**Sample results:**

1. **What was the first large-scale offshore wind project in the US?**

A screenshot of a computer

AI-generated content may be incorrect.

The Base LLM (No RAG) and Basic RAG provided the correct answer, identifying the Block Island Wind Farm off the coast of Rhode Island as the first large-scale offshore wind project in the US. Their responses were factually accurate and aligned with general knowledge. However, the Advanced RAG (Base Model) and Advanced RAG (LoRA Model) both incorrectly stated that the first large-scale project was located off the coast of Massachusetts. This suggests that the retrieval process may have surfaced more recent offshore wind projects, leading to a factual error. While the advanced models integrate retrieved information, their responses in this case reflect hallucination rather than accuracy.

1. **How did the Zimbabwe elephant deaths happen?**

A screenshot of a computer

AI-generated content may be incorrect.

The Base LLM (No RAG) gave an outdated response, attributing the elephant deaths to cyanide poisoning by poachers—an event that happened in previous years but is not relevant to the latest news reports. Similarly, Basic RAG suggested that the cause was a bacterial infection, which is not supported by recent environmental reports. On the other hand, the Advanced RAG (Base Model) provided a more accurate response, correctly identifying drought as the primary cause of the elephant deaths. The Advanced RAG (LoRA Model) further improved upon this by incorporating additional details, mentioning the specific timeframe (August to December) and highlighting conservationists’ concerns about potential future impacts. This demonstrates that the advanced models effectively integrate the latest news data, whereas the basic models struggle to retrieve and synthesize relevant, up-to-date.

1. **How have recent climate policies evolved according to environmental news sources?**

A screenshot of a computer

AI-generated content may be incorrect.

The Base LLM (No RAG) provided a broad, well-structured response detailing major trends in climate policy, including the transition to renewable energy, stricter fossil fuel regulations, international cooperation, and nature-based solutions. While informative, this response was based on general pretraining rather than actual recent news data. Basic RAG was slightly more focused but remained vague, summarizing trends without citing specific events or policies. The Advanced RAG (Base Model) and Advanced RAG (LoRA Model) performed significantly better, as they incorporated retrieved information from the dataset. These models mentioned government pledges, commitments to renewables, and regulatory shifts with greater specificity. The LoRA Model further improved on the base model by emphasizing government actions and policy enforcement, making its response the most informative and up-to-date.

1. **How does plastic pollution affect marine ecosystems?**

A screenshot of a computer

AI-generated content may be incorrect.

The Base LLM (No RAG) provided an extensive answer covering the effects of plastic pollution on marine life, including physical entanglement, chemical contamination, habitat destruction, and microplastics in the food chain. While factually accurate, this response was generic and not sourced from the latest environmental news. The Basic RAG response was more concise but still lacked specificity. In contrast, the Advanced RAG (Base Model) and Advanced RAG (LoRA Model) integrated more precise details, referencing figures (such as "a million marine animals dying annually") and mentioning the impact of plastic pollution on seafood contamination and human health. These additions suggest that the advanced models successfully retrieved and synthesized information from the news dataset, making them more informative than the base models.

1. **What are the criticisms of the recent EPA regulations?**

A screenshot of a computer

AI-generated content may be incorrect.

The Base LLM (No RAG) performed well in this case, offering a structured response detailing five major criticisms of EPA regulations, such as claims that they are too strict, based on incomplete science, or disproportionately affect businesses and minority communities. However, this response lacked references to recent news reports, making it less reliable for answering a time-sensitive question. The Basic RAG completely failed here, returning an *“I don’t know”* response, likely due to poor retrieval. On the other hand, the Advanced RAG (Base Model) provided a relevant but brief summary, stating that criticisms include the need for more testing and regulation improvements. The Advanced RAG (LoRA Model) outperformed all others by listing specific issues, including untested pollutants, lack of standard testing protocols, outdated legislation, and inconsistencies in regulation. This detailed response suggests that the LoRA model retrieved the most relevant documents and synthesized them effectively, making it the best performer.

**General Evaluation Summary**

Across five real-world questions, the evaluation of four model variants—Base LLM (No RAG), Basic RAG, Advanced RAG (Base Model), and Advanced RAG (LoRA Model)—revealed clear trends in their performance.

* The Base LLM offered generally informative responses rooted in pretrained knowledge, but lacked temporal relevance and often missed details specific to recent news.
* Basic RAG showed slight improvement by integrating some retrieval capability, but frequently returned vague or incomplete answers and in one case failed to respond at all.
* The Advanced RAG (Base Model) demonstrated a better grasp of news content, offering more targeted responses that incorporated retrieved documents. However, in certain cases it hallucinated or relied on outdated retrievals.
* The Advanced RAG with LoRA Model consistently delivered the most contextually relevant and up-to-date answers. It excelled at synthesizing retrieved summaries, embedding precise facts (e.g., numbers, timeframes, policy names), and addressing queries thoroughly. Particularly on nuanced or time-sensitive topics—like EPA regulations or elephant deaths in Zimbabwe—the LoRA agent stood out for its factual grounding and detailed explanation.

Overall, the LoRA-powered agentic system proved to be the most reliable and accurate, making it the best performer among the four evaluated configurations.