Subject: Internship Proposal

ID	PTI_Distefano Salvatore_19/04/2025 10.16.24
Data	19/04/2025 10.16.24

Project Supervisor

Surname	Distefano
Name	Salvatore
Department	01/04/2025
Laboratory	HPCA
E-mail	sdistefano@unime.it
Phone number	

Project Co-Supervisor

Surname	
Name	
Job Position	
Department	

Laboratory	
E-mail	
Phone number	

Project details

Title	Prompt Engineering

Detailed description: Context:

The field of Artificial Intelligence has witnessed a paradigm shift with the emergence of Large Language Models (LLMs). These models demonstrate remarkable capabilities in understanding and generating human-like text, opening up unprecedented opportunities across various domains, including scientific research, data analysis, and software development. However, effectively harnessing the power of LLMs hinges on the ability to craft precise and nuanced instructions, known as "prompts." Prompt engineering has thus become a critical discipline, focusing on the art and science of designing prompts that elicit desired, accurate, and reliable outputs from these powerful models. This internship offers an exciting opportunity to delve into the core principles and advanced techniques of prompt engineering within the vibrant academic environment of the University of Messina.

Issues and Challenges:

While LLMs possess impressive general capabilities, realizing their full potential in specific, complex tasks presents several challenges that prompt engineering seeks to address:

Eliciting Specific and Accurate Information: Ensuring that LLMs provide factual, relevant, and detailed responses tailored to the user's needs, especially in technical or scientific contexts.
Controlling Output Format and Style: Guiding LLMs to generate outputs in specific formats (e.g., code snippets, structured data, reports) and adhere to desired writing styles.
Mitigating Bias and Hallucinations: Reducing the generation of biased or factually

incorrect information, which is crucial for trustworthy applications.
Optimizing for Efficiency and Cost: Understanding how prompt design impacts the computational resources (e.g., tokens, processing time) required by LLMs.
Enabling Complex Reasoning and Problem-Solving: Designing prompts that encourage LLMs to perform multi-step reasoning, planning, and problem-solving.
Adapting to Different Model Architectures: Recognizing that optimal prompting strategies can vary across different LLM architectures and sizes.

Approaches and Methodologies:

This internship will explore a range of contemporary prompt engineering approaches and methodologies, including:

Instruction-Based Prompting: Crafting clear, concise, and unambiguous instructions that directly guide the LLM.
Few-Shot Learning: Providing the LLM with a limited number of input-output examples within the prompt to enable it to learn the desired task.
Chain-of-Thought Prompting: Guiding the LLM through intermediate reasoning steps to arrive at a final answer for complex problems.
Retrieval-Augmented Generation (RAG): Integrating external knowledge sources into the prompting process to enhance the accuracy and contextuality of LLM outputs.
Prompt Iteration and Refinement: Systematically experimenting with different prompt variations and analyzing their impact on the quality of the generated text.
Meta-Prompting: Using prompts to guide the LLM in generating effective prompts for other tasks.
Adversarial Prompting: Investigating the vulnerabilities of LLMs to malicious prompts and developing strategies for robust prompt design.
The student will gain practical experience in applying these methodologies to various tasks and evaluating their effectiveness.

Measurable Targets and Metrics:

During the internship, the student will be expected to contribute to the following measurable targets:

[Target 1 (Task Performance): Achieve a specific level of performance (e.g., accuracy, F1-score) on a defined task using optimized prompts compared to baseline prompts.]
Metric: Task-specific evaluation metrics (e.g., accuracy for classification, BLEU score for

text generation).
[Target 2 (Efficiency Gain): Reduce the average token usage or inference time for a specific task by a certain percentage through prompt optimization.] Metric: Average tokens per response, inference time.
[Target 3 (Robustness Evaluation): Demonstrate a certain level of resilience in LLM output quality when subjected to variations in prompt phrasing or input.] Metric: Consistency of performance across different prompt variations.
[Target 4 (Knowledge Integration Effectiveness): Quantify the improvement in the relevance and accuracy of LLM responses when using RAG compared to prompting without external knowledge.] Metric: Relevance scores (e.g., using embedding similarity), accuracy against a ground truth.
Progress towards these targets will be assessed through systematic experimentation, data analysis, and regular discussions with the supervisor.

Tools:

The student will have access to and gain experience with the following tools:

Large Language Model APIs and Platforms: Interaction with various LLM APIs (e.g., OpenAI, Google AI), and potentially open-source LLMs via platforms like Hugging Face Transformers.
Prompt Engineering Frameworks: Libraries and tools designed to facilitate prompt creation, management, and experimentation (e.g., LangChain, PromptFlow).
Evaluation Metrics and Libraries: Tools for evaluating the quality and performance of LLM-generated text (e.g., BLEU, ROUGE, accuracy metrics from scikit-learn).
Data Manipulation and Analysis Libraries: Python libraries such as Pandas and NumPy for working with datasets.
Version Control: Git for collaborative development and tracking changes.

