nature water

Perspective

https://doi.org/10.1038/s44221-023-00069-6

Rewards, risks and responsible deployment of artificial intelligence in water systems

Received: 11 November 2022

Accepted: 13 March 2023

Published online: 11 May 2023

Check for updates

Catherine E. Richards[1,2], Asaf Tzachor[1,3], Shahar Avin[1] & Richard Fenner[2]

Artificial intelligence (AI) is increasingly proposed to address deficiencies across water systems, which currently leave about 25% of the global population without clean water, about 50% without sanitation services and about 30% without hygiene facilities. AI is poised to enhance supply insights, catchment management and emergency response, improve treatment plant and distribution network design, operation and maintenance, and advance service availability, demand management and water justice. However, proliferation of this nascent technology could trigger serious and unexpected problems, including system-wide compromise owing to design errors, malfunction and cyberattacks as well as exposures to cascading socio-ecological, water–energy–food nexus and coupled critical infrastructure failures. In response, we make three recommendations for safe and responsible deployment of AI across potable water supply and sewage disposal systems: address gaps in foundational infrastructure and digital literacy; establish institutional, software and hardware mechanisms for trustworthy AI; and prioritize applications based on our proposed systematic benefit and risk assessment framework.

Early scientific developments in potable water supply and wastewater disposal systems (encapsulated hereafter as 'water systems') enabled ancient societies to transform into urban metropolises beyond their riverside origins and build resilience to weather perturbations, including wet and dry spells[1]. For instance, the Nazcans constructed subterranean aqueducts to transport drinking water long distances while mitigating evaporation losses[2], and the Indus Valley civilization constructed brick sewers to drain baths and latrines into isolated soak pits to mitigate exposure of people to sewage[3].

While engineering feats have produced manifold benefits, some instances of technological innovation have resulted in 'progress traps': events where human ingenuity to solve a given problem inadvertently manifests unanticipated problems that outpace society's—and technology's—capacity to then solve them[4]. For instance, Ancient Rome's lead plumbing was an engineering marvel, connecting its vast population to reliable water and wastewater networks, but its outflows have also been linked to contaminating harbour water with lead, potentially poisoning marine life and people[5].

More recently, artificial agricultural irrigation has depleted groundwater aquifers[6] and caused salination[7]. Wastewater treatment has inadvertently contributed to global warming, toxicity and acidification[8]. Desalination of sea water has caused air, marine and land pollution[9]. Innovations in adjacent sectors realizing short-term benefits have created longer-term problems for water resources, such as hydroelectric dams for energy production degrading aquatic ecosystems, biogeochemical dynamics and water quality[10]. Despite successful, and essential, innovations across water systems, our thirst for technology-based problem-solving has often locked us into chronic progress traps.

Today, some 25% of the global population lack access to clean water, 50% lack access to sanitation services and 30% lack access to hygiene facilities[11]. Anthropogenic climate change threatens to exacerbate these issues, with higher temperatures increasing water scarcity globally and extreme events, including storms, floods and droughts, damaging water systems infrastructure in developed nations and undermining water, sanitation and hygiene (WASH) efforts in developing nations[12].

[1]Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK. [2]Department of Engineering, University of Cambridge, Cambridge, UK. [3]School of Sustainability, Reichman University (IDC Herzliya), Herzliya, Israel. e-mail: cer76@cam.ac.uk; at875@cam.ac.uk

