

Akademija umetne inteligence za poslovne aplikacije

Vpeljava velikih jezikovnih
modelov (ChatGPT, BERT)
v poslovne aplikacije



V treh dneh delavnic se bodo udeleženci spoznali ne le z osnovami velikih jezikovnih modelov ter koraki vpeljave, temveč bo poudarek predvsem na praktičnih primerih, skozi katere bomo predstavili dejanske aplikacije modelov kot sta BERT in ChatGPT.



8. 4., 15. 4. in
22. 4. 2024.



9:00 - 15:00



INFORMACIJE IN
PRIJAVA



Začnemo 8. 4. 2024

PRVA DELAVICA - Kdaj: 8. 4. 2024, od 9:00 do 14:00, dvorana B (1. nad)

Priložnosti in pasti generativne umetne inteligence in jezikovnih modelov (1. del)

Program delavnice:

- Uvod v umetno inteligenco:
 - Generativni in diskriminativni pristopi;
- Osnove velikih jezikovnih modelov (LLMs):
 - Osnove delovanja;
 - Pregled razvoja in področja;
- Priložnosti vpeljave jezikovnih modelov v podjetjih:
 - Koraki uvedbe;
 - Primeri uporabe iz prakse;
- Pasti pri vpeljavi jezikovnih modelov (ter kako jih ublažiti):
 - Halucinacije, natančnost, transparentnost, stroški.

- Pregled velikih jezikovnih modelov (OpenAI ChatGPT in GPT-4, Google Gemini, razne odprtokodne alternative)

Opis:

Predavanje služi kot uvod v uvajanje generativne umetne inteligence v podjetjih. Udeleženci se bodo spoznali z osnovami umetne inteligence ter strojnega učenja s fokusom na generativnih pristopih. Kot trenutno najbolj razširjeni družini pristopov generativne umetne inteligence, se bomo osredotočili predvsem na velike jezikovne modele. Predstavili bomo glavne koncepte njihovega delovanja, za katere verjamemo, da so ključni za razumevanje dobrih praks uporabe ter vpeljave te tehnologije. Skozi primere iz prakse se bodo udeleženci spoznali s priložnostmi vpeljave jezikovnih modelov, prav tako bomo predstavili najboljše prakse za uvedbo v poslovanje. Pregledali bomo pasti in probleme, s katerimi se lahko srečamo ob vpeljavi takih modelov, kot na primer halucinacija, tj. generacija netočnih informacij, ki so videti verjetne. Predstavili bomo, kako lahko v praksi rešujemo takšne težave. Zaključili bomo s pregledom obstoječih modelov, kot je ChatGPT, in predstavili nekatere alternative (tako lastniške kot odprtokodne).

Učinkovita komunikacija z velikimi jezikovnimi modeli (*prompt engineering*) (2. del)**Program delavnice:**

- Osnove pozivanja:
 - Osnove učenja velikih jezikovnih modelov in vpliv na pozivanje;
 - Sistemski, uporabniški pozivi;
 - Pregled dobrih praks pozivanja;
 - Napredno pozivanje - med drugim:
 - Pregled razvoja pozivanja (od učenja z nekaj primeri do učenja brez primerov);
 - Dodeljevanje vlog;
 - Tok misli (*chain of thought*);
 - Samodoslednost (*self-consistency*);
 - Praktični primeri pozivanja (primeri za povzemanje, transformacijo, klasifikacijo, obravnavo dolgih dokumentov, itd.);
 - Pregled "hekanja" pozivov - primeri napadov in kako jih preprečiti.

Opis:

Prompt Engineering oziroma inženiring pozivov/ukazno inženirstvo se je v zadnjem času izoblikovalo kot nova disciplina razvoja pozivov, s katerimi lahko velike jezikovne modele učinkovito uporabljamo za najrazličnejše naloge. Razvijanje pozivov zahteva različna znanja povezana z jezikovnimi modeli, kot so razumevanje njihovih zmognosti in omejitev, poznavanje potencialnih zlonamernih pozivov in zaščita pred njimi. Udeleženci se bodo na predavanju spoznali z opisani koncepti in pridobili znanje na učinkovitejšo rabo jezikovnih modelov v svojem delu in vsakdanjiku.

V uvodnem delu predavanja se bomo dotaknili učenja velikih jezikovnih modelov v luči vpliva le-tega na kasnejše pozivanje. Predstavili bomo razlike med sistemskimi in uporabniškimi pozivi, ter se spoznali z dobrimi praksami. Udeleženci se bodo prek praktičnih primerov spoznali z naprednejšimi tehnikami pozivanja, kot sta med drugim tok misli (*chain of thought*) in samodoslednost (*self-consistency*). Pregledali bomo razvoj pozivanja od učenja z nekaj primeri (*few-shot learning*) do učenja brez primerov (*zero-shot learning*), torej od modelov kot sta GPT-2 in GPT-3 do modernih modelov specializiranih za sledenje uporabnikovim navodilom (kot sta ChatGPT in GPT-4). Udeležencem bomo predstavili glavne skupine problemov, na katere se preslika veliko nalog, ki jih rešujemo z jezikovnimi modeli (recimo povzemanje, klasifikacija, ipd.). Na koncu bomo udeležence spoznali tudi s potencialnimi zlonamernimi napadi na jezikovne modele (recimo injiciranje, razkrivanje pozivov) ter z rešitvami, kako se pred njimi zaščititi.

