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

[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 under https://www.oeko.de/publikationen/p-details/anwendung-der-kuenstlichen-intelligenz-ki-fuer-die-standortauswahl-von-tiefen-geologischen-endlagern-aki

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.