Modul 2: Fiktion oder Realität?

KI, Fake News und Deepfakes im Kontext von Extremismus

In diesem Modul geht es darum, wie Künstliche Intelligenz (KI) zur Verbreitung von Fake News, Deepfakes und Verschwörungserzählungen eingesetzt wird. Dabei werden deren Bedeutung für die Extremismusprävention sowie Möglichkeiten zur Förderung eines kritischen und reflektierten Umgangs mit digitalen Medien thematisiert.

1. Begriffe & Definitionen (Die Täuschungs-Matrix)

- **Fake News:** Gefälschte oder erfundene Nachrichten, die bewusste Lügen verbreiten und oft reißerisch gestaltet sind.
- **Deep Fakes:** Reales Bild-, Video- oder Audiomaterial, das mit KI manipuliert wurde und dadurch echt wirkt.
- **KI-generierte Bilder:** Bilder, die mit einer KI-Bildgenerierung aus Texteingaben erzeugt werden.
- **Verschwörungserzählungen:** Eine geschlossene Welterklärung mit klaren Feindbildern.
- **Cheap Fakes:** Manipulierte Bilder, die mit einfachen Bearbeitungswerkzeugen (Videomitschnitte, veränderte Geschwindigkeit, falscher Kontext) ohne komplexe KI erstellt wurden.
- **Desinformation vs. Falschinformation:** Falschinformationen entstehen durch Versehen oder Missverständnisse (keine böse Intention). Desinformation (Fake News) verbreitet absichtlich Unwahrheiten.

2. Die Plattform-Realität & Social Media-Nutzung

Plattform	Tägliche Nutzung (12-19 J.)[1]	Risiko-Profil
WhatsApp	~82%	Kettennachrichten, Peergroup-Verstärkung, persönlicher Austausch.
YouTube	~76%	Lange Erklärvideos, Verschwörungsinhalte, pseudowissenschaftliche Modelle.
TikTok	~64%	Algorithmische Sogwirkung, menschenfeindliche Ideologien, eigene KI-Inhalte.

[1] Saferinternet.at (2026): Jugend-Internet-Monitor 2026. Verfügbar unter: https://www.saferinternet.at/services/jugend-internet-monitor [Zugriff: 15.05.2026].

Instagram	~64%	Reichweite durch Likes, Vertrauen in Influencer*innen trotz Fake News.
Discord / Telegram	Wachsend	Geschlossene Gruppen, kaum Moderation, unzensierte gewaltvolle Inhalte.

3. Wie KI-Inhalte erstellt

- **KI-generierte Bilder:** Text-zu-Bild-Generatoren (z.B. DALL-E, Midjourney) kombinieren Merkmale aus riesigen Internet-Datenbanken zu täuschend echten Bildern.
- **Deepfakes:** Entstehen durch Lipsync, Face Swap oder Face Reenactment. Stimmen und Gesichter werden täuschend echt manipuliert (z.B. Personen sagen Dinge, die sie nie gesagt haben).
- **Verbreitung:** Algorithmen und KI helfen, Fake News zu produzieren und zu verbreiten.

4. Warum Fakes viral gehen und Jugendliche besonders empfänglich dafür sind

Algorithmus & Stickiness | Filterblasen | Popularität & Strategie

- **Algorithmen & Plattform-Design:** Soziale Medien bevorzugen emotionale, aufmerksamkeitserregende Inhalte, da diese mehr Interaktionen erzeugen und die Verweildauer dadurch länger ist („Stickiness“). Dadurch können sich Fake News schneller verbreiten.
- **Filterblasen:** User*innen werden vor allem Inhalte angezeigt, die ihren Meinungen und Überzeugungen entsprechen. Dadurch werden Fake News möglicherweise weniger kritisch hinterfragt.
- **Psychologische Faktoren in Krisenzeiten:** In unsicheren Situationen steigt das Bedürfnis nach Orientierung und einfachen Erklärungen. Fake News nutzen dies gezielt aus, indem sie vermeintlich exklusives Wissen vermitteln, Ängste schüren oder einfache Antworten auf komplexe Fragen liefern.
- **Besondere Empfänglichkeit von Jugendlichen:** Jugendliche befinden sich in einer Phase der Identitätsfindung und orientieren sich häufig an Gleichaltrigen und sozialen Vorbildern. Likes, Kommentare und Shares können das Bedürfnis nach Anerkennung verstärken. Gleichzeitig werden Informationen nicht immer ausreichend auf ihre Glaubwürdigkeit überprüft.

5. Erkennungsmerkmale von Fakes

- **KI-Bilder:** Anatomische Fehler (Hände, Zähne, Haare), unscharfe Übergänge, intensive Farben, unnatürliche Posen.
- **Deepfakes:** Fehlendes Blinzeln, unnatürliche Mimik, unpassende Lippenbewegungen oder roboterhafte/monotone Stimmen.
- **Fake News / Desinformation:** Extreme Formulierungen in Großbuchstaben (SKANDAL, SENSATION), fehlende Belege, obskure URLs, imitierte Logos.

6. Extremismus & die STOP-Methode für die Praxis

Extremismusbezug: Populistische und extremistische Gruppierungen nutzen diese Mechanismen gezielt für ihre Online-Propaganda. Durch emotionale Botschaften, vereinfachte Feindbilder und klare Freund-Feind-Schemata versuchen sie, Aufmerksamkeit zu erzeugen, bestehende Vorurteile zu verstärken und ihre Reichweite zu erhöhen. I

Die STOP-Methode:

S - Stopp: Kurz innehalten (10 Sekunden warten) vor dem Teilen.

T - Tiefer schauen: Wer ist die Quelle? Profil, Impressum und Datum prüfen.

O - Original suchen: Gibt es den Inhalt woanders oder in anderem Kontext? (Rückwärts-Bildsuche / Google Lens nutzen).

P - Prüfen: Echtheit sicherstellen und bei Unsicherheit Faktenchecker-Seiten nutzen.

7. Anlaufstellen in Österreich

- **Beratungsstelle Extremismus:** Radikalisierung & Extremismus (beratungsstelleextremismus.at|0800202044
- **Saferinternet.at:** Beratung & Medienkompetenz (saferinternet.at)
- **Bundesstelle für Sektenfragen:** Verschwörungserzählungen (bundesstelle-sektenfragen.at)
- **Mimikama:** Faktenprüfung & aktuelle Berichte (mimikama.at)

Impressum
bOJA - Beratungsstelle Extremismus
Redaktion: Alexander Fontó & Carolin Luiprecht | Projektmitarbeit: Rahel Willmer
Gefördert von: Bundeskanzleramt – Sektion Familie und Jugend
Bundeskanzleramt