

## Die Chatbot-Challenge – Spielend mit KI von der Idee zum Dialogsystem

Ingo Siegert<sup>17</sup>, Stefan Hillmann<sup>27</sup>, Philine Thalia Kowol<sup>37</sup>, Matthias Busch<sup>47</sup>, Jan Nehring<sup>57</sup> und Xenia Klinge<sup>67</sup>

**Abstract:** In der Chatbot-Challenge sollen Gruppen von vier bis fünf Studierenden ein spannendes Escape-Spiel erstellen, das durch einen Dialog mit einem Conversational Interface erlebt wird. Aufgabe ist es, den Spielverlauf zu planen, mittels des Rasa-Frameworks ansprechend umzusetzen und die Ergebnisse in einer Studie zu prüfen. Hierzu muss nicht nur das KI-gestützte Framework erlernt und angewandt werden, sodass der Chatbot Eingaben verstehen und passend beantworten kann, um das Spiel voranzubringen, sondern auch das Design des Spiels und der Interaktionen iterativ mithilfe von Probanden evaluiert werden. Zum Abschluss der Challenge wird aus allen erstellten Bots der Gewinner in einer Nutzungsstudie ermittelt. Die Chatbot-Challenge wird als hybrides, paralleles Master-Modul an der OvG-Universität Magdeburg und der TU Berlin durchgeführt.

**Keywords:** AI-Literacy, Chatbot, Escapespiel, Hochschule

### 1 Einleitung

Der zunehmende Einsatz von KI in Freizeit und Beruf erfordert eine stärkere Förderung im Umgang mit angewandter KI in der Bildung. Es ist wichtig, dass Studierende in der Lage sind, KI-Methoden und -Technologien effektiv zu nutzen und zu verbessern. Dazu gehört auch, KI-Systeme zu entwickeln und zu evaluieren, sowie das Verständnis von deren Grenzen und Möglichkeiten. Wichtig ist hierbei die Entwicklung von Lehrmaterialien, die sich nicht nur auf Grundlagen und technische Hintergründe konzentrieren, sondern insbesondere auch die sichere Verwendung von KI als Werkzeug zum Erreichen bestimmter Ziele vermitteln, sodass Studierende nicht allein die abstrakte Funktionsweise der Algorithmen verstehen, sondern auch, wie und wozu diese in der Praxis verwendbar sind.

Das hier vorgestellte Studienprojekt ist eine auf ein Semester ausgelegte Projektarbeit, in der die Anwendung von KI-Methoden praktisch geübt und in Kombination mit Fähigkeiten des Projektmanagements und dem wissenschaftlichen Evaluieren und Präsentieren gebracht

---

<sup>1</sup> Otto-von-Guericke-Universität Magdeburg, Fachgebiet Mobile Dialogsysteme, Universitätsplatz 2, 39106 Magdeburg, Deutschland siegert@ovgu.de

<sup>2</sup> TU Berlin, Q&U Lab, Ernst-Reuter-Platz 7, 10587 Berlin, Deutschland stefan.hillmann@tu-berlin.de

<sup>3</sup> TU Berlin, Q&U Lab, Ernst-Reuter-Platz 7, 10587 Berlin, Deutschland p.goerzig@tu-berlin.de

<sup>4</sup> Otto-von-Guericke-Universität Magdeburg, Fachgebiet Mobile Dialogsysteme, Universitätsplatz 2, 39106 Magdeburg, Deutschland matthias.busch@ovgu.de

<sup>5</sup> DFKI, Speech and Language Technologies Lab, Alt-Moabit 91c, 10559 Berlin, Deutschland jan.nehring@dfki.de

<sup>6</sup> DFKI, COS, Alt-Moabit 91c, 10559 Berlin, Deutschland xenia.klinge@dfki.de

<sup>7</sup> Alle Autoren haben gleichermaßen beigetragen.

wird. Dies geschieht anhand einer klar umrissenen Zielsetzung: Ein Escape-Spiel mit einem Chat-Interface zu erzeugen, das sich in einer abschließenden Challenge gegen andere Spiele in Bezug auf Spielspaß und Qualität der technischen Realisierung durchsetzen kann. Die Challenge wird als hybrides, paralleles Lehr-Modul an der Otto-von-Guericke-Universität Magdeburg und der Technischen Universität Berlin angeboten.

## 2 Die Chatbot-Challenge

Die Aufgabe ist, mithilfe des KI-gestützten Chatbot-Frameworks Rasa ein Escape-Room-Spiel zu entwickeln. In einem Escape-Room müssen die Spielenden eine Reihe von Rätseln lösen, um sich aus einem Raum zu befreien. In der Chatbot-Challenge interagieren die Spielenden mit dem Spiel lediglich über Textnachrichten. Sie schreiben beispielsweise "Öffne die Tür", worauf das Spiel antwortet "Die Tür ist verschlossen". Das Setting eines Escape-Spiels ist allseits bekannt, bietet damit auch internationalen Studierenden einen Zugang, erlaubt auch kurze Spiele ohne limitiert zu wirken und erlaubt trotzdem viel Freiheit in der Entwicklung. Ergebnis des Moduls sollte ein bedienbarer Chatbot mit spielbarer und konsistenter Story sein. Im Fokus steht hier aber vor allem die ganzheitliche Erstellung eines KI-Projekts von der Idee über die Umsetzung bis zur Evaluation im "produktiven Einsatz".

