



# Artificial intelligence in the water domain: Opportunities for responsible use

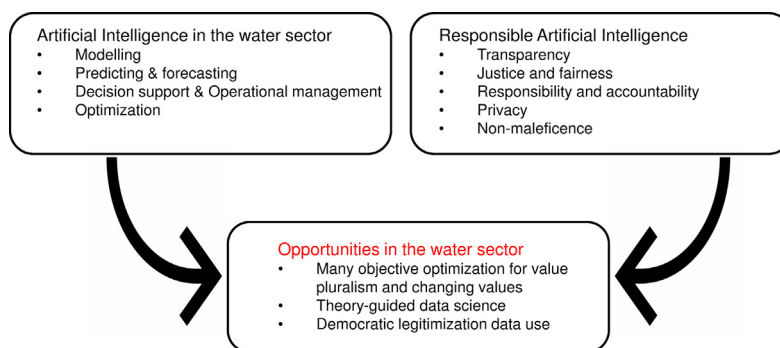
Neelke Doorn

Delft University of Technology, School of Technology, Policy and Management, Department of Values, Technology and Innovation, PO Box 5015, 2600 GA Delft, the Netherlands

## HIGHLIGHTS

- Increasing focus on responsible AI.
- Use of AI techniques in water domain relatively underdeveloped.
- Responsible use depends on the values included in the problem formulation.
- Promising directions include AI for value pluralism and changing values.
- AI in water sector can draw from experience in participatory approaches.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 13 May 2020

Received in revised form 28 August 2020

Accepted 21 September 2020

Available online 29 September 2020

Editor: Martin Drews

### Keywords:

Artificial intelligence

Responsible AI

Data science

Ethics

Water domain

Many-objective optimization

## ABSTRACT

Recent years have seen a rise of techniques based on artificial intelligence (AI). With that have also come initiatives for guidance on how to develop “responsible AI” aligned with human and ethical values. Compared to sectors like energy, healthcare, or transportation, the use of AI-based techniques in the water domain is relatively modest. This paper presents a review of current AI applications in the water domain and develops some tentative insights as to what “responsible AI” could mean there. Building on the reviewed literature, four categories of application are identified: modeling, prediction and forecasting, decision support and operational management, and optimization. We also identify three insights pertaining to the water sector in particular: the use of AI techniques in general, and many-objective optimization in particular, that allow for a pluralism of values and changing values; the use of theory-guided data science, which can avoid some of the pitfalls of strictly data-driven models; and the ability to build on experiences with participatory decision-making in the water sector. These insights suggest that the development and application of responsible AI techniques for the water sector should not be left to data scientists alone, but requires concerted effort by water professionals and data scientists working together, complemented with expertise from the social sciences and humanities.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Digitalization is permeating society in many ways and the role of digital technologies is only expected to increase. While digitalization comes in many forms, in recent years artificial intelligence (AI), including machine learning, has gained enormous traction. Notable

application domains include transportation (autonomous cars), energy, healthcare, and manufacturing. Compared to these other fields, the use of AI in the water domain is still relatively modest.

Although AI in itself is not new (cf. Crevier, 1993), its current use and impact are unprecedented. With that, AI has also become a topic people have strong opinions about. Some present dystopian views about autonomous systems “taking control”, while others see AI as a panacea for many of today’s societal challenges (Russell, 2019). While neither of these extremes seem to be constructive, with the current rise of AI

E-mail address: [N.Doom@tudelft.nl](mailto:N.Doom@tudelft.nl).

techniques it is important to think about the desired role of AI in society (IEEE, 2019). Under the term “responsible AI”, many initiatives are now being launched to compile guidelines for the principles and values according to which AI is to be developed.

The aim of this paper is to present a review of current AI applications in the water domain and to develop some tentative insights on what “responsible AI” could mean there. Following a brief explanation of artificial intelligence and related terms in Section 2, Section 3 describes our review methodology. The results are presented in Section 4. This is followed by a discussion (Section 5) and concluding remarks (Section 6).

## 2. Artificial intelligence

The idea of artificial intelligence is based on the assumption that the process of human thought can be mechanized, which goes as far back as to the ancient Chinese, Indian and Greek philosophers (Berlinski, 2000). Whereas the ancient and medieval philosophers were still primarily concerned with formalization of reasoning, which would allow to reduce argumentation to calculation (Buchanan, 2005), from the early 19th century onwards the work focused on programmable computers. In the first half of the 20th century, progress in the fields of mathematics, psychology, engineering, economics, and neurology (Crevier, 1993), and the development of massive code breaking machines in the Second World War (McCorduck, 2004), as well as analytical work that allowed to describe what it means to conceive of a machine as “thinking” (Turing, 1950), ultimately led to the establishment of artificial intelligence as an academic field. It is at the 1956 Dartmouth Summer Research Project on Artificial Intelligence that the term artificial intelligence was coined (Moor, 2006; for some of the seminal work at that time, see Shannon, 1950; Minsky, 1961; Newell, 1965; Simon, 1967; McCarthy and Hayes, 1969). Since the 1950s, the field of AI has gone through several “waves” of rapid progress and hype, followed by periods of cooling in investment and declining interest (often referred to as the “AI winters” (Floridi, 2020; Hendler, 2008)).

Unavoidably, with such a long history, a precise and stable definition of the term AI is difficult to provide, depending also on the state of the field. For the purpose of this paper, we will focus on what the US Defense Advanced Research Projects Agency (DARPA), one of the major public funders of AI research in the past decades, has recently labeled the second and third wave of AI (DARPA, 2016). The earlier phases of the field, which primarily focused on high-level “symbolic” representations of problems (“Good Old-Fashioned AI” or GOF AI (Haugeland, 1985)), fall beyond the scope of this review. Applications developed in these early days were good at reasoning and heuristics but had limited ability to learn or generalize.

The second wave of AI took off at the turn of the 21st century, with statistical models that were trained with “big data”. These approaches allow for nuanced classification and predication capabilities but there is no contextual capability and only minimal reasoning ability. To make this more concrete: second wave AI algorithms are very well able to recognize a cat from a picture, but they are not able to explain *why* it is a cat. These approaches are dominated by statistical learning, relying heavily on advances in machine learning (ML), including, amongst others, deep learning and evolutionary computation techniques.

ML allows for automatic improvement through experience (Jordan and Mitchell, 2015). It focuses on the development of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead (Koza et al., 1996). For this, ML algorithms build a mathematical model based on sample data, known as “training data” (Hastie et al., 2009). A special class of ML techniques involves deep learning, which refers to algorithms that use multiple layers to progressively extract higher-level features from raw input; this makes them particularly

suitable for learning from unstructured and unlabeled data (Deng and Yu, 2013).

Evolutionary computation forms a family for global optimization inspired by biological evolution. A special case of evolutionary computation concerns genetic computation, which relies on bio-inspired operators such as mutation, crossover, and selection (Goldberg, 1989). Mimicking biological evolution, an algorithm generates an initial set of solutions (“population”) to a problem. This set of solutions is iteratively updated, where each new generation is produced by removing solutions that perform less on the chosen fitness function of the algorithm, and introducing small random changes, analogous to the biological processes of natural selection and mutation. As a result, the fitness of the population, that is, the quality of the solutions to the initially defined problem, will gradually increase (Bäck et al., 1997).

One of the points of criticisms raised against these second wave AI techniques is that the workings of these techniques remain black-boxes and that they are heavily dependent on the quality of the training data. Partly in response to this problem of opacity, DARPA’s third wave of AI techniques focuses on explainability and the development of so-called “white box AI” (Zhu et al., 2018). These AI systems differ from the second wave in their ability for contextual adaptation. Third wave AI techniques understand context and meaning, and are able to adapt accordingly. To use the example of the picture of the cat again, techniques in the third wave will not only be able to recognize the cat, but they will also be able to explain *why* it is a cat and *how* they arrived at that conclusion. If successful, this may reduce some of problems associated with the black box-nature of current ML techniques. Additionally, third wave AI techniques are less dependent on large sets of training data. At the time of writing its overview in 2016, DARPA positioned the third wave of AI prospectively in the 2020s, so whether and how this third wave will actually develop is still an open question.

DARPA’s second and third wave of AI techniques are consistent with definitions in contemporary computer science, where the ability to learn and respond to a changing environment is often seen as key. An intelligent system is able to increase the initial knowledge base through a process of inference or “learning”, and the AI system can be “trained” to recognize patterns or signals and to respond to them (Tareen et al., 2012).

