

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

Contribution ID: 11

Type: **not specified**

## Deployment considerations and wrap-up

*Tuesday, 7 July 2026 15:00 (15 minutes)*

This wrap-up session discusses how a Colab-based RAG prototype can be prepared for real-world deployment. Participants will review the main components of a deployed RAG application, including user interface, API or orchestration chain, retriever service, vector or hybrid store, LLM service, monitoring and logging. The session highlights production concerns such as latency, cost, data privacy, access control, source updates, model versioning, failure logging and regression testing. It also revisits the specific challenges of Greek-language public-service RAG, including morphology, administrative terminology, tables, articles, codes, appendices and vague user questions. Participants will receive an end-to-end design checklist covering information need, risk level, document preparation, chunking, metadata, retrieval strategy, grounding, citations, evaluation and deployment safeguards. The session closes the webinar by connecting the theoretical concepts and hands-on notebook into a coherent pathway for operational RAG applications. It also identifies next steps for adapting the tutorial to participants' own organisational domains and use cases.

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