### 2.1 Inhalte

Die größte Herausforderung bei der Erstellung des Chatbots ist es, dass die Menge der möglichen Texteingaben nicht im Voraus bekannt ist. Es lassen sich nicht alle Intents (Handlungsabsichten<sup>8</sup>) und Entities (Elemente der "Spiel-"Welt<sup>9</sup>), die in den Eingaben vorkommen können, vorhersehen. Daher kommen KI-Methoden zum Einsatz, um die Interaktion mit dem Chatbot mittels NLU (Natural Language Understanding) zu verbessern. Als Plattform für den Chatbot wird Rasa<sup>10</sup> genutzt. Dessen NLU-Komponente basiert auf einem zweistufigen Ansatz. Zunächst werden mit einem Transformer-basierten [Va17, De19] vortrainiertem Sprachmodell (bspw. [Fe22]) die Eingaben der Nutzenden in sog. Embeddings (eine numerische Repräsentation, die es z.B. erlaubt semantische Ähnlichkeit zu messen) überführt. Diese werden dann von einem zweiten Klassifikator verwendet, dem Dual Intent und Entity Classifier (DIET) [Bu20], der für die eigentliche Erkennung der Intents und der Entities trainiert bzw. verwendet wird.

Die Teilnehmenden lernen ein KI-Modell zu trainieren, zu evaluieren und direkt anzuwenden. Sie üben am praktischen Beispiel Escape-Room, wie Daten für das Training eines Klassifikators gesammelt und aufbereitet werden müssen (benötigte Datenmenge, Datenbalance in verschiedenen Klassen) und wie die trainierten Modelle hinsichtlich ihrer Klassifikationsleistung evaluiert werden können (geeignete Maße, Kreuzvalidierung).

Neben der reinen Erkennungsrate sind in der Challenge auch User Experience (UX) und Spielspaß (Game Experience) wichtige Metriken der Evaluation des Chatbots. Nicht nur die Erzeugung des KI-Modells wird erlernt, das den Chatbot befähigt, Eingaben in natürlicher

---

<sup>8</sup>Ein Intent beschreibt das Ziel einer Interaktion, z.B. "Tür öffnen", "Licht anschalten".

<sup>9</sup>Eine Entity bezieht sich auf Personen, Orte oder Objekte, die der Chatbot "kennen" soll, z.B. Tür, Schlüssel.

<sup>10</sup>Ein kostenloses Open-Source-Chatbot-Framework basierend auf maschinellem Lernen.

Sprache semantisch korrekt zu interpretieren, sondern auch, wie die UX eines interaktives Dialogsystem evaluiert werden kann. In dem Studienprojekt wird also die äußerst relevante Verbindung zwischen dem Erzeugen eines KI-gestützten Systems und dessen Nutzung durch und Wirkung auf die Nutzenden hergestellt.

## 2.2 Ablauf und Regeln

Die Veranstaltung ist in mehrere Praxisblöcke, Ergebnis-Vorstellungen sowie Selbstlernphasen aufgeteilt. Die Praxisblöcke umfassen die Vorstellung der Chatbot-Challenge, der Lehrenden und der Regularien. Weiterhin gibt es Blöcke zur Einführung Rasa, zum Trainieren der NLU, zum Planen, Durchführen und Auswerten von Nutzungstudien und zum Deployment im Frontend. Die Ergebnisvorstellungen finden immer als parallele hybride Veranstaltung an der Uni Magdeburg und der TU Berlin statt. Die Termine und erwarteten Ergebnisse sind als entsprechende Meilensteine (MS) vordefiniert und umfassen: (MS1) Vorstellung Gesamtidee, 1. Prototyp, (MS2) 1. Version vollständiger Escape-Room, Konzept Nutzerstudie, (MS3) Escape-Room nach Nutzerstudie, Auswertung Nutzerstudie, (MS4) Auswertung finale Nutzerstudie, Lessons-Learnt. Am Ende schreiben alle Gruppen einen Bericht, der die Ergebnisse der Projektarbeit zusammenfasst.

Damit die Ausgangsbedingungen zwischen den Teams vergleichbar sind und die Aufgabe insgesamt nicht zu groß wird, gibt es klare Regeln, die am Beginn der Veranstaltung kommuniziert werden. Hierzu gehört, dass alle Gruppen Rasa verwenden müssen, dass in den Antworten keine Bilder erlaubt sind und keine externen LLMs (Large Language Models) benutzt werden dürfen. Weiterhin ist der Fragebogen für die UX-Evaluation festgelegt (UEQ+[Sc21]), für die Evaluierung des Spielspaßes sollen die Teams Vorschläge geben.

