



Fight Fakes

Zwischenbericht | Call 18 | Projekt ID 6823

Lizenz: CC BY

Inhalt

1	Einleitung.....	3
2	Status der Arbeitspakete.....	3
2.1	Arbeitspaket 1 - Projektmanagement	3
2.2	Arbeitspaket 2 - Konzeption	3
2.3	Arbeitspaket 3 - Entwicklung.....	6
2.4	Arbeitspaket 4 - Didaktische Einbettung und Verwertung.....	9
3	Zusammenfassung Planaktualisierung	9
4	Öffentlichkeitsarbeit/ Vernetzung.....	9

1 Einleitung

Laut der von Saferinternet.at durchgeführten Studie [„Jugendliche und Falschinformationen im Internet“](#) fällt es 70% der Jugendlichen in Österreich schwer Informationen im Internet zu bewerten. Erst wenn Informationen unglaubwürdig erscheinen, werden sie von ihnen überprüft. Das ist wenig überraschend angesichts der Informationsflut durch TikTok & Co. – es wäre nicht machbar, alle Inhalte zu prüfen. Der Boom generativer Künstlicher Intelligenz verschärft das Problem durch das vermehrte Aufkommen von Deep Fakes.

Gleichzeitig fehlen geeignete deutschsprachige Materialien, um mit Jugendlichen an der bessere Informationsbewertung in Zeiten sozialer Medien und Deep Fakes zu arbeiten. Das Ziel des vorliegenden Projekts ist diese Lücke zu schließen, indem eine Lernumgebung entwickelt wird, in der Jugendliche spielerisch lernen, Informationen schnell und intuitiv zu bewerten.

In der ersten Projekthälfte wurden vorhandene Spielumgebungen und Tools zur Erstellung von Deep Fakes getestet und evaluiert sowie erste Prototypen erstellt und validiert. Diese Recherchen und Evaluationen bilden die Grundlage für das erstellte Konzept, das Lastenheft und den Start der Entwicklung durch unseren Partner. In der anstehenden zweiten Projekthälfte liegt der Schwerpunkt auf der Entwicklung der Spielumgebung und der Erstellung von Spielen.

2 Status der Arbeitspakete

2.1 Arbeitspaket 1 - Projektmanagement

Im Arbeitspaket 1 „Projektmanagement“ wurde Formales zum Projektstart erfolgreich erledigt. Darunter fallen die Vertragsunterzeichnung, die Erstellung einer Projektergebnis-Liste (inkl. Lizenz), die Inbetriebnahme des Projektblogs und die Beantragung der ersten Förderrate.

Bislang wurden zwei Blogeinträge erstellt – der nächste Blogeintrag wird im Juni veröffentlicht. Insgesamt sind im Projektverlauf 6 Blogeinträge geplant. Die erste Projekthälfte legte den Schwerpunkt auf die Konzeptionsarbeiten und der Vorbereitung der Entwicklung der Lernumgebung mit Lernszenarien sowie auch der Einbettung. Aktuell liegen die verbrauchten Kosten noch knapp unter 50% - es ist allerdings aufgrund der für die zweite Projekthälfte vorgesehenen Aufgaben mit einem deutlich höheren Aufwand für das Projektteam zu rechnen.

2.2 Arbeitspaket 2 - Konzeption

Das Arbeitspaket 2 „Konzeption“ konnte bereits erfolgreich abgeschlossen werden, folgende Tätigkeiten wurden durchgeführt: (1) Evaluation bestehender Tools, (2) Erstellung Prototyp, (3) Erstellung Deepfake-Quiz, (4) Erstellung Konzept.

