



Bundesamt
für die Sicherheit
der nuklearen Entsorgung

Quo Vadis Künstliche Intelligenz in der nuklearen Entsorgung?

Impressum

Bundesamt
für die Sicherheit
der nuklearen Entsorgung
(BASE)

Wegelystraße 8
10623 Berlin

Telefon: 030 184321 0
E-Mail: info@base.bund.de
www.base.bund.de

Bleiben Sie über das BASE informiert:

✉ www.base.bund.de/newsletter
🐦 www.twitter.com/BASE_bund
📷 www.instagram.com/_base_bund/

Gestaltung: quermedia GmbH, Kassel
Abbildungen: BASE und genannte Quellen,
Titelbild: iStock.com/OlgaSalt
Druck: Bonifatius GmbH, Paderborn

urn:nbn:de:0221-2024021441650

Diese Publikation wurde auf Papier gedruckt,
welches die Kriterien des Blauen Engels einhält.

Stand: Mai 2023

Quo Vadis Künstliche Intelligenz in der nuklearen Entsorgung?

**Eine Sammlung transdisziplinärer
Perspektiven herausgegeben vom
Bundesamt für die Sicherheit
der nuklearen Entsorgung**

- 4 Einführung
Ina Richter und Annika Froese
Bundesamt für die Sicherheit der nuklearen Entsorgung (BASE)

- 8 Potentiale von KI im Umgang mit Geo-Daten
Judith Krohn
Öko-Institut e.V. Darmstadt

- 10 Anmerkungen zu Voraussetzungen einer erfolgreichen Adaption
von KI-Modellen für die nukleare Entsorgung
Manfred Krafczyk
Institut für rechnergestützte Modellierung im Bauingenieurwesen,
TU Braunschweig

- 12 Digitaler Zwilling für ein geologisches Tiefenlager:
KI-Methoden für reaktiven Transport
Vinzenz Brendler
Institut für Ressourcenökologie,
Helmholtz-Zentrum Dresden-Rossendorf e.V.

- 14 Unterstützung politischer Entscheidungen durch
KI-gestützte Auswertung von Bürger:innenbeteiligungsverfahren
Julia Romberg
Heinrich-Heine-Universität Düsseldorf

- 18 Zwischen Effizienz und Kontrollverlust:
Der Einsatz von Künstlicher Intelligenz durch die Verwaltung
im Spiegel des Rechts
Stephanie Schiedermair und Johannes Weil
Juristenfakultät der Universität Leipzig,
ScaDS.AI Dresden/Leipzig

- 22 Quo Vadis Künstliche Intelligenz in der nuklearen Entsorgung?
Eine Schlussbetrachtung
Annika Froese und Ina Richter
Bundesamt für die Sicherheit der nuklearen Entsorgung (BASE)

Die Publikation entstand im Kontext eines Workshops
am BASE. Sie gibt die Auffassung und Meinung der
Autor:innen wieder und muss nicht mit der Meinung des
Herausgebers übereinstimmen.

Einführung

**Ina Richter und Dr. Annika Froese
Bundesamt für die Sicherheit der
nuklearen Entsorgung (BASE)**

Am Thema Künstliche Intelligenz (KI) kommt heutzutage kaum noch jemand vorbei. Man kann von einem regelrechten Hype sprechen; und das nicht erst, seit OpenAI im Frühling 2023 die neueste Version des Chatbots ChatGPT veröffentlichte. Es gibt kaum noch Bereiche, in denen wir nicht mit KI in Berührung kommen. Autonomes Fahren verspricht, den Menschen hinterm Steuer zu unterstützen oder gar zu ersetzen. KI-Anwendungen können Texte übersetzen sowie Bilder erkennen. KI soll Krankheiten diagnostizieren, Kreditbetrug entdecken und Gebäude energieeffizienter machen. Angesichts dieser vielfältigen Einsatzmöglichkeiten stellt sich die Frage: Was kann KI im Bereich der nuklearen Entsorgung beitragen? Diesem Thema widmet sich die vorliegende Publikation.

Was genau aber ist KI, wenn es doch so viele verschiedene Gesichter hat? Bei dem Begriff KI denken manche an futuristische Szenarien, in denen intelligente Maschinen eine alternative Lebensform verkörpern und vom Menschen nicht mehr maßgeblich zu unterscheiden sind. Solche Maschinen können autonom handeln, besitzen wohlmöglich Empfindungsvermögen und Bewusstsein. Diese so genannte ‚starke KI‘ ist derzeit allerdings nur ein theoretisches Konzept, das viele ungelöste Fragen aufwirft, und dessen Realisierbarkeit umstritten ist. Die obengenannten Anwendungsbeispiele gehören zur ‚schwachen KI‘.¹ Hier übernimmt KI eine bestimmte Aufgabe, wie z. B. die Textverarbeitung oder das Fahren eines Autos, für die bisher Menschen verantwortlich waren. Innerhalb der schwachen KI gibt es wiederum eine Spannbreite an Methoden und Ansätzen, bei denen maschinelles Lernen das aktuell dominante Paradigma ist.

Die Anzahl und Vielfalt an Definitionen des KI-Begriffs sind allerdings so groß, dass man sie auch anhand anderer Merkmale kategorisieren kann. So unterscheidet ein gängiger Ansatz² verschiedene Definitionen anhand zweier Dimensionen. Zunächst kann man fragen, ob die Definition sich auf Denkprozesse oder das Verhalten bezieht. Zudem kann man Definitionen anhand der Erfolgskriterien, an der KI gemessen wird, unterscheiden. Manche Definitionen bewerten KI als erfolgreich, wenn das Ergebnis der KI-Anwendung die entsprechende menschliche Leistung wiedergibt, während andere Definitionen voraussetzen, dass erfolgreiche KI-Anwendungen eine zu bestimmende ideale Leistung erbringen, die der menschlichen oftmals überlegen ist. Soll eine KI-Anwendung beispielsweise bei der bildgebenden Diagnostik so viel leisten, wie der/die durchschnittliche Radiolog:in, oder soll die KI-Anwendung die menschliche Leistung übertreffen?

Aber auch dieser Ansatz erschöpft nicht alle Möglichkeiten der Kategorisierung. Abgesehen von der Bewertung des Endresultats unterscheiden sich nämlich einige Definitionen auch hinsichtlich der Rolle, die sie dem Prozess zuschreiben, mit dem eine KI-Anwendung ihre Ergebnisse produziert. Manche Definitionen beinhalten die Erwartung, dass KI menschliche Denkmuster nachbildet, bei anderen Definitionen spielt die Vorgehensweise, mit der eine KI-Anwendung ihre Ergebnisse erzielt, keine Rolle. Soll zum Beispiel eine KI-Anwendung Bilder anhand derselben Kriterien kategorisieren, die Menschen bei der Bilderkennung heranziehen, oder ist es nebensächlich, wie die KI-Anwendung zu ihrem Ergebnis kommt? Schon dieser kurze Überblick zeigt, dass es schwer ist, sich auf ein gemeinsames Verständnis des KI-Begriffs zu einigen.

Der KI-Begriff umfasst außerdem eine Vielzahl von Methoden. Ein veranschaulichendes Beispiel ist das sogenannte maschinelle Lernen, bei dem ein System zunächst eine Lernphase durchläuft, um im Anschluss neues Wissen zu generieren.³ In der Lernphase wird das System mit Trainingsdaten versorgt, in denen es Muster bzw. Gesetzmäßigkeiten feststellt. Wenn es dann mit neuen Daten konfrontiert wird, kann es diese Daten anhand der erlernten Muster beurteilen. So kann man zum Beispiel einem Videoüberwachungs-System beibringen, ‚verdächtiges‘ Verhalten zu erkennen, indem man es zuvor mit vom Menschen als verdächtig eingestuftes Bildmaterial trainiert hat.⁴

Das zuletzt genannt Beispiel der Videoüberwachung wirft direkt wichtige und ebenso komplexe Fragen auf, die KI-Anwendungen im Allgemeinen betreffen: Welche Annahmen und Werte liegen der Auswahl der Daten sowie der Zielvorgabe des KI-Systems zugrunde? Inwiefern beeinflussen die Vorurteile der Entwickler:innen eine KI-Anwendung? Sind die verfügbaren Daten quantitativ und qualitativ ausreichend, um eine KI-Anwendung zu ermöglichen? Wie wahrscheinlich ist es, dass der KI-Methode dennoch ein Fehler unterläuft? Was sind die zu erwartenden Folgen eines solchen Fehlers? Wer haftet, wenn ein Schaden verursacht wird? Ist der Einsatz von KI ethisch vertretbar? Gibt es Aufgaben, die grundsätzlich nicht aus menschlicher Hand an eine KI-Anwendung abgegeben werden sollten? Solche Fragen verdeutlichen, dass es notwendig ist, sich mit den Entwicklungen von KI zu befassen und nach den Chancen und Risiken ihres Einsatzes zu fragen.

Mit dem Thema KI beschäftigt sich auch das Bundesamt für die Sicherheit der nuklearen Entsorgung (BASE). Als zentrale Fachbehörde des Bundes für den sicheren Umgang mit den Hinterlassenschaften der Atomenergie nimmt es Regulierungs-, Genehmigungs- und Aufsichtsaufgaben im Bereich Endlagerung, Zwischenlagerung sowie für den Umgang und Transport von hochradioaktiven Abfällen wahr. Um diese Aufgaben bestmöglich zu erfüllen, gilt es sich kontinuierlich mit Entwicklungen in Wissenschaft und Technik zu befassen. Mit Blick auf KI zeigt sich, dass deren Potentiale und Risiken für die nukleare Entsorgung von radioaktiven Abfällen bisher wenig adressiert sind.

Ein Konsortium aus Wissenschaftler:innen verschiedener Disziplinen hat im November 2021 eine erste wissenschaftliche Bestandsaufnahme veröffentlicht.⁵ Diese ‚Roadmap zur Entwicklung und Adaption von Methoden der Künstlichen Intelligenz (KI) für die Endlagerforschung‘ gibt erstmals einen Überblick über KI-Anwendungen, die für die Endlagerung als potentiell nutzbar erachtet werden. Die Autor:innen sehen vor allem bei der Visualisierung von Arbeitsabläufen, beim Umgang mit Ungewissheiten und selbstlernenden Systemen sowie beim Wissensmanagement mögliche Anwendungsfelder für KI. Jenseits dessen gibt es aktuell nur wenige Forschungsk Kooperationen und Forschungsprogramme, die KI im Kontext der nuklearen Entsorgung verfolgen.⁶

Das BASE greift die Entwicklungen im Feld KI und Endlagerung auf und veranstaltete zum Thema am 1. Dezember 2022 ein transdisziplinäres Panel ‚Quo Vadis Künstliche Intelligenz in der nuklearen Entsorgung?‘. Im Rahmen der Fachdiskussion ging es um die Frage, welchen Beitrag KI für die nukleare Entsorgung leisten kann. Expert:innen aus verschiedenen wissenschaftlichen Disziplinen und Bereichen der Gesellschaft kamen zusammen, um über den Stand der Entwicklungen von KI für die nukleare Entsorgung zu berichten. Darüber hinaus wurden Erfahrungen und Erkenntnisse aus anderen Anwendungsfeldern vorgestellt, die für die Endlagerung relevant sind, und über rechtliche und ethische Anforderungen reflektiert, die es beim Einsatz von KI zu berücksichtigen gilt. Mit der Veranstaltung brachte das BASE wissenschaftliche Expertise aus der Politikwissenschaft, Geographie, Informatik, der rechnergestützten Modellierung und Rechtswissenschaft mit praxisnahen Erfahrungen wie etwa aus dem Umweltbundesamt und der Bundesgesellschaft für Endlagerung mbH (BGE mbH) zusammen, um theoretische Potentiale mit den Bedingungen der Praxis zu konfrontieren.

Die vorliegende Publikation gibt einen Überblick über die Inhalte und Positionen, die einige der Expert:innen beim Panel zur Diskussion stellten, und fasst die Diskussionen im Rahmen eines abschließenden Fazits kurz und prägnant zusammen. Ziel ist es, einen Einblick in die vorgestellten Inhalte und Diskussionspunkte zu geben und für die Öffentlichkeit zugänglich zu machen.

