



netidee

STIPENDIEN

Advancing Education with AI

Zwischenbericht | Call 19 | Stipendium ID 7326

Lizenz: CC BY-SA

Inhalt

1	Einleitung	3
2	Status	3
2.1	Meilenstein 1 - Literaturrecherche	3
2.2	Meilenstein 2 - Datensammlung	3
2.3	Meilenstein 3 - Umfrage zur Ermittlung von Kriterien für gute Arbeitsblätter	
2.4	Meilenstein 4 - Proof of Concept Programmierung der Plattform	
2.5	Meilenstein 5 - Plattform wird im schulischen Umfeld eingesetzt	
2.6	Planänderung	
3	Zusammenfassung Planaktualisierung	4
3.1	Fazit	3
3.2	Planaktualisierung	

1 Einleitung

Dieser Zwischenbericht dokumentiert den aktuellen Entwicklungsstand der KI-gestützten Bildungsplattform, die im Rahmen meiner Masterarbeit entsteht. Die Plattform ermöglicht Lehrkräften die automatisierte Erstellung personalisierter Arbeitsblätter für Schüler*innen im Alter von 10-18 Jahren. Das System kombiniert modernste KI-Technologie mit den spezifischen Anforderungen des österreichischen Bildungssystems und setzt dabei neue Standards für inklusive Bildung. Das System unterstützt derzeit acht Fächer mit unterschiedlichen Implementierungsständen: Mathematik und Englisch sind vollständig funktionsfähig mit umfangreichen, qualitätsgesicherten Aufgabensammlungen. Deutsch und Informatik weisen eine vollständige Curriculum-Integration auf, benötigen jedoch noch den Aufbau einer ausreichenden Aufgabenbasis, da öffentlich verfügbare Materialien in diesen Fächern stark limitiert sind.

Die Plattform generiert Aufgaben in drei Schwierigkeitsstufen und über 20 verschiedenen Aufgabentypen. Ein weltweit einzigartiges Gender/Gleichstellungs-Moderationssystem stellt sicher, dass alle generierten Inhalte den aktuellen Standards für inklusive Bildung entsprechen.

Seit Projektbeginn wurden erhebliche technische und pädagogische Fortschritte erzielt. Die Plattform umfasst nun über 60.000 Zeilen Code. Besonders hervorzuheben ist die implementierte Integration für die Fächer Deutsch, Englisch, Mathematik und Informatik des österreichischen Curriculums, basierend auf den RIS Bundesnormen der AHS, die eine präzise Abstimmung aller generierten Aufgaben auf spezifische Kompetenzbereiche ermöglicht.

2 Status

2.1 Meilenstein 1 - Literaturrecherche

Kurzbeschreibung der Haupttätigkeiten

Die Literaturrecherche umfasste eine systematische Untersuchung der bestehenden Bildungstechnologielandschaft im deutschsprachigen Raum sowie aktuelle Forschungstrends im Bereich KI-gestützter personalisierter Lernumgebungen. Der Fokus lag auf der Identifikation von Marktlücken, der Analyse bestehender Plattformen und der wissenschaftlichen Fundierung des Projektansatzes.

Die Hauptaktivitäten beinhalteten:

- Analyse führender EdTech-Plattformen im DACH-Raum (ANTON, Sofatutor, Bettermarks, StudySmarter, to_teach)
- Recherche aktueller Forschung zu AI-gestütztem personalisiertem Lernen (2023-2025)
- Untersuchung der Arbeitsbelastung von Lehrkräften und Zeitaufwand für Unterrichtsvorbereitung
- Evaluation internationaler KI-Bildungsframeworks (EC/OECD AI Literacy Framework 2025)
- Identifikation wissenschaftlicher Arbeiten österreichischer und deutscher Forschungsinstitutionen die über KI-gestützten

Erkenntnisse zur Vorgangsweise

Bestehende Plattformen und deren Limitierungen

Die Analyse bestehender Bildungsplattformen ergab signifikante Marktlücken:

ANTON App: Kostenlose Lernplattform für Klassen 1-10, entwickelt von solocode GmbH (Berlin). Die App bietet über 100.000 Aufgaben und wird als "completely free without any extra costs involved" beworben (anton.app, 2024). Bietet strukturierte Übungen, jedoch keine KI-gestützte Personalisierung oder Aufgabengenerierung.

Sofatutor: Kommerzielle Plattform mit über 11.000 selbstproduzierten Lernvideos. Die Preise beginnen bei 15,96€/Monat, wobei historische Preismodelle zwischen 8-14€ lagen (Sofatutor Pricing, 2024; TechCrunch, 2010). Fokus auf Videoinhalte, keine intelligente Aufgabenerstellung oder Lehrplanintegration.

Bettermarks: Adaptive Mathematik-Lernplattform aus Berlin mit über 100.000 interaktiven Übungen für Klassen 4-10. Die Plattform erkennt über 2.800 Fehlermuster und bietet personalisiertes Feedback. Laut der EVA-CBTM Studie 2012 war bettermarks Testsieger unter deutschen und englischen Produkten. In Uruguay verbesserten sich Schülerleistungen um 10%, bei sozial benachteiligten Kindern sogar um bis zu 30% (bettermarks.com, 2024). Beschränkt auf Mathematik, keine fächerübergreifende Lösung.

