



netidee

PROJEKTE

Fight Fakes

Endbericht | Call 18 | Projekt ID 6823

Lizenz: CC BY 4.0

Inhalt

1	Einleitung.....	3
2	Projektbeschreibung.....	3
3	Status der Arbeitspakete.....	4
3.1	Arbeitspaket 1 - Projektmanagement.....	4
3.2	Arbeitspaket 2 - Konzeption.....	4
3.3	Arbeitspaket 3 - Entwicklung.....	7
3.4	Arbeitspaket 4 – Didaktische Einbettung und Verwertung.....	13
3.5	Arbeitspaket 5 – Dokumentation und Formales am Projektende.....	14
4	Umsetzung Förderauflagen.....	14
5	Liste Projektendergebnisse.....	15
6	Verwertung der Projektergebnisse in der Praxis.....	16
7	Öffentlichkeitsarbeit/ Vernetzung.....	16
8	Eigene Projektwebsite.....	16
9	Geplante Aktivitäten nach netidee-Projektende.....	17
10	Anregungen für Weiterentwicklungen durch Dritte.....	17

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, überprüfen sie diese. Das ist wenig überraschend, denn angesichts der Informationsflut durch TikTok & Co. wäre es nicht machbar, alle Inhalte zu prüfen. Der Boom generativer Künstlicher Intelligenz verschärft das Problem durch das vermehrte Auftreten von Deep Fakes.

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

Hier setzen wir mit dem Projekt „Fight Fakes“ an und konnten im Laufe der Projektzeit eine quizähnliche Spielumgebung entwickeln, in der Jugendliche spielerisch lernen, Informationen schnell und intuitiv zu bewerten und Lehrkräften ein einfach zu bedienendes Werkzeug an die Hand zu geben, um das Thema spielerisch im Unterricht zu behandeln.

2 Projektbeschreibung

„Fight Fakes“ ist ein Lernspiel, das Jugendliche darin unterstützt, ihr Bauchgefühl zu trainieren und digitale Desinformation besser zu erkennen. Es kombiniert aktuelles Wissen über Deep Fakes mit praxisnahen Übungen in einer spielbasierten Lernumgebung.

Die simulierten Szenarien orientieren sich visuell und funktional an gängigen Social-Media-Plattformen wie TikTok und SnapChat. In diesen Szenarien üben Schüler:innen, Inhalte wie Videos, Bilder und Texte zu bewerten.

Die Aufgaben greifen reale Alltagssituationen auf, um den Wissenstransfer in die digitale Lebenswelt der Jugendlichen zu erleichtern. Neben dem visuellen Aspekt zeigt sich das auch in der Verwendung eines Timers, der die Dynamik sozialer Netzwerke widerspiegelt, in denen Entscheidungen oft in Sekundenschnelle getroffen werden. Der simulierte Zeitdruck im Spiel trainiert das Bauchgefühl der Spieler:innen und sensibilisiert sie für den bewussten Umgang mit Informationen im Internet.

3 Status der Arbeitspakete

Das Projekt musste aufgrund interner Personalwechsel und durch Verzögerungen im Entwicklungsprozess um 2 Monate kostenneutral verlängert werden. Die Änderung wurde vom netidee-Vorstand freigegeben. Im Laufe des Projekts stellte sich heraus, dass die Entwicklung am meisten Ressourcen benötigte, daher wurde mit dem AP3 bereits früher als angedacht begonnen und lief bis Ende März, da bis zum Ende der Projektlaufzeit Spiele entwickelt, evaluiert und korrigiert wurden. Das AP2 benötigte dafür weniger Stunden. Auch AP4 lief bis ans Ende der Projektlaufzeit, da die neuen Spiele auch in die Anwendungsdokumentation für Lehrkräfte sowie auf Saferinternet.at eingebettet werden mussten.

Die unterschiedlichen Personalwechsel bedingten auch eine Änderung im Team. So verließ Louise Beltzung das ÖIAT mit Ende Dezember, die Projektleitung wurde von Valentine Auer übernommen. Als Junior Researcher wurde im Jänner Natalie Trell eingestellt, die einen großen Teil der Game-Entwicklung übernahm. Die ebenso neu am ÖIAT eingestellten Personen Iris Strasser und Hannah Reindl unterstützten insbesondere beim Lektorat der Games. Außerdem wurden aufgrund des Personalwechsels zusätzliche Stunden von Corina Kindler (Back Office Managerin) und Bernhard Jungwirth (Geschäftsführer) notwendig, um Valentine Auer als neue Projektleitung zu unterstützen.

3.1 Arbeitspaket 1 - Projektmanagement

Im Arbeitspaket 1 „Projektmanagement“ wurden die formalen Schritte zum Projektstart und für den Projektabschluss erfolgreich durchgeführt. Dazu gehören die Vertragsunterzeichnung, die Erstellung der Projektergebnisliste (inkl. Lizenz), die Inbetriebnahme des Projektblogs, die Beantragung der Förderrate sowie das Verfassen des Zwischenberichts.