## 3. Methods

Our study applied a systematic review methodology (1) to make an inventory of the AI applications in the water domain and (2) to assess the ethical values and principles that are discussed in the context of AI. Using the ISI Web of Knowledge, the largest and most comprehensive research publication database, two searches were carried out: for English-language journal articles on AI applications in the water domain and on responsible AI, respectively.

### 3.1. Selection of search terms

To allow for a set of articles that is large enough to cover the broadness of AI applications in the water domain but still manageable in terms of number of papers, the following search terms were used. First a search was performed on papers having a relevant AI-term and the word “water” in the title. The relevant AI-terms were: Artificial Intelligence, Machine Learning, Hidden Markov Model, Evolutionary Computation/Algorithm and Genetic Computation/Algorithm. Machine Learning was added as it is characteristic for most current AI techniques. To include more technical papers that do not use general terms like “AI” or “Machine Learning”, the terms “Hidden Markov Model”, “Evolutionary Computation/Algorithm” and “Genetic Computation/Algorithm” were added. Especially evolutionary and genetic computation are widely used in the water domain. The Hidden Markov Model is a statistical technique that can be used to recover a data sequence (time series) that is not immediately observable, but where other data that depend

on the sequence is observable. It is therefore also considered of relevance for modeling activities in the water domain. After that, the same search string was applied again but now to the "topic", which includes not only article title but also article abstract and keywords. This way, also more technical papers that did not use any of the more general AI-related terms or the application field in the title were identified. Since this second search led to a too large set of papers to consider (>12.000), only the 100 most-cited papers of this second search were included. While the search strategy may have a slight bias towards less technical and more applied papers, we consider this justified as the aim is to identify the most relevant type of applications of AI. Also, with the addition of the more specific AI-techniques in the search string, these terms are expected to cover a sufficiently broad range of AI-applications in the water domain.

### 3.2. Search strategy

All subject areas in the database were included. The search was performed on all article types published before 2020.

The following search strategy was used.

- (1) TITLE/TOPIC: ("Artificial Intelligen\*" OR AI OR "Machine Learning" OR ((Evolutionary OR Genetic) AND (Algorithm\* OR Computation\*))) OR "Hidden Markov") AND TITLE/TOPIC: Water AND TIMESPAN: 1900–2019 (hereafter: "Search\_Water")
- (2) TITLE: ("Artificial Intelligen\*" OR AI OR "Machine Learning" OR ((Evolutionary OR Genetic) AND (Algorithm\* OR Computation\*))) OR "Hidden Markov") AND TITLE: (ethic\* OR responsible) AND TIMESPAN: 1900–2019 (hereafter: "Search\_Ethics")

After both searches, the academic literature found was assessed in two steps. This resulted in 855 unique publications in *Search\_Water* and 216 in *Search\_Ethics*. Their abstracts were then sifted for relevance, to exclude papers not about AI and water (*Search\_Water*) or AI and ethics (*Search\_Ethics*). This produced a set of 601 relevant papers on AI and water (*Search\_Water*), and 187 on AI and ethics (*Search\_Ethics*), which were then scanned on a full-paper basis. Papers in *Search\_Water* that did not describe a specific application of AI were excluded from further analysis. Also papers that did not focus on the water domain explicitly were excluded. Papers in *Search\_Ethics* that did not discuss specific ethical issues, values, principles, or guidelines of AI were excluded from the search set. See Fig. 1 for a flow diagram of the selection process.

## 4. Results

### 4.1. Search\_Water

Looking at the results of *Search\_Water*, the applications vary widely in terms of scope and real-world impact, but it is still possible to define some basic categories of application nevertheless. In Section 5, the scope and impact of the different applications is discussed in more detail. The four main categories of application currently being discussed in the literature are modeling, prediction and forecasting, decision support and operational management, and optimization. A small number of papers discussed cybersecurity, but these were too few to warrant discussion as a separate category. In this section, only the four main categories are discussed. But these should not be considered as exclusive or exhaustive.

#### 4.1.1. Modeling

Within the set of articles in *Search\_Water*, 130 (out of 601; 22%) referred to modeling. In these articles, the use of AI was typically described in terms of improving accuracy or reducing uncertainty, or as an efficient way to gather information that would otherwise be difficult to establish. Within this category, a few papers that used AI in computational fluid

mechanics were also found. Here, AI was mostly used as an alternative to solving the algebraic equations numerically (e.g. Shang, 2005).

One example where AI has been used to improve the accuracy of a model is a modeling study of the Andong Dam watershed in Korea. Seo et al. (2015) describe two approaches to model water reservoir systems. Conventional modeling (that is, not using AI techniques) uses statistical models based on time series analysis. However, most of these are linear models, which makes them less suitable for modeling complex hydrological systems that show highly non-linear and non-stationary behavior. The authors compare two AI techniques to model the non-linear behavior of the Andong Dam watershed: artificial neural networks and adaptive neuro-fuzzy inference systems. Artificial neural networks (also known as connectionist models) contain a set of algorithms loosely inspired by the way the human brain works. These models are designed to recognize patterns by considering examples, from which they generate identifying characteristics, without being programmed with task-specific rules (Dawson and Wilby, 2001). Adaptive neuro-fuzzy inference systems combine neural networks with fuzzy logic models, which are mathematical means of representing vagueness and imprecise information and which are therefore especially suitable when handling data that is vague and lacks certainty (Jang, 1993). Seo et al. observe improved behavior of their models as compared to conventional methods. The added value of AI here lies in its ability to model non-linear systems, which improves the accuracy of these models compared to conventional ones based on linear input-output relations.

Many applications used genetic or evolutionary computation to reduce uncertainty. For example, Yu et al. (2019) used a so-called noisy genetic algorithm (NGA) in the context of sustainable water reservoir operation under stochastic inflow conditions. Sustainable reservoir operation requires that operation rules are not only based on satisfying utility demands but also take into account the necessary environmental flow conditions downstream. Meeting both the utility demand and the environmental flow demand is challenging given the stochastic nature of the inflow conditions. Existing studies in sustainable reservoir operation are often based on deterministic inflow conditions (such as historical inflows) as an input for optimization, but these prove to be inaccurate. Alternatively, Monte Carlo simulation or comparable simulation tools can be used to handle the stochastic variables. However, these often require several runs of the tool, which is computationally demanding. By contrast, NGAs can run well without sampling plenty of realizations in different optimizations. The NGA differs from a standard GA in that the fitness function is replaced by a sampling fitness function based on expected fitness rather than deterministic fitness. In an empirical study of the Tanghe reservoir in China, Yu et al. found a reduction in the computation time of 90% compared to a Monte Carlo simulation of the same reservoir and an increased performance in terms of the utility demand and the environmental flows demand.

Studies that use ML to interpret images or other products of remote sensing to obtain spatial information about some relevant property are an example where AI generates information that would otherwise be difficult to establish. Huang et al. (2015) used ML techniques to infer different types of bodies of water from urban high-resolution remote-sensing images. In particular, they found added value of AI compared to conventional methods based on in-situ measurements in classifying different types of bodies of water and associated water extraction. Similar applications are reported by Acharya et al. (2019), who compared different ML techniques to provide information on surface water extraction in Nepal, and by Bair et al. (2018), who used ML techniques to estimate the so-called snow water equivalent (SWE) in mountain watersheds. What these papers have in common is that the spatial distribution of the property being looked for is heterogeneous, so that sparse networks of sensors fail to characterize that heterogeneity.

