

## **Protokoll des Workshops: Die „AG Lernerfolg lädt ein“: Bewertung von Prüfungsleistungen im Kontext von KI und andere Fragen rund um das Thema Prüfen.**

Donnerstag, 7. März 2024, 13.00-14.30 Uhr

Moderation: Heide Stiebeler

Teilnehmende aus der AG Lernerfolg: Christina Thunstedt, Rosane Werkhausen, María Jesús García, Tania Tapia Pérez, Kristina Gauss, Hao Zhou, Maddy Stapel, Anne Schlömer

Protokoll: Anne Schlömer

Einleitende Worte: Christina Thunstedt und Heide Stiebeler

### **Dokumentation der Rückmeldungen zu den Arbeitsaufträgen 1 bis 4 aus den Gruppen:**

#### **1. Lesen Sie die Portfolioprfüfungsaufgabe und überlegen Sie: An welchen Stellen ist es sinnvoll, mit KI als Hilfsmittel zu arbeiten? Mit welcher Funktion?**

- Wie war die Aufgabenstellung gemeint? Ist KI hier sinnvoll für den Lernenden oder für Lehrende? Antwort: „Sinnvoll für den Lernerfolg“ ist gemeint – Wichtig ist in diesem Zusammenhang immer in ein Gespräch mit den Studierenden zu treten.
- ChatGPT liefert meist neue Wörter. Was macht man/machen die Studierenden damit? Gehören die Wörter zum Niveau, das gelernt wird? Wie bringt man die neuen Wörter in den aktiven Wortschatz?
- Wir halten KI-Einsatz bei Teil 4 für sinnvoll: Wortschatzarbeit, Grammatik und Überprüfung sowie bei Teil 2. Bei Teil 1 wäre eher google einzusetzen.
- Wir können uns auch bei Teil 3 KI-Einsatz vorstellen: Den Text kann man auch sprechen lassen oder die Aussprache üben, indem man sich den Text erstmal vorlesen lässt.
- KI-Nutzung ist bei Teil 2 sinnvoll und bei Teil 1/5. KI-Nutzung ist also immer dann sinnvoll, wenn es um eigene Ideen, Formulierungen geht.
- Wir haben Bedenken bei KI-Einsatz in Teil 2: Diesen komplett an die KI abzugeben – würde das akzeptiert? Wie kann man es verhindern? Die Aufgabenstellung ist deswegen möglicherweise problematisch. Antwort: Das kann man weder verhindern noch kontrollieren.
- Die Aufgabenstellung ist deswegen möglicherweise problematisch.

- Zu Teil 1/5: Wie kann ich das bewerten? Kreativität kann man nicht bewerten. Antwort: Nur Überprüfbares und Beobachtbares kann bewertet werden, aber die Versprachlichung der (kreativen) Idee kann bewertet werden.

