

# PODPORA ZÁKAZNÍKOV A POUŽÍVATEĽOV

Scenár digitalizácie s využitím umelej inteligencie



## POPIS

Virtuálny Digitálny Asistent, tzv. **Chatbot** je expertný systém, ktorý pomocou metód umelej inteligencie napodobňuje komunikáciu človeka. Pojmom expertný systém je myslené to, že znalosti chatbota (z angl. skills) sú vytvorené manuálne. Chatbot môže byť doplnený o znalostné vyhľadávanie v poskytnutých dokumentoch alebo poskytovanie informácií z databáz, informačných systémov, vyvolávať akcie či skripty, alebo prepojiť používateľa na agenta call centra. Primárne je vstupom chatbota text v chatovacom (dialógovom) okne. Odpoveďami chatbota môžu byť text, tlačidlo alebo audiovizuálny obsah. Chatbota je možné doplniť o systém automatického rozpoznania reči a systém prevodu textu na reč.

Vo všeobecnosti sa chatbot skladá z viacerých blokov.

**Možnosti a limitácie chatbotov môžu byť v skratke popísané nasledovne:**

### Komunikačný kanál:

Komunikačným kanálom môže byť webové rozhranie alebo časť webovej stránky (iframe). Chatbotové frameworky často poskytujú možnosti jednoduchého prepojenia s populárnymi aplikáciami alebo sociálnymi sieťami ako napr. Facebook, Slack, atď. Do tohto bloku je možné pridať aj autentifikáciu.

### Systém automatického rozpoznania reči (z angl.

Automatic Speech Recognition, skr. ASR)

(voliteľný doplnok):

Cieľom tohto systému je prevod vyslovenej reči používateľom na text. Systémy ASR fungujú v súčasnosti so slovnou chybovosťou <5 %. Chybovosť je závislá od kvality zvukovej nahrávky, typu mikrofónu, vzdialenosti rečníka od mikrofónu, hluku v pozadí ale aj od artikulácie, tempa reči alebo nárečia a celkovo od kvality rečového prejavu. Štandardne systémy ASR vedia rozpoznať iba slová v slovníku, ktorý býva z viacerých dôvodov rozumne limitovaný. V prípade vyslovenia mimoslovníkového slova je vybraté najpravdepodobnejšie slovníkové slovo alebo slová. Navrhnuť si vlastný systém ASR je veľmi zložitý a okrem expertov na danú oblasť vyžaduje aj množstvo tréningových dát, ktoré sú veľmi drahé. V praxi sa na automatické rozpoznávanie reči používajú cloudové služby alebo on-premise softvéry (na vlastných serveroch).

V prípade, že slovná chybovosť nie je dostačujúca, je možné požiadať dodávateľa o použitie špecifického (doménového) jazykového modelu. Ten napomáha pri rozpoznávaní reči tak, že slová a slovné spojenia často vyskytujúce sa v doméne majú vyššiu pravdepodobnosť pri rozpoznávaní reči, a preto môžu znížiť slovnú chybovosť. To platí hlavne pre špecifické domény ako napr. zdravotníctvo, právo, hovorová reč, atď.

### Porozumenie prirodzeného jazyka

(z. angl. Natural Language Understanding, skr. NLU):

Cieľom tohto bloku je pochopenie textu zadaného používateľom, čo je nazývané ako zámer (z angl. intent). Výsledkom tohto bloku je rozpoznanie jedného z konečnej množiny zámerov v datasete. V prípade, že klasifikačný algoritmus neprekročí stanovený prah dôveryhodnosti, je možné vypísať záložnú odpoveď. Napr. „Nerozumiem vám, prosím, skúste vetu preformulovať“.

Každý text zadaný používateľom je spracovávaný týmto blokom. Jadrom tohto bloku je model strojového učenia pre textovú klasifikáciu. Trénovanie tohto modelu zabezpečuje chatbot framework automaticky a pre jednoduchšie aplikácie nie sú potrebné špeciálne vedomosti.

### Rozpoznávanie pomenovaných entít

(z angl. Named Entity Recognition, skr. NER)

(voliteľný doplnok, v praxi používaný takmer vždy):

Tento blok zabezpečuje rozpoznanie entít v texte zadanom používateľom. Entity môžu byť akékoľvek textové dáta – meno, priezvisko alebo číslo objednávky. Na rozpoznanie jednoduchých entít ako meno, priezvisko, adresa jestvuje viacero nástrojov. Na rozpoznávanie vlastných entít je potrebné vytvoriť vlastný dataset (označením entity v texte) a natrénovať z neho model. Natrénovanie takého modelu sa štandardne robí automaticky v chatbot frameworkoch a pre jednoduchšie aplikácie nie sú potrebné špeciálne vedomosti.

