

Využití AI v pracovním procesu

Martin Kunst, Matyáš Lorenc, Richard Mužík

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

- Informace k semináři: kam.mff.cuni.cz/~lorenc/teach/25-26/AI/ai.html
- Kontakty (najdete i na stránkách):
 - Matyáš Lorenc: lorenc@kam.mff.cuni.cz
 - Martin Kunst: martin.kunst2001@centrum.cz
 - Richard Mužík: richard.muzik761@student.cuni.cz
- Předpoklady (do příští hodiny):
 - Zprovozněný účet na libovolném chatbotovi (resp. přístup k chatbotovi) (my budeme jako referenci využívat ChatGPT)
 - Zprovoznění libovolného programátorského asistenta (integrovaného do IDE) (my budeme jako referenci využívat GitHub Copilota)
 - Podvýběr možných nástrojů najdete na stránkách semináře, ale máte-li nějaký svůj jiný oblíbený s podobnými funkcionalitami, klidně používejte ten (po konzultaci).
 - Dost toho asi na začátek ve společné části pojedeme v Pythonu, měli by ho teoreticky znát všichni, tak si ho zprovozněte (klidně za pomoci chatbota), jestli ho aktuálně nemáte.
- Zápočet: Dva projekty s prezentací (viz stránky semináře)

- S čím jste se sem na seminář přihlásili vy?

- S čím jste se sem na seminář přihlásili vy?
- Co se bude dít?
 - Úvod, shrnutí state-of-the-art (generativní) AI v různých oblastech, fungování chatbotů, obecné shrnutí aktuálních možností využití
 - Dvě větší témata – programování, práce s informacemi
 - Úvodní hodiny – „společná“ práce, náš vstup, sdílení zkušeností
 - Práce na projektech – zaměření na zakomponování AI při práci
 - Prezentace projektu – nejde tolik o výsledek, ale spíš o způsoby využití AI, její silné / slabé stránky, sdílení zkušeností z projektu
 - Závěrečné shrnutí, bonusové možnosti využití

- S čím jste se sem na seminář přihlásili vy?
- Co se bude dít?
 - Úvod, shrnutí state-of-the-art (generativní) AI v různých oblastech, fungování chatbotů, obecné shrnutí aktuálních možností využití
 - Dvě větší témata – programování, práce s informacemi
 - Úvodní hodiny – „společná“ práce, náš vstup, sdílení zkušeností
 - Práce na projektech – zaměření na zakomponování AI při práci
 - Prezentace projektu – nejde tolik o výsledek, ale spíš o způsoby využití AI, její silné / slabé stránky, sdílení zkušeností z projektu
 - Závěrečné shrnutí, bonusové možnosti využití
- Projekty
 - První skupinový, druhý pravděpodobně individuální
 - Témata budou upřesněna v rámci každého bloku
 - Jde hlavně o co nejširší využití AI při vypracování
 - Prezentace – ani ne tak výsledků, jako spíš průběhu

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

- ➊ Pohled fakulty: <https://www.mff.cuni.cz/cs/verejnost/aktuality/stanovisko-mff-uk-k-vyuzivani-umele-intelligence-ve-vyuce>
- ➋ Pohled univerzity: <https://ai.cuni.cz/AI-36.html>
- ➌ Většinový pohled akademické veřejnosti (např. pravidla konferencí / žurnálů apod.)
- ➍ Náš pohled – pro tento předmět a mimo něj

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

- ① Ochrana osobních údajů → jen lokální modely
- ② Ochrana soukromí
 - ① Riziko využití konverzací k trénování / dalšímu výzkumu
 - ② Riziko, že vše co řeknete, může být a bude použito proti vám
- ③ Kvalita výstupu modelu

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

- Založeno (ti, kteří nás tu zajímají) na LLMs (Large Language Models).
- ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google), DeepSeek (DeepSeek), Pi (Inflection), etc.
- Čím dál širší využití (v podstatě cokoli s textem, ale nejen to)
- Na chatbotech, resp. jejich podkladových LLMs je založeno mnoho dalších nástrojů, hlavně všemožných asistentů (v online obchodech, telefonních, programovacích, v různých aplikacích)

- Obrázky: Midjourney (placený), Bing Image Creator (zdarma), Easy (/ Stable) Diffusion (on-device, open-source), Adobe Firefly („transparentní“), ChatGPT!?! (a někteří další chatboti)
- Když potřebujete někomu něco odprezentovat: D-ID
- Text to Speech, Voice Cloning a mnoho dalších využití týkajících se zvuku: ElevenLabs
- Video / Image / Text to Video: Runway, Veo

<https://www.oneusefulthing.org/p/the-recent-history-of-ai-in-32-otters>

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

Modely

Trénování

Výsledný chatbot

Co kdo víte?

