

Kompetenzzentrum
Öffentliche IT

Frühjahrstagung "KI in der Evaluation" der DeGEval AK Verwaltung & Professionalisierung am 21. Mai 2026 in Wien

KI und Evaluation: Lessons learned aus der öffentlichen Verwaltung

Dorian Wachsmann



Gefördert durch:



Bundesministerium für Digitales und Staatsmodernisierung



Fraunhofer FOKUS

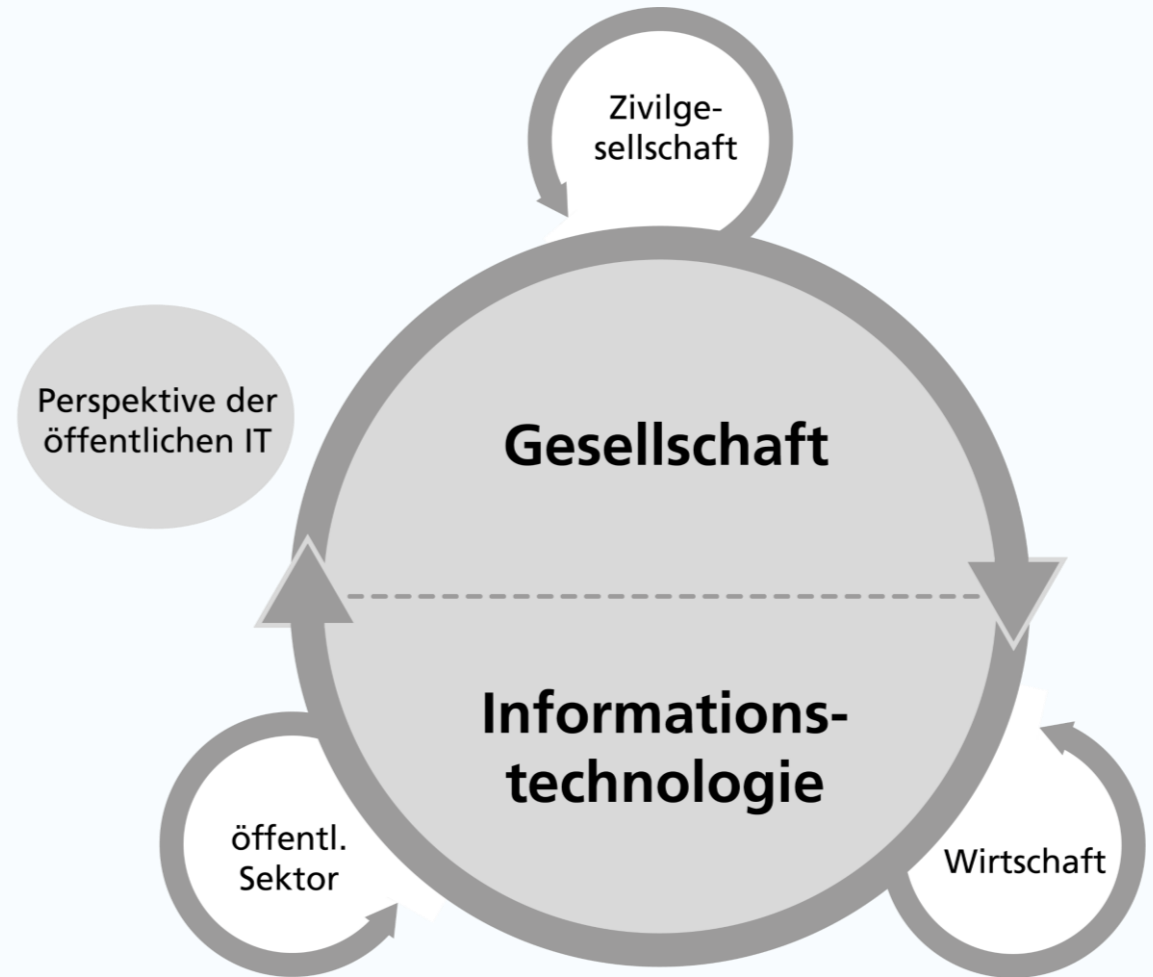
aufgrund eines Beschlusses des Deutschen Bundestages

Kompetenzzentrum Öffentliche IT (ÖFIT)

Die Perspektive der Öffentlichen IT verbindet staatliche, gesellschaftliche und technische Aspekte miteinander.