 https://doi.org/10.1038/s44221-023-00069-6

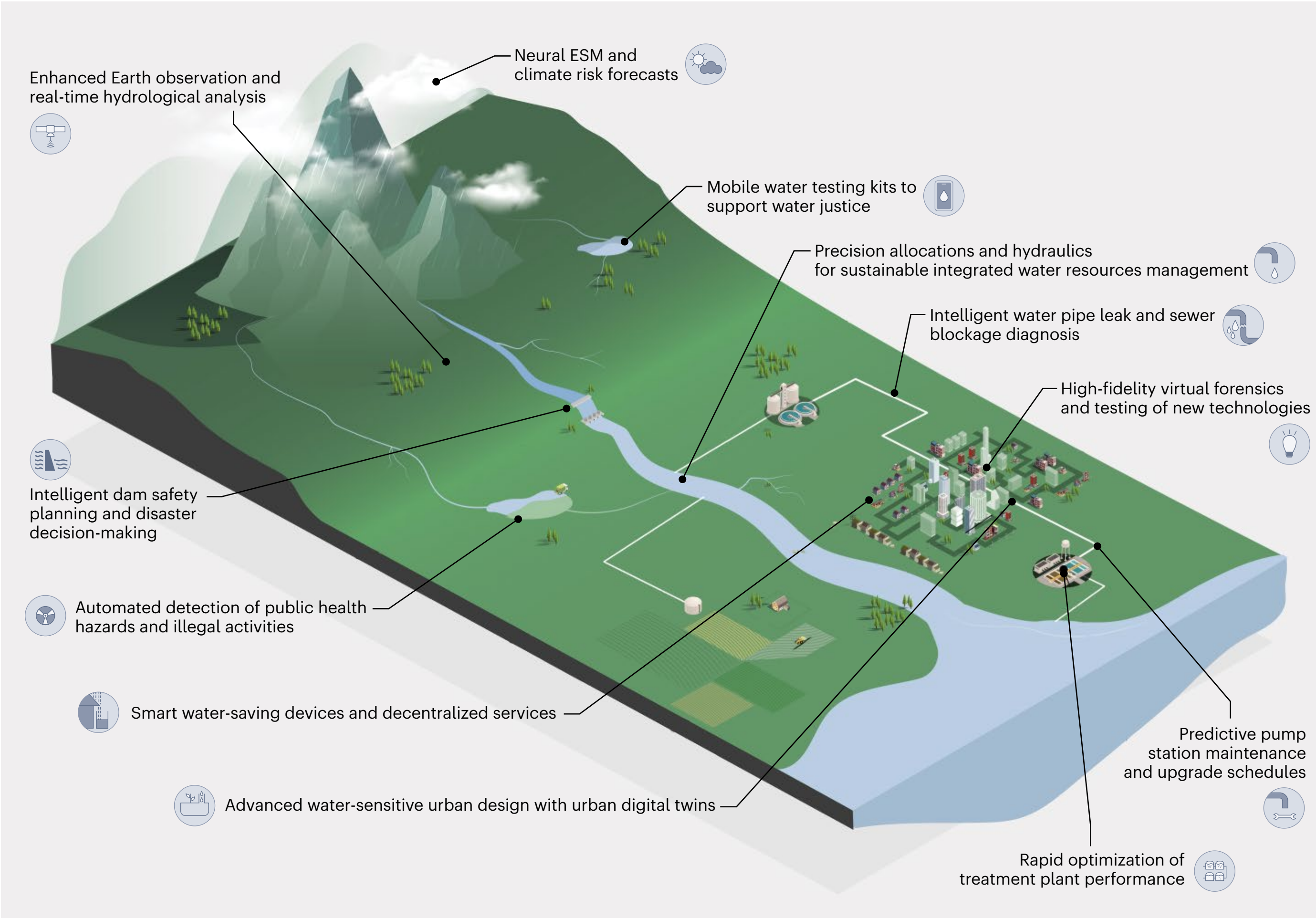

Fig. 1 | Example benefits of AI for solving problems across water systems. AI has the potential to yield system-wide benefits ranging from enhanced catchment insights to optimized network efficiency to improved service for end-users.

Against this backdrop, artificial intelligence (AI), and its subdivision of machine learning (ML), is the latest technological intervention proposed to solve problems across water systems by building climate resilience, enhancing performance of infrastructure and, in limited cases, assisting WASH efforts. However, burgeoning applications of AI may give rise to serious, and unexpected, problems that are underappreciated and must be responsibly, and pre-emptively, managed to avoid unintentionally undermining efforts to meet Sustainable Development Goal 6.

In this Perspective, we provide a balanced consideration of AI in water systems. We survey potential system-wide benefits of AI applications from catchment to end-user. Then we highlight potential systemic barriers, direct risks and exposures to cascading failures, which may prove catastrophic for communities. Finally, we propose a three-tiered risk mitigation approach, necessary to prevent proliferation of this currently nascent technology perpetuating the progress-trap phenomenon.

Here we define AI as a machine-based 'intelligent agent' capable of interacting with its environment with the aid of sensors, interpreting information for decision-making and autonomously taking actions to achieve goal-oriented outcomes via a human or robotic actuator, while ML refers to the subset of algorithmic models that learn and predict outcomes through passive observation of the environment[13].

Benefits from catchment to end-user

Given the current lack of wide-scale deployment in the 'real world', we highlight presumed benefits from AI applications across three levels covering water systems end-to-end: (1) water supply (catchment level), including enhanced supply insights, catchment management and emergency response; (2) water distribution and disposal (network level), including efficient treatment and network infrastructure design, operations and maintenance; and (3) water demand (end-user level), including improved service availability, demand management and water justice (Fig. 1).

Beyond component-specific applications detailed below, advanced AI may also eventually be used to simulate, inform and optimize operating policy for whole water systems in line with integrated water resources management principles[14].

Enhanced insights at catchment level

Over 10% of people worldwide are exposed to high and critical water stress, and climate change is expected to worsen this exposure in urban and rural areas alike[15]. As such, complete, high-resolution and reliable analysis of Earth's natural water resources, hydrologic cycles and anthropogenic perturbations is essential to monitor and manage water supply[16].

ML models may process big datasets, such as interferometric synthetic aperture radar imagery, and (re-)construct missing data[17] to provide precise quantitative estimates of historical freshwater location and persistence, including withdrawal and replenishment, which aids forensic identification of water stress and scarcity drivers[18]. Complementary algorithms analysing satellite, drone, terrestrial and reservoir data may support real-time observation, anomaly detection, and swift short-term predictions of the hydrological cycle and weather patterns[19]. This includes quantity parameters, such as evapotranspiration[20], condensation, precipitation[21], infiltration, surface run-off, streamflow[22], subsurface flow and soil moisture[23], as well as quality factors, for example, nutrients such as phosphorus and nitrogen[24] and minerals such as flouride[25].

Such AI applications may be used for optimizing aquifer drawdown schedules to maintain water tables within sustainable limits[26] and dam-filling schedules to minimize harm to aquatic ecosystems related to upstream and downstream hydrology alterations[27]. These enable automated detection of public health hazards, including pollution plumes, waterborne disease pathogens, such as protozoa (for example, giardia), bacteria (for example, dysentery), viruses and parasitic worms[28], as well as eutrophication and harmful algal blooms[29]. Similarly, they may help detect illegal and harmful accidental activities, such as dumping or discharging hazardous chemicals into reservoirs or recreational water bodies[30].

In emergency prevention, preparedness and response, by integrating real-time rainfall data with early warning systems and control technologies, AI may monitor reservoir inflows and communicate with dam telemetry to manage safe spillway releases[31]. Such technologies may intervene in human mismanagement of dams, mitigating events like the 2011 Brisbane flood, which resulted in damages of over AU$2 billion[32]. Meanwhile, intelligent 'rain cloud to stormwater' monitoring systems, utilizing remote-sensing and community observations, could improve flood disaster response[33]. Smart groundwater management, leveraging borehole sensors, satellite data and ML, can also improve resilience through early warning action in drought prone regions, such as Kenya[34].

The drive for integrated catchment management necessitates understanding of water cycle dynamics within Earth system models (ESMs) to forecast short-period weather and long-period climate variability and associated influences on drought, desertification, storm surge, and water insecurity prevalence and intensity[35]. While still in its infancy, neural ESMs[36] may improve understanding of underpinning physics, uncover hidden parameters and expand simulation options[37].