3.2 Arbeitspaket 2 - Konzeption

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

- Evaluation vorhandener Tools

Ziel dieses Tasks war es, bestehende 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 zu lassen.



Abbildung 1: Screenshot der User-Journey für Lehrer:innen

Der Prototyp zeigt auch, 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) verortet werden kann und welche Spielmechaniken zur Bewertung des Contents notwendig sind.

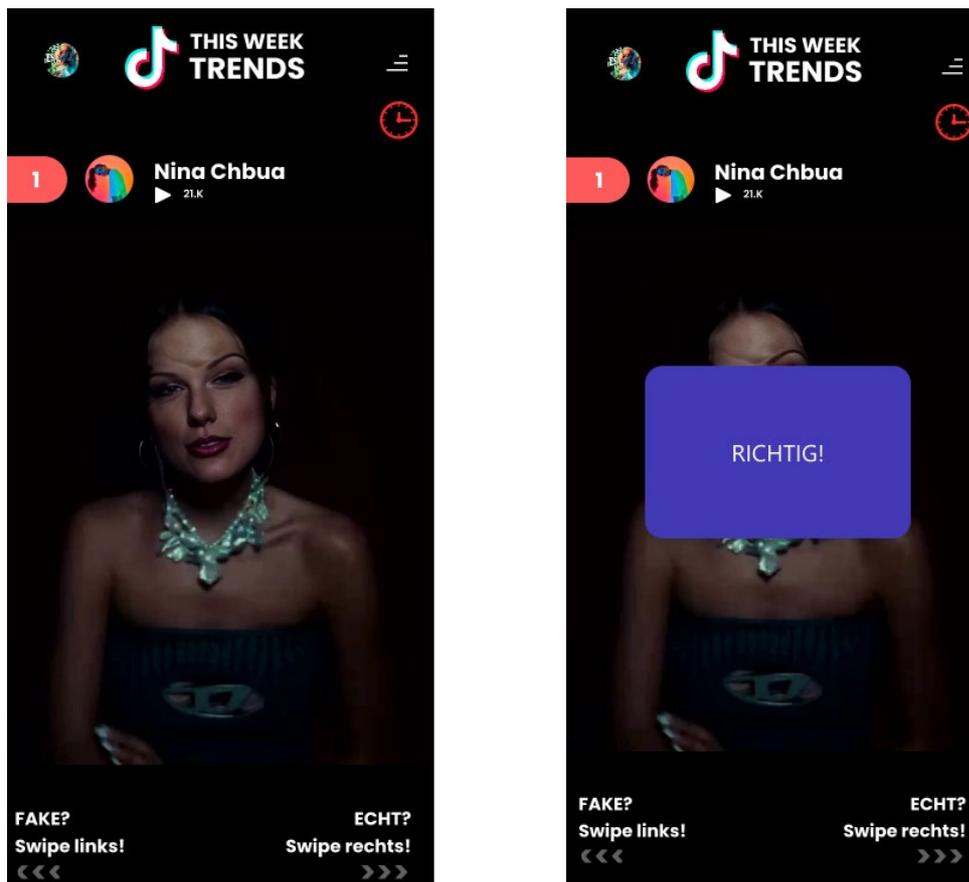


Abbildung 2: Screenshot der User-Journey für Schüler:innen

- 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 (Bilder/Audio/Videos) erstellt.

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 erste evaluierte Quiz¹ bilden die Grundlage für weitere Evaluierungen an, 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.

3.3 Arbeitspaket 3 - Entwicklung

Das Arbeitspaket 3 „Entwicklung“ umfasst folgende Tätigkeiten: (1) Entwicklung von Content zum Einsatz in Lernszenarien, (2) Entwicklung der situierten Lernszenarien, (3) Entwicklung der Lernumgebung und (4) Entwicklung der Spiele. Während der gesamten Entwicklungszeit wurden Feedbackschleifen sowohl mit Jugendlichen als auch mit Lehrenden durchgeführt, um möglichst nahe an der Zielgruppe entwickeln zu können.

- Entwicklung von Content

Die Entwicklung von Content basiert auf dem Austausch mit Lehrer:innen und Expert:innen von 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 auch von den Lehrkräften als sinnvoll erachtet werden. Daher wurden in einem ersten Schritt die übergreifenden Themen des Lehrplans für Mittelschulen als Orientierung für die Erstellung der Inhalte herangezogen. Diese sind:

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

Insgesamt wurden im Laufe des Projekts rund 250 Deepfakes (Video, Audio- und Bilddateien) erstellt, die jedoch bei weitem nicht alle in den Spielen verwendet wurden. Während vor allem die Qualität der erstellten Bildern sehr gut ist, befindet sich das Medium Video noch in einem früheren Entwicklungsstadium.² Dennoch waren sowohl bei den Bildern als auch bei den Videos zahlreiche Iterationen und Promptanpassungen notwendig, um zum gewünschten Ergebnis zu gelangen. Die Ergebnisse mehrerer Iterationen, die sich u.a. in der Qualität bzw. „Authentizität“ unterscheiden, konnten zwar teilweise für verschiedene Kompetenzstufen verwendet werden, am Ende wurde jedoch eine Auswahl der „besten“ Ergebnisse pro Spiel getroffen. Der Großteil des erstellten Contents wurde unter folgendem Link Open Source zur Verfügung gestellt:

