

# Příprava a implementace RAG řešení do produkce v jazyce Python

Kód kurzu: PYTHON\_RAG

Účastníci se naučí navrhnout, implementovat, vyhodnotit a nasadit produkční řešení RAG (Retrieval-Augmented Generation) včetně: ingestion pipeline, embeddingů a vector DB, retrieval strategií, orchestrace LLM, evaluace, observability, security, deploymentu a škálování. Kurz je zaměřen prakticky - každý den obsahuje hands-on workshopy a průběžné budování jednoho produkčního projektu.

## Pro koho je kurz určen

- AI/ML Engineers
- Backend developéři
- Data Engineers
- Solution Architects
- Tech Leads
- Innovation teams

## Co Vás naučíme

Po absolvování bude mít účastník:

- produkční RAG projekt
- deployment pipeline
- eval framework
- monitoring dashboard
- architektonické best practices

## Osnova kurzu

Den 1 — Architektura RAG a příprava dat

Téma dne

Pochopení architektury moderního systému RAG a návrh ingestion pipeline.

Learning outcomes

Účastník bude umět:

- vysvětlit komponenty architektury RAG
- identifikovat vhodné use cases
- připravit dokumentové zdroje
- navrhnout chunking strategii
- implementovat základní ingestion pipeline

Agenda

1. Úvod do GenAI a RAG

- Limity čistého LLM
- Halucinace
- Kontextové okno
- Proč RAG funguje
- Typy systémů RAG:

### GOPAS Praha

Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

### GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

### GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved

# Příprava a implementace RAG řešení do produkce v jazyce Python

- o Naive RAG
- o Hybrid RAG
- o Agentic RAG
- o Graph RAG

## 2. Architektura produkčního systému RAG

- Data ingestion
- Parsing
- Chunking
- Embeddings
- Vector database
- Retriever
- Reranker
- Prompt orchestration
- LLM serving
- Monitoring

## 3. Příprava dat

- PDF parsing
- HTML a Confluence ingestion
- OCR pipeline
- Metadata enrichment
- Deduplication
- Verzování dokumentů

## 4. Chunking strategie

- Fixed-size chunking
- Semantic chunking
- Recursive chunking
- Parent-child documents

## 5. Praktický workshop

Implementace:

- document loader
- chunking pipeline
- embedding generation

Technologie:

- Python
- LangChain / LlamaIndex
- OpenAI embeddings / BGE
- Unstructured.io

**GOPAS Praha**  
Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Brno**  
Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Bratislava**  
Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved

# Příprava a implementace RAG řešení do produkce v jazyce Python

Praktické zadání

Připravit ingestion pipeline pro interní knowledge base.

Den 2 — Embeddings, retrieval a vector databáze

Téma dne

Jak funguje retrieval vrstva a jak optimalizovat kvalitu vyhledávání.

Learning outcomes

Účastník bude umět:

- vybrat embedding model
- pracovat s vector DB
- implementovat hybrid search
- optimalizovat kvalitu retrieval

Agenda

1. Embeddings deep dive

- Dense embeddings
- Sparse embeddings
- Multilingual embeddings
- Domain adaptation

2. Vector databáze

Porovnání:

- Pinecone
- Weaviate
- Qdrant
- Milvus
- PostgreSQL + pgvector

3. Retrieval strategie

- Similarity search
- MMR retrieval
- BM25
- Hybrid retrieval
- Multi-query retrieval
- Self-query retrieval

4. Reranking

- Cross-encoders
- Cohere rerank
- BGE reranker

## GOPAS Praha

Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

## GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

## GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved

# Příprava a implementace RAG řešení do produkce v jazyce Python

- Context compression

## 5. Praktický workshop

Implementace:

- vector DB
- semantic search
- hybrid retrieval
- reranking pipeline

Praktické zadání

Optimalizovat retrieval precision pro firemní dokumentaci.