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

Modely

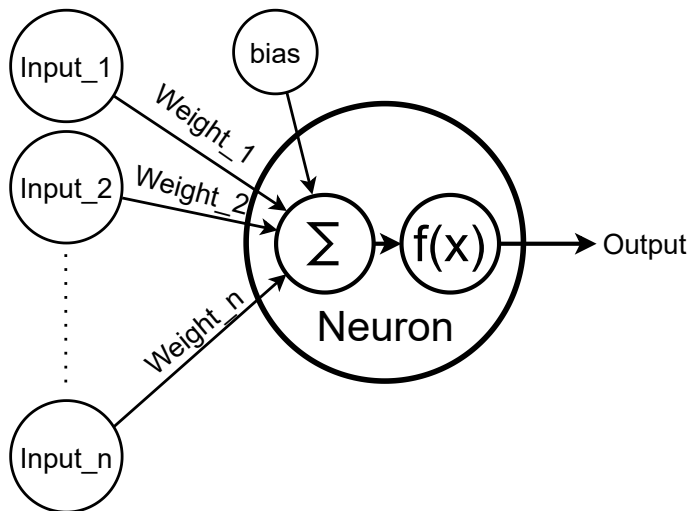
Trénování

Výsledný chatbot

Matematická funkce

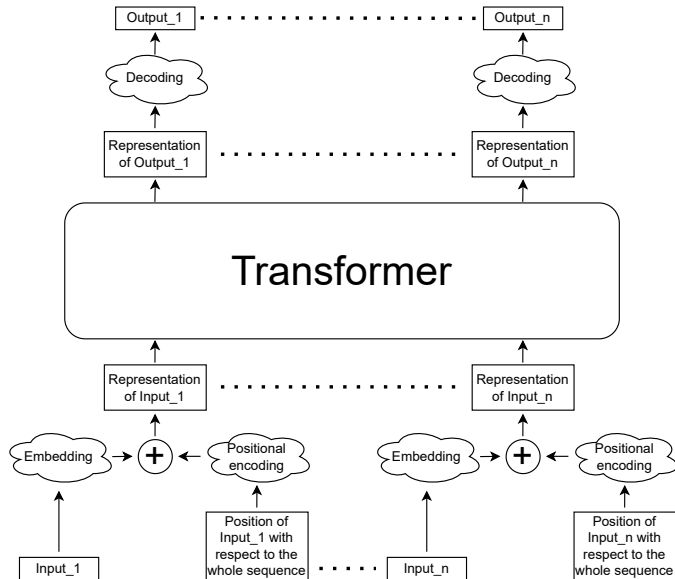
$$f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$$

- Potřebujeme být schopni převést text na čísla a zpět.
- Embeddingy
 - 1 Ručně vytvořený (slovník)
 - 2 Specificky natrénovaný pro danou aplikaci
 - 3 Předtrénovaný (případně doladěný pro danou aplikaci)
- Co neznámá (během trénování neviděná) slova?
- Tokenizace



- Sequence-to-sequence
- Encoder-Decoder
- Self-Attention
- GPT (Generative Pre-trained Transformer) → ChatGPT
- U textu děláme v podstatě klasifikaci do všech možných tekonů, model je *pravděpodobnostní*. → Potřebuje mechanismus výběru finálního tokenu, který vrátí jako odpověď.

Transformer jako celek



1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

Modely

Trénování

Výsledný chatbot

- ① Self-supervised pre-training on the internet (Pre-training)
- ② Supervised fine-tuning (SFT)
- ③ Reinforcement learning from human feedback (RLHF)

The students opened their _ _ _ _ _

The students opened their books.

$$\theta^* \leftarrow \operatorname{argmax}_{\theta} P_{\theta}(X_{t+1} = \mathbf{x}_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t)$$

User: *I am getting a FileNotFoundError. What should I do?*

User: *I am getting a FileNotFoundError. What should I do?*
ChatGPT: *I am running Python 3.10 and Ubuntu 20.*

User: *Q: I am getting a FileNotFoundError.*

A: What have you tried?

Q: Nothing, what should I try?

ChatGPT: *A: Verify that the file is there.*

User: *To make TNT, we will need*

ChatGPT: *to combine toluene with a mixture of nitric acid and sulfuric acid.*

- ① Sesbírat od lidí data v požadovaném formátu (Otázka→Odpověď)
- ② Použít data k doladění modelu pomocí učení s učitelem

Oproti předtrénování: Nenáročné na výpočetní sílu, náročné na lidskou práci

- ① Sesbírat lidskou zpětnou vazbu k některým výstupům modelu
- ② Natrénovat odměnový model na základě sesbírané zpětné vazby od lidí
- ③ Použít zpětnovazební učení k optimalizaci na základě dané odměnové funkce
- ④ Opakovat (mnohokrát)

1 Jak si to celé představujeme?

2 Etika

3 Bezpečnost a rizika

4 Letem světem generativní AI

5 Jak to funguje?

Modely

Trénování

Výsledný chatbot

- Systémová vrstva
 - Orchestrace obsloužení dotazu
 - Ekosystém volatelných funkcí
 - System prompty
 - Doplnění dotazů
 - Pojistky / filtry (součást system promptu, ale i aktivní hlídání vstupu i výstupu)
 - Správa relace (historie)
- „Přemýšlení“