<https://cloud.oiat.at/s/YS3Yj5jPXsqNoD4>

² siehe auch Blog-Eintrag zu Deep Fails: <https://www.netidee.at/fight-fakes/deepfails-mit-pica-10-dream-machine-co>



Abbildung 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“)



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

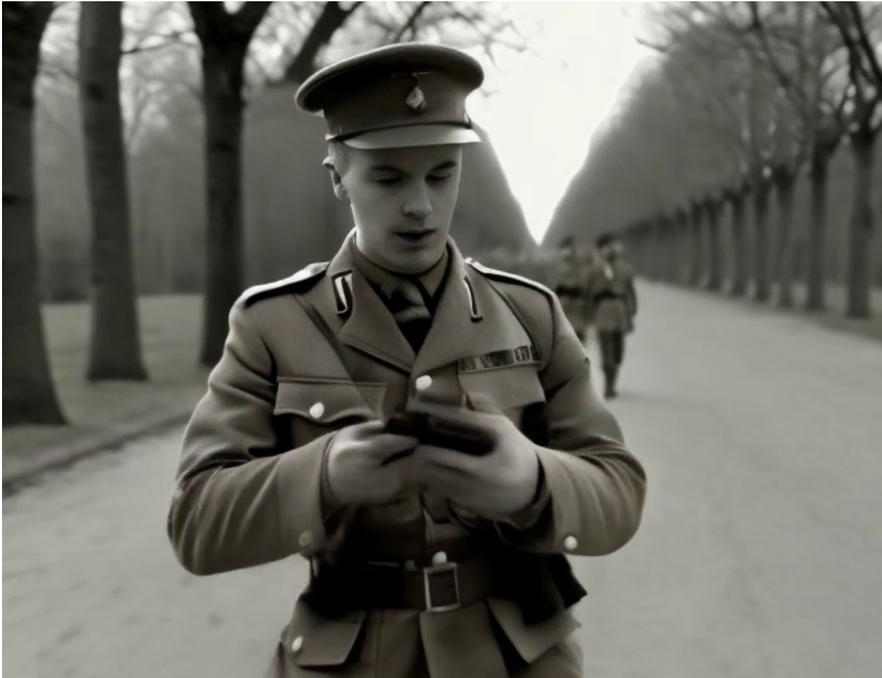


Abbildung 5: Screenshot eines Videos, das mit Hilfe von Sora erstellt wurde (Prompt: „Create a black and white video of a First World War soldier walking down an avenue, holding a smartphone and looking at it.“)

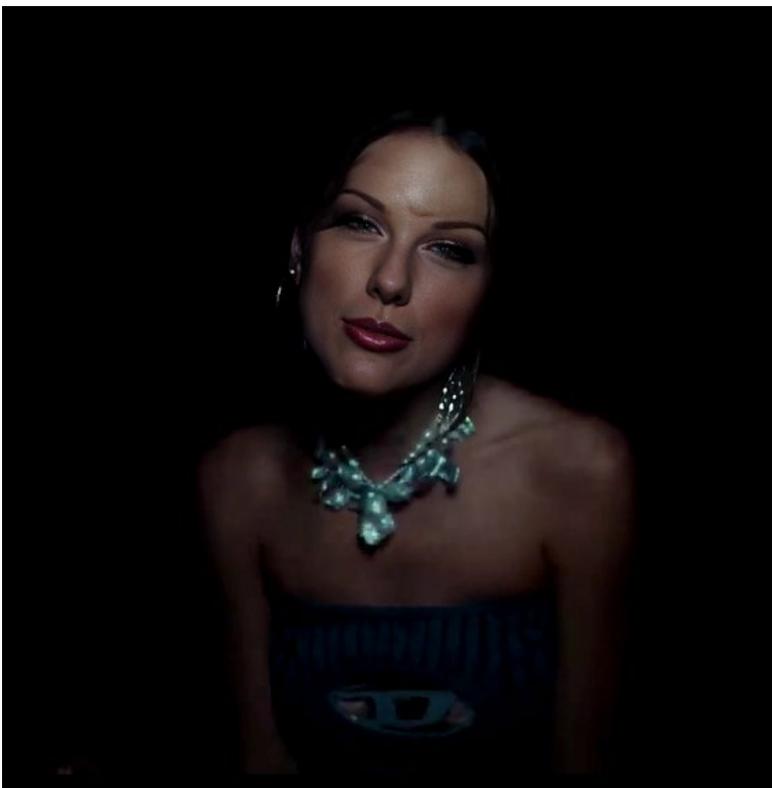


Abbildung 6: Screenshot eines Deep Fake Videos, Face Swap mit Hilfe von artguru.ai (Ausgangsvideo: Nina Chuba, ausgetauschtes Gesicht: Taylor Swift)

Darüber hinaus musste auch „echter Content“, der nicht von einer KI erstellt wurde, recherchiert werden. Dazu wurden verschiedene Datenbanken für Stock-Bilder und Stock-Videos durchsucht. Je nach Thema war es jedoch relativ schwierig, insbesondere Videos zu finden, die ästhetisch und inhaltlich zu den jeweiligen Spielen passen.

