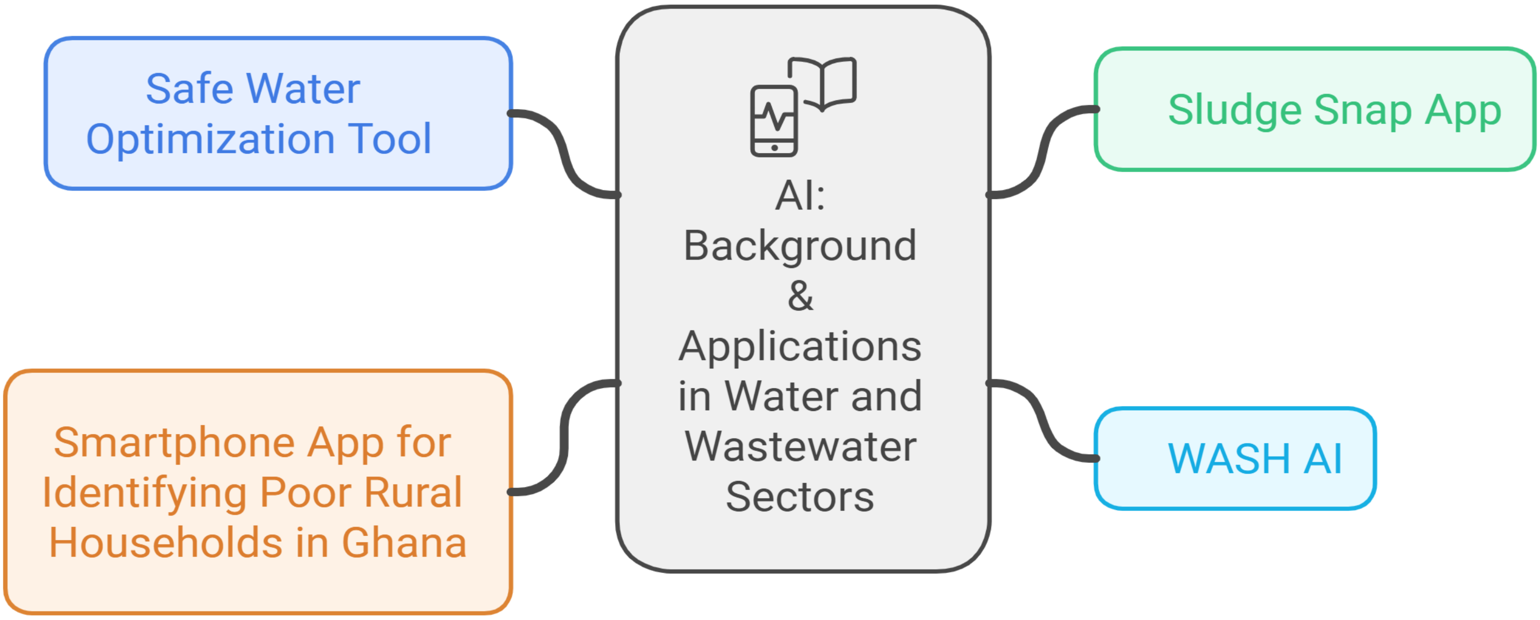
AI in Water and Wastewater Sectors – Background, Potential & Emerging Applications



Introduction

Artificial intelligence (AI) has been with us since the 1950s. Its application in water quality, quantity and access can be traced to 1990. However, it is only now gaining traction in day-to-day applications.