## Den 3 — Prompt orchestration a generování odpovědí

Téma dne

Jak správně skládat pipeline retrieval + prompt + generation.

Learning outcomes

Účastník bude umět:

- navrhnout robustní prompt pipeline
- implementovat citation-aware odpovědi
- minimalizovat hallucinations
- implementovat guardrails

Agenda

### 1. Prompt engineering pro RAG

- System prompts
- Context injection
- Few-shot prompting
- Citation prompting

### 2. Kontextové strategie

- Sliding window
- Context compression
- Retrieval filtering
- Conversational memory

### 3. Guardrails a bezpečnost

- Prompt injection
- Jailbreak protection
- Sensitive data filtering
- PII masking

#### GOPAS Praha

Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved

# Příprava a implementace RAG řešení do produkce v jazyce Python

## 4. Agentic workflows

- Tool calling
- Multi-step retrieval
- Query planning
- Reflection loops

## 5. Praktický workshop

### Implementace:

- conversational RAG chatbot
- citations
- anti-hallucination mechanismů
- guardrails

### Technologie:

- LangGraph
- Guardrails AI
- OpenAI / Anthropic APIs

### Praktické zadání

Vytvořit enterprise chatbot s citacemi zdrojů.

## Den 4 — Evaluate, observability a kvalita

### Téma dne

Jak měřit kvalitu systému RAG a monitorovat produkci.

### Learning outcomes

#### Účastník bude umět:

- evaluovat retrieval
- měřit hallucinations
- implementovat tracing
- monitorovat produkční systém

### Agenda

#### 1. Evaluate RAG

- Retrieval metrics
- Precision@K
- Recall@K
- Groundedness
- Faithfulness

#### 2. LLM-as-a-judge

- Automatická evaluace

#### **GOPAS Praha**

Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

#### **GOPAS Brno**

Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

#### **GOPAS Bratislava**

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved

# Příprava a implementace RAG řešení do produkce v jazyce Python

- Human feedback loops

- Golden datasets

## 3. Observability

Nástroje:

- LangSmith
- Weights & Biases
- Arize AI
- OpenTelemetry

## 4. Cost optimization

- Token optimization
- Caching
- Smart retrieval
- Model routing

## 5. Praktický workshop

Implementace:

- evaluation pipeline
- tracing
- dashboards
- hallucination detection

Praktické zadání

Vyhodnotit kvalitu systému RAG na test datasetech.

## Den 5 — Deployment a produkční provoz

Téma dne

Nasazení enterprise-grade systému RAG do produkce.

Learning outcomes

Účastník bude umět:

- nasadit RAG do cloudu
- škálovat inference
- zabezpečit produkční systém
- připravit CI/CD pipeline

## Agenda

### 1. Produkční architektura

- API gateway
- Async processing
- Queue systems

#### GOPAS Praha

Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved

# Příprava a implementace RAG řešení do produkce v jazyce Python

- Microservices vs monolith

## 2. Deployment strategie

- Docker
- Kubernetes
- Serverless
- GPU inference

## 3. Security a governance

- IAM
- Secrets management
- Audit logging
- Compliance (GDPR)

## 4. CI/CD pro systémy AI

- Model versioning
- Prompt versioning
- Automated evaluation
- Canary releases

## 5. Závěrečný workshop

Nasazení kompletního řešení RAG:

- ingestion
- retrieval
- orchestration
- monitoring
- deployment

Technologie používané v rámci kurzu:

- Python
- FastAPI
- LangChain
- LlamaIndex
- LangGraph
- Docker
- Kubernetes
- PostgreSQL + pgvector
- Qdrant
- OpenAI / Anthropic APIs

**GOPAS Praha**  
Na Strži 2097/63  
140 00 Praha 4 - Krč  
Tel.: +420 226 201 390  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Brno**  
Nové sady 996/25  
602 00 Brno  
Tel.: +420 530 513 590  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Bratislava**  
Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2026 GOPAS, a.s.,  
All rights reserved