- Entwicklung situierter Lernszenarien

Nicht nur die Inhalte selbst, sondern auch die Umgebung, in der die Inhalte präsentiert werden, sollte sich an dem Alltag der Zielgruppe orientieren. Daher wurden situierte Lernszenarien entwickelt, die ein Lernen in vertrauter Umgebung ermöglichen.

Die situierten Lernszenarien waren als Softwaremodul konzipiert. In Gesprächen mit Lehrenden stellte sich jedoch heraus, dass die Szenarien - ebenso wie die Inhalte - möglichst niederschwellig und einfach bedien- und anpassbar zur Verfügung gestellt werden sollten. Daher wurde entschieden, die Lernszenarien als editierbare Dateien zu erstellen.

Entlang der von Jugendlichen am häufigsten genutzten Social-Media-Plattformen wurden die Social-Media-Timelines von TikTok und Snapchat imitiert. Im Cloud-Ordner von Fight Fakes wurden jeweils Hintergründe im Stil von Snapchat und TikTok sowie Templates der beiden Plattformen als Photoshop- und Premiere-Dateien für die Bild- und Videobearbeitung zur Verfügung gestellt (<https://cloud.oiat.at/s/7ZPXkLPCDH3f9Gk>). Darüber hinaus werden in den jeweiligen Themenordnern die in den Lernszenarien eingesetzten Inhalte (aufgeschlüsselt nach Kompetenzstufen) zur Verfügung gestellt.

Da die Lernszenarien nicht als Softwarelösung angeboten wurden, wurden die geplanten Interaktionselemente erst bei der Entwicklung der Lernumgebung selbst implementiert. Außerdem hat sich die Lizenz dadurch von MIT License auf CC BY 4.0 geändert.