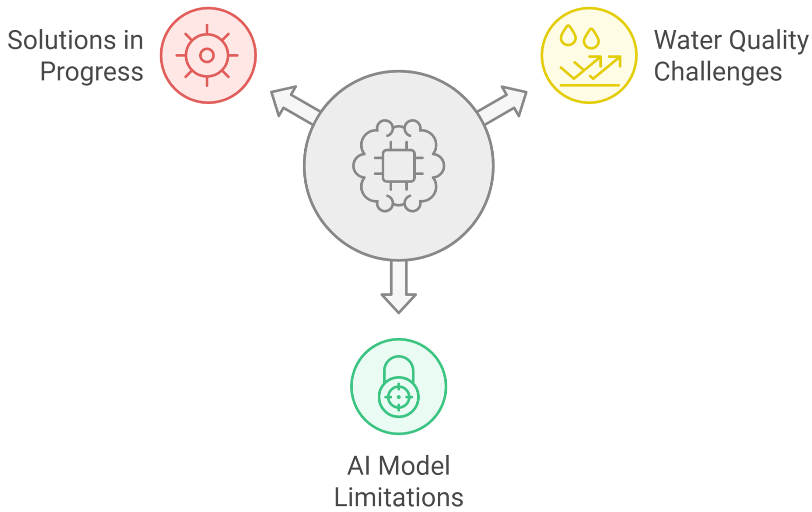
This brief highlights select AI, limitations and potential for research and practice in the prediction of the quality of surface and subsurface waters. It further demonstrates AI's applications that have emerged for day-to-day use in the water and wastewater sectors and their potential impact.

While knowledge of AI and its applications is still rapidly evolving, its potential impact on laboratory analysis in wastewater treatment plants, the synthesis of water and wastewater research, efforts to ensure safety in drinking water quality, and support reaching vulnerable communities continues to develop.

Despite this potential, underlying challenges remain. These include transparency of how AI works behind the scenes, and limits in data quality and quantity, which still need to be addressed.

> "...deep learning has the potential to solve challenges by filling in spatial and temporal data through training to improve its predictions about surface water."

Background



A recent review paper by Joel Podgorski from Eawag and others highlighted the applications and potential of AI, specifically deep learning[1], and inland surface and subsurface waters' (including drinking and polluted water) research and practice [6]. This was essential since intensive and extensive water quality data collection is manually demanding, expensive, and often not conducted in accredited laboratories where independent verification is possible. Therefore, its quality is questionable. In addition, over three-quarters of global data on total suspended solids is sampled from less than a fifth of the rivers worldwide, and the sampling predominantly takes place in North America. Lastly, water quality monitoring does not capture the water quality of the whole river in time and space.

The researchers concluded that deep learning has the potential to solve these challenges by filling in spatial and temporal data through training to improve its predictions about surface water. It would also be able to predict data-scarce variables from similar datasets that are richer and subsurface water quality from catchment properties.

While this is promising, these AI models in general face two challenges:

1. They can only correctly predict within the constraints of their training data. This is also the case for large language models, such as Chat GPT, which often cannot provide meaningful information beyond their training data.
2. They have been criticised as "black boxes", i.e. users often have no idea what is happening in the background.

These two reasons reduce people's confidence and trust in them and, consequently, limit AI uptake and use.

Researchers are, however, working on the two challenges. The training-data-constraints challenge is being resolved by process-guided deep learning and differentiable modeling while the black box challenge is being tackled through explainable deep learning approaches.

- Process-guided learning aims to ensure that the model has domain-specific training and "punishes" deviation from established processes by stakeholders.
- Differentiated models include physically meaningful parameters and equations that can be inspected and manipulated.
- Explainable deep learning approaches aim to resolve the black box challenge by evaluating model 'reasoning', interpreting model decisions, and extracting patterns and drivers.

Yet, this is a work in progress in the water field and beyond. Both these solutions are well elaborated in the paper for those interested [6].

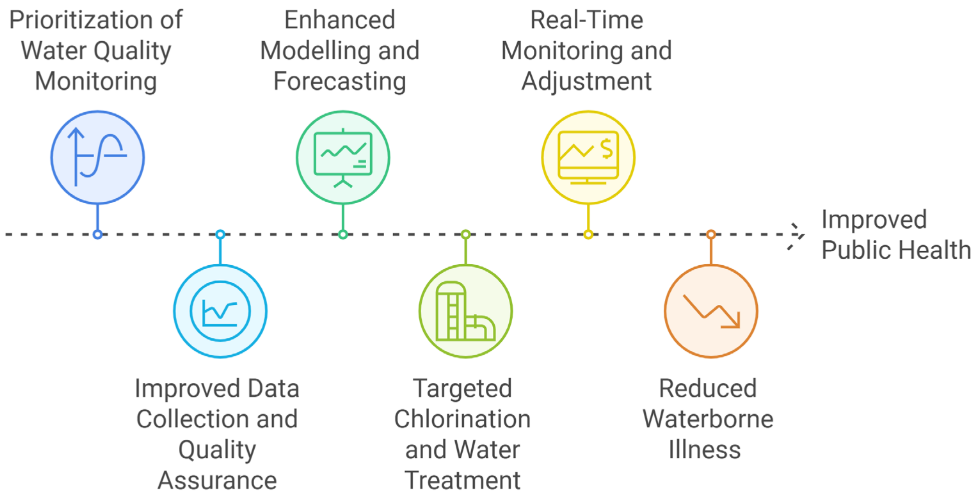
Selected Examples of AI Use in the Water and Wastewater Sectors

