

**Bewerbung für studentischen DINI-Wettbewerb:  
“Mit KI lernen: Intelligente Bildung für eine intelligente Zukunft”**

# **AI-Lab: Evaluation Game**

**KI-gestütztes Spiel zur Evaluierung in der Lehre**

**30.06.2024**

**KONTAKT:**

**StAI Hub - student AI Hub Marburg  
ai.lab@uni-marburg.de**





# ÜBERSICHT

Mit dem Evaluationsspiel entwickeln wir ein KI-basiertes Evaluationssystem für universitäre Lehrveranstaltungen. Es handelt sich um ein 2D-Spiel mit einem integrierten KI-gestützten Fragebogen, der den Evaluationsprozess gamifiziert und interaktiv gestaltet.

Unser Fragebogen besteht aus Skalenfragen und offenen Fragen. Um den Benutzer:innen die Möglichkeit zu geben, ihre Antworten zu verfeinern, nutzen wir ein Large Language Model (LLM), das dialogähnlich auf die Antworten von Benutzer:innen eingeht und gezielt weiterführende Fragen generiert. Dies ermöglicht eine tiefere Reflexion und ein präzises Feedback. Das dialogorientierte Format trägt dazu bei, von den Erfahrungen der Benutzer:innen zu hören, die in dem Fragebogen nicht erfasst wurden.

---

# TEAM

Wir sind ein interdisziplinäres studentisches Team von der Philipps Universität Marburg, das Teil der **stAI Hub** Hochschulgruppe an der Universität Marburg ist. Unser Team besteht aus:

- **Ali Kholmovaia** (B.Sc. Informatik)
- **Tim Ushakov** (B.Sc. Informatik)



## MOTIVATION

Im Wintersemester 2023/2024 haben wir ein studentisches projektorientiertes Modul "AI-Lab" zur Frage "Wie können wir mit KI gut zusammenleben?" erfolgreich durchgeführt. Nach Abschluss des Moduls stellten wir uns die Frage, wie wir das Modul qualitativ so auswerten können, dass möglichst viele Teilnehmer:innen ihre Erfahrungen reflektieren und den Fragebogen ausfüllen.

Da wir selbst Studierende sind, die regelmäßig Fragebögen zur Lehrevaluation ausfüllen müssen, ist uns bewusst, wie schwer es manchmal ist, die eigenen Erfahrungen zu reflektieren und die Reflexion in einen standardisierten Fragebogen zu übertragen. Deshalb überlegten wir uns, wie wir dieses Format aus Sicht der Studierenden verbessern können. Dabei stellten wir uns drei zentrale Fragen:

- 1) Wir können wir die Reflexion der Studierenden anregen?
- 2) Wir können wir mehr Raum für die individuellen Erfahrungen geben, die in einen standardisierten Fragebogen nicht passen würden?
- 3) Und wie kann uns KI dabei helfen?

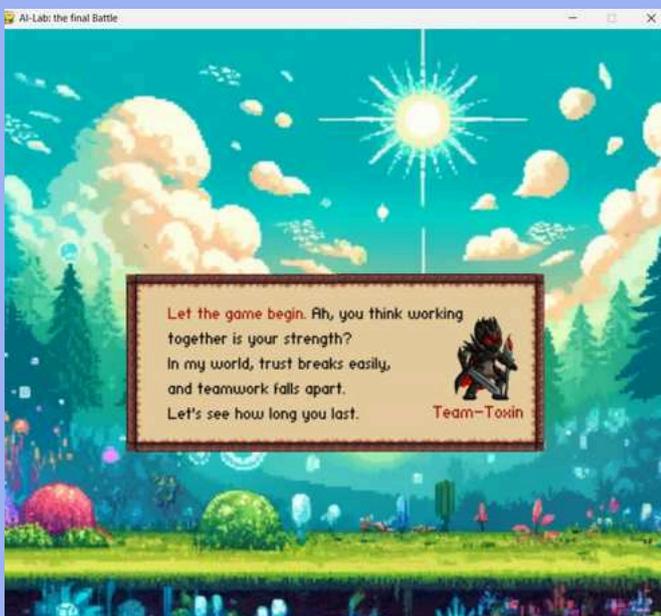
So kamen wir auf die Idee, den Prozess zu gamifizieren und ein einfaches Spiel zu entwickeln. In diesem Spiel werden sowohl vorgefertigte Fragen als auch dynamisch generierte Fragen, die auf den vorherigen Antworten basieren, durch ein Large Language Model (LLM) gestellt.

Dieser Ansatz erschien vielversprechend, weil er die Spieler:innen aktiv in den Bewertungsprozess einbeziehen und zu kohärenten, durchdachten und detaillierten Antworten führen kann.