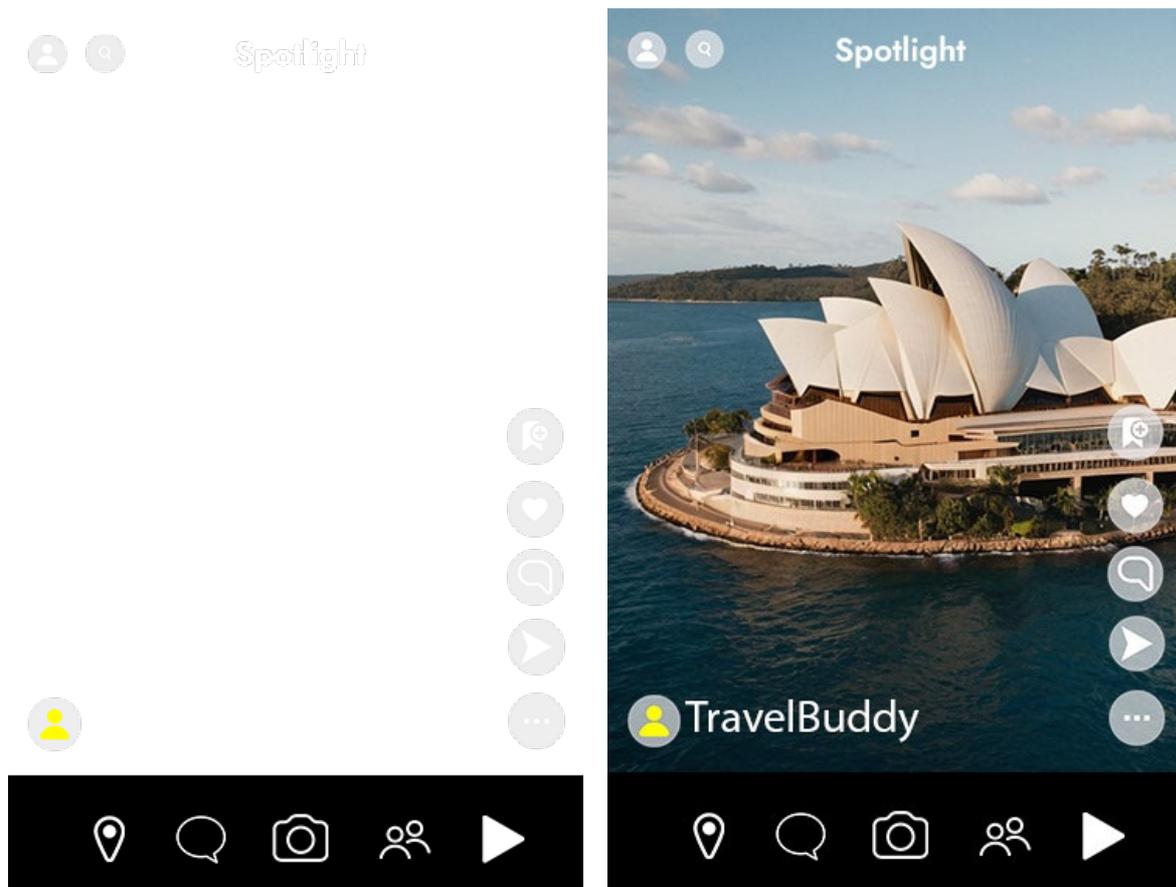


Abbildung 7: Snapchat-Vorlage mit (re) und ohne (li) Content

- Entwicklung Lernumgebung

Auf Basis des in AP2 entwickelten Prototyps und Konzepts wurde ein Subauftrag an das Wiener Unternehmen bytewood zur Entwicklung der Lernumgebung bzw. der Spiele vergeben. Es wurden zwei Open Source Games entwickelt und auf github veröffentlicht – das Swipe Game (<https://github.com/oiat/SwipeGame>) und das Chat Game (<https://github.com/oiat/ChatGame>).

Das Swipe Game imitiert eine ähnliche User Experience wie Social Media Timelines, indem ein Bild nach dem anderen angezeigt wird. Dabei wurden folgende Interaktions- und Gamificationelemente integriert: Auf mobilen Geräten (Smartphones, Tablets) können die Spieler:innen mit einem Swipe entscheiden, ob es sich um ein Fake handelt (Swipe nach rechts) oder nicht (Swipe nach links) - zusätzlich gibt es die Möglichkeit, mit Emoji zu reagieren. Für diese Entscheidungen haben die Spieler:innen jedoch nur kurz Zeit (die Zeit ist je Spiel anpassbar). Nach jedem angezeigten Bild oder Video wird eine kurze Auflösung angezeigt und am Ende jedes Spiels können Highscore und Weiterleitungen angezeigt werden.

Während das Swipe-Game von den Spielenden eine schnelle und duale Fake/Nicht-Fake-Entscheidung verlangt, ermöglicht das Chat-Game eine tiefere Auseinandersetzung mit den Themen Deepfakes und Fake-News. Simuliert werden beim Chat-Game Messenger-Apps wie WhatsApp, in denen Textnachrichten, Bilder und Videos verschickt werden können. Die Möglichkeiten bezüglich der zeitlichen Begrenzung sowie der Anzeigen am Ende des Spiels sind

identisch mit dem Swipe Game. Als zentrales Interaktionselement steht den Spielenden die Auswahl von Antworten (Text, Bild, Videos – auch kombiniert) zur Verfügung.

- Entwicklung Games

Insgesamt wurden 18 Swipe-Games in 6 Kategorien mit jeweils 3 unterschiedlichen Kompetenzlevels (EASY, PRO und PRO EXTREME) erstellt. Nach jedem EASY und PRO Spiel wird auf das nächsthöhere Kompetenzlevel verwiesen, nach jedem PRO EXTREME Spiel wird auf ein Lernmaterial oder ein nützliches Tool verlinkt.

Bei der Entwicklung lag der Fokus auf dem Swipe-Game, da es der ursprünglichen Projektidee entspricht, Social Media Timelines nachzubilden und eine intuitive und schnelle Bewertung von Inhalten zu ermöglichen. Ein erstes Chat-Game wurde prototypisch umgesetzt und nur in der Anwendungsdokumentation, nicht aber auf der Projektwebsite veröffentlicht (<https://app.pixelshakes.com/gameviewer.php?gameid=710>). Dabei wurde nicht auf Kompetenzlevels im Sinne von Schwierigkeitsgraden gesetzt, sondern jeweils zwei Spiele pro Thema entwickelt, in der die Spielenden das Game einmal aus der „bösen (evil)“ (bspw. wie erstelle ich ein Deepfake und kann dabei Richtlinien umgehen) und einmal aus der „guten (good)“-Perspektive (bspw. wie entlarve ich Deepfakes) bearbeiten können.

Eine zentrale Schwierigkeit bei der Erstellung der Spiele war die Einschätzung der Kompetenzen. Die Frage, welche Bilder oder Videos leicht als Deepfake erkennbar sind und welche nur schwer von echten Inhalten zu unterscheiden sind, wurde laufend mit unterschiedlichen Expert:innen diskutiert und anhand der von uns erstellten Inhalte in die Kompetenzstufen eingeordnet.

