

# Tvorba LLM (Large Language Models) v jazyce Python

Kód kurzu: PYTHON\_LLM

Kurz je určen pro programátory, kteří chtějí pochopit, jak velké jazykové modely (LLM) fungují „uvnitř“, a prakticky si projít cestu od jednoduchých statistických jazykových modelů přes RNN/LSTM až po transformer a mini GPT. V druhé části kurzu se účastníci naučí pracovat s hotovými modely (Hugging Face), provést fine-tuning (včetně LoRA) a postavit praktickou aplikaci nad vlastními dokumenty pomocí RAG (retrieval-augmented generation). Součástí jsou i produkční aspekty: latence, optimalizace, kvantizace a nasazení jako API včetně Dockeru.

| Pobočka    | Dnů | Cena kurzu | ITB |
|------------|-----|------------|-----|
| Praha      | 5   | 28 500 Kč  | 50  |
| Brno       | 5   | 28 500 Kč  | 50  |
| Bratislava | 5   | 1 140 €    | 50  |

Uvedené ceny jsou bez DPH.

## Termíny kurzu

| Datum | Dnů | Cena kurzu | Typ výuky | Jazyk výuky | Lokalita |
|-------|-----|------------|-----------|-------------|----------|
|-------|-----|------------|-----------|-------------|----------|

Uvedené ceny jsou bez DPH.

### Pro koho je kurz určen:

- Programátory se základní znalostí Pythonu, kteří chtějí porozumět LLM a umět je prakticky využít.
- Vývojáře, kteří chtějí implementovat a trénovat malé modely, a následně pracovat s existujícími LLM (fine-tuning, RAG, nasazení).
- Data/ML nadšence, kteří chtějí získat ucelený praktický přehled od „nuly“ po aplikaci.

### Požadované vstupní znalosti

- Základní znalost jazyka Python na úrovni kurzu PYTHON\_INTRO
- Základní znalost NumPy na úrovni kurzu PYTHON\_DATAN
- Základy lineární algebry jsou výhodou (nejsou podmínkou)

### Metody výuky

- Odborný výklad s praktickými ukázkami, cvičení na počítačích.
- Práce v Jupyter Notebook a skriptech, průběžné miniprojekty.
- Důraz na pochopení principů a reprodukovatelnost.

### Studijní materiály

- Prezentace probírané látky v tištěné nebo online formě.
- Notebooky a referenční implementace k jednotlivým dnům.
- Ukázkové datasety a šablony pro trénování, vyhodnocení a nasazení.

### Osnova

Základy neuronových sítí a NLP

- Co je jazykový model
- Pravděpodobnost a predikce dalšího slova / tokenu
- Tokenizace
- Embeddingy
- Neuronová síť (perceptron, vrstvy, aktivace)
- Backpropagation (intuice)
- Python + NumPy
- Implementace unigram modelu

#### GOPAS Praha

Kodaňská 1441/46  
101 00 Praha 10  
Tel.: +420 234 064 900-3  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 542 422 111  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 248 282 701-2  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2020 GOPAS, a.s.,  
All rights reserved

# Tvorba LLM (Large Language Models) v jazyce Python

- Implementace bigram modelu
- Trénink malé sítě v PyTorch
- Výstup dne: malý jazykový model predikující další slovo na malé vzorce textu

## Rekurentní sítě a první generování textu

- RNN
- LSTM
- Problém miznouceho gradientu
- Teacher forcing
- Sampling (temperature, top-k)
- Implementace LSTM modelu v PyTorch
- Trénink na malém datasetu (např. Shakespeare)
- Generování textu
- Výstup dne: model generující krátké texty

## Transformer architektura

- Attention mechanismus
- Self-attention
- Multi-head attention
- Positional encoding
- Encoder vs Decoder
- Proč je transformer škálovatelný
- Implementace mini-transformera
- Vytvoření malého GPT-like modelu
- Trénink na malém datasetu
- Výstup dne: fungující mini GPT model

## Trénink, fine-tuning a práce s hotovými modely

- Pretraining vs Fine-tuning
- Transfer learning
- LoRA a parameter-efficient fine-tuning (PEFT)
- Tokenizéry (BPE)
- Použití knihovny Hugging Face Transformers
- Fine-tuning malého modelu
- Práce s modely (např. LLaMA / kompatibilní open-source modely dle dostupnosti)
- Vytvoření vlastního chatbot skriptu
- Výstup dne: fine-tuned model na vlastních datech

## RAG, nasazení a produkční aspekty

- Embeddingy pro vyhledávání
- Vektorové databáze (FAISS)
- RAG architektura
- Latence a optimalizace
- Kvantizace modelů
- Nasazení (API, Docker)
- Generování embeddingů
- Uložení do FAISS
- Implementace RAG pipeline
- Vytvoření jednoduchého API (FastAPI)
- Finální projekt: interní chatbot nad vlastními dokumenty

## Používané technologie

### GOPAS Praha

Kodaňská 1441/46  
101 00 Praha 10  
Tel.: +420 234 064 900-3  
[info@gopas.cz](mailto:info@gopas.cz)

### GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 542 422 111  
[info@gopas.cz](mailto:info@gopas.cz)

### GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 248 282 701-2  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2020 GOPAS, a.s.,  
All rights reserved

# Tvorba LLM (Large Language Models) v jazyce Python

- Python 3.11+
- PyTorch
- Hugging Face Transformers
- FAISS
- FastAPI
- Jupyter Notebook

## Co si účastník odnese

- Rozumí, jak LLM fungují vnitřně a jaké jsou jejich stavební bloky
- Umí postavit jednoduchý transformer a mini GPT na malém datasetu
- Umí fine-tunovat model včetně LoRA/PEFT
- Umí postavit RAG aplikaci nad vlastními dokumenty
- Umí nasadit model jako API a chápe základní produkční aspekty (latence, optimalizace, kvantizace)

**GOPAS Praha**  
Kodaňská 1441/46  
101 00 Praha 10  
Tel.: +420 234 064 900-3  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Brno**  
Nové sady 996/25  
602 00 Brno  
Tel.: +420 542 422 111  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Bratislava**  
Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 248 282 701-2  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2020 GOPAS, a.s.,  
All rights reserved