

Innovative Ideas Seminar - Jasper Slowik

Einleitung

Mein Name ist Jasper Slowik und ich studiere Wirtschaftsinformatik im Bachelor. Besonders faszinieren mich die Bereiche Machine Learning und Künstliche Intelligenz. Durch Coursera-Kurse zu Convolutional Neural Networks (CNNs) und allgemeinen Neuronalen Netzen konnte ich mir erste Kenntnisse aneignen. Mit dem Flutter-Framework habe ich ebenfalls bereits Erfahrungen gesammelt, was mir bei der geplanten Nutzeroberfläche zugutekommt. Meine Python- und TensorFlow-Erfahrung decken Grundlagen für die KI-Entwicklung ab.

Aktueller Stand

Künstliche Intelligenz (KI) hat riesiges Potenzial, das in vielen Bereichen noch ungenutzt ist. Unternehmen verlieren Zeit und Kunden, wenn Informationen schwer zu finden sind oder Interaktionen unpersönlich wirken. Meine Idee ist es, ein intuitives Werkzeug zu erstellen, das spezialisierte KI-Lösungen auf Basis bereitgestellter Datendokumente erstellt. Dieses Werkzeug ist insbesondere für den ersten Kundenkontakt gedacht. Das Besondere an diesem Werkzeug ist, dass es keinerlei technische Vorkenntnisse erfordert und die Nutzung von KI so einfach wie einen File-Upload macht.

Die erstellten KI-Modelle sollen über reine FAQ-Beantwortung hinausgehen und eine aktive, individuelle Kundenberatung unterstützen. Zum Beispiel könnten sie auf ‚Welches Fahrrad passt zu mir?‘ mit Nachfragen zu Budget oder Einsatzort reagieren und präzise, produktspezifische Empfehlung geben. Sie entsprächen damit einem virtuellen Berater, der über typische Chatbot-Funktionen hinausgeht.

Ergänzend lässt sich ebenso ein Gamification-Ansatz integrieren, etwa in Form eines Quizzes wie ‚Finde dein perfektes Fahrrad in 3 Schritten‘, das Kunden spielerisch durch Budget, Einsatzort und Präferenzen führt. Dies würde Kunden spielerisch einbinden, die Kundenbindung stärken und könnte damit letztlich auch die Kaufwahrscheinlichkeit erhöhen.

Im Prinzip laden Nutzer in der App/Weboberfläche ihre Dokumente und Informationen hoch, wählen gewünschte Parameter und die Aufgabe, die das Modell erfüllen soll (FAQ, Kundenberatung, Gamification-Ansatz). Mit einem Klick ist das Modell erstellt und mit einem Link abrufbar.

Technische Umsetzung

- Die Nutzeroberfläche wird mit Flutter umgesetzt, ist plattformübergreifend nutzbar und soll optisch und technisch ansprechend gestaltet sein.
- Als Basis dienen sog. distilled Models, wie TinyLLama (Open Source), die leistungsstark sind und gleichzeitig Ressourcen schonen.
- Das Tuning basiert auf dem RAG-Ansatz, um hochgeladene Dokumente als erweiterten Kontext zu nutzen, ergänzt durch Few-Shot-Learning und Parameter-Tuning.

- Das Modell wird über eine einfache Chat-Oberfläche bereitgestellt und wird über einen Link zugänglich gemacht.
- Noch offen sind Details zu Speichieranforderungen, Rechenleistung und der sicheren Bereitstellung der Nutzerdaten.

Geplante Schritte im Seminar

1. Analyse der Modellanpassung: Untersuchung, wie Large Language Models (LLMs) an spezifische Daten angepasst werden können, um aktive Kundenberatung und ggf. Gamification zu ermöglichen. Dazu gehören Datenaufbereitung (Abruf, Bereinigung, Vektorisierung), Auswahl geeigneter Inferenzmethoden und Modelloptimierung. Methoden wie Prompt-Engineering und Retrieval-Augmented-Generation (RAG) werden analysiert, um die besten Ansätze zu ermitteln.
2. Entwicklung eines Integrationsprozesses: Erstellung eines automatisierten Workflows, um hochgeladene Daten und Dateien effizient in ein bestehendes LLM zu integrieren. Der Fokus liegt auf einer technisch robusten Umsetzung, etwa durch Vektor-Datenbanken und automatische Datenverknüpfung, um reibungslose Funktionalität zu gewährleisten.
3. Erste Prototypen und Tests: Entwicklung von Prototypen, um den Prozess zu testen. Ziel ist es zu validieren, ob die Modelle aktive Kundenberatung (z.B. gezielte Nachfragen, Empfehlungen) und optional Gamification-Funktionen korrekt umsetzen sowie Schwachstellen und Optimierungspotenziale zu erkennen.
4. Integration einer benutzerfreundlichen grafischen Oberfläche: Implementierung einer intuitiven Oberfläche in Flutter für plattformübergreifende Nutzung. Nutzer können Daten hochladen, Parameter für Aufgaben wie Kundenberatung oder Gamification definieren und so ein bestehendes Modell anpassen, ohne technisches Vorwissen zu benötigen.
5. Bereitstellung des Modells: Bereitstellung als Web-Anwendung, die per Link direkt nutzbar ist oder über APIs in Unternehmenssysteme integriert werden kann. Dies ermöglicht schnelle Einsatzbereitschaft sowie Flexibilität.

