

TRAINING VON KI-MODELLEN MIT HILFE KIRCHLICHER WEBSEITEN

Kristof Lehoczky

Stabsstelle Digitalisierung, Kirchenamt der EKD

Gliederung



Bildquelle:
https://cf.katholisch.de/adobestock_295876765_digital.jpg

KI und das Training von KI-
Modellen (+ Kirche)

DSGVO-konformität von
Webscraping

Kirchliche
Webseitenbetreiber und ihre
Pflichten gem. DSGVO-EKD.

KI und das Training von KI-Modellen

Künstliche Intelligenz ist ein Oberbegriff für Methoden, die auf Automatisierung von Entscheidungsvorgängen abzielen, die traditionell den Einsatz menschlicher Intelligenz erfordern. (*Bundesamt für Sicherheit und Informationstechnik*)

Erforderlichkeit von Trainingsdaten

Warum ist dieses Thema für die Kirche wichtig?

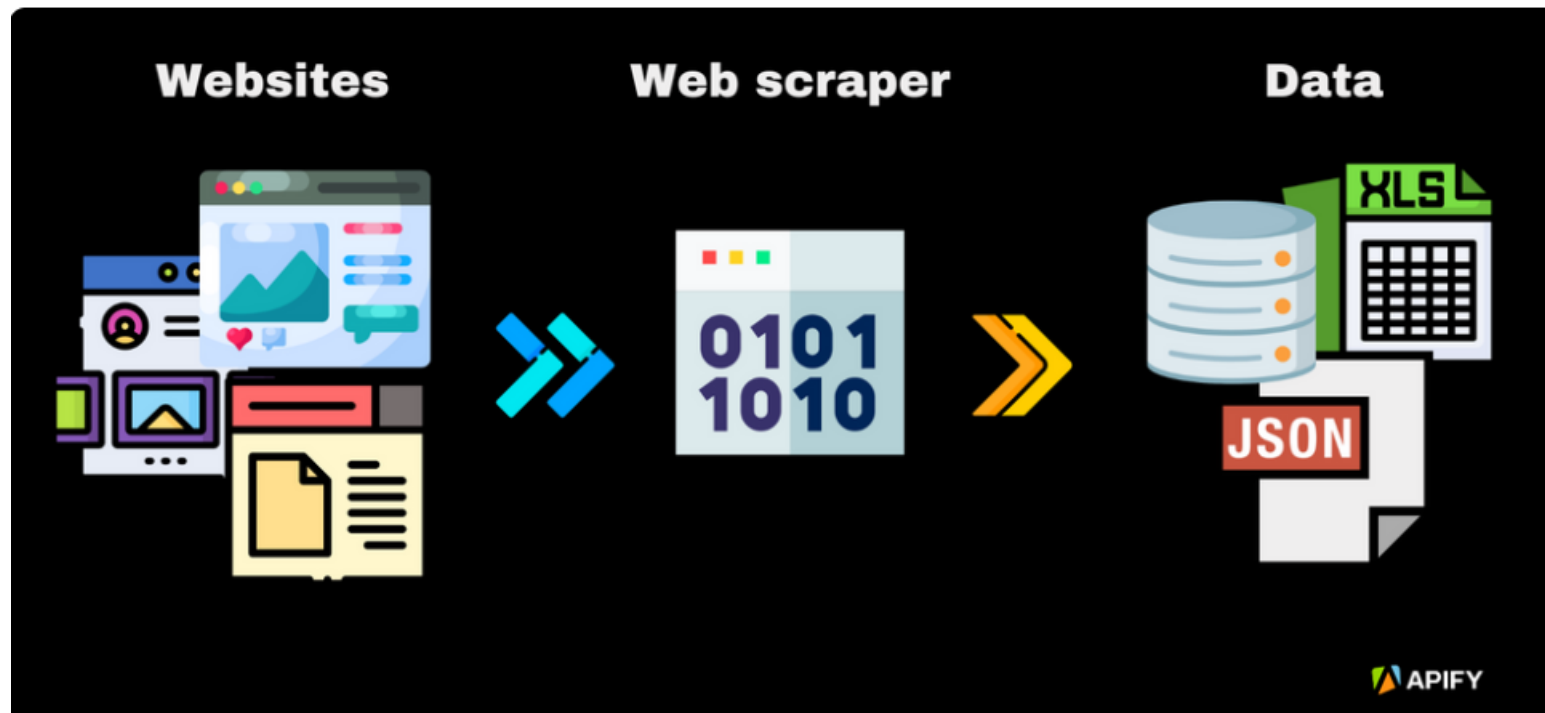
Erwachsenen- und Familienbildung⁴

Art der Veranstaltung	Anzahl	Teilnahmen
Religion und Ethik	27.979	622.405
Kultur und Gestalten	22.699	553.749
Familie und Generationen	21.958	400.801
Politik und Gesellschaft	15.042	403.444
Gesundheit	14.207	209.242
Qualifikationserwerb	3.220	49.145
Engagementqualifizierung	3.300	45.060
Insgesamt	108.405	2.283.846

