

FIGHT FAKES

Anwendungsdokumentation

1. Hintergrund

Generative KI boomt und mit ihr die Erstellung von Fakes und Deep Fakes. Konfrontiert damit sind auch zunehmend Jugendliche: Durch TikTok, Snapchat & Co. ist die Zielgruppe einer Flut an Informationen ausgesetzt. Die Bewertung, ob einem Inhalt vertraut werden kann oder nicht, wird dadurch immer schwieriger.

Das bestätigt auch die saferinternet.at-Studie [„Jugendliche und Falschinformationen im Internet“](#): 70% der Jugendlichen haben Schwierigkeiten damit, Informationen im Internet zu bewerten. Das Bauchgefühl ist ein zentraler Wegweiser, um Fakes zu erkennen. Nur wenn Informationen intuitiv unglaubwürdig erscheinen, werden sie überprüft. Gleichzeitig zeigt die Studie auch, dass die wichtigste Informationsquelle für Jugendliche Soziale Netzwerke sind und das, obwohl die dort geposteten Inhalte für die Jugendlichen am wenigsten glaubwürdig sind.

Um Jugendliche bei ihrem Dilemma im Umgang mit Deep Fakes zu unterstützen, benötigt es niederschwellige Werkzeuge. Derzeit mangelt es jedoch an Unterrichtsmaterial, da sich die Inhalte rund um Digitalisierung laufend ändern. So gibt es derzeit kaum geeignete deutschsprachige Materialien zum Thema Deep Fakes, die gleichzeitig kurzweilig und mit wenig Vorbereitung in der Schule einsetzbar sind.

Hier setzen wir mit dem Game [„Fight Fakes“](#) an und bauen auf unser erfolgreichstes Unterrichtsmaterial auf: Unsere digitalen Quiz wurden seit 2022 bereits 1,5 Millionen Mal gespielt. Daran angelehnt haben wir eine quizähnliche Spielumgebung entwickelt, in der Jugendliche spielerisch lernen, Informationen schnell und intuitiv zu bewerten.

2. Fight Fakes: das Swipe-Game

„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.

2.1. Spielbeschreibung

Die simulierten Szenarien orientieren sich visuell und funktional an gängigen Social-Media-Plattformen wie TikTok und SnapChat. In diesen Umgebungen 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.

2.2. Spielsteuerung

Auch die Spielsteuerung orientiert sich an die digitale Welt: Um Inhalte als Fake oder Nicht-Fake zu bewerten gibt es zwei unterschiedliche Möglichkeiten.

Möglichkeit 1: Bewertung per Swipe

Die Swipe-Mechanik ist für mobile Endgeräte optimiert. Dabei können die Spieler:innen einfach nach links oder rechts wischen, um die Inhalte als fake oder echt zu bewerten:

- **Swipe links:** Der Inhalt wird als echt bewertet
- **Swipe rechts:** Der Inhalt wird als fake bewertet

Möglichkeit 2: Bewertung per Emoji

Aber auch in der Desktopversion kann mittels Computermaus der jeweils angezeigte Inhalt nach links oder rechts gewischt werden, um eine Bewertung vorzunehmen. Alternativ können Bewertung auch mit Emojis vorgenommen werden, welche Emojis für welche Bewertung steht wird zu Beginn jedes Spiels erklärt.

In jedem Spiel stehen beide Möglichkeiten zur Verfügung, die Schüler:innen wählen die Steuerung, die ihnen am besten gefällt.

2.3. Kompetenzlevel

Zu jeder Thematik gibt es 3 Niveaus: Easy, Pro und Pro Extreme. Am Ende der ersten zwei Levels wird man automatisch zum nächsten weitergeleitet, um so die Kompetenz nach und nach zu trainieren und zu verfeinern.

- Im Level *Easy* sind die Deepfakes noch gut zu erkennen, oft sind sie gänzlich unsinnig, oder sie haben bestimmte Fehler drinnen – vor allem wenn Menschen zu sehen sind.
- Beim Niveau *Pro* wird es schon schwieriger, einerseits weil die echten Bilder teilweise komischer werden, wenn beispielweise selten, komisch aussehende Tiere mit dabei sind. Und andererseits, weil die Deepfakes besser sind, oder so geschnitten wurden, dass man offensichtliche Fehler (wie Menschen mit verformten Körpern) nicht sehen kann.
- Letztlich kommt man zu *Pro Extreme*, wo man oft den Unterschied überhaupt nicht mehr erkennen kann, oder nur bei längerer, intensiverer Betrachtung. Dies kann zwar

frustrierend sein, soll aber darauf hinweisen, dass Deepfakes immer besser werden, und man beim schnellen drüber schauen diese KI generierten Bilder manchmal sehr leicht mit Echten verwechseln kann.

