

UNTERRICHTSEINHEIT 03

KI-Ethik, Bias, Datenschutz, Umwelt & digitale Souveränität

Die tiefergehenden Verantwortlichkeiten von KI;
und was Bibliotheken verstehen, hinterfragen
und aktiv vertreten müssen

🕒 60–90 Min. | Alle Bibliotheksmitarbeitenden

Warum diese Fragen nicht warten können

Die Verbreitung von KI schreitet schneller voran als die ethische Reflexion darüber. Bibliotheken müssen dieser Entwicklung voraus sein.



Ethische Fehlentwicklungen passieren bereits heute

Systeme zur Gesichtserkennung haben unschuldige Menschen fälschlicherweise identifiziert. Recruiting-Algorithmen haben Frauen systematisch benachteiligt. Das sind keine hypothetischen Zukunftsszenarien.



Bias ist strukturell, nicht zufällig

KI-Bias entsteht aus Trainingsdaten, die historische Ungleichheiten widerspiegeln. Er verschwindet nicht von selbst, sondern erfordert gezielte Gegenmaßnahmen.



Nutzer*innendaten sind gefährdet

Jedes KI-System, das Suchanfragen, Rechercheverhalten oder Lesehistorien verarbeitet, wirft erhebliche Vertraulichkeitsfragen auf, die Bibliotheken rechtlich und ethisch berücksichtigen müssen.



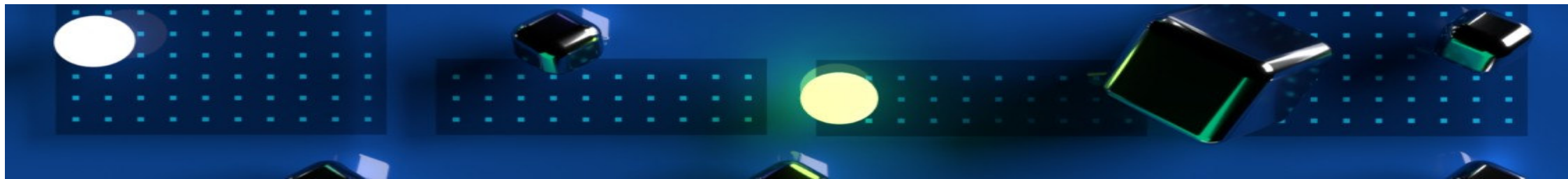
Die ökologischen Kosten sind real

Das Training eines einzelnen großen KI-Modells kann ähnlich viele CO₂-Emissionen verursachen wie fünf Autos während ihrer gesamten Lebensdauer. Bibliotheken mit Nachhaltigkeitsanspruch müssen dies berücksichtigen.



Wer KI kontrolliert, kontrolliert Wissen

Eine kleine Anzahl von Unternehmen kontrolliert die KI-Systeme, die weltweit beeinflussen, wie Informationen gefunden, gefiltert und dargestellt werden. Das ist eine Frage der digitalen Souveränität.



KI-Ethik

Was ist richtig? Was ist schädlich? Wer entscheidet darüber?
Ethik ist keine Einschränkung guter KI. Sie ist ihre Voraussetzung.

Wichtige ethische Rahmen für KI

Globale Institutionen haben Leitlinien entwickelt. Bibliotheksmitarbeitende sollten diese kennen und anwenden.

UNESCO (2021)

Menschenwürde und Inklusion

Das erste globale normative Instrument für KI-Ethik. Es betont Menschenrechte, Geschlechtergerechtigkeit, ökologische Nachhaltigkeit und friedliche Gesellschaften. KI in Bildung und Informationsvermittlung wird ausdrücklich berücksichtigt.

EU AI Act (2024)

Risikobasierte Regulierung

Klassifiziert KI-Systeme nach Risikostufen (inakzeptabel → hoch → begrenzt → minimal). Bibliotheken, die KI etwa für Nutzer*innenprofiling oder Personalentscheidungen einsetzen, können in Hochrisikobereiche fallen.

IFLA (2020)

Zugang, Datenschutz, Vertrauen

Die Internationale Vereinigung bibliothekarischer Verbände betont offenen Zugang, Datenschutz, Schutz vor Überwachung und das Recht auf Informationszugang ohne algorithmische Vermittlung.

