

PHAROS Training Series - Course 9 "RAG End-to-End: Architecture, Retrieval, Generation and Evaluation"

Contribution ID: 10

Type: **not specified**

RAG evaluation: retrieval quality, faithfulness, groundedness and RAGAS/DeepEval concepts

Tuesday, 7 July 2026 14:40 (20 minutes)

This evaluation session explains how to assess RAG systems beyond a single final answer score. Participants will learn why evaluation must decompose the pipeline into retrieval, generation, citation quality and end-to-end behaviour. The session introduces retrieval metrics such as Precision@k, Recall@k, Mean Reciprocal Rank and NDCG, showing how they describe the evidence made available to the model. It then discusses answer-level criteria, including correctness, faithfulness, groundedness, citation quality, completeness and relevance. Participants will also be introduced to RAGAS and DeepEval concepts for automated RAG evaluation, regression testing and structured comparison of system variants. The emphasis is diagnostic: metrics should identify failure modes and guide concrete improvements, such as better chunking, query handling, filtering, reranking or prompt constraints. By the end, participants will understand how to evaluate and iterate RAG systems systematically. This prepares them to maintain RAG quality as data, prompts and models evolve in operational settings, after deployment too.

Presenters: DROSATOS, George (ATHENA RC); GYFTOPOULOS, Sotiris (ATHENA RC)