Agenda 1. dan:

1. dan:

8:45 - 9:00 - Zbor udeležencev

9:00 - 11:00 - Uvodno predavanje (Priložnosti in pasti generativne umetne inteligence in jezikovnih modelov)

11:00 - 11:15 - Odmor

11:15 - 12:00 - Predavanje (Prompt inženiring - uvod v pozivanje)

12:00 - 13:00 - Odmor za kosilo

13:00 - 14:15 - Predavanje (Prompt inženiring - pregled tehnik pozivanja)

14:15 - 14:30 - Odmor

14:30 - 15:30 - Predavanje (Prompt inženiring - pregled tehnik pozivanja)

DRUGA DELAVNICA - Kdaj: 15. 4. 2024, od 9:00 do 15:00, dvorana F (medetaža)

Sodobna obdelava naravnega jezika: BERT prek praktičnih primerov

Program delavnice:

- Teoretično predavanje
 - Motivacija za arhitekturo transformer,
 - Pregled razvoja jezikovnih modelov;
 - BERT in GPT - razlike, prednosti in slabosti;
 - Učenje jezikovnih modelov;
 - Praktične vaje z uporabo BERT-a:
 - Pregled HuggingFace ekosistema za razvoj in uporabo velikih jezikovnih modelov;
 - Praktični primeri:
 - Zaznava imenskih entitet v besedilih;
 - Klasifikacija besedil in prepoznavanje sentimenta;
 - Iskanje (podobnih) besedil z BERT-ovimi vložitvami.

Opis:

Delavnica je razdeljena na teoretični in praktični del. V teoretičnem delu bomo prek pregleda razvoja področja motivirali vpeljavo arhitekture transformer, na kateri slonijo veliki jezikovni modeli kot sta BERT in ChatGPT. Udeležencem bomo predstavili dva načina razvoja modelov na področju obdelave jezika: kodirnike kot je BERT in dekodirnike kot je GPT družina modelov. Predstavili bomo prednosti in slabosti obeh pristopov ter primere uporabe, hkrati pa se bomo dotaknili učenja takšnih jezikovnih modelov.

V praktičnem delu bomo udeležencem predstavili platformo HuggingFace. Gre za skupek knjižnic, ki tvorijo temelj sodobne obdelave jezika, saj poenostavljajo uporabo vnaprej naučenih velikih jezikovnih modelov. Udeleženci bodo pridobljeno znanje lahko preuzkusili na vnaprej pripravljenih realnih primerih, ki se rešujejo z uporabo BERT modelov. Za ta namen bodo uporabili pripravljene Jupyter beležke, podatkovne množice in dodatno učene jezikovne BERT modele. Vsi primeri bodo imeli tako slovensko kot angleško različico, tako da se bodo udeleženci lahko seznanili z različnimi modeli, ter bodo lahko pridobljeno znanje uporabili na bolj splošnih problemih. Najprej se bomo lotili prepoznavanja imenskih entitet v besedilih, to je avtomatska zaznava osebnih, zemljepisnih imen, itd. Nadaljevali bomo s klasifikacijo besedil, kjer vhodnemu besedilu z naučenim BERT modelom določimo enega izmed vnaprej določenih razredov. Primer tega je prepoznavanje sentimenta, kjer besedilu določimo ali gre za pozitivno ali negativno naravnano besedilo. Zaključili bomo z uporabo BERT-a za iskanje podobnih besedil, tu gre za primer semantičnega iskanja, ki je hrbtenica modernih spletnih iskalnikov kot je Google.

Celoten praktični del delavnice je zastavljen tako, da omogoča neposredno uporabo pridobljenega znanja v realnih aplikacijah.

Komuniciranje aplikacij in jezikovnih modelov: GPTs

Program delavnice:

- Pregled obstoječih GPT-jev;
- Razvoj lastnih GPT-jev (praktični primeri):
 - Osnovni primer: TODO seznam;
 - Interakcija s spletno trgovino;
 - Interakcija z internimi dokumenti (Excel, Word, ...);

Opis:

V sklopu tega dela delavnic bomo spoznali GPT-je; le-ti so način, da prilagodimo ChatGPT za različne naloge. Med drugim nam omogočajo komunikacijo z internimi dokumenti in interakcijo z našimi aplikacijami. Na delavnici bomo predstavili GPT-je prek praktičnih primerov, kot sta interakcija s spletno trgovino in odgovarjanje na vprašanja o internih dokumentih.