- Evaluation vorhandener Tools

Ziel dieses Tasks war es, vorhandene Spielumgebungen, die unter anderem im Schulunterricht eingesetzt werden, zu sichten und zu testen, um darauf aufbauend Anforderungen für die Entwicklung einer eigenen Spielumgebung definieren zu können. Dabei konnte sowohl auf die Erfahrungen des Projektteams selbst als auch auf jene von saferinternet.at - Trainer:innen aufgebaut werden. Nach ersten Tests wurden einige Tools aussortiert, detailliert evaluiert wurden die Tools Kahoot! (kahoot.it), Riddle (riddle.com) und Pixelshakes (pixelshakes.com).

- Erstellung Prototyp

Die Erkenntnisse der evaluierten Spielmechaniken flossen in einen ersten Prototyp, der sowohl die User Journey der Lehrer:innen als auch die User Journey der Schüler:innen abbildet. Dabei wurde (ähnlich der Mechanik von Kahoot!) darauf geachtet, dass der Einstieg für die Lehrer:innen möglichst einfach gehalten wird und neben einer einmaligen Registrierung lediglich ein QR-Code gescannt werden muss, um Jugendliche einer Klasse gegeneinander antreten lassen zu können.



Abb.1 Screenshot der User-Journey für Lehrer:innen

Außerdem bildet der Prototyp ab, wie die Lernumgebung auf saferinternet.at eingebettet werden kann. Die User-Journey der Schüler:innen gibt einen ersten Einblick, wie der Content in alltagsnahe Lernszenarien (siehe AP3) eingebettet werden kann und welche Spielmechaniken zur Bewertung des Contents notwendig sind.