StudySmarter: Intelligente Lern-App mit Community-generierten Inhalten. Primär für Studenten konzipiert, keine spezifische Anpassung an österreichische Lehrpläne.

Die Forschung zu KI im Bildungsbereich zeigt laut aktuellen Studien positive Effekte, aber mit differenzierten Ergebnissen:

Die RAND Corporation berichtet von Verbesserungen um 2 % in Mathematik und Lesen durch personalisiertes Lernen (RAND, 2024)

Eine Gates Foundation Studie fand Effektgrößen von .41 in Mathematik und .29 im Lesen bei personalisierten Lernansätzen (EdWeek, 2014)

92% der Studenten nutzen KI-Tools in ihrem Studium (HEPI/Kortext Survey, 2025), 86% nutzen KI laut Digital Education Council (2024)

Kritische Herausforderungen: Datenschutz, algorithmische Voreingenommenheit, digitale Kluft
Empirische Daten zur Lehrerbeltung unterstreichen die Dringlichkeit effizienter

Unterrichtsvorbereitung:

Durchschnittliche Arbeitszeit: 54 Stunden/Woche (EdWeek Research Center; TASB, 2024)

84% der Lehrkräfte berichten über Zeitmangel für alle erwarteten Aufgaben (Pew Research Center, April 2024)

77% empfinden ihren Beruf als stressig, 68% als überwältigend (Pew Research Center, 2024)

94% der Lehrkräfte, die einen Berufswechsel erwägen, nennen hohe Arbeitsbelastung als wichtigen Faktor (UK Government Working Lives Report, 2024)

Kurzbeschreibung der erreichten Ergebnisse

Die Literaturrecherche validierte die Grundannahmen des Projekts und identifizierte klare Alleinstellungsmerkmale:

Marktlücke bestätigt: Keine existierende Plattform bietet KI-gestützte, lehrplankonforme Aufgabengenerierung für den österreichischen Bildungsraum!

Wissenschaftliche Fundierung: Integration des EC/OECD AI Literacy Framework "Empowering Learners for the Age of AI" (Launch Event 22. Mai 2025, ailiteracyframework.org) als theoretische Basis gewährleistet internationale Standards

Differenzierungsfaktoren dokumentiert:

Vollständige österreichische Lehrplanintegration (Alleinstellungsmerkmal)

Gender-Moderationssystem basierend auf behördlichen Richtlinien (weltweit einzigartig)

Fokus auf Zeitersparnis bei gleichzeitiger Qualitätssteigerung

Besondere Erfolge/ Probleme

Erfolge:

- Identifikation von 117 offiziellen Beispielen für geschlechtergerechte Sprache aus DACH-Behörden als Grundlage für das Moderationssystem
- Bestätigung der Relevanz durch aktuelle Forschung: Generative KI revolutioniert adaptive Lernsysteme (Frontiers in Education, 2024; Emerald Insight AI in Education Studies, 2024)
- Nachweis des Bedarfs: Lehrkräfte berichten von durchschnittlich 54 Stunden Arbeitszeit pro Woche, wobei weniger als 46% der Zeit für tatsächlichen Unterricht verwendet wird (EdWeek, 2024)

Probleme:

- Begrenzte öffentlich verfügbare Daten zu kommerziellen EdTech-Plattformen im DACH-Raum
- Wenige spezifische Studien zur KI-Akzeptanz bei österreichischen Lehrkräften
- Fehlende standardisierte Metriken für Qualitätsmessung KI-generierter Bildungsinhalte

Gab es große Abweichungen zum Plan? Warum?

Die Literaturrecherche verlief weitgehend planmäßig mit folgenden Anpassungen:

- Erweiterter Fokus auf internationale Frameworks: Ursprünglich war nur die Analyse österreichischer Standards geplant. Die Entdeckung des EC/OECD AI Literacy Framework führte zu einer wertvollen Erweiterung der theoretischen Grundlage.
- Tiefere Analyse der Lehrerbelastung: Die alarmierenden Statistiken zur Arbeitsbelastung (54 Stunden/Woche) erforderten eine ausführlichere Untersuchung als geplant, bestätigten aber die Projektrelevanz.
- Zusätzliche Recherche zu Generativer KI: Die rapiden Entwicklungen im Bereich Generative AI machten eine Aktualisierung der technologischen Grundlagen notwendig, führten aber zu wertvollen Erkenntnissen für die Systemarchitektur.

2.2 Meilenstein 2 - Datensammlung

Kurzbeschreibung der Haupttätigkeiten

Die Datensammlung umfasste eine systematische Recherche von Bildungsmaterialien für die KI-gestützte Plattform. Als Hauptaktivität wurde das vollständige Scraping der österreichischen Aufgabenpools durchgeführt: <https://prod.aufgabenpool.at/amn/index.php?id=M> für Mathematik und https://prod.aufgabenpool.at/srdp_lfs/index.php?id=en für Englisch. Zusätzlich wurden vertiefende Recherchen zur Verfügbarkeit digitaler Bildungsressourcen in Österreich und Deutschland durchgeführt. Ergänzend erfolgte die Akquisition von Materialien der Bildungsstandards-Aufgabenbeispiele Mathematik über <https://www.iqs.gv.at/downloads/archiv-des-bifie/unterstuetzende-materialien-zu-bildungsstandards/aufgabenbeispiele-mathematik>. Direkte Kontaktaufnahmen mit Lehrkräften und Bildungseinrichtungen zur Materialgewinnung bildeten den dritten Schwerpunkt der Aktivitäten.