### Dialógový manažér

(z angl. Dialog Manager, skr. DM):

Tento blok zabezpečuje riadenie dialógu. DM sleduje aké zámary boli rozpoznané v NLU, aké entity boli rozpoznané a následne predpovedá nasledujúcu akciu. Štandardné akcie môžu byť:

- Odpoveď (text, audiovizuálny obsah, informácia z externého systému, výsledok skriptu alebo vyhľadaná informácia v znalostiach – dokumentoch);
- Vyžiadanie si viac informácií – entít;
- Čakanie na ďalšiu správu od používateľa;
- Ukončenie konverzácie.
- Jadrom tohto bloku je model strojového učenia. Z praxe však vyplynulo, že čím kvalitnejšie sú tréningové dáta pre DM, tým lepšie funguje. Inak povedané – čím viac možností konverzácie je zapísaných v tréningových dátach, tým lepšie funguje riadenie dialógu.

### Systém prevodu textu na reč

(z angl. Text to Speech, skr. TTS) (voliteľný doplnok):

Systém prevodu textu na reč (známy aj ako syntéza reči) prevádza textové odpovede na reč. V súčasnosti dosahuje veľmi dobré výsledky a reč znie skoro prirodzene. Vytvorenie vlastného systému TTS nie je jednoduché a vyžaduje spracovanie expertami na danú oblasť a kvalitné rečové dáta s presnými časovo zarovnanými prepismi. Pre tento

blok je vhodné radšej využiť cloudové služby alebo on-premise riešenia.

## PREDPOKLADY

### Proces

Chatboty sa štandardne používajú v oblastiach:

- **Call centrá** – časť zákazníkov kontaktujúcich call centrum (volaním alebo písaním) môže byť obslužená chatbotom. Jednoduchšie úlohy zvládajú chatboty veľmi dobre, ako napríklad zobrazenie výšky faktúry. V prípade, že chatbot nevie naplniť požiadavky používateľa, je možného používateľa automaticky prepojiť na operátora.
- **Poskytovanie informácií používateľom** – systém môže používateľom poskytovať informácie, odpovedať na frekventované otázky alebo vyhľadať poruchu u používateľa a postupne smerovať k vyriešeniu, prípadne pomôcť s hlbšou diagnostikou problému alebo vyhľadávať v dokumentoch.
- **Samoobslužné systémy** – pomocou samoobslužných chatbotov si môže používateľ objednať alebo upraviť nejakú službu, prípadne zmeniť nastavenie systému. V tomto prípade je nutná aj autentifikácia.
- **Reklamné kampane** – chatbot na reklamné kampane môže napríklad informovať používateľa o akciách, pričom táto konverzácia je vyvolaná chatbotom, nie používateľom. Z etického hľadiska je v tomto prípade vhodné dať používateľovi možnosť sa jednoducho odhlásiť z tejto komunikácie. Komunikácia môže byť uskutočnená textom alebo telefonickým rozhovorom (v slovenčine sa zatiaľ veľmi nepoužíva).

Z praxe vyplynulo, že používatelia dobre prijímajú komunikáciu s chatbotom. V niektorých prípadoch používatelia priam preferujú komunikáciu s chatbotom pred živým operátorom. Chatbot tak môže značne zredukovať počet rozhovorov s operátorom a znížiť tak náklady na pracovnú silu.

### Dáta

Tréningové dáta pre porozumenie prirodzeného jazyka (NLU) je potrebné vytvoriť manuálne. V prípade jednotlivých zámerov je potrebné napísať niekoľko možností, akými by ich mohol napísať používateľ. Čím viac možností sa uvedie, tým bude komunikácia s chatbotom prirodzenejšia. Nie je však nutné uvádzať všetky možnosti. Napríklad pre zámer „výška faktúry“ môže stačiť 10 rôznych príkladov. NLU je vlastne textový klasifikátor na báze strojového učenia. S vhodným modelom vektorovej sémantickej reprezentácie slov môže fungovať rozpoznávanie zámeru aj pre také texty, ktoré nie sú uvedené medzi príkladmi.

