Federal Office
for the Safety of
Nuclear Waste Management

Quo Vadis Artificial Intelligence in Nuclear Waste Management?

Imprint

Bundesamt für die Sicherheit der nuklearen Entsorgung
(BASE)

Wegelystraße 8
10623 Berlin

Phone: +49 30 184321 0
E-mail: sowi@base.bund.de
www.base.bund.de

Stay informed about BASE:
www.base.bund.de/newsletter
www.twitter.com/BASE_bund
www.instagram.com/_base_bund/

Design: quermedia GmbH, Kassel
Illustrations: BASE and named sources
Cover: iStock.com/OlgaSalt

urn:nbn:de:0221-2024021441667

Correct in May 2023

Quo Vadis Artificial Intelligence in Nuclear Waste Management? A Collection of Transdisciplinary Perspectives published by the Federal Office for the Safety of Nuclear Waste Management

This publication was produced in the context of a workshop at BASE. The present text is a translation of the German original.

Introduction

Ina Richter and Dr Annika Froese
Federal Office for the Safety of Nuclear Waste Management (BASE)

In recent years, few technologies have experienced a hype like artificial intelligence (AI)–and not just since OpenAI published the latest version of its chatbot ChatGPT in spring 2023. AI is everywhere: autonomous driving promises to support or even replace the human behind the wheel. AI applications can translate texts and recognise images. AI is supposed to diagnose diseases, detect credit fraud, and make buildings more energy efficient. Given such diverse fields of application, one might ask: What can AI contribute in the field of nuclear waste management? The present publication is dedicated to this topic.

But what exactly is AI, given that it has so many different faces? For some, the term AI conjures up futuristic scenarios in which intelligent machines embody an alternative form of life, and are barely distinguishable from humans. Such machines can act autonomously, and may well possess sentience and consciousness. However, this so-called 'strong AI' is currently only a theoretical concept which raises many as yet unsolved questions, and its feasibility is a matter of debate. By contrast, the above-mentioned examples of AI applications belong to what is known as 'weak AI'.[1] This is where AI is used to carry out specific tasks previously done by humans, such as word processing or driving a car. Weak AI includes a wide range of methods and approaches, among which machine learning is currently the dominant paradigm.

However, the variety in how AI is defined also permits a categorization on the basis of other characteristics. For example, one common approach[2] distinguishes between different definitions based on, first, whether the definition refers to thought processes or behaviour; and second, the success criteria against which AI is measured. Some definitions consider AI successful if the result of the AI application mirrors the corresponding human performance, while other definitions require successful AI applications to deliver an ideal performance that is often superior to human performance. For instance, do we want an AI application to perform as well in diagnostic imaging as the average radiologist, or do we want the AI application to surpass human performance?

Yet even this more nuanced approach does not exhaust all possibilities for categorisation. Apart from assessing the end result, some definitions also differ in terms of the role they ascribe to the process by which an AI application produces its results. Some definitions articulate the expectation that AI should replicate human thought patterns; in other definitions, the method by which an AI application produces its results does not matter. For example, should an AI application categorise images using the same criteria that humans use for image recognition, or is the way the AI application obtains its result irrelevant? This brief overview alone shows how difficult it is to agree on a common understanding of the term AI.

In addition, the term AI also encompasses a variety of methods. A good example is machine learning, where a system initially completes a learning phase to generate new knowledge.[3] During the learning phase, the system is provided with training data, from which it infers patterns or regularities. When confronted with new data, the system can then evaluate this data on the basis of the patterns it has learned. For example, a video surveillance system can be taught to recognise 'suspicious' behaviour by training it using image material that humans have classified as suspicious.[4]

The latter example immediately raises important and equally complex questions about AI applications in general: What assumptions and values underlie the selection of data and the specification of the AI system's goal? To what extent do developers' biases influence an AI application? Is the available data quantitatively and qualitatively sufficient for a successful AI application? How likely is it that the AI system will nonetheless make mistakes? What are the expected consequences of such mistakes? Who is liable if damage is caused? Is the use of AI ethically justifiable? Are there tasks that, in principle, should remain in the hands of humans, and should not be handed over to an AI application? Questions like these underline the need to study AI developments and to critically examine the opportunities and risks of its use.

The Federal Office for the Safety of Nuclear Waste Management (BASE) is also concerned with the topic of AI. As the central federal authority for the safe handling of the legacy of nuclear energy, it performs regulatory, licensing and supervisory tasks in the areas of final storage, interim storage and the handling and transport of high-level radioactive waste. To perform these tasks in the best possible way, it is important to keep abreast of developments in science and technology. The potential and risks of AI in nuclear waste management have not yet been adequately addressed.

A consortium of scientists from various disciplines published an initial scientific review in November 2021.[5] This 'Roadmap for the Development and Adaptation of Artificial Intelligence (AI) Methods for Repository Research' provides the first overview of AI applications deemed potentially useful for final disposal. Most notably, the authors suggest potential applications for AI in the visualisation of workflows, in dealing with uncertainties and self-learning systems, as well as in knowledge management. Beyond this, there are only a few research collaborations and research programmes currently exploring AI in the context of nuclear waste management.[6]

BASE responded to developments in the field of AI and final disposal by organising a transdisciplinary panel entitled 'Quo Vadis Artificial Intelligence in Nuclear Waste Management' which took place on 1 December 2022. The expert discussion addressed the potential contributions of AI to nuclear waste management. Experts from various scientific disciplines and social sectors came together to report on the current status of AI developments in nuclear waste management. Relevant experiences and findings from other fields of application were also presented. The discussion also covered the legal and ethical requirements that must be considered when using AI.

With this event, BASE brought together scientific expertise from political science, geography, computer science, computational modelling and jurisprudence on the one hand, and practical experience such as that provided by the German Environment Agency (UBA) and the Federal Company for Radioactive Waste Disposal on the other. The idea was to confront theoretical potentials with practical conditions. This publication gives an overview of the contents and positions that some of the experts presented for discussion at the panel, and provides a concise summary of the discussions. The aim is to give an insight into the key discussion points and to make them accessible to the public. The contributions collected in this publication can be divided into three thematic areas.