On the basis of such forecasts, optimization algorithms could support sustainable, long-term catchment watershed and infrastructure planning. For instance, AI-enabled ESM outputs paired with geographic information systems could efficiently examine climate risks to dams and downstream damages associated with dam failure[38]. It may inform expansion of artificial water sources, such as desalination or recycled water[39], where water scarcity is predicted. Furthermore, AI-enhanced hydraulic models, characterizing drainage basin water-flow pathways and velocities, flooding footprints and tidal levels, can hone river engineering, dam weir and wall upgrades, and storm surge barrier implementation[40].

Optimized efficiency at network level

Considering growing population demands on water systems, AI may support development of new potable water, stormwater and sewerage infrastructure shaped by engineering innovations[33] alongside effective management of ageing critical assets[41].

Goal-driven AI systems, paired with virtual testing environments, may accelerate prototyping and testing of more sustainable materials[42], such as graphene-based nanomaterial membranes for desalination[43] or metal–organic frameworks for desert water harvesting[44].

Optimization algorithms could be implemented to enhance reliability, longevity and expenditure minimization—critical for public utilities—in the design, construction and upgrade of treatment and distribution facilities[45]. AI-powered digital twins of cities[46] may also help to rapidly scale water-sensitive urban design[47], including prioritized placement of bioretention systems, buffer strips and swales, infiltration trenches, porous paving, sedimentation retention, artificial wetlands, rainwater harvesting systems, and aquifer storage and recovery systems.

Together, AI, Internet of Things devices and robotics may enhance operational efficiency across water and wastewater facilities. For instance, coagulation, flocculation, sedimentation, filtration (for example, reverse osmosis) and disinfection (for example, chlorination) processes in water treatment plants could be intelligently fine-tuned to meet drinking water standards by leveraging sensor data on microbial and contaminant content of inflows and outflows at any given time[48].

Similarly, the performance of wastewater treatment plants could be advanced through self-adaptive unit processes, including preliminary screening and grit removal, primary phase-separation (for example, clarification), secondary (for example, fixed film) and tertiary treatment (for example, activated carbon) and disinfection (for example, ultraviolet light), based on the real-time organic and inorganic content of sewerage inflows and effluent discharge requirements[49]. In addition, intelligent anaerobic digesters may boost biogas and electricity production from by-product sludge[50], while intelligent classification and sorting could maximize the efficacy and safety of biosolids for agricultural reuse[51].

Smart distribution systems also provide advantages over traditional supervisory control and data acquisition systems[52]. ML models utilizing real-time data from network sensors could measure, monitor and optimize flow pressure and velocity to improve energy efficiency and operating costs by autonomously controlling and configuring water pump stations without human oversight[53]. Advanced computational systems may help to prevent harmful sewage overflows during wet weather events by fine-tuning utilization of storage in wastewater pump stations, pipes and manholes, and expedite alerts to clean-up crews where unplanned discharges do occur[54].

Intelligent technologies may transform routine maintenance activities and reduce downtime. Network leakage results in the loss of 45 billion litres of potable water per day in developing countries, which is equivalent to hydrating 180 million people, and major pipeline leaks can short circuit high-intensity cables, posing a lethal threat to people[55]. Predictive analytics, supported by sensors and cloud computing, can detect anomalies, pinpoint locations and prioritize the severity of leaks to accelerate isolations and repairs in real time[56], with efforts to forecast pipe deterioration[57] and resolve algorithm transferability across heterogeneous pipelines[58] already improving the accuracy of AI-enabled leak identification applications. ML models paired with traditeional CCTV data, used for image classification, object identification and semantic segmentation, may similarly be implemented to predict, diagnose and fix wastewater network defects and blockages[59].

Furthermore, AI could extend asset life and optimize capital expenditures by automating maintenance operations, such as the cleaning of ultrafiltration membranes in treatment plants, and designing predictive upgrade schedules based on historical and real-time asset condition assessments[60].

Improved services at end-user level

At the community level, computational intelligence could contribute to more sustainable, resilient and equitable access to water systems. For instance, AI-based analysis of historical, smart meter, satellite imagery and water consumption forecast data may inform management of conflicting sectoral and transboundary demands with precise allocations as well as monitor withdrawal compliance[61].

The agriculture sector is responsible for 70% of annual freshwater withdrawals, of which 60% (that is, 42% of global total) is wasted[62]. Targeted AI applications could help reduce this unnecessary consumption. AI may enable rapid experimentation in 'virtual farms' to determine minimum irrigation volumes and schedules to maximize crop yields under various conditions[63]. Such programmes implemented alongside digital twin and robotic technologies could enable precision farming with smart irrigation systems[64]. Autonomous processing of satellite or drone hyperspectral imaging, enabled by computer vision and ML algorithms[65], may provide detailed maps of soil moisture and crop