In practice, four examples of the use of AI (machine learning and large language models) in the water and sanitation sectors seem promising. These are expounded in the sections that follow, but are not an exhaustive list. They include the Safe Water Optimization Tool, the Sludge Snap App, WASH AI, and a smartphone App for Identifying Poor Rural Households in Ghana.

[1] *Deep learning falls under the category of machine learning enabled by neural networks with more than two layers. Neural networks mimic how our brain works and enable machines to learn from data and make predictions or decisions without being explicitly programmed to do so.*

1. Safe Water Optimization Tool

The Safe Water Optimization Tool (SWOT) is a web-based water quality modelling and assurance platform that helps WASH teams ensure drinking water safety over the last mile of distribution – from collection to the point-of-consumption. It is a collaborative initiative of the Dahdaleh Institute for Global Health Research at York University (Toronto, Canada) and Médecins Sans Frontières/Doctors Without Borders (Amsterdam, The Netherlands) and provides evidence-based, site-specific chlorination targets to help water system operators in humanitarian, and other resource-poor settings reliably maintain free residual chlorine levels. This subsequently protects water from pathogenic recontamination after it is collected from public tap stands and during household storage and use.



Research shows that the impact of this last mile on drinking water quality is site- and population-specific. Status-quo guidance and standards fail to protect water supplies, and this can lead to considerable public health risks.

The SWOT uses routine water quality monitoring data and combines process-based and machine learning modelling to forecast chlorine decay, generate chlorination recommendations, and provide real-time insight into the effectiveness of current water treatment practices. The first ever post-distribution machine learning models integrate water quality data that is not useable in conventional process-based models and generate probabilistic forecasts to measure water safety risks under various treatment and source water quality scenarios.

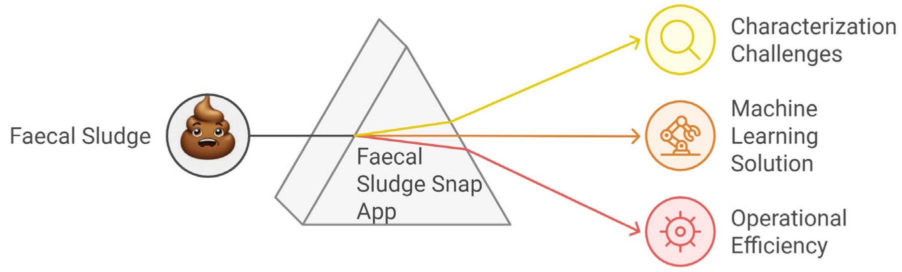
It is built on a user-friendly, low-bandwidth web platform that is accessible in resource-constrained settings. The tool is free-to-use for the humanitarian sector, and we provide a knowledge base, training, and broad-ranging technical support to help users improve water treatment and water quality monitoring.

The SWOT has been used in humanitarian settings around the world to support effective water chlorination programmes across a range of contexts, including in piped systems, water trucking, hand pump chlorinators, and institutions, as well as in systems reliant on both surface water and groundwater. Work is being done on expanding the capabilities of the SWOT to support coordination, and to assess drinking water-related health risks using quantitative microbial risk assessment (QMRA) methods.

A particular challenge to scaling this further is that water quality monitoring is often under-prioritised in WASH programming, resulting in insufficient quantity and quality of data to understand water safety. While new modelling approaches can help get the most out of the data that is collected, more progress must be made in implementing robust monitoring and quality assurance, particularly when it comes to critical health risks, such as drinking water quality.