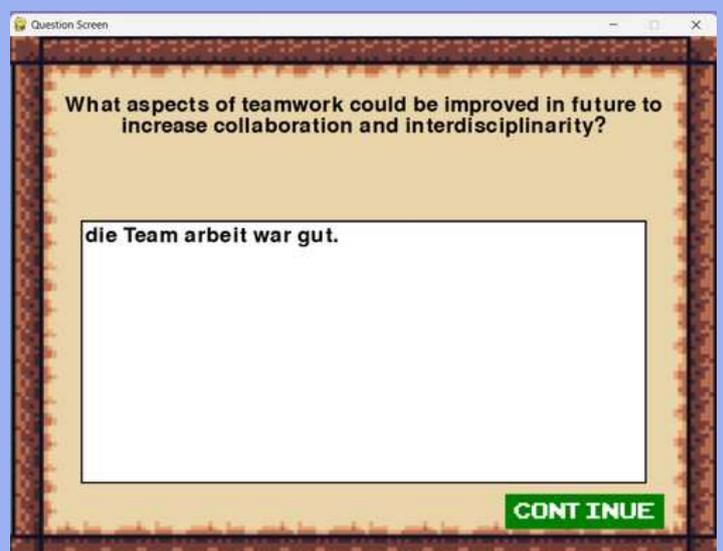
# SPIELSABLAUF

Das Spiel besteht aus drei Level, die jeweils ein spezifisches Thema unseres Moduls behandeln: Teamarbeit, Projektdurchführung und Lernunterstützung. Jedes Level folgt einer einheitlichen Struktur:

1) **Einführungsdialog:** Zu Beginn jedes Levels werden die Spieler:innen durch einen Dialog in das Levelthema eingeführt. Dies schafft den Rahmen und Kontext für die folgenden Fragen und soll die Erfahrung immersiver machen.



2) **Fragerrunde:** Den Spieler:innen werden dann Fragen zum Thema des Levels gestellt. Diese Fragen umfassen sowohl Skalen als auch offene Fragen. Die Spieler:innen geben zunächst ihre Antworten, die dann von einem Large Language Model verarbeitet werden. Das LLM generiert Folgefragen, um die Antworten zu verfeinern und zu vertiefen.

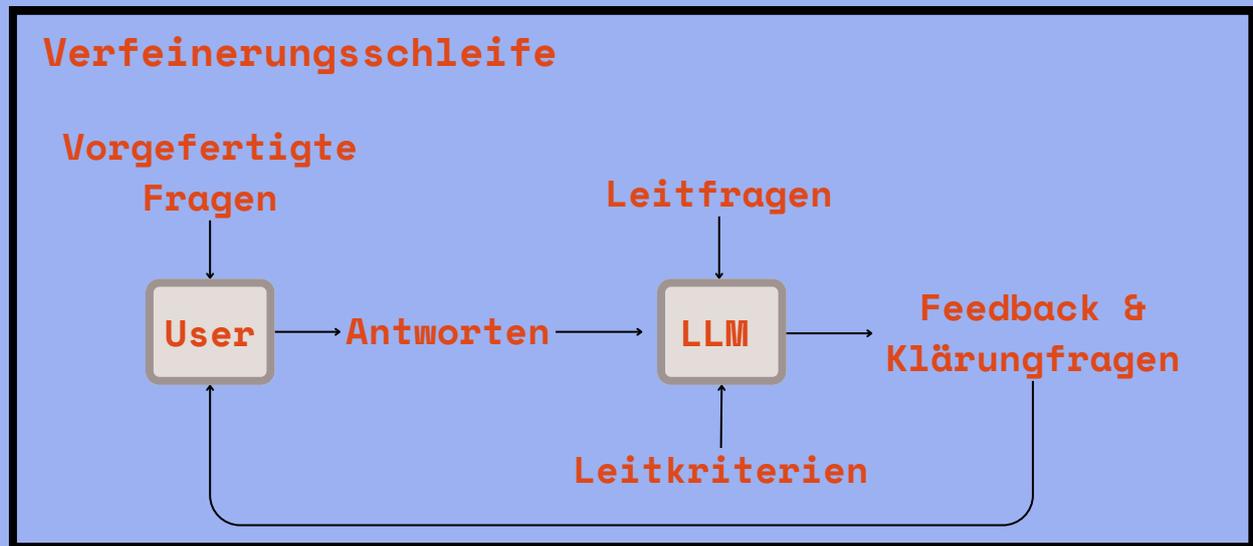


# SPIELSABLAUF

3) **Kampf:** Nach der Fragerunde nehmen die Spieler:innen an einem Kampf mit dem Boss des Levels teil. Der Schwierigkeitsgrad des Kampfes steigt mit jedem Level. Wenn die Spieler:innen den Boss nicht besiegen können oder keine Lust haben, am Kampf teilzunehmen, können sie den Kampf überspringen und nur die Fragen beantworten.