Die ersten drei Beiträge adressieren den Ist-Stand der KI-Entwicklung im Feld der Endlagerung. Judith Krohn gibt in ihrem Beitrag ‚Potentiale von KI im Umgang mit Geo-Daten‘ einen Überblick, in welchen Themenfeldern Methoden der KI im Umgang mit Geo-Daten bereits erforscht sind. Potentiale für die Endlagersuche, so die Autorin, liegen beim KI-gestützten Datenmanagement und Auswerten von Bildern und Datenräumen. Da Daten jedoch vielfach Unsicherheiten in sich bergen und etwa einem Bias unterliegen können, sind auf diesen Daten basierende KI-Anwendungen gerade in der Endlagersuche zur Unterstützung von Entscheidungen, als Kontrollinstanz und ergänzend zu klassischen Verfahren anzuwenden. Die Bewertung durch Menschen werde hier nicht durch KI ersetzt, sondern unterstützt. Manfred Krafczyk erläutert zunächst, wie KI-Methoden die Entwicklung und Anwendung von Rechenmodellen und Simulationen unterstützen und damit potentiell auch im Bereich der nuklearen Entsorgung von Nutzen sein können. In seinem Beitrag ‚Anmerkungen zu Voraussetzungen einer erfolgreichen Adaption von KI-Modellen für die nukleare Entsorgung‘ benennt er aber auch die Herausforderungen, die mit dem Einsatz von KI-Methoden

einhergehen, bedingt u. a. durch technische Aspekte und der Menge und Qualität der zur Verfügung stehenden Trainingsdaten. Vinzenz Brendler gibt einen Einblick, welchen Mehrwert eine konkrete KI-Methode – der digitale Zwilling oder englisch auch digital twin (DT) – für die nukleare Entsorgung birgt. In seinem Beitrag ‚Digitaler Zwilling für ein geologisches Tiefenlager: KI-Methoden für reaktiven Transport‘ gibt er einen Einblick in ein frühes von der Europäischen Union gefördertes Projekt, das interaktive Visualisierung erforscht.

Der vierte Beitrag fokussiert auf Erfahrungen mit KI in einem Bereich, von dem für die nukleare Entsorgung gelernt werden kann. Julia Romberg legt in ‚Unterstützung politischer Entscheidungen durch KI-gestützte Auswertung von Bürger:innenbeteiligungsverfahren‘ dar, wie Methoden der KI Verfahren der Bürgerbeteiligung unterstützen können. Potentiale sieht die Autorin in der KI-gestützten Auswertung von Texten, also etwa den Eingaben von Bürger:innen in Beteiligungsprozessen.

Der fünfte Beitrag greift Rahmenbedingungen auf, die beim Einsatz von KI zu berücksichtigen sind. Stephanie Schiedermaier und Johannes Weil reflektieren in ‚Zwischen Effizienz und Kontrollverlust: Der Einsatz von Künstlicher Intelligenz durch die Verwaltung im Spiegel des Rechts‘ über Spannungsverhältnisse, die sich beim Einsatz von KI in öffentlichen Verwaltungen aus rechtlicher Sicht ergeben. Aus Sicht der Autor:innen besteht aktuell weiterer Klärungsbedarf, bevor die KI-Technologien breite Anwendung in staatlichen Einrichtungen finden.

Die Publikation schließt mit einer Schlussbetrachtung aus Sicht des BASE. Diese legt dar, dass Methoden der KI durchaus Potentiale für die Entsorgung radioaktiver Abfälle bergen. Allerdings gibt es auch eine Reihe an Grenzen, insbesondere mit Blick auf das vorliegende Datenmaterial und die Verfügbarkeit von Daten. Darüber hinaus ist unklar, wie robust die Ergebnisse von KI-Methoden sind und ob KI-Anwendungen nicht eher als ‚Kontrollinstanzen‘ fungieren sollten, die Fehler und Unsicherheiten reduzieren. Zudem stellen sich für Akteure der Endlagersuche ganz eigene Fragen: Was wäre z. B., wenn die BGE mbH bei der Suche nach dem am geeignetsten Standort für ein Endlager KI tatsächlich einsetzt? Als Aufsichts- und Genehmigungsbehörde im Verfahren der Endlagersuche wäre für das BASE zu klären, welcher Regulierungsbedarf besteht.

1 Für eine zugängliche Diskussion des Unterschieds zwischen ‚starker‘ und ‚schwacher‘ KI, siehe Ramge, T. (2018): Mensch und Maschine Wie künstliche Intelligenz und Roboter unser Leben verändern, S. 18f. Reclam, Ditzingen.

2 Russell, S; Norvig, P. (2012): Künstliche Intelligenz Ein moderner Ansatz, S. 22f. Pearson, München.

3 Für Methoden, bei denen KI anhand von Beispielen lernt, siehe die Kapitel 18, 19 und 20 in Russell, S; Norvig, P. (2012): Künstliche Intelligenz. Ein moderner Ansatz, Pearson, München.

4 Siehe z. B. das Projekt „Intelligente Videoüberwachung“ des Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung in Partnerschaft mit dem Polizeipräsidium Mannheim, der Stadt Mannheim und dem Ministerium für Inneres, Digitalisierung und Migration Baden-Württemberg, bei dem der Einsatz einer algorithmenbasierten Videoüberwachung zur Bekämpfung von Straßensriminalität im öffentlichen Raum getestet wird. <https://www.iosb.fraunhofer.de/de/projekte-produkte/intelligente-videoeueberwachung.html>.

5 Krafczyk, M.; Brendler, V.; Czaikowski, O.; Gruner, M.; Hoth, N.; Kolditz, O.; Nagel, T.; Herold, P.; Müller, C.; Seher, H.; Simo, E.; Stahlmann, J. (2021): Eine Roadmap zur Entwicklung und Adaption von Methoden der Künstlichen Intelligenz (KI) für die Endlagerforschung. 2021. TU Braunschweig; Helmholtz-Zentrum Dresden-Rossendorf e.V.; Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH; TU Bergakademie Freiberg; Helmholtz Zentrum für Umweltforschung; BGE TECHNOLOGY GmbH. o. O. DOI: <https://doi.org/10.5281/ZENODO.5752277>.

6 Siehe u.a. das vom BASE beauftragte Vorhaben ‚Anwendung der künstlichen Intelligenz (KI) für die Standortauswahl von tiefen geologischen Endlagern‘ (https://www.base.bund.de/DE/themen/fa/soa/documents/Kuenstliche_Intelligenz.html); die IGD-TP3, eine Forschungskoooperation von europäischen Vorhabenträger:innen zur Unterstützung der Umsetzung der nationalen Entsorgungsprogramme zur tiefergeologischen Endlagerung radio-aktiver Abfälle (igdtp.eu); und das gemeinsame Forschungsprogramm der Europäischen Union zu Entsorgung und Management radioaktiver Abfälle (EURAD4) (ejp-eurad.eu).

Potentiale von KI im Umgang mit Geo-Daten

Judith Krohn
Öko-Institut e.V. Darmstadt

Kurzvorstellung des Vorhabens AKI – Methodisches Vorgehen

Im Vorhaben „Anwendung der künstlichen Intelligenz (KI) für die Standortauswahl von tiefen geologischen Endlagern“ (AKI)¹ wurden auf Basis einer umfassenden internationalen Literaturrecherche zunächst Einsatzbereiche künstlicher Intelligenz in den Geowissenschaften allgemein identifiziert und diese dann im Hinblick auf den potentiellen Einsatz für die geowissenschaftlichen Fragestellungen im Standortauswahlverfahren (StandAV) bewertet. Mit einem für diesen Zweck entwickelten Bewertungsschema können Qualität, Eignung und Relevanz des in der Literatur beschriebenen KI-Einsatzes zur Lösung einer geowissenschaftlichen Fragestellung von einem interdisziplinär zusammengesetzten Expertenteam eingeordnet werden.

Die Bewertung erfolgt dafür in zwei Stufen. In der ersten Stufe wird beurteilt, wie stark grundsätzliche Stärken und Schwächen der KI für das jeweilige KI-Einsatzgebiet zutreffen. In der zweiten Stufe erfolgt die Beurteilung des in der Literatur beschriebenen KI-Einsatzes konkret anhand von Bewertungsfragen mit Positiv- und Negativmustern, die eine eindeutige Aussage zu Größe, Verfügbarkeit und Güte der verwendeten Datengrundlage sowie zum technologischen Reifegrad, zur Nachvollziehbarkeit und zur speziellen Relevanz für das StandAV ermöglichen. An Fallbeispielen aus der Literatur, die aufgrund einer ersten Einschätzung relevant für das StandAV sein können, wurde dieses Bewertungsschema angewandt und die Ergebnisse kontinuierlich durch interdisziplinären Austausch auf Plausibilität geprüft. Aus einer Aggregation der Ergebnisse konnten erste Erkenntnisse zu Chancen und Risiken eines potentiellen KI-Einsatzes im StandAV gewonnen und weiterer Forschungsbedarf identifiziert werden.

Standortauswahlverfahren setzt den Rahmen für potenziellen KI-Einsatz

Der Rahmen des wissenschaftsbasierten StandAV legt fest, dass alle Entscheidungen und Schlussfolgerungen sicher, datengestützt und nachvollziehbar sein sollen. Für die bestmögliche Langzeitsicherheit des Endlagers müssen darüber hinaus mögliche geologische Entwicklungen über einen Nachweiszeitraum von einer Million Jahre analysiert werden. Derartige Prognosen zur räumlichen und zeitlichen Veränderung der geologischen Gegebenheiten bauen auf eine umfangreiche und verlässliche Datengrundlage auf. Vorhersagen erfordern komplexe Berechnungen und Modellierungen im mehrdimensionalen Raum auf unterschiedlichen Skalenebenen und unter Einbezug verschiedener geochemischer und geophysikalischer Wechselwirkungen. Eine zentrale Rolle spielt dabei auch die Beschreibung und Bewertung von Ungewissheiten. Hohe Ansprüche müssen zudem an die Nachvollziehbarkeit der angewendeten KI-Methoden gestellt werden. KI-Methoden, die dem Transparenzanspruch des StandAV nicht genügen, bergen beträchtliche Risiken, das Vertrauen der Bevölkerung in den Auswahlprozess aufs Spiel zu setzen.

KI-Einsatz in Geowissenschaften in der Literatur

In den Geowissenschaften gibt es bereits heute eine hohe und stetig steigende Anzahl an Literatur, die sich mit dem Einsatz von KI beschäftigt. Die Quellen setzen sich unter anderem mit dem Erkennen, Segmentieren, Generieren und Verarbeiten von Daten in digitalen Bildern (Computer Vision), der Klassifizierung (Classification) oder der Zuordnung von Daten (Clustering) auseinander. Darüber hinaus zeigt die Literatur Einsatzgebiete auf, in denen mit KI-Methoden Ersatzmodelle (Surrogate Modelle) entwickelt, Schätzungen (Prediction), Vorhersagen oder Prognosen (Forecast) erstellt, sowie komplexe Zusammenhänge vereinfacht (Dimension Reduction), in komplexen Lösungsräumen nach besseren Lösungen gesucht (Optimization) oder Ausreißer und ungewöhnliche Muster erkannt und gekennzeichnet (Anomaly Detection) werden.

Es gibt demnach bereits umfangreiche Forschung auf dem Gebiet KI in den Geowissenschaften, sehr selten wird sie bisher auf reale Probenkampagnen angewandt. Aber erst mit einem konkreten direkten Einsatz in der Geologie können Erfahrungen mit KI in diesem Einsatzgebiet valide und zukunftsweisend mit Fokus auf das StandAV ausgewertet werden. Dabei ist insbesondere die Eignung der KI-Anwendung für die jeweilige Problemstellung zu prüfen und das Vorgehen im Hinblick auf den zusätzlichen Nutzen von KI im Vergleich zu herkömmlichen Methoden sowie in einer Gegenüberstellung von Potenzial und Risiken zu evaluieren und abzuwägen. Anschließend muss im Sinne der Übertragbarkeit für Einsatzfelder im StandAV geprüft werden, welche Anpassungen erforderlich sind und die Risiken dieser Anpassungen zusätzlich eingeschätzt werden.