⁴ Quelle: Statistik 2019 der Deutschen Evangelischen Arbeitsgemeinschaft für Erwachsenenbildung.

KI und das Training von KI-Modellen (2): Webscraping

- ▶ Webscraping: eine Sammlung von Verfahren (Computerprogramm), mit denen Webseiten durchgesucht (Indexieren der Inhalte) und Daten aus denen extrahiert werden können (=Erhebung von Daten zu Trainingszwecke)



Bildquelle: <https://blog.apify.com/what-is-web-scraping/>

DSGVO-konformität von KI-Training

- ▶ Ausgangspunkt: Massenhafte Erhebung von öffentlich zugänglichen Daten von Webseiten für Trainingszwecke
- ▶ DSGVO-konformität des KI-Trainings (Orientierungshilfe der DSK)
 - ▶ Wurden für das Training personenbezogene Daten verwendet?
 - ▶ Falls ja, gab es eine Rechtsgrundlage für die Nutzung der Daten für das Training?
 - ▶ Inwiefern weist die KI-Anwendung zum Zeitpunkt des Einsatzes selbst Personenbezug auf?

DSGVO-konformität von KI-Training (2)

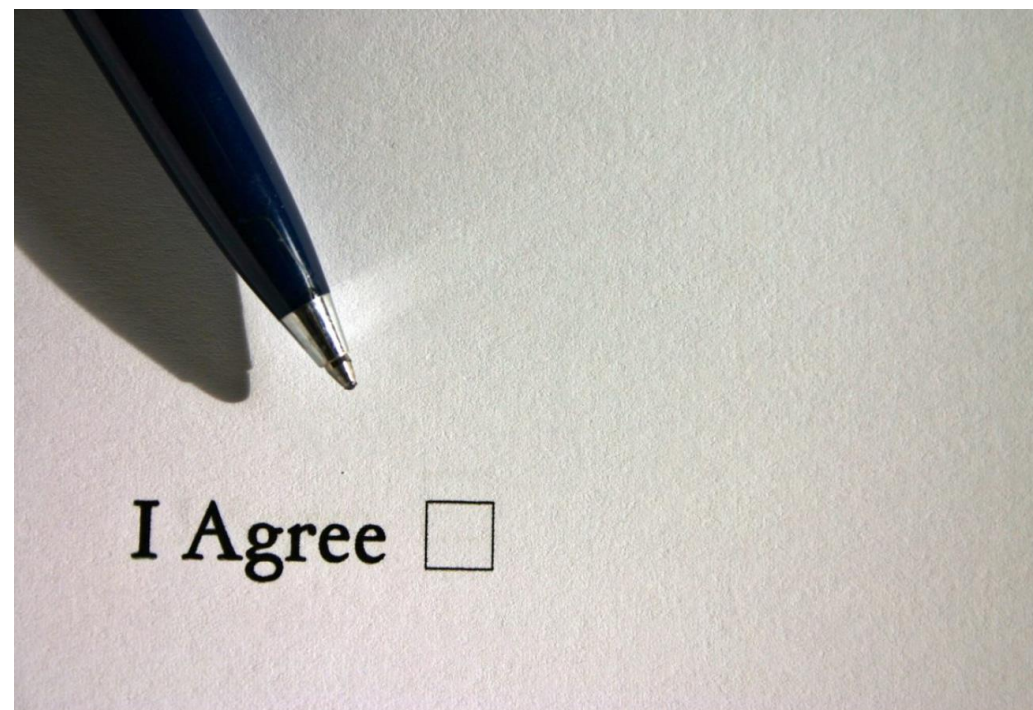
- ▶ Personenbezogene Daten auf kirchlichen Webseiten: wahrscheinlich ja
z.B.:
 - ▶ Namen, Fotos von Kirchenvorstehern
 - ▶ Fotos von der Jugendfreizeit

