Subject: Development of an Autonomous AI Agent with Uncertainty Assessment and Dynamic Knowledge Graph Retrieval

Project Duration: 6 months

Research Overview:

This research aims to address the limitations of Large Language Models (LLMs) in managing uncertainty and mitigating hallucinations by developing an autonomous AI agent. The agent will assess its uncertainty during text generation and dynamically retrieve information from a knowledge graph to enhance the accuracy and coherence of its responses. This approach is crucial for improving the reliability of AI systems in various applications where precise and trustworthy information is essential.

Research Objectives:

1. **Develop a Mechanism for Uncertainty Assessment**: Create a module that evaluates the confidence of the AI agent in its generated text, using techniques such as confidence scoring and attention mechanisms.
2. **Implement Dynamic Knowledge Graph Retrieval**: Design a system that efficiently queries a knowledge graph based on the assessed uncertainty, ensuring the retrieval of relevant and accurate information.
3. **Integrate into LLM for Enhanced Text Generation**: Seamlessly incorporate the retrieved information back into the text generation process of the LLM, improving the quality and coherence of the output.
4. **Evaluate System Performance**: Conduct comprehensive evaluations to measure the effectiveness of the system in terms of accuracy, coherence, and reduction in uncertainty, comparing it against existing models.

Expected Outcomes:

- A functional AI agent prototype that demonstrates improved text generation with reduced uncertainty.
- Empirical evidence showing enhanced performance metrics, such as accuracy and coherence, in comparison to traditional LLMs.
- Insights into the effectiveness of uncertainty-driven information retrieval in LLMs, contributing to the theoretical understanding of AI agent capabilities.

References :

[1] Su, W., Tang, Y., Ai, Q., Wu, Z., & Liu, Y. (2024). DRAGIN: Dynamic Retrieval Augmented Generation based on the Information Needs of Large Language Models. *Journal of Artificial Intelligence*, 10(1), 1-20.

[2] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 1877-1901.

[3] Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 6268-6281.

4. Izacard, G., & Grave, E. (2020). Leveraging passage retrieval with generative models for open domain question answering. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6532-6543.