KI im Kontext Standortauswahlverfahren

Nach Auswertung der Chancen und Risiken von KI-Methoden zeigt sich, dass zu den Kernstärken der KI das Datenmanagement sowie das Auswerten von Bildern und hochdimensionalen Datenräumen gehört. Im Bereich der Geodatenverarbeitung gewinnen zunehmend die Analyse von großen Datenmengen sowie auch die Interpretation ungenauer Daten mit Hilfe von KI an Bedeutung. Ein weiterer großer Nutzen der KI-Methoden in den geowissenschaftlichen Kategorien besteht in der genaueren Abbildung von zeitaufwendigen numerischen Berechnungen oder in der Optimierung der an die Datenverarbeitung anknüpfenden geostatistischen Analyse. Viele KI-Einsatzbereiche bieten darüber hinaus die Chance, ein besseres Verständnis von Gesetzmäßigkeiten und Zusammenhängen der realen Prozesse erzielen zu können und damit die Objektivität der Bewertungen zu erhöhen. Allerdings steht letzteres im Widerspruch zu dem teilweise gleichzeitig vorhandenen Risiko, dass durch die ebenfalls häufig vorhandene Gefahr eines Daten- oder Entwickler-Bias bewusst oder unbewusst Einfluss auf die Entscheidungen von KI-Verfahren genommen werden kann. Aber die Chance, bisher nicht erfassbare Prozesse mit KI abbilden zu können, sollte in einem Vorhaben, das auf das Ziel „Endlagerstandort mit der bestmöglichen Sicherheit für mindestens eine Million Jahre“ ausgerichtet ist, zumindest näher geprüft werden.

Voraussetzungen für tragfähige Lösungen sind allerdings eine ausreichend qualitativ hochwertige Datenbasis sowie die Auswahl von KI-Anwendungen, deren Qualität und Eignung für die konkrete geowissenschaftliche Fragestellung ausführlich validiert sind. Ein großer Nachteil von KI ist deren meistens vorhandene Intransparenz. Diese birgt zum einen die Gefahr, dass Unsicherheiten bei Berechnungen über

große Zeiträume verschleiert oder sogar zu nicht einsehbaren Fehlerketten akkumuliert werden können. Zum anderen ist eine negative Voreingenommenheit der Öffentlichkeit, die häufig aus einer Intransparenz der Verfahren resultiert, gefährlich für den gesamten Auswahlprozess. Unsere Analyse zeigt, dass selbst mit neuen KI-Anwendungen der sog. „erklärbaren künstlichen Intelligenz“ (XAI) nicht in allen KI-Einsatzgebieten gewährleistet werden kann, dass die angewendeten Methoden und die erreichten Ergebnisse nachvollziehbar darzustellen sind. Zusätzlich muss darauf hingewiesen werden, dass die Methoden der XAI selbst noch am Anfang stehen.

Der Einsatz von KI-Anwendungen eignet sich deshalb im StandAV ausschließlich dazu, Entscheidungen zu unterstützen, klassische Verfahren zu ergänzen oder als Kontrollinstanz zum Aufspüren von Fehlern und zur Bewertung von Ungewissheiten zu fungieren.

1 Krohn, J. et al. (2022): Anwendung der künstlichen Intelligenz (KI) für die Standortauswahl von tiefen geologischen Endlagern (AKI). Vorhaben mit FKZ 4721E03210 für das Bundesamt für die Sicherheit der nuklearen Entsorgung (BASE). Online verfügbar unter <https://www.oeko.de/publikationen/p-details/anwendung-der-kuenstlichen-intelligenz-ki-fuer-die-standortauswahl-von-tiefen-geologischen-endlagern-aki>.

Anmerkungen zu Voraussetzungen einer erfolgreichen Adaption von KI-Modellen für die nukleare Entsorgung

Prof. Dr. Manfred Krafczyk
Institut für rechnergestützte Modellierung
im Bauingenieurwesen, TU Braunschweig

Computational Engineering hat sich in den letzten Jahrzehnten als ein interdisziplinäres Paradigma entwickelt, welches sich mit der Entwicklung und Anwendung von Rechenmodellen und Simulationen, oft unter Verwendung von Hochleistungsrechnern, beschäftigt, um komplexe Mehrfeldprobleme zu lösen, die bei der technischen Analyse und Konstruktion von Ingenieurproblemen auftreten. Die rechnergestützte Simulation bietet dabei die Möglichkeit, in Bereiche vorzudringen, die entweder für traditionelle Experimente unzugänglich sind oder in denen die Durchführung empirischer Untersuchungen zu aufwändig ist. Die Entwicklung eines konkreten Simulationsmodells oder Verfahrens beginnt dabei typischerweise mit der Systemidentifikation, bei der ein System als ein Konglomerat individueller Entitäten und ihrer Wechselwirkung gedacht wird. Ingenieursysteme im Bereich der nuklearen Entsorgung zeichnen sich mehrheitlich durch die Beschreibung von Systemen als Kontinua über sog. Felder aus, die miteinander in Raum und Zeit interagieren. Die mathematische Beschreibung erfolgt bisher meist in Form von partiellen Differenzialgleichungen, die es unter Annahme geeigneter Rand- und Anfangsbedingungen (meist approximativ) zu lösen gilt. Da neben der Anzahl der Dimensionen in Raum und Zeit oft noch weitere Parameter im Sinne einer Problemdimension variiert werden, ist man oft mit den Auswirkungen des sogenannten Fluches der Dimensionen konfrontiert, der für viele praxisrelevante Probleme zu inakzeptablen Rechenzeiten bei der Auffindung einer hinreichend genauen Näherungslösung der Modellgleichungen führt. Daher besteht weiterhin der Bedarf nach innovativen Methoden zur (näherungsweisen) Lösung von partiellen Differenzialgleichungen oder alternativer Modellierungsansätze wie sie beispielsweise durch verschiedene Varianten des sog. Maschinellen Lernens bzw. der künstlichen Intelligenz in den letzten Jahren sehr populär geworden sind.

Maschinelles Lernen als die Wissenschaft von rechnergestützten Methoden, die ihre Problemlösungskapazität durch sukzessive Verarbeitung größerer Datenmengen „selbständig“ erweitern und insbesondere sog. Neuronale Netze (NN) haben in der letzten Dekade eine eindrucksvolle Erfolgsgeschichte in Forschung und Technik geschrieben, da sie unsere Fähigkeit, aus Daten Strukturen und Wissen abzuleiten extrem erweitert haben.

Mathematisch betrachtet handelt es sich bei einem NN meist um eine Abbildung zwischen zwei Räumen (beispielsweise reeller Zahlen) von potenziell unterschiedlicher Dimension. Gemäß dem universal approximation theorem kann ein sog. Feedforward Neural Network (FNN) unter gewissen allgemeinen Annahmen asymptotisch jede beliebige Funktion mit beliebig kleinem Fehlerbetrag approximieren. Im Gegensatz zu klassischen Modellierungsansätzen müssen NN jedoch zur Bestimmung der gesuchten Abbildung mit geeigneten Trainingsdaten vor der eigentlichen Nutzung „konfiguriert“ werden. Die Leistungsfähigkeit des NN hängt daher neben vielen technischen Aspekten grundsätzlich von der Anzahl seiner Freiheitsgrade (Anzahl und Größe seiner Schichten) und der Menge und Qualität der zu seiner Konditionierung zur Verfügung stehenden Trainingsdaten (u. U. auch von dem/der Trainierenden) ab. Setzt man voraus, dass die Trainingsdaten den Teilsystemzustand eines oder mehrerer potenzieller Endlagerstandorte charakterisieren, können daraus ggf. Eigenschaften eines (anderen) Standortes mit geringem Berechnungsaufwand extrapoliert werden. Die resultierenden Unsicherheiten bzgl. der Qualität der Lösung sind jedoch von vielfältigen Faktoren abhängig und im Gegensatz zu Modellierungsansätzen basierend auf Differenzialgleichungen schwieriger separierbar, da insbesondere die Trainingsdaten potenziell implizite Informationen enthalten, deren Einfluss auf berechnete Lösungen potenziell schwer zu identifizieren sind. Auch wenn Modelle auf Basis von Differenzialgleichungen auch Unsicherheiten insbesondere in Bezug auf Anfangs-, Rand- und Werkstoffgesetze beinhalten, ist das Lösungsverhalten bei einem definierten Setup asymptotisch konvergent, was für datengetriebene Modelle nicht zwingend gilt.

Auch ist anzumerken, dass datengetriebene Modelle unter Umständen vorsätzlich „in die Irre“ geführt werden können¹, indem das Eingangssignal auf subtile Weise gestört wird, wobei diese Störung die menschliche Charakterisierung nicht notwendigerweise beeinflussen würde.

Für viele Ingenieurprobleme (bisher außerhalb der Domäne der nuklearen Entsorgung) existieren seit einigen Jahren datengetriebene Ansätze zur Lösung von Differenzialgleichungen, für welche sich der Begriff der Physics Informed Neural Networks (PINNs) etabliert hat. Die in der jüngeren Literatur vorgestellten Beispiellösungen legen nahe, dass die Lösung partieller Differenzialgleichungen mit PINNs sehr effizient sein kann, auch wenn wesentliche Aspekte wie asymptotische Konvergenz und Stabilität bisher nur heuristisch untersucht wurden. Die Frage nach der Gültigkeit solcher Ergebnisse außerhalb der Trainingsdomäne wird aktiv erforscht.² Unklar ist generell auch, wie die zu erzielende Genauigkeit einer PINN-basierten Näherungslösung von der Anzahl und Größe der verborgenen Schichten (und damit auch des assoziierten Trainings- und Berechnungsaufwandes) abhängt. Hier könnten auch klassische Methodenvarianten wie raumzeitliche Adaption basierend auf a priori Fehlerschätzern von sample-Punkten auf PINNs übertragen werden und zu einer signifikanten Effizienzsteigerung beitragen. Während bei Überschreitung von Stabilitätsgrenzen klassische numerische Verfahren einfach versagen, tritt ein solcher Effekt bei PINNs nicht ein. Dies kann bei unreflektiertem Einsatz von PINNs dazu führen, dass die erzielten Lösungen von unzureichender Qualität sind. Neben der Lösung von partiellen Differenzialgleichungen im Sinne eines Vorwärts-Problems konnten datengetriebene Ansätze jedoch schon erfolgreich zur Lösung inverser nicht-linearer Probleme adaptiert werden. Dies umfasst z.B. die Rekonstruktion physikalischer Parameter in Differenzialgleichungen aus Messdaten oder die Rekonstruktion physikalischer Felder aus komplementären Teildaten. Diese Problemklasse ist für klassische numerische Verfahren (wenn überhaupt) nur mit signifikant höherem Aufwand zugänglich. Unabhängig davon, ob PINNs mittelfristig belastbare Effizienz- und Genauigkeitsvorteile bei der Lösung von Vorwärts-Problemen zeitigen, steht schon jetzt fest, dass sie insbesondere im Bereich der inversen Probleme neue Wege im Computational Engineering zu beschreiten erlauben.

Die zukünftige Untersuchung komplexer Probleme in der nuklearen Entsorgung dürfte insbesondere von Ansätzen profitieren, bei der eine Kombination von klassischen Modellierungsansätzen, datengetriebener Modellierung und Unsicherheitsanalysen zum Einsatz kommt. Dies setzt aber eine besondere Sorgfalt bei der Erhebung, Bearbeitung und Persistierung der entsprechenden Trainingsdaten voraus, um eine Reproduzierbarkeit der Ergebnisse sicherzustellen. Weitere generelle Empfehlungen zu systematischen Forschungsansätzen hierzu sind dem Artikel „Eine Roadmap zur Entwicklung und Adaption von Methoden der Künstlichen Intelligenz (KI) für die Endlagerforschung“³ zu entnehmen.

1 Juyeon Heo, Sunghwan Joo, Taesup Moon, Fooling. (2019). Neural Network Interpretations via Adversarial Model Manipulation. <https://arxiv.org/abs/1902.02041v3>.