# FRAGERUNDE



In jeder Fragerunde werden den Spieler:innen Fragen zu spezifischen Themen der Evaluation gestellt, die sowohl Skalen- als auch offene Fragen umfassen. Das ermöglicht es, sowohl quantitative als auch qualitative Daten zu sammeln.

Um die Antworten der Spieler:innen zu evaluieren und das Feedback mit zusätzlichen Fragen zur Verfeinerung zu stellen, bilden wir in jeder Fragerunde einen Prompt, der aus mehreren Teilen besteht:

1) Die Spieler:innen geben **ihre Antworten auf die vorgefertigten Fragen**. Diese Antworten werden als Grundlage für die weitere Verarbeitung und Verfeinerung genutzt.

Question Screen

Tasks were fairly and appropriately assigned to team members.

Strongly Disagree Disagree Neutral Agree Strongly Agree

Our team was successful in combining diverse skills and knowledge to achieve our goals.

Strongly Disagree Disagree Neutral Agree Strongly Agree

All team members were actively and equally involved in the work process.

Strongly Disagree Disagree Neutral Agree Strongly Agree

CONTINUE

Question Screen

What aspects of teamwork could be improved in future to increase collaboration and interdisciplinarity?

die Team arbeit war gut.

CONTINUE

# FRAGERUNDE

2) **Leitkriterien:** Das LLM bewertet die Antworten anhand mehrerer Kriterien wie Klarheit, Relevanz, Beschreiben von spezifischen Beispielen und konstruktive Kritik. Jedes Kriterium wird eine Anzahl an Punkten zugewiesen. Diese Kriterien helfen dabei, die Qualität der Antworten zu bestimmen und Bereiche für Verbesserungen zu identifizieren. Die Spieler:innen haben keinen Zugang zu diesen Punktzahlen, um zu vermeiden, dass sie versuchen, unbedingt mehr Punkte zu bekommen, was zu einem Bias in den Antworten führen könnte.

```
Clarity and Coherence: _ points
Relevance: _ points
Specific Examples: _ points
Constructive Criticism: _ points
```

3) **Leitfragen:** Um dem LLM Richtungen zur Fragengenerierung vorzugeben, führen wir in den Prompt auch vorgefertigte für das Thema des Levels relevante Beispielfragen ein.

