/poetry* dataset impact the GPT-2 model's ability to generate coherent, grammatically correct, and contextually relevant poetry compared to its base performance?**
   1. **Specificity:** Examines the impact of fine-tuning on fundamental linguistic qualities (coherence, grammar, relevance) within the context of poetry.
   2. **Relevance to Generative AI:** Explores the effects of transfer learning and domain adaptation on model output quality, a key aspect of LLM application.
   3. **Investigable:** Can be assessed through automated metrics (e.g., perplexity, n-gram overlap) and qualitative review of generated outputs before and after fine-tuning.
2. **To what extent can a fine-tuned GPT-2 model using LoRA effectively emulate the distinctive poetic styles of time-period and topic from the *merve/poetry* dataset, as perceived by human evaluators?**
   1. **Specificity:** Focuses on GPT-2, specific authors from the dataset, and human perception of style emulation.
   2. **Relevance to Generative AI:** Directly addresses the core generative AI concept of style transfer and content generation.
   3. **Investigable:** Can be investigated by training the model on genre-specific subsets of the dataset and conducting qualitative evaluations (e.g., human-in-the-loop assessments, Turing-test-like evaluations of generated poems).
3. **Given the distribution of poets within time-period and topic, how can filtering certain poets based on relative contribution to the dataset, influence the model to be able to mimic different poetic styles?**
   1. **Specificity:** Targets impact of fine tuning based on filtered set of authors to identify stylistic accuracy
   2. **Relevance to Generative AI:** Addresses the impact of overrepresentation of source material for fine tuning (comparable distribution of authors in fine tuning dataset)
   3. **Investigable:** Comparing generative output from a filtered, fine-tuned dataset versus a non-filtered fine-tuned dataset.

**Plan of Action:**

* **Data Preprocessing (1 Week Duration)**
  + The goal of data preprocessing is to transform raw poetic text into a clean, consistent, and model-ready format that facilitates effective style learning. Preprocessing techniques may include extraneous character removal, normalizing formatting, and whitespace removal. Poem text may be formatted to include clear start and end tokens for the language model. Data will be split into training, validation, and test sets.
* **Model Implementation and Interaction Methods (1 Week Duration)**
  + This phase focuses on setting up the LLM, fine-tuning, and preparing for interactive generation. LoRA (Low-Rank Adaptation) will be employed for efficient fine-tuning. This method reduces the number of trainable parameters significantly, making fine-tuning faster and less memory-intensive while maintaining performance.
* **Experimental Design and Evaluation Metrics (1 Week Duration)**
  + This phase focuses on designing experiments to test the model's performance and defining concrete metrics for evaluation. Separate LoRA adapters will be trained for each selected poet persona/genre. Hyperparameters (learning rate, batch size, number of epochs, LoRA rank) will be optimized using a validation set.
  + Evaluation metrics will consist of qualitative human evaluation using a Likert Scale rating criteria where a small group of independent evaluators rate the stylistic fidelity, coherence, creativity, fluency, and authenticity of the generated poems compared to the baseline LLM and some authentic poems.
* **Analysis Techniques (1 Week Duration)**
  + This phase involves interpreting the results and drawing conclusions. A qualitative analysis of the human evaluations will be performed where the human rating scores are aggregated and analyzed to identify the strengths and weaknesses in the generated poems.
  + Additionally, quantitative metric interpretations can be performed analyzing style classification, stylistic features, and embedding similarities between generated poems and authentic poems.

**Team Contribution:** The responsibilities of each team member will be determined by individual strengths and familiarity with the data. However, each team member will jointly participate and collaborate on activities related to the **Plan of Action** to provide valuable insights into alternative methods for model development, training, and evaluation.

Defined roles and descriptions include:

* **Project Lead and Data Architect:** This role is responsible for the overall strategic direction, project management, and ensuring the technical architecture of the data pipeline and model is sound.
* **Data Scientist and Modeler:** This role will focus on the core machine learning aspects, including model selection, training, and evaluation.
* **Data Engineer and Preprocessing Specialist:** This role is crucial for acquiring, cleaning, and preparing the large historical dataset, ensuring data quality and availability for the modeling phase.
* **Research and Documentation Specialist:** This role will focus on the theoretical underpinnings, literature review, and comprehensive documentation of the project.

**References:**

1. Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, *15*(8), 260.
2. Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., ... & Azam, S. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE access*, *12*, 26839-26874.
3. Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2023, June). Prompt engineering in large language models. In *International conference on data intelligence and cognitive informatics* (pp. 387-402). Singapore: Springer Nature Singapore.
4. Alhafni, B., Kulkarni, V., Kumar, D., & Raheja, V. (2024). Personalized text generation with fine-grained linguistic control. *arXiv preprint arXiv:2402.04914*.
5. Ajwani, R. D., Zhu, Z., Rose, J., & Rudzicz, F. (2024). Plug and Play with Prompts: A Prompt Tuning Approach for Controlling Text Generation. *arXiv preprint arXiv:2404.05143*.