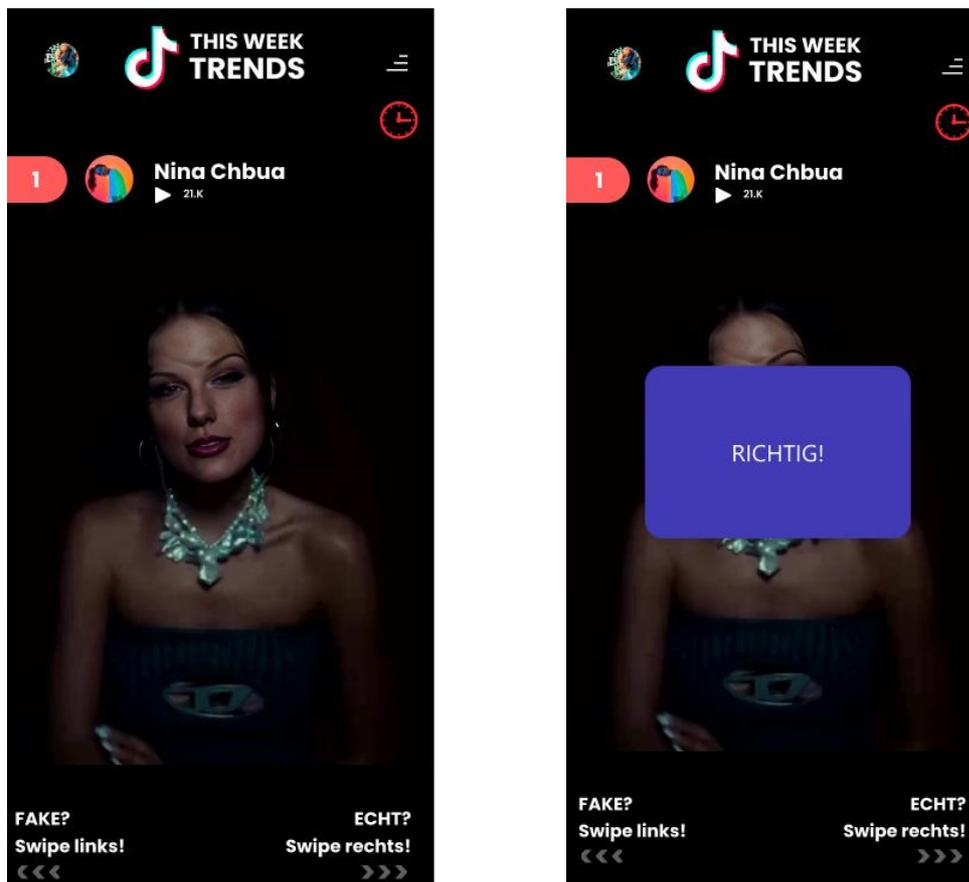


Abb.2 Screenshot der User-Journey für Schüler:innen

- Erstellung Deepfake-Quiz

Um nicht nur Spielmechaniken und User Journeys, sondern auch geeignete Inhalte für die Spielumgebung testen und konzipieren zu können, wurde bereits ein Deepfake-Quiz erstellt – das außerhalb der zu entwickelten Lernumgebung läuft. Dafür wurden geeignete Deepfake-Tools (Gen-AI Tools) recherchiert und evaluiert, datenschutzrechtliche Fragen geklärt (sowohl in Bezug auf die Tools als auch die zu erstellenden Inhalte) und schließlich Deepfakes erstellt (Bilder/Audio/Videos).

Dies lieferte wichtige Erkenntnisse für den weiteren Verlauf des Projekts. Unter anderem zeigt sich, dass das Ziel nicht sein kann Jugendliche nur beizubringen, zu erkennen, ob ein Inhalt mit KI generiert wurde. Vielmehr muss es auch darum gehen, dass sie die Inhalte bewerten können.

- Erstellung Konzept

Der Prototyp sowie das evaluierte erste Quiz¹ stellten die Grundlage für weitere Evaluierungen dar, welche gemeinsam mit den bisherigen Recherchen und Erkenntnissen in ein Konzept für die Spielumgebung einfließen. Die Spezifikation der technischen Anforderungen wurde im Austausch mit dem technischen Partner bytewood erarbeitet. Für die didaktischen Spezifikationen konnten auf die Expertise von saferinternet.at-Trainer:innen und Lehrer:innen zurückgegriffen werden.

¹ Das Quiz wurde bereits von Saferinternet.at-Trainer:innen und einzelnen Lehrer:innen getestet.

2.3 Arbeitspaket 3 - Entwicklung

Das Arbeitspaket 3 „Entwicklung“ umfasst folgende Tätigkeiten: (1) Entwicklung von Content zum Einsatz in Lernszenarien, (2) Entwicklung von situierten Lernszenarien, (3) Entwicklung der Lernumgebung.

- Entwicklung von Content

Die Entwicklung von Content basiert auf dem Austausch mit Lehrer:innen und Saferinternet.at und umfasst eine Sammlung von Fakes, Deep Fakes und echtem Content. Von zentraler Bedeutung ist dabei, dass der Content einerseits eigenständig verwendet werden kann, bspw. auf Social Media, und andererseits in unterschiedlichen Kombinationen in die Lernszenarien und somit in die Spielumgebung eingebettet werden kann. Der Content soll nicht nur nah am Alltag der Jugendlichen sein, sondern muss auch von den Lehrkräften als sinnvoll erachtet werden. Daher wurden als Orientierung zur Erstellung des Contents die übergreifenden Themen des Lehrplans für Mittelschulen genutzt. Dazu zählen:

- Sprachen
- Mathematik und Naturwissenschaften
- Wirtschaft und Gesellschaft
- Musik, Kunst und Kreativität
- Gesundheit und Bewegung

Die vorgenommene Einteilung erlaubt es, sowohl den Realitäten der Schüler:innen (beispielsweise Fitnessvideos im Bereich Gesundheit und Bewegung, Musikvideos im Bereich Kunst und Kreativität) als auch den Bedürfnissen der Lehrer:innen gerecht zu werden. Letztere können auf diese Weise für ihr Fach relevante Inhalte finden.

Erstellt wurden bislang Video-, Audio- und Bilddateien. Es zeigt sich, dass die Qualität von Audio und Bild teilweise sehr gut ist, allerdings das Medium Video noch in der Anfangsphase steht. Laufend werden auch Beispiele von Deepfakes gesammelt, die auf Social Media kursieren.