The first three contributions address the current state of AI development in the field of final disposal. Judith Krohn's contribution 'Potentials of AI in dealing with geodata' gives an overview of the fields where AI methods analysing geodata have already been researched. According to the author, AI-supported data management and the evaluation of images and data spaces both offer potential for the search for a repository site. However, data often contain uncertainties, and can be subject to bias. Therefore, as far as the search for a repository site is concerned, AI applications based on such data should be used to support decisions, as a control authority and as a supplement to conventional methods. Human data analysis will not be replaced by AI in this case, but rather supported by it. Manfred Krafczyk first explains how AI methods support the development and application of computational models and simulations, demonstrating how they can potentially also be useful in the field of nuclear waste disposal. In his contribution 'Notes on prerequisites for a successful adaptation of AI models for nuclear waste management' he also identifies the challenges associated with the use of AI methods, which are due to technical aspects and the quantity and quality of the available training data, among other things. Vinzenz Brendler discusses the added value of a specific AI method - the digital twin (DT) - for nuclear waste management. In his contribution 'Digital twin for a deep geological repository: AI methods for reactive transport', he introduces an early project funded by the European Union that explores interactive visualisation.

The fourth contribution focuses on experiences with AI in a field that is particularly relevant to nuclear waste management: public participation. In her article 'Supporting political decisions through AI-supported evaluation of citizen participation processes', Julia Romberg explains how AI methods can support citizen participation processes.

The author sees potential in the AI-supported evaluation of texts, such as the input of citizens in participation processes.

The fifth article addresses the larger framework conditions that must be considered when using AI. In their article ‘Between Efficiency and Loss of Control: The Use of Artificial Intelligence by the Public Administration in the Mirror of the Law’, Stephanie Schiedermair and Johannes Weil provide a legal perspective on the tensions arising from the use of AI in public administrations. The authors see a need for further clarification before AI technologies can find widespread application in state institutions.

The publication concludes with a final comment that reflects the perspective of BASE. It argues that AI methods certainly hold potential for the management of radioactive waste. However, there are also a number of limitations, especially with regard to the quality and availability of data. Furthermore, there is uncertainty as to how robust the results of AI methods are, and whether AI applications should not rather function as ‘control instances’ that reduce errors and uncertainties. In addition, actors in the search for a repository site face questions of their own: What would happen, for example, if the Federal Company for Radioactive Waste Disposal (BGE mbH) actually used AI in its search for the most suitable site for a final repository? As a supervisory and licensing authority in the repository site selection procedure, BASE would have to determine whether there is a need for regulation in this case.

1 For an accessible discussion of the difference between ‚strong‘ and ‚weak‘ AI, see 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 For methods where AI learns from examples, see chapters 18, 19, and 20 in Russell, S; Norvig, P. (2012): Künstliche Intelligenz Ein moderner Ansatz, S. 22f. Pearson, München.

4 See, for example, the ‘Intelligente Videoüberwachung’ project of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, in partnership with the Mannheim Police Headquarters, the City of Mannheim and the Ministry of the Interior, Digitalisation and Migration of Baden-Württemberg, which is testing the use of algorithm-based video surveillance to combat street crime in public spaces. (https://www.iosb.fraunhofer.de/de/projekte-produkte/intelligente-videoueberwachung.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 See, among others, the BASE-commissioned project “ 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); the IGD-TP3, a research cooperation of European project sponsors to support the implementation of national disposal programmes for deep geological disposal of radioactive waste (igdtp.eu); and the Joint European Union Research Programme on Radioactive Waste Disposal and Management (EURAD4) (ejp-eurad.eu).

The potential of AI in dealing with geodata

Judith Krohn
Öko-Institut e.V. Darmstadt

Short presentation of the AKI project - methodological procedure

Based on a comprehensive international literature review, the project entitled "Application of Artificial Intelligence (AI) for the Site Selection of Deep Geological Repositories (AKI)"[1] initially identified applications of artificial intelligence in the geosciences in general, which were then evaluated with regard to their potential use for geoscientific issues in the German site selection procedure (StandAV). An evaluation scheme developed for this purpose will allow an interdisciplinary team of experts to classify the quality, suitability and relevance of the AI application described in the literature for solving a geoscientific problem.

This evaluation was carried out in two steps. Step one assessed the extent to which strengths and weaknesses of AI quite generally apply to the respective AI applications in particular. The second step involved a detailed assessment of the AI applications described in the literature, based on evaluation questions with positive and negative patterns that allow a clear statement regarding the size, availability and quality of the data basis used, as well as the level of technological maturity, the comprehensibility and the specific relevance for the StandAV site selection procedure. This evaluation scheme was applied to case studies found in the literature that, based on an initial assessment, may be relevant for the StandAV procedure. There was an ongoing interdisciplinary exchange to check the plausibility of the results. Aggregation of the findings provided an initial overview of the opportunities and risks of a potential use of AI in the StandAV, and identified areas for further research.

Site selection procedure sets the framework for potential AI use

The framework of the science-based StandAV procedure stipulates that all decisions and conclusions must be certain, data-based and comprehensible. Furthermore, potential geological developments over a detection period of one million years must be analysed to ensure the best possible long-term safety of the repository. Such prognoses of spatial and temporal changes in the geological conditions are based on a comprehensive and reliable data basis. Predictions require complex calculations and modelling in multidimensional space on different scales and including various geochemical and geophysical interactions. The description and evaluation of uncertainties also play a key role. Furthermore, the comprehensibility of the respective AI methods must meet high standards. AI methods that do not meet the StandAV's transparency standards carry considerable risks of jeopardising the public's trust in the site selection process.

The literature on the use of AI in the geosciences

There is a large and growing body of literature in the geosciences that deals with the use of AI. Among other things, the sources deal with the recognition, segmentation, generation and processing of data in digital images (computer vision), the classification or clustering of data. In addition, the literature presents applications where AI methods are used to develop surrogate models, create predictions, forecasts or prognoses, simplify complex relationships (dimension reduction), search for better solutions in complex solution spaces (optimisation) or recognise and identify outliers and unusual patterns (anomaly detection).

While extensive research in the field of AI in the geosciences has already been carried out, it has very rarely been applied to real-life sampling campaigns. Yet it is only with a concrete direct application in geology that experience with AI in this field of application can be evaluated in a valid and future-oriented way and with a focus on the StandAV procedure. The suitability of an AI application for a given problem must be examined in particular, and the method must be evaluated and weighed with regard to the additional benefits of AI compared to conventional methods. The potentials and risks must also be compared. Subsequently, with regard to transferability for applications in the StandAV procedure, it needs to be checked which adaptations are necessary and the risks of these adaptations must also be assessed.