Als interdisziplinäres Team unterstützen wir die Gestaltung des digitalen Wandels von Staat und Gesellschaft.

- Denkfabrik für den digitalen Wandel
- Analyse von Digitalisierungstrends
- Anbieter- und technologieneutrale wissenschaftliche Evaluation
- Wissenschaftliche Expertise zur Gestaltung der Digitalisierung für Politik und Verwaltung



Outline

1. Stand der Lage

Wie sieht die Realität in der öffentlichen Verwaltung aus? Wo und in welchen Kontexten kommt KI tatsächlich heute schon zum Einsatz, wo werden aktuell Anwendungen für was entwickelt

2. Evaluation und GenKI am Beispiel einer Studie zu digitaler Souveränität

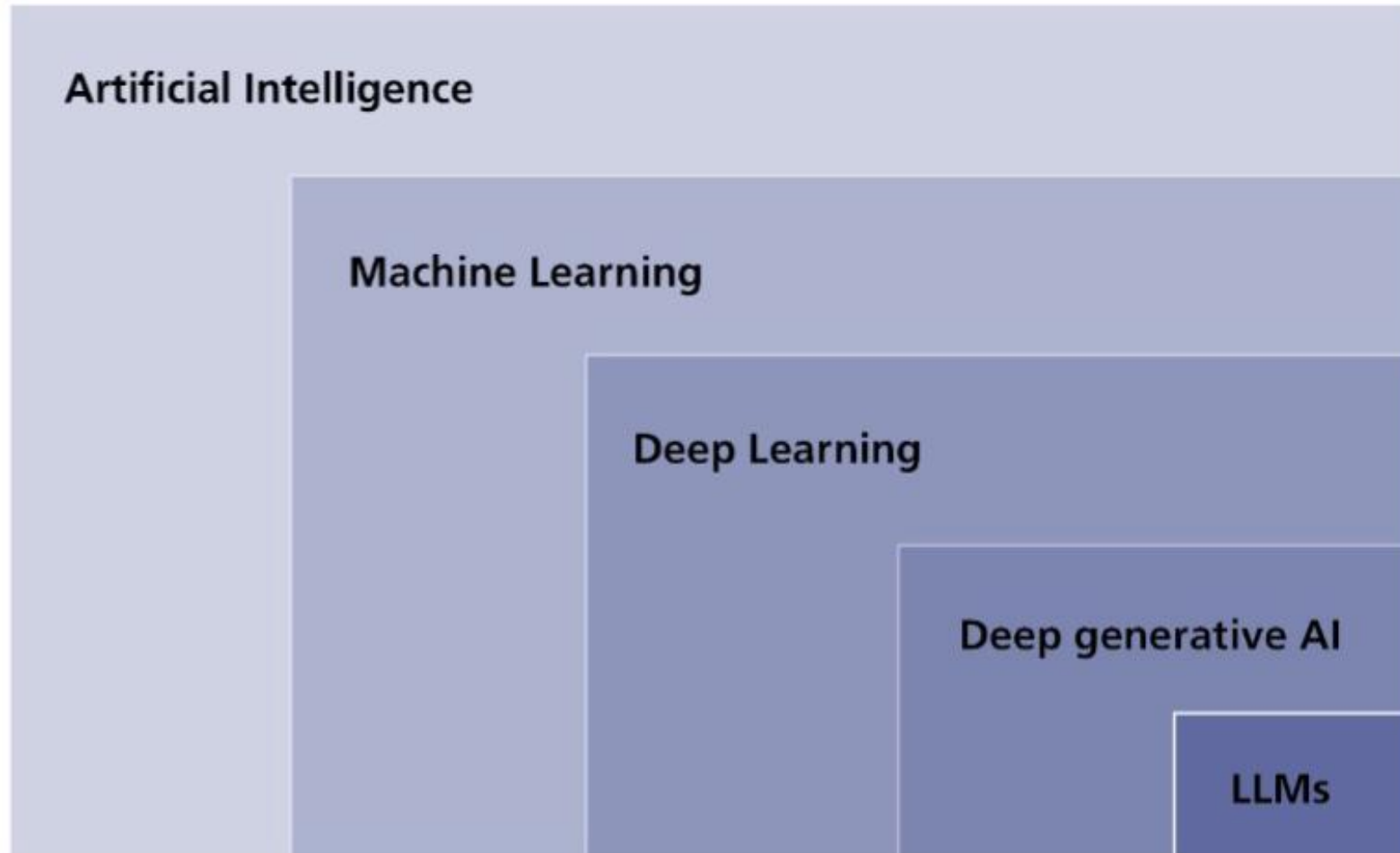
- a) Evaluation von GenKI
- b) Evaluation mit GenKI