#### 4.1.2. Prediction and forecasting

Within the set of articles in *Search\_Water*, 78 (out of 601; 13%) referred to prediction and/or forecasting. In these articles, the use of AI

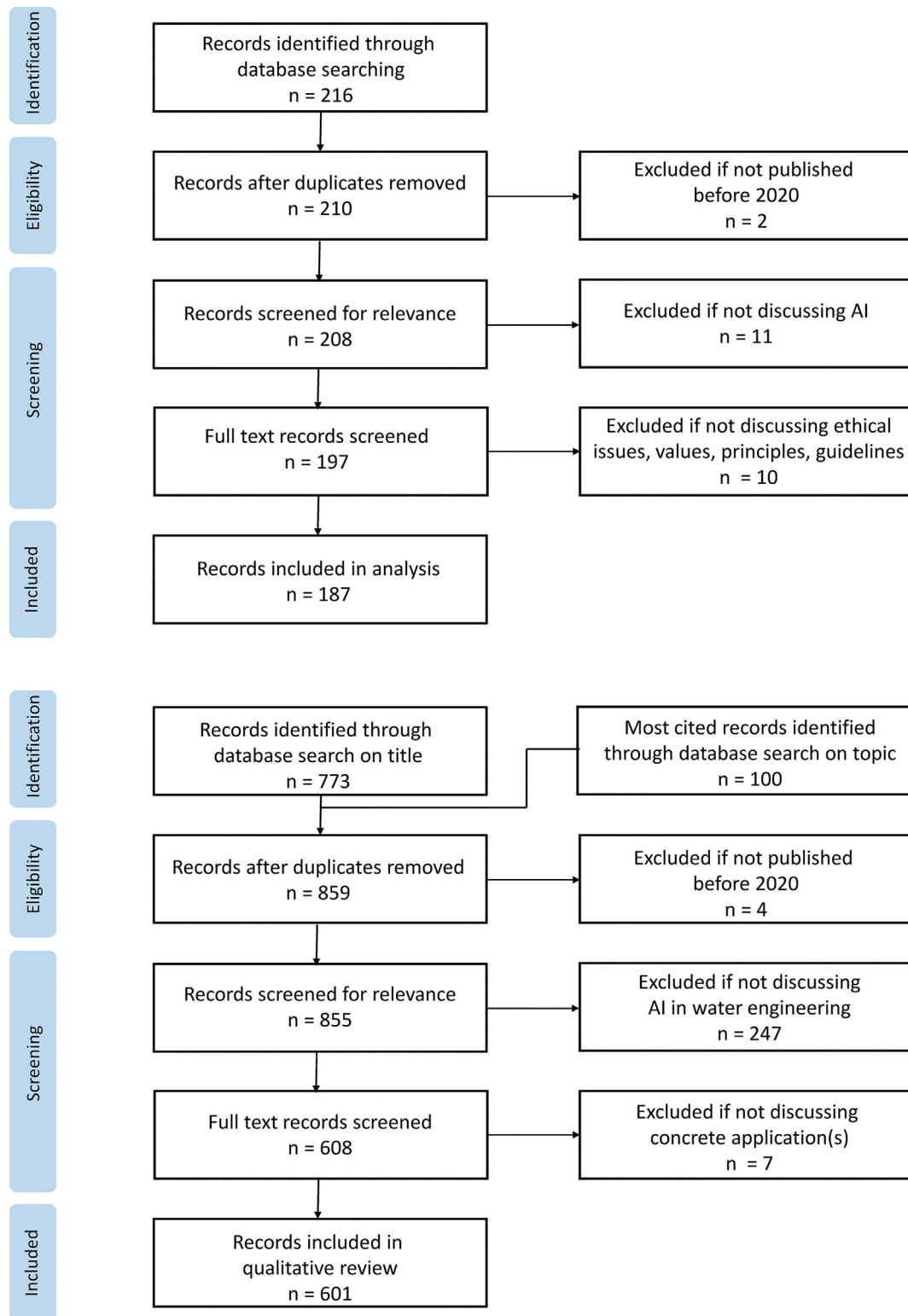


Fig. 1. Flow diagram of the selection process for relevant literature in Search\_Ethics (top) and Search\_Water (bottom) (based on: Moher et al., 2009).

was typically described in terms of its ability to develop predictive models that do not rely on site-specific characteristics or that are applicable to a wide range of environmental conditions. Although there is a clear overlap with the modeling category discussed above, and quite some papers are identified as referring both to modeling and to prediction/forecasting, the papers discussed in the category prediction and forecasting all have a time-dimension, which is not the case for all those categorized under modeling.

For example, Singh and Gupta (2012) developed different predictive models to forecast the formation of trihalomethanes (THMs) in chlorinated waters, which poses a high risk to humans. For adequate control of THMs, their levels in the water need to be known, but determining those levels in the laboratory is too costly and time-consuming. With AI-based predictive models, THM formation can be predicted based on parameters that are easier to determine, such as water pH, temperature, and the concentration of easier to establish chemical compounds, such



as bromide. The results indicate that the AI-based models are capable of capturing the complex non-linear relationship between water conditions and corresponding THM formation.

Another example of prediction is a study that explores the performance of ML models for the forecasting of water quality parameters in coastal waters fed by contributing streams that carry potentially polluted water (Alizadeh et al., 2018). In this case, hourly recorded water quality parameters of salinity, temperature and turbidity in Hilo Bay (Hawaii, USA) were combined with flow data from the contributing Waialuku River. Several ML techniques were used to investigate the impact of the river's flow on the water quality parameters from the current time up to 2 h ahead. The researchers found that water quality parameters can be properly forecasted up to several hours in advance, which may provide valuable information for environmental management and monitoring in coastal areas.

#### 4.1.3. Decision support and operational management

Within the set of articles in *Search\_Water*, 183 (out of 601; 30%) referred explicitly or implicitly to decision support and operational management. This category again partly overlaps with the others, but because of its own specific focus it deserves separate consideration. Many articles in this category focus on event detection and early warning.

An example where AI was used for the detection of accidental water contamination is found in the paper by Arnon et al. (2019). It focuses on water contamination from organic components, which is traditionally detected by inferring "routine" water patterns from measurements of indicator or surrogate physical and chemical parameters of presumably non-contaminated water. Deviations from these routine patterns are then taken as a sign of contamination. Typical physical and chemical parameters used to describe these routine patterns are the water's turbidity, oxidation-reduction potential, free chlorine, and conductivity. However, non-contaminating processes, such as changes in maintenance and operational procedures, may also lead to changes in the physical and chemical parameters, rendering identification of "real" contamination difficult. Hence, this procedure is conditional on stable background conditions. In this paper, a methodology is developed to defer water contamination from the spectroscopic properties of water (UV absorbance spectra). The challenge is to select the relevant features for the given classification problem ("contaminated" or "potable"). Based on an affinity measure combining Pearson correlation and Euclidean distance between the tested absorbance spectrum and the characteristic absorbance of the water types flowing in the water distribution system, an algorithm was developed to detect water anomalies. The innovative part was in the use of a feature selection algorithm that strengthened the relevant differences between the potable and contaminated water. The authors used Gram-based amplification, which allows for computationally efficient feature selection (Ramona et al., 2012). A database of uncontaminated water sources was used to train the detection algorithm. The algorithm proved successful in detecting contaminants at relatively low concentrations with a very low rate of false-positives, under stochastic and highly varying water characteristics.

Bagriacik et al. (2018) compared different AI-based models to detect damage to pipes in the aftermath of an earthquake. The models were evaluated in terms of their ability to accurately predict the total number and approximate spatial distribution of damaged pipes, to correctly classify each individual pipe as damaged or not, and to describe the relative importance of pipe and earthquake attributes in predicting damage.

Yuan et al. (2019) used ML techniques to distinguish eruption and precursory signals of a geyser under noisy environmental conditions. Building on seismic data from the Chimayó geyser (New Mexico, USA), the authors developed a method to filter the relevant signals from background noises, such as daily temperature variations, animal movement near the geyser, and human activity, and to classify the

filtered data into three classes of geyser state: remnant noise, precursor, and eruption states. The proposed ML approach demonstrates an ability to extract eruption and precursory signals from background noise, which makes it very suitable for providing real-time actionable information. Although developed for the detection of geyser eruptions, this model has strong usage potential for other noisy environments where the detection of anomalies from time series is challenging (Yuan et al., 2019).

A few articles combined the quantitative and physical engineering models of some socio-physical system with models that simulate the behavior of the human actors within the system. In the context of the detection and prevention of water contamination, Zechman (2013) developed a model that combined agent-based modeling with evolutionary algorithms to develop and evaluate different threat management strategies. Mewes and Schumann (2019) developed an agent-based irrigation planning model with a machine learning-based training component that is able to identify the current hydrological situation and adapt irrigation and cropping schemes accordingly within the model at runtime.