DSGVO-konformität von KI-Training (3) Rechtsgrundlage der Verarbeitung: Art.6 I DSGVO

1. Die Verarbeitung ist nur rechtmäßig, wenn mindestens eine der nachstehenden Bedingungen erfüllt ist:
 1. *Die betroffene Person hat ihre **Einwilligung** zu der Verarbeitung der sie betreffenden personenbezogenen Daten für einen oder mehrere bestimmte Zwecke gegeben;*
 2. *die Verarbeitung ist für die Erfüllung eines Vertrags, dessen Vertragspartei die betroffene Person ist, oder zur Durchführung vorvertraglicher Maßnahmen erforderlich, die auf Anfrage der betroffenen Person erfolgen;*
 3. *die Verarbeitung ist zur Erfüllung einer rechtlichen Verpflichtung erforderlich, der der Verantwortliche unterliegt;*
 4. *die Verarbeitung ist erforderlich, um lebenswichtige Interessen der betroffenen Person oder einer anderen natürlichen Person zu schützen;*
 5. *die Verarbeitung ist für die Wahrnehmung einer Aufgabe erforderlich, die im öffentlichen Interesse liegt oder in Ausübung öffentlicher Gewalt erfolgt, die dem Verantwortlichen übertragen wurde;*
 6. *die Verarbeitung ist zur **Wahrung der berechtigten Interessen** des Verantwortlichen oder eines Dritten erforderlich, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt.*

DSGVO-konformität von KI-Training (4): Einwilligung des Betroffenen

- ▶ Hohe Anzahl der Betroffenen (+ indirekte Erhebung)
- ▶ Zweck der Verarbeitung kann sich ändern
- ▶ Schwierigkeit bezüglich der Widerrufbarkeit (Löschen, Vergessen, Umtrainieren)



Bildquelle: <https://blog.melchers-law.com/wp-content/uploads/Einwilligung-1200x829.jpg>

DSGVO-konformität von KI-Training (5): Berechtigtes Interesse des KI-Entwicklers

- ▶ „ Die Verarbeitung ist nur rechtmäßig, wenn [...] die Verarbeitung [...] zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich [ist], *sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt.* (Art. 6. I lit. f DSGVO)



DSGVO-konformität von KI-Training (6): Berechtigtes Interesse des KI-Entwicklers, Erforderlichkeit, keine behördliche Aufgabenerfüllung

- ▶ **Jedes ideelle und wirtschaftliche Interesse** (Art. 29-Datenschutzgruppe)
 - ▶ Gewinnerzielung (z.B.: Entwicklung und Verbesserung eines Produktes)
 - ▶ Wissenschaftliche Forschung
- ▶ **Erforderlichkeit** zur Entwicklung eines trainierten KI-Modells
- ▶ **Keine behördliche Aufgabenerfüllung** (passende Grundlage: Art. 6. I. lit. e)

DSGVO-konformität von KI-Training (7): Interessen, Grundrechte und Grundfreiheiten des Betroffenen, kein wirksamer Widerspruch

- ▶ **Aspekte zur Berücksichtigung:**
 - ▶ **Vernünftige Erwartungen** bzgl. potenzieller, weiterer Verarbeitung (Erstveröffentlichung)
 - ▶ **Intimitätsgrad** der Daten
 - ▶ Mögliche, **negative Auswirkungen** z.B.: Diskriminierung, Identitätsdiebstahl.... (Erwg. 75. DSGVO)
 - ▶ **Begünstigende Auswirkungen** auf die Betroffenen
 - ▶ **Möglichkeit eines Widerspruchs** gem. Art. 21 DSGVO
- ▶ **Personenbezogene Daten von Kindern:** besonderer Schutz, aber die Rechtfertigung ist möglich

DSGVO-konformität von KI-Training (8): Personenbezug beim Einsatz von KI-Modellen

- ▶ Resultat eines Hackerangriffes oder Softwarefehlers
 - ▶ Eintrittswahrscheinlichkeit durch Einsatz von geeigneten TOMs senken
- ▶ Neue Entwicklung: Im LLM-System selbst ist kein Personenbezug vorhanden (*Hamburger Thesen zum Personenbezug in Large Language Models*)

Zwischenfazit: Ein DSGVO-konformes Training von KI-Modellen mithilfe von aus Webseiten erhobenen Inhalten ist möglich.