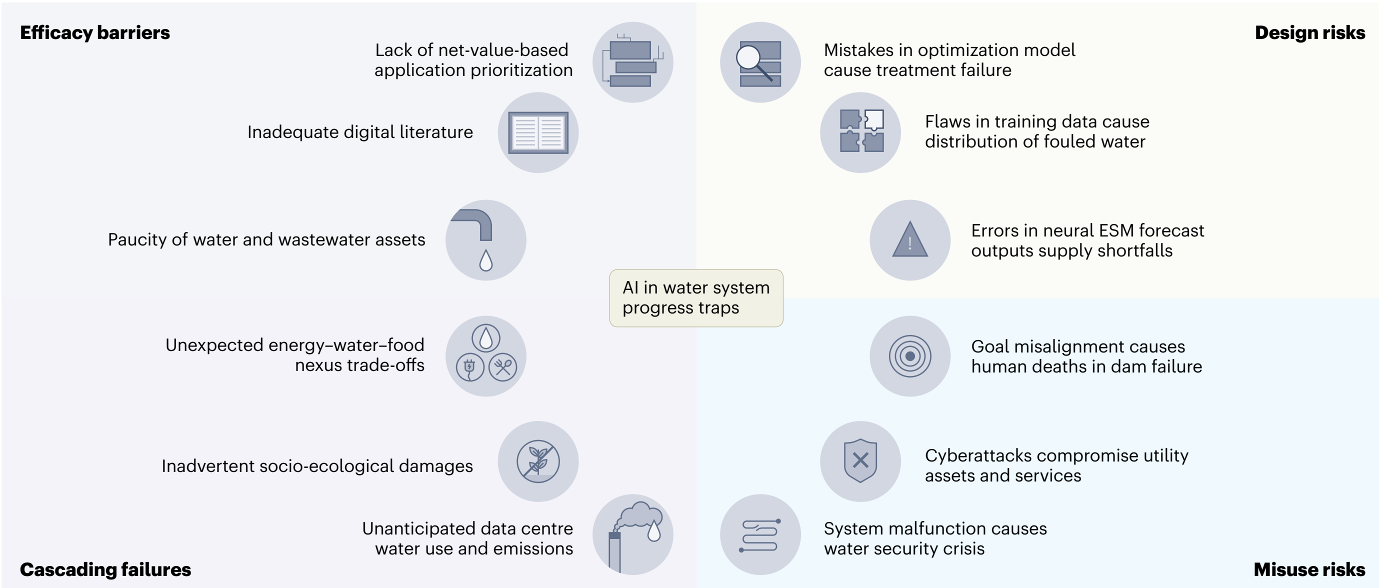

Fig. 2 | Example risks of AI across water systems that may lead to progress traps. Infrastructure and human capital barriers, direct risks related to design errors and misuse and indirect risks related to cascading system failures may undermine the potential benefits of AI if not managed responsibly.

conditions, which water authorities could use to monitor irrigation shortfalls or excesses and adjust supply allocations accordingly[66].

At the household level, smart water-saving devices, such as intelligent toilets, taps and sprinklers, may curtail household water consumption, while smart meters coupled with predictive demand and pricing analytics could provide incentives to drive behavioural change towards water conservation[67]. Furthermore, AI may control safe, decentralized potable water, stormwater and sewerage systems, such as automated rainwater tanks, domestic water recycling and home biodigesters[68]. Household units comprising real-time fluorescence sensors coupled with ML can accurately predict and intervene faecal contamination of drinking water in line with World Health Organization risk levels to prevent disease outbreaks common in both high- and low-income countries[69].

While most of these applications rely on established water systems infrastructure, AI also has potential to improve water justice. Neural ESMs and optimization algorithms could support international development agencies and governments in determining where to prioritize investment in WASH efforts to effectively address the most pressing problems while building climate resilience.

Intelligent water technologies, including off-grid facilities, such as solar-powered 'water ATMs'[70], and portable devices, such as 'smart handpumps'[71], could be distributed and monitored remotely to improve safe water access, especially for women and girls. The proliferation of personal smart phones in developing nations[72] could also enable mass communication of drinking water contamination or educational information about menstrual and hygiene practices similarly to that during the COVID-19 pandemic[73]. Portable AI systems may be trained to evaluate drinking water quality, based on free residual chlorine content, to prevent outbreak of waterborne diseases in humanitarian settlements[74].

Barriers and systemic risks

As evidenced throughout history, technological problem-solving can bring about unintended consequences, which may prove more challenging than the original problem. Given the potential proliferation of AI across water systems, as highlighted above, it is important to understand the risk landscape. To this end, we highlight issues that may undermine potentially beneficial applications of AI in water systems, including: barriers related to infrastructure and human capital; direct risks related to design errors and malicious use; and indirect exposure to cascading failures (Fig. 2).

Infrastructure and human capital barriers

AI is only as good as the systems into which it is integrated and the people responsible for its development. Many of the potential AI applications outlined above require established water systems infrastructure, supporting information and communications technology (ICT) infrastructure and domain expertise. In this vein, we highlight below anticipated instances of how infrastructure and human capital requirements may create technical and socioeconomic barriers that limit the deployment of, and give rise to unintended problems associated with, AI in the water sector.

The lack of foundational and safely managed infrastructure, including dams, treatment plants, pipes, toilets, showers and taps—currently leaving one in four people without clean drinking water and two in four people being without adequate sanitation services—will undermine the capacity of AI to address these water systems deficiencies in low-income regions, thereby precluding the most vulnerable populations from its associated benefits[75].

Even in developed countries with well-established water systems infrastructure, the complexity and cost associated with integration of advanced digital technologies, which are necessary to support AI applications, across the water sector may limit the feasibility of deploying AI in the short term[76]. Indeed, the water industry brands itself as slow and painful, more so than other sectors, when it comes to innovation owing to its long project timelines, investment constraints and conservative nature[77].

Currently, AI applications must be tailor-made to the specific context, and further algorithmic development is needed for accuracy in most situations. While there may exist successful real-world applications in other sectors, or demonstrations of water-system-specific algorithms in the literature, these are unlikely to be readily transferable to water systems in practice.

Wide-scale deployment of AI will require human capital with both AI and water sector subject matter expertise. As such, shortfalls in digital literacy among water sector workers and consumers, and lack of human capital at non-governmental organizations (NGOs) in WASH contexts, may further impair access to the potential benefits of AI in the water sector[78].

Beyond limiting the deployment of AI in the water sector, particularly in developing regions, the unequal distribution of these barriers may also give rise to unexpected problems as AI begins to proliferate in more digitally capable hubs around the world. For instance, 'digital divides' in human capital, where high-skilled labour forces are advantaged while low-skilled labour forces are disadvantaged, could reinforce global inequalities[76].

In addition, deployment of AI where barriers only partially exist but have not been fully resolved, such as where water sector practitioners develop skills sufficient to implement AI systems but insufficient to effectively identify and correct errors and malfunction, could undermine any potential benefits by giving rise to serious consequences, as outlined below.