2. Faecal Sludge Snap App

In urban areas with nonsewered sanitation, wastewater is stored onsite in containments prior to transport for treatment (commonly referred to as faecal sludge). Faecal sludge characteristics are highly variable, which is problematic for the design and operation of treatment technologies, and also for estimating quantities and qualities of faecal sludge for sanitation planning in cities.



For instance, data from the Lubigi treatment plant in Kampala, Uganda, showed that influent chemical oxygen demand (COD) measurements for influent faecal sludge can be up to two times more variable than influent wastewater from sewers.

Characterisation of treatment-relevant metrics, such as total solids, chemical oxygen demand, ammonia, and dewatering performance, is difficult to achieve in many treatment plants because access to analytical laboratories can be prohibitively expensive and time consuming, or it is simply just not available. Solutions are, therefore, needed.

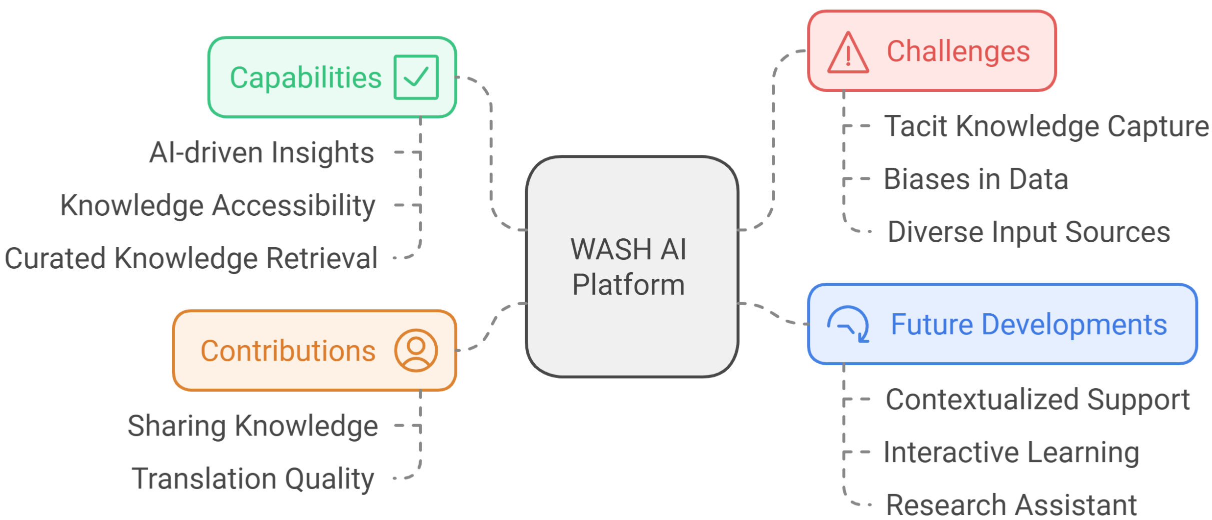
The Sludge Snap App, which is in its Beta version, was developed at Eawag. It is a promising solution that employs a machine learning approach to handle the characterisation of wastewater. The user takes a picture of the sludge, and the app integrates image processing to use colour and texture with a machine learning model to provide rapid, on-site predictions of faecal sludge characteristics, aiming to support operational efficiency.

In the initial development, researchers used a dataset of 465 samples collected from 420 different locations in Lusaka, Zambia. Samples were characterised and onsite predictors were evaluated for their power to predict treatment-relevant analytical parameters. The findings demonstrate that colour and texture data extracted from photographs substantially predicted the qualities of faecal sludge (R^2 of 0.6). In general, machine learning models based on photos and probe measurements worked quite well to predict not only total solids, but also ammonium and dewatering performance in the sludge.

Further piloting and research in additional cities is needded before the app can be scaled. Dependent on obtaining funding, future plans include expanding data collection to build a global database for more robust and universally applicable predictive models through the Sludge Snap App.

3. WASH AI

WASH AI, an initiative of Baobab Tech, seeks to transform knowledge management in the Water, Sanitation, and Hygiene sector using generative artificial



intelligence. The platform's capabilities are designed to offer comprehensive, context-specific insights in the following ways. First, it uses AI-driven insights to facilitate informed decision-making of diverse practitioners, from community-based organisations, government agencies, funders and consultants. Second, it enhances knowledge accessibility by building support in over 20 languages and adapting information to the user's context and expertise level. Lastly, it provides a curated knowledge retrieval service.