Ziele für das Seminar

Im Seminar soll ein funktionales Tool entwickelt werden, mit dem Nutzer ohne KI-Erfahrung bestehende Large Language Models (LLMs) an ihre Bedürfnisse anpassen können. Die Entwicklung gliedert sich in folgende Schritte:

1. Technische Umsetzung: Umsetzung des zuvor beschriebenen Plans. Der Fokus liegt auf einer robusten und intuitiven Lösung, die eine einfache Modellkonfiguration ermöglicht.
2. Validierung der Modellfunktionalität: Nach Fertigstellung des Prototyps wird geprüft, ob die angepassten Modelle die versprochenen Aufgaben (z. B. präzise Antworten, kontextgerechte Ergebnisse) zuverlässig erfüllen. Ziel ist es, Schwachstellen in der Funktionalität aufzudecken und sicherzustellen, dass die Modelle ihren Zweck erfüllen.
3. Analyse weiterführender Aspekte: Sobald die Grundfunktionalität validiert ist, werden folgende Fragen untersucht:

- **Praxisnutzen:** Ist das Tool in realen Unternehmensszenarien einsetzbar? Hierzu werden Aspekte wie Stabilität, Praktikabilität und realer Nutzen untersucht.
- **Wirtschaftlicher Nutzen:** Bietet es messbare Vorteile, wie Kostensenkung, Effizienzsteigerungen oder wäre ein Umsatzwachstum erwartbar?
- **Kundenerfahrung:** Verbessern die eingesetzten Modelle und optionale Gamification-Elemente die Kundenzufriedenheit, Interaktionen und Bindung?

Innovation und Abgrenzung zu bestehenden Lösungen

Der Markt kennt bereits Anbieter wie Zendesk, Tidio, Freshworks, Intercom und Drift, die KI-Chatbots für Kundenservice anbieten. Meine Lösung unterscheidet sich jedoch durch innovative Ansätze, die sie klar abgrenzen:

1. **Sofortige Einsatzbereitschaft:** Durch die Bereitstellung einer benutzerfreundlichen No-Code-Oberfläche können Unternehmen Ihre Chatbot-Lösung ohne zusätzliche Schulungen oder technische Vorabsprachen, wie bei Drift, sofort implementieren.
2. **No-Code KI-Anpassung:** Im Gegensatz zu Plattformen wie Zendesk oder Intercom, die technisches Know-how für die Integration voraussetzen, ermöglicht Ihr Ansatz die Anpassung von Large Language Models (LLMs) per File-Upload und Parameterauswahl, automatisiert durch Retrieval-Augmented Generation (RAG) und Few-Shot-Learning.
3. **Dialogbasierte Kundenberatung:** Während Tools wie Tidio oder Freshworks oft auf statische FAQ-Antworten beschränkt sind, nutzt mein Lösung RAG, um kontextbezogene Nachfragen und produktspezifische Empfehlungen zu generieren, was zu einer personalisierten Kundenberatung führt.
4. **Gamification-Integration:** Die Einbindung spielerischer Elemente wie „3 Schritte zu deiner Lösung“ fördert die Kundenbindung und steigert die Conversion-Raten, ein Feature, das bei anderen Lösungen fehlt.
5. **Ressourcenschonende Modelle:** Durch den Einsatz von destillierten Modellen, wie TinyLLama, reduzieren Sie Rechenaufwand und Kosten, ohne die Leistung zu beeinträchtigen, was einen Vorteil gegenüber ressourcen- und kostenintensiven Lösungen darstellt.

Abschluss und Formalia

Mit dieser Bewerbung möchte ich am Seminar teilnehmen. Ich freue mich auf eine mögliche Zusammenarbeit!