Erkenntnisse zur Vorgangsweise

Die Literaturrecherche und praktische Datenakquisition offenbarten fundamentale strukturelle Probleme im österreichischen und deutschen Bildungswesen. Österreich investierte zwar €278 Millionen in die K-12-Digitalisierung bis 2024, jedoch existieren praktisch keine öffentlichen APIs oder programmatische Zugriffsmethoden für Bildungsinhalte. Die Eduthek-Plattform mit über 6.000 Lernmaterialien wurde ausschließlich für menschliche Nutzer:innen konzipiert, ohne standardisierte Methoden für automatisierte Inhaltsverarbeitung. Die rechtliche Situation erwies sich als komplex: Während [aufgabenpool.at](https://prod.aufgabenpool.at) unter Creative Commons BY 4.0 und dem Informationsweiterverwendungsgesetz (IWG 2022) verfügbar ist, herrscht bei anderen Materialien rechtliche Unsicherheit. Lehrkräfte zeigten erhebliche Zurückhaltung bei der Materialweitergabe, was auf systemische Barrieren wie DSGVO-Compliance-Ängste, unklare Urheberrechtsrichtlinien und mangelnde institutionelle Unterstützung zurückzuführen ist.

Kurzbeschreibung der erreichten Ergebnisse

Die Datensammlung erbrachte eine solide Basis für die Plattform: Vollständige Akquisition aller SRDP-Mathematik- und Englisch-Aufgaben aus dem österreichischen Aufgabenpool, umfassende Sammlung der IQS-Bildungsstandards-Beispiele für Mathematik, sowie rechtlich abgesicherte Nutzungsgrundlage durch Creative Commons-Lizenzierung. Darüber hinaus entstand eine fundierte Analyse der österreichischen Bildungslandschaft, die die systematischen Hürden für

EdTech-Entwickler dokumentiert. Die Erkenntnisse zur mangelnden Verfügbarkeit offener Bildungsressourcen bestätigten die Notwendigkeit einer generativen KI-Lösung und validierten das Projektkonzept.

Besondere Erfolge/ Probleme

Der größte Erfolg lag in der rechtlich einwandfreien Akquisition hochwertiger, standardisierter Aufgaben direkt vom österreichischen Bildungsministerium. Die Creative Commons BY 4.0-Lizenzierung ermöglicht nicht nur die Nutzung, sondern auch die kommerzielle Weiterverwendung bei korrekter Quellenangabe ("Datenquelle: Bundesministerium für Bildung"). Problematisch erwies sich jedoch die extreme Schwierigkeit, darüber hinaus systematisch nutzbare Bildungsressourcen zu finden. Nur 19,9% der österreichischen Lehrkräfte fühlen sich im IKT-Einsatz gut vorbereitet, der niedrigste EU-Wert. Die Recherche deckte auf, dass Österreich im Gegensatz zu Estland oder den Niederlanden keine umfassende OER-Politik (Open Educational Resources) verfolgt. Lehrkräfte verweigerten größtenteils die Materialteilung aus Angst vor Rechtsverletzungen und aufgrund zeitlicher Überlastung. Die föderale Struktur mit unterschiedlichen Regelungen für Bundes- und Landesschulen verkomplizierte systematische Lösungsansätze zusätzlich.

Gab es große Abweichungen zum Plan? Warum?

Erhebliche Planabweichungen ergaben sich aus der unerwarteten Knappheit verfügbarer Bildungsdatenbanken. Der ursprüngliche Plan sah vor, mehrere bestehende Ressourcen zu integrieren und zu erweitern. Stattdessen musste auf eine primär generative Strategie umgeschwenkt werden. Die geplante Zusammenarbeit mit Lehrkräften zur Materialsammlung scheiterte weitgehend an den identifizierten systemischen Barrieren. Diese Abweichung erwies sich jedoch als glücklicher Umstand: Die Notwendigkeit einer vollständig generativen Lösung bestätigte die Innovationskraft des gewählten Ansatzes. Anstatt nur bestehende Materialien zu digitalisieren, entsteht ein System, das personalisierte, qualitätsgesicherte Aufgaben nach österreichischen Bildungsstandards erzeugt. Die ursprünglich als Problem wahrgenommene Materialknappheit wurde zum Kernargument für die Projektrelevanz und untermauerte die Notwendigkeit einer KI-gestützten Lösung für das österreichische Bildungswesen.

2.3 Meilenstein 3 - Umfrage zur Ermittlung von Kriterien für gute Arbeitsblätter (Klarheit, Relevanz, Schwierigkeitsgrad, Aufgabenvielfalt) bei Lehrkräften..