K bloku rozpoznávania entít (NER) je možné pristúpiť dvomi spôsobmi. Prvým je použitie voľne dostupných NER modelov, ktoré však rozpoznávajú len obmedzené množstvo rôznych entít. Druhá možnosť je označiť jednotlivé entity v tréningových dátach, konkrétne v dátach pre tréningovanie NLU. Príklad: „Chcel by som zistiť vlakové spojenie z Košíc do Bratislavy“ môžeme označiť „vlakové“ ako typ dopravy,

„Košíc“ ako štartovacie mesto a „Bratislavy“ ako destináciu. Pre jednotlivé entity môže byť vytvorená vyhľadávacia tabuľka, v ktorej môže byť zoznam všetkých slovenských obcí a miest.

Najkomplikovanejšie a najzdĺhavejšie je vytváranie dát pre DM. V týchto dátach je potrebné symbolicky zapísať všetky (alebo väčšinu) možností vývoja dialógu. Symbolicky znamená, že sa zapisujú iba názvy zámerov a názvy odpovedí. Chatbotovi je možné definovať viacero možností odpovede, aby komunikácia bola prirodzenejšia a menej strojová. Chatbot náhodne vyberá odpoveď z množiny odpovedí. Napr. odpoveď pre rozlúčenie môže byť: „Dovidenia“, „Pekný deň“, atď. Tieto odpovede je potrebné taktiež manuálne zadefinovať.

V tréningových dátach je možné zadať rôzne dialógové sledy v prípade, že bola detegovaná špecifická entita. Takisto je možné v dialógu zadať, ktoré entity sú povinné a chatbot sa na ne môže postupne opýtať – napr. „Zadajte číslo OP“, „Zadajte RČ“.

Jadrom DM je model strojového učenia, ktorý je trénovaný z hore uvedených dát. Model sa snaží v maximálnej miere kopírovať symbolicky popísané dialógové sledy. V prípade, že neexistuje konkrétny sled v tréningových dátach, model sa snaží nájsť najvhodnejšiu alternatívu, poprípade vypísať chybovú hlášku alebo prepojiť používateľa na operátora.

### Aplikácie

Chatbot riešenia je možné vytvoriť pomocou cloudovej služby alebo na vlastných zariadeniach. Cloudové služby sú štandardne platené podľa počtu relácií, v prípade ASR/TTS počtu sekúnd nahrávky. Existuje aj množstvo frameworkov (aj open source) na vytvorenie chatbot riešení na vlastnom hardvéri. V dnešnej dobe sú cloudové služby, ale aj frameworky vytvorené tak, aby boli čo najjednoduchšie a prípadne, aby nebolo potrebné spracovanie expertom (data scientist).

### Infraštruktúra

Systémy na vlastnom hardvéri sa delia na frontend a backend. Frontend je komunikačným rozhraním a backend štandardne vykonáva NLU, NER a DM. Rečové rozhrania ASR a TTS sa väčšinou implementujú do frontendu. Výpočtové zaťaženie backendu závisí od metód strojového učenia použitých v NLU a NER.

Čím komplexnejší model, tým je potrebný aj väčší počet výpočtov. Väčšinou sa používajú jednoduchšie modely, ktoré pracujú rýchlejšie, a teda jednotlivý server zvládne viac relácií. V prípade používania on-premise ASR/TTS riešenia je výpočtové zaťaženie veľmi veľké. Na spracovanie sú potrebné stroje s hardvérovými akcelerátormi, ako sú GPU, FPGA, NNP.

### Ľudia

Na nasadenie riešenia je potrebný frontend developer. Pre backend na vlastných serveroch je potrebný skúsený

IT developer, ktorý si naštuduje konkrétny framework. Pre produkčné riešenie je potrebné myslieť na automatické testovanie a monitoring.

Na úspešné nasadenie a udržiavanie chatbota je potrebný manažér obsahu, ktorý bude vytvárať tréningové dáta pre NLU, NER a DM. Počet manažérov obsahu závisí od komplexnosti riešenia a dostupných schopností chatbota.

Pre uchovanie konverzačných dát je potrebné myslieť aj na anonymizáciu dát tak, aby spĺňali nariadenia GDPR. Preto je potrebná manuálna alebo automatická anonymizácia textov.

#### Profily na strane zadávateľa:

- Vedúci podpory zákazníkov spolupracujú s dátovými vedcami na určení problémov a cieľov;
- Informatici zodpovedajú za prevádzku výpočtovej techniky a sieťovej infraštruktúry;
- Pracovníci zodpovedajú za obsluhu zákazníka, za predaj, za kvalitu a včasnosť servisu v existujúcom helpdesk pracovišku, call centre, atď.