3. Lessons learned

Stolpersteine, Herausforderungen, Evaluation von GenKI

Kurze Einordnung: „Künstliche Intelligenz“

- **Künstliche Intelligenz** als Oberbegriff vereint unterschiedlichste Ansätze, intelligentes Verhalten nachzubilden
- **Maschinelles Lernen** als Ansatz, auf Basis von Daten zu „lernen“
- **Deep Learning** ist eine Methode, die auf tiefen Neuronalen Netzen basiert
- **Generative KI** ist in der Lage, „neue“ Datenpunkte zu generieren
- **Große Sprachmodelle** (LLMs)

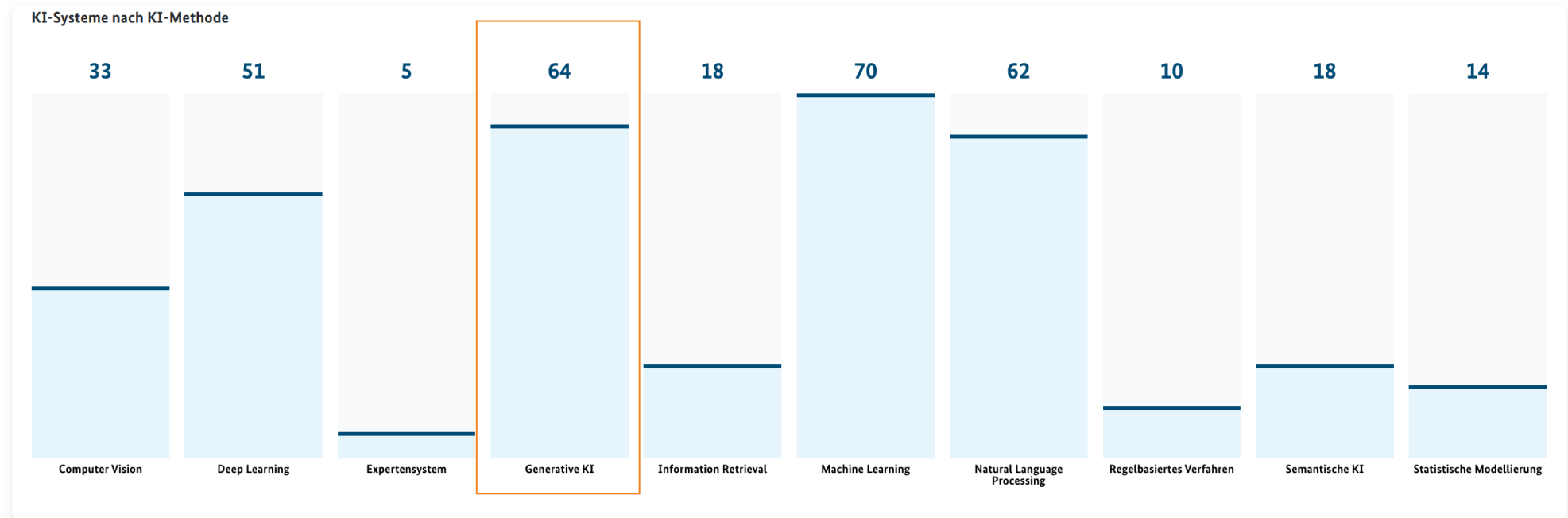


Blick in die Praxis der öffentlichen Verwaltung: Was gibt es, was wird entwickelt?