Kurzbeschreibung der Haupttätigkeiten

Von Jänner bis März 2025 wurde eine systematische Online-Umfrage zur Ermittlung von Qualitätskriterien für KI-generierte Arbeitsblätter durchgeführt. 11 Lehrkräfte aus AHS, MS und BHS (Altersgruppe der Schülerinnen überwiegend 10-14 Jahre) bewerteten sieben zentrale Qualitätskriterien mittels Likert-Skalen (1 = stimme gar nicht zu, 5 = stimme voll zu): Task Relevance and Alignment (Lehrplanbezug und Anknüpfung an Vorwissen), Task Difficulty and Challenge (angemessener Schwierigkeitsgrad), Clarity and Accessibility (Klarheit der Anweisungen), Accuracy and Quality (Richtigkeit und Strukturierung), Equity and Inclusion (Inklusivität und Bias-Freiheit), Time Appropriateness (zeitliche Angemessenheit) sowie Engagement and Motivation (Schülerinnen-Engagement). Parallel erfolgte eine umfassende Literaturrecherche zur wissenschaftlichen Fundierung jedes einzelnen Kriteriums durch peer-reviewte Studien aus den Bereichen Bildungspsychologie, Curriculum-Design und Instruktionsdesign.

Erkenntnisse zur Vorgangsweise

Die Kombination aus empirischer Befragung und theoretischer Fundierung erwies sich als methodisch robust. Die Likert-Skala ermöglichte eine präzise Quantifizierung der Lehrkräfte-Prioritäten, während die offenen Fragen qualitative Einsichten lieferten. Die wissenschaftliche Literaturrecherche bestätigte alle gewählten Kriterien durch substantielle Evidenz: Jin et al. (2019) belegten die Bedeutung der Lehrplanalignment, Sweller et al. (2019) demonstrierten optimale Lerneffekte bei angemessener kognitiver Belastung, und Freeman et al. (2014) zeigten in einer Meta-Analyse von 225 Studien 6% höhere Prüfungsleistungen bei engagierenden Materialien. Die Forschung offenbarte Effektstärken von $d = 0.30-0.80$ für qualitativ hochwertige Bildungsmaterialien, was diese als "high-leverage intervention" klassifiziert. Besonders wertvoll war die Erkenntnis, dass die Kriterien synergistisch wirken, Arbeitsblätter, die in mehreren Dimensionen excellieren, erzeugen multiplikative statt nur additive Lernvorteile.

Kurzbeschreibung der erreichten Ergebnisse

Die Umfrage erbrachte klare Prioritäten: Richtigkeit und Lehrplanrelevanz erhielten die höchsten Bewertungen (Durchschnitt >4.5), während Zeiteffizienz den niedrigsten, aber immer noch positiven Wert erzielte. Die qualitative Analyse der Freitexte identifizierte drei zentrale Anforderungen: ein obligatorischer Verifizierungs-Workflow ("Vier-Augen-Prinzip"), granulare Adaptivität der Schwierigkeitsgrade, und erweiterte Fach- und Formatvielfalt. Die wissenschaftliche Fundierung bestätigte alle sieben Kriterien durch substantielle Evidenz: Hines et al. (1985) belegten 20-30% Leistungssteigerung durch klare Instruktionen, King-Sears et al. (2023) zeigten moderate positive Effekte ($g = 0.43$) für UDL-basierte inklusive Materialien, und Li & Xue (2023) dokumentierten signifikante Korrelationen zwischen Engagement und Lernerfolg. Diese Erkenntnisse führten zur Implementierung eines strukturierten Qualitätssicherungssystems mit automatisierter Bewertung aller generierten Aufgaben entlang der validierten Kriterien.

Besondere Erfolge/ Probleme

Der größte Erfolg lag in der wissenschaftlichen Validierung aller gewählten Kriterien durch hochrangige peer-reviewte Studien, was dem Plattform-Design eine solide evidenzbasierte Grundlage verlieh. Die Bereitschaft zur Beta-Teilnahme (9 von 11 Befragten) demonstrierte das Interesse der Lehrkräfte an der Lösung. Problematisch erwies sich jedoch die geringe Teilnehmerzahl (n=11), die statistische Generalisierbarkeit limitiert. Darüber hinaus wollten trotz positiver Bewertungen keine Lehrkräfte eigenes Unterrichtsmaterial zur Verfügung stellen, was die bereits in Meilenstein 2 identifizierten systemischen Sharing-Barrieren bestätigte. Die Komplexität der simultanen Berücksichtigung aller sieben Kriterien stellte hohe Anforderungen an das technische System-Design. Besonders herausfordernd war die Integration der Equity-Kriterien in die automatisierte Bewertung, da Bias-Erkennung subtile kulturelle und kontextuelle Nuancen erfordert.

Gab es große Abweichungen zum Plan? Warum?

Eine signifikante Planabweichung ergab sich durch die Entscheidung, von ursprünglich geplanten vier auf sieben Bewertungskriterien zu erweitern. Die initiale Konzeption sah nur Klarheit, Relevanz, Schwierigkeitsgrad und Aufgabenvielfalt vor. Die Literaturrecherche und ersten Lehrkräfte-Gespräche identifizierten jedoch Equity/Inclusion, Time Appropriateness und Engagement als ebenso kritische Faktoren. Diese Erweiterung war notwendig, da aktuelle Bildungsforschung zeigt, dass umfassende Qualitätsbewertung alle Dimensionen effektiven Lernens berücksichtigen muss. Die zusätzlichen Kriterien erhöhten zwar die Systemkomplexität, ermöglichten aber eine weltweite einzigartige holistische Qualitätsbewertung KI-generierter Bildungsmaterialien. Die kleinere Stichprobe (11 statt der ursprünglich geplanten 20–30 Teilnehmer*innen) resultierte aus der bereits dokumentierten Zurückhaltung österreichischer Lehrkräfte bei Forschungs Kooperationen. Die Befragung ist jedoch auf der Projektwebseite öffentlich einsehbar und die Zahl der Teilnehmenden wird voraussichtlich noch steigen, da ich sie auf der Website pushen werde. Zudem können relevante Metriken bei Bedarf entsprechend angepasst werden. Die geringe Zahl wurde zudem durch die intensive qualitative Analyse und die solide wissenschaftliche Fundierung kompensiert.