Abb. 3 Deep Fake Bild erstellt mit Dall-E 3 (Prompt: „A photorealistic black and white image capturing a moment in Victorian England. A woman dances gracefully in front of the Crystal Palace, her Victorian dress swirling around her. She is uniquely wearing over-ear headphones, a stark contrast to her period attire. The scene is lively, with other pedestrians in the background, amidst trees and meadows. The photograph-style image should evoke a sense of nostalgia and wonder, with a focus on detail and the surreal blend of time periods“)



Abb.4 Screenshot eines Deep Fake Videos mit Billie Eilish, Audio wurde mit HeyGen erstellt, anschließend wurden die Lippen mit dem Fake-Audio synchronisiert (Wav2Lip)

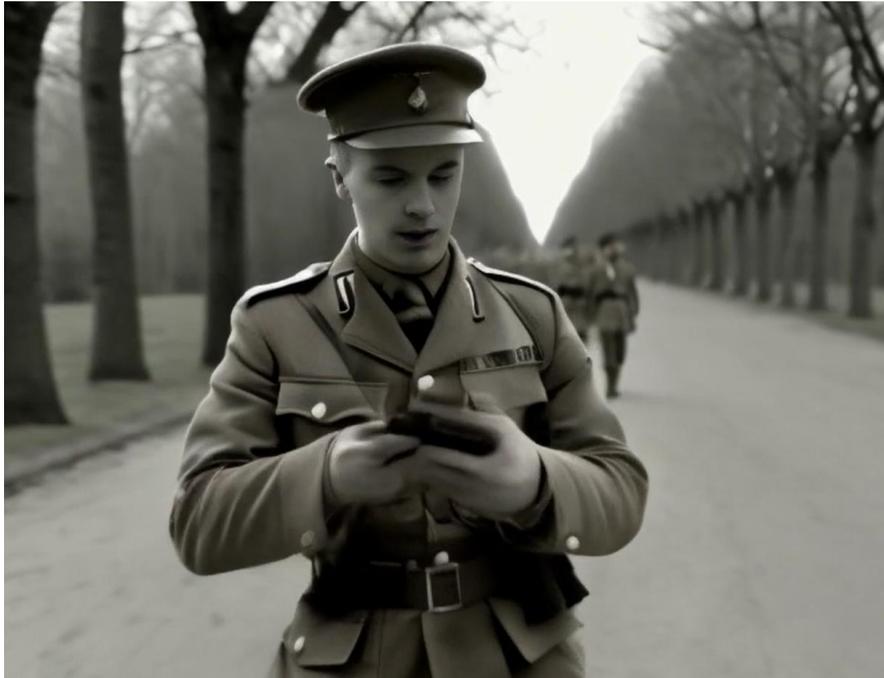


Abb.5 Screenshot eines Videos, das mit Hilfe von Sora erstellt wurde (Prompt: „Create a black and white video of a First World War soldier walking down an avenue, holding a smartphone and looking at it.”)