#### 4.1.4. Optimization

Within the set of articles in *Search\_Water*, 211 (out of 601; 35%) referred to optimization. In these articles, the use of AI was typically described in terms of its ability to provide necessary information to optimize processes or to identify optimal solutions to complex, often ill-structured problems. Although the distinction between modeling and prediction and forecasting on the one hand and optimization on the other is not watertight, optimization tools generally have the capability of systematically determining optimal water planning and management decisions. However, this may come at the expense of sacrificing accurate representation of the underlying water system behavior. Although advances in AI may challenge the very distinction between optimization and the applications discussed under modeling (Labadie, 2014), for the sake of providing an overview of different types of applications, it may still be useful to describe optimization as a separate category. Guo et al. (2017) describe the use of AI for precision irrigation. Accurate information on the water status of a plant root system is essential for precision irrigation, but also difficult to establish. In this study, ML techniques were combined with phenotyping to develop a discrimination method for plant root zone water status in a greenhouse, thereby allowing for more precise and more economical allocation of water. A different type of optimization was used in the study by Strobl and Robillard (2006), who explored the usefulness of AI techniques in the design of a water quality monitoring network: a classical yet complex optimization problem that requires an optimal configuration of sensors to ensure maximum information extraction from the water quality data collected. In this study, the authors explored potentially applicable AI techniques, including expert systems, artificial neural networks, genetic algorithms, and fuzzy logic systems. The review included relatively many similar applications that used AI for the design of some system or water network.

Another example of optimization is found in the paper by Kumar et al. (2018), where AI was used to allocate resources for pro-active maintenance. Around the world, many vital infrastructures are showing their age, which prompts the question as to what constitutes an adequate maintenance scheme for them. Although maintenance in a healthy condition is usually cheaper than repairing failing infrastructures, resources are often limited. The authors focus on water supply infrastructures, which are particularly vulnerable to water main breaks. These not only cause major disruptions to everyday life for residents and businesses, but are costly as well. They also typically occur without any prior warning, making it difficult to schedule maintenance and to decide which mains should be replaced proactively before a break occurs. Moreover, the magnitude and history of several factors contributing to deterioration and failure of the mains are often unknown. These include factors such as the relevant chemical parameters of the water

inside the pipe and in the soil, pipe production parameters, external loading. Using historical data on which mains have failed previously, descriptors of pipes, and other data sources, the authors applied ML techniques to assess the risk of a water main breaking within the next three years. The resulting model provides risk scores for each of the city's blocks, allowing the authorities to schedule maintenance proactively rather than reactively repairing or replacing mains that have broken. In addition to optimized allocation of resources, the authors also see a potential for the model to provide insights into factors that are important in predicting water main failures.

Lastly, quite a number of papers referred to the use of multi-objective and many-objective optimization techniques for trade-off analysis. Multi-objective (in the case of two or three objectives) and many-objective optimization (in the case of more than three objectives) use genetic computation techniques to generate alternatives for complex planning problems, enabling the discovery of the key trade-offs that need to be made between relevant decision parameters (Kasprzyk et al., 2013). This approach is especially relevant for design processes that are multi-dimensional and characterized by uncertainty, of which there are many in the water domain. Examples discussed include water supply (Kasprzyk et al., 2013), urban storm-water runoff (di Pierro et al., 2006), and water quality management (Chatterjee et al., 2017). Strikingly, these were also the articles that paid most attention to how end users (planners, water managers) could use these techniques in daily practice.

This rough overview shows the potential of AI-based techniques in the water domain. Because of the complexity of many processes in this sector, the strength of data-driven over theory-driven models may be that they are better capable of capturing non-linearity in the relevant physical processes. This is facilitated by the increasing availability of relevant sensor data. When it comes to design and planning, we see many applications focusing on the optimal design of water distribution networks. Many of the approaches described in the academic papers have also found their way to real-life applications (Van Thienen et al., 2018). When it comes to the application of ML to decision-making, Hadjimichael et al. (2016) report that only a very small proportion of the academic literature on ML and water deal with decision-making, which suggests that water is lagging behind broadly comparable sectors like energy and logistics (Van Thienen, 2019).

#### 4.2. Search\_Ethics

One of the discussions within the field of AI generally is how to develop "responsible" AI; that is, how to make sure that it is consistent with important ethical values. To assess the extent to which this discussion is also taking place within the subset of AI literature focusing on water, we evaluated how many of the papers in the list of publications from Search\_Ethics are cited by any of the papers in Search\_Water. Web of Science identified 782 articles that cited one or more of those in the set Search\_Ethics, but there was no overlap between this set of citing articles and the publications in the set Search\_Water. In other words, none of the identified articles on AI in the water domain cited any of the articles on AI and ethics in Web of Science. This suggests that there is little or no discussion of the ethical aspects of AI in the water domain.<sup>1</sup>

Looking at how ethics in relation to AI is discussed in the literature, we see two main strands. One focuses on the ethical challenges that AI and ML pose, often linked to specific application domains, such as warfare (e.g. Russell, 2015), medicine (e.g. Char et al., 2018), or labor (e.g. Torresen, 2018). The other strand focuses on the values or principles

that should guide AI and ML developments (Cath, 2018; Dignum, 2018). In the past five years, we have also seen a growth of non-academic institutions developing principles and guidelines for AI. Combining these different strands suggests a general convergence on the relevant ethical principles, but at the same time also substantive divergence on how these principles are to be interpreted, let alone implemented in different domains (Jobin et al., 2019).

Arguably, the five most prominent principles mentioned in the literature are transparency, justice and fairness, responsibility and accountability, privacy, and non-maleficence, each of which can be linked to concerns raised by AI applications.

*Transparency* is often mentioned in response to the perceived black-box nature of AI in general and ML in particular. Decision-making by AI systems, especially those based on deep learning techniques, is frequently said to be opaque, in the sense that, once the algorithms are trained with data, it is practically impossible to examine the internal structure of those algorithms. This makes it difficult to understand why and how they produce a particular outcome (Winfield and Jirotko, 2018).

*Justice and fairness* are mentioned to counter bias in algorithms and to secure equal access to AI-based technologies and their benefits. Standard ML techniques can acquire stereotyped biases from their training data and can inadvertently disseminate and further strengthen these biases (Caliskan et al., 2017). Solutions to bias are sought both in diversifying the training data (Park et al., 2019) and in securing the possibility to appeal or challenge decisions informed by AI-based technologies (Veale and Brass, 2019).

*Responsibility and accountability* are mentioned as a counterbalance to the autonomous character of AI. Discussions about the responsibility for undesirable outcomes of technologies have a long history (e.g. various textbooks on engineering ethics (Davis, 2005; Harris et al., 2005; Van de Poel and Royakkers, 2011)), but with the development of autonomous technologies this discussion has regained relevance. Some scholars stress that responsibility is a matter concerning the people constructing and using the technology and not something that applies to the technology itself. Despite their ability to learn, AI systems are ultimately artifacts constructed by people (Dignum, 2018). Others develop guidelines on how to have humans or society "in the loop" (Rahwan, 2018) or conceptualize what it means to have "meaningful human control" (Santoni de Sio and Van den Hoven, 2018), so that ultimately there are people who could intervene and who could also be held accountable in case of undesirable decisions.

*Privacy* relates to the use of data and the need to protect people's right to privacy. Approaches like "privacy by design" stress the need to minimize the collection of data and to prevent privacy issues from materializing rather than solving them once they occur (Cavoukian, 2010). Others view privacy primarily as a value or right to be protected through regulation (Gürses and del Alamo, 2016).

*Non-maleficence* is a general principle stemming from medical ethics (Beauchamp and Childress, 2001[1994]), indicating an intention to avoid needless harm or injury. In the context of AI, it is often interpreted as the need to protect safety and security (Beil et al., 2019). While the appeal to safety and security could be read as a principle independent from the ones discussed above, many appeals to non-maleficence concern the prevention of infringements of personal privacy and the responsibility to avoid misuse (Floridi and Cowsli, 2019), which suggests that the different principles have some overlap.

#### 5. Discussion

Application of the five ethical principles to the use of AI techniques in the water domain shows that those principles require further operationalization into action-guiding design recommendations for specific AI systems in the water domain.