2.4 Meilenstein 4 - Plattform wird im schulischen Umfeld eingesetzt

Technische Architektur

Die Entwicklung der intelligenten Aufgabenerstellungsplattform umfasste den Aufbau eines produktionsreifen Systems, das auf einer Multi-LLM-Strategie und einer vierlagigen Softwarearchitektur basiert. Als primäres Sprachmodell wurde Mistral-medium-latest aufgrund seiner hohen Effizienz, und kontextuellen Genauigkeit gewählt, während GPT-4o mini für den empirischen Vergleich verwendet wird, um einer meiner wissenschaftlichen Fragen zu beantworten. Die Verwaltung der komplexen Prompt-Infrastruktur erfolgt über Langfuse v3, das ein Cloud-gehostetes, versioniertes Management von Prompt-Templates ermöglicht und somit Reproduzierbarkeit und schnelles Prototyping sicherstellt.

Die Architektur gliedert sich in vier logische Schichten:

Generation Layer: Dieses Kernmodul orchestriert den gesamten Erstellungsprozess. Es integriert eine anspruchsvolle Moderations-Engine zur Sicherstellung von Gender-Gerechtigkeit und pädagogischen Standards sowie eine automatisierte Regeneration-Schleife, die bei Qualitätsmängeln selbstständig neue Aufgabenvarianten anfordert.

Context Layer: Das Herzstück der Plattform ist ein progressives Kontext-Retrieval-System, das eine achtstufige Relaxation-Strategie anwendet, um auch bei sehr spezifischen Anforderungen relevanten Kontext zu finden.

Enhancement Layer: Dieses Modul ist für die nachgelagerte Qualitätsverbesserung zuständig. Es führt automatisierte Korrekturen an der LaTeX-Syntax durch und optimiert die sprachliche Klarheit der generierten Aufgaben.

Quality Assurance Layer: Eine automatisierte Pipeline, die sowohl technische Validierungen (z.B. JSON-Formatierung) als auch inhaltliche Prüfungen (z.B. Moderations-Scores) durchführt.

Prompt-Management-System

Das methodische Fundament bildet ein umfangreiches Prompt-Management-System, das 29 spezialisierte Jinja2-Vorlagen für verschiedene Fächer und Aufgabentypen umfasst. Darunter fallen komplexe Anweisungen wie `math_open_question` für schrittweise mathematische Problemlösungen, `english_vocabulary_test` für strukturierte Vokabelprüfungen und `german_grammar` für Grammatikübungen mit spezifischem kulturellem Kontext für Österreich.

Technische Innovationen

Der entscheidende technische Durchbruch wurde durch die Entwicklung des progressiven Kontext-Retrieval-Systems erzielt. Dieses implementiert eine achtstufige Kaskade von Relaxationsstrategien, die von einer exakten Übereinstimmung (`EXACT_MATCH`) bis hin zu minimalen Randbedingungen (`MINIMAL_CONSTRAINTS`) reicht. Diese Methode stellt sicher, dass das System stets den bestmöglichen Kontext liefert, anstatt bei zu restriktiven Filtern zu versagen. Die Retrieval-Augmented Generation (RAG) wurde mittels `pgvector` auf einer PostgreSQL-Datenbank realisiert. Hierfür werden 1024-dimensionale Embeddings über Mistral's Embedding-API generiert und in multiplen, für verschiedene Inhaltstypen optimierten Vektor-Indizes gespeichert. Die Validierung von ähnlichen Aufgaben als Kontext erfolgt über eine Cosine-Similarity-Suche, die semantische Relevanz garantiert.

Fachspezifische Lösungsansätze

Die fachspezifischen Anforderungen erforderten maßgeschneiderte Lösungsansätze:

Mathematik: Aufgrund der häufigen Fehler in KI-generiertem LaTeX wurde ein robustes Post-Processing-System entwickelt, das auf Regex-basierten Mustern beruht, um häufige Syntaxfehler automatisch zu korrigieren.

Englisch: Für "Reading Comprehension"-Aufgaben wurde ein System implementiert, das eine strukturierte JSON-Ausgabe erzwingt und so die maschinelle Auswertbarkeit sicherstellt.

Deutsch: Um den Anforderungen des österreichischen Bildungssystems gerecht zu werden, wurde eine tiefgreifende kulturelle Kontextualisierung in die Prompts integriert.

Insgesamt wurden über 200 Curriculum-Kompetenzen des österreichischen Lehrplans digital abgebildet und entlang der Hierarchie Fach → Jahrgangsstufe → Themenbereich → Kompetenz → Unterziel strukturiert.