AI in the context of the site selection procedure

An evaluation of the opportunities and risks of AI applications shows the core strengths of AI to include data management and the evaluation of images and high-dimensional data spaces. In the field of geospatial data processing, AI is gaining importance both for the analysis of large amounts of data as well as for the interpretation of imprecise data. Another major benefit of AI applications in geoscientific categories is the more accurate mapping of time-consuming numerical calculations or the optimisation of geostatistical analysis linked to data processing. Many AI applications also have the potential to achieve a better understanding of the laws and interrelationships of real processes and thus increase the objectivity of assessments. However, the latter is at odds with the concurrent risk of consciously or unconsciously influencing the decisions of AI processes due to the equally frequent danger of data or developer bias. Nonetheless, the prospect of being able to map previously undetectable processes by means of AI should at least be examined more closely in a process that is geared towards the goal of a "repository site with the best possible safety for at least one million years".

The prerequisites for viable solutions, however, are a sufficiently high-quality database and the selection of AI applications that have been extensively validated in view of their quality and suitability for the specific geoscientific problem. A major disadvantage of AI is its frequent lack of transparency. On the one hand, this harbours the danger of uncertainties in calculations being concealed over long periods of time or even accumulating into error chains that cannot be detected. On the other hand, a negative public bias, which often results from a lack of transparency in the procedures, is dangerous for the entire site selection process. Our analysis shows that even with new AI applications of so-called "explainable artificial intelligence" (XAI), it may not be possible for all fields of AI application to ensure that the methods used and the results achieved can be presented in a comprehensible manner. In addition, it must be pointed out that the methods of XAI themselves are still in their infancy.

The use of AI applications in the StandAV procedure is therefore only suitable for supporting decisions, supplementing conventional procedures or acting as a checking tool for detecting errors and evaluating uncertainties.

1 Krohn, J. et al. (2022): Anwendung der künstlichen Intelligenz (KI) für die Standortauswahl von tiefen geologischen Endlagern (AKI). Project number 4721E03210, commissioned by the Federal Office for the Safety of nuclear Waste Management (BASE). Available at https://www.oeko.de/publikationen/p-details/anwendung-der-kuenstlichen-intelligenz-ki-fuer-die-standortauswahl-von-tiefen-geologischen-endlagern-aki

Remarks on the prerequisites for a successful adaptation of AI models for nuclear waste management

Prof. Dr Manfred Krafczyk
Institute for Computational Modelling in Civil Engineering, TU Braunschweig

Over the past decades, computational engineering has emerged as an interdisciplinary paradigm, which deals with the development and application of computational models and simulations, often using high performance computers, to solve complex multi-field problems encountered in the technical analysis and design of engineering problems. In this context, computational simulation offers the possibility of exploring areas that are either inaccessible by means of traditional experiments or where it is too costly to conduct physical investigations. The development of a specific simulation model or process typically begins with a system identification that thinks of a system as a conglomerate of individual entities and their interactions. Most engineering systems in the field of nuclear waste management are characterised by the description of systems as continua across so-called fields that interact in space and time. The mathematical description has so far mostly taken the form of partial differential equations that must be solved (mostly approximatively) under the assumption of suitable boundary and initial conditions. Since, in addition to the number of dimensions in space and time, other parameters are often varied in the sense of a problem dimension, one is often confronted with the effects of the so-called curse-of-dimensions, which leads to unacceptable computation times for many problems relevant to practice with regard to finding a sufficiently accurate approximate solution for the model equations. Therefore, there is still a need for innovative methods for the (approximate) solution of partial differential equations or alternative modelling approaches, such as those that have become very popular in recent years through different variants of so-called machine learning or artificial intelligence.

Machine learning as the science of computer-aided methods that "independently" expand their problem-solving capacity by successively processing larger amounts of data and especially so-called neural networks (NN) have been an impressive success story in research and technology over the last decade, as they have significantly expanded our ability to derive structures and knowledge from data.

From a mathematical point of view, an NN is usually a mapping between two spaces (for example, real numbers) of potentially different dimensions. The universal approximation theorem states that, based on certain general assumptions, a feedforward neural network (FNN) can asymptotically approximate any function with an arbitrarily small error. In contrast to conventional modelling approaches, however, NNs must be "configured" with suitable training data to determine the desired mapping prior to actual use. Apart from various technical aspects, the performance of the NN, therefore, depends on the number of its degrees of freedom (number and size of its layers) and the quantity and quality of the training data available for its conditioning (possibly also on the person training the NN). Assuming that the training data characterises the sub-system state of one or more potential repository sites, properties of a (different) site can be extrapolated from it with little computational effort. Yet, the resulting uncertainties regarding the quality of the solution depend on many factors and, in contrast to modelling approaches based on differential equations, are more difficult to separate, since the training data in particular contains potentially implicit information whose influence on the solutions to be computed may be difficult to identify. Even though models based on differential equations contain uncertainties, especially with regard to initial, boundary and material laws, the solution behaviour for a defined setup is asymptotically convergent, which is not necessarily true for data-driven models.
It should also be noted that data-driven models can be deliberately "misled" in certain circumstances[1], by subtly disturbing the input signal, although this disturbance would not necessarily affect the human characterisation.

Data-driven approaches to solving differential equations, known as Physics Informed Neural Networks (PINNs), have been available for many engineering problems (so far outside the domain of nuclear waste management) for several years now. The sample solutions presented in the recent literature suggest that the solution of partial differential equations with PINNs can be very efficient, even if essential aspects such as asymptotic convergence and stability have so far only been investigated heuristically. The validity of such results outside the training domain is the subject of ongoing research.[2] Another unresolved issue is how the accuracy of a PINN-based approximate solution depends on the number and size of the hidden layers (and thus also on the associated training and computational effort). Traditional application variants such as spatio-temporal adaptation based on a priori error estimators of sample points could also be transferred to PINNs and contribute to a significant increase in efficiency. While conventional numerical methods will just fail if stability limits are exceeded, such an effect will not occur with PINNs.

