

## Bližší informace k AI chatbotu pro projekt "Chci na vejšku"

Hledáme řešení chatbotu, který pomůže studentům při rozhodování o výběru vysoké školy a studijního oboru. Chatbot musí být integrován s naší databází, podporovat dlouhodobou paměť a umožňovat dynamické zpracování scénářů.

Rádi bychom stihli chatbota vyvinout a otestovat do ledna 2026.

### Funkční požadavky

#### 1. Dlouhodobá paměť:

Chatbot by měl podporovat dlouhodobou paměť, aby si pamatoval informace z předchozích interakcí uživatelů. To umožní chatbotu poskytovat personalizované odpovědi na základě dřívějších informací (např. preference, odpovědi na kvízy, dříve probírané programy).

**Poznámka:** Kvízy nebudou součástí chatbotu ale součástí širší aplikace - tyto interakce by se tedy také měly ukládat a chatbot by s nimi měl pracovat.

#### 2. Dynamické zpracování scénářů:

Při zahájení chatu by měl chatbot zobrazit uvítací zprávu, která popisuje jeho schopnosti (např. pomůže studentům vybrat studijní obor, odpovědět na otázky o vysokých školách apod.). Úvodní zprávu dodáme. Předpokládáme, že se studenti budou nejčastěji ptát na:

- pomoc s rozhodováním, na jakou vysokou jít, jakou fakultu nebo si vybrat (zde dodáme 3 scénáře připravené kariérovými poradci - scénáře bychom časem chtěli rozšiřovat o další). Chatbot by měl dynamicky s nimi tyto scénáře procházet a na závěr jim pomoci zúžit výběr.
- obecné informace jako například datum dne otevřených dveří na X fakultě nebo podmínky přijetí na Y obor (informace budou v DB)

Předpokládáme, že v budoucnu budeme chtít scénáře rozšiřovat / přidávat.

#### 3. Rychlost a kvalita

Potřebujeme, aby chatbot poskytoval kvalitní výstupy, které skutečně pomohou studentům. Jinak to nedává smysl. Odpovědi musí být rychlé díky naší cílovce.

#### 4. Web Search:

Možná povolíme chatbotu hledat na webu informace, které nejsou k dispozici v naší databázi. Pokud bude tato funkce implementována, chatbot by měl prioritně čerpat z oficiálních zdrojů (např. weby vysokých škol, oficiální vzdělávací portály).

**Poznámka:** Nejsme si jisti, zda web search bude nezbytný. Zvažme náročnost a nákladnost tohoto.

#### 5. Podpora jazyků:

Stránky budou v CZ a SK. Tím pádem i chatbot by měl komunikovat v jednom z těchto jazyků. Potřebujeme pořešit, aby úvodní zpráva od chatbota byla v patřičném jazyce. Potom už bot odpovídá v jazyce dotazujícího.

#### 6. **Infrastruktura:**

Zbytek aplikace bude hostován na Azure a u boťa bychom preferovali, aby tomu tak bylo také. Nejedná se ale o striktní podmínku - pokud by dávala větší smysl jiná varianta, jsme otevřeni i té.

#### 7. **Zpětná vazba od uživatelů:**

Po konverzaci by měl chatbot požádat o zpětnou vazbu, aby zjistil, zda byly odpovědi užitečné pro studenta.

#### 8. **Prioritizace zdrojů dat:**

Chatbot by měl dávat prioritu datům z naší DB. Pokud informace v databázi nejsou, může použít vlastní knowledge base, případně web search.

#### 9. **LLM a spotřeba tokenů:**

Co se týče modelů, preferujeme OpenAI nebo Claude, zejména kvůli kvalitnímu generování textu v češtině, které máme ověřené.

Naším cílem je, aby jedna konverzace se studentem v roli kariérového poradce netrvala déle než 10-15 minut, a to z důvodu udržení pozornosti. Student se pak může doptávat i na konkrétní informace z naší databáze (například kdy je den otevřených dveří na určité fakultě, nebo aby si mohl porovnat školy, fakulty či obory). Předpokládáme, že se student bude k chatbotovi opakovaně vracet, zejména při rozhodování o kariéře a volbě školy.

Při vývoji chceme klást důraz na flexibilitu řešení a zároveň optimalizaci nákladů - například využitím cachování, pokud není nastaveno automaticky, případně i dalších mechanismů. V rámci odhadů popřemýšlejme, zda u delších konverzací nedosáhneme maximální velikosti cache a co s tím. Celkově bychom k řešení měli přistupovat tak, aby zůstalo dlouhodobě nákladově efektivní (ale ne na úkor kvality).

#### 10. **Analytika:**

Potřebujeme analytiku na uživatelskou interakci, ideálně přes GA4:

- **Míra konverze:** Kolik lidí začne konverzaci, ale zavře ji před jejím dokončením?
- **Úspěšnost odpovědí:** Měření úspěšnosti interakcí na základě zpětné vazby od uživatelů.

### **Odhad návštěvnosti**

- Rok 1: Očekáváme 320 platících uživatelů.
  - Prvních několik měsíců bude chatbot zdarma. Během těchto měsíců musíme nastavit max. měsíční spend. Po dosažení tohoto limitu se služba na daný měsíc zastaví. Také budeme muset nastavit guardrail, aby to studenti nepoužívali namísto chatGPT pro standardní dotazy, resp. popřemýšlejme o tomto.
- Rok 2: Cíl 1 000 platících uživatelů.
- Rok 3: Cíl 2 000 platících uživatelů.

**Co potřebujeme:**

- Detailní návrh řešení
- Odhad input a output tokenů na 1 uživatele na měsíc a kalkulaci nákladů + input tokeny pro dlouhodobou paměť
- Odhad fixních a variabilních nákladů při např. 1 000 uživatelích
- Nacenění vývoje