Darüber hinaus war es uns wichtig, bei der Entwicklung der Spiele nicht das Narrativ zu bedienen, dass Deepfakes immer problematisch sind bzw. dass auch nicht KI-generierter Content problematisch und fake sein kann. Daher haben wir einerseits bewusst nicht-emotionale und unproblematische Themen gewählt und andererseits im Themenkomplex Gesundheit nicht nur mit Deepfakes, sondern auch mit Fake News gearbeitet, unabhängig davon, ob das entsprechende Bild oder Video KI-generiert ist oder nicht.

- Planänderungen

Aufgrund der Feedbackschleifen bei der Entwicklung der Spiele und der engen Abstimmung mit Saferinternet.at-Trainer:innen musste die Spielumgebung öfter angepasst werden. So wurde beispielsweise das Erstellen von Instanzen erleichtert, die Möglichkeit zum Einspielen von Musik hinzugefügt und die Highscore-Funktion angepasst.

Durch die zusätzlichen Anpassungen erhöhten sich die externen Kosten von den im Zwischenbericht genannten 13.200 Euro auf 15.400 Euro. Zusätzlich fielen 3.000 Euro für Baits (siehe Arbeitspaket 4) für Beratung und die Erstellung von Begleitmaterial an. Damit erhöhten sich die Entwicklungskosten gegenüber dem Antrag von insgesamt 14.600 Euro exkl. USt auf 18.470 Euro exkl. USt.

Das Projekt konnte dennoch erfolgreich durchgeführt werden, da die Anzahl der entwickelten Games minimal reduziert wurde – vor allem im Hinblick auf die geplanten Aktivitäten nach Projektende von netidee (siehe Kapitel 9) wird die im Antrag geplante Anzahl an Spielen im Laufe des Jahres 2025 auf jeden Fall erreicht werden.

3.4 Arbeitspaket 4 – Didaktische Einbettung und Verwertung

Im Arbeitspaket 4 „Didaktische Einbettung & Verwertung“ wurden (1) Inhalte erstellt, die ergänzend zu den vorhandenen Saferinternet.at Materialien Lehrer:innen beim Einsatz des Spiels im Unterricht unterstützen, (2) Materialien für Schüler:innen erstellt und (3) die Spiele inkl. Begleitmaterialien auf saferinternet.at eingebettet.

Saferinternet.at verfügt bereits über Unterrichtsmaterialien zum Thema Fake News und Deepfakes („Wahr oder falsch im Internet“, „Fakt oder Fake? Hoaxes, bearbeitete Bilder und Deepfakes im Internet“), einen News-Artikel zum Thema Deepfakes („Deepfakes – Wie Foto-, Video und Audioinhalte manipuliert werden!“) sowie zu zum Thema Informationskompetenz. Diese Materialien wurden zusammen mit den entwickelten Games als Begleitmaterial auf einer eigenen Projektwebsite auf saferinternet.at veröffentlicht.

Zusätzlich wurden folgende Materialien und Texte erstellt:

- **Kurzanleitung:** Damit Lehrer:innen das Spiel auch kurzfristig und ohne lange Vorbereitung im Unterricht einsetzen können, wurde den Games eine Kurzanleitung vorangestellt.
- **Anwendungsdokumentation:** Eine ausführliche Anleitung inkl. Hintergrund zum Spiel, Informationen zu den Kompetenzlevels, Begleit- und Bonusmaterial sowie eine Level-Übersicht finden sich in der Anwendungsdokumentation.
- **Unterrichtsbeispiel:** Fight Fakes kann sowohl als schnelles Spiel in z.B. Supplierstunden eingesetzt werden, als auch als Möglichkeit zur vertiefenden Auseinandersetzung mit dem Thema. Zur Gestaltung einer solchen Unterrichtseinheit wurde ein Unterrichtsbeispiel erstellt.
- **Erklärvideo:** Der Faktencheck-Kanal auf TikTok, Baits, erstellte im Rahmen eines Subauftrags ein Erklärvideo für Schüler:innen zum Thema „Deepfakes erkennen“, das ebenfalls auf der Projektseite veröffentlicht wurde.
- **Faktenchecks:** Zusätzlich wurden von Baits 10 Faktencheck-Videos zur Verfügung gestellt, in denen Alltagsmythen, die in sozialen Medien kursieren, entlarvt werden. Die Videos dienen zusammen mit den anderen Begleitmaterialien als Grundlage, um mit Jugendlichen über Fakes zu sprechen.

- Letzte Feedbackrunde

Am Ende des Projektes wurde noch eine letzte Feedbackrunde von Lehrer:innen gesammelt. Während in den vorangegangenen Interviews und Feedbackrunden die technische Umsetzung der Spielumgebung, die inhaltliche Ausrichtung des Spiels sowie die Kompetenzstufen getestet wurden, ging es in der letzten Feedbackrunde darum, Feedback von Lehrer:innen zu erhalten, die

das Spiel im Unterricht einsetzen und dafür keine weiteren Erklärungen oder Vorgaben als die Projektseite erhalten. Ziel war es, einerseits zu überprüfen, ob die Informationen auf der Projektseite ausreichen, um das Spiel sinnvoll im Unterricht einsetzen zu können, und andererseits, ob das Spiel selbst für den Unterricht geeignet ist.