## 3 Lernziele

Künstliche Intelligenz	Benutzung des Rasa Frameworks Generierung von Trainingsdaten für maschinelles Lernen Evaluation von KI-basierten Klassifikatoren
UX-Design	Design eines textbasierten Spiels Iterative Durchführung von Nutzungstudien
Projektmanagement	Selbstorganisation in Gruppen, Management Tools Selbstständige Erarbeitung von Teilzielen
Evaluierung der Interaktion	User Experience und Game Experience Fragebögen

Tab. 1: Lernziele der Lehrveranstaltung

In der Challenge haben die Studierenden Gelegenheit, sich mit dem Thema KI in einem anwendungsorientierten Rahmen auseinanderzusetzen, der von ihnen einerseits ein Verständnis der zugrunde liegenden Prinzipien verlangt (z.B. beim Trainieren eines NLU-Modells), andererseits aber die Aufmerksamkeit direkt auf das Ergebnis lenkt, das sich in Gesprächen mit relativ schnell iterierbaren Versionen des Chatbots einfach überprüfen lässt. Somit liegt der Fokus auf der bewussten und souveränen Anwendung des KI-gestützten Dialogsystemframeworks, wie sie in der heutigen Softwarelandschaft mehr und mehr erforderlich wird.

Darüber hinaus werden im Gesamtprojektrahmen weitere wichtige Fähigkeiten und Kenntnisse in UX-Design, Projektmanagement und Evaluierung vermittelt (Tab. 1). Durch die Wettbewerbsrahmung setzten sich die Teilnehmenden mit den anderen Lösungen auseinander und identifizierten die Vor- und Nachteile, um diese im eigenen Chatbot zu berücksichtigen.

## 4 Ausblick

Die Challenge hat im April 2023 als universitätsübergreifende Lehrveranstaltung begonnen. Dies ist das erste Mal, dass sie durchgeführt wird, in kommenden Jahren soll sie jedoch in dieser oder angepasster Form weitergeführt werden.

Während die Studierenden in ihrer Gruppe jeweils einen Chatbot entwickeln, wurde durch uns ein Portal entwickelt, in das die Chatbots integriert werden können. Dies erlaubt es, dass alle Teilnehmenden alle Chatbots unter gleichen Bedingungen testen können – und die interessierte Öffentlichkeit sowie Teilnehmende auch zukünftig die Spiele spielen können. Die Challenge ist in vielen Punkten eingeschränkt, damit sich die Studierenden auf die zu bewertenden Aspekte konzentrieren und die abschließende Bewertung der Ergebnisse fair bleibt, da alle dieselben Mittel nutzen. Dennoch bieten sich Möglichkeiten, die Materialien in zukünftigen Challenges anzupassen, insbesondere im Bereich von Textgenerierung. So hat sich bereits in diesem Jahr in den Fragen Studierender ein großes Interesse am Einsatz von ChatGPT gezeigt und auch für Rasa gibt es erste Arbeiten LLMs zu integrieren<sup>11</sup>.

Die bisherigen Erfahrungen mit der diesjährigen Challenge zeigen, dass alle Gruppen in der Lage waren, mit Rasa erste Prototypen zu entwerfen, und NLU-Modelle zu trainieren. Unterschiede zeigen sich, wie die Gruppen den Chatbot im Spiel interpretieren (Erzähler, Mit-/Gegenspieler). Auch zeigen sich hier aber bereits erste Herausforderungen bei der Entwicklung. Der Ansatz von Versuch und Irrtum, der sonst für kleine Demonstratoren ein gangbarer Weg ist, stößt hier schnell an zeitliche Grenzen.

## Literaturverzeichnis

- [Bu20] Bunk, T.; Varshneya, D.; Vlasov, V.; Nichol, A.: , DIET: Lightweight Language Understanding for Dialogue Systems, 2020.
- [De19] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proc. 2019 Conf. of the North American Chap. ACL: Human Lang. Tech. ACL, Minneapolis, Minnesota, S. 4171–4186, Juni 2019.
- [Fe22] Feng, F.; Yang, Y.; Cer, D.; Arivazhagan, N.; Wang, W.: Language-agnostic BERT Sentence Embedding. In: Proc. of the 60th Annual Meeting of ACL. Dublin, Ireland, S. 878–91, 2022.
- [Sc21] Schrepp, M.: Measuring User Experience with Modular Questionnaires. In: 2021 Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS). 2021.
- [Va17] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I.: Attention is All you Need. In: Advances in Neural Information Processing Systems. Jgg. 30. Curran Associates, Inc., 2017.

---

<sup>11</sup>etwa in Projekten wie RasaGPT (<https://github.com/paulpierre/RasaGPT>).