OECD (2019)

Inklusives Wachstum und Rechenschaftspflicht

Fünf Grundprinzipien: inklusives Wachstum, menschenzentrierte Werte, Transparenz, Robustheit und Verantwortung. Die Prinzipien wurden von den G20 übernommen und prägen zahlreiche nationale KI-Strategien.

Wie KI Schaden verursacht: Typologie

Die verschiedenen Arten verstehen, auf die KI-Systeme reale Schäden verursachen können.

Schadensart	Beispiel	Relevanz für Bibliotheken
Verteilungsbezogener Schaden	KI verweigert Kredite, Jobs oder Sozialleistungen auf Basis verzerrter Vorhersagen.	KI verweigert Dienstleistungen oder markiert Konten aufgrund demografischer Muster.
Repräsentativer Schaden	KI verbindet bestimmte Gruppen mit negativen Eigenschaften.	Discovery-Systeme machen Literatur aus dem Globalen Süden weniger sichtbar.
Qualitätsbezogener Schaden	Spracherkennung funktioniert schlechter für Frauen oder Nicht-Muttersprachler*innen.	KI-gestützte Auskunftssysteme funktionieren für englischsprachige Personen besser als für andere.
Zwischenmenschlicher Schaden	Deepfakes werden zur Belästigung, Verleumdung oder Manipulation eingesetzt.	Nutzer*innen werden auf Basis bibliotheksbezogener Daten gezielt mit Desinformation angesprochen.
Erkenntnisbezogener Schaden	KI liefert überzeugend falsche Informationen und untergräbt das Vertrauen in Fakten.	Nutzer*innen zitieren KI-Halluzinationen in wissenschaftlichen oder gesellschaftlichen Kontexten.



Algorithmischer Bias

Bias ist kein Fehler. Er ist oft Bestandteil der Daten.

Wenn die Welt ungerecht ist und KI von dieser Welt lernt, dann lernt KI auch deren Ungerechtigkeiten.

Wie Bias in KI-Systeme gelangt

Bias entsteht entlang der gesamten Entwicklungskette von KI, nicht nur in den Daten.

Daten- erhebung

Unterrepräsentation von Minderheiten; historische Diskriminierung in Datensätzen; Internetdaten, die wohlhabende englischsprachige Bevölkerungsgruppen überrepräsentieren

Daten- annotation

Annotierende Personen bringen kulturelle Vorannahmen ein; Kategorien spiegeln häufig die Perspektive der Annotierenden wider, nicht die der betroffenen Gemeinschaften

Modelldesign

Optimierungsziele können schlechte Ergebnisse für Minderheiten verdecken; Architekturentscheidungen enthalten implizite Annahmen

Einsatzkontext

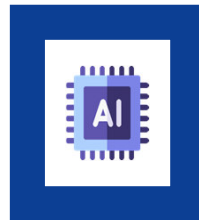
Modelle werden oft in anderen Kontexten eingesetzt als jenen, für die sie entwickelt wurden

Feedback- schleifen

Verzerrte Ergebnisse erzeugen neue verzerrte Daten und verstärken das Problem im Lauf der Zeit.

Algorithmischer Bias

„Eine Gruppe von
Golfspieler*innen“



Algorithmischer Bias



Algorithmischer Bias

≡ *Translate*

English ▼

Turkish ▼

He is a doctor



0 bir doktor



Algorithmischer Bias

Garbage in, garbage out (GIGO)

Fehlerhafte Eingabedaten führen zu fehlerhaften Ergebnissen.





