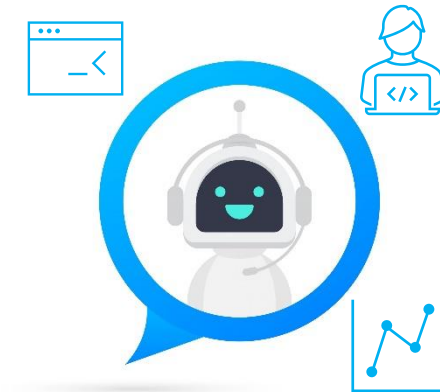


# Benchmarking eines Chatbots für die Produktion | Benchmarking of a chatbot for production

## Ausgangssituation

In modernen Produktionsumgebungen werden Menschen zunehmend mit anspruchsvollen Aufgaben konfrontiert, die Fähigkeiten wie das Navigieren in großen Informationsbeständen oder die Bedienung komplexer Maschinen erfordern. Um dieser Herausforderung zu begegnen, wird im Rahmen eines Forschungsprojekts ein modularer Chatbot entwickelt, der Menschen bei verschiedenen Produktionsaufgaben unterstützt, indem er Daten abrufen und Aktionen ausführt. Um den Chatbot in einer realen Produktionsumgebung zu bewerten, ist eine strukturierte Testmethodik erforderlich.



## Zielsetzung

Ziel dieser sehr praktischen Arbeit ist es, einen bestehenden Chatbot hinsichtlich seiner Eignung für den Einsatz in Produktionsumgebungen zu evaluieren. Die notwendigen Schritte zur Erstellung von Testanfragen, Auswertung der Antworten und Analyse der Ergebnisse wurden bereits theoretisch beschrieben und sollen nun implementiert werden. Dies umfasst das Entwickeln automatisierter Testing-Tools, das Durchführen von Umfragen und die Auswertung der erhobenen Daten.

Relevante Forschungsfragen sind:

- Wie kann das Testing von Chatbots weitestgehend automatisiert werden?
- Wie muss eine Tool-Bibliothek für das Testing aussehen?
- Wie gut ist der aktuelle Chatbot?

Bei sehr guten Ergebnissen besteht die Möglichkeit, als Co-Autor an einer wissenschaftlichen Veröffentlichung mitzuwirken.

## Anforderungsprofil

- hohes Engagement, Interesse an generativer KI und LLMs wie z.B. ChatGPT
- selbständige, methodische und strukturierte Arbeitsweise
- sehr gute Kenntnisse in Deutsch oder Englisch
- Programmiererfahrung in Python

Ich freue mich auf Ihre Bewerbung, idealerweise mit einem kurzen Lebenslauf und einem aktuellen Notenauszug. Die Arbeit kann ab sofort begonnen werden.

----- English version -----

## **Initial situation**

In modern production environments, people are increasingly confronted with demanding tasks that require skills such as navigating large amounts of information or operating complex machines. To meet this challenge, a modular chatbot is being developed as part of a research project that supports people in various production tasks by retrieving data and performing actions. In order to evaluate the chatbot in a real production environment, a structured test methodology is required.

## **Objective**

The aim of this very practical work is to evaluate an existing chatbot in terms of its suitability for use in production environments. The necessary steps for creating test prompts, evaluating the responses and analysing the results have already been described in theory and are now to be implemented. This includes developing automated testing tools, conducting surveys and analysing the collected data.

Relevant research questions are:

- How can the testing of chatbots be automated as far as possible?
- What should a tool library for testing look like?
- How good is the current chatbot?

For very good results, there is the possibility of co-authoring a scientific publication.

## **Requirements profile**

- high commitment, interest in GenAI and LLMs such as ChatGPT
- independent, methodical, and structured way of working
- very good proficiency in German or English
- programming experience in python

I look forward to receiving your application, ideally with a short CV and a recent transcript of records.

The thesis can be started immediately.

## **Kontakt/Contact**

[M.Sc. Lasse Streibel](mailto:lasse.streibel@iwb.tum.de)

Tel.: 089 / 289 15497

[lasse.streibel@iwb.tum.de](mailto:lasse.streibel@iwb.tum.de)