Agenda 2. dan:

2. dan:

8:45 - 9:00 - Zbor udeležencev

9:00 - 10:30 - Predavanje (uvod in motivacija - Sodobna obdelava naravnega jezika: BERT prek praktičnih primerov)

10:30 - 10:45 - Odmor

10:45 - 12:00 - Predavanje (praktični primeri - Sodobna obdelava naravnega jezika: BERT prek praktičnih primerov)

12:00 - 13:00 - Odmor za kosilo

13:00 - 14:15 - Predavanje (praktični primeri - Sodobna obdelava naravnega jezika: BERT prek praktičnih primerov)

14:15 - 14:30 - Odmor

14:30 - 16:30 - Predavanje (Komuniciranje aplikacij in jezikovnih modelov: GPTs)

TRETJA DELAVNICA Kdaj: 22. 4. 2024, od 9:00 do 15:00, dvorana F (medetaža)

Komuniciranje aplikacij in jezikovnih modelov: ChatGPT funkcije

Program delavnice:

- Razlika med GPT-ji in ChatGPT funkcijami;
- Replikacija GPTs funkcionalnosti s funkcijami:
 - Osnovni primer: TODO seznam;
 - Interakcija s spletno trgovino;
 - Interakcija z internimi dokumenti (Excel, Word, ...);

Opis: V tej delavnici bomo podrobneje raziskali ChatGPT funkcije, katerih funkcionalnost je podobna funkcionalnosti GPT-jev. GPT-ji, ki smo jih že spoznali, so integrirani v produkt ChatGPT (spletna pogovorna aplikacija). Po drugi strani so nam ChatGPT funkcije na voljo prek programskega vmesnika (OpenAI API). Omogočajo nam, da se ChatGPT sklicuje na funkcije v naši kodi, ki mu jih predstavimo - tako ga lahko povežemo na zunanje API-je, funkcionalnosti znotraj naših aplikacij, itd. Kot take so funkcije predvsem uporabne pri lastnih postavitvah chatbotov, recimo pri virtualnih asistentih na lastnih spletnih straneh.

V praktičnem delu delavnice bomo z uporabo ChatGPT funkcij replicirali primere, ki smo jih predhodno implementirali z uporabo GPT-jev.

Sodobna obdelava naravnega jezika: nadgradnja ChatGPT-ja s knjižnico LangChain

Program delavnice:

- Pregled LangChain knjižnice za razvoj aplikacij z velikimi jezikovnimi modelov:
 - Glavni koncepti: modeli, spomin, pozivi, verige (*chains*), razčlenjevalniki izhoda;
 - Praktični primeri uporabe knjižnice LangChain:
 - Asistent za odgovarjanje na vprašanja o internih dokumentih;
 - Pogovorni robot s spominom: reševanje problema omejene dolžine konteksta;
 - Poizvedovanje po SQL bazah v naravnem jeziku (ali po Excel dokumentih).

Opis:

V uvodnem delu delavnice bomo predstavili knjižnico LangChain in njene glavne gradnike, (kot so modeli, spomin, verige, razčlenjevalniki izhoda, itd). Skupaj z udeleženci bomo v interaktivnem praktičnem delu nato razvili rešitve za 3 primere iz prakse, s katerimi se na področju obdelave jezika pogosto srečamo podatkovni znanstveniki. Razvili bomo **virtualnega asistenta**, ki bo lahko odgovarjal na vprašanja o internih dokumentih. Za ta namen bomo naprej postavili vektorsko bazo internih dokumentov (le-ti bodo pripravljene vnaprej), na katero se bo lahko skliceval virtualni asistent. Drugi praktični primer bo **pogovorni robot s spominom**, kar pomeni, da si bo lahko zapomnil predhodne interakcije z uporabnikom. Zadnji primer iz prakse pa bo **asistent za poizvedovanje po tabularičnih podatkih**, bodisi v SQL bazi ali Excel dokumentih, kjer uporabnik ne bo potreboval znanja SQL, pač pa bo lahko poizvedbo asistentu napisal kar v naravnem jeziku. Asistent bo nato to poizvedbo prevedel v ustrezno obliko in uporabniku vrnil rezultat.

Agenda 3. dan

3. dan:

8:45 - 9:00 - Zbor udeležencev

9:00 - 11:00 - Predavanje (Komuniciranje aplikacij in jezikovnih modelov: ChatGPT funkcije)

11:00 - 11:15 - Odmor

11:15 - 12:00 - Predavanje (uvod in motivacija - Sodobna obdelava naravnega jezika: nadgradnja ChatGPT-ja s knjižnico LangChain)

12:00 - 13:00 - Odmor za kosilo

13:00 - 14:45 - Predavanje (praktični primeri - Sodobna obdelava naravnega jezika: nadgradnja ChatGPT-ja s knjižnico LangChain)

14:45 - 15:00 - Odmor

15:00 - 16:30 - Predavanje (praktični primeri - Sodobna obdelava naravnega jezika: nadgradnja ChatGPT-ja s knjižnico LangChain)



Združenje za
informatiko in
telekomunikacije



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSKI
RAZVOJ IN TEHNOLOGIJO



Sofinancira
Evropska unija

Naložbo sofinancira Evropska unija iz Evropskega sklada za regionalni razvoj