Marktplatz der KI-Möglichkeiten: (Gen)KI-Systeme Bund

- **KIPITZ** - KI-Plattform des ITZBund
- **SPARK** - KI-Module zur Unterstützung und Beschleunigung von Planungs- und Genehmigungsverfahren (in Umsetzung)
- ...

Übersicht über KI-Systeme der Bundesverwaltung



<https://maki.beki.bund.de/a/bmi-makimo-app?kiosk>

Vorstudie »Nachnutzung bestehender KI-Systeme«

GenKI-Systeme Länder und Kommunen

- **LLMoin** - KI-System der Hansestadt Hamburg
- **NRW.Genius** - KI-Plattform des Landes NRW
- **F13** - Open Source KI-System von BW
- **AIGuide** - KI-System des Landes Hessen
- **MUCGPT** - KI-System der Stadt München
- **BärGPT** – KI-System Berlin
-



Agentic AI Hub

Verbindung von **Start-ups und Kommunen**

Schnelle Umsetzung (Pilotierung in drei Monaten) und spätere Skalierung

Auswahl von 18 Pilotprojekten in 17 Kommunen mit 10 Startups

KI-Funktionen:

Verarbeitung unstrukturierter Daten, Steigerung der Datenqualität, Antragsprüfung (Vollständigkeit und Korrektheit), Entscheidungsvorbereitung, Antragsunterstützung, Postfachbearbeitung, interne und externe Dokumentation, ...

Weiter Austauschformate und Ausschreibungsrunden geplant



© 09.03.2026, Pressemitteilung

Agentic AI Hub: Pilotierung in Kommunen startet

18 Pilotprojekte wurden ausgewählt, KI-Agenten in die Verwaltung zu bringen / Wildberger: „KI nicht nur diskutieren, sondern machen“

Pressemitteilung 12/2026

Evaluation und GenKI:

Am Beispiel einer Studie zu digitaler Souveränität

Digitale Souveränität im Bereich großer Sprachmodelle (LLMs)

- Studie initiiert mit BMI (Referat DG II 2 Digitale Souveränität) mit der Zielsetzung:
 1. **Analyse**, inwieweit die Bundesverwaltung ihre digitale Souveränität im Bereich großer Sprachmodelle entweder:
 - wiederherstellen muss, oder
 - weiterhin beibehalten kann.
 2. Erarbeitung von **Handlungsempfehlungen** zur Sicherung oder Wiedererlangung dieser Souveränität



Veröffentlicht unter:
<https://www.oeffentliche-it.de/publikationen/digitale-souveraenitaet-und-grosse-sprachmodelle-in-der-bundesverwaltung/>

Idealtypischer Evaluierungsprozess

- (1) Vorbereitung/Konzeption
- (2) Durchführung (Datenerhebung und Datenanalyse)
- (3) Berichterstellung und Berichtslegung
- (4) Umsetzung von Empfehlungen



1. Vorbereitung und Konzeption

Souveränitäts-Indikatoren im Kontext von LLMs

Wechselmöglichkeit	Gestaltungsfähigkeit	Einfluss auf Anbieter
Modularität	Softwaredokumentation	Geschlossene Ökosysteme
Nachnutzbarkeit	Technische Kompetenzen	Rechtliche Kompetenzen
Open Source	Austausch und gegenseitiges Lernen	Zertifizierungen
Anzahl alternativer Modelle	Informationsbereitstellung	Verhandlungsmacht
Flexible Modellwahl	Open Source: Modell	Sitz des Anbieters
Cloud vs. On-Premise	Zusammenarbeitsstrukturen	Standort / Anbieter der Cloudserver
Dateiformate	Datenquellen & Verfügbarkeit	
	Datenanpassbarkeit	