In order to do so, let us first look at what characterizes the water domain and discuss a few typical AI applications in context. Typical for the

<sup>1</sup> Of course, it cannot be ruled out that some relevant gray literature on ethics and AI/ML or some that uses slightly different terminology is cited by the articles in Search\_Water, but the lack of any citations to the articles in Search\_Ethics at least suggests that the topic of ethics does not feature prominently in the literature on water and AI/ML.

water domain is that it is a largely public domain (Rogers et al., 2002) and that water is seen as a resource that should be available to all people, at least in some sufficient quantity and quality (Gleick, 1998). Secondly, many management issues in the water domain are characterized by scarcity, both of the water itself and in terms of the space and funds available for building and upgrading the relevant infrastructure, both of the water itself and in terms of the space and funds available for implementing and upgrading relevant infrastructure, and conflicting uses (Hajkowicz and Collins, 2007). Additionally, many water problems have a spatial character (Gerlak et al., 2018), often also in combination with a dedicated infrastructure (Clark et al., 2011). Lastly, and this holds especially in relation to climate change, many problems in the water sector are characterized by uncertainty (WWAP, 2012). All these aspects pose ethical challenges in themselves, also without the use of AI (e.g. Groenfeldt, 2013; Brown and Smith, 2010; Doorn, 2013). While these features are not unique to water and not all papers in the review explicitly address all these aspects, many of the papers touch upon at least some of them.

Based on these characterizations of the water domain and the AI applications discussed in Section 4, let us look at some concrete applications that are already in use or which use in the near future is conceivable and that may have a substantial impact on society and the natural environment. While the scientific literature is obviously concerned with the most cutting-edge AI techniques, more conventional data science techniques are already deployed by the sector at a larger scale. A first typical example concerns wastewater-based (or sewage-based) epidemiology (WBE), which refers to the analysis of pollutants and biomarkers in raw wastewater, using techniques from biochemistry and bioinformatics (Lorenzo and Picó, 2019). It can be used to provide real-time information on the production and consumption of legal and illegal drugs, exposure to certain agents (e.g. pesticides or persistent organic pollutants), incidence of specific diseases (e.g. diabetes or cancer), and determination of some lifestyle consequences (e.g. exposure to personal care products or consumption of doping substances). Recently, WBE has been used in the early detection of SARS-Cov-2 (Orive et al., 2020). Although current WBE is based on rather conventional data science, the data may be used as input for algorithmic decision-making and profiling, for example for the detection of consumption or production of illegal drugs. The technique then becomes ethically-sensitive, prompting questions related to privacy and non-discrimination. Increased police surveillance in areas with allegedly high illegal drug use may go against principles of non-discrimination. While the use of WBE for early infection detection may warrant low aggregation levels of relevant data to allow for specific profiling, this may be problematic when done for the purpose of surveillance in criminal investigations. Currently, samples for WBE are mostly taken at the wastewater treatment plant, and thus at high aggregation levels. However, the technique in principle allows for analysis at lower aggregation levels, such as streets or households, which prompts the need to implement safeguards so that personal data is not used for illegitimate purposes in those cases.

A second example concerns algorithmic water allocation. Many applications discussed in Section 4 involved the optimization of water allocation and a better match of water supply and demand. This may eventually result in algorithmic decisions on how to allocate water. While the prudent allocation of scarce resources is to be welcomed, optimization ultimately depends on the formulation of the - to be optimized - objective function. The water ethics literature already discusses how existing approaches aimed at improving “water efficiency” and “water productivity” often deprive smallholders of water use rights and harm local livelihood and production strategies (Boelens and Vos, 2012) and this effect may be exacerbated with increased focus on optimization. Hence, naïve optimization may have negative effects on distributive justice and even conflict with the legally recognized human right to water. A more radical critique along these lines could be derived from the so-called commonwealth-of-life idea.

While the use of AI or data science is not an explicit topic within the water ethics literature, defendants of the commonwealth-of-life idea argue that the very notion of water allocation already reflects a too anthropocentric attitude towards water as a “resource” to be managed by and to the benefit of humans (Brown, 2010). This denies the intrinsic value of water itself. The use of algorithms may add to this concern.

Another example of optimization concerns water user profiling. Water demand data at the household level combined with, for example, GIS-data and census data, allow for evaluation of demand information in a way that it “can help water utilities to plan the water supply system in an optimal manner to meet demand” (Ghavidelfar et al., 2017: p. 2184). However, the use of this same information may also prompt ethical questions when it affects water accessibility of some specific users. Also related to water use is the increased use of nudging tools. Largely drawing from psychological research in decision-making, nudges aim at influencing people to make better decisions by presenting choices in specific ways, still leaving intact their freedom of choice (Thaler and Sunstein, 2008). Nudges are often presented as an effective alternative to traditional monetary or regulatory policy interventions to achieve some public goal, even though critics worry that nudges are manipulative (White, 2013) or coercive (Hausman and Welch, 2010) and thereby a violation of human dignity and autonomy. Generic nudging in the water-sector is already common-place, for example households or farmers who are provided with information on how their own water use relates to that of comparable users to “nudge” them into more conservative behavior (e.g. Chabé-Ferret et al., 2019; Miranda et al., 2019). While generic nudges seem justifiable conditional on due respect for personal autonomy and human dignity (Schmidt and Engelen, 2020), AI-based techniques open the possibility for targeted micro-nudging, where nudges are tailored to specific individuals. This opens a whole new box of ethical questions beyond human dignity and autonomy, especially when used by public or semi-public actors to achieve a public goal, such as water saving. These personalized nudges often involve data or tools from private companies, which are not under democratic control in the way that public or semi-public actors are (Schmidt, 2017). Moreover, when some users are treated differently than others, or when some user groups are nudged into behavior that is beneficial to other people whereas others are not, targeted nudging may conflict with basic demands of justice and equal treatment.

A last example concerns the use of AI techniques for algorithmic investment prioritization generally and maintenance scheduling specifically. The discussion of pipe leak detection in Section 4 is an example where AI is used to detect which parts of a water infrastructure are in need of improvement (Kumar et al., 2018). While it seems in principle a rational approach to schedule maintenance based on failure probabilities, when combined with data about the expected damage, the algorithm may put the infrastructure in more affluent areas higher on the list. This prompts questions about justice: who gets better infrastructure and for what reasons?

These examples show that responsible AI in the water sector is not only a matter of constraining some type of uses or applications, but also one of harnessing its potential (Taddeo and Floridi, 2018). The use of AI techniques may help us to address some of the challenges that water domain is currently faced with, but it may at the same time pave the way to a world characterized by profiling and one-sided optimization. Responsible AI is therefore as much a question of responsible AI techniques as one of responsible governance. And while the use of AI techniques may introduce new ethical concerns related to privacy, transparency, and non-maleficence, it may also alter ongoing discussions in the water ethics literature. Strikingly, the legally recognized human right to water is already formulated in terms of access to relevant information (CESCR, 2002) and this may become only more relevant with the increased use of AI and data science techniques. Discussions on responsibility in relation to water now primarily focus on the forward-looking responsibility to act, for example in relation to flood prevention (Doorn, 2016) or water service delivery (Koehler,



2018), but with the increased use of AI and algorithmic decision-making, the attention may shift towards backward-looking notions of responsibility, such as accountability and liability, which focus not so much on “who should act” but more on “who should repair” or “who should explain how a decision came about”.

Based on these concrete examples and the discussion of general AI principles, we can now formulate some general insights on what it means to develop responsible AI specifically for the water domain, including some promising opportunities for the use of AI.

First, the opportunities for responsible use. At the abstract level, the different ethical principles call for AI techniques that are aligned to human values (Blumenstock, 2018). The challenge, therefore, is to build systems purposed to maximize the realization of these values (Russell, 2016). In the literature, it has been suggested that systems be equipped with suitable utility functions that govern how options are evaluated by the system (Bostrom, 2014).

The use of evolutionary algorithms for many-objective optimization may provide a concrete starting point to broaden the implementation of values in AI systems. Although many-objective optimization is in itself primarily an optimization algorithm, and by that also prone to some of the drawbacks of optimization discussed above, on a more abstract level the method reflects the ideal of value pluralism, that is, the recognition that there is a plurality of legitimate values or goals worth striving for (Doorn, 2019). A study on water allocation in the Lower Rio Grande Valley, Texas, found that the use of many-objective optimization resulted in water management strategies that were more adaptive to conditions of water shortage and that would not have been identified as promising with traditional optimization techniques (Kasprzyk et al., 2016). Recently, distributive justice has been included as a new decision criterion in a flood risk management many-objective optimization problem in the Netherlands (Ciullo et al., 2020).