#### Profily s IKT zameraním či už v internom tíme alebo na strane dodávateľa, poprípade pracovníci:

- Dátoví inžinieri spravujú dáta a príslušnú dátovú platformu, aby bola plne funkčná pre analýzu;
- Dátoví vedci, ktorí pripravujú, študujú, vizualizujú a modelujú dáta na platforme dátovej vedy;
- Experti na technológiu konkrétneho chatbota;
- IT architekti spravujú základnú infraštruktúru potrebnú pre podporu dátovej vedy;
- Vývojári aplikácií so znalosťou machine learning algoritmov (Python, R a iných.), ktorí nasadzujú modely do aplikácií s cieľom vytvárania produktov založených na dátach;
- Programátori a ďalší IKT špecialisti zodpovedajú za implementáciu riešenia tak po stránke softvéru aj hardvéru;
- Konzultant so znalosťami GDPR.

#### Organizácia

Chatbot riešenie je možné po softvérovej stránke prideliť iným dodávateľom (cloudové riešenia, externí dodávatelia). Obsah chatbota je potrebné vytvoriť. Firma ako vlastník znalostí a procesov je najkompetentnejšia na tvorbu obsahu chatbota. Tvorba dát nie je zložitá, ale môže byť zdĺhavá. Externalizácia tvorby tréningových dát chatbota je možná,

je však na zvážení či náklady na tvorbu a prípadné zmeny nepresiahnu náklady na vlastných zamestnancov.

## PRÍNOSY A RIZIKÁ

### Kvalitatívne prínosy

- Zvýšenie kapacity obslužených zákazníkov;
- Možné zvýšenie kvality, vďaka odstráneniu ľudského faktora;
- Rýchlejšie obsluženie zákazníka;
- Dostupnosť 24/7.

### Kvantitatívne prínosy

- Úspory;
- Náklady na zamestnancov sú oveľa vyššie ako na chatbot riešenie. Úspory sú hlavne na menšom počte zamestnancov call centier;
- Vyššie výnosy;
- Relatívne skorý ROI. V prípade vlastného riešenia pomocou napr. open-source frameworkov je ROI trochu dlhší. V prípade cloudových riešení sa platí za každú reláciu, čiže náklady sú platené postupne (samozrejme je potrebné vytvoriť znalosti chatbota v oboch prípadoch);
- V prípade používania on-premise ASR/TTS riešenia je ROI neskorší. Najväčšie náklady sú na zakúpenie alebo vývoj vlastného ASR/TTS riešenia, zakúpenie vlastných serverov s hardvérovými akcelerátormi a aj operačné náklady na správu týchto serverov.

### Riziká

- Najväčším rizikom je zlé vytvorenie datasetu hlavne pre DM. V takom prípade môže byť komunikácia s chatbotom chaotická a akceptácia používateľmi nízka, až žiadna;
- V prípade nízkej akceptácie používateľmi je pravdepodobné neskoršie ROI, ak vôbec nastane;
- V prípade použitia cloudových riešení je potrebné rátať s malou mierou nedostupnosti služby. Tá je však pri cloudových riešeniach extrémne nízka. V prípade použitia cloudových riešení je nutné anonymizovať dáta pred vstupom do cloudu a následne po výstupe z cloudu ich naspäť rekonštruovať tak, aby boli naplnené požiadavky GDPR, poprípade interných smerníc na dodržiavanie ochrany osobných údajov.

*Tento text je súčasťou dokumentu ANALÝZA A NÁVRH MOŽNOSTÍ VÝSKUMU, VÝVOJA A APLIKÁCIE UMELEJ INTELIGENCIE NA SLOVENSKU - DIELO Č. 2 - MANUÁL PRE FIRMY NA ZAVEDENIE UMELEJ INTELIGENCIE. Dielo bolo vypracované pre Úrad podpredsedu vlády SR pre investície a informatizáciu autorským kolektívom zo Slovenskej technickej univerzity v Bratislave na základe Zmluvy o dielo č. 1024/2019 zo dňa 29. 10. 2019. Počas tvorby tejto štúdie boli jednotlivé výstupy posudzované expertným tímom združeným pod Slovenským centrom pre výskum umelej inteligencie - Slovak.AI, ktorého členom je aj Slovenská technická univerzita v Bratislave. Všetky závery a komentáre v správe odzrkadľujú názory a postoje autorského kolektívu, ktoré sa opierajú o výsledky analýz opísaných v správe a o diskusie s odborníkmi na problematiku umelej inteligencie spolupracujúcimi na tejto správe. Všetky údaje v tomto texte, ak nie je uvedené inak, sú aktuálne k dátumu odovzdania správy.*

© 2019, 2020 Slovenská technická univerzita v Bratislave, Úrad podpredsedu vlády SR pre investície a informatizáciu. Všetky práva vyhradené.