2 Andrea Bonfanti, Roberto Santana, Marco Ellero, Babak Gholami. (2023). On the Hyperparameters influencing a PINN's generalization beyond the training domain. <https://doi.org/10.48550/arXiv.2302.07557>.

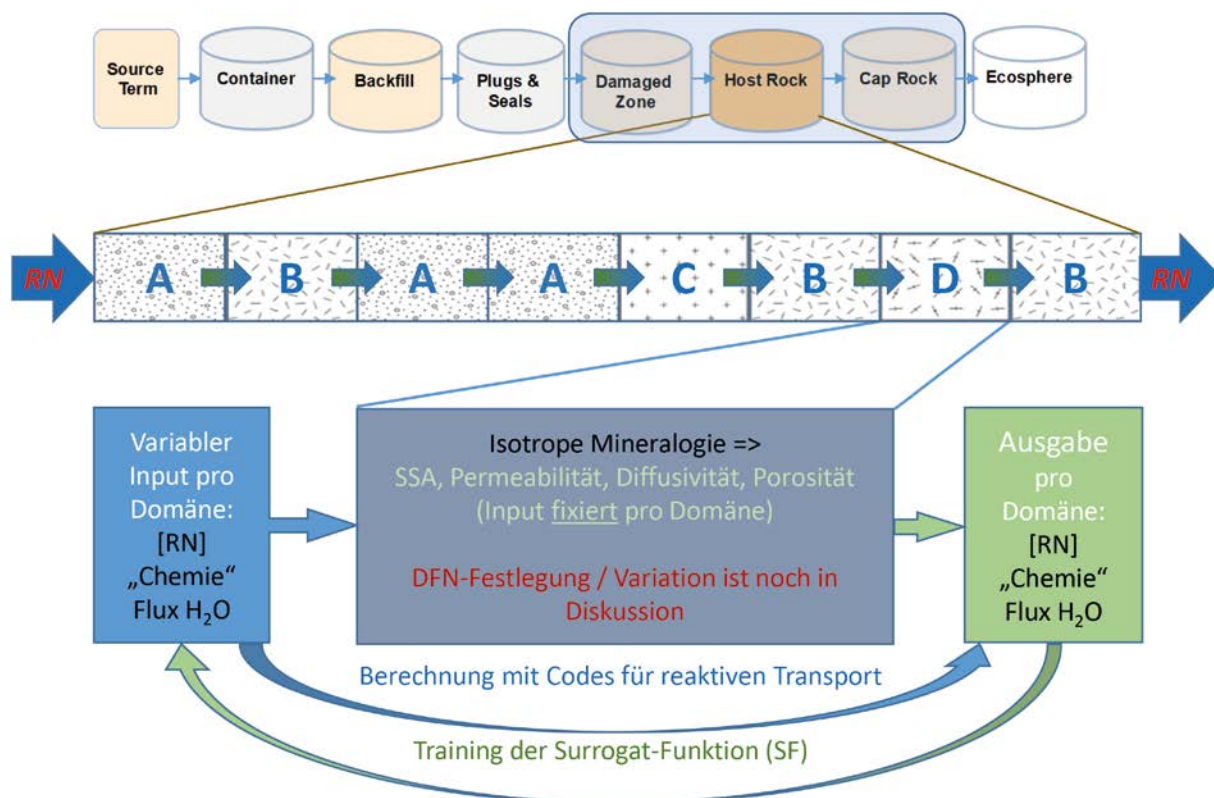
3 Krafczyk, Manfred, Brendler, Vinzenz, Czaikowski, Oliver, Gruner, Matthias, Hoth, Niels, Kolditz, Olaf, Nagel, Thomas, Herold, Philipp, Müller, Christian, Seher, Holger, Simo, Eric, & Stahlmann, Joachim. (2021). Eine Roadmap zur Entwicklung und Adaption von Methoden der Künstlichen Intelligenz (KI) für die Endlagerforschung. Zenodo. <https://doi.org/10.5281/zenodo.5752277>.

Digitaler Zwilling für ein geologisches Tiefenlager: KI-Methoden für reaktiven Transport

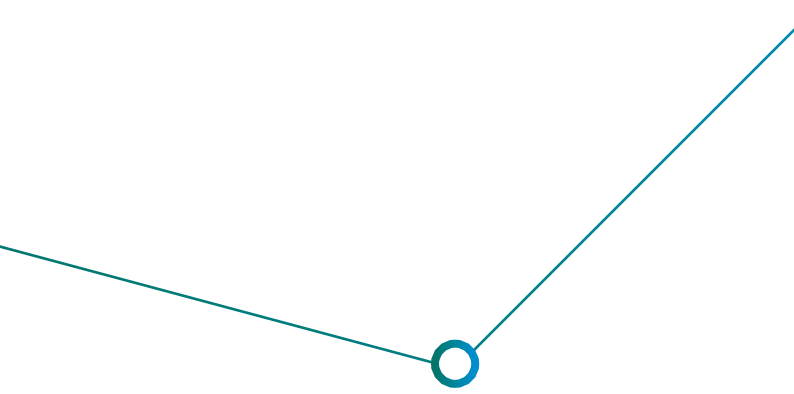
Prof. Dr. Vinzenz Brendler
Institut für Ressourcenökologie,
Helmholtz-Zentrum Dresden-Rossendorf
e.V.

Offensichtlich besteht ein Bedarf an digitalen Zwillingen (DT) für ein nukleares Endlager (NWR), wie z.B. von der EURAD-Gemeinschaft /PRA 20/¹ und in Deutschland /KRA 21/² hervorgehoben wurde. Ein solcher ganzheitlicher Ansatz kann unerwartete Querwirkungen überprüfen, erleichtert die Aktualisierung sowohl der Eingabedaten als auch der Modellspezifikation (wie sie im Laufe des Standortauswahlprozesses zu erwarten ist), unterstützt die Transparenz aller Schritte, fördert eine effiziente Nutzung von Ressourcen, ermöglicht eine Beschleunigung der Berechnungen und bietet die Grundlage für globale Sensitivitäts- und Unsicherheitsanalysen.

Es gibt jedoch mehrere Herausforderungen: die Komplexität der Aufgabe, die Arbeit auf sehr unterschiedlichen Skalen, die Allgegenwart von Ersatzdatenmatrizen, hohe Rechenkosten, die Notwendigkeit einer wirklich interdisziplinären Zusammenarbeit, ein Mangel an Standardisierung, Bestimmungen zur Nachhaltigkeit in Bezug auf Daten, Codes und Modellierungsansätzen. Und nicht zuletzt zögern alle Interessengruppen und insbesondere die Öffentlichkeit, auf Black-Box-Tools zu vertrauen. Deshalb setzt die EU im Bereich der künstlichen Intelligenz auf Exzellenz und Vertrauen, um Forschung und industrielle Kapazitäten zu fördern und gleichzeitig Sicherheit und Grundrechte zu gewährleisten: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.



„Chemie“ = pH, E_H, Ionenstärke, K_d, [wichtige Elemente], p_{CO₂}, ... ; für jede Domäne sind spezifische Mittelwerte und Varianzen dieser Parameter gegeben. K_d-Werte resultieren selbst wieder aus speziellen SF!



Ein frühes Projekt (VIRTUS – Virtuelles Untertage-labor im Steinsalz /GRS 14/³) konzentrierte sich auf die interaktive Visualisierung. Erste DT-Ansätze auf EURATOM-Ebene – „Pre-disposal management of radioactive waste (PREDIS)“ – betreffen die Zwischenlagerung ab 2021. Um weitere Fortschritte bei der Erstellung eines digitalen Zwillings für NWR zu erzielen, wurde kürzlich ein Vorschlag entwickelt, der sich auf die Kompartimente eines NWR konzentriert, in denen die geologischen Bedingungen die Radionuklidmigration bestimmen, d. h. die in der Abbildung blau umrahmten Felder. Hier wird ein Bottom-up-Ansatz gegenüber einer Top-down-Strategie bevorzugt. Zu diesem Zweck wird jedes der Kompartimente in eine Reihe von geologischen Bereichen unterteilt. In einer ersten Stufe wird die in Verteilungskoeffizienten ausgedrückte Retardierung innerhalb einer jeden solchen Domäne durch Training von Surrogatfunktionen über den Smart-Kd-Ansatz zugänglich gemacht /STO 17/⁴. Mögliche Surrogate umfassen Entscheidungsbäume (Decision Trees: /DEL 21/⁵), algorithmische Ansätze wie Arbeiten über angepasste Taylorreihen, z. B. /LEA 20/⁶ oder multivariate adaptive Regressionssplines /KEA 16/⁷, Kernel Ridge Regression /HAS 09/⁸ oder Gaussian Process Regression /RAS 06/⁹, aber auch Reduced Basis Methods /CHE 22/¹⁰ oder Physics-Informed Neural Networks (PINN: /HOR 91/¹¹, /LAL 21/¹²).

In einer zweiten Stufe (siehe Abbildung) werden diese Retardationsparameter dann mit allen geologischen Informationen kombiniert, um den reaktiven Transport von Radionukliden zu modellieren. ML-Methoden werden insbesondere für die Modellierung physikalischer Systeme mit Hilfe nichtlinearer, gewöhnlicher und partieller gekoppelter Differentialgleichungen (ODEs und PDEs) benötigt. Auch für solche Probleme sind PINNs anpassungsfähig /RAI 19/¹³. Die durch PINNs induzierten Ersatzmodelle müssen eine hohe Approximationsqualität im Vergleich zum Grundwahrheitsmodell (d.h. zur Referenzlösung) aufweisen und mit vertretbarem Trainingsaufwand erzeugt werden können. Beide Herausforderungen können für eine große Anzahl von Problemen /HEC 20/¹⁴, die in dem hier vorgestellten Rahmen zum Tragen kommen, gewährleistet werden.

- 1 /PRA 20/ Prasianakis, N. (2020) „Towards digital twins: machine learning based process coupling and multiscale modelling of reactive transport phenomena“, Goldschmidt Conference 2020, www.psi.ch/en/media/70145/download?attachment.
- 2 /KRA 21/ Krafczyk, M.; Brendler, V.; Czaikowski, O.; Gruner, M.; Hoth, N.; Kolditz, O.; Nagel, Th.; Herold, Ph.; Müller, Ch.; Seher, H.; Simo, E.; Stahlmann, J. (2021) „Eine Roadmap zur Entwicklung und Adaption von Methoden der Künstlichen Intelligenz (KI) für die Endlagerforschung.“ zenodo.org/record/5752277#.Yaon4NDMKPo.
- 3 /GRS 14/ VIRTUS – Virtuelles Untertagelabor im Steinsalz (GRS Report 354) (2014); www.grs.de/en/research-and-assessment/disposal/virtus-virtual-underground-laboratory.
- 4 /STO 17/ Stockmann, M., Schikora, J., Becker, D.-A., Flügge, J., Noseck, U., Brendler, V. (2017) *Chemosphere* 187, 277–285. doi.org/10.1016/j.chemosphere.2017.08.115.
- 5 /DEL 21/ De Lucia, M., Kühn, M. (2021) *Geosci. Model Dev.* 14, 4713–4730. doi.org/10.5194/gmd-14-4713-2021.
- 6 /LEA 20/ Leal, A.M.M., Kyas, S., Kulik, D.A., Saar, M.O. (2020) *Transp. Porous Media* 4, 161–204. doi.org/10.1007/s11242-020-01412-1.
- 7 /KEA 16/ Keating, E.H., Harp, D.H., Dai, Z., Pawar, R.J. (2016) *Intern. J. Greenhouse Gas Control* 46, 187–19. doi.org/10.1016/j.ijggc.2016.01.008.
- 8 /HAS 09/ Hastie, T., Tibshirani, R., Friedman, J. (2009) „The Elements of Statistical Learning. Data Mining, Inference, and Prediction“, Springer.
- 9 /RAS 06/ Rasmussen, C.E., Williams, C.K.I. (2006) „Gaussian processes for machine learning“, MIT Press.
- 10 /CHE 22/ Chellappa, S., Feng, L., Benner, P. in „Realization and Model Reduction of Dynamical Systems“, pp 137–155, Springer, 2022.
- 11 /HOR 91/ Hornik, K. (1991) *Neural Networks* 4, 251. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T).
- 12 /LAL 21/ Laloy, E., Jacques, D. (2021) „Speeding up reactive transport simulations in cement systems by surrogate geochemical modeling: deep neural networks and k-nearest neighbors“. arxiv.org/pdf/2107.07598.pdf.
- 13 /RAI 19/ Raissi, M., Perdikaris, P., Karniadakis, G.E. (2019) *J. Comput. Phys.* 378, 686–707. doi.org/10.1016/j.jcp.2018.10.045.
- 14 /HEC 20/ Hecht, M., Gonciarz, K., Michelfeit, J., Sivkin, V., Sbalzarini, I.F. (2020) „Multivariate interpolation in unisolvent nodes – lifting the curse of dimensionality“. arxiv.org/abs/2010.10824.