If PINNs are used without due consideration, this can lead to the solutions achieved being of inadequate quality. However, apart from solving partial differential equations in the sense of a forward problem, data-driven approaches have already been successfully adapted to solve inverse non-linear problems. This includes, for example, the reconstruction of physical parameters in differential equations from measurement data or the reconstruction of physical fields from complementary partial data. This class of problems is only (if at all) accessible for classical numerical processes with significantly higher effort. Irrespective of whether PINNs will yield reliable efficiency and accuracy advantages in solving forward problems in the medium term, it is already clear that they will open up new avenues in computational engineering, especially in the area of inverse problems. The future investigation of complex problems in nuclear waste management is likely to benefit, in particular, from approaches that use a combination of conventional modelling approaches, data-driven modelling and uncertainty analyses. Nevertheless, to ensure the reproducibility of the results, special care must be taken when collecting, processing and persisting the corresponding training data. The article “A roadmap for the development and adaptation of artificial intelligence (AI) methods for repository research”[3] contains further general recommendations on systematic research approaches.

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

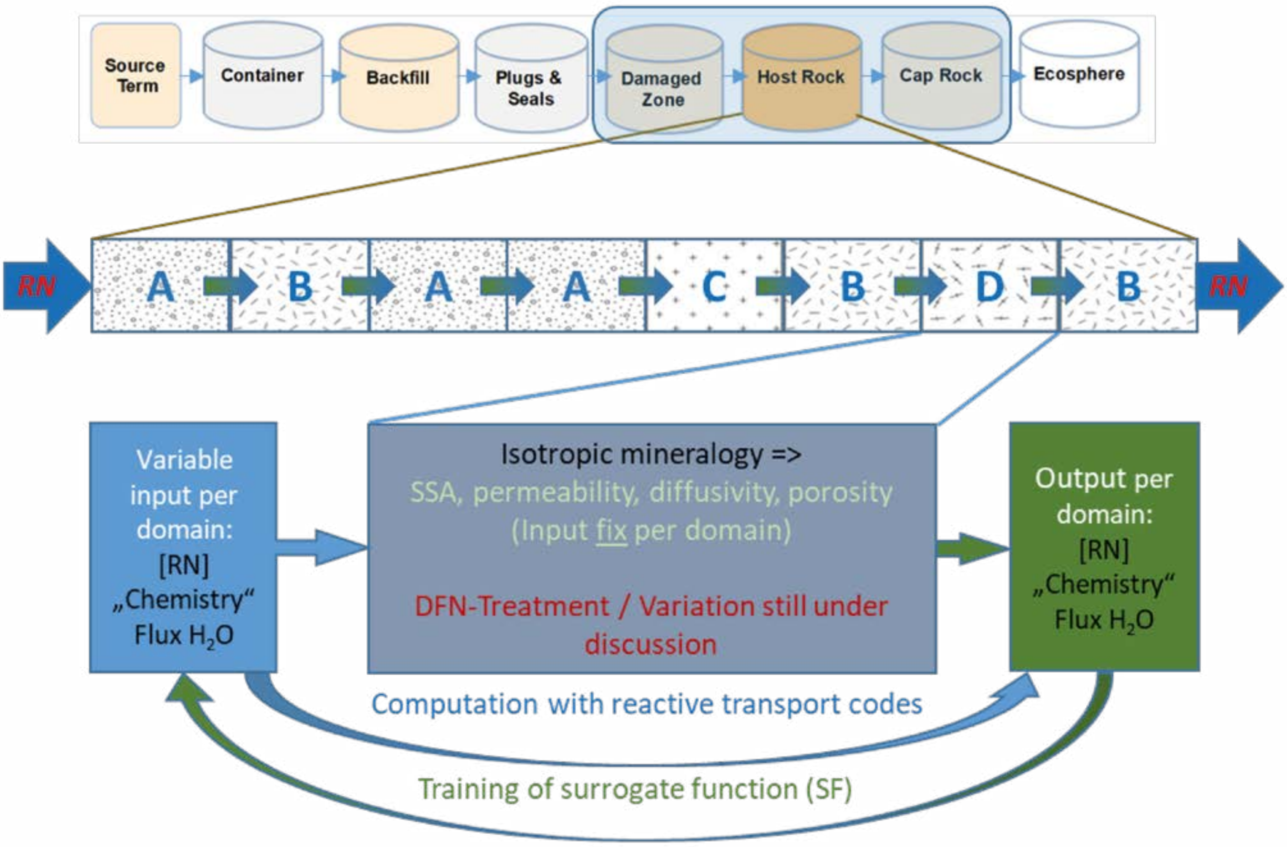
3 Krafczyk, Manfred, Brendler, Vinzenz, Czaikowski, Oliver, Gruner, Matthias, Hoth, Niels, Kolditz, Olaf, Nagel, Thomas, Herold, Philipp, Müller, Christian, Seher, Holger, Simo, Eric, & Stahlmann, Joachim. (2021). Eine Roadmap zur Entwicklung und Adaption von Methoden der Künstlichen Intelligenz (KI)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

Digital twin for a deep geological repository: AI methods for reactive transport

Prof. Dr Vinzenz Brendler
Department for Actinide Thermodynamics, Helmholtz-Zentrum Dresden-Rossendorf e.V.

There is an obvious need for digital twins (DT) for a nuclear waste repository (NWR), as highlighted e.g. by the EURAD community /PRA 20/[1] and in Germany /KRA 21/[2]. Such a holistic approach can check for unexpected cross-effects. It facilitates updating both input data and model specification (as is to be expected during the site selection process), promotes transparency of all steps, encourages efficient use of resources, enables acceleration of computations, and provides the basis for global sensitivity and uncertainty analyses.

There are several challenges, however: the complexity of the task, the work on very different scales, the ubiquity of surrogate data matrices, high computational costs, the need for truly interdisciplinary collaboration, a lack of standardisation, sustainability regulations regarding data, codes and modelling approaches. Last but not least, all stakeholders, and especially the public, hesitate to rely on black-box tools. That is why the EU's approach to artificial intelligence centers on excellence and trust, aiming to boost research and industrial capacity while ensuring safety and fundamental rights: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.



„Chemistry" = pH, E_H, Ionic strength, K_d, [important elements], p_{CO2}, ... ; for each domain respective specific mean values and variances are given. K_d values originate from other surrogate functions

An early project (VIRTUS - Virtual Underground Laboratory in Rock Salt /GRS 14/[3]) focused on interactive visualisation. First DT approaches at EURATOM level - “Pre-disposal management of radioactive waste (PREDIS)” - concern interim storage starting in 2021. To make further progress in establishing an NWR digital twin, a recent proposal focuses on the compartments of an NWR where geological conditions determine radionuclide migration, i.e. the fields inside the blue frame in the figure below. Here, a bottom-up approach is preferred over a top-down strategy. For this purpose, each of the compartments is divided into a sequence of geological domains. In a first stage, the retardation expressed in distribution coefficients is made accessible within each such domain by training surrogate functions via the smart-Kd approach /STO 17/[4]. Possible surrogates include decision trees (Decision Trees: /DEL 21/)[5], algorithmic approaches such as adapted Taylor series, e.g. /LEA 20/[6] or multivariate adaptive regression splines /KEA 16/[7], Kernel Ridge Regression /HAS 09/[8] or Gaussian Process Regression /RAS 06/[9], but also Reduced Basis Methods /CHE 22/[10] or Physics-Informed Neural Networks (PINN: /HOR 91/[11], /LAL 21/[12]).

