**Empowering OGD-users with an AI-powered Chatbot**

|  |  |  |  |
| --- | --- | --- | --- |
| Nutzen | Rahmenbedingungen | Instrument | Design |
| Warum? | Für wen? | Was? | Wie? |
| Was ist das Problem?  Was wollen wir damit erreichen?  Wer verlangt dies? | Wer sind die Kunden?  Arbeitsweise der Kunden?  Woher kommen die Daten?  In welchem Format haben wir die Daten? | Welche Systeme können wir für die Problemlösung verwenden? | Wie soll das Design aussehen?  Was soll das System können? |
| **Problem:**   1. Gewisse User finden nicht die gesuchten Datensätze, weil sie nicht wissen, wie mit der Suche umgehen. 2. Die User stellen mehrheitlich keine Fragen und wissen nicht wie formulieren 3. Keine Kontextsuche bsp. Wetter / Meteo 4. Es gibt keine Treffer, wenn die Wörter falsch eigegeben werden. (Fehlertoleranz, Tippfehler) 5. Trefferliste ist zu lang (nicht klar, ob der relevanteste Eintrag ganz oben ist)   **Ziel:**   * **Hauptziel: User journey zu optimieren, um den gesuchten Datensatz aufzufinden Fokus Problem 1-5)** * Suchanfrage (Prompt) mit LLM aufwerten, um semantisch zu suchen **(Problem c.)** * Prompt und Suchergebnisse analysieren lassen und Empfehlungen generieren **(Problem a. & b.)** * Spezifische Lösung vorschlagen für zu kurze und zu lange Trefferliste **(Problem d. & e.)**   **Lösungsansätze:**   * Bessere Fragen vorschlagen * Ursprüngliche Quelle soll angezeigt werden * Hilfe, wie die Fragen verfeinert werden können. Ähnlich wie bei Google. * Technisch: App nativ auf Webseite anbinden   **Problemstellungen (kein Fokus)**   1. *Die Anzahl der Schritte bis zum Datensatz ist relativ lang* 2. *User verwenden oder übersehen die Filter (nicht intuitiv)* 3. *Es ist nicht gleich ersichtlich, ob das angezeigte Suchresultat, dass ist was ich gesucht habe* 4. *User wissen nicht, dass es noch weitere Daten zu einem Thema gibt, die sie auf einer anderen Plattform suchen müssen* 5. *Es gibt Datenquellen die auf anderen Plattform liegen, die über opendata nicht zugänglich sind (Vernetzung)* 6. *Für User unklar wann und wieso keine Resultate angezeigt werden* 7. *Metadaten Qualität* | **User:**   * Verwaltung * Data Journalisten * Forschende * Firmen * Hobby   **Typische Suchanfrage:**   * Stichwörter * Bsp. Ich suche nach relativen Schattierungen nach EMS * Was sind die typischen Suchanfragen auf Opendata.swiss   **Datenquelle:**   * Opendata.swiss CKAN API [API guide — CKAN 2.9.10 documentation](https://docs.ckan.org/en/2.9/api/#making-an-api-request) * Team AI 2: [opendata.swiss API nutzen — Handbuch Opendata.swiss 1.0 Dokumentation](https://handbook.opendata.swiss/de/content/nutzen/api-nutzen.html) | * Vorhandene Suche (Index basierte) verwenden   **Team AI 1: Suche verbessern (synonyme Suche erlauben)**   1. Erhalten Suchanfrage von Webseite 2. Werten Prompt mit OpenAI API auf 3. Müssen auf CKAN API zugreifen 4. Liefern Prompts und Trefferliste an Team AI 2 5. Muss sich Testfälle überlegen, anhand derer geprüft werden kann, ob die Suche besser (semantischer) ist   **Team AI 2: Suchtreffer analysieren und Empfehlungen generieren**   * Prompts und Trefferliste von Team AI 1 analysieren * Empfehlungen für User generieren. Mindestens für zu kurze und zu lange Trefferlist     **Systeme**   * OpenAI * Streamlit * Llangchain   **Team UX**   * Empfehlung für UX/UI und User Journey. * Integration: Wie taucht es auf, wie Tonalität/Persona der Empfehlung * Was passiert, wenn der User der Empfehlung folgt * Wie messen wir (Metriken) das, dass sich die Customer Journey verbessert hat * Wie können wir die Nutzung des AI-Assistenten promoten * Wie könnten wir User-Feedback integrieren * Können wir aus der User-Historie lernen und diese bei einem weiteren Website Besuch berücksichtigen * Datenschutz (Website Cookies) | **User Journey**   * 1. User gibt einen Suchbegriff oder Suchanfrage (Satz) ein * 2. technisch: LLM wertet Prompt semantisch auf und wir lösen eine Suche auf der API aus. * 3. technisch: Original-Prompt + verbesserter Prompt + Suchergebnisse gehen ein zweites Mal an ein LLM. Das LLM analysiert und generiert Vorschläge für die Nutzer, um seine Suche zu verbessern * 4. User bekommt Trefferliste und Empfehlung   **Crazy shit**   * Code Generierung * Starter code * Datenqualitäts-Checker * Vorschlag, mit was der Datensatz verbunden werden könnte * Visualisierungsvorschlag der Daten |
|  |  |  |  |