Unterstützung politischer Entscheidungen durch KI-gestützte Auswertung von Bürger:innen-beteiligungungsverfahren

Julia Romberg*
Heinrich-Heine-Universität Düsseldorf

Politische Partizipation beschreibt das Verhalten von Bürger:innen, das entweder alleine oder zusammen freiwillig ausgeführt wird, um politische Entscheidungen zu beeinflussen¹. Eine wichtige Form der politischen Partizipation ist die Öffentlichkeitsbeteiligung, die von Behörden genutzt wird, um die Bevölkerung zu bestimmten politischen Themen zu konsultieren.² Zielt die Partizipation primär auf die Bürger:innen ab, die nicht im Rahmen organisierter Interessensgruppen und Verbände agieren, bezeichnet man sie spezifischer auch als Bürger:innenbeteiligung.

Bürger:innenbeteiligungsverfahren verfolgen in der Regel zwei Ziele.³ Einerseits soll die Einbeziehung des kollektiven Wissens zu einem informierten Entscheidungsprozess führen, damit wirksame politische Maßnahmen abgeleitet werden können. Andererseits wird antizipiert, dass die Möglichkeit, Wissen einzubringen, Bedenken zu äußern und (bis zu einem gewissen Grad) die finalen Politikentscheidungen mitzugestalten, zu einer höheren Akzeptanz und idealerweise auch zu einer höheren Zufriedenheit mit den getroffenen Entscheidungen bei den Bürger:innen führt. Letzten Endes soll so die Legitimität der politischen Maßnahmen gestärkt werden.

Partizipation kann auf allen Verwaltungsebenen und in verschiedenen Formaten umgesetzt werden. Dabei wird der Input der Bürger:innen oft in textueller Form gesammelt, beispielsweise wenn Online-Plattformen oder Fragebögen mit Freitextfeldern eingesetzt werden.

Die anschließende Auswertung dieser Textbeiträge ist von entscheidender Bedeutung, um handlungsrelevante Informationen zu identifizieren und als Grundlage für den weiteren politischen Prozess zu nutzen.

In der Regel erfolgt die Auswertung der gesammelten Beitragsbeiträge manuell durch die Verwaltung oder beauftragte Dienstleister:innen. Die einzelnen Schritte der Auswertung umfassen unter anderem das wiederholte Lesen der Beiträge, die Erkennung von Duplikaten, die Vorsortierung der Beiträge nach Themen und Verwaltungseinheiten sowie eine tiefergreifende Analyse der einzelnen Standpunkte.⁴ Um demokratischen Normen gerecht zu werden, ist hierbei die Sicherung eines fairen und transparenten Entscheidungsprozesses wichtig, in dem alle eingegangenen Beiträge mit gleicher Sorgfalt begutachtet werden.⁵ Die Wahrnehmung der Legitimität von Entscheidungen ist dabei direkt von dem Vorgehen der Behörden bei der Auswertung der Beiträge abhängig⁶: Werden die zuvor genannten Kriterien in der öffentlichen Wahrnehmung nicht eingehalten, kann die daraus resultierende Politik als weniger legitim angesehen werden.⁷ Diesen Anforderungen gerecht zu werden, erfordert einen hohen Einsatz von Zeit, Personal, und folglich auch finanziellen Mitteln, was im schlimmsten Fall zu einer scheinbar unüberwindbaren Hürde für die erfolgreiche Umsetzung von Öffentlichkeitsbeteiligung werden kann.⁸

Als Lösung bietet sich der Einsatz von Künstlicher Intelligenz (KI) an. Die Idee ist dabei, verschiedene Teilaufgaben vollständig oder teilweise zu automatisieren, wie beispielsweise die Erkennung von Duplikaten⁹, die Erkennung von Themen und eine entsprechende Gruppierung der Beiträge¹⁰, die Analyse der öffentlichen Stimmung¹¹ und von Argumenten¹² sowie die Zusammenfassung von Inhalten.¹³

In der BMBF-geförderten Forschungsgruppe CIMT¹⁴ (Citizen Involvement in Mobility Transitions, zu Deutsch Bürger:innenbeteiligung an der Verkehrswende) liegt ein Schwerpunkt auf der Entwicklung solcher KI-Techniken, die die Auswertung von

* Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung im Förderschwerpunkt Sozial-ökologische Forschung (Förderkennzeichen 01UU1904) gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin/ beim Autor. Kontakt: Julia.Romberg@uni-duesseldorf.de.

Bürger:innenbeiträgen erleichtern. Im Kontext von Verkehrsplanungsverfahren wurden relevante Teilaufgaben in Interviews mit Praktiker:innen identifiziert¹⁵ und KI-basierte Lösungen entwickelt. Im Folgenden wird eine Auswahl dieser Verfahren vorgestellt.

Omnipräsent zeichnet sich der Bedarf einer Unterstützung durch KI bei der thematischen Kategorisierung. Hierzu können Algorithmen des überwachten maschinellen Lernens eingesetzt werden, die mittels händisch kodierter Trainingsdaten die Klassifikation in vorgegebene Themencluster erlernen. Um den verbleibenden manuellen Aufwand möglichst gering zu halten, wurden in Romberg und Escher¹⁶ spezifische Methoden für die Auswertung von Öffentlichkeitsbeteiligungen entwickelt. Das vorgeschlagene Verfahren kann in acht von zehn Fällen die richtigen Themen zuordnen, während es den manuellen Kategorisierungsaufwand gleichzeitig bis zu einem Fünftel senkt. Am praktischen Anwendungsfall dreier kommunaler Beteiligungsverfahren veranschaulicht bedeutet das, dass lediglich 120 von 459, 180 von 366, und 500 von 2314 Beiträgen des jeweiligen Verfahrens von Hand kategorisiert werden müssen.

Neben der thematischen Strukturierung sind die einzelnen Meinungen und Standpunkte der Bürger:innen für die Auswertung von Interesse. Eine automatisierte Kennzeichnung von Argumenten kann insbesondere bei langen Texten die Auswertung beschleunigen, da andere Textpassagen, die für diesen Aspekt der Analyse nicht relevant sind, ausgeblendet werden können. In Romberg und Conrad¹⁷ wurden deshalb Methoden präsentiert, mit denen argumentative Textbereiche in Bürger:innenbeiträgen identifiziert werden können. Diese erreichen eine Genauigkeit von 77 Prozent. Weiterhin können die argumentativen Passagen danach aufgeteilt werden, ob sie einen Zustand beschreiben (beispielsweise einen aktuellen Missstand) oder einen Vorschlag enthalten (beispielsweise einen Lösungsansatz zur Behebung des Missstandes). Diese Unterscheidung kann von einer KI bereits in neun von zehn Fällen richtig getroffen werden.

Anknüpfend ist für die Ableitung von politischen Maßnahmen aus den gesammelten Beiträgen auch die Frage danach relevant, wie konkret Zustände und Vorschläge beschrieben werden. Je eindeutiger die Beschreibung ist, desto leichter gestaltet sich die Auswertung. Bei vagen Ideen oder unklar formulierten Problembeschreibungen ist hingegen deutlich mehr Arbeit nötig, bevor Maßnahmen formuliert werden können, da der Interpretationsspielraum groß ist und offene Fragen geklärt werden müssen. Einfache KI-Modelle können hier bereits eine 79-prozentige Genauigkeit liefern.¹⁸

1 Max Kaase. 2000. Politische Beteiligung/Politische Partizipation. In: Andersen, U., Woyke, W. (eds) Handwörterbuch des politischen Systems der Bundesrepublik Deutschland. VS Verlag für Sozialwissenschaften, Wiesbaden. https://doi.org/10.1007/978-3-322-93232-7_105.

2 Stephanie Bock und Bettina Reimann. 2021. Mit dem Los zu mehr Vielfalt in der Bürgerbeteiligung? Chancen und Grenzen der Zufallsauswahl. Kursbuch Bürgerbeteiligung# 4, 184-199. <https://repository.difu.de/handle/difu/583064>.

3 Vivien A. Schmidt. 2013. Democracy and Legitimacy in the European Union Revisited: Input, Output and "Throughput." *Polit. Stud.* 61, 1 (2013), 2-22. <https://doi.org/10.1111/j.1467-9248.2012.00962.x>

4 Julia Romberg und Tobias Escher. 2020. Analyse der Anforderungen an eine Software zur (teil-)automatisierten Unterstützung bei der Auswertung von Beteiligungsverfahren. Working Paper 1, CIMT Research Group, Institute for Social Sciences, Heinrich Heine University Düsseldorf. https://www.cimt-hhu.de/wp-content/uploads/2020/12/cimt_working_paper1.pdf; Mahmood Jasim, Enamul Hoque, Ali Sarvaghad, und Narges Mahyar. 2021. Communitypulse: Facilitating community input analysis by surfacing hidden insights, reflections, and priorities. In *Designing Interactive Systems Conference 2021, DIS '21*, page 846-863, New York, NY, USA. Association for Computing Machinery. ISBN 9781450384766. <https://doi.org/10.1145/3461778.3462132>; Stuart Shulman, Jamie Callan, Eduard Hovy, und Stephen Zavestoski. 2004. SGER collaborative. *Journal of E-Government*, 1(1):123-127. https://doi.org/10.1300/J399v01n01_09; Anthony Simonofski, Jérôme Fink, und Corentin Burnay. 2021. Supporting policy-making with social media and e-participation platforms data: A policy analytics framework. *Government Information Quarterly*, 38(3):101590.; Michael A. Livermore, Vladimir Eidelman, und Brian Grom. 2017. Computationally assisted regulatory participation. *Notre Dame L. Rev.*, 93:977.

5 Robert A. Dahl. 1989. *Democracy and its critics*. Yale university press, Yale.

6 Vivien A. Schmidt. 2013. Siehe Fußnote 3.

7 Peter Esaiasson. Will citizens take no for an answer? What government officials can do to enhance decision acceptance. *European Political*

Science Review, 2(3):351-371, 2010.; Michael A. Strebelt, Daniel Kübler, und Frank Marcinkowski. The importance of input and output legitimacy in democratic governance: Evidence from a population-based survey experiment in four west european countries. *European Journal of Political Research*, 58(2):488-513, 2019.

8 Narges Mahyar, Diana V. Nguyen, Maggie Chan, Jiayi Zheng, und Steven P. Dow. The civic data deluge: Understanding the challenges of analyzing large-scale community input. In *Proceedings of the 2019 on Designing Interactive Systems Conference, DIS '19*, page 1171-1181, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450358507. doi: 10.1145/3322276.3322354. URL <https://doi.org/10.1145/3322276.3322354>.

9 Helen Yang und Jamie Callan. 2005. Near-duplicate detection for eRulemaking. In *Proceedings of the 2005 National Conference on Digital Government Research*, pages 78-86. Digital Government Research Center.

10 Namhee Kwon, Stuart W. Shulman, und Eduard Hovy. 2006. Multidimensional text analysis for eRulemaking. In *Proceedings of the 7th Annual International Conference on Digital Government Research*, pages 157-166. Digital Government Research Center.; Helen Yang und Jamie Callan. 2009. Ontocop: Constructing ontologies for public comments. *IEEE Intelligent Systems*, 24(5):70-7.

11 Kwon et al. 2006. Siehe Fußnote 10; Jasmin et al. 2021. Siehe Fußnote 4.

12 Namhee Kwon et al. 2006. Siehe Fußnote 10; Joonsuk Park und Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. In *Proceedings of the First Workshop on Argument Mining*, pages 29-38. Association for Computational Linguistics.