In a second stage (see figure), these retardation parameters are then combined with all the geological information to model the reactive transport of radionuclides. ML methods are particularly needed for modelling physical systems by means of non-linear, ordinary and partial coupled differential equations (ODEs and PDEs). PINNS are also adaptable for such problems /RAI 19/[13]. The surrogate models induced by PINNs must have a high approximation quality compared to the ground truth model (i.e. to the reference solution), and it must be possible to generate them with reasonable training effort. Both challenges can be guaranteed for a large number of problems /HEC 20/[14], which come into play in the framework presented here.

1 /PRA 20/ Prasianakis, N. (2020) „Towards digital twins: machine learning based process coupling and multiscale modelling of reactive transport phenomena“, Goldschmidt Conference 2020, http://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.

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.

Supporting political decisions through AI-assisted evaluation of citizen participation processes

*Julia Romberg**
Heinrich-Heine-Universität Düsseldorf

Political participation describes the voluntary actions of citizens, taken either alone or with others, to influence political decisions.[1] One important form of political participation is public participation, which public authorities employ to consult the population on specific political issues.[2] If participation is primarily aimed at citizens who do not act within the framework of organised interest groups and associations, it is also referred to, more specifically, as citizen participation.

Citizen participation processes usually pursue two goals.[3] On the one hand, the inclusion of collective knowledge aims to facilitate an informed decision-making process so that effective policies can be derived. On the other hand, it is expected that the opportunity to contribute knowledge, voice concerns and (to a certain extent) help shape the final policy decisions will lead to greater public acceptance and, ideally, also to greater satisfaction with the decisions taken. The ultimate aim is to strengthen the legitimacy of policy measures.

Participation can be implemented at all administrative levels and in various formats. Citizen input is often obtained in textual form, for example by using online platforms or questionnaires with free-text fields. The subsequent evaluation of these textual contributions is crucial to identify information relevant for action, and to use it as a basis for the further political process.

Public administrations or contracted service providers usually carry out the evaluation of the contributions manually. The individual steps of the evaluation include, among other things, a repeated reading of the contributions, the detection of duplicates, the pre-sorting of the contributions according to topics and administrative units, as well as an in-depth analysis of individual points of view.[4]

To meet democratic norms, it is important to ensure a fair and transparent decision-making process in which all contributions received are examined with equal care.[5] The authorities' approach to evaluating the contributions directly affects how the legitimacy of decisions is perceived:[6] If, in the public eye, the aforementioned criteria are not met, the resulting policy may be seen as less legitimate.[7] Meeting these requirements requires a high commitment in terms of time, personnel, as well as financial resources, that - in the worst case - can become a seemingly insurmountable hurdle for the successful implementation of public participation.[8]

One solution is the use of artificial intelligence (AI). The idea is to fully or partially automate various subtasks, such as the detection of duplicates,[9] the identification of topics and corresponding grouping of contributions,[10] the evaluation of the public sentiment[11] and of arguments[12] as well as the summary of contents.[13]

The BMBF-funded research group CIMT[14] (Citizen Involvement in Mobility Transitions) focuses on the development of AI techniques that facilitate the evaluation of citizen contributions. In the context of transport planning procedures, relevant subtasks were identified in interviews with practitioners[15], and AI-based solutions were developed. A selection of these methods is presented below.

The need for AI support in thematic categorisation is omnipresent. For this purpose, supervised machine learning algorithms can be trained, with the help of manually coded data, to classify input according to predefined topic clusters. To keep the remaining manual effort as low as possible, in Romberg and Escher[16], we developed specific methods for evaluating public participation. The proposed algorithm can assign the correct topics in eight out of ten cases, while at the same time reducing the manual categorisation effort by up to one fifth. Using the practical application case of three municipal participation procedures, this means that only 120 of 459, 180 of 366, and 500 of 2314 contributions of the respective procedure have to be categorised manually.

In addition to thematic structuring, the individual opinions of the citizens are of interest for the

* The project on which this report is based was funded by the Federal Ministry of Education and Research in the funding priority Social-Ecological Research (funding code 01UU1904). The responsibility for the content of this publication lies with the author. Contact: Julia.Romberg@uni-duesseldorf.de.

evaluation. An automated identification of arguments can accelerate the evaluation, especially in the case of long texts, as text passages that are not relevant for this aspect of the analysis can be put aside. In Romberg and Conrad[17], we therefore presented methods that help identify argumentative text passages in citizen contributions. These achieve an accuracy of 77 percent. Furthermore, the argumentative passages can be divided according to whether they describe a current state (for example, the bad condition of a street) or propose a measure (for example, a solution to remedy the grievance). AI is able to make the correct distinction in nine out of ten cases.

The question of how concrete conditions and propositions are described can also be relevant for deriving political measures from the collected contributions. The clearer the description, the easier the evaluation. In the case of vague ideas or unclear wording, on the other hand, far more work is needed before measures can be developed, as the scope for interpretation is large and open questions have to be clarified. Simple AI models can already deliver a 79 percent accuracy in such cases.[18]

In summary, it can be said that methods for an AI-supported evaluation of text contributions are already showing promising results. Especially with regard to thematic pre-sorting, the time required for the evaluation can be drastically reduced. This is particularly beneficial for processes with a larger number of contributions. Nevertheless, important aspects of the practical applicability of AI in public participation still need to be clarified.

This includes questions about how reliable the applications need to be so that practitioners can use them with trust, and what application-specific user-friendly implementations should look like.[19] Similarly, support should not be limited to textual participation, but should include other formats such as oral input. Speech recognition methods that translate speech into text seem conceivable for this purpose. Finally, the potential of AI is not limited to the evaluation phase, but can also provide support during or before the participation phase, for example to simplify the moderation of the processes, or to improve the input.