Die „echten“ Bilder sind oft überarbeitet, beispielsweise mit Farb- und Helligkeitskontrasten, oder mit Filtern. Wir haben uns bewusst dazu entschieden, auch solche Bilder in dem Spielkontext als „echt“ zu kennzeichnen, da dies die Realität auf Social Media und Online widerspiegelt.

3. Fight Fakes: das Chat-Game

3.1. Spielbeschreibung

Während das Swipe-Game von den Spielenden eine schnelle und duale Fake/Nicht-Fake-Entscheidung verlangt, ermöglicht das Chat-Game eine tiefergehende Auseinandersetzung mit den Themen Deepfakes und Fake-News. Simuliert wird ein Chat-Szenario, das sich visuell und funktional an gängigen Messenger-Apps orientiert.

In dieser Umgebung kommunizieren die Spieler:innen mit einer fiktiven Person und treffen Entscheidungen in einem dialogbasierten Format – ähnlich wie im Alltag auf Plattformen wie WhatsApp, Instagram oder Discord.

Die Nachrichten der fiktiven Figur sollen sie dabei in eine Geschichte hineinziehen – je nach Kampagne entweder als Teil eines „guten“ („good“) oder „bösen“ („evil“) Vorhabens. Im „evil“-Pfad helfen sie z. B. dabei, ein fragwürdiges Produkt mit Hilfe von Deepfake-Technologien zu bewerben. Dabei erleben sie Schritt für Schritt, wie Deepfakes erstellt werden, wo Fehler passieren und wie leicht man mit solchen Inhalten manipulieren kann.

Das Chat-Game soll Jugendlichen spielerisch vermitteln, wie Deepfakes entstehen, woran man sie erkennt und welche technischen und ethischen Fragen dabei eine Rolle spielen.

3.2. Spielsteuerung

Das Spiel wird über eine intuitive **Chat-Oberfläche** gesteuert, wie sie viele Jugendliche aus Apps wie WhatsApp oder Instagram kennen. Entscheidungen werden direkt im Gesprächsverlauf getroffen. Die einzelnen Aufgaben sind in kurze Chat-Abschnitte unterteilt. Spieler:innen wählen dabei zwischen mehreren Antwortmöglichkeiten – teils rein textbasiert, teils mit unterstützenden Bildern oder Videos.

- Zu jeder Nachricht gibt es **2–4 Antwortmöglichkeiten**, aus denen die Spieler:innen auswählen können.
- Einige Antwort-Optionen sind zusätzlich mit **Bildern oder Videos (Deepfakes)** versehen, welche die Erstellung der Deepfakes simulieren sollen.

- Nach jeder Antwort bekommen die Spieler:innen eine kurze Rückmeldung (teils auch mit Emoji-Reaktionen).

Ziel ist es, durch Entscheidungen den Verlauf der Geschichte zu beeinflussen, dabei Wissen über Deepfakes aufzubauen und ein Gespür für Manipulation, Fake-Inhalte und mediale Verantwortung zu entwickeln. Das Spielprinzip fördert die kritische Reflexion im Umgang mit digitalen Inhalten wie Deepfakes.

4. Einsatz im Unterricht

4.1 Anleitung für das Spielen in der Klasse

„Fight Fakes“ ist niederschwellig und modular gestaltet. Dadurch kann das Spiel im Unterricht auf verschiedene Arten eingesetzt werden. Die Spiele sind sowohl eigenständig zu spielen als auch entlang eines Kompetenzpfads (siehe 2.3. Kompetenzlevel). Je nach zeitlichen Ressourcen ist es zudem möglich, einzelne Spiele ohne zusätzliche Vorbereitung in der Klasse zu spielen oder mit Hilfe der Begleitmaterialien (siehe 3.2. Begleitmaterialien) tiefer in das Thema einzutauchen.

Die Durchführung des Spiels verläuft unabhängig von der Gestaltung gleich ab:

Schritt 1: Auswahl eines Spiels

Auf der Projektseite <https://www.saferinternet.at/projekte/fight-fakes> finden Sie die unterschiedlichen Spiele. Starten Sie am besten mit dem Niveau „Easy“.

Schritt 2: Erstellen einer Instanz

Mit Klick auf ein Spiel gelangen Sie zu folgender Ansicht:

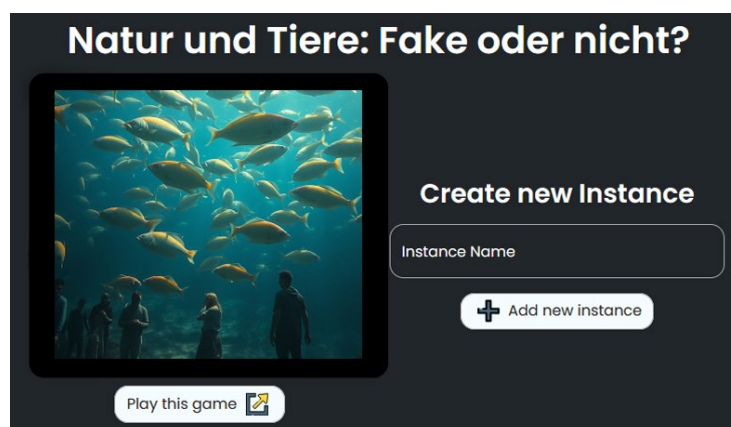


Abb. 1: Erstellen Sie eine Instanz für Ihre Klasse

Hier können Sie eine neue Instanz anlegen. Damit stellen Sie sicher, dass innerhalb einer Klasse gespielt wird und die Schüler:innen nur gegeneinander antreten. Benennen Sie Ihre

Instanz im Feld „Instance Name“ (z.B. Klasse 1B) und klicken Sie anschließend auf „Add new instance“.

Schritt 3: Scannen des QR-Codes

Nach dem Erstellen einer Instanz gelangen Sie zu folgender Ansicht:

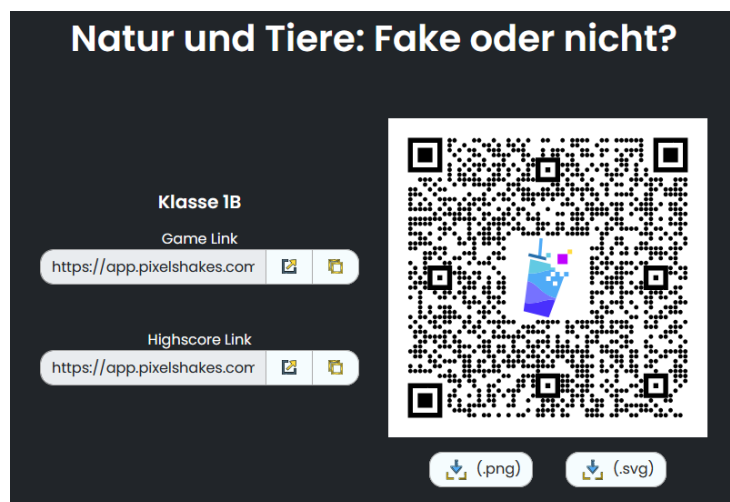


Abb. 2: Lassen Sie Ihre Schüler:innen den QR-Code scannen, damit diese gegeneinander antreten können

Nun bitten Sie Ihre Schüler:innen den angezeigten QR-Code zu scannen. Alternativ können Sie auch den Link an Ihre Schüler:innen versenden. In beiden Fällen gelangen die Schüler:innen zur Spielansicht.

Schritt 4: Schüler:innen spielen

In der Spielansicht müssen die Spieler:innen nur noch auf „Start“ klicken und schon beginnt das Spiel: Den Schüler:innen werden dabei nacheinander Bilder oder Videos angezeigt. Innerhalb weniger Sekunden müssen Sie entscheiden, ob es sich dabei um ein Deep Fake handelt oder nicht (siehe 2.2. Spielsteuerung).

Achtung: Wird das Spiel auf dem Handy über mobile Daten gespielt, kann es vorkommen, dass die Videos nicht automatisch abgespielt werden. Stattdessen erhalten die Spielenden eine Warnung, dass mobile Daten verwendet werden. Mit einem Klick auf den Play-Button verschwindet die Warnung und das Video wird abgespielt.

Schritt 5: Abschlusscreen & Highscores

Am Ende werden die Schüler:innen gebeten ihren Namen für den Highscore einzugeben. Die Schüler:innen können sich den Highscore auf ihrem Gerät anzeigen lassen. Sie können den Highscore aber auch für alle projizieren, indem Sie auf den „Highscore Link“ (siehe Abb. 2) aufrufen.

Zusätzlich gibt es am Ende jedes Spiels entweder einen Link zum nächsten Level (siehe 2.3 Kompetenzlevel), oder beim letzten Level *Pro Extreme* noch einen Link zu weiterführenden Informationen für die Schüler:innen, die dabei helfen, Fakes zu erkennen. Diese können auch Sie für weitergehende Informationen und Diskussionen verwenden.

4.2 Begleitmaterial

„Fight Fakes“ ist so konzeptioniert, dass das Spiel ohne große Vorbereitung für die Lehrpersonen funktioniert. Wenn Sie die Spiele jedoch zum Anlass nehmen wollen, um sich mit ihren Schüler:innen genauer mit dem Thema Deepfakes auseinanderzusetzen empfehlen wir folgende Begleitmaterialien:

- Im Themenschwerpunkt Informationskompetenz finden Sie News, FAQs, Tipps sowie weitere Quiz zum Thema Informationskompetenz:
<https://www.saferinternet.at/themen/informationskompetenz>
- Der Artikel [Deepfakes – Wie Foto-, Video- und Audioinhalte manipuliert werden](#) gibt weiterführende Informationen speziell zum Thema Deepfakes:
- Das Unterrichtsmaterial „Wahr oder falsch im Internet? Informationskompetenz in der digitalen Welt“ unterstützt Lehrende bei der Vermittlung dieser wichtigen Fähigkeiten im Unterricht:
https://www.saferinternet.at/fileadmin/categorized/Materialien/WEB_Wahr_oder_falsch_im_Internet.pdf
- Der ISPA-Flyer „Fakt oder Fake? Hoaxes, bearbeitete Bilder und Deepfakes im Internet“ informiert über Hoaxes, bearbeitete Bilder und Fake News und wie diese effektiv erkannt werden können:
https://www.saferinternet.at/fileadmin/categorized/Materialien/ISPA_Fakt_oder_Fake.pdf
- Die Faktencheck-Videos am Ende unserer Projektseite wurden von „Digitaler Kompass“ erstellt und entlarven hartnäckige Alltagsmythen, die auf TikTok & Co. verbreitet werden: <https://www.saferinternet.at/projekte/fight-fakes>

4.3 Bonus: Open Source

Sowohl der Fake- und Deepfake-Content als auch die Vorlagen für die an Social Media angelehnten Szenarien stehen als Open Source (CC BY 4.0) frei zur Verfügung.

Den Content können Sie verwenden, um die Spiele nachzubesprechen und die einzelnen Bilder und Videos gemeinsam mit den Schüler:innen zu analysieren. Gleichzeitig können Sie den Content und die Vorlagen selbst als Lehrmaterialien nutzen – unabhängig von „Fight Fakes“.

Sie finden die Dateien hier (eingeordnet in thematische Kategorien und dort jeweils noch nach Kompetenzlevel): <https://cloud.oiat.at/s/YS3Yj5jPXsqNoD4>

4. Level-Übersicht

Aktuell können auf saferinternet.at/projekte/fight-fakes 18 Games gespielt werden – jeweils drei Spiele in 6 Kategorien (Gesundheit, Biologie, Geographie) und mit drei unterschiedlichen Kompetenzlevels (EASY, PRO, PRO EXTREME). Die nachfolgende Übersicht hilft Ihnen dabei die jeweiligen Bilder (auch jene, die in der Cloud zur Nachbesprechung gespeichert sind) den Levels zuzuordnen und gleichzeitig bei der Auflösung den Überblick zwischen Echt und Fake zu behalten.

Gesundheit

Link zur Erstellung einer Instanz (siehe Punkt 3 – Einsatz im Unterricht)

Easy: <https://app.pixelshakes.com/instancegenerator.php?gameid=698>

Pro: <https://app.pixelshakes.com/instancegenerator.php?gameid=700>

Pro Extreme: <https://app.pixelshakes.com/instancegenerator.php?gameid=701>

Individueller Link direkt zum Spiel:

Easy: <https://app.pixelshakes.com/gameviewer.php?gameid=698>

Pro: <https://app.pixelshakes.com/gameviewer.php?gameid=700>

Pro Extreme: <https://app.pixelshakes.com/gameviewer.php?gameid=701>

Die Fake-Inhalte der Gesundheitsspiele enthalten sowohl Fakes (falsche Aussagen) als auch Deepfakes. Dabei haben wir uns an Fake-Gesundheits-Trends orientiert, denen Jugendliche auf Social-Media begegnen:

- Virale Diät-Hacks & Abnehm-Mythen
- Detox & Cleansing Scams
- „Das Geheimnis, das Ärzte dir nicht sagen wollen“ – Verschwörungsmythen
- Biohacking & Extreme „Wellness“-Trends
- Fake-Supplements & Vitamin-Mythen
- Fitness-Mythen & Fake-Trainingsmethoden
- Lebensmittel-Panik & Fake-Giftstoff-Mythen
- Falsche Informationen zu mentaler Gesundheit

Biologie (Tiere)

Link zur Erstellung einer Instanz (siehe Punkt 3 – Einsatz im Unterricht)

Easy: <https://app.pixelshakes.com/instancegenerator.php?gameid=688>

Pro: <https://app.pixelshakes.com/instancegenerator.php?gameid=677>

Pro Extreme: <https://app.pixelshakes.com/instancegenerator.php?gameid=694>

Individueller Link direkt zum Spiel:

Easy: <https://app.pixelshakes.com/gameviewer.php?gameid=688>

Pro: <https://app.pixelshakes.com/gameviewer.php?gameid=677>

Pro Extreme: <https://app.pixelshakes.com/gameviewer.php?gameid=694>

Die Fake-Inhalte der Biologiespiele enthalten nur Deepfakes.

Geographie (Sehenswürdigkeiten)

Link zur Erstellung einer Instanz (siehe Punkt 3 – Einsatz im Unterricht)

Easy: <https://app.pixelshakes.com/instancegenerator.php?gameid=693>

Pro: <https://app.pixelshakes.com/instancegenerator.php?gameid=692>

Pro Extreme: <https://app.pixelshakes.com/instancegenerator.php?gameid=691>

Individueller Link direkt zum Spiel:

Easy: <https://app.pixelshakes.com/gameviewer.php?gameid=693>

Pro: <https://app.pixelshakes.com/gameviewer.php?gameid=692>

Pro Extreme: <https://app.pixelshakes.com/gameviewer.php?gameid=691>

Die Fake-Inhalte der Geographiespiele enthalten nur Deepfakes.

Kunst

Link zur Erstellung einer Instanz (siehe Punkt 3 – Einsatz im Unterricht)

Easy: <https://app.pixelshakes.com/instancegenerator.php?gameid=711>

Pro: <https://app.pixelshakes.com/instancegenerator.php?gameid=712>

Pro Extreme: <https://app.pixelshakes.com/instancegenerator.php?gameid=713>

Individueller Link direkt zum Spiel:

Easy: <https://app.pixelshakes.com/gameviewer.php?gameid=711>

Pro: <https://app.pixelshakes.com/gameviewer.php?gameid=712>

Pro Extreme: <https://app.pixelshakes.com/gameviewer.php?gameid=713>

Die Fake-Inhalte der Kunstspiele enthalten nur Deepfakes.

Profilbilder

Link zur Erstellung einer Instanz (siehe Punkt 3 – Einsatz im Unterricht)

Easy: <https://app.pixelshakes.com/instancegenerator.php?gameid=716>

Pro: <https://app.pixelshakes.com/instancegenerator.php?gameid=715>

Pro Extreme: <https://app.pixelshakes.com/instancegenerator.php?gameid=714>

Individueller Link direkt zum Spiel:

Easy: <https://app.pixelshakes.com/gameviewer.php?gameid=716>

Pro: <https://app.pixelshakes.com/gameviewer.php?gameid=715>

Pro Extreme: <https://app.pixelshakes.com/gameviewer.php?gameid=714>

Die Fake-Inhalte der Profilbilderspiele enthalten nur Deepfakes.

Technikgeschichte

Link zur Erstellung einer Instanz (siehe Punkt 3 – Einsatz im Unterricht)

Easy: <https://app.pixelshakes.com/instancegenerator.php?gameid=695>

Pro: <https://app.pixelshakes.com/instancegenerator.php?gameid=696>

Pro Extreme: <https://app.pixelshakes.com/instancegenerator.php?gameid=697>

Individueller Link direkt zum Spiel:

Easy: <https://app.pixelshakes.com/gameviewer.php?gameid=695>

Pro: <https://app.pixelshakes.com/gameviewer.php?gameid=696>

Pro Extreme: <https://app.pixelshakes.com/gameviewer.php?gameid=697>

Die Fake-Inhalte der Profilbilderspiele enthalten nur Deepfakes.

Prototyp Chat-Game

Link zur Erstellung einer Instanz:

<https://app.pixelshakes.com/instancegenerator.php?gameid=710>

Link direkt zum Spiel:

<https://app.pixelshakes.com/instancegenerator.php?gameid=695>

Fight Fakes ist im Rahmen einer netidee-Förderung der Internet Stiftung entstanden (<https://www.netidee.at/fight-fakes>). Die Spiele haben wir in Zusammenarbeit mit Bytewood (<https://bytewood.com>) entwickelt.