Tabelle 3: Indikatoren digitale Souveränität

1. Vorbereitung/Konzeption

- **Literature Review & Synthese:** 

Automatisches Zusammenfassen von Vorstudien und gesetzlichen Grundlagen

- **Design von Instrumenten:** 

Entwurf von Interviewleitfäden oder Fragebögen basierend auf den Evaluationsfragen

2. Durchführung

Datenquelle für die Evaluation

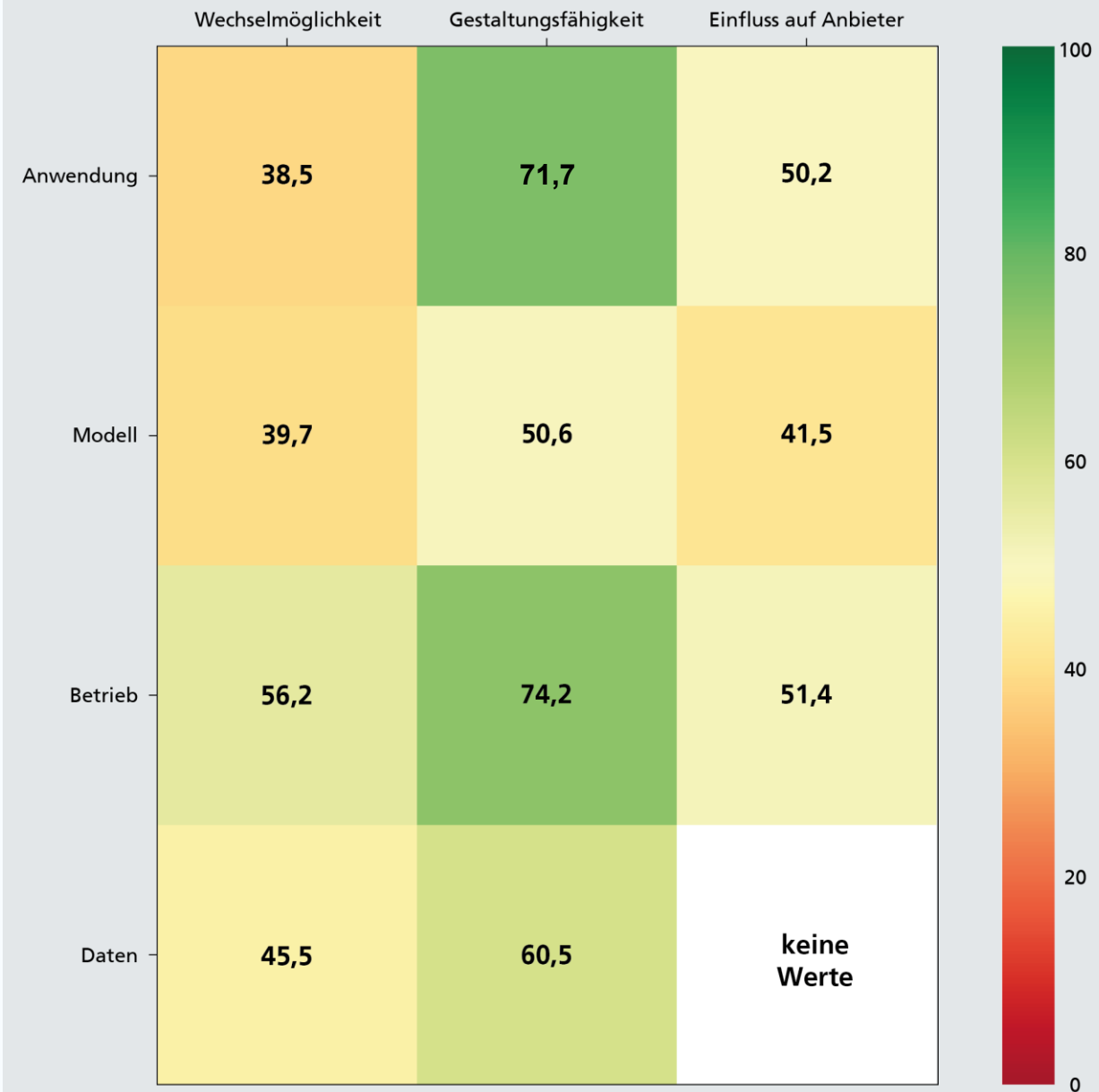
Marktplatz der KI-Möglichkeiten:

- Erhebungsbogen zur Erfassung der Informationen für Indikatoren
- 85 angefragt, 60 Rückmeldung, **33** relevant und ausgefüllt
- Vertiefende Interviews mit 5 Projekten



2. Durchführung Indikatormatrix

- Insgesamt positiv, besonders *Gestaltungsfähigkeit*
- Verbesserungspotential beim Ziel *Wechselmöglichkeit*
- Begrenzte Aussagekraft bei *Einfluss auf Anbieter*



2. Durchführung (Datenerhebung und Datenanalyse)

- **Transkription & Kodierung:** 

Automatisches Verschlagworten von Interviewtranskripten (Qualitative Inhaltsanalyse)

- **Anonymisierung:**

Automatisches Entfernen von personenbezogenen Daten aus Erhebungsprotokollen

- **Datenstrukturierung:** 

Umwandlung von unstrukturierten Texten (z.B. Freitextantworten) in strukturierte Tabellen

3. Berichterstellung und Berichtslegung

Veröffentlichung der Studie

Schritte:

- Executive Summary
- Rechtschreibprüfung
- Barrierefreiheit des gesamten Dokuments

Möglichkeiten:

- Übersetzung
- Leichte Sprache



3. Berichterstellung und Berichtslegung

- **Drafting & Refinement:** 

Erstellen von Erstentwürfen für Kapitel oder das „Executive Summary“

- **Zielgruppenadaption:**

Umformulierung komplexer Befunde in Leichte Sprache oder Kurzformate für das Management

- **Übersetzung:**

Schnelle Bereitstellung von Berichten in mehreren Amtssprachen

4. Umsetzung von Empfehlungen

- **Maßnahmen-Tracking:**

Automatisierte Analyse von Fortschrittsberichten im Vergleich zu den ursprünglichen Empfehlungen

- **Wissensmanagement:**

Aufbau eines Chatbots, der auf dem Evaluationsbericht basiert, damit Stakeholder Fragen zu den Ergebnissen stellen können („Wie wurde Empfehlung X begründet?“)

Lessons learned?

Was bleibt zu bedenken, was
macht die Zukunft?

Stolpersteine

- Sicherstellung der **Datenqualität und -verfügbarkeit** für KI-Systeme
- **Transparenz und Erklärbarkeit** von KI-Entscheidungen (Black Box-Problem)
- Risiko der Verstärkung von **Bias und Diskriminierung** durch Algorithmen
- **Datenschutz und Datensicherheit** bei sensiblen Verwaltungsdaten
- KI ist **kein Allheilmittel**

Best guess: Veränderung der Rolle von Evaluatoren

- Weniger manuelle Datenarbeit, Fokus auf **konzeptionelle Arbeit**:
 - Entwicklung und Anpassung von **Bewertungsmodellen**
- Verstärkter Bedarf an **interdisziplinären Kompetenzen (IT, Ethik, ...)**:
 - Kompetenzanforderungen für einen **kritischen Umgang** mit KI-generiertem Output
 - KI als ein neues Werkzeug: Stärken und Schwächen kennen und korrekt einsetzen

Take-Aways

1. In der öffentlichen Verwaltung „sprießen“ aktuell überall neue GenKI-Systeme
2. Der tatsächliche Nutzen, die Verhältnismäßigkeit dieser Systeme ist abzuwarten / zu evaluieren
3. GenKI-Systeme können, wenn richtig angewendet, Evaluationsprozesse unterstützen



Kompetenzzentrum
Öffentliche IT

Bleiben wir in Kontakt

Kompetenzzentrum Öffentliche IT (ÖFIT)

Tel. +49 30 3463-7173

info@oeffentlich-it.de

www.oeffentliche-it.de

[@OeffentlicheIT](https://twitter.com/OeffentlicheIT)