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 Sarvghad, and Narges Mahyar. 2021. Communitypulse: Facilitating community input analysis by surfacing hidden insights, reflections, and priorities. In Designing Interactive Systems Conference 2021, DIS ‘21, pages 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, and Stephen Zavestoski. 2004. SGER collaborative. Journal of E-Government, 1(1):123-127. https://doi.org/10.1300/J399v01n01_09; Anthony Simonofski, Jerôme Fink, and 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, and 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. See footnote 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. Strebel, Daniel Kübler, and Frank Marcinkowski. The importance of input and output legitimacy in democratic governance: Evidence from a population-based survey experiment in four western European countries. European Journal of Political Research, 58(2):488-513, 2019.

8 Narges Mahyar, Diana V. Nguyen, Maggie Chan, Jiayi Zheng, and Steven P. Dow. The civic data deluge: Understanding the challenges of analysing large-scale community input. In Proceedings of the 2019 on Designing Interactive Systems Conference, DIS ‘19, pages 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 and 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, and 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 and Jamie Callan. 2009. Ontocop: Constructing ontologies for public comments. IEEE Intelligent Systems, 24(5):70-7.

11 Kwon et al. 2006. See footnote 10; Jasmin et al. 2021. See footnote 4.

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

13 Miguel Arana-Catania, Felix-Anselm V. Lier, Rob Procter, Nataliya Tkachenko, Yulan He, Arkaitz Zubiaga, and 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 and Tobias Escher. 2020. See footnote 4.

16 Julia Romberg and 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 and 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. See footnote 4.

Between Efficiency and Loss of Control: The Use of Artificial Intelligence by the Public Administration in the Mirror of the Law

Prof. Dr Stephanie Schiedermair and Johannes Weil*
Faculty of Law, Leipzig University, ScaDS.AI Dresden/Leipzig

Opportunities and challenges for law and public administration

The many forms and applications of artificial intelligence (AI) do not stop at public administration, given that they are associated with the promise of increasing the efficiency of administrative action while at the same time reducing costs and the workload of employees.[1]

Existing and potential applications result both from citizen contact (front office) – e.g. through the use of chatbots or assistance with applications – and the automation of internal processes within public authorities (back office). Systems that contribute to the preparation of externally effective decisions through comprehensive data evaluation, forecasting or verification, or are even intended to generate fully automated decisions, are being discussed in particular detail.[2] In the context of nuclear waste disposal, the application of AI for the control, analysis and evaluation of geodata in the site selection process is currently being investigated.[3]

However, the great promise of AI technologies is countered by various practical problems that pose great challenges to the law and thus to the strictly law-bound administration in particular: For one, the systems are dependent on considerable amounts of data.

Secondly, there are difficulties with the comprehensibility and explainability of the decision-making process ("black box" phenomenon), especially in the case of more complex applications in the area of "deep learning", meaning that adequate control and legal review are more difficult. In addition, AI systems often have a certain error rate or can even produce discriminatory results ("bias in AI").[4]

Even though the sovereign use of AI cannot be assessed in a general legal manner, and must instead be judged according to the respective field of application, the concrete facts of the case and the relevant regulations, the problems outlined are connected to overarching legal questions that typically relate to constitutional law (II) as well as data protection law (III) and anti-discrimination law (IV). The AI Regulation envisaged at EU level will play a decisive role in the future (V).

Constitutional framework

At a constitutional level, the principle of democracy and the rule of law require that there is sufficient state control over the AI system used.[5] It is not imperative that the responsible official fully understands its functioning, nor does the software have to work absolutely flawlessly. After all, administrative practice has so far been based on human decision-making, which is also hard to penetrate and subject to certain error rates. In this respect, suitable precautions must be taken to ensure that the desired results will generally be achieved. Possible remedies include prior testing or human oversight.[6]

Especially in the case of onerous decisions, the possibility of judicial review in the sense of effective legal protection in accordance with Article 19(4) of the German Basic Law (Grundgesetz, GG) must be guaranteed for the benefit of those affected. The requirement of justification under the rule of law is of central importance here. It is true that there may be a certain amount of tension in the case

---

* Stephanie Schiedermair holds the Chair of European Law, International Law and Public Law at the Faculty of Law at Leipzig University. Johannes Weil is her research assistant and doctoral student at ScaDS.AI Dresden/Leipzig. The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research „Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI.

of non-transparent systems. It should be noted, however, that it is not necessary to have information on all aspects of the decision, which also applies to automated decision-making. Nevertheless, the statement of reasons must be designed in such a way that those affected can effectively appeal against the decision.[7]

Depending on the case constellation, special requirements may also arise from fundamental rights.[8] As far as the right to informational self-determination is concerned, the Federal Constitutional Court recently clarified that the use of AI regularly constitutes a particularly serious interference and that procedural safeguards must be taken with regard to its specific dangers.[9]

In general, the level of legal requirements depends on various factors, namely the specific situation in which the system is used, the relevance of fundamental rights to the task assigned to the system, the controllability, the reversibility of the process, the potential for damage and the availability of alternative methods.[10] To this extent, stronger requirements will have to be imposed in the context of externally effective decisions, while the mere use of AI in the front or back office appears to be less problematic. The use of AI to support the selection of sites for nuclear waste disposal is likely to be subject to more stringent requirements, primarily due to the high relevance of fundamental rights.[11]

Tension under data protection law

The highly data-driven AI technologies inevitably come into conflict with data protection law. The provisions of the European General Data Protection Regulation (GDPR), which is central in this respect, nevertheless only apply to the processing of personal data, Art. 2(1) GDPR. If the search for a suitable repository site only involves the evaluation of geodata that do not allow conclusions to be drawn about natural persons, the regulatory regime of the GDPR does not need not be observed.