Algorithmischer Bias

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

SCORING OF WELFARE BENEFICIARIES: THE INDECENCY OF CAF'S ALGORITHM NOW UNDENIABLE

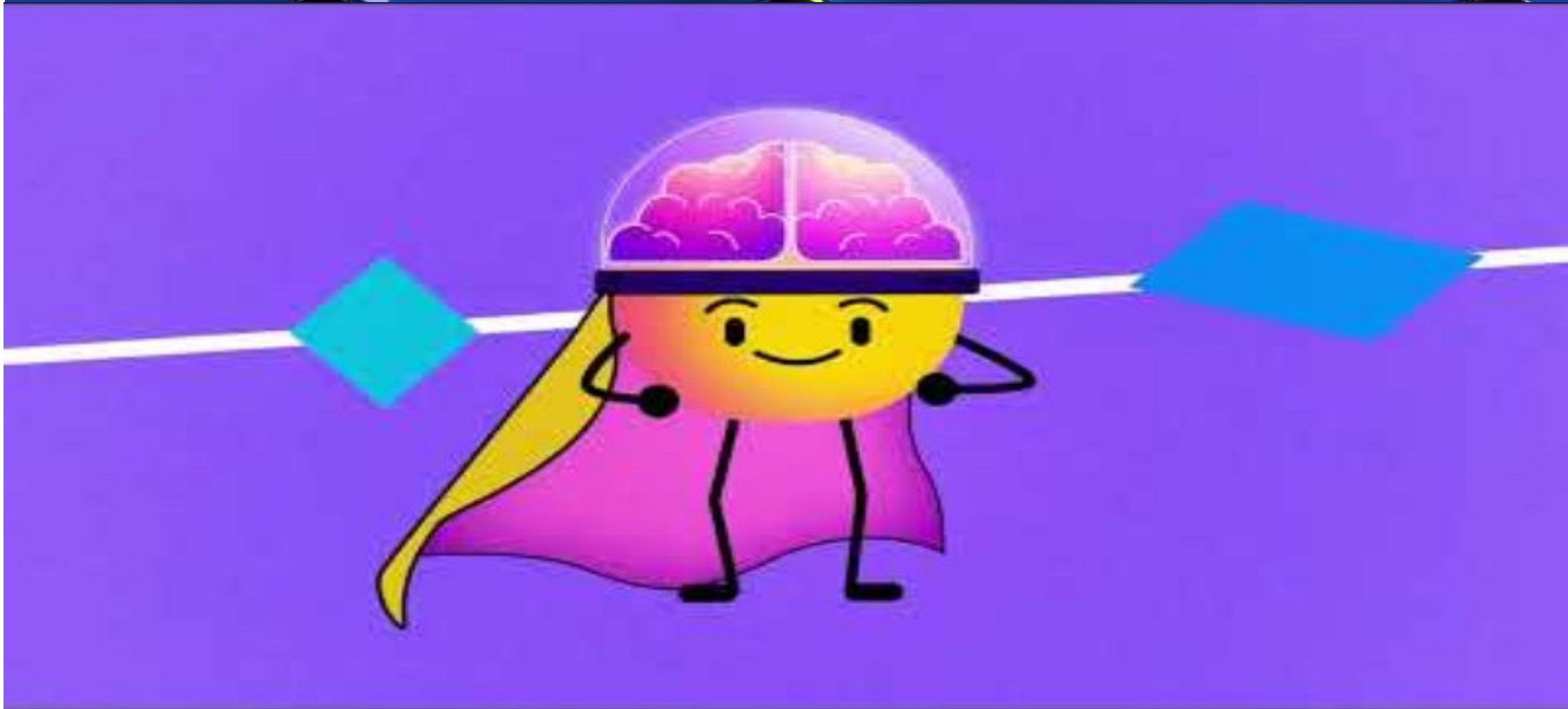
27 November 2023



Algorithmischer Bias

- Das KI-gestützte Recruiting-System von Amazon wurde als benachteiligend gegenüber Frauen eingestuft, da es auf historisch männlich dominierten Daten aus der Technologiebranche trainiert wurde.
- Untersuchungen aus dem Bereich Strafverfolgung zeigen deutlich höhere Fehlerraten bei der Erkennung dunkler Hauttöne.
- Medizinische Algorithmen haben Verzerrungen bei der Einschätzung von Behandlungsbedarfen schwarzer Patient*innen gezeigt.
- Wohnungs- und Stellenanzeigen können algorithmisch anhand demografischer Merkmale ausgespielt werden.

Algorithmischer Bias





Datenschutz und Überwachung

„Wenn du nicht für das Produkt bezahlst, bist du das Produkt.“

Daten von Bibliotheksnutzer*innen sind keine Ressource. Sie sind ein Recht,
das geschützt werden muss.

Wie KI-Systeme Daten erfassen und verwenden

Die Datenflüsse sind oft unsichtbar. Ihre Folgen sind es nicht.