Bildquelle: https://datenschutz-hamburg.de/fileadmin/_processed_/b/c/csm_AdobeStoc_k_745718842_fd9d77b4d1.jpeg

Die Lage aus der Sicht der kirchlichen Webseitenbetreiber

- ▶ **Einschlägige Rechtsnorm:** Kirchengesetz über den Datenschutz der Evangelischen Kirche in Deutschland (DSG-EKD) (Grundlage für die eigene Datenschutzregelung: Art. 91 DSGVO)
- ▶ **Rolle der Webseitenbetreiber hinsichtlich des KI-Trainings gem. DSGVO:**
 - ▶ Verantwortlicher: NEIN
 - ▶ Auftragsverarbeiter: NEIN
 - ▶ Betroffener: NEIN
- ▶ **Rolle gem. DSG-EKD: verantwortliche Stelle** gem. § 4. Nr. 9. DSG-EKD

Pflichten der verantwortlichen Stelle gem. DSGVO-EKD

- ▶ Kein Verbot durch DSGVO-EKD
- ▶ Informationspflicht
- ▶ Schutz personenbezogener Daten



Bildquelle:
https://www.ekd.de/ekd_de/ds_img/assets_3420_x/20180322_Thinkstock_thanasus_smartphone_3420_rdax_1140x541_60s.jpg

Informationspflicht

- ▶ Gebot der **Transparenz** und **Informationspflicht** (§ 16 und § 17 DSGVO-EKD)
 - ▶ Mögliche Lösung: Ergänzung der Datenschutzerklärung
- ▶ **Informierte Einwilligung** (gem. § 6 Nr. 1 DSGVO-EKD)
 - ▶ Hinweis auf die Datenschutzerklärung
 - ▶ Information bezüglich der Änderung der Datenschutzerklärung?

Impressum

Datenschutz

Cookie-Einstellungen

Transparenz

Bildquelle: Evangelische Kirche in Deutschland – EKD

Schutz vor Webscraping

- ▶ *“Consider if you publish personal data on your own website. This may be from your staff or from your website users. You may need to ensure that you protect that personal data from being collected and used for AI training or other processing where you have not already agreed that purpose with your staff or users, or if they do not have a reasonable expectation it will be used for AI training.” (DPC, Irische Datenschutzkommission)*
- ▶ Entscheidung über Veröffentlichung der personenbezogenen Daten.
- ▶ Schutz der veröffentlichten Daten: **Robots.txt**

Schutz vor Webscraping (2): Robots.txt

```
User-agent: *  
Disallow: /test/  
User-agent: Applebot  
Disallow: /magazine/  
Allow: /magazine/in-eigener-sache/  
Allow: /magazine/unicef/  
Allow: /magazine/so-arbeitet-die-redaktion/  
User-agent: ChatGPT-User  
Disallow: /magazine/  
Allow: /magazine/in-eigener-sache/  
Allow: /magazine/unicef/  
Allow: /magazine/so-arbeitet-die-redaktion/  
User-agent: GPTBot  
Disallow: /magazine/  
Allow: /magazine/in-eigener-sache/  
Allow: /magazine/unicef/  
Allow: /magazine/so-arbeitet-die-redaktion/  
User-agent: Google-Extended  
Disallow: /magazine/  
Allow: /magazine/in-eigener-sache/  
Allow: /magazine/unicef/  
Allow: /magazine/so-arbeitet-die-redaktion/
```

Quelle: [gmx.net/robots.txt](https://www.gmx.net/robots.txt)

- ▶ Spielregeln für Webscraping
- ▶ Kein 100-prozentiger Schutz

Fazit:

Das datenschutzkonforme Training von KI-Modellen mithilfe kirchlicher Webseiten ist unter Berücksichtigung der Rahmenbedingungen möglich.

Vielen Dank für Ihre Aufmerksamkeit