Das Ergebnis zeigt, dass beide Fragen positiv beantwortet werden konnten. Die Testpersonen hatten keine Fragen zum Einsatz des Spiels im Unterricht, gleichzeitig wurden die Fragen nach der Bewertung u.a. der Benutzerfreundlichkeit, der Auswahl der Inhalte und Themen, der Spielmechanik mit gut oder sehr gut beantwortet und auch die Frage, ob Fight Fakes auch in Zukunft im Unterricht eingesetzt werden wird, wurde zu einem großen Teil positiv beantwortet.

3.5 Arbeitspaket 5 – Dokumentation und Formales am Projektende

Im Arbeitspaket 5 „Dokumentation und Formales am Projektende“ wurden die formalen Schritte zum Projektabschluss erfolgreich durchgeführt. Dazu gehören die Endabrechnung, das Erstellen der Entwicklerdokumentation, des Endberichts und des Summaries sowie das Verfassen eines abschließenden Blogbeitrags und das Übermitteln der entsprechenden Dokumente.

4 Umsetzung Förderauflagen

Es wurden keine speziellen Förderauflagen festgelegt.

5 Liste Projektergebnisse

1	Projektzwischenbericht	CC BY 4.0	https://www.netidee.at/fight-fakes
2	Projektendbericht	CC BY 4.0	https://www.netidee.at/fight-fakes
3	Entwickler_innen- DOKUMENTATION	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-03/netidee_call18_6823_Entwicklerdokumentation.pdf
4	Anwender_innen- DOKUMENTATION	CC BY 4.0	https://www.netidee.at/sites/default/files/2025-03/netidee_call18_6823_Anwendungsdokumentation.pdf
5	Veröffentlichungsfähiger Einseiter / Zusammenfassung	CC BY 4.0	https://www.netidee.at/sites/default/files/2024-06/netidee_call18_6823_Konzept_FightFakes.pdf
6	Dokumentation Externkommunikation (Teil des Endberichtes)	CC BY 4.0	https://www.netidee.at/fight-fakes
7	Konzept Lernumgebung	CC BY 4.0	https://www.netidee.at/sites/default/files/2024-06/netidee_call18_6823_Konzept_FightFakes.pdf
8	Content	CC BY 4.0	https://www.netidee.at/fight-fakes https://cloud.oiat.at/s/YS3Yj5jPXsqNoD4
9	Situierte Lernszenarien	CC BY 4.0	https://www.netidee.at/fight-fakes https://cloud.oiat.at/s/YS3Yj5jPXsqNoD4
10	Swipe-Game	MIT License	https://www.netidee.at/fight-fakes https://github.com/oiat/SwipeGame
11	Chat-Game	MIT License	https://www.netidee.at/fight-fakes https://github.com/oiat/ChatGame
12	Projektseite, inkl. Lernumgebung und Begleitmaterialien	CC BY 4.0	https://www.saferinternet.at/projekte/fight-fakes

6 Verwertung der Projektergebnisse in der Praxis

Die Games wurden und werden auch weiterhin aktiv über das Netzwerk von Saferinternet.at beworben, um die Projektergebnisse zu verwerten: In einem ersten Schritt wurde das Trainer:innen-Netzwerk aktiviert, um die Spiele in Workshops (bspw. an Schulen oder in Jugendzentren) einzusetzen. Derzeit wird das Spiel unter Lehrer:innen verbreitet, vor allem über verschiedene Newsletterverteiler, die Lehrer:innen erreichen.

Darüber hinaus konnte das Projektteam Erkenntnisse über Deepfakes gewinnen, die auch in anderen Forschungsprojekten Anwendung finden. Die zahlreichen iterativen Generationen von Deepfakes gaben dem Team laufend einen für das ÖIAT zentralen Einblick in den Stand generativer KI, in die Erstellung von Prompts, in mögliche Biases sowie in Richtlinien (und wie diese umgangen werden können) seitens der Tools.

7 Öffentlichkeitsarbeit/ Vernetzung

Die Öffentlichkeitsarbeit wird erst nach Projektende beginnen, da wir bis zum Projektende noch Feedback von den verschiedenen Zielgruppen einholen wollten, um dieses in mögliche weitere Projekte einfließen lassen zu können (siehe Punkt 9). In diesem Zusammenhang ist insbesondere die Nutzung von Social Media Kanälen durch die Vernetzung mit Baits geplant. So sollen die Faktenchecks über Saferinternet-Kanäle verbreitet und das Spiel beworben werden.