13 Miguel Arana-Catania, Felix-Anselm V. Lier, Rob Procter, Nataliya Tkachenko, Yulan He, Arkaitz Zubiaga, und Maria Liakata. Citizen participation and machine learning for a better democracy. *Digital Government: Research and Practice*, 2(3):1-22, 2021.

14 <https://www.cimt-hhu.de/>.

15 Julia Romberg und Tobias Escher. 2020. Siehe Fußnote 4.

Zusammenfassend lässt sich festhalten, dass Verfahren zur KI-gestützten Auswertung von Textbeiträgen bereits vielversprechende Ergebnisse zeigen. Insbesondere in der thematischen Vorsortierung kann der Zeitaufwand für die Auswertung drastisch reduziert werden, wovon insbesondere Prozesse mit einer größeren Anzahl von Beiträgen profitieren können. Dennoch sind wichtige Aspekte der praktischen Einsetzbarkeit von KI in der Öffentlichkeitsbeteiligung noch zu klären.

Dazu gehören Fragen danach, wie zuverlässig die Methoden sein müssen, damit Praktiker:innen sie mit Vertrauen nutzen können, und wie anwendungsspezifische benutzer:innenfreundliche Implementierungen aussehen sollten.¹⁹ Ebenso sollte die Unterstützung nicht auf textuelle Beteiligung begrenzt bleiben, sondern möglichst weitere Formate, wie gesprochene Meinungsäußerungen, unterstützen. Hierzu sind Methoden der Spracherkennung denkbar, die Sprache in Text übersetzen. Schließlich beschränkt sich das Potential von KI nicht nur auf die Auswertungsphase, sondern kann auch während oder vor der Beteiligungsphase unterstützen, beispielsweise um die Moderation der Verfahren zu vereinfachen oder den Input zu verbessern.

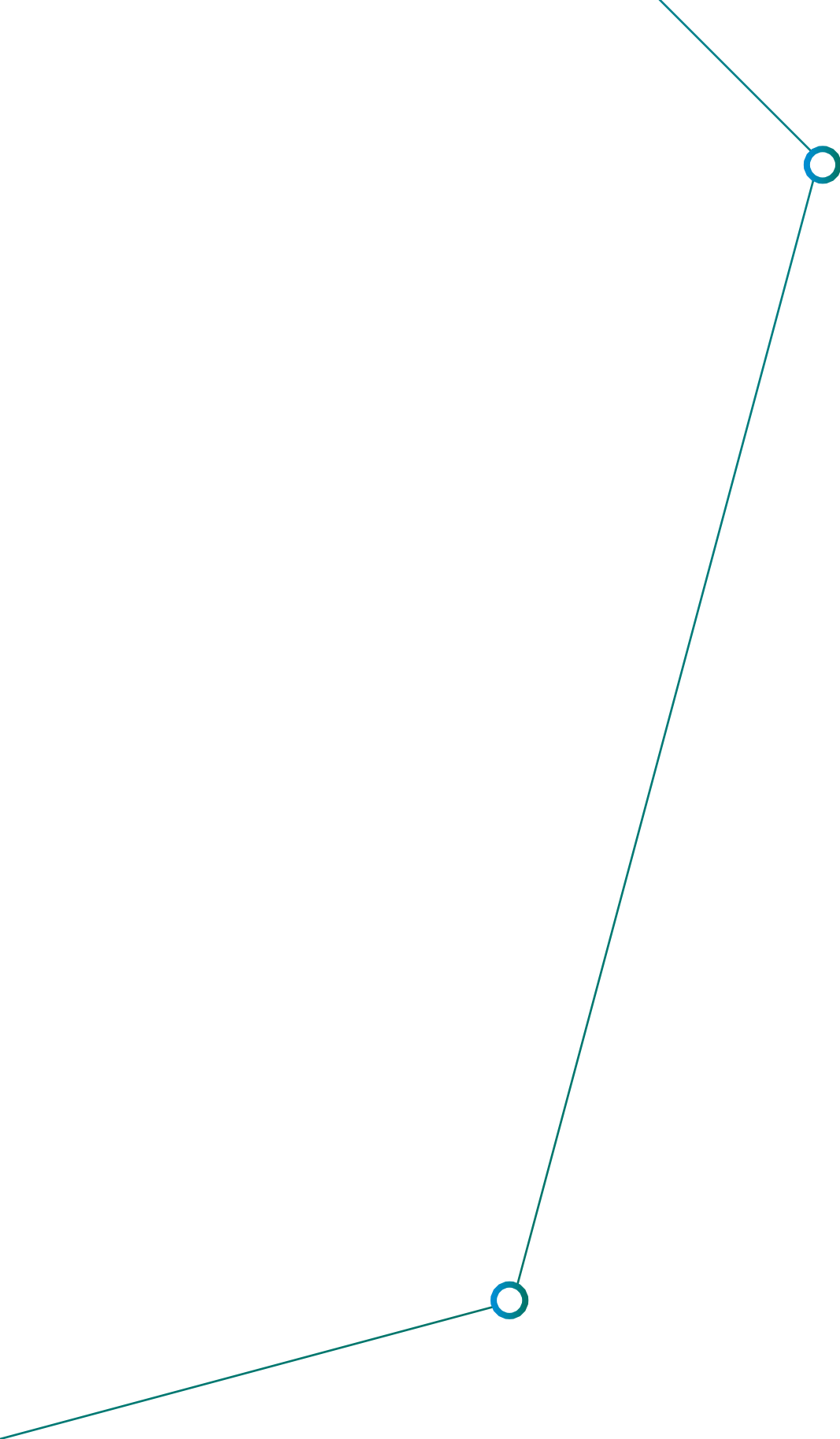
16 Julia Romberg und Tobias Escher. 2022. Automated Topic Categorisation of Citizens' Contributions: Reducing Manual Labelling Efforts Through Active Learning. In: Marijn Janssen, et al. Electronic Government. EGOV 2022. Lecture Notes in Computer Science, vol 13391. Springer, Cham. https://doi.org/10.1007/978-3-031-15086-9_24.

17 Julia Romberg und Stefan Conrad. 2021. Citizen Involvement in Urban Planning – How Can Municipalities Be Supported in Evaluating Public Participation Processes for Mobility Transitions?. In Proceedings of the 8th Workshop on Argument Mining (pp. 89-99), Punta Cana, Dominican Republic. Association for Computational Linguistics. <https://aclanthology.org/2021.argmining-1.9>.

18 Julia Romberg. 2022. Is Your Perspective Also My Perspective? Enriching Prediction with Subjectivity. In Proceedings of the 9th Workshop on Argument Mining (pp.115-125), Gyeongju, Republic of Korea. Association for Computational Linguistics. <https://aclanthology.org/2022.argmining-1.11>.

19 Jasim et al. 2021. Siehe Fußnote 4.





Zwischen Effizienz und Kontrollverlust: Der Einsatz von Künstlicher Intelligenz durch die Verwaltung im Spiegel des Rechts

**Prof. Dr. Stephanie Schiedermair und
Johannes Weil***
*Juristenfakultät der Universität Leipzig,
ScaDS.AI Dresden/Leipzig*

Chancen und Herausforderungen für Recht und Verwaltung

Die vielfältigen Formen und Einsatzfelder von Künstlicher Intelligenz (KI) machen auch vor der Verwaltung keinen Halt, sind sie doch mit dem Versprechen verbunden, eine Effizienzsteigerung des Verwaltungshandelns bei gleichzeitigen Kosteneinsparungen und Entlastungen des Personals zu bewirken.¹

Bereits bestehende und potenzielle Einsatzfelder ergeben sich sowohl im Bürgerkontakt (Front Office) – z.B. über den Einsatz von Chatbots oder zur Antragsunterstützung – als auch bei der Automatisierung von behördeninternen Prozessen (Back-Office). Besonders intensiv diskutiert werden Systeme, die zur Vorbereitung von außenwirksamen Entscheidungen durch umfassende Datenauswertung, Prognosen oder Überprüfung beitragen oder sogar vollautomatisiert Entscheidungen generieren sollen.² Im Kontext der nuklearen Endlagerung wird derzeit die Anwendung von KI zur Kontrolle, Analyse und Auswertung von Geodaten im Rahmen des Standortauswahlverfahrens untersucht.³

Den großen Verheißungen von KI-Technologien stehen dabei aber verschiedene praktische Probleme gegenüber, welche das Recht und damit gerade die strikt gesetzgebundene Verwaltung vor große Herausforderungen stellen: So sind die Systeme zum einen auf erhebliche Datenmengen angewiesen. Zum anderen bestehen vor allem bei komplexeren Anwendungen im Bereich des „Deep Learning“ Schwierigkeiten bei der Nachvollziehbarkeit und Erklärbarkeit des Entscheidungsprozesses („Black Box“-Phänomen), was eine hinreichende Kontrolle und rechtliche Überprüfbarkeit erschwert. Hinzu kommt, dass KI-Systeme häufig auch eine gewisse Fehlerrate aufweisen oder sogar diskriminierende Ergebnisse ausgeben können („Bias in AI“).⁴

Wenngleich der hoheitliche Einsatz von KI rechtlich nicht pauschal bewertet werden kann, sondern vielmehr nach dem jeweiligen Einsatzfeld, dem konkreten Sachverhalt und den danach einschlägigen Vorschriften zu beurteilen ist, schließen sich an die dargelegten Problemstellungen übergreifende Rechtsfragen an, die sich typischerweise im Verfassungsrecht (II) sowie Datenschutz- (III) und Antidiskriminierungsrecht (IV) bewegen. Eine maßgebliche Rolle wird künftig der auf EU-Ebene anvisierten KI-Verordnung zukommen (V).

Verfassungsrechtliche Steuerungsvorgaben

Auf verfassungsrechtlicher Ebene setzen Demokratie- und Rechtsstaatsprinzip voraus, dass eine hinreichende staatliche Kontrolle über das verwendete KI-System besteht.⁵ Dabei ist es weder zwingend notwendig, dass sich dem zuständigen Amtsträger die Funktionsweise vollends erschließt, noch muss das Programm absolut fehlerfrei arbeiten. Schließlich beruht die Verwaltungspraxis bislang auf menschlichem Entscheidungsverhalten, welches ebenfalls nicht gänzlich durchdrungen werden kann und gewisse Fehlerraten aufweist. Insofern sind

* Die Autorin ist Inhaberin des Lehrstuhls für Europarecht, Völkerrecht und Öffentliches Recht an der Juristenfakultät der Universität Leipzig. Der Autor ist ihr wissenschaftlicher Mitarbeiter und Doktorand am ScaDS.AI Dresden/Leipzig. Diese Arbeit wurde ermöglicht durch die finanzielle Unterstützung des Bundesministeriums für Bildung und Forschung und des Sächsischen Staatsministeriums für Wissenschaft, Kultur und Tourismus im Rahmen des Programms Kompetenzzentren für KI-Forschung für das „Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig“, Förderkennzeichen: ScaDS.AI.

geeignete Vorkehrungen zu treffen, um sicherzustellen, dass in der Regel die erwünschten Ergebnisse erzielt werden. Abhilfe schaffen können etwa vorherige Testungen oder menschliche Aufsicht.⁶

Vor allem bei belastenden Entscheidungen muss zudem eine gerichtliche Überprüfbarkeit im Sinne eines effektiven Rechtsschutzes nach Art. 19 Abs. 4 GG zugunsten der Betroffenen gewährleistet sein. Von zentraler Bedeutung ist hierbei das rechtsstaatliche Begründungsgebot. Zwar besteht bei intransparenten Systemen insoweit ein gewisses Spannungsverhältnis. Zu beachten ist aber, dass keine Informationen über sämtliche entscheidungsursächliche Aspekte erforderlich sind, was ebenso für die maschinelle Entscheidungsfindung gilt. Gleichwohl muss die Begründung so gestaltet sein, dass Betroffene effektiv gegen die Entscheidung vorgehen können.⁷

Je nach Fallkonstellation können sich auch spezielle Anforderungen aus den Grundrechten ergeben.⁸ So hat das Bundesverfassungsgericht mit Blick auf das Recht auf informationelle Selbstbestimmung jüngst klargestellt, dass dem Einsatz von KI regelmäßig ein besonderes Eingriffsgewicht zukommt und verfahrensrechtliche Vorkehrungen im Hinblick auf die spezifischen Gefahren zu treffen sind.⁹

Generell hängt das rechtliche Anforderungsniveau von verschiedenen Faktoren ab, namentlich der konkreten Einsatzsituation, der Grundrechtsrelevanz der dem System überantworteten Aufgabe, der Kontrollierbarkeit, der Umkehrbarkeit des Vorgangs, dem Schadenspotential und der Verfügbarkeit von alternativen Methoden.¹⁰ Insofern werden tendenziell stärkere Vorgaben im Kontext von außenwirkenden Entscheidungen zu stellen sein, während der bloße Einsatz von KI im Front- oder Back-Office unproblematischer erscheint. Für die Verwendung von KI zur Unterstützung der Standortauswahl bei der nuklearen Entsorgung dürften danach – vor allem aufgrund der hohen Grundrechtsrelevanz¹¹ – stärkere Anforderungen gelten.