Direct risks from design errors and misuse

Technical robustness, governance and ethics of AI[79], which are increasingly explored in other sectors such as agriculture[80], engender a distinct risk landscape in the water systems context. We provide below possible examples of how errors and biases in data and algorithmic models, including goal misalignment[81], as well as increased exposure to misuse by malicious actors, may see potential applications of AI in the water sector cause substantial social, economic and environmental harm.

At the catchment level, AI applications necessitate thorough knowledge of highly complex Earth system processes, including the water cycle and climate change. Extraction of flawed hydrological data from satellite feeds or weather forecast data from a neural ESM by an unsupervised ML model optimizing water allocations in high-competition regions could result in unexpected shortfalls of water supply for human consumption. An AI algorithm tasked with minimizing damages in the event of a dam failure could inadvertently prioritize reduction of economic losses at the expense of human life if it were accidentally programmed to optimize for wrong or overly narrow goal ranges[81].

At the network level, mistakes in the programming of intelligent wastewater treatment plant models, such as for automated maintenance of biological secondary treatment units, could cause a process crash, leading to downstream discharge of untreated effluent or upstream network overflows[82]. Such an event could expose human and marine life to untreated sewage as well as result in environmental discharge fines for the operator. Meanwhile, goal-oriented AI for optimizing water pipe maintenance based on failure probability and damage prediction could inadvertently undermine the water security of low-income populations by prioritizing infrastructure in more affluent areas[83].

At the end-user level, errors in training datasets, a sensor fault or failures in algorithmic generalizations of an ML model directing recycled water to different end-uses based on real-time quality data could result in a public health crisis if non-potable water, or worse, water contaminated with pathogens, is distributed to households for consumption. AI tools collecting household water consumption data may raise privacy concerns over use profiling, while AI-enabled water-demand-reduction applications may inadvertently employ biased 'micro-nudging' leading to undemocratic water access, undermining the dignity and autonomy of at-risk populations[83].

Incorporation of AI into water systems collectively heightens the risk of network-wide failures. Specifically, overreliance on AI, either in critical components of water systems or through high coupling between water systems components, could lead to systemic risks, where the isolated risks described above could potentially result in compromising an entire utility's assets and services.

Furthermore, water is already the subject of intra- and international geopolitics and corporate competition—where appropriation of fresh water is associated with agricultural land grabbing estimated at 310 billion cubic metres of green water (that is, rainwater) and 140 billion cubic metres of blue water (that is, irrigation water) per year[84]—and cybersecurity is of growing concern given the recent increase in events of system compromise across the sector[85]. While previous generations of cyberattacks, including distributed denial-of-service (DDoS), ransomware, structured query language (SQL) injection and Trojan horse, were disruptive, the presence of embedded AI with minimal human oversight may provide hackers the opportunity to take full control of highly interconnected systems[86].

Such network-wide failures may put entire communities at risk of water insecurity and could quickly translate to humanitarian crises and conflicts due to the relatively localized nature of water resources that—unlike the established global supply chains of energy and food resources—are not easily substituted or traded en masse internationally[87].

Cascading system failures

The direct failures of AI within water systems highlighted above may indirectly cascade into local and regional catastrophes outside the water industry. In highlighting such instances below with presumed examples, we note that these failures could occur independently but that flawed AI, alongside lack of human oversight, may exacerbate the frequency and severity of such indirect risks.

Heavy reliance on AI could create fragile interdependencies between critical infrastructure systems. Scenarios could include three-way coupling where cloud computing underpins AI in water systems and energy systems, while water cooling is needed for cloud computing and electricity generation, and grid power is required for water systems and data centre operation. Such tight connectivity amplifies the risk of accidental failures or malicious cyberattacks cascading across systems and makes recovery from otherwise isolated events substantially more challenging.

Notwithstanding the above, seemingly successful applications of AI in water systems may have unexpected negative repercussions. Inadvertent socio-ecological consequences may occur where an AI-enabled digital twin optimizes the processes of a seawater desalination facility but does not accurately account for brine effects on ecosystems at the discharge point, resulting in damage to the marine environment and biodiversity[88]. Similarly, energy and food security problems could arise where AI models implemented by the water industry are biased such that water–energy–food nexus trade-offs are not appropriately represented[89].

Furthermore, while ML advances can reduce computational energy demand[90], expanded use of inefficient AI systems may increase the power intensity of data centres, thereby increasing water usage in liquid cooling technologies and greenhouse gas emissions that feedback to undermine water security[91].

Responsible AI to prevent progress traps

To ensure that potential applications of AI in water systems realize intended benefits and do not unintentionally perpetuate progress traps, we make three sets of recommendations for the water industry to safely deploy this nascent digital technology. Detailed in turn below, the first addresses gaps in infrastructure and digital literacy, the second outlines technical mechanisms for trustworthy AI, and the third proposes a six-layer framework to guide benefit and risk assessment of AI applications across water systems in practice.

Address gaps in infrastructure and literacy

Where foundational catchment, treatment and distribution assets and hygiene facilities are lacking, there is little potential for AI to solve current water systems deficiencies. Governments, development funds, philanthropists and start-ups seeking to further WASH efforts in developing countries must consider social equity and economic efficiency when evaluating AI applications in place of, or complementary to, 'brick and mortar' projects[92].

In developed countries, water utilities should ensure that adequate ICT infrastructure, such as sensors and cloud computing capabilities, are accounted for in AI application planning. Furthermore, water utilities must develop clear strategies and architectures, such as application programming interfaces, for systems integration and interoperability, given the need to maintain legacy infrastructure, including both physical structures and electromechanical equipment as well as heterogeneous data siloes containing operations critical information, alongside new digital technologies[93].

Where AI applications are deemed appropriate, the water industry will need to manage the upskilling, reskilling and new skilling of its workforce, to ensure that the sector is equipped with the human capital necessary to design, operate and manage AI systems. Professional associations and trade unions should engage with academic institutions and NGOs with expertise in AI to develop new educational courses and certifications.