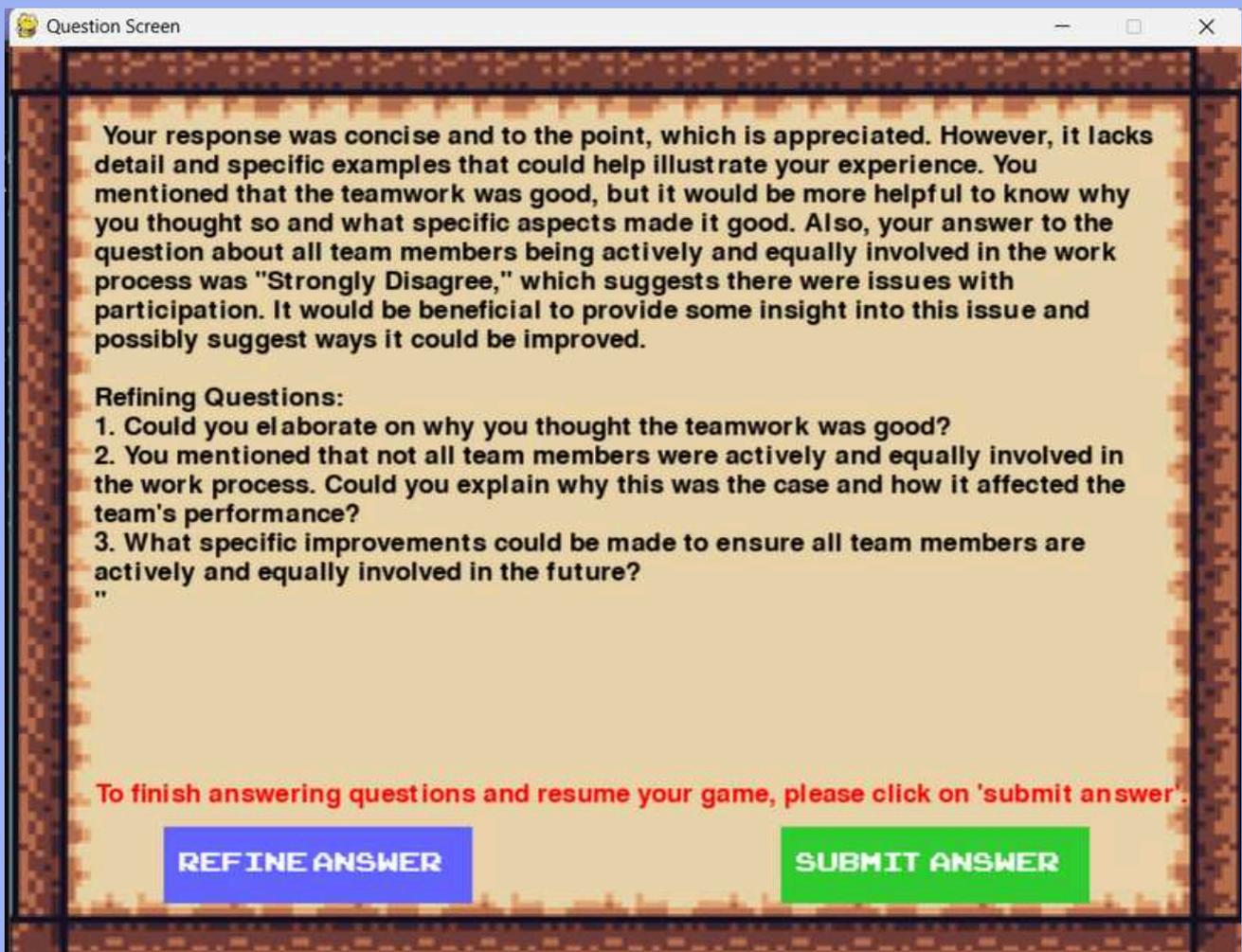
```
def get_extra_questions_teamwork(self):
    return """
- Roles: "Did everyone know what they were supposed to do in the team?"
- Talking: "Was it easy to talk and listen to each other in the team?"
- Solving Problems: "When there was a disagreement, how did the team solve it?"
- Helping Out: "Did everyone do their part in the team's work?"
- Leading: "How well did your team leader or mentor help the team?"
- Goals: "Did everyone agree on what the team was trying to achieve?"
- Getting Better: "Did the team discuss ways to improve for next time?"
- Trusting: "Could you count on everyone in the team?"
- New Ideas: "Was the team open to hearing new ideas?"
```

# FRAGERUNDE

07

Anhand dieses Prompts generiert das LLM ein Feedback zu den Antworten und offene Fragen ("refining questions"), die die Spieler:innen beantworten können.

Nach der Beantwortung der neuen Fragen kann eine weitere Verfeinerungsiteration durchgeführt werden. Spieler:innen können selbst entscheiden, ob sie die Antwort ergänzen oder mit dem Spiel fortfahren möchten.



The screenshot shows a window titled "Question Screen" with a yellow background and a brown border. The text inside is as follows:

Your response was concise and to the point, which is appreciated. However, it lacks detail and specific examples that could help illustrate your experience. You mentioned that the teamwork was good, but it would be more helpful to know why you thought so and what specific aspects made it good. Also, your answer to the question about all team members being actively and equally involved in the work process was "Strongly Disagree," which suggests there were issues with participation. It would be beneficial to provide some insight into this issue and possibly suggest ways it could be improved.

**Refining Questions:**

1. Could you elaborate on why you thought the teamwork was good?
2. You mentioned that not all team members were actively and equally involved in the work process. Could you explain why this was the case and how it affected the team's performance?
3. What specific improvements could be made to ensure all team members are actively and equally involved in the future?

"

To finish answering questions and resume your game, please click on 'submit answer'.

REFINE ANSWER      SUBMIT ANSWER

Unsere Pläne sehen vor, das Spiel im nächsten Semester in den großen Vorlesungen des Fachbereichs Informatik und Mathematik der Universität Marburg zu testen, um unser Evaluierungsspiel mit dem standardisierten Fragebogen vergleichen zu können.

Außerdem wollen wir dafür ein lokal laufendes Open-Source-Modell benutzen, um die Daten der Proband:innen lokal zu speichern.



Wir würden uns sehr freuen, wenn unsere Idee durch die Anerkennung mit dem Preis noch mehr Aufmerksamkeit bekommen würde und wir dadurch noch mehr Impulse für die Verbesserung des Spiels bekommen würden.

Wir freuen uns auf Ihr Feedback und hoffen auf eine positive Rückmeldung!