Entwicklung des Gender-Moderationssystems

Die Entwicklung des Gender-Moderationssystems basierte auf einer systematischen Analyse offizieller Richtlinien des BMBWF zur geschlechtergerechten Schule. Das System scannt generierte Aufgaben und erkennt potenzielle Diskriminierungen, indem es den Text mit einer umfassenden Datenbank von Negativbeispielen vergleicht. Statt starrer Regeln nutzt es eine vektorbasierte Ähnlichkeitssuche, um auch subtile Verletzungen zu identifizieren. Das System erkennt aus Korrekturen Fehler und hat mittels Fine-tuning das Potenzial kontinuierlich sich zu verbessern.

Technische Implementierung des Moderationssystems

Das System nutzt eine pgvector-Datenbank mit Embeddings und speichert 58 indizierte Guideline-Chunks. Diese können jederzeit erweitert werden, in der Zukunft auch von Lehrkräften. Die technische Implementierung kombiniert Mistral AI für Embeddings und Analyse mit strukturierten JSON-Antworten für Violation-Detection. Die größte Herausforderung war der Aufbau einer umfassenden Violation-Datenbank. Durch iterative Entwicklung entstanden mehrere Service-Versionen mit einer stetig wachsenden Anzahl an Beispielen. Das finale System deckt über 15 Diskriminierungskategorien ab, darunter Geschlechterstereotype, veraltete Rollenbilder und exkludierende Sprache.

Performance und Qualitätssicherung

Die technische Infrastruktur ist auf maximale Performance und Stabilität ausgelegt. Sie zeichnet sich durch eine durchgängige Async/Await-Implementierung für nicht-blockierende Operationen aus.

Die Qualitätssicherung erfolgt vollautomatisiert durch drei Kernkomponenten: einen universellen LaTeX-Normalizer, eine proaktive Gender-Equality-Moderation und eine Regenerierungs-Routine, die bei der Unterschreitung vordefinierter Qualitätsmetriken automatisch eingreift.

Erreichte Erfolge

Adaptives Kontextsystem: Die Entwicklung einer adaptiven Kontextstrategie, die je nach Fach unterschiedliche Prioritäten setzt (in Mathematik wird die Kompetenzdeckung, in Englisch die Konsistenz des Aufgabentyps priorisiert), führte zu einer signifikanten Steigerung der Relevanz und Qualität der generierten Aufgaben.

Umfassende Testinfrastruktur: Eine robuste Testsuite mit 59 automatisierten Evaluationsdateien sichert die Regressionsfreiheit bei Weiterentwicklungen und validiert die End-to-End-Funktionalität.

Produktionsreifes Monitoring: Die tiefe Integration von Langfuse-Tracing für LLM-Call-Analytics und Sentry-Monitoring für Error-Tracking bietet eine umfassende Observability und ermöglicht proaktives Performance-Management.

Bewältigte Herausforderungen

LaTeX-Dialekt-Inkonsistenzen: Verschiedene LLMs produzierten unterschiedliche und oft fehlerhafte LaTeX-Dialekte. Die Lösung war die Entwicklung eines universellen LaTeX-Normalizers, der mit einem umfangreichen Set an Regex-Mustern die verschiedenen Varianten vereinheitlicht und korrigiert.

Balance zwischen Kontextqualität und -verfügbarkeit: Anfänglich führten zu restriktive Matching-Kriterien zu häufigen Kontextlücken. Dieses Problem wurde durch die Einführung dynamischer, fachspezifischer Quality-Thresholds gelöst, die eine optimale Balance zwischen Präzision und Abdeckung herstellen.

Skalierung bei hoher Parallelität: Die Verarbeitung zahlreicher gleichzeitiger Anfragen erforderte den Übergang zu einer vollständig asynchronen Processing-Pipeline, die durch QStash-gestützte Hintergrundjobs entkoppelt wurde, um die Systemlast zu verteilen und die Antwortzeiten zu stabilisieren.

Architekturevolution

Die Systemarchitektur erfuhr eine signifikante Evolution, die über den ursprünglichen Entwurf hinausging:

Architektur:

Wandel von einem monolithischen Service, der einfache LLM-Prompts ausführt, zu einem entkoppelten, vierlagigen System (Generierung → Kontext → Enhancement → Qualität).

Prompt-Engineering: Ersetzung eines einzelnen, universellen Prompts durch 29 spezialisierte und kontextsensitive Jinja2-Vorlagen.

Kontext-Retrieval: Weiterentwicklung von einem direkten Kompetenz-Matching zu einer flexiblen, achtstufigen Relaxationsstrategie.

Qualitätssicherung: Übergang von einer manuellen Validierung zu einer vollautomatisierten Moderations- und Korrektur-Pipeline.

Begründung der Abweichungen

Die signifikanten Abweichungen vom ursprünglichen Plan waren eine direkte Konsequenz aus der tiefen Auseinandersetzung mit der Komplexität des österreichischen Bildungssystems. Die Notwendigkeit, pädagogisch validierte, qualitativ hochwertige und diskriminierungsfreie Aufgaben zu erstellen, machte eine weitaus differenziertere und robustere technische Lösung erforderlich.

Der strategische Mehraufwand in die Entwicklung dieser anspruchsvollen Architektur führte zu einer erheblichen Steigerung der Systemqualität, Skalierbarkeit und Produktionsreife.