Additionally, many-objective optimization is currently also deployed to deal with uncertainty (Nicklow et al., 2010). However, uncertainty not only applies to climatic and demographic conditions, but also to what humans value. So the real challenge is not to optimize for specific values upfront, but rather to develop AI techniques that are flexible with regard to the values to be optimized (Van de Poel, 2018). This seems especially relevant in the water domain where often long-term decisions need to be made, for example regarding infrastructures that are designed to last for a relatively long lifespan. Ideally, we should be able to account for changing use or changing appreciation of those infrastructures (e.g. Linsen et al., 2015). Where current adaptive algorithms may provide the necessary flexibility for real-time decision support (Wong and Kerkez, 2016), we are still a long way from developing algorithms that have this more distant time horizon. First steps in this direction are the introduction of models for decision-making under deep uncertainty; for example, the use of many-objective optimization tools to discern decision-relevant scenarios within a wide range of future scenarios (Kwakkel, 2019), where scenarios can also account for variation of values and changing societal values. Ultimately, the practical and responsible use of optimization techniques in general, and evolutionary algorithms in particular, depends on the quality of the problem formulations (Maier et al., 2014).

Regarding the ethical concerns voiced, AI models are usually contrasted with theory-driven models, which is seen both as an advantage and as a drawback. The advantage is that data-driven models are capable of providing useful outcomes in complex situations where existing theories are not able to capture the non-linear aspects of the system. The drawback of these data-driven models is that they are black boxes with no physical meaning and little explanatory power. Here, a combination of data-driven models and physics-based solutions may provide a fruitful middle way (Sun and Scanlon, 2019). In the water domain, a field with a strong physical science base, so-called theory-guided data science (TGDS) may provide a promising way to keep the best of both approaches. By introducing dependencies with sufficient grounding in physical principles, TGDS models have a better chance to represent causal relationships. Since TGDS develops models consistent

with scientific principles, they also achieve better generalizability than models that are purely data-driven (Karpatne et al., 2017). This may address some of the transparency concerns that apply to applications that have no or less physical grounding.

Lastly, many of the dystopian scenarios involving AI really boil down to the undemocratic use of data and lack of involvement by stakeholders actually affected by the outcomes of AI-based technologies. Especially given the spatial and public nature of water and many water infrastructure, participation and public engagement is high on the water ethics agenda (Sharp, 2017). In recent decades, the water domain has gained ample experience in democratizing water policy and in involving stakeholders in decision-making on water-related issues, so as to improve its democratic legitimacy and to avoid water policies that go against important public and ethical values (e.g. Pigmans et al., 2019; Basco-Carrera et al., 2017; Mostert, 2003). While a few articles in the review already addressed the issue of participation (Smith et al., 2019; Lerma et al., 2015; Lewis and Randall, 2017), this aspect could still be further strengthened. Using AI in the water domain requires similar attention to these democratic questions and so should be approached as something requiring more than just data expertise (Blumenstock, 2018). It is therefore important that AI-based water interventions are developed in collaboration between the people who understand the problems and context and the data experts. Ideally, some basic training in data science will be part of the training of the future generation of water professionals (Sun and Scanlon, 2019).

## 6. Conclusions

This paper presents a review and a rough categorization of different types of applications of AI in the water domain. These categories are: modeling, prediction and forecasting, decision support and operational management, and optimization. Comparison of the reviewed literature on AI in the water domain and recent literature on AI and ethics shows that little attention is paid to the ethical aspects of AI applications in the water domain. Recent literature on AI and ethics suggests a list of five principles that should be taken into account when developing AI applications. These are: transparency, justice and fairness, responsibility and accountability, privacy, and non-maleficence.

Application of AI in the water domain remains somewhat limited, compared to its use in other sectors (Hadjimichael et al., 2016). This provides an opportunity to avoid and to learn from the mistakes made elsewhere (Van Thienen, 2019). This paper has discussed three insights pertaining to the water domain in particular. The first is the use of AI-based optimization techniques that optimize on several dimensions. Many-objective optimization may be instrumental in addressing some of the pressing ethical challenges pertaining to water, namely competing uses and uncertainty (Doorn, 2019). Its success in this regard strongly depends on the quality of the problem formulations and the values included in it (Maier et al., 2014). Second, some of the drawbacks of AI in other sectors can be partly compensated because many applications in the water domain are also informed by physics. This allows for theory-guided data science, which can avoid some of the pitfalls of strictly data-driven models. Third, many of the ethical concerns with respect to AI deal with the lack of stakeholder involvement in data-driven decision-making. Fortunately, this is increasingly recognized and there is ample experience with participatory decision-making in the water domain. This suggests that the development and application of responsible AI techniques for the water domain should not be left to data scientists alone, but requires concerted effort by water professionals and data scientists working together, complemented with expertise from the social sciences and humanities.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



## Acknowledgements

This research was supported by a grant from the Dutch National Research Council NWO (grant no. VI.Vidi.195.119).