To this end, it is important that AI in water systems is explainable. On the one hand, explainable AI is necessary for domain experts to meaningfully confirm, challenge and transfer knowledge across use-cases[94]. Meanwhile, the move from research-based settings to practical applications makes deployment and communication of white box (that is, algorithms providing understandable results), rather than black box (that is, algorithms that can hardly be understood even by domain experts), algorithms by AI experts essential to ensure that water industry practitioners and civilian end-users lacking expertise can trust and interact with AI in water systems by understanding its functionality[95].

Finally, the water industry must develop and enact sectoral-specific legislation, regulations and policies fit for purpose in handling the nuances of AI in water systems, especially when it comes to matters of technical standards and transparency, human agency and oversight, safety and security, accountability and liability, and diversity and inclusion. Where such governance frameworks are lacking, utilities looking to implement AI across their water systems should give due consideration to evolving issues such as insurances and liabilities related to failure of AI systems.

Establish mechanisms for trustworthy AI

At the current time, our understanding of AI is in a state of rapid development, so the water industry must keep abreast of issues related to technical robustness, governance and ethics as the field of AI safety evolves (for example, see the European Union's ethics guidelines for trustworthy AI)[96]. Here we outline institutional, software and hardware mechanisms that the water industry should develop and maintain in its 'toolbox' to ensure responsible deployment of AI across water systems.

Institutional mechanisms shape knowledge, incentives and accountability[97]. Routine red team exercises, where cybersecurity experts are internally engaged to find vulnerabilities in AI systems, should be conducted by water utilities to keep ahead of malicious actors seeking to compromise critical water systems infrastructure. Bug, bias and safety bounty schemes can also be used to incentivize external stakeholders and beneficiaries to disclose problems with AI systems in practice. Such schemes may be especially useful when it comes to mapping systemic exposures to cascading socio-ecological consequences and water–energy–food nexus trade-offs. Beyond these exercises, as the water industry gains real-world experience with AI, collaborative cross-sector knowledge banks of implementation best practices, safety incidents and lessons learned should be maintained.

Software mechanisms address specificities, understanding and oversight of AI systems themselves[97]. The water industry should work with academic experts to establish design standards, interpretability manuals and user-testing methods to ensure reproducibility, privacy preservation and verification. Processes of human-centred design, safe by design and secure by design could help mitigate several of the risks outlined above. Cross-sector knowledge banks of validated standard AI source codes, for typical applications such as pipe leakage detection and treatment process optimization, could be maintained to accelerate best practice. Water utilities must also maintain audit trails of problem definition, design, development and operation of AI systems and should have these traceable logs analysed by expert third-party auditors to maximize the capture of incidents and lessons learned.

Hardware mechanisms confront the capability, accessibility and reliability of physical resources[98]. Water utilities must ensure that AI applications incorporate fail-safes enabling automated or human-initiated shutdown and workarounds to mitigate potential catastrophes caused by malfunctioning or compromised systems. Quality control inspectors should be engaged to regularly measure and report on performance of AI and smart cyber-physical systems on a case-by-case and water-system-wide basis. Water sector stakeholders should also consider establishing open-access research and development partnerships with academic experts, who generally lack access to commercial scale hardware, to accelerate collaborative cross- and intersector advances in trusted AI for water systems.

Framework for benefit and risk assessment

The water industry must establish a transparent framework for responsible deployment of AI in relation to control of its infrastructure and services that provides balanced assessment of benefits and risks[99]. AI is not an end goal itself and should be treated as a technological response to clearly defined problems. 'Don't start with moon shots': a holistic approach must include thorough understanding of water systems deficiencies, evaluation of typical AI systems that will safely address such deficiencies, and staged prototyping, pilot and roll-out processes[100].

To this end, we propose an exemplar six-layer framework (Table 1), which elaborates general concepts from the trustworthy AI guidelines on technical robustness, governance and ethics, that addresses theoretical screening, proof of concept and practical scale-up considerations for the deployment of AI in water systems. The example considerations provided here are intended to inspire water sector practitioners with the basis of a 'live scorecard' to qualify the net value proposition of a given AI application pre-, mid- and post-implementation in the real world.

Closing reflections

The world is not on track to meet Sustainable Development Goal 6. Over 1.6 million people are dying annually from unsafe and inaccessible drinking water, stormwater and sewerage services, and climate change is expected to exacerbate water-related issues. In response, AI has been proposed as the latest technological innovation to help address water systems deficiencies. However, technology alone cannot solve water supply and wastewater disposal problems. Poorly managed proliferation of AI across water systems may give rise to progress traps that could further undermine and complicate water security.

As such, this Perspective sought to cross-pollinate domain-siloed knowledge, and contextualize synthesized insights, from the technical water systems and AI safety literature, to raise awareness among academics, water and AI industry practitioners and layman end-users of the need to prioritize 'responsible deployment' of AI in water systems to mitigate risks.

Given that AI has not yet proliferated, and its deployment is particularly nascent in the water sector, empirical data on real-world applications is relatively scarce. As such, the example water-system-wide AI applications outlined here, to highlight potential system-wide reach of AI, are based on demonstrations of AI algorithms or isolated case studies in the academic literature. Similarly, the examples of AI risks provided are speculative, albeit informed by cutting-edge AI safety literature.