Datenschutzrechtliche Spannungsverhältnisse

In Konflikt geraten die stark datengetriebenen KI-Technologien zwangsläufig mit dem Datenschutzrecht. Die Regelungen der insoweit zentralen europäischen Datenschutzgrundverordnung (DS-GVO) gelten gleichwohl nur für die Verarbeitung von personenbezogenen Daten, Art. 2 Abs. 1 DS-GVO. Werden für die Suche nach einem geeigneten Standort für die Endlagerung lediglich Geodaten ausgewertet, die keine Rückschlüsse auf natürliche Personen zulassen, ist das Regelungsregime der DS-GVO demnach nicht zu beachten.

Bewegt sich eine KI-basierte Datenverarbeitung aber im Anwendungsbereich der DS-GVO, tun sich Spannungsfelder im Hinblick auf die Grundsätze des Datenschutzes auf, die jeweils durch spezielle Regelungen in der DS-GVO ausgeformt werden. Zunächst müssen personenbezogene Daten transparent, d.h. in einer für die betroffene Person nachvollziehbaren Weise, verarbeitet werden (Art. 5 Abs. 1 lit. a DS-GVO). Konflikte können sich hier speziell bei undurchsichtigen, komplexen Systemen ergeben. Vor allem mit Blick auf die Informationspflichten und Auskunftsrechte der Art. 13 ff. DS-GVO ist der Umfang, in welchem über die Funktionsweise eines KI-Systems aufzuklären ist, im Einzelnen umstritten.¹²

Die für KI-Systeme charakteristische Angewiesenheit auf große Datenmengen steht in Kontrast zu dem in Art. 5 Abs. 1 lit. c DS-GVO verankerten Grundsatz der Datenminimierung. Kompromisse können hier dadurch gefunden werden, dass die verwendeten Daten soweit wie möglich pseudonymisiert oder gar anonymisiert werden.¹³ Zu einem entsprechenden Maß an Datenqualität zwingt das Prinzip der Datenrichtigkeit (Art. 5 Abs. 1 lit. d DS-GVO). Auch hier können sich Probleme bei Black Box-Systemen ergeben, weil Fehler sich regelmäßig nur schwer identifizieren lassen.¹⁴

„Bias in, Bias out“ – Antidiskriminierungsrecht und KI

Ein weiterer Problembereich – der zumindest bei der bloßen Auswertung von Geodaten weniger relevant sein dürfte – resultiert daraus, dass KI-Anwendungen diskriminierend wirken können. Mögliche Gründe sind etwa eine unvollständige Datengrundlage, in der eine bestimmte Gruppe über- bzw. unterrepräsentiert ist, oder Trainingsdaten, die auf diskriminierenden menschlichen Entscheidungen beruhen.¹⁵

Diskriminierungsverbote, welche eine Ungleichbehandlung wegen geschützter Merkmale wie ethnischer Herkunft, Religion oder Geschlecht durch staatliche Stellen im Grundsatz verbieten, existieren sowohl auf unions- als auch auf verfassungsrechtlicher Ebene (z.B. Art 3 Abs. 3 GG, Art. 21 Abs. 1 GRC). Diese sind auch anwendbar, wenn das verwendete KI-System nicht unmittelbar an eine solche Kategorie, sondern an sog. Stellvertretermerkmale anknüpft, welche typischerweise mit diesen korrelieren.¹⁶

Häufig lässt sich die Diskriminierung aber nicht ohne weiteres nachweisen, vor allem wenn es an der Nachvollziehbarkeit des Systems mangelt.⁴⁷ Wird ein KI-System eingesetzt, das nach seiner Beschaffenheit eine diskriminierende Wirkung potenziell möglich erscheinen lässt, müssen daher spezifische Vorkehrungen getroffen werden, die etwa eine angemessene Qualität der Trainingsdaten sicherstellen.⁴⁸ Liegt eine Diskriminierung vor, dürfte sich diese zumindest mit Blick auf die von Art. 3 Abs. 3 GG geschützten Merkmale nur in seltenen Ausnahmefällen rechtfertigen lassen, weil hier außerordentlich gewichtige Gründe zu fordern sind, eine bloße Effizienzsteigerung durch KI kann nicht genügen.

Spezifische Rechtsentwicklung auf EU-Ebene

Eine spezifische Antwort auf die Problemstellungen, die typischerweise mit dem Einsatz von KI verbunden sind, kündigt sich auf EU-Ebene an. Die Kommission hat im April 2021 einen Entwurf für eine KI-Verordnung (KI-VO) veröffentlicht, der sich aktuell noch im Gesetzgebungsverfahren befindet.⁴⁹ Der Entwurf verfolgt einen gestuften, risikobasierten Ansatz. Einsatzformen von KI, die als schlechthin zu risikobehaftet gelten – etwa staatliche Social Scoring-Systeme, werden verboten (Art. 5 KI-VO).

Ein erhöhtes Pflichtenprogramm gilt für sog. Hochrisiko-Systeme, die vor allem durch ihren Einsatzbereich charakterisiert werden (Art. 6 u. Anhang III KI-VO), wozu etwa Strafverfolgung und Personalmanagement zählen. Insbesondere werden auch die Verwaltung und der Betrieb kritischer Infrastrukturen erfasst. Als Hochrisiko-Systeme sind danach u.a. solche Systeme zu qualifizieren, die bestimmungsgemäß als Sicherheitskomponenten in der Stromversorgung verwendet werden sollen. Durch die Verbindung zur Atomenergie und den starken Grundrechtsbezug ist es denkbar, dass hierunter auch KI-Systeme mit Bezug zur nuklearen Entsorgung fallen könnten. Als Anforderungen an Hochrisiko-KI-Systeme schreiben die Art. 9 ff. KI-VO u.a. ein Risikomanagementsystem mit vorheriger Risikoevaluation, eine Qualitätssicherung hinsichtlich der relevanten Datensätze, eine technische Dokumentation, Transparenz- und Informationspflichten, die Gewährleistung durchgängiger menschlicher Aufsicht und ein angemessenes Maß an Genauigkeit, Robustheit und Cybersicherheit vor.

Unabhängig davon werden spezielle Transparenzpflichten für bestimmte gefahrgeneigte Anwendungen wie Chatbots, Emotionserkennungssysteme und Deepfakes festgelegt (Art. 52 KI-VO). Systeme, die den vorgenannten Kategorien nicht unterfallen, also nach der Konzeption des Entwurfs nicht mit einem erhöhten Risiko behaftet sind, bleiben hingegen unreguliert.

Schlussbetrachtung

Aus rechtlicher Sicht bestehen mit Blick auf den Einsatz von KI durch die Verwaltung noch viele offene Fragen, die einer Klärung bedürfen, bevor die Technologien staatlicherseits wirklich breite Verwendung finden. Ziel sollte es dabei sein, einen robusten rechtsstaatlichen und grundrechtskonformen Rahmen zu schaffen, der gleichzeitig Raum für Innovation lässt, um eine effiziente und bürgerfreundliche Ausgestaltung der Verwaltung voranzutreiben. Neben spezifischen Regulierungsvorhaben wie der europäischen KI-Verordnung braucht es dafür auch einen interdisziplinären Austausch, der u.a. technische, rechtliche und ethische Aspekte zusammenbringt.

- 1 Hierzu Wirtz/Weyerer, Künstliche Intelligenz im öffentlichen Sektor: Anwendungen und Herausforderungen, Verwaltung & Management 25 (2019), 37 (39); Coglianese/Ben Dor, AI in Adjudication and Administration, Brooklyn Law Review 86 (2021), 791 (827 f.).
- 2 Im Überblick Etscheid/von Lucke/Stroh, Künstliche Intelligenz in der öffentlichen Verwaltung, Anwendungsfelder und Szenarien, 2020, <https://publica-rest.fraunhofer.de/server/api/core/bitstreams/d3d9f520-1fd4-4516-98d6-a3370c134155/content> (Stand aller Links: 6.3.2023).
- 3 S. Öko-Institut e.V., Künstliche Intelligenz in der Endlager-suche: nur unter transparenten Bedingungen, 7.12.2022, <https://www.oeko.de/presse/archiv-pressemeldungen/presse-detailseite/2022/kuenstliche-intelligenz-in-der-endlagersuche-nur-unter-transparenten-bedingungen>.
- 4 Instruktiv zu den KI-bezogenen Herausforderungen für das Recht Wischmeyer, Regulierung intelligenter Systeme, AöR 2018, 1.
- 5 Vgl. BVerfG, Urt. v. 16.2.2023 – 1 BvR 1547/19, Rn. 100 im Kontext des Grundrechts auf informationelle Selbstbestimmung.
- 6 Zu allem Busche, Staatliche Verantwortungsübernahme beim hoheitlichen Einsatz intransparenter Algorithmen, in: Ammann et al. (Hrsg.), Verantwortung und Recht, 2022, S. 45 (55 ff.).
- 7 S. Wischmeyer, in: Ebers/Heinze/Krügel/Steinrötter (Hrsg.), Künstliche Intelligenz und Robotik, Rechtshandbuch, 2020, § 20 Rn. 56.
- 8 Für einen Überblick in Bezug auf die EU-Grundrechte s. Gonzáles Fuster, Artificial Intelligence and Law Enforcement, Impact on Fundamental Rights, 2020, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf).
- 9 S. BVerfG, Urt. v. 16.2.2023 – 1 BvR 1547/19, Rn. 100 ff.
- 10 S. Unger, Demokratische Herrschaft und künstliche Intelligenz, in: Unger/von Ungern-Sternberg (Hrsg.), Demokratie und künstliche Intelligenz, 2019, S. 113 (126); Busche, Staatliche Verantwortungsübernahme beim hoheitlichen Einsatz intransparenter Algorithmen, in: Ammann et al. (Hrsg.), Verantwortung und Recht, 2022, S. 45 (58 Fn. 56).
- 11 Dazu Emanuel, Rechtliche Rahmenbedingungen für die Suche nach einem Endlager für hochradioaktiven Abfall (HAW), 2021, S. 115 ff.
- 12 Ausführlich Hoeren/Niehoff, KI und Datenschutz – Begründungserfordernisse automatisierter Entscheidungen, RW 2018, 47 (54 ff.); Kumkar/Roth-Isigkeit, Erklärungs-pflichten bei automatisierten Datenverarbeitungen nach der DS-GVO, JZ 2020, 277; Hoffmann/Kevekordes, Das Right to Explanation, DuD 2021, 609.
- 13 S. Kloos/Schmidt-Bens, in: Hartmann (Hrsg.), KI & Recht kompakt, 2020, S. 165 (183).
- 14 S. Paal, in: Kaulartz/Braegelmann (Hrsg.), Rechtshandbuch Artificial Intelligence und Machine Learning, 2020, Kap. 8.7 Rn. 17.
- 15 S. Wischmeyer, in: Ebers/Heinze/Krügel/Steinrötter (Hrsg.), Künstliche Intelligenz und Robotik, Rechtshandbuch, 2020, § 20 Rn. 58.
- 16 Dazu Tischbirek, Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, 2020, S. 103 (Rn. 14 ff.).
- 17 S. Borgesius, Strengthening legal protection against discrimination by algorithms and artificial intelligence, The International Journal of Human Rights 24 (2020), 1572 (1577).
- 18 In diese Richtung BVerfG, Urt. v. 16.2.2023 – 1 BvR 1547/19, Rn. 77 u. 100.
- 19 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final; s. zuletzt den vorläufigen Standpunkt des Rats der EU v. 6.12.2022, ST 15698 2022 INIT; ausführlich Ebert/Spiecker gen. Döhmann, Der Kommissionsentwurf für eine KI-Verordnung der EU, NVwZ 2021, 1188.