## References

- Acharya, T.D., Subedi, A., Lee, D.H., 2019. Evaluation of machine learning algorithms for surface water extraction in a Landsat 8 scene of Nepal. *Sensors* 19, 15.
- Alizadeh, M.J., Kavianpour, M.R., Danesh, M., Adolf, J., Shamshirband, S., Chau, K.W., 2018. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. *Engineering Applications of Computational Fluid Mechanics* 12, 810–823.
- Arnon, T.A., Ezra, S., Fishbain, B., 2019. Water characterization and early contamination detection in highly varying stochastic background water, based on Machine Learning methodology for processing real-time UV-spectrophotometry. *Water Res.* 155, 333–342.
- Bäck, Thomas, David B. Fogel, and Zbigniew Michalewicz (eds.). 1997. *Handbook of Evolutionary Computation* (Oxford University Press: Oxford, UK).
- Bagriaciak, A., Davidson, R.A., Hughes, M.W., Bradley, B.A., Cubrinovski, M., 2018. Comparison of statistical and machine learning approaches to modeling earthquake damage to water pipelines. *Soil Dyn. Earthq. Eng.* 112, 76–88.
- Bair, E.H., Calfa, A.A., Rittger, K., Dozier, J., 2018. Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *Cryosphere* 12, 1579–1594.
- Basco-Carrera, Laura, Warren, Andrew, van Beek, Eelco, Jonoski, Andreja, Giardino, Alessio, 2017. Collaborative modelling or participatory modelling? A framework for water resources management. *Environ. Model. Softw.* 91, 95–110.
- Beauchamp, T.L., and J.F. Childress. 2001[1994]. *Principles of Biomedical Ethics* [Fifth Edition] (Oxford University Press: New York).
- Beil, Michael, Proft, Ingo, van Heerden, Daniel, Sviri, Sigal, van Heerden, Peter Vernon, 2019. Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensive Care Medicine Experimental* 7, 70.
- Berlinski, David. 2000. *The Advent of the Algorithm: The 300-year Journey from an Idea to the Computer* (Harcourt: San Diego, USA).
- Blumenstock, Joshua, 2018. Don't forget people in the use of big data for development. *Nature* 561, 1.
- Boelen, Rutgerd, Vos, Jeroen, 2012. The danger of naturalizing water policy concepts: water productivity and efficiency discourses from field irrigation to virtual water trade. *Agric. Water Manag.* 108, 16–26.
- Bostrom, N. 2014. *Superintelligence* (Oxford University Press: Oxford).
- Brown, Peter G., 2010. Are there any natural resources? In: Brown, Peter G., Schmidt, Jeremy J. (Eds.), *Water Ethics: Foundational Readings for Students and Professionals*. Island Press, Washington / Covelo / London.
- Brown, Peter G., and Jeremy J. Smith (eds.). 2010. *Water Ethics: Foundational Readings for Students and Professionals* (Island Press: Washington).
- Buchanan, Bruce G., 2005. A (very) brief history of artificial intelligence. *AI Mag.* 26, 53–60.
- Caliskan, Aylin, Bryson, Joanna J., Narayanan, Arvind, 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 183–186.
- Cath, C., 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges introduction. *Philos. Trans. R. Soc. A. Math. Phys. Eng. Sci.* 376, 8.
- Cavoukian, Ann, 2010. Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D. *Identity in the Information Society* 3, 247–251.
- CESCR. 2002. Right to water, general comment no. 15, 20 January 2003, E/C.12/2002/11. Committee on Economic, Social and Cultural Rights - United Nations.
- Chabé-Ferret, Sylvain, Le Coent, Philippe, Reynaud, Arnaud, Subervie, Julie, Lepercq, Daniel, 2019. Can we nudge farmers into saving water? Evidence from a randomised experiment. *Eur. Rev. Agric. Econ.* 46, 393–416.
- Char, D.S., Shah, N.H., Magnus, D., 2018. Implementing machine learning in health care - addressing ethical challenges. *N. Engl. J. Med.* 378, 981–983.
- Chatterjee, S., Sarkar, S., Dey, N., Sen, S., Goto, T., Debnath, N.G., Ilee, 2017. Water quality prediction: multi objective genetic algorithm coupled artificial neural network based approach. *IEEE 15th International Conference on Industrial Informatics*.
- Clark, Robert M., Simon Hakim, and Avi Ostfeld (eds.). 2011. *Handbook of Water and Wastewater Systems Protection* (Springer: New York, NY).
- Crevier, Daniel. 1993. *AI: The Tumultuous Search for Artificial Intelligence* (Basic Books: New York, NY).
- DARPA. 2016. 'A DARPA Perspective on AI [Accessed July 20, 2020]'. <https://www.darpa.mil/about-us/darpa-perspective-on-ai>.
- Davis, M. (Ed.) 2005. *Engineering Ethics* (Ashgate Publishing: Aldershot).
- Ciullo, A., Kwakkel, J.H., De Bruijn, K., Doorn, N., Klijn, F., 2020. Efficient or fair? Operationalizing ethical principles in flood risk management: A case study on the Dutch-German Rhine. *Risk Analysis* 40 (9), 1844–1862. <https://doi.org/10.1111/risa.13527>.
- Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. *Progress in Physical Geography: Earth and Environment* 25, 80–108.
- Deng, Li, Yu, Dong, 2013. Deep learning: methods and applications. *Foundations and Trends in Signal Processing* 7, 197–387.
- di Pierro, F., Khu, S.T., Savić, D., 2006. From single-objective to multiple-objective multiple-rainfall events automatic calibration of urban storm water runoff models using genetic algorithms. *Water Sci. Technol.* 54, 57–64.
- Dignum, V., 2018. Ethics in artificial intelligence: introduction to the special issue. *Ethics Inf. Technol.* 20, 1–3.
- Doorn, Neelke, 2013. Water and justice: towards an ethics for water governance. *Public Reason* 5, 95–111.
- Doorn, Neelke, 2016. "Distributing responsibilities for safety from flooding." In *FLOODrisk 2016 - 3rd European Conference on Flood Risk Management* edited by M. Lang, F. Klijn and P. Samuels, DOI: 10.1051/e3sconf/20160724002. Lyon.
- Doorn, Neelke, 2019. *Water Ethics: An Introduction* (Rowman & Littlefield: New York).
- Floridi, Luciano, 2020. AI and its new winter: from myths to realities. *Philos. Technol.* 33, 1–3.
- Floridi, Luciano, and Josh Cowls. 2019. 'A unified framework of five principles for AI in society', *Harvard Data Science Review*, 1.
- Gerlak, Andrea K., House-Peters, Lily, Varady, Robert G., Albrecht, Tamee, Zúñiga-Terán, Adriana, de Grenade, Rafael Routson, Cook, Christina, Scott, Christopher A., 2018. *Water security: a review of place-based research*. *Environ. Sci. Pol.* 82, 79–89.
- Ghavidelfar, Saeed, Shamseldin, Asaad Y., Melville, Bruce W., 2017. A multi-scale analysis of single-unit housing water demand through integration of water consumption, land use and demographic data. *Water Resour. Manag.* 31, 2173–2186.
- Gleick, Peter H., 1998. The human right to water. *Water Policy* 1, 487–503.
- Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley/Pearson PLC: Boston, MA).
- Groenfeldt, D., 2013. *Water Ethics: A Values Approach to Solving the Water Crisis* (Earthscan / Routledge: New York).
- Guo, D.D., Juan, J.X., Chang, L.Y., Zhang, J.J., Huang, D.F., 2017. Discrimination of plant root zone water status in greenhouse production based on phenotyping and machine learning techniques. *Sci. Rep.* 7, 12.
- Gürses, S., del Alamo, J.M., 2016. Privacy engineering: shaping an emerging field of research and practice. *IEEE Security & Privacy* 14, 40–46.
- Hadjimichael, A., Comas, J., Corominas, L., 2016. Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. *AI Commun.* 29, 747–756.
- Hajkowicz, Stefan, Collins, Kerry, 2007. A review of multiple criteria analysis for water resource planning and management. *Water Resour. Manag.* 21, 1553–1566.
- Harris, C.E., M.S. Pritchard, and M.J. Rabins. 2005. *Engineering Ethics: Concepts and Cases* (Wadsworth: Belmont, CA).
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer New York: New York, NY).
- Haugeland, John. 1985. *Artificial Intelligence: The Very Idea* (MIT Press: Cambridge, MA).
- Hausman, Daniel M., Welch, Brynn, 2010. Debate: to nudge or not to nudge\*. *J. Polit. Philos.* 18, 123–136.
- Hendler, James, 2008. Avoiding another AI winter. *IEEE Intell. Syst.* 23, 2–4.
- Huang, X., Xie, C., Fang, X., Zhang, L.P., 2015. Combining pixel- and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8, 2097–2110.
- IEEE. 2019. "Ethically Aligned Design: A Vision for Prioritizing Human Well-being With Autonomous and Intelligent Systems [First Edition]." In: <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>: The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.
- Jang, J.-S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cyber.* 23, 665–685.
- Jobin, Anna, Ienca, Marcello, Vayena, Effy, 2019. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* 1, 389–399.
- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260.
- Karpatne, A., G. Atluri, J.H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar. 2017. 'Theory-guided Data Science: A New Paradigm for Scientific Discovery From data', *IEEE Transactions on Knowledge and Data Engineering*, 29.
- Kasprzyk, J.R., Nataraj, S., Reed, P.M., Lempert, R.J., 2013. Many objective robust decision making for complex environmental systems undergoing change. *Environ. Model. Softw.* 42, 55–71.
- Kasprzyk, Joseph R., Reed, Patrick M., Hadka, David M., 2016. Battling arrow's paradox to discover robust water management alternatives. *J. Water Resour. Plan. Manag.* 142, 04015053. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000572](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000572).
- Koehler, Johanna, 2018. Exploring policy perceptions and responsibility of devolved decision-making for water service delivery in Kenya's 47 county governments. *Geoforum* 92, 68–80.
- Koza, John R., Bennett, Forrest H., Andre, David, Keane, Martin A., 1996. Automated design of both the topology and sizing of analog electrical circuits using genetic programming. In: Gero, John S., Sudweeks, Fay (Eds.), *Artificial Intelligence in Design '96*. Springer Netherlands, Dordrecht.
- Kumar, A., S. A. Rizvi, B. Brooks, R. A. Vanderveld, K. H. Wilson, C. Kenney, S. Edelstein, A. Finch, A. Maxwell, J. Zuckerbraun, R. Ghani, and Acm. 2018. Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks.
- Kwakkel, Jan H., 2019. A generalized many-objective optimization approach for scenario discovery. *Futures & Foresight Science* 1, e8.
- Labadie, J.W., 2014. *Advances in Water Resources Systems Engineering: Applications of Machine Learning*.
- Lerma, N., Paredes-Arquiola, J., Andreu, J., Solera, A., Sechi, G.M., 2015. Assessment of evolutionary algorithms for optimal operating rules design in real water resource systems. *Environ. Model. Softw.* 69, 425–436.
- Lewis, A., Randall, M., 2017. Solving multi-objective water management problems using evolutionary computation. *J. Environ. Manag.* 204, 179–188.
- Linsen, Max, Erik Mostert, and Pieter Van der Zaag. 2015. 'Infrastructure and adaptive management in an eco-hydrological Delta: Lessons learned from design and