Datasets:

The internship may involve working with a variety of datasets depending on the specific focus of the research, including:

Publicly Available Text Datasets: Datasets used for training and evaluating language models (e.g., Common Crawl, Wikipedia).
Task-Specific Datasets: Datasets designed for evaluating performance on specific tasks like question answering, text summarization, or code generation.

Curated Datasets: Potentially creating or utilizing smaller, focused datasets to evaluate specific aspects of prompt engineering.
The emphasis will be on understanding the characteristics of these datasets and how they can be used to evaluate the effectiveness of different prompting strategies.

Application Domains:

The principles and techniques learned during this internship are applicable across a wide range of domains, including:

Content Creation: Generating articles, blog posts, marketing copy, and creative writing.
Education: Developing personalized learning experiences and automated feedback systems.
Customer Service: Building intelligent chatbots and virtual assistants.
Software Development: Assisting with code generation, documentation, and debugging.
Scientific Research: Automating literature reviews, generating hypotheses, and assisting with data analysis.
Information Retrieval: Enhancing search engines and knowledge management systems.

Case Studies:

The student will be introduced to relevant case studies that highlight the impact of effective prompt engineering:

Case Study 1: Improving Question Answering Accuracy:

Study: "Generated Knowledge Prompting for Commonsense Reasoning" by Yasufumi Zhang, Kazuki Irie, and Antoine Bosselut (2022).
Relevance: This paper explores how generating relevant knowledge snippets and incorporating them into prompts can significantly improve the performance of LLMs on commonsense question-answering tasks. While not strictly focused on scientific QA, the principles of knowledge-augmented prompting are highly relevant.
Verified Link (ACL Anthology): https://aclanthology.org/2022.emnlp-main.557/

Case Study 2: Using Chain-of-Thought for Complex Reasoning:

Study: "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models" by Jason Wei et al. (2022).

Relevance: This foundational paper introduces the chain-of-thought prompting technique, demonstrating its effectiveness in enabling LLMs to perform multi-step reasoning for various tasks, including arithmetic, commonsense, and symbolic reasoning.
Verified Link (arXiv): https://arxiv.org/abs/2201.11903 (Note: arXiv is a pre-print server but this paper has been highly influential and widely cited).

Case Study 3: Enhancing Code Generation with Prompt Engineering:

Study: "Do Prompt Programming Languages Need Types?" (2023).
Relevance: This paper investigates the impact of prompt design, specifically the inclusion of type information, on the quality of code generated by large language models. It highlights how carefully crafted prompts can lead to more accurate and reliable code.
Verified Link (arXiv): https://arxiv.org/abs/2305.19824 (Note: arXiv is a pre-print server).

Relevant Literature:

The student will be encouraged to explore the following key literature to build a strong foundation in prompt engineering:

[Reference 1 (Introduction to Prompting): Brown, T. B., et al. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901. (DOI: 10.48550/arXiv.2005.14165)]
Relevance: This seminal paper demonstrates the power of in-context learning and highlights the importance of prompt design for few-shot performance.

[Reference 2 (Chain-of-Thought Prompting): Wei, J., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903. (DOI: 10.48550/arXiv.2201.11903)]
Relevance: Introduces a crucial technique for enabling complex reasoning in LLMs by guiding them through intermediate steps.

[Reference 3 (Retrieval-Augmented Generation): Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 1771-1781. (URL: https://proceedings.neurips.cc/paper/2020/file/6b49323545041c114992a92e1e393338-Paper.pdf)]
Relevance: Explores a key method for grounding LLM responses in external knowledge, improving accuracy and reducing hallucinations.

[Reference 4 (A Survey of Prompting Methods): Liu, P., et al. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. ACM Computing Surveys, 55(10), 1-35. (DOI: 10.1145/3560831)]
Relevance: Provides a comprehensive overview of various prompting techniques and their applications in natural language processing.

Further Investigation:

During the internship, the student will be encouraged to delve deeper into these and other relevant publications to:

Gain a comprehensive understanding of the theoretical foundations of prompt engineering.
Explore the empirical evidence supporting the effectiveness of different prompting techniques.
Investigate the limitations and potential biases associated with various prompting strategies.
Identify novel and emerging trends in prompt engineering research.
Apply prompt engineering principles to address specific challenges in intelligent applications.

Duration (month – max 12)	12
Duration (hours)	undefined
Open positions	8

Internship Skills

Technical requirements:

Other skills	