# Instagram-Scraper und KI-Check

Welche Parteien und Politiker/innen setzen besonders viel KI ein beim Posten? Mit KI-Detektoren kann man viele Fälschungen erkennen**. (Aber bitte unbedingt den Abschnitt ganz unten dazu lesen!)** Spinoff eines Recherche-Projekts im ARD-KI-Netzwerks, sehr spät, aber vielleicht doch noch nützlich.

## Einmal checken

* Einzelne Bilder und Videos checken: <https://frankruft.de/apps/bildcheck>
* Einzelne Instagram-Konten checken: <https://frankruft.de/apps/Instalupe>
* Einzelne Telegram-Channels(!): <https://frankruft.de/apps/Klingeldraht>

Unter <https://frankruft.de/apps/Chemtrail> gibt es auch noch eine App für Bluesky, aber die ist veraltet und funktioniert womöglich nicht mehr zuverlässig.

## Immer wieder checken

Auf meinem frankruft.de-Server läuft ein kleines Programm, das abends um 22 Uhr unsere Liste mit interessanten Instagram-Accounts durchgeht, neue Posts, Highlights und Stories archiviert und analysiert, und das Ergebnis verlinkt.

Zentrale Steuerungs-Instanz für das Programm ist **ein Google-Doc** mit der ID:

* [1Tr1YU8zVu7AFBWy8HS9ZWVxFUgQPc51rvf-UlrXRXXM](https://docs.google.com/spreadsheets/d/1Tr1YU8zVu7AFBWy8HS9ZWVxFUgQPc51rvf-UlrXRXXM/edit?gid=835993928#gid=835993928)

## Welche Instagram-Konten werden überwacht?

**Alle, die in dem Google-Doc aufgelistet sind.**

Das Programm arbeitet sich abends durch die Liste durch, ruft alle der Konten auf, prüft, ob es neue Posts, Stories und Highlights gibt. Medien werden auf dem Server archiviert und mit unseren KI-Detektoren überprüft, ob sie KI-generiert sind.

## Wenn ich ein anderes Konto überprüfen lassen möchte?

Wenn du nur einmal prüfen willst: Nutze den [KI-Check für Instagram](https://frankruft.de/apps/Instalupe/) (er greift übrigens auf dieselben Datenbanken zu und prüft beim ersten Mal 30 Posts).

Wenn ein Konto dauernd überwacht werden soll: wende dich an eine/n der Admins, die Schreibzugriff auf das Google-Doc haben – z.B. [jan.eggers@hr.de](mailto:jan.eggers@hr.de)

Wir müssen mit den Prüfungen **allerdings sehr restriktiv umgehen**, weil wir nur begrenzt Zugriff auf die KI-Erkennung haben. Mehr als 200 sollten wir nicht beobachten.

## Wie zuverlässig sind die KI-Detektoren?

**Leider nicht sehr.** Unveränderte Bilder werden gut erkannt und haben in der Regel Werte von 90% oder höher (0,99 in der Tabelle). Aber sobald das Bild zugeschnitten oder verändert wird, sinken die Erkennungswerte. Und mitunter erzeugen völlig harmlose, natürliche Bilder einen KI-Erkennungswert von 0,6 oder höher.

Noch schlechter sieht es bei der KI-Text-Erkennung aus, die prinzipbedingt nicht so gut funktionieren kann. Hier gibt es eine Menge falsch-positive und falsch-negative Scans – und daran denken: **wenn der Detektor sagt, dass etwas mit einer Wahrscheinlichkeit von 90% KI generiert ist (0,9), dann heißt das: in 1 von 10 Fällen liegt er falsch!** Ein einzelnes Bild kann man nie allein mit KI-Detektoren als KI-Fälschung belegen.

## Wie komme ich an die Daten aus den Scans?

Der Scraper schreibt die Ergebnisse der Scans in die Google-Tabelle – insbesondere in die Spalte 4 („xlsx“). Dort findet sich für jedes gescannte Konto eine **Excel-Datei zum Herunterladen, die alle bisher gescannten Posts und Medien auflistet**. (Neben KI-Erkennungswerten finden sich dort auch Transkriptionen von Videos/Audios und Beschreibungen von Bildern.)

In der Excel-Datei hat jedes Bild, jedes Video, jedes Audio eine Zeile – und in der Excel-Tabellen-Spalte „file“ findet sich ein Link, über den man das jeweilige Bild/Video direkt aus dem Server-Archiv abrufen kann. So kommt man auch an die Stories von vorgestern!

## Wo kommt das eigentlich her?

Die Apps haben wir im Recherche-Team gebaut; sie laufen auf meinem (Jans) Server frankruft.de. Die Programme setzen drei KI-Detektoren ein:

* [Detectora](https://detectora.de/) – für deutsche KI-Texte; eher experimentell
* [Hive](https://hivemoderation.com/) – ein größerer kommerzieller KI-Erkennungsdienst
* [AIORNOT](https://aiornot.com/) – ein kleinerer kommerzieller KI-Erkennungsdienst

Je mehr Detektoren man im Ensemble einsetzt, desto sicherer kann man bei den Ergebnissen sein.

Fun Fact zum Schluss: Die Nutzung von AIORNOT ist möglich, weil der WDR für zwei Monate den Zugang gekauft hat; die Hive-Nutzung läuft über ein sehr eingeschränktes Test-Account der Tagesschau. Wir sind eine ARD!