## 2. An welchen Stellen sehen Sie besondere Herausforderungen für die Bewertung?

- Wir könnten uns eine Änderung in der Aufgabenstellung vorstellen: Man könnte die Textproduktion und den Grammatikteil trennen. In der vorliegenden TN-Antwort ist der Blogbeitrag nicht so authentisch, weil die vorgegebene Grammatik abgearbeitet wird. Man sollte die Grammatik separat prüfen (in einem E-Test z.B.). Beim Blog sollte man weniger Leitfragen vorgeben. Eine Schwierigkeit in der Bewertung kann außerdem auftreten, wenn z.B. Grammatikstrukturen von manchen mehrfach benutzt werden, teilweise richtig, teilweise falsch, von anderen werden sie aber bspw. nur einmal richtig benutzt: Wie soll man das fair bewerten?
- Warum werden Grammatikstrukturen überhaupt abgefragt? Man könnte diese auch von ChatGPT machen lassen. Was ist das Ziel der Aufgabe? Warum werden die sprachlichen Strukturen überhaupt vorgegeben? Ein Ergebnis ohne ChatGPT wäre sehr schön für die Beurteilung. Aber wenn mit ChatGPT gearbeitet wird, ist es schwierig zu beurteilen. Wo ist dann die Eigenleistung?
- Wir schlagen das Trennen der eigenen Produktion vom Arbeiten mit ChatGPT vor. Die Aufgabenstellung verführt zu Vermischung von Eigenproduktion und KI. Wir können uns vorstellen, die Aufgabe nicht nur als Hausaufgabe zu geben, sondern Teile (z.B. den Text zu schreiben) in den Unterricht zu verlegen. (Kommentar: Die Aufgabe war als Hausaufgabe in einem Zeitraum von 10-12 Tagen gedacht.)
- Wir fragen uns, ob die Prüfungsform Hausaufgabe/ Portfolioaufgabe als Prüfungsform noch relevant ist? Man müsste Aufgaben anbieten, die nicht mit ChatGPT lösbar sind? Gibt es überhaupt solche Aufgaben, die man ohne ChatGPT machen kann?
- Ich habe Übungsaufgaben, die im Kurs gemacht wurden, mit Hausaufgaben verglichen. Im Ergebnis gab es sehr große Unterschiede. Das führt zu Frustration, weil die Ergebnisse nicht realistisch waren. Am liebsten lasse ich in der Klasse schreiben.
- Ich akzeptiere in der Prüfung keine Ergebnisse, die über dem unterrichteten Level liegen. Dafür gebe ich 0 Punkte und kündige es auch vorher so an.
- Zu Teil 2 und der Reflexion: Wenn nicht alle das gleiche Tool verwenden, kann man nicht gerecht bewerten. Man muss vorher klären, welche Hilfsmittel in welchem Teil von allen verwendet werden dürfen.

Workshop-Tage März 2024: KI meets Fremdsprachenlehre

- Wenn wir explizit „open book“ prüfen, dann heißt das, dass auch Hilfsmittel genutzt werden können. Wenn es erlaubt ist, dann muss man auch damit leben. Vielleicht gibt es eine Zwischenlösung?
- Schön an der Aufgabe ist, dass die Studierenden bei der Aufgabe eigene Interessen berücksichtigen und eigene Ideen entwickeln können. Es wird für jeden Teilnehmer:in eine sehr individuelle Lösung geben. Diesen individuellen Teil könnte man im Unterricht produzieren lassen, um die Bewertung fairer machen zu können.
- Wir haben an der Uni Münster ein Formblatt am Sprachenzentrum, das bei Prüfungen ausgefüllt und unterschrieben werden muss. Darin muss u.a. angegeben werden, wie die Aufgabe gelöst wurde, welche KI benutzt wurde, u.U. auch die Prompteingabe. Im Text müssen die Stellen markiert werden, die über Tools produziert wurden. Wir wollen die Nutzung der KI fördern, aber auch fair beurteilen können.
- KI ist bei uns an der LMU (Studiengang Master DaF, Fach Medien), also im wissenschaftlichen Kontext, Thema im Seminar. Bei Seminararbeiten darf KI benutzt werden, aber bestimmte Textstrukturen kann die KI nicht produzieren. Wir erwarten z.B., dass die Theorie mit Praxisbeispielen kombiniert wird. Das funktioniert nicht mit ChatGPT. Im Unterricht haben wir das gemeinsam ausprobiert und verdeutlicht. Insofern kann ich sicher sein, dass die Seminararbeiten eine Eigenleistung sind. Sprachliche Fehler, die leicht durch Hilfsmittel verbessert werden können, spielen allerdings keine Rolle mehr in der Bewertung. Wir verwenden auch einen Fragebogen und ein Bewertungsmuster.
- Was kann die KI nicht? Was könnte man in der Prüfung abfragen, das nur in Eigenleistung entstehen kann?
- Im Moment kann die KI vielleicht einiges noch nicht, aber bald kann die KI immer mehr.
- Es gibt nicht mehr den Kontrast „eigenständig schreiben“ vs. „KI Benutzung“. Kann man „eigenständig“ überhaupt noch sagen? Eine Leistung, die mit Hilfsmitteln erstellt wurde, ist doch in Teilen auch „eigenständig“. „Eigenständig“ wird zu einem graduellen Begriff.
- Müssen Studierende, die eine Fremdsprache lernen, überhaupt noch einen Text ohne Hilfsmittel produzieren können?