Notwithstanding the above, one real-world example of a failed AI application causing harm to people was reported in November 2022[101]. In this instance, the public health department of Toronto, Canada, replaced its traditional method of using day-old laboratory tests with an ML-based predictive water quality assessment tool to determine whether the water quality at local beaches was safe for swimming. Rather than being more accurate as purported, the ML tool identified

Table 1 | Framework for holistic benefit and risk assessment to support water utilities in the responsible deployment of AI across water systems

Stage of AI deployment		Benefit maximization considerations	Risk minimization considerations
(1) Theoretical screening	**(1a) Needs assessment** Determining where AI interventions might be needed to provide benefits in the water system	**(1a.i)** Have needs for AI intervention within the water system been scanned for using both a top-down (that is, considering the water sector/utility strategic needs) and bottom-up (that is, considering the end-user needs) lens across the catchment, network and end-user levels? **(1a.ii)** Have any identified needs (that is, benefits) been qualitatively or quantitatively ranked, considering technical (for example, technology readiness level) economic (for example, associated capex and opex) and social (for example, water security) factors, to determine an order of prioritization? **(1a.iii)** Do relevant water system stakeholders (for example, regulatory bodies, water association, utility management, engineers and operators, industrial, commercial and residential end-users) recognize the identified need as a real, high-priority need?	**(1a.iv)** Is AI an appropriate intervention or would an alternative, more established and less risky, intervention be more appropriate? **(1a.v)** Is the water utility (that is, AI system owner) familiar with the general risks associated with the failure or compromise of AI systems (that is, as outlined in this paper and the AI safety literature), and are they prepared to mitigate and/or bear responsibility for such risks? **(1a.vi)** Have negative consequences (for example, employee redundancy) or exposure to novel contexts associated with successful implementation of the AI intervention been considered? Have any such consequences been justified or clear strategies been put in place to mitigate harm (for example, retraining programmes)?
	(1b) Efficacy assessment Determining which AI system is most appropriate to achieve the needed benefits	**(1b.i)** Have similar AI interventions previously demonstrated the needed benefits in the water sector? If so, are the algorithms for/developers of the AI systems accessible for use/consultation? If not, satisfy (1b,iii). **(1b.ii)** Have similar AI interventions previously demonstrated the needed benefits in other sectors (for example, energy sector) or in research (that is, commercial, NGO or academic) studies? If so, are the algorithms for/developers of the AI systems accessible for use/consultation? If not, satisfy (1b,iii). **(1b.iii)** Have AI experts (including AI developer and safety professionals) been consulted and validated the theoretical plausibility of an AI system design to achieve the needed benefits in this specific context?	**(1b.iv)** Is the water system, and associated ICT, infrastructure sufficiently developed to support integration of the AI system? Is the water utility prepared for the cost, complexity and any evolving legal requirements associated with integration of advanced digital technologies across their assets? **(1b.v)** Does the water utility have in-house access to the human capital (that is, employees with skills and knowledge) necessary to design, test, implement, operate and manage AI systems should external experts/consultants become unavailable/prohibitively expensive? If not, have clear strategies been put in place to upskill, reskill and new skill its workforce? **(1b.vi)** Can the data requirements of the AI system be met reliably (that is, availability, accessibility, quality)? Does the water utility have clear measures in place for data privacy and security?
(2) Proof of concept	**(2a) Validity assessment** Appraising the performance of the prototype AI system prototype in a laboratory setting	**(2a.i)** Have relevant water system stakeholders been consulted on the specific benefits to be provided by the AI system and indicators of success specific to their context? **(2a.ii)** Have a clear set of testing metrics (for example, precision, accuracy and explainability criteria) been established against which to measure the deployability of the AI system? Have a clear set of deployment metrics (for example, water quantity, water quality, cost saving targets and digital literacy targets) been established against which to measure the success of the proposed AI intervention? **(2a.iii)** Have AI experts qualified that the AI system prototype's performance has satisfied the testing criteria?	**(2a.iv)** Have relevant water system stakeholders been consulted on the risks that may be introduced by the AI system specific to their context, and is the water utility prepared to mitigate and/or bear responsibility for such risks? **(2a.v)** Does the AI system comply with any established design standards, interpretability manuals and user-testing methods, and have processes of human-centred design, safe by design and secure by design been adopted? Is the AI system fit for purpose, specific to the water system context, and capable of accommodating conditions that fall outside the training regime? **(2a.vi)** Have clear data architectures, including concept design of an application programming interface, been established to ensure successful interoperability between legacy infrastructure and the AI system? Have clear cybersecurity strategies been established to minimize the AI-enabled water system's exposure to malicious actors?
	(2b) Feasibility assessment Appraising the performance of the pilot AI system in a comparable real-world setting	**(2b.i)** Have a range of water system parameter permutations and combinations (for example, biochemical parameters in an anaerobic wastewater treatment unit) been tested to determine optimum operating conditions for the AI system? **(2b.ii)** Are mechanisms in place for water system stakeholders to provide real-time feedback on the AI system once it is deployed? Are clear processes/resources in place for the water utility to action response to such feedback to improve the AI system? **(2b,iii)** Have AI experts and water system stakeholders qualified that the AI system pilot's performance has satisfied the testing criteria?	**(2b.iv)** Can the AI system provide explanations for decisions, actions or accidents (that is, is it explainable), in the form of accurate and actionable information about its inputs, internal decision-making and outputs that enable meaningful human oversight and intervention? **(2b.v)** Are outcomes of the AI system consistent across variable real-world operating conditions? Have the causes of any discrepancies between the prototype and pilot system performance been identified and resolved? **(2b.vi)** Have AI system fail-safe and workaround mechanisms been implemented, tested and qualified? Can the system fail safely in the presence of technical disruptions, lack of human maintenance, extreme environmental conditions (for example, floods) or malicious interference?

Table 1 (continued) | Framework for holistic benefit and risk assessment to support water utilities in the responsible deployment of AI across water systems