The platform's parsing module ingests and curates WASH information from various sources (reports, research, webinars, etc.), enhancing the Language Model's knowledge base with sector-specific content. This process helps to provide more accurate and relevant responses to user queries.

Despite these capabilities, the platform still faces a number of challenges. These include capturing the tacit knowledge held by practitioners in geographically-localised contexts. This kind of knowledge is rarely recorded, is crucial for understanding nuanced, contextual realities, but is difficult to capture and integrate into AI systems. AI systems also risk perpetuating the existing biases inherent in their training data or retrieved information. This includes reinforcing dominant paradigms and neglecting minority or local perspectives.

Ensuring diverse and representative input sources is essential to mitigate this risk. Furthermore, the platform must continually adapt to the local context, considering that environmental, economic, political, and socio-cultural dynamics influence WASH practices. Static or generalised information could lead to ineffective strategies and outcomes.

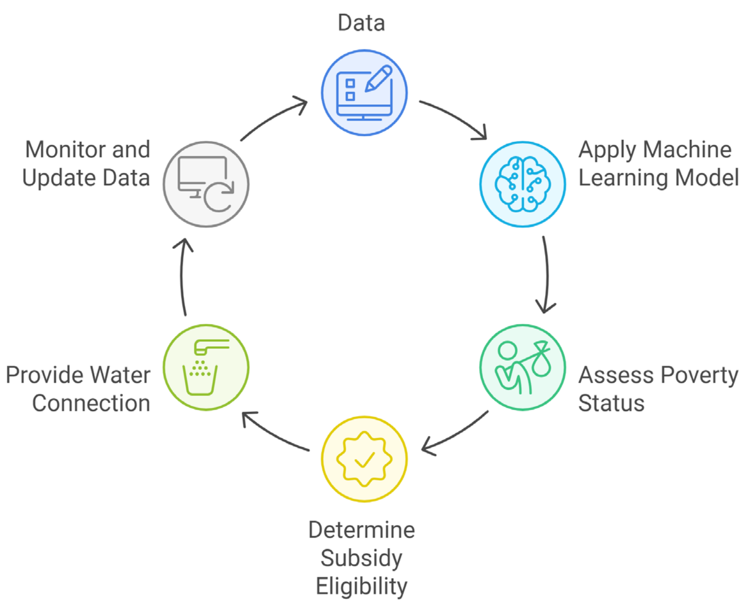
To develop the platform further, some new features will be added. These include the ability to provide more contextualised support by focusing on specific countries and sub-topics within the WASH sector. In addition, a more interactive learning mode aimed at helping practitioners expand their knowledge will be introduced. This mode will offer guided educational experiences based on their specific needs. Lastly, a research assistant that will utilise AI agents for deeper dives into research topics and to provide comprehensive insights and support evidence-based practices will be introduced.

To capture context specific nuances, the platform developers are looking for practitioners to contribute their knowledge and experiences, particularly those with a country-level focus. Additionally, they seek assistance with machine translation quality assessments, especially for WASH-specific terminology. If you would like to contribute, please reach out and help us enhance the platform's diversity, local-relevancy and depth.

4. Smartphone App for Identifying Poor Rural Households in Ghana

The Aquaya Institute's innovative use of artificial intelligence aims to streamline identification of households eligible for low-cost water connections. Determining which households require financial support remains challenging due to variable definitions of poverty, differences in living standards among locations, and subjective eligibility criteria. Further, poverty is not a static condition; circumstances may shift at any time, for example given changes in employment, disability, household size, or broader economic trends.

While we consider water services as a fundamental human right, suppliers with limited resources often face difficulty in determining which



unconnected households to prioritize for improved access. Many service providers rely on broad indicators with little follow-up validation, such as neighborhood affiliation or short-term income. This approach may not accurately reflect the household's level of need, leading to potential misallocation of limited subsidy funds.

To address these challenges, Aquaya developed a machine learning–based tool to support subsidy eligibility screening at the rural household level. By applying feature reduction techniques to the Ghana Living Standard Survey, which holds over 600 proxy questions related to a household's assets and living conditions, Aquaya reduced the number of indicators to 47 that effectively correlated with poverty status.