However, if AI-based data processing falls within the scope of the GDPR, there are tensions with regard to the principles of data protection, which are all shaped by specific regulations in the GDPR. First of all, personal data must be processed transparently, i.e. in a manner that is comprehensible to the data subject (Article 5(1)(a) GDPR). Conflicts can arise here, especially in the case of non-transparent, complex systems. As far as the obligation to provide information and the right to information under Article 13 et seq. of the GDPR are concerned, the extent to which information on the functioning of an AI system must be provided is disputed in detail.[12]

The reliance on large amounts of data, which is characteristic of AI systems, contrasts with the principle of data minimisation enshrined in Article 5(1)(c) GDPR. Compromises might be to pseudonymise or even anonymise the data used as far as possible.[13] The principle of data accuracy (Article 5(1)(d) GDPR) requires a corresponding level of data quality. This can also lead to problems with black box systems, as errors are often difficult to identify.[14]

“Bias in, Bias out” – Anti-discrimination law and AI

Another problem, which should be less relevant for the mere evaluation of geodata, results from the fact that AI applications may have a discriminatory effect. Possible causes for this are, for example, an incomplete data basis where a particular group is over- or under-represented, or training data that is based on discriminatory human decisions.[15]

Non-discrimination rules, which principally prohibit unequal treatment by state authorities on the grounds of protected characteristics such as ethnic origin, religion or gender, exist both at EU level and in constitutional law (e.g. Article 3(3) GG, Article 21(1) CFR). These are also applicable if the AI system used is not directly linked to such a category, but to so-called proxy characteristics, which typically correlate with them.[16]

Yet it is often difficult to prove discrimination, especially if the system is not transparent.[17] When using an AI system which, by its very nature, has the potential to be discriminatory, specific precautions must be taken to ensure that the quality of the training data is adequate.[18] Any discrimination is likely to be justifiable only in rare cases, at least with regard to the characteristics protected by Article 3(3) GG, as exceptionally serious reasons are required here; a mere increase in efficiency through AI does not suffice.

Specific legal developments at EU level

One specific answer to the problems typically associated with the use of AI is on the horizon at EU level. The Commission published a draft AI Regulation in April 2021 (AI Act), but the legislative process is still ongoing.[19] The draft follows a tiered, risk-based approach. AI applications considered too risky – such as state social scoring systems – are to be banned (Article 5 AI Act).

So-called high-risk systems, which are primarily characterised by their area of application (Article 6 and Annex III AI Act), including law enforcement and human resources management, are subject to a more stringent programme of obligations. This includes, in particular, the management and operation of critical infrastructures. Accordingly, systems that are intended to be used as safety components in power supply are to be qualified as high-risk systems. Due to the connection to nuclear energy and the high relevance of fundamental rights, this could also include AI systems related to nuclear waste disposal. As requirements for high-risk AI systems, Article 9 et seq. of the AI Regulation prescribe, among other things, a risk management system, quality assurance with regard to the relevant data sets, technical documentation, transparency and information requirements, the guarantee of effective human oversight and an appropriate level of accuracy, robustness and cyber security.

Irrespective of the above, specific transparency requirements are laid down for certain risk-prone applications such as chatbots, emotion recognition systems and deep fakes (Article 52 AI Act). On the other hand, systems that do not fall under the aforementioned categories, i.e. are not subject to an increased risk according to the conception of the draft, will remain unregulated.

Concluding remarks

From a legal perspective, many unresolved issues regarding the use of AI in administration remain. They need to be addressed before the state can really make widespread use of these technologies. The goal should be to create a robust framework that complies with the rule of law and fundamental rights, while leaving room for innovation to promote efficient and citizen-friendly administration. In addition to specific regulatory projects such as the European AI Regulation, this also calls for an interdisciplinary exchange that brings together technical, legal and ethical aspects.

1 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 See generally 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 (links accessed on 6 March 2023).

3 Öko-Institut e.V., Künstliche Intelligenz in der Endlagersuche: 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 On AI-related legal challenges, see generally Wischmeyer, Regulierung intelligenter Systeme, AöR 2018, 1.

5 See BVerfG, Judgement of 16.2.2023 – 1 BvR 1547/19, para. 100 in the context of the fundamental right to informational self-determination.

6 Busche, Staatliche Verantwortungsübernahme beim hoheitlichen Einsatz intransparenter Algorithmen, in: Ammann et al. (eds.), Verantwortung und Recht, 2022, p. 45 (55 et seq.).

7 Wischmeyer, in: Ebers/Heinze/Krügel/Steinrötter (eds.), Künstliche Intelligenz und Robotik, Rechtshandbuch, 2020, § 20 para. 56.

8 For an overview in relation to EU fundamental rights, see 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.

9 BVerfG, Judgement of 16.2.2023 – 1 BvR 1547/19, para. 100 et seq.

10 Unger, Demokratische Herrschaft und künstliche Intelligenz, in: Unger/von Ungern-Sternberg (eds.), Demokratie und künstliche Intelligenz, 2019, p. 113 (126); Busche, Staatliche Verantwortungsübernahme beim hoheitlichen Einsatz intransparenter Algorithmen, in: Ammann et al. (eds.), Verantwortung und Recht, 2022, p. 45 (58 fn. 56).

11 See generally Emanuel, Rechtliche Rahmenbedingungen für die Suche nach einem Endlager für hochradioaktiven Abfall (HAW), 2021, p. 115 et seq.

12 In detail Hoeren/Niehoff, KI und Datenschutz - Begründungserfordernisse automatisierter Entscheidungen, RW 2018, 47 (54 et seq.); Kumkar/Roth-Isigkeit, Erklärungspflichten bei automatisierten Datenverarbeitungen nach der DS-GVO, JZ 2020, 277; Hoffmann/Kevekordes, Das Right to Explanation, DuD 2021, 609.

13 Kloos/Schmidt-Bens, in: Hartmann (ed.), KI & Recht kompakt, 2020, p. 165 (183).

14 Paal, in: Kaulartz/Braegelmann (eds.), Rechtshandbuch Artificial Intelligence und Machine Learning, 2020, ch. 8.7 marginal no. 17.

15 Wischmeyer, in: Ebers/Heinze/Krügel/Steinrötter (eds.), Künstliche Intelligenz und Robotik, Rechtshandbuch, 2020, § 20 para. 58.

16 Tischbirek, Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems, in: Wischmeyer/Rademacher (eds.), Regulating Artificial Intelligence, 2020, p. 103 (para. 14 et seq.).

17 Borgesius, Strengthening legal protection against discrimination by algorithms and artificial intelligence, The International Journal of Human Rights 24 (2020), 1572 (1577).

18 See BVerfG, Judgement of 16.2.2023 – 1 BvR 1547/19, para. 77 and 100.

19 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (AI Regulation) and amending certain Union acts, COM(2021) 206 final; see most recently the provisional position of the Council of the EU on 6.12.2022, ST 15698 2022 INIT; in detail Ebert/Spiecker called Döhmann, Der Kommissionsentwurf für eine KI-Verordnung der EU, NVwZ 2021, 1188.