Die Datenflussskette

- 1 Nutzer*innen stellen eine Anfrage über ein KI-System.
- 2 Die Anfrage wird über eine externe KI-Schnittstelle verarbeitet.
- 3 Der Anbieter speichert die Anfrage zur Verbesserung des Modells.
- 4 Nutzungsdaten werden aggregiert und möglicherweise weiterverwertet.
- 5 Es entstehen Profile, die Such-, Lese- und Anfrageverhalten verknüpfen.
- 6 Diese Profile können unter bestimmten rechtlichen Bedingungen für Behörden zugänglich sein.

Bestehende Schutzmechanismen

- ▶ DSGVO: Auskunft, Löschung und Datenminimierung
- ▶ Nationale Bibliotheksgesetze zum Schutz der Vertraulichkeit
- ▶ IFLA Privacy Framework

Zentrale Risiken

- ▶ Datenübermittlungen in Drittländer
- ▶ Umfangreiche Nutzungsrechte für Anbieter
- ▶ Offenlegung sensibler Informationen

Wie KI-Systeme Daten erfassen und verwenden

Die Vertraulichkeit der Nutzung ist eine rechtliche und ethische Grundlage. KI hebt sie nicht auf.

Grundprinzip: Die Verpflichtung einer Bibliothek zum Schutz der Privatsphäre endet nicht dort, wo KI ins Spiel kommt. Jedes KI-System muss dieselben Standards erfüllen wie andere Bibliothekssysteme.

1

Bevor ein KI-System eingeführt wird, das Nutzer*innendaten verarbeitet, sollte eine dokumentierte **Datenschutz-Folgenabschätzung (DPIA)** erstellt werden. Dabei sollte geprüft werden, welche Daten erhoben werden, wohin sie fließen und welche Risiken mit ihrer Verarbeitung verbunden sind.

2

KI-Anwendungen sollten so konfiguriert werden, dass sie **nur die unbedingt notwendigen Daten erfassen**. Personenbezogene Informationen sollten nach Möglichkeit niemals in Prompts oder Anfragen

3

Die **Datenschutzbestimmungen und Nutzungsbedingungen** jedes KI-Anbieters sollten vollständig gelesen und bewertet werden. Wichtige Fragen: Speichert der Anbieter die eingegebenen Anfragen? Wie lange werden diese Daten aufbewahrt? Dürfen die Daten zum Training zukünftiger Modelle verwendet werden?

4

Wenn KI-Systeme in Dienstleistungen für Nutzer*innen eingesetzt werden, muss dies **klar und verständlich kommuniziert** werden. Nutzerinnen sollten die Möglichkeit haben, KI-gestützte Angebote abzulehnen und stattdessen gleichwertige menschliche Unterstützung in Anspruch zu nehmen.

5

Jede*r Mitarbeiter*in, muss verstehen, welche Arten von Daten bei der Nutzung von KI im Arbeitsalltag eingegeben werden dürfen und welche nicht. Bibliotheken sollten hierfür Listen zulässiger und unzulässiger Datentypen erstellen und regelmäßig kommunizieren.



Umweltauswirkungen

KI hat einen ökologischen Fußabdruck. Bibliotheken mit Nachhaltigkeitsanspruch müssen ihn berücksichtigen.

Die Energiekosten von „Intelligenz“ sind nicht unsichtbar. Sie befinden sich lediglich in Rechenzentren.

Umweltauswirkungen

Die Größenordnung ist entscheidend: Moderne KI-Systeme haben einen außergewöhnlich hohen Ressourcenbedarf.

626,000

kg CO₂-Äquivalent

Geschätzte Emissionen für das Training eines großen Sprachmodells.

10×

mehr Energieverbrauch

Eine ChatGPT-Anfrage benötigt etwa zehnmal mehr Energie als eine Google-Suche.

6.4 m

Kubikmeter Wasser

Geschätzter Verbrauch von Microsofts KI-Rechenzentren für Kühlsysteme im Jahr 2022.

2–4%

Globaler CO₂-Emissionen

Anteil des Tech-Sektors, der mit wachsender KI-Nutzung weiter steigt.