Geplant ist die Bewerbung des Games mit Hilfe der von Baits zur Verfügung gestellten Faktenchecks. Über Instagram erreicht Saferinternet.at rund 4.500 Personen. Außerdem werden unterschiedliche Newsletter genutzt, um das Spiel zu verbreiten – u.a. der allgemeine Saferinternet.at Newsletter mit knapp 10.000 Abonnent:innen, sowie Newsletter, die sich speziell an Schulen und Partner:innen richten (insgesamt etwa 2.100 Abonnent:innen). Außerdem wird eine News auf Saferinternet.at veröffentlicht, die Website hat monatlich rund eine halbe Million Seitenaufrufe und erreicht dabei vor allem Lehrkräfte und Eltern.

Darüber hinaus wird das Spiel auch im Zuge von den weiteren geplanten Aktivitäten in anderen Forschungsprojekten (siehe Kapitel 9) verwendet und beworben werden. Insbesondere die Projekte TGuard und defame Fakes stoßen aktuell auf ein großes Medieninteresse, das im Zuge der Dissemination der erstellten Games genutzt werden kann.

8 Eigene Projektwebsite

Die Spiele wurden als Angebot für Lehrkräfte auf saferinternet.at eingebettet. Um auch die Begleitmaterialien hinzufügen zu können, wurde eine eigene Projektseite erstellt:

<https://www.saferinternet.at/projekte/fight-fakes>

9 Geplante Aktivitäten nach netidee-Projektende

Neben der bereits erwähnten Öffentlichkeitsarbeit und der Verwertung der Projektergebnisse innerhalb der Zielgruppe (Saferinternet-Trainer:innen sowie Lehrkräfte) können wir im Zuge verschiedener aktuell laufender Projekte weitere Games erstellen und somit „Fight Fakes“ in verschiedenen Szenarien einsetzen.

- **TGuard – Bekämpfung von Desinformation durch die Analyse von Fake-News-Netzwerken:** Das Projekt bekämpft die Verbreitung von Fake News und entwickelt innovative Methoden zur Erkennung von Desinformation in sozialen Medien sowie wirksame Strategien zur Verhinderung von KI-generierten Falschmeldungen. Fight Fakes könnte in diesem Projekt als eine mögliche Methode zur Bekämpfung von Fake News entlang des erhobenen Feedbacks weiterentwickelt werden.
- **SafeNEM – Sicherheit & Qualität am Markt für Nahrungsergänzungsmittel:** In diesem ACR-strategischen Projekt beschäftigt sich das Projektteam mit dem betrügerischen und problematischen Online-Vertrieb von Nahrungsergänzungsmitteln. Ein besonderer Fokus liegt dabei auf der Bewerbung von Nahrungsergänzungsmitteln mit unlauteren Gesundheitsaussagen. Die Ergebnisse sollen auch an Konsument:innen disseminiert werden. Fight Fakes ist ein geeignetes Tool, um spielerisch zu erklären, was problematische (und vielleicht sogar gefährliche) Gesundheitsaussagen sind.
- **defame Fakes – Detektion von Deepfakes und medialen Manipulation von Bildern und Videos:** In diesem KIRAS-Projekt (FFG) werden geeignete und verständliche Assessment-Werkzeuge zur Erkennung von Deepfakes und Manipulationen erforscht. Die entwickelten Tools können einerseits in Fight Fakes als mögliche Unterstützung zum Erkennen von Deepfakes integriert werden. Andererseits ist auch in diesem Projekt ein Disseminationsteil geplant, um Nutzer:innen zum Thema Deepfakes zu sensibilisieren. Daher sollen auch in diesem Projekt weitere Games entwickelt werden.

Auch eine englischsprachige Ausweitung ist angedacht, um die Spielumgebung im europäischen Saferinternet-Netzwerk zu verbreiten.

10 Anregungen für Weiterentwicklungen durch Dritte

Vor allem aus der letzten Feedback-Runde, in der Lehrkräfte das Game direkt im Unterricht eingesetzt haben, konnten mögliche Weiterentwicklungen durch Dritte gesammelt werden:

- **Weiterführung der Instanzen:** Eine zentrale Funktion des Games ist, dass Lehrkräfte eine Instanz erstellen können, um einzelne Games nur innerhalb der Klasse zu spielen und dementsprechend den Highscore nur innerhalb der Klasse anzeigen zu lassen. Hier wurde der

Wunsch geäußert, dass diese Möglichkeit über 3 Level (je Kategorie und entlang des Kompetenzpfades von easy zu pro extreme) gespeichert werden kann.

- **Weiterentwicklung Chat Game:** Das Chat Game ist aktuell nur als Prototyp vorhanden, hier wäre eine Weiterentwicklung wünschenswert, um das Spiel weniger linear verlaufen zu lassen. Aktuell kann das Gespräch nicht je nach Antwort der Spieler:innen unterschiedlich weitergeführt werden. Dies wäre jedoch im Sinne einer größeren Flexibilität der Dialoge sinnvoll.
- **Eigenes Backend für Lehrkräfte:** Denkbar wäre auch ein eigenes Backend für Lehrer:innen, damit diese ihre verschiedenen Instanzen selbst pflegen und bearbeiten können, die Highscores im Blick haben oder sogar eigene Spiele erstellen können.