Quo Vadis Artificial Intelligence for Nuclear Waste Management? A Final Review

Dr Annika Froese and Ina Richter
Federal Office for the Safety of Nuclear Waste Management (BASE)

The experts at the BASE panel on 1 December 2022 spent a whole day discussing the implications of AI for nuclear waste management. At the end of the day, they were largely in agreement: AI does hold potential that can be useful in the field of nuclear waste management. However, the experience with AI in this domain is still too limited to accurately assess this potential. Promising developments can be seen, however, in the handling of geoscientific data. To ensure the safety of a repository for high-level radioactive waste, the geological developments that can be expected within the next one million years must be simulated and analysed. This can only be done on the basis of extensive and reliable data. AI could help to classify data, process digital images and simplify complex processes. Another promising application of AI is in the creation and analysis of a digital twin, which is a digital representation of the entire repository system.

Whether AI is used to generate new knowledge or to manage and verify existing knowledge, these applications show great potential in technical and scientific fields. But AI can also support public institutions in the area of citizen participation. Contributions to participation processes can nowadays often be submitted electronically. This results in extensive data sets consisting of text contributions from individual citizens, which usually have to be manually evaluated by staff. Such manual processing requires resources that are often unavailable in public institutions. This is where AI can provide support by categorising text contributions, assigning them to overarching clusters or topics, and eliminating duplications. This pre-processing can make later evaluations by staff more efficient and save resources. Public participation plays a central role in nuclear waste management, especially in the search for a repository site for high-level radioactive waste. This opens up further potential for AI applications.

Nevertheless, using AI in the field of nuclear waste management also poses several challenges. During the discussion, for example, it became clear that the use of the term "artificial intelligence" itself must be subject to further scrutiny. As mentioned in the introduction, this term encompasses a variety of different methods. Hence, there is the question of whether it serves clarity to speak of 'artificial intelligence' in general instead of naming a specific method. The fact that the term is used widely and is associated with different methods, evaluations and expectations in different circles confirms the need for a critical examination.

For one thing, the vagueness of the term AI complicates the regulation of the technologies that are subsumed under this heading. On the other hand, it is associated with different public expectations. It is not clear what the use of AI would trigger in those involved in participation processes in the context of nuclear waste management. Are the results of a participation process more likely to be accepted if a machine rather than a human evaluates the comments and remarks submitted? Do technologies lend more legitimacy to such processes because the public ascribes more 'objectivity' to them than to the responsible authorities? Such considerations also touch on fundamental - as yet little discussed - questions of democratic decision-making processes that could limit the use of AI to certain fields of application. One example is the question of who assumes responsibility for decisions and the associated consequences of actions that are accompanied or prepared by a machine. How can we ensure that developer bias, i.e. the opinions of those who train an AI, does not unilaterally influence political decisions?

In the attempt to define the term AI more clearly, we can also see that some of the so-called AI methods are not new or build on methods that have been available for a long time. For instance, there is no clear separation between AI approaches and traditional methods of numerical modelling that are relevant to the field of nuclear waste management. The ideal would be to use the best of both worlds. One of the reasons why this has not happened so far is that the potential of already available methods and technologies is often not fully exploited. This is due to a lack of resources and the necessary infrastructure. One way forward may therefore be to check which traditional methods are already available and to exploit them first, before investing in AI.

The question of whether to fully exploit the potential of already proven technologies first is exacerbated when one considers the role of time factors in nuclear waste management. As a general rule, the work of BASE should be firmly grounded in the state of the art in science and technology. However, scientific and technical knowledge is constantly advancing and it is not possible to reliably predict when certain developments will actually become applicable, especially in a high-risk area such as nuclear waste management. However, the work of BASE calls for decisions to be made at a certain point in time, even though possible, future developments might influence or even improve these decisions. Finally, the search for a final repository site is also tied to a legal mandate that includes certain time horizons. In short, this begs the question: is it worth investing in AI at all if one has to wait a long time for the returns on these investments? And who defines how long one can wait for such a technical development, especially when the time of applicability cannot be determined with certainty? Can state institutions that use taxpayers' money to find solutions for societal problems take the risk of such investments?

It is not just a question of whether AI will be sufficiently advanced in time to be useful in nuclear waste management. Any AI application is only as good as the data it is based on. And the devil is in the detail here. Every AI system needs data that corresponds to the system's main objective. When it comes to research, for example, this means that data collection, data maintenance and AI must go hand in hand. However, the data needed in nuclear waste management is often incomplete and not comparable because it was not systematically collected and stored in the past. In this respect, we need to ask ourselves whether we currently have the necessary data at all to train AI systems. And if so, who has it? Beyond that, data can have both qualitative and quantitative deficiencies that may prevent the successful application of AI. We can use a simple example from medicine to illustrate such a deficiency: if you train an AI system to recognise malignant skin lesions based on photos, but these photos only include one skin type, there is a risk that the AI system will fail as soon as it is confronted with other skin types. In other use cases, it may be even more difficult to assess whether the data shows such a deficiency. Reliable data is of particular importance in the field of nuclear waste management, where safety is the top priority. However, the resilience or susceptibility to error of individual AI methods is currently difficult to assess. Accordingly, assessing the potential risk of using AI is equally difficult.

Apart from the technical challenges, the use of AI also raises ethical and legal questions. As mentioned earlier, it is difficult to establish an effective regulation for applications that do not have clear boundaries. For example, if we regulate the use of certain chemicals, it is possible to determine whether a product contains these chemicals. However, if we were to ask whether an application was AI, we might not necessarily find a direct answer to this question. Moreover, it is unclear whether general rules for AI are appropriate when there are so many different methods covered by the term AI. The process of a large infrastructure project such as the search for a repository site is regulated by law. Legal certainty is indispensable here to ensure that the process is in accordance with the law and, if necessary, to initiate legal action. As the supervisory and licensing authority in the search for a final repository site, BASE would have to clarify, for example, whether there was need for regulation if the project implementer - the Bundesgesellschaft für Endlagerung (BGE) - were to use AI.

All in all, it is evident that many issues regarding the use of AI in the field of nuclear waste management remain unresolved. Addressing these challenges requires a transdisciplinary exchange that brings together actors from research and practice and deals with various technical, legal and ethical questions. With its transdisciplinary panel 'Quo Vadis Artificial Intelligence and Nuclear Waste Disposal?', BASE initiated such a dialogue and brought together different perspectives. Building on this first step will be crucial if AI is to be used profitably in the field of nuclear waste management, and especially in the process of finding a repository site.