Stage of AI deployment		Benefit maximization considerations	Risk minimization considerations
(3) Practical scale-up	**(3a) Usability assessment** Evaluating whether the AI system is being used as intended	**(3a.i)** Are water system stakeholders utilizing the mechanisms to provide feedback on the AI system and are they satisfied with the water utility's provision of these mechanisms and any response to any feedback? **(3a.ii)** Has the digital literacy of water system stakeholders interacting with the AI system improved such that they have developed a working understanding of its functionality? **(3a.iii)** Has the water utility engaged in collaborations with water associations and academic institutions and NGOs with AI expertise to develop new professional development courses and certifications for water sector practitioners?	**(3a.iv)** Have red team cybersecurity exercises regularly been carried out, and are bug, bias and safety bounty schemes being maintained in line with contextual changes and AI system upgrades? Have any institutional-related incidents or near-misses been addressed in a timely manner to mitigate repeat occurrence? **(3a.v)** Have software audit trails been kept and traceable logs analysed by third-party expert auditors on a regular basis? Have any software-related incidents or near-misses been addressed in a timely manner to mitigate repeat occurrence? **(3a.vi)** Have hardware systems undergone regular maintenance and quality assurance and quality control (QA/QC) inspections, and have fail-safe and workaround mechanisms regularly been tested? Have any hardware-related incidents or near-misses been addressed in a timely manner to mitigate repeat occurrence?
	(3b) Success assessment Evaluating whether the AI system is performing as intended	**(3b.i)** Have AI experts and water system stakeholders qualified that the AI system is fulfilling the needed benefits, that is, meeting its deployment metric targets? Have any unintended benefits been realized? **(3b.ii)** Have realized benefits, best practices and associated information (for example, validated source code) been recorded in knowledge banks for sharing across the water sector as well as with other sectors and academics? **(3b.iii)** Has the AI system regularly been reviewed by AI experts and has the water utility implemented recommended but non-essential updates to maximize its performance against the deployment metrics?	**(3b.iv)** Have anticipated risks (for example, direct failures in the water system, cascading failures into the energy or water sector or the environment), vulnerabilities (for example, exposure to cyberattacks) and new fragilities (for example, additional reliance on external ICT assets) been mitigated and/or related incidents resolved successfully? Have unanticipated risks been mitigated and/or related incidents resolved successfully? **(3b.v)** Have benefits of the AI system been realized equally by stakeholders across the water system? Have any incidents or near-misses related to bias or inequality been addressed in a timely manner to mitigate repeat occurrence? **(3b.vi)** Have unanticipated and anticipated risk-related incidents, near-misses and lessons learned (for example, case studies) been recorded in knowledge banks for sharing across the water sector as well as with other sectors and academics?

only about 30% of unsafe beach water days, resulting in 50 instances of public bathers being exposed to dangerous bacteria levels over the summer. This highlights the very real harm that AI failure may cause if responsible deployment principles are not prioritized and effectively executed.

As empirical data on successful and unsuccessful applications of AI across water systems becomes more readily available, we encourage researchers and practitioners to rigorously database and evaluate such information to build a thorough understanding of the real benefits and real risks, necessary to evolve risk management practices as AI proliferates. We hope that the framework conceptualized here provides a basis from which multidisciplinary academics and water and AI industry practitioners can develop proactive and critical risk management practices, informed by participatory approaches that engage and educate end-users.

Finally, the water industry must take a tiered approach to risk anticipation and mitigation to ensure responsible deployment of AI in water systems, including addressing barriers related to infrastructure and digital literacy, establishing institutional, software and hardware mechanisms for trustworthy AI, and prioritizing applications based on rigorous benefit and risk assessment. With US$6.3 billion projected investment in AI water technologies, we urge the water sector, particularly the larger and more developed utilities driving the foray into digital water, to allocate a substantial portion of this funding away from purely technical capacity building to AI safety initiatives that will help ensure potential benefits of AI in water systems are realized at scale.

References

1. Hosseiny, S. H., Bozorg-Haddad, O. & Bocchiola, D. in *Economical, Political, and Social Issues in Water Resources* (Ed. Bozorg-Haddad, O.) 189–216 (Elsevier, 2021).
2. Schreiber, K. J. & Rojas, J. L. The Puquios of Nasca. *Lat. Am. Antiq.* **6**, 229–254 (1995).
3. Coningham, R. & Young, R. *The Archaeology of South Asia: From the Indus to Asoka, c.6500 BCE–200 CE* 101–278 (Cambridge Univ. Press, 2015).
4. Wright, R. *A Short History of Progress* (House of Anansi, 2004).
5. Price, M. Origins of ancient Rome's famed pipe plumbing system revealed in soil samples. *Science* (28 August 2017).
6. Bierkens, M. F. P. & Wada, Y. Non-renewable groundwater use and groundwater depletion: a review. *Environ. Res. Lett.* **14**, 063002 (2019).
7. Gordon, L., Dunlop, M. & Foran, B. Land cover change and water vapour flows: learning from Australia. *Phil. Trans. R. Soc. Lond. B. Biol. Sci. B* **29**, 358 (2003).
8. Liu, W., Iordan, C. M., Cherubini, F., Hu, X. & Fu, D. Environmental impacts assessment of wastewater treatment and sludge disposal systems under two sewage discharge standards: a case study in Kunshan, China. *J. Clean. Prod.* **287**, 125046 (2021).
9. Elsaid, K. et al. Environmental impact of desalination technologies: a review. *Sci. Total Environ.* **748**, 141528 (2020).
10. da Silva, G. C. X. et al. Environmental impacts of dam reservoir filling in the East Amazon. *Front. Water* https://doi.org/10.3389/frwa.2020.00011 (2020).

11. UN-Water *Summary Progress Update 2021: SDG 6—Water and Sanitation for All* (United Nations, 2021).
12. IPCC *Climate Change 2022*: Impacts, Adaptation and Vulnerability (eds Pörtner, H.-O. et al.) (Cambridge Univ. Press, 2022).
13. Russell, S. J. & Norvig, P. *Artificial Intelligence: A Modern Approach* (Prentice Hall, 2020).
14. Rozos, E. Machine learning, urban water resources management and operating policy. *Resources* **8**, 173 (2019).
15. *The Sustainable Development Goals Report 2022* (United Nations, 2022).
16. Yamazaki, D. & Trigg, M. A. The dynamics of Earth's surface water. *Nature* **540**, 348–349 (2016).
17. Kadow, C., Hall, D. M. & Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **13**, 408–413 (2020).
18. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).
19. Larson, A. A clearer view of Earth's water cycle via neural networks and satellite data. *Nat. Rev. Earth Environ.* **3**, 361 (2022).
20. Koppa, A