**Title: How Different AI Model Architectures Influence Responses to same Benchmark Questions**

**Abstract:**

Artificial Intelligence (AI) models are evaluated using standardized benchmark datasets to assess reasoning, factual recall, and problem-solving capabilities. However, different AI architectures—such as transformer-based and mixture-of-experts models—yield varying accuracy rates on the same benchmark questions. This research explores how different AI models respond to identical evaluation prompts and investigates the architectural factors that contribute to these performance variations.

Using API access to multiple AI models, we systematically collect and compare their responses to benchmark dataset inputs. Additionally, we analyze research papers detailing each model’s architecture to identify key differences in size, training techniques, and retrieval mechanisms. Factors such as parameter count, fine-tuning strategies, and reinforcement learning techniques are examined to assess their impact on factual recall and reasoning.

This study bridges the gap between AI performance metrics and practical deployment, offering insights into how model architecture influences reasoning and decision-making. Understanding these differences can guide future improvements in AI model development and benchmark design.