Quo Vadis Künstliche Intelligenz in der nuklearen Entsorgung? Eine Schlussbetrachtung

Dr. Annika Froese und Ina Richter
Bundesamt für die Sicherheit der nuklearen Entsorgung (BASE)

Einen Tag lang haben die Expert:innen beim Panel des BASE am 1. Dezember 2022 diskutiert. Am Ende des Tages waren sie sich weitgehend einig: Künstliche Intelligenz (KI) birgt Potenziale, die im Bereich der nuklearen Entsorgung von Nutzen sein können. Allerdings gibt es nach wie vor noch zu wenig Erfahrung, um die Potenziale genau abschätzen zu können. Wo es allerdings vielversprechende Entwicklungen gibt, ist beim Umgang mit geowissenschaftlichen Daten. Um die Sicherheit eines Endlagers für hochradioaktive Abfälle zu gewährleisten, müssen die geologischen Entwicklungen, die innerhalb der nächsten Millionen Jahre zu erwarten sind, erfasst und analysiert werden. Dies kann nur auf Grundlage von umfangreichen und verlässlichen Daten geschehen. KI könnte dabei helfen, Daten zu klassifizieren, digitale Bilder zu verarbeiten und komplexe Prozesse zu vereinfachen. Eine weitere Anwendung von KI besteht in der Erstellung und Analyse eines digitalen Zwillings, bei dem es sich um eine digitale Darstellung des gesamten Endlagersystems handelt.

Ob mit KI nun neues Wissen generiert wird oder aber KI als Kontrollinstanz fungiert und zur Überprüfung bereits bestehender Erkenntnisse genutzt wird, zeigen diese Anwendungen in technisch-naturwissenschaftlichen Bereichen großes Potenzial. Aber auch im Bereich der Bürger:innenbeteiligung kann KI öffentliche Einrichtungen unterstützen. Beiträge im Rahmen von Beteiligungsverfahren können heutzutage oftmals digital eingereicht werden. Daraus ergeben sich umfangreiche Datensätze, die aus Textbeiträgen einzelner Bürger:innen bestehen und üblicherweise durch Mitarbeitende manuell ausgewertet werden müssen. Solch eine manuelle Verarbeitung beansprucht Ressourcen, die in öffentlichen Einrichtungen oftmals nicht vorhanden sind. KI kann hier unterstützen, indem es Textbeiträge kategorisiert, übergreifenden Gruppen zuordnet und Dopplungen eliminiert. Dies kann die weitere Bearbeitung durch Mitarbeitende effizienter gestalten und Ressourcen schonen. In der nuklearen

Entsorgung, insbesondere bei der Suche nach einem Standort für ein Endlager für hochradioaktive Abfälle, spielt Öffentlichkeitsbeteiligung eine zentrale Rolle. Hier besteht weiteres Potenzial für KI-Anwendungen.

Abgesehen von dem Mehrwert, den der Einsatz von KI im Bereich der nuklearen Entsorgung schaffen könnte, gibt es aber auch einige Herausforderungen. So zeigte sich während der Diskussion, dass bereits die Verwendung des Begriffs ‚Künstliche Intelligenz‘ hinterfragt werden muss. Wie einleitend erwähnt, umfasst dieser Begriff nämlich eine Vielzahl verschiedener Methoden. Es stellt sich daher die Frage, ob es der Klarheit dienlich ist, wenn generell von KI gesprochen wird, statt eine konkrete Methode zu benennen. Die Tatsache, dass der Begriff weitläufig gebraucht wird und in unterschiedlichen Kreisen mit unterschiedlichen Methoden, Wertungen und Erwartungshaltungen verbunden ist, bestätigt die Notwendigkeit einer kritischen Auseinandersetzung.

Die Unschärfe des KI-Begriffs erschwert zum einen die Regulierung der Technologien, die man versucht darunter zusammenzufassen. Zum anderen verbinden sich damit unterschiedliche Erwartungshaltungen in der Öffentlichkeit. Es ist nicht klar, was der Einsatz von KI bei den Menschen, die sich in Beteiligungsprozessen im Rahmen der nuklearen Entsorgung einbringen, auslöst. Werden beispielsweise die Ergebnisse eines Beteiligungsprozesses eher akzeptiert, wenn eine Maschine statt eines Menschen die eingebrachten Kommentare und Anmerkungen auswertet? Verleihen Technologien solchen Prozessen mehr Legitimität, weil die Öffentlichkeit ihnen mehr ‚Objektivität‘ zuspricht als etwa den zuständigen Behörden? Mit solchen Überlegungen werden auch grundsätzliche – noch wenig diskutierte – Fragen demokratischer Entscheidungsprozesse tangiert, die den Einsatz von KI auf bestimmte Anwendungsfelder beschränken könnten. Zum Beispiel stellt sich die Frage, wer die Verantwortung übernimmt für Entscheidungen und damit verbundene Handlungskonsequenzen, die durch eine Maschine begleitet bzw. vorbereitet werden. Wie kann sichergestellt werden, dass der Entwickler:innen-Bias, also die Meinungen derjenigen, die eine KI trainieren, politische Entscheidungen nicht einseitig beeinflussen?

Bei dem Versuch, den KI-Begriff klarer zu definieren, lässt sich außerdem feststellen, dass einige sogenannte KI-Methoden nicht neu sind bzw. auf schon lange verfügbare Methoden aufbauen. Es gibt beispielsweise keine klare Trennung zwischen KI-Ansätzen und klassischen Methoden numerischer Modellierung, die für das Feld der nuklearen Entsorgung relevant sind. Idealerweise sollte das Beste aus beiden Welten genutzt werden. Dass dies bisher nicht geschieht, liegt auch daran, dass die Potenziale bereits verfügbarer Methoden und Technologien oftmals nicht ausgeschöpft werden. Dies etwa, weil es an Ressourcen und dafür notwendiger Infrastruktur fehlt. Ein Weg vorwärts kann daher sein zu prüfen, welche klassischen Methoden bereits vorliegen und – bevor neu in KI investiert wird – diese erst einmal auszuschöpfen.

Die Frage, ob man zunächst das Potenzial bereits erprobter Technologien voll ausschöpfen sollte, wird verschärft, wenn man die Rolle zeitlicher Faktoren in der nuklearen Entsorgung beachtet. Generell gilt zwar, dass die Arbeit des BASE auf dem neuesten Stand von Wissenschaft und Technik fußen soll. Aber wissenschaftliche und technische Fortschritte entwickeln sich fortlaufend und man kann nicht verlässlich voraussagen, wann bestimmte Entwicklungen tatsächlich anwendungsfähig werden, insbesondere in einem risikobehafteten Bereich wie der nuklearen Entsorgung. Die Arbeit des BASE fordert aber Entscheidungen, die zu einem bestimmten Zeitpunkt getroffen werden müssen, auch wenn mögliche, zukünftige Entwicklungen diese Entscheidungen unter Umständen beeinflussen oder sogar verbessern würden. Nicht zuletzt ist die Endlagersuche auch an einen gesetzlichen Auftrag gebunden, der bestimmte Zeithorizonte beinhaltet. Kurz gesagt stellt sich die Frage: Lohnt es sich überhaupt in KI zu investieren, wenn man auf die Erträge dieser Investitionen lange warten muss? Und wer definiert, wie lange auf eine solche technische Entwicklung gewartet werden soll, gerade wenn der Zeitpunkt der Anwendbarkeit nicht mit Sicherheit bestimmt werden kann? Können staatliche Institutionen, die Steuergelder beanspruchen, um Lösungen für gesellschaftliche Probleme zu finden, das Risiko solcher Investitionen eingehen?

Es geht dabei nicht nur um die Frage, ob KI rechtzeitig weit genug fortgeschritten ist, um in der nuklearen Entsorgung von Nutzen zu sein. Jede KI-Anwendung ist immer nur so gut, wie die Datengrundlage auf der sie basiert. Und hier steckt der Teufel im Detail. Denn für jede KI-Anwendung braucht es die erforderlichen Daten. Das heißt, dass etwa in der Forschung Datenerhebung, Datenpflege und KI Hand in Hand gehen müssen. Die Daten, die gerade in der nuklearen Entsorgung benötigt werden, sind aber oftmals unvollständig und nicht miteinander vergleichbar, weil sie in der Vergangenheit nicht systematisch

erfasst und hinterlegt wurden. Insofern stellt sich die Frage, haben wir aktuell überhaupt die erforderlichen Daten, um KI-Systeme zu trainieren? Und wenn ja, wer verfügt darüber? Jenseits dessen können Daten sowohl qualitative und quantitative Mängel aufweisen, die eine erfolgreiche Anwendung von KI verhindern. Ein einfaches Beispiel, um einen solchen Mangel zu veranschaulichen, kommt aus der Medizin: Wenn man eine KI-Methode anhand von Fotos darauf trainiert, bösartige Hautveränderungen zu erkennen, diese Fotos aber nur einen Hauttyp abbilden, riskiert man, dass die KI-Methode versagt, sobald sie mit Fotos anderer Hauttypen konfrontiert ist. In anderen Anwendungsfällen mag es noch schwerer sein, einschätzen zu können, ob die Daten einen solchen Mangel aufweisen. Gerade im Bereich der nuklearen Entsorgung, wo Sicherheit das oberste Primat ist, braucht es verlässliche Daten. Allerdings kann man derzeit die Belastbarkeit bzw. Fehleranfälligkeit von einzelnen KI-Methoden nur schwer bewerten. Welches Risiko man womöglich mit der Anwendung von KI eingeht, ist demnach nicht direkt zu beurteilen.

Abgesehen von den technischen Herausforderungen stellen sich bei dem Einsatz von KI auch ethische und rechtliche Fragen. Wie bereits erwähnt, ist es schwer effektive Regulierung zu etablieren für einen Anwendungsbereich, der keine klaren Grenzen aufweist. Reguliert man beispielsweise die Nutzung bestimmter Chemikalien, lässt sich feststellen, ob ein Produkt diese Chemikalien enthält. Wenn man sich nun aber fragt, ob es sich bei einer Anwendung um KI handelt, ist diese Frage nicht unbedingt direkt zu beantworten. Außerdem ist es fraglich, ob allgemeine Regeln für KI an sich passend sind, wenn sich hinter dem Begriff KI doch so viele verschiedene Methoden verstecken. Der Ablauf eines großen Infrastrukturvorhabens wie der Endlagersuche ist gesetzlich geregelt. Rechtssicherheit ist hier unabdingbar, um den gesetzmäßigen Ablauf zu gewährleisten und notfalls einzuklagen. Für das BASE als Aufsichts- und Genehmigungsbehörde im Verfahren der Endlagersuche wäre etwa zu klären, welcher Regulierungsbedarf besteht, wenn der Vorhabenträger KI anwendet.

Insgesamt wird deutlich, dass es beim Thema KI im Feld der nuklearen Entsorgung noch viele offene Fragen gibt. Es braucht dafür einen transdisziplinären Austausch, der Akteur:innen aus Forschung und Praxis zusammenbringt und sich mit technischen, rechtlichen und ethischen Aspekten beschäftigt. Mit dem transdisziplinären Panel ‚Quo Vadis Künstliche Intelligenz in der nuklearen Entsorgung?‘ hat das BASE diesen Dialog aufgenommen und verschiedene Perspektiven zusammengebracht. Dies kann jedoch nur ein erster Schritt sein, wenn KI im Bereich der nuklearen Entsorgung und insbesondere im Prozess der Endlagersuche gewinnbringend zum Einsatz kommen soll.