Using the TensorFlow Lite library in Python (a toolkit that enables on-device machine learning) and neural network algorithms, Aquaya further reduced the number of proxy questions to only 14 with the highest predictive accuracy. Then, integrated this with a free, open-source survey application (Jetsurvey) enabling it to function offline on Android smartphones, including in remote rural areas. To ensure community acceptability, Aquaya added on five questions reflecting locally gathered perspectives on dimensions of poverty (such as food insecurity or chronic illness).

In total, the new screening app contains 19 simple, yes-or-no questions that a rural supplier can use to assess eligibility for a household water connection subsidy within minutes. Aquaya will continue to test and pilot the method in several rural communities in Ghana, through a partnership with the Conrad N. Hilton Foundation and the social enterprise Safe Water Network. While determining how to fund the expansion of household water connections remains challenging, service providers now have one extra tool in their toolbox. This innovation may ultimately serve as a model for using AI technology to support urban water service expansion and other public services and locations outside of Ghana.

Conclusion

Going forward, researchers have to handle the following:

1. Consider data sources including hydrology data, remote sensing, citizen science, and social media, among others, which is essential to augment publicly available time and space data to establish how these relate with surface and subsurface water data for spatio-temporal prediction.
2. Consult with stakeholders to gather new ideas to discover new patterns, processes, and relationships that regulate water quality and service provision dynamics.
3. Find ways to predict future and unmonitored water quality conditions to enable mitigation and support adaptive capacities of communities in the context of climate change.

Further Reading

1. Ali, S. I., Ali, S. S., & Fesselet, J. F. (2021). Evidence-based chlorination targets for household water safety in humanitarian settings: recommendations from a multi-site study in refugee camps in South Sudan, Jordan, and Rwanda. Water Research, 189, 116642.
2. Auffray, V., Poulin, C., Setty, K., Murray, A., & Delaire, C. (2024, June). Simplified Smartphone App for Identifying the Poorest Households in Ghana. The Aquaya Institute. https://aquaya.org/wp-content/uploads/3July_AI-based-Poverty-Screening-Tool-for-Ghana-.pdf
3. Poulin, C., Trimmer, J., Press-Williams, J., Yachori, B., Khush, R., Peletz, R., & Delaire, C. (2022). Performance of a Novel Machine Learning-Based Proxy Means Test in Comparison to Other Methods for Targeting Pro-Poor Water Subsidies in Ghana. Development Engineering, 7, 100098. https://doi.org/10.1016/j.deveng.2022.100098
4. Ward, B. J.; Allen, J.; Escamilla, A.; Sivick, D.; Sun, B.; Yu, K.; Dahlberg, R.; Niu, R.; Ward, B. C.; Strande, L. (2021) Sludge snap: a machine learning approach to fecal sludge characterization in the field, In: Equitable and sustainable WASH services: future challenges in a rapidly changing world. Proceedings of the 42nd WEDC international conference, 3307 (2 pp.)
5. Ward, B. J.; Andriessen, N.; Tembo, J. M.; Kabika, J.; Grau, M.; Scheidegger, A.; Morgenroth, E.; Strande, L. (2021) Predictive models using “cheap and easy” field measurements: can they fill a gap in planning, monitoring, and implementing fecal sludge management solutions?, Water Research, 196, 116997 (12 pp.), doi:10.1016/j.watres.2021.116997
6. Zhi, W., Appling, A. P., Golden, H. E., Podgorski, J., & Li, L. (2024). Deep learning for water quality. Nature Water, 2(3), 228-241.

Further information:
www.sandec.ch

Authorship contribution:
George Wainaina, Joel Podgorski, Linda Strande, Christoph Lüthi (Eawag), James Elliot Brown (SWOT), Olivier Mills (WASH AI), Karen Setty, Chloé Poulin (Aquaya)

Reviewed:
Anna Murray, Duncan Penfold-Brown, and Rachel Peletz

layout:
George Wainaina

Address:
Eawag, Überlandstrasse 133, CH-8600 Dübendorf, Schweiz

+41 58 765 5511

info@sandec.ch

2024

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

aquaya WASH AI Safe Water Optimization Tool eawag aquatic research