**3. Lesen Sie die Portfolioaufgabe und die Teilnehmer:innenantwort noch einmal und diskutieren Sie mindestens eine der folgenden Fragen:**

- a. Wie würden Sie die einzelnen Aufgabenteile gewichten?**
- b. Nach welchen Kriterien würden Sie die Teilnehmer:innenantworten bewerten?**
- c. Was würden Sie - auch nach Lektüre der Teilnehmer:innenantwort – an der Aufgabenstellung ändern?**

- Zunächst würde man die klassischen Bewertungskriterien zugrunde legen (Vollständigkeit, Korrektheit, Aufbau). Diese werden allerdings redundant bei KI-Erstellung. Kann man überhaupt noch Texte bewerten? Man sollte die Mündlichkeit aufwerten (z.B. Präsentation).
  - Für Teil 3 würde ich die wenigsten Punkte geben, weil an der TUM die Aussprache nicht bewertet wird. Kommentar: Wir dürfen am TUM Sprachenzentrum mündliche Prüfungen aufgrund der Komplexität des Mündlichen nur unter besonderen Bedingungen (u.a. zwei Prüfer:innen) durchführen, deshalb gibt es sie nicht. Audioaufnahmen sind aber möglich. Man kann z.B. Feedback auf die Audioaufnahme geben, aber keine Punkte vergeben.
  - Ich habe gute Erfahrungen mit mündlichen Teilen in Portfolioaufgaben gemacht. Z.B. in der A1 eine Smalltalkaufgabe, in der u.a. die Interaktion bewertet wird, aber z.B. nicht die Aussprache. Ergänzt wird dies durch eine klassische Prüfung (Grammatik etc.).
  - Man kann sich bei mündlichen Leistungen auf nur einen Aspekt konzentrieren; damit grenzt man sich gegenüber einer mündlichen Prüfung klar ab.
  - KI-generierte Texte kann man entlarven, indem man sie mündlich „verteidigt“. Das ist allerdings mit hohem Aufwand verbunden.
  - Teil des Portfolios kann eine Audioaufnahme sein, in der über den Kurs reflektiert wird und Stellung zur eigenen Produktion bezogen wird. In meinem Schwedisch B2-Kurs müssen die Studierenden u.a. sagen, was sie mit Schwedisch vorhaben (was ist das Ziel von dem Kurs, Lernergebnis). Dies führt zur Festigung der Inhalte und es ist eine Winwin-Situation für Prüfende und Studierende, denn der Dozierende erhält so auch Feedback zum Kurs. Wichtig ist, dass die Passung stimmen muss. Man sollte auch nicht um jeden Preis KI „reindrücken“. Die Transparenz im Kurs ist sehr wichtig.
  - Was wird bewertet bei 10 Min. Sprechen? Antwort: Es gibt sehr gute und einfache Kriterien aus der Aufnahmeprüfung für schwedische Universitäten (TISUS-Test), die sehr valide sind. Z.B.: Kann ich dem Redner folgen? Wie ist der Redefluss? Damit kann man sehr gut arbeiten.
  - Eine Idee für Audioaufnahmen ist ein Interview zu verlangen (evtl. auch unter zwei Studierenden). Die Bewertung ist gut möglich. Einen Text nur vorlesen zu lassen, ist nicht sinnvoll.
- 4. Was nehmen Sie für Ihren Unterricht und die Erstellung von Prüfungsaufgaben aus diesem Workshop mit? Notieren Stichwörter auf den ausliegenden Kommunikationskarten.**

Ist aus Zeitmangel entfallen.