- construction of the Haringvliet Sluices.' in, *EGU General Assembly Conference Abstracts* (EGU: Vienna).
- Lorenzo, Maria, Picó, Yolanda, 2019. Wastewater-based epidemiology: current status and future prospects. *Current Opinion in Environmental Science & Health* 9, 77–84.
- Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D.P., Vrugt, J.A., Zechin, A.C., Minsker, B.S., Barbour, E.J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., Reed, P.M., 2014. Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions. *Environ. Model. Softw.* 62, 271–299.
- McCarthy, John, and Patrick J. Hayes. 1969. Some Philosophical Problems From the Standpoint of Artificial Intelligence (Stanford University).
- McCorduck, Pamela. 2004. *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence* (Taylor & Francis: Boca Raton, FL).
- Mewes, B., Schumann, A., 2019. The potential of combined machine learning and agent-based models in water resources management. *Hydrol. Wasserbewirtsch.* 63, 332–338.
- Minsky, Marvin, 1961. Steps toward artificial intelligence. *Proc. IRE* 49, 8–30.
- Miranda, Juan Jose, Datta, Saugato, Zoratto, Laura, 2019. Saving water with a nudge (or two): evidence from Costa Rica on the effectiveness and limits of low-cost behavioral interventions on water use. *World Bank Econ. Rev.* 34, 444–463.
- Moher, David, Liberati, Alessandro, Tetzlaff, Jennifer, Altman, Douglas G., Prisma Group The, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6, e1000097.
- Moor, J., 2006. The Dartmouth College artificial intelligence conference: the next fifty years. *AI Mag.* 27, 87–89.
- Mostert, Erik, 2003. The challenge of public participation. *Water Policy* 5, 179–197.
- Newell, Allen, 1965. Limitations of the current stock of ideas about problemsolving. In: Kent, A., Taulbee, O. (Eds.), *Proceedings of a Conference on Electronic Information Handling*. Spartan, New York.
- Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., Zechman, E., Asce Task Comm Evolutionary, 2010. State of the art for genetic algorithms and beyond in water resources planning and management. *J. Water Resour. Plan. Manag.* 136, 412–432.
- Orive, Gorka, Lertxundi, Unax, Barcelo, Damia, 2020. Early SARS-CoV-2 outbreak detection by sewage-based epidemiology. *Sci. Total Environ.* 732, 139298.
- Park, S.H., Kim, Y.H., Lee, J.Y., Yoo, S., Kim, C.J., 2019. Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review. *Science Editing* 6, 91–98.
- Pigmans, Klara, Aldewereld, Huib, Dignum, Virginia, Doorn, Neelke, 2019. Value deliberation to improve stakeholder participation in water governance. *Water Resour. Manag.* 33, 4067–4085.
- Rahwan, Iyad, 2018. Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf. Technol.* 20, 5–14.
- Ramona, Mathieu, Richard, Gaël, David, Bertrand, 2012. Multiclass feature selection with kernel gram-matrix-based criteria. *IEEE Transactions on Neural Networks and Learning Systems* 23, 1611–1623.
- Rogers, Peter, de Silva, Radhika, Bhatia, Ramesh, 2002. Water is an economic good: how to use prices to promote equity, efficiency, and sustainability. *Water Policy* 4, 1–17.
- Russell, Stuart J., 2015. Ethics of artificial intelligence. *Nature* 521, 415–416.
- Russell, Stuart J., 2016. Should we fear supersmart robots. *Sci. Am.* 314, 58–59.
- Russell, Stuart J. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking - Penguin Random House: London).
- Santoni de Sio, Filippo, Van den Hoven, Jeroen, 2018. Meaningful human control over autonomous systems: a philosophical account. *Frontiers in Robotics and AI* 5.
- Schmidt, Andreas T., 2017. The power to nudge. *Am. Polit. Sci. Rev.* 111, 404–417.
- Schmidt, Andreas T., Engelen, Bart, 2020. The ethics of nudging: an overview. *Philos Compass* 15, e12658.
- Seo, Y., Kim, S., Kisi, O., Singh, V.P., 2015. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* 520, 224–243.
- Shang, Z., 2005. Application of artificial intelligence CFD based on neural network in vapor-water two-phase flow. *Eng. Appl. Artif. Intell.* 18, 663–671.
- Shannon, Claude E., 1950. XXII. Programming a computer for playing chess. *London Edinburgh Dublin Philos. Mag. J. Sci.* 41, 256–275.
- Sharp, Liz, 2017. *Reconnecting People and Water: Public Engagement and Sustainable Urban Water Management*. Oxon / New York, Routledge.
- Simon, Herbert A., 1967. The logic of heuristic decision making. In: Rescher, N. (Ed.), *The Logic of Decision and Action*. University of Pittsburgh, Pittsburgh.
- Singh, K.P., Gupta, S., 2012. Artificial intelligence based modeling for predicting the disinfection by-products in water. *Chemom. Intell. Lab. Syst.* 114, 122–131.
- Smith, R., Kasprzyk, J., Dilling, L., 2019. Testing the potential of multiobjective Evolutionary algorithms (MOEAs) with Colorado water managers. *Environ. Model. Softw.* 117, 149–163.
- Strobl, R.O., Robillard, P.D., 2006. Artificial intelligence technologies in surface water quality monitoring. *Water Int.* 31, 198–209.
- Sun, A.Y., Scanlon, B.R., 2019. How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environ. Res. Lett.* 14, 28.
- Taddeo, Mariarosaria, Floridi, Luciano, 2018. How AI can be a force for good. *Science* 361, 751.
- Tareen, J.A.K., Abbasi, T., Abbasi, S.A., 2012. Water-Quality Indices Based on Fuzzy Logic and Other Methods of Artificial Intelligence.
- Thaler, Richard H., Sunstein, Cas R., 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, Yale University Press.
- Torresen, J., 2018. A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI* 4, 10.
- Turing, Alan. 1950. 'Computing machinery and intelligence', *Mind*, LIX: 4330460.
- Van de Poel, Ibo R., 2018. Design for value change. *Ethics and Information Technology* <https://doi.org/10.1007/s10676-018-9461-9>.
- Van de Poel, Ibo R., Royakkers, L.M.M., 2011. *Ethics, Technology, and Engineering: An Introduction*. West-Sussex, Wiley-Blackwell.
- Van Thienen, P. 2019. "responsible AI for the water sector?" in. <https://www.kwrwater.nl/actueel/responsible-ai-for-the-water-sector/>: KWR.
- Van Thienen, Peter, Henk-Jan Van Alphen, Andrea Brunner, Yuki Fujita, Bram Hillebrand, Rosa Sjerps, Joost Van Summeren, Anthony Verschoor, and Bart Wullings. 2018. "Explorations in data mining for the water sector." in. Nieuwegein, The Netherlands: KWR Watercycle Research Institute.
- Veale, Michael, Brass, Irina, 2019. Administration by algorithm? Public management meets public sector machine learning. In: Yeung, Karen, Lodge, Martin (Eds.), *Algorithmic Regulation*. Oxford University Press, Oxford.
- White, M.D., 2013. *The Manipulation of Choice: Ethics and Libertarian Paternalism*. New York, Palgrave Macmillan.
- Winfield, A.F.T., Jirotko, M., 2018. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philos. Trans. R. Soc. A. Math. Phys. Eng. Sci.* 376, 13.
- Wong, Brandon P., Kerkez, Branko, 2016. Real-time environmental sensor data: an application to water quality using web services. *Environ. Model. Softw.* 84, 505–517.
- WWAP, 2012. "the United Nations World Water Development Report 4: Managing Water under Uncertainty and Risk." in. UNESCO - World Water Assessment Programme, Paris.
- Yu, C.X., Yin, X.N., Yang, Z.F., Dang, Z., 2019. 'Sustainable Water Resource Management of Regulated Rivers under Uncertain Inflow Conditions Using a Noisy Genetic Algorithm', *International Journal of Environmental Research and Public Health*, 16.
- Yuan, B.C., Tan, Y.J., Mudunuru, M.K., Marcillo, O.E., Delorey, A.A., Roberts, P.M., Webster, J.D., Gammans, C.N.L., Karra, S., Guthrie, G.D., Johnson, P.A., 2019. Using machine learning to discern eruption in noisy environments: a case study using CO2-driven cold-water geyser in Chimayo, New Mexico. *Seismol. Res. Lett.* 90, 591–603.
- Zechman, E.M., 2013. Integrating evolution strategies and genetic algorithms with agent-based modeling for flushing a contaminated water distribution system. *J. Hydroinf.* 15, 798–812.
- Zhu, Jichen, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G. Michael Youngblood. 2018. "Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation." In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1–8.