KI-Souveränität und Macht

Wer entwickelt KI? Wem gehört sie? Welche Werte spiegeln sich in ihr wider?

Kontrolle über KI-Infrastrukturen bedeutet Kontrolle über die Zukunft des Wissens.

Das Oligopol-Problem der KI

Der Markt für leistungsfähige Basismodelle wird derzeit von wenigen Unternehmen dominiert. Deren Entscheidungen über Trainingsdaten, Zugang, Preise, Moderation und Verfügbarkeit prägen, was Milliarden Menschen wissen oder nicht wissen können.

Konzentration von Rechenleistung	Training großer Modelle erfordert enorme Infrastruktur, die von wenigen Cloud-Anbietern kontrolliert wird (AWS, Azure, GCP).
Konzentration von Daten	Die meisten Trainingsdaten stammen überwiegend aus wohlhabenden englischsprachigen Ländern.
Konzentration von Arbeit	Datenannotation und Inhaltsmoderation werden häufig von gering bezahlten Arbeitskräften im „Globalen Süden“ durchgeführt.
Konzentration von Werten	Kleine Teams innerhalb weniger Unternehmen in einer Handvoll Ländern bestimmen, welche Inhalte KI erzeugen, zeigen oder empfehlen darf.
Regulatorisches Ungleichgewicht	Die Regulierung entwickelt sich insbesondere in der EU schnell, während in vielen anderen Regionen nur begrenzte Aufsicht existiert.

KI-Souveränität: Bedeutung für Bibliotheken

Sammlungssouveränität: Wenn KI genutzt wird, um Bibliotheksbestände zu kuratieren, zu filtern oder zu priorisieren, werden Entscheidungen über Sichtbarkeit teilweise von Bibliothekar*innen auf kommerziell entwickelte Algorithmen übertragen.

→ *Menschliche Kontrolle über KI-beeinflusste Sichtbarkeit wahren. Rankings regelmäßig überprüfen.*

Suchsouveränität: KI verändert, was Nutzer*innen finden, ohne zu verändern, was existiert. Nutzer*innen, die nur das sehen, was eine KI präsentiert, verfügen über ein eingeschränktes Informationsumfeld – geprägt durch das Training der KI und nicht durch tatsächliche Bedürfnisse.

→ *Alternative Recherchemöglichkeiten ohne KI anbieten. Kontrollierte Vokabulare vermitteln.*

Wissenssouveränität: KI-Systeme, die auf Daten aus dem „Globalen Norden“ basieren, können das Wissen einer bestimmten kulturellen Tradition als universell darstellen. Bibliotheken, die vielfältige Gemeinschaften bedienen, müssen kritisch hinterfragen, was KI weiß und was sie ausblendet.

→ *Unterrepräsentierte Wissensformen sichtbar machen. Vereinheitlichung durch KI kritisch hinterfragen.*

Rechtliche Souveränität: Daten, die von KI-Anbietern verarbeitet werden, können ausländischen Überwachungsgesetzen unterliegen. Bibliotheken müssen sicherstellen, dass die grenzüberschreitende Nutzung von KI den nationalen Datenschutzbestimmungen entspricht.

→ *Rechtliche Anforderungen an Datenstandorte vor der Beschaffung prüfen.*

Group Gruppenaktivität: Das ethische Audit: The Ethical Audit

Gruppen von 4–6 Personen | 25 Minuten | Wählen Sie ein KI-System, das Ihre Bibliothek bereits nutzt.

Ethik

Gibt es ein veröffentlichtes Ethik-Konzept? Wer hat es verfasst? Welche Risiken und Schadenspotenziale sind dokumentiert?

Bias

Wurde das System unabhängig auf Verzerrungen untersucht? Welche Sprachen oder Gemeinschaften sind in den Trainingsdaten unterrepräsentiert?

Datenschutz

Welche Daten werden erhoben, gespeichert und wie lange? Ist das System DSGVO-konform?

Umwelt

Veröffentlicht der Anbieter Daten zu Energieverbrauch oder Emissionen? Werden erneuerbare Energien genutzt?

Souveränität

In welchen Ländern werden die Daten verarbeitet? Entspricht die Inhaltsrichtlinie den Werten und Rechten unserer Gemeinschaft?