2.5 Meilenstein 5 - Plattform wird im schulischen Umfeld eingesetzt

Testmethodik

Die Pilottests umfassten strukturierte 15-minütige Sessions mit fünf Lehrkräften unterschiedlicher Fachrichtungen und Erfahrungsstufen. Die Tests wurden remote via Screen-Sharing durchgeführt, wobei quantitative Metriken (Task-Completion-Time, Error-Rate, Feature-Discovery) und qualitative Daten (Think-Aloud-Protokoll, semi-strukturierte Interviews) erhoben wurden.

Basierend auf Real-Time-Feedback wurde während der Testphase die "Rapid-Feature-Implementation" praktiziert: Identifizierte Verbesserungen(und Fehler) wurden innerhalb von 24-48 Stunden implementiert.

Zentrale Erkenntnisse

Die wichtigste Erkenntnis war die Bedeutung des Onboardings. Ohne strukturierte Einführung entdeckten Nutzer:innen nur 40% der Features. Die Implementierung eines interaktiven, in diesem Fall vor gesprochenen Tutorial-Systems mit Progress-Tracking erhöhte die Feature-Discovery-Rate auf 90%. Um das volle Potenzial der Plattform auch wirklich allen zu gewährleisten, werden "Loom"-Videos beim Initial Launch vorhanden sein.

Das Chatbot-System erwies sich als überraschend populär. Ursprünglich gar nicht oder nur als Nebenfunktion geplant, wurde es von allen Testpersonen als "Killer-Feature" bezeichnet. Die spontane Entwicklung des "Qrious Learner" Chatbots während der Testphase demonstrierte die Flexibilität der Systemarchitektur. Die quantitativen Ergebnisse übertrafen ebenfalls allen Erwartungen.

Institutionelle Partnerschaften

Besonders bedeutsam ist der gewonnene institutionelle Partner: Die VHS Wien. Von Theresia Manas, einer der Leiterinnen von <https://www.vhs.at/de/e/wiener-lernhilfe/sommerangebote>, erhielt ich ein sehr starkes Interesse an der Plattform. Sie möchte diese bei allen Lernstationen in Verwendung haben. Des Weiteren hat sie gemeint, dass dieses Tool in der Erwachsenenbildung ebenfalls sehr interessant sei, und sie hat natürlich von mir die Zusage erhalten, es bei den Lernstationen verwenden zu können. Da ich noch über ein gutes Budget verfüge, ist es mir selbstverständlich möglich, dieses Tool komplett kostenfrei für alle zur Verfügung zu stellen. Des Weiteren bin ich gerade im Gespräch mit der AHS Goethe Gymnasium Astgasse (900 Schüler*innen und über 100 Lehrkräfte) als Pilotschule für eine langfristige Projektumsetzung.

Methodische Anpassungen

Der ursprünglich geplante Umfang von 20 Testpersonen wurde strategisch auf fünf intensive Testsessions reduziert. Diese Anpassung erwies sich als methodisch überlegen, da die qualitative Tiefe des erhaltenen Feedbacks den quantitativen Ansatz deutlich übertraf. Die konzentrierten Sessions ermöglichten substantiellere Erkenntnisse und resultierten in präziseren, zielgerichteten Systemverbesserungen. Mit der geplanten öffentlichen Verfügbarkeit der Plattform wird sich die Nutzerbasis und damit das Feedback-Volumen organisch und signifikant erweitern.

Technische Herausforderungen und aktueller Status

Aufgrund verschiedener technischer Herausforderungen konnte die Plattform noch nicht vollständig deployed werden. Skalierungsissues, Mailserver aufsetzen und die generelle Usability für mehrere hunderte User zur gleichen Zeit, erwies sich als größere Aufgabe als initial gedacht. Derzeit konzentrieren sich alle Entwicklungsressourcen darauf, das System für den Online-Launch vorzubereiten. Ein separater Blogbeitrag mit der offiziellen Ankündigung wird zeitnah veröffentlicht.

3 Zusammenfassung Planaktualisierung

3.1 Fazit

Das Projekt hat sich von einer ursprünglich einfacheren Konzeption zu einer hochkomplexen, aber dafür deutlich robusteren und pädagogisch fundierten Lösung entwickelt. Die Abweichungen vom ursprünglichen Plan waren notwendig, um den hohen Qualitätsansprüchen des österreichischen Bildungssystems gerecht zu werden und ein wirklich praxistaugliches Tool zu schaffen.

3.2 Planaktualisierung

Laut der aktualisierten Stipendienplanung wurden mehrere Meilensteine um etwa einen Monat nach hinten verschoben, was auf unerwartete technische Herausforderungen zurückzuführen ist. Der schulische Einsatz der Plattform (Meilenstein 5) musste von April-Juni auf Juni verschoben werden, da die ursprünglich entwickelten Prompts nicht die erwarteten Qualitätsergebnisse lieferten und eine grundlegende Überarbeitung der Prompt-Architektur notwendig wurde. Zusätzlich erwies sich die initiale Systemarchitektur als nicht skalierbar – die Plattform war nur für 5-10 gleichzeitige Nutzer:innen ausgelegt, was völlig unzureichend war, als die VHS Wien starkes Interesse an einer institutionellen Partnerschaft zeigte.

Wesentlich mehr Zeit musste auch in die Curriculum-Anpassungen investiert werden, da das ursprünglich AHS-fokussierte System zu einer abstrakteren, schultyp-übergreifenden Lösung weiterentwickelt werden musste, um Lehrkräften von verschiedenen Schultypen (AHS, MS, BHS) eine optimale Nutzung zu ermöglichen. Die Verbesserungsphase der Plattform wurde ebenfalls um

einen Monat vorgezogen, da umfangreiche Deployment-Schwierigkeiten und die Implementierung von Rate-Limiting-Mechanismen zur Gewährleistung konstanter Verfügbarkeit für alle Nutzer:innen mehr Zeit in Anspruch nahmen als ursprünglich geplant. Die Schreib- und Korrekturphasen wurden entsprechend angepasst, um eine fundierte Analyse der erweiterten Systemarchitektur zu ermöglichen. Diese strategischen Verzögerungen stellen sicher, dass die Plattform über den Sommer und Anfang September optimal an Lehrkräfte vermittelt werden kann und dabei höchsten technischen und pädagogischen Standards entspricht.

Erwähnte Quellenangaben von der Literaturrecherche:

Quellenverzeichnis

Bildungsplattformen

ANTON App

- ANTON - the free learning app for school. (2024). https://anton.app/en_us/
- solocode GmbH. (2024). ANTON: Kindergarten - Grade 8. App Store. <https://apps.apple.com/us/app/anton-kindergarten-grade-8/id1180554775>

Sofatutor

- Sofatutor. (2024). Preisverzeichnis. <https://www.sofatutor.com/legal/preise>
- Sofatutor. (2024). Learning offer and pricing. <https://us.sofatutor.com/pricing>
- TechCrunch. (2010). Sofatutor closes Series A round to expand video courses. <https://techcrunch.com/2010/01/12/sofatutor-closes-series-a-round-to-expand-video-courses/>
- Frog Capital. (2021). Sofatutor replaces lessons as German school students remain at home. <https://frogcapital.com/think-frog/sofatutor-replaces-lessons-as-german-school-students-remain-at-home/>

Bettermarks

- Bettermarks. (2024). Adaptive Learning System. <https://bettermarks.com/>

- Bettermarks. (2024). Platform. <https://bettermarks.com/platform/>
- EU Startup News. (2024). Startup Showcase: Bettermarks - Revolutionizing Mathematics Education. <https://eustartup.news/startup-showcase-bettermarks-revolutionizing-mathematics-education/>
- FASE. (2024). Bettermarks. <https://fa-se.de/en/projects/finalised/bettermarks/>

Forschung zu personalisiertem Lernen und KI in der Bildung

RAND Corporation

- RAND. (2024). How Does Personalized Learning Affect Student Achievement? https://www.rand.org/pubs/research_briefs/RB9994.html

Gates Foundation Studie

- EdWeek. (2014). Personalized Learning May Boost Achievement, Charter Study Says. <https://www.edweek.org/technology/personalized-learning-may-boost-achievement-charter-study-says/2014/11>

Studentenumfragen zu KI-Nutzung

- HEPI. (2025). Student Generative AI Survey 2025. <https://www.hepi.ac.uk/2025/02/26/student-generative-ai-survey-2025/>
- Digital Education Council. (2024). What Students Want: Key Results from DEC Global AI Student Survey 2024. <https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024>
- Campus Technology. (2024). Survey: 86% of Students Already Use AI in Their Studies. <https://campustechnology.com/articles/2024/08/28/survey-86-of-students-already-use-ai-in-their-studies.aspx>

Weitere Forschungsquellen

- Frontiers in Education. (2024). Crafting personalized learning paths with AI for lifelong learning: a systematic literature review. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1424386/full>
- Emerald Insight. (2024). AI-enabled personalized learning: empowering management students for improving engagement and academic performance. <https://www.emerald.com/insight/content/doi/10.1108/xjm-02-2024-0023/full/html>

Lehrerbelastung und Arbeitszeit

Arbeitszeit-Statistiken

- EdWeek Research Center. (2024). Here's How Many Hours a Week Teachers Work. <https://www.edweek.org/teaching-learning/heres-how-many-hours-a-week-teachers-work/2022/04>
- TASB. (2024). Typical Teacher Works 54 Hours per Week. <https://www.tasb.org/news-insights/typical-teacher-works-54-hours-per-week>
- Pew Research Center. (2024). How K-12 public teachers manage their workload. <https://www.pewresearch.org/social-trends/2024/04/04/how-teachers-manage-their-workload/>

Internationale Studien

- Tandfonline. (2023). Workload, work intensification and time poverty for teachers and school leaders: a systematic research synthesis. <https://www.tandfonline.com/doi/full/10.1080/00131911.2023.2196607>
- UK Government. (2024). Working lives of teachers and leaders: wave 2 summary report. <https://www.gov.uk/government/publications/working-lives-of-teachers-and-leaders-wave-2/working-lives-of-teachers-and-leaders-wave-2-summary-report>

AI Literacy Framework

EC/OECD Initiative

- AI Literacy Framework. (2025). AI Literacy Framework for Primary & Secondary Education. <https://ailiteracyframework.org/>
- European Education Area. (2025). Empowering learners for the age of AI: launch of the draft AI literacy framework and stakeholder consultations. <https://education.ec.europa.eu/event/empowering-learners-for-the-age-of-ai-draft-ai-literacy-framework-launch>
- Media and Learning Association. (2025). OECD and EC Launch AI Literacy Framework for Schools. <https://media-and-learning.eu/subject/artificial-intelligence/oecd-and-ec-launch-ai-literacy-framework-for-schools/>