

GPT-4 und Alternativen in der Praxis: Qualität, Implementierung, Kosten und Datenschutz

Succeet 2023 Intensive Workshop

Paul Simmering
Senior Data Scientist
paul.simmering@teamq.de



Viele Marktforschende haben bereits mit ChatGPT experimentiert und vielversprechende Resultate erzielt. Doch wie gelingt die Übertragung in die Praxis? In der Praxis zählen Genauigkeit, IT-Implementierung, Kosten und Datenschutz. Anhand dieser Aspekte liefert Ihnen dieser Workshop Fakten und Entscheidungskriterien für den erfolgreichen Einsatz von Sprachmodellen. Wir nehmen Projektbeispiele in den Blick, in denen Sprachmodelle zur Analyse von Umfrageantworten, Produktbewertungen und Social Media Beiträgen genutzt werden. Neben den kommerziellen Modellen GPT-3.5 und GPT-4 von OpenAI betrachten wir auch Open Source Modelle wie Llama 2 von Meta, sowie die Möglichkeit, Modelle mit eigenen Daten zu optimieren. Vorkenntnisse sind nicht erforderlich, allerdings ist es von Vorteil, wenn Sie bereits Erfahrung mit ChatGPT oder einem ähnlichen Modell gemacht haben.

Agenda

1. Einleitung (5 Minuten)
2. Arbeit anhand des Fallbeispiels (25 Minuten)
 - a. Zielsetzung und Anforderungen
 - b. Lösungsfindung
3. Besprechung (15 Minuten)

Fragebogen für die Entwicklung eines Projekts mit großen Sprachmodellen

Die folgenden Fragen helfen dabei ein fiktives oder reales Projekt mit großen Sprachmodellen wie GPT-4 oder Llama 2 zu planen. Im Workshop gehen wir gemeinsam durch ein Fallbeispiel. Sie können mit einem eigenen Beispiel ebenfalls die Fragen beantworten. Der ausgefüllte Fragebogen kann als Basis für eine Empfehlung für ein Vorgehen und eine Architektur dienen

Ziel

1. Was soll das Modell tun?

2. Wie wird die Aufgabe bisher erledigt?

- Es ist eine neue Aufgabe
- Manuell
- Durch eine vereinfachte Methode, nämlich _____
- Mit einem anderen Modell, nämlich _____

3. Welche Daten soll das Modell zur Bearbeitung der Aufgabe nutzen?
Bitte alle zutreffenden Optionen ankreuzen.

- Umfrageantworten
- Kundenbewertungen
- Social Media Posts
- Reporte, Newsartikel oder andere Texte
- Webseiten
- Andere Daten: _____
- Das Modell soll nur sein bisheriges Wissen nutzen

4. Wie soll der Output aussehen?

- Tabelle
- Text
- Chat
- Audio
- Bild
- Video
- Anderes: _____

5. Warum soll die Aufgabe mit einem großen Sprachmodell gelöst werden?
Bitte alle zutreffenden Optionen ankreuzen.

- Kosten einsparen
- Schneller oder mehr Daten verarbeiten
- Höhere Qualität
- KI ausprobieren
- Anderer Grund: _____

Qualität

6. Was macht die erfolgreiche Bearbeitung der Aufgabe aus?

- Hohe Genauigkeit anhand der Metrik: _____
- Ich kann es nur umschreiben, es geht um: _____
- Keine Vorstellungen

7. Gibt es Beispiel Input-Output Paare zur Evaluation?

- Ja, in tabellarischer Form
- Ja, aber nicht als Tabelle
- Nein, aber können hergestellt werden
- Nein, und Herstellung ist nicht absehbar

8. Kann ChatGPT die Aufgabe mit Beispiel-Inputs lösen?

- Ja, auf Anhieb
- Ja, aber die zu bearbeitenden Daten passen nicht in den Chat
- Manchmal, aber es macht Fehler
- Nein, gar nicht

9. In welchen Sprachen liegen die Inputs vor?

Bitte alle zutreffenden Optionen ankreuzen.

- Englisch
- Deutsch
- Andere Sprache, nämlich _____

Skalierung

10. Wie lange soll das Modell genutzt werden?

- Einmaliges Projekt, ad-hoc
- Langfristiges Projekt
- Unklar

11. Wie häufig wird das Modell im ersten Jahr voraussichtlich aufgerufen?

Ein Aufruf ist ein verarbeiteter Text, oder eine Frage im Chat.

- Weniger als 10.000 Aufrufe
- Mehr als 10.000 Aufrufe
- Mehr als 100.000 Aufrufe
- Mehr als 10 Millionen Aufrufe

Kompetenzen

12. Ist Personal mit Programmierkenntnissen verfügbar?

- Ja, Profi
- Ja, ein bisschen
- Nein, aber Interesse mit No-Code Lösungen zu arbeiten
- Nein, Implementierung soll von Dienstleister übernommen werden

13. Hat Ihr Unternehmen eine Cloud?

- Eigenes Rechenzentrum
- Amazon Web Services
- Microsoft Azure
- Google Cloud Platform
- Andere Cloud: _____

Datenschutz

14. Welche Attribute treffen auf die Daten zu?

Bitte alle zutreffenden Optionen ankreuzen.

- Öffentlich
- Personenbezogen auf EU-Bürger
- Medizinisch oder anderweitig sensibel

15. Wo sollen diese Daten verarbeitet werden?

- Nur auf eigenen Server
- Deutschland
- Europa
- USA
- Andere Anforderung: _____

Falls Sie sich weiter zu Ihrem Projekt oder anderen KI-Themen austauschen möchten, kontaktieren Sie mich gerne per E-Mail. Sie können gerne Fotos Ihres ausgefüllten Fragebogens anhängen.

Paul Simmering
Senior Data Scientist
paul.simmering@teamq.de