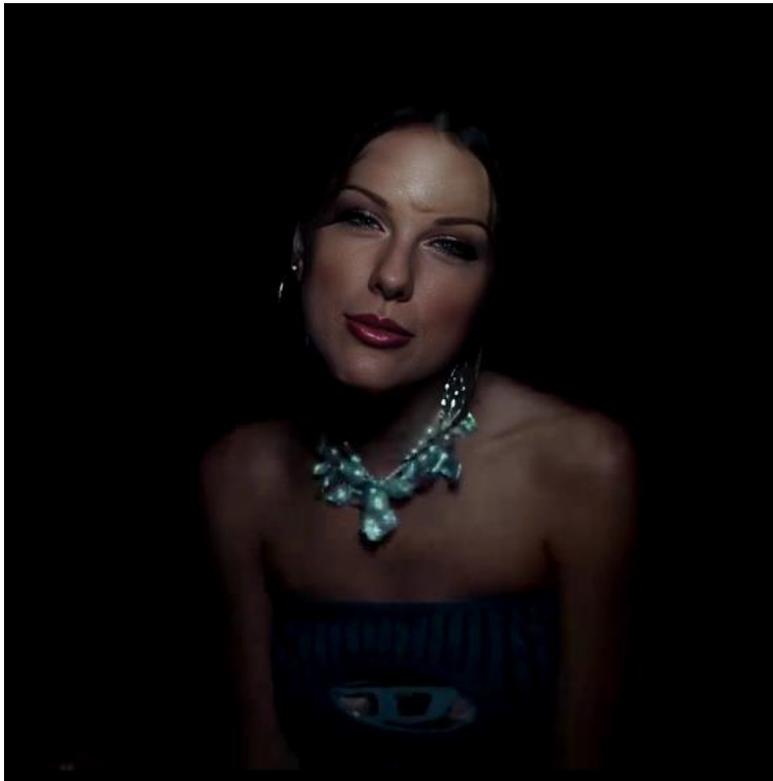


Abb.6 Screenshot eines Deep Fake Videos, Face Swap mit Hilfe von artguru.ai (Ausgangsvideo: Nina Chuba, ausgetauschtes Gesicht: Taylor Swift)

- Weitere Entwicklungen

Es konnte in der ersten Projekthälfte bereits der Auftrag für die Entwicklung der Lernumgebung erteilt werden. Plan ist eine Entwicklung von Umgebung und Szenarien bis in den Herbst 2024. Bis dahin werden Content und Begleitmaterialien erstellt. Offen ist noch, wie die Modularität der Übungen inkl. Lernpfade gestaltet werden kann – Das Freischalten von Elementen über die Absolvierung von Übungen könnte eventuell nicht technisch umgesetzt werden, sondern im Sinne einer Selbsteinschätzung durch die Lehrpersonen. Dies wird im Laufe der zweiten Projekthälfte iterativ anhand von Feedback durch Lehrer:innen und pädagogische Expert:innen sowie dem technischen Partner entschieden.

2.4 Arbeitspaket 4 – Didaktische Einbettung und Verwertung

In diesem Arbeitspaket soll Content erstellt werden, der ergänzend zu vorhandenem Saferinternet.at Material, Lehrer:innen bei der Umsetzung unterstützt. Es sollen Materialien auch für Schüler:innen verständlich sein. Dieses Arbeitspaket startet laut Plan im August 2024.

3 Zusammenfassung Planaktualisierung

Es sind keine Aktualisierungen im Zeitplan geplant. Beim Subauftrag für die Entwicklung der Lernumgebung wurden Drittkosten umgeschichtet – geplant waren 11.000 Euro für die Entwicklung und 3.000 Euro für Beratung zur Konzeption der Entwicklungsarbeiten. Die enge Kooperation mit dem Subauftragnehmer Bytewood umfasst mehr als nur ein Auftrag für Entwicklung – und entsprechend ist nicht damit zu rechnen, dass weitere Konzeptionskosten für Externe zu tragen kommen. **Es wurde ein Subauftragung abgeschlossen mit ihnen in der Höhe von 13.200 Euro exkl. USt.**

4 Öffentlichkeitsarbeit/ Vernetzung

Das Gros der Öffentlichkeitsarbeit wird bei Veröffentlichung der Lernumgebung stattfinden – Ziel ist die Einbettung der Spiele auf Saferinternet.at.

Es fand bereits ein erstes Austauschtreffen mit BAIT statt – dem Team des Faktencheck-Kanal für Jugendliche auf TikTok. Geplant ist, sobald Ergebnisse vorliegen, dies auch mit Partnern des ÖIAT bzw. Saferinternet.at zu verbreiten, wie z. B. Mimikama oder das BIK-Netzwerk (Netzwerk der Saferinternet-Zentren in der Europäischen Union).