



Lektion 4.

Anti-
RUMQUR

stiftung
digitale
chancen



CITIZENS
IN POWER

challedu
Inclusion | gender | education

Co-funded by the
Erasmus+ Programme
of the European Union

Finanziert von der Europäischen Union. Die Ansichten und Meinungen, die hier zum Ausdruck gebracht werden, stammen jedoch ausschließlich von den Autoren und spiegeln nicht unbedingt die der Europäischen Union oder der Nationalen Agentur wider. Weder die Europäische Union noch die Bewilligungsbehörde können dafür verantwortlich gemacht werden.

Modul 3: KI-generierte Inhalte erkennen

Der Aufstieg von KI-generierten Inhalten hat es schwieriger gemacht, gegen Falschinformationen und Desinformation vorzugehen. KI kann in rasantem Tempo äußerst realistische gefälschte Videos, Bilder und Texte erstellen, so dass es für Faktenprüfer*innen schwer ist, mitzuhalten. Das bedeutet, dass sich falsche Informationen schnell verbreiten und sehr überzeugend wirken können. Es ist schwierig zu erkennen, was wahr ist und was nicht. Infolgedessen können sich irreführende Inhalte leicht viral verbreiten, das Vertrauen der Öffentlichkeit beschädigen und Verwirrung stiften.

Beim derzeitigen Stand der KI ist es immer noch möglich, KI-generierte Inhalte zu erkennen, indem man entweder auf bestimmte Fehler achtet oder hilfreiche Tools einsetzt. In diesem Modul werden sowohl die Herausforderungen der KI als auch die Möglichkeiten der Nutzer*innen, diese Herausforderungen zu bewältigen, untersucht.

Aktivität 1: Input – Einführung und KI-generierter Text

Typ: Input

Dauer: 10 Minuten

Lernergebnisse: Die Teilnehmenden werden ein besseres Verständnis von generativer KI, ihren Herausforderungen bei der Verbreitung von Desinformation und den Schwierigkeiten bei der Erkennung von KI-generierten Texten haben.

Benötigte Materialien/Ressourcen: Notizbücher und Stifte für die Teilnehmenden. Input-Skript für die Workshopleitung. Großer Bildschirm oder Projektor und ein Laptop, der damit verbunden ist, für die Workshopleitung.

Optional: Die Workshopleitung kann eine PowerPoint-Präsentation vorbereiten, die an das Alter und das Hintergrundwissen der Gruppe angepasst ist.

Die Workshopleitung verwendet die folgenden Informationen für ihren Input.

Einführung:

KI kann täuschend echt aussehende Bilder von Ereignissen generieren, die nie stattgefunden haben, künstlich erzeugte Stimmen und anspruchsvolle Texte ohne menschliche Autor*innen. Es wird immer schwieriger zu unterscheiden, was real und was gefälscht ist. Dies unterstreicht eine allgemeine Wahrheit über Technologie: Sie kann sowohl ein Segen als auch ein Fluch sein, abhängig davon, wie und zu welchem Zweck sie genutzt wird.

Während ein Großteil der heute online verfügbaren KI-generierten Inhalte das Ergebnis kreativer Erkundungen durch Benutzer*innen ist, kann KI auch verwendet werden, um zu täuschen und Desinformation zu verbreiten.

Generative KI ist eine Art künstlicher Intelligenz, die neue Inhalte wie Texte, Bilder, Musik oder Videos erstellen kann, indem sie aus vorhandenen Daten lernt. Sie verwendet intelligente Computerprogramme, sogenannte neuronale Netzwerke, um Dinge zu erstellen, die aussehen oder klingen, als wären sie von Menschen gemacht. Beispiele sind Chatbots wie ChatGPT, die Texte schreiben, KI-Tools, die Kunstwerke erstellen, und Programme, die gefälschte Videos erzeugen.

- Video zeigen: <https://www.youtube.com/watch?v=oxXpB9pSETo>



KI-generierter Text:

KI-generierter Text wird mithilfe einer Art künstlicher Intelligenz namens Sprachmodelle erstellt. Diese Modelle werden mit großen Mengen an Textdaten aus Büchern, Artikeln und Websites trainiert. Indem sie Muster und Strukturen in den Daten lernen, kann die KI neuen Text generieren, der klingt, als wäre er von einem Menschen geschrieben worden. Wenn die KI einen Prompt oder ein Thema erhält, sagt sie die nächsten Wörter und Sätze basierend auf dem, was sie gelernt hat, vorher.

KI-generierter Text stellt eine Herausforderung in Bezug auf Desinformation dar, da er realistische und überzeugende Artikel oder Social-Media-Beiträge erstellen kann. Dies erleichtert es, absichtlich irreführende Inhalte zu erstellen. Die Inhalte können in großen Mengen produziert und auf bestimmte Zielgruppen zugeschnitten werden.

Es ist auch wichtig zu bedenken, dass KI-generierte Inhalte nicht unbedingt wahr sein müssen. Ein Grund dafür ist, dass die Sprachmodelle hinter den Chatbots selbst kein Gefühl für die Realität haben und nicht wissen, was wahr und was falsch ist. Wir müssen uns bewusst sein, dass eine KI – egal wie authentisch ihre Antworten klingen mögen – keine Kenntnis der Welt um uns herum und keine menschliche Wahrnehmung hat. Während eine KI berechnen kann, welches Wort am wahrscheinlichsten auf ein anderes folgt, kann die KI selbst die Bedeutung des resultierenden Textes nicht erfassen. Wie wahrscheinlich es ist, dass der Text faktisch korrekt ist, hängt davon ab, wie oft eine Wortsequenz in den Trainingsdaten erschienen ist.

Erkennung von KI-generiertem Text: Es wird immer schwieriger, KI-generierten Text zu identifizieren, insbesondere für die durchschnittlichen Benutzer*innen. Einige Online-Tools können jedoch helfen, KI-Inhalte zu erkennen. Zum Beispiel:

GPTZero:

<https://gptzero.me/>

AI Detector von Content at Scale:

<https://contentatscale.ai/ai-content-detector/>

Zeige einen der KI-Detektoren und erkläre kurz, wie man ihn benutzt:

Einfach den Text in das Textfeld kopieren oder eine URL in das URL-Feld einfügen.

Du erhältst dann eine Einschätzung, wie wahrscheinlich es ist, dass dieser Inhalt menschlich generiert wurde.

Es ist jedoch wichtig zu wissen, dass diese Einschätzung nicht vollständig zuverlässig ist! Es ist nur eine Vorhersage und nicht 100% genau. Es kann helfen, eine Vorstellung davon zu bekommen, ob ein Text von KI erstellt wurde, aber man sollte ihm nie blind vertrauen. Diese Tools funktionieren auch besser auf Englisch als in anderen Sprachen.

Anweisungen für Teilnehmende: Hört euch den Input der Workshopleitung an, macht euch Notizen und stellt Fragen, wenn etwas unklar ist.

Mögliche Anpassungen: Abhängig vom Alter, Wissen und Hintergrund der Teilnehmenden kann der Input angepasst werden. Im Allgemeinen kann die Workshopleitung die Aktivität in einer Weise implementieren, die am besten zur Gruppe passt. Es sollten jedoch alle notwendigen Informationen vermittelt werden, damit die folgenden Aktivitäten durchführbar sind.

Nachbesprechung: Frage die Teilnehmenden, ob sie Fragen haben.

Hintergrundlektüre/externe Ressourcen:

<https://www.digital-mobil-im-alter.de/wissen/kuenstliche-intelligenz/ki-generierte-inhalte>

<https://www.makeuseof.com/accurate-ai-text-detectors/>

<https://research.ibm.com/blog/what-is-generative-AI>



Mit KI erstellt

Aktivität 2: KI-Detektoren testen

Typ: Gruppenaktivität und Diskussion

Dauer: 10 Minuten

Lernergebnisse: Die Teilnehmenden wissen, wie man KI-Text-Detektor-Tools benutzt, und sind sich der Einschränkungen dieser Tools bewusster.

Benötigte Materialien/Ressourcen:

Ein Gerät pro Gruppe mit einem ChatGPT-Account (oder einem anderen KI-Text-Generator).
Laptop und großer Bildschirm/Projektor für die Workshopleitung.

Hinweise für Workshopleitung:

Teile die Teilnehmenden in Gruppen von 2-3 Personen auf. Jede Gruppe sollte ein Gerät mit einem ChatGPT-Account haben.

Nach 6-7 Minuten sollen die Teilnehmenden ihre Arbeit beenden und ihre Erkenntnisse mit den Nachbesprechungsfragen diskutieren.

Anweisungen für Teilnehmende:

- Loggt euch in ChatGPT ein. Er stellt einen Artikel zu einem Thema eurer Wahl. Ihr könnt absichtlich einen falschen Artikel erstellen, zum Beispiel über Chemtrails.
- Kopiert den Text in eines der KI-Detektor-Tools. Wie genau identifiziert es den KI-generierten Text?
- Versucht, euren ursprünglichen Text mit Aufforderungen wie „Formuliere so um, dass der Text weniger wie ein KI-generierter Text klingt“ anzupassen, oder versucht es mit verschiedenen Sprachen, indem ihr Aufforderungen wie „Übersetze diesen Text ins Englische“ verwendet.
- Kopiert den angepassten Text in den KI-Detektor. Wie hat sich die Bewertung verändert?

Mögliche Anpassungen:

Wenn die Teilnehmenden keine ChatGPT-Accounts haben oder keinen einrichten möchten, kann diese Aktivität auch gemeinsam mit der gesamten Gruppe durchgeführt werden. Die Teilnehmenden geben der Workshopleitung, die ihren Bildschirm teilt, vor, welche Eingaben gemacht werden sollen. Die Workshopleitung kopiert dann den Text in einen KI-Text-Detektor und bespricht das Ergebnis mit der Gruppe.

Nachbesprechung:

- Was habt ihr bemerkt?
- Wie zuverlässig waren die Detektoren in eurem Test?
- Was zeigt euch das über die Herausforderungen bei der Erkennung von KI-generierten Inhalten?

Hintergrundlektüre / externe Ressourcen:

<https://www.makeuseof.com/accurate-ai-text-detectors/>

Aktivität 3: Input und Quiz – KI-generierte Bilder

Typ: Input und Quiz

Dauer: 10 Minuten

Lernergebnisse: Die Teilnehmenden werden ein besseres Verständnis für KI-generierte Bilder haben und wissen, wie man sie erkennt.

Benötigte Materialien/Ressourcen:

- Notizbücher und Stifte für die Teilnehmenden
- Input-Skript für die Workshopleitung
- Großer Bildschirm oder Projektor und ein Laptop, der damit verbunden ist, für die Workshopleitung
- Lektion 4 Anhang 3: Bilder

Zeige die Beiträge entweder auf einem großen Bildschirm (Projektor/TV) über einen Laptop und eine PDF/PPT-Datei.

Oder (bevorzugt): Mentimeter Quiz (PC + Projektor/TV, ein Smartphone/Tablet pro Teilnehmer*in)

1. Richte ein Quiz auf www.mentimeter.com (oder einer vergleichbaren Plattform) ein
2. Erstelle fünf Folien und fügt die fünf Beispielbilder aus Anhang 4 hinzu, jeweils mit der Frage "Glaubst du, dass dieses Bild echt oder KI-generiert ist?"
3. Füge auf jeder Folie ein Bild hinzu (es wird möglicherweise nicht ideal angezeigt aufgrund des Formats der Folie, aber die Teilnehmenden sehen es korrekt auf ihren Geräten). Füge die Antwortmöglichkeiten "Echt" und "KI-generiert" hinzu und markiere die richtige Option
4. Deaktiviere "More points for fast correct answers"

Hinweise für Workshopleitung:

Gib eine kurze Einführung zum Thema KI-generierte Bilder und zeige die Checkliste auf einem großen Bildschirm an.

Give a short introduction to the topic of AI-generated images and show the checklist on a large screen

Erkennen von KI-generierten Bildern und Videos

KI-generierte Inhalte wie Bilder und Videos werden mit fortschrittlichen neuronalen Netzwerken erstellt, die auf großen Datensätzen trainiert wurden. Diese Netzwerke lernen, neue Inhalte zu erstellen, indem sie Muster in den ihnen zugeführten Daten erkennen. Mit einer Beschreibung oder einer Idee können sie hochrealistische Bilder oder Videos erzeugen. Trotz ihrer zunehmenden Raffinesse gibt es jedoch Möglichkeiten, KI-generierte Inhalte zu erkennen, indem man auf bestimmte Hinweise achtet.

Wichtige Indikatoren zur Identifizierung von KI-generierten Inhalten

- **Anatomie und Gliedmaßen:** KI hat oft Schwierigkeiten, die menschliche Anatomie genau darzustellen. Bei Bildern sollte man auf Hände mit zu vielen Fingern oder ungewöhnlich geformte Gliedmaßen achten. In Videos sollte man auf unnatürliche Körperbewegungen achten, da Deepfakes mechanische oder sich wiederholende Bewegungen erzeugen können.
- **Text und Symbole:** KI-generierte Bilder können Text unsachgemäß behandeln, was zu falsch geschriebenen Wörtern, verschwommenen Buchstaben oder merkwürdigen Symbolen führen kann. Ähnlich kann im Video jeder auf dem Bildschirm sichtbare Text, der ungewöhnlich erscheint, ein Hinweis sein.

- **Haar und Haut:** Menschenhaare in KI-generierten Bildern können wie Pinselstriche aussehen, anstatt natürlich zu fließen. In Videos kann die Haut zu glatt oder zu faltig erscheinen, und Inkonsistenzen zwischen Haut und anderen Merkmalen wie Haaren und Augen können auf Manipulation hinweisen.
- **Symmetrie und Konsistenz:** KI kann Schwierigkeiten mit der Symmetrie haben. In Bildern und Videos kann eine Seite des Gesichts anders aussehen als die andere. Konsistenzprobleme sind ebenfalls ein Hinweis; wenn es mehrere Bilder oder Frames desselben Subjekts gibt, können Inkonsistenzen darauf hindeuten, dass sie KI-generiert sind.
- **Texturen und Muster:** Die realistische Darstellung von Texturen, Mustern und Stoffen ist für KI herausfordernd. Kleine Abweichungen in der Art, wie Oberflächen erscheinen, können darauf hinweisen, dass ein Bild KI-generiert ist. In Videos sollte man auf unnatürliche Texturen bei Objekten oder Hintergründen achten, die fehl am Platz wirken.
- **Geometrie und Winkel:** Überprüfen Sie, wie Objekte in der Szene zusammenpassen. In Bildern können ungewöhnliche Winkel oder Elemente, die nicht natürlich ausgerichtet sind, ein Zeichen für KI-Generierung sein. In Videos sollte man auf Inkonsistenzen in räumlichen Beziehungen und Objektinteraktionen achten.
- **Gesichtszüge und Bewegungen:** Deepfakes beinhalten oft Gesichtstransformationen. Achten Sie auf unrealistische Hauttexturen, nicht übereinstimmende Altersmerkmale zwischen verschiedenen Gesichtszügen und unnatürliche Schatten oder Reflexionen. Achten Sie auf die Blinzelmuster, da Deepfakes möglicherweise zu viel oder zu wenig blinzeln. Überprüfen Sie auch die Lippenbewegungen, um zu sehen, ob sie mit den gesprochenen Worten übereinstimmen.
- **Details und Verfeinerungen:** KI-generierte Inhalte können manchmal feine Details übersehen. In Bildern kann sich das in der Art und Weise zeigen, wie Licht mit Oberflächen interagiert und unrealistische Highlights oder Schatten produziert. In Videos sollte man auf Reflexionen in Linsen und kleine Merkmale wie Muttermale oder Gesichtsbehaarung achten, die möglicherweise nicht so natürlich erscheinen.

Fahre fort mit dem Quiz:

Wenn du eine Menti-Umfrage verwendest, teile den Code mit der Gruppe. Zeige den "Präsentations"-Modus auf [mentimeter.com](https://www.mentimeter.com) auf dem großen Bildschirm. Bitte die Teilnehmenden, eine Antwort auszuwählen. Sobald alle abgestimmt haben, werden die Ergebnisse auf dem Hauptbildschirm angezeigt.

Wenn du keine Menti-Umfrage verwendest, zeige die Beiträge nacheinander auf einem großen Bildschirm und lasse die Teilnehmenden abstimmen, ob sie denken, dass ein Bild echt oder KI-generiert ist, z.B. durch Handzeichen.

Nach jedem Bild, besprecht kurz, welche Anzeichen auf KI-generierte Bilder hingewiesen haben. Hier ist die Lösung des Quiz und einige Punkte, die zu beachten sind:

1: Olaf Scholz: Echt

- Bundeskanzler Olaf Scholz hatte im September 2023 einen Unfall beim Laufen im Wald. Er stürzte und verletzte sich im Gesicht. Um sein verletztes Auge zu bedecken, trug er eine Augenklappe und postete dieses Foto öffentlich auf seiner X-Seite.
- Quelle: [Olaf Scholz auf X](#)
- Bildnachweis: Steffen Kugler/Bundesregierung

2: Donald Trump: KI-generiert

- Dieses Bild wurde von Anhängern von Donald Trump erstellt.
- Faktencheck hier: [BBC News](#)
- Indikatoren, dass es gefälscht ist:
 - Personen fehlen Finger

- Der Raum zwischen zwei Fingern von Trumps linker Hand sieht unrealistisch aus
- Der Text auf dem Hut links ist verschwommen und unlesbar
- Textur der Haut: zu glänzend

3: Katy Perry: KI-generiert

- Dieses Bild kursierte während der MET Gala 2024. Allerdings war Katy Perry nicht einmal anwesend.
- Faktencheck hier: [Politifact](#)
- Indikatoren, dass es gefälscht ist:
 - Im Hintergrund sind die Gesichter, Hände, Arme und Körper der Fotografen nicht klar voneinander oder von der Umgebung getrennt, sondern fließen ineinander über.

4: Giorgia Meloni: Echt

- Dieses Bild wurde am 9. Februar 2023 im Elysée-Palast aufgenommen.
- Quelle: [FAZ](#)
- Bildnachweis: EPA

5: Tom Cruise: KI-Generiert

- Faktencheck: [Snopes](#)
- Indikatoren, dass es gefälscht ist:
 - Die Hand des angeblichen Tom Cruise-Doubles in der Mitte scheint keine Fingernägel zu haben
 - Die Textur der Haut ist zu glatt, glänzend und wachsartig

Nach dem Quiz, gib ein paar weitere Ratschläge und Tipps

Fortgeschrittene Techniken zum Erkennen von KI

- **AI-Based Detection Software:** Verschiedene Tools werden entwickelt, um KI-generierte Inhalte zu erkennen. Diese Programme analysieren Unstimmigkeiten in Bildern und Videos, die möglicherweise für das menschliche Auge nicht sichtbar sind. Sie können besonders effektiv bei der Identifizierung von Deepfakes sein, indem sie feine Details untersuchen und mit bekannten Mustern von Manipulationen vergleichen.

Für Bilder z.B.:

[Hive Detection Tool](#)

Für Videos z.B.:

[Deepware Scanner](#) or [DeepFake-o-meter](#).

- **Menschliche Analyse:** Trotz der Fortschritte in der KI-Erkennungssoftware bleibt menschliche Intuition und Analyse von entscheidender Bedeutung.
- **Gesunder Menschenverstand:** Das gilt für alle Informationen im Internet. Pause, bleibe skeptisch und kritisch und hinterfrage, was du siehst und hörst. Welche Motive könnten dahinterstecken, bestimmte Personen solche Aussagen machen zu lassen? Wie realistisch ist das Szenario?

Anweisungen für die Teilnehmenden:

Öffnet menti.com auf euren Handheld-Geräten. Gebt den Code ein, den die Workshopleitung euch gibt. Schaut euch das Bild an, das auf eurem Gerät angezeigt wird, und wählt aus, ob ihr denkt, dass es echt oder von KI-generiert ist.

Mögliche Anpassungen:

Die Workshopleitung kann auch eigene Beispiele für das Quiz verwenden, wenn sie andere Bilder zeigen möchte.

Nachbesprechung:

Fandet ihr diese Aktivität herausfordernd oder einfach? Was ist eure wichtigste Erkenntnis aus dieser Aktivität?

Hintergrundlektüre:

<https://www.theguardian.com/technology/2024/apr/08/how-to-tell-if-an-image-is-ai-generated>

Aktivität 4: Prompt Battle – Wer erstellt das realistischste KI-Bild

Typ: Spiel / Kreativwettbewerb

Dauer: 15 Minuten

Lernergebnisse: Die Teilnehmenden werden ein besseres Verständnis dafür entwickeln, wie KI-Bilder erstellt werden und wie die Anweisungen („Prompts“), die man gibt, das Ergebnis beeinflussen.

Materialien/Ressourcen:

- Notizbücher und Stifte für die Teilnehmenden.
- Einführungsskript für die Workshopleitung.
- Großbildschirm oder Projektor und ein Laptop, der damit verbunden ist, für die Workshopleitung.

Anweisungen für die Workshopleitung:

Teilt die Gruppe in zwei Teams auf. Mit der kostenlosen Microsoft Copilot Designer Software (oder kostenpflichtiger Software wie Dall-E oder Midjourney) treten die Teams gegeneinander an, um das realistischste KI-Bild zu einem vorgegebenen Thema zu erstellen.

Es wird in drei Runden gespielt. In jeder Runde kann ein Team ein Thema aus der folgenden Liste wählen oder ein eigenes Thema erfinden. Danach hat jedes Team 3-5 Minuten Zeit, um ein detailliertes Prompt zu erstellen, um ein realistisches Bild zu erhalten. Achtet darauf, dass die Teams das Prompt des anderen Teams nicht hören oder sehen können. Nach Eingabe des Prompts erstellt die Software vier ähnliche Bilder, aus denen die Teams ihr Lieblingsbild auswählen. Wenn die Zeit abgelaufen ist, präsentieren beide Gruppen die Ergebnisse. Die ganze Gruppe kann die Bilder betrachten und entscheiden, welches Bild realistischer aussieht. Nach drei Runden wird das Team gekürt, das die realistischsten Bilder erstellt hat.

Themenliste:

- Eine Szene einer Protestaktion von Klimaaktivisten
- Ein Parfüm-Werbeplakat mit einem Modell am Strand
- Ein Politiker im Wahlkampf, der vor einer großen Menschenmenge spricht (Hinweis: Viele KI-Generatoren stellen keine echten Persönlichkeiten dar, z.B. Donald Trump, um gegen Desinformation zu schützen)
- Eine verheerende Waldbrand
- Paparazzi, die Prominente jagen

Tipps zur Erstellung eines guten Prompts (können im Voraus präsentiert werden):

- Beschreibe so präzise wie möglich, was im Bild zu sehen sein soll (Motiv im Vordergrund, Hintergrund, Details, wie was angeordnet ist...)
- Gib den Stil des Bildes an: fotorealistisch, Malerei, schwarz-weiß, etc.
- Kontext: Gib Hintergrundinformationen, Daten oder Kontext für die präzise Erstellung des Inhalts an.
- Schlüsselwörter: Nenne wichtige Schlüsselwörter oder Phrasen, die enthalten sein sollen.
- Beispiele: Gib Beispiele für den gewünschten Stil, die Struktur oder den Inhalt.
- Verwende Lichteffekte und Stimmungen, um deinem Bild einen bestimmten Stil zu verleihen, z.B. Kino-Beleuchtung, Weihnachtsbeleuchtung oder Beschreibungen wie düster, fröhlich, hell, verschwommen, etc.
- Wichtig: Prompts funktionieren derzeit besser auf Englisch.

Nachbesprechung: Diskutiere anschließend die verschiedenen verwendeten Prompts und versuche, Fehler in den Bildern und Anzeichen für KI-Generierung zu erkennen. Ähnlich wie beim „Get Bad News“-Spiel wechselt diese Aktivität die Perspektiven. Indem man sich in die Rolle des Erstellens eines täuschend echten Bildes mit KI versetzt, lernt man, wie schnell und einfach es heutzutage ist, realistisch aussehende Bilder mit den richtigen Prompts zu erstellen. Die Teilnehmenden üben, Fehler in ihren eigenen Bildern zu erkennen, die auf ein KI-Produkt hinweisen. Dies ist ein unterhaltsamer und spielerischer Ansatz, um junge Menschen dafür zu sensibilisieren, Bilder im Internet genauer zu betrachten.

Mögliche Anpassungen: Statt zwei Teams können auch nur zwei Personen gegeneinander antreten. In diesem Fall wechseln sich die konkurrierenden Personen in jeder Runde ab. Neben den vorgegebenen Themen können natürlich auch andere Themen oder Ideen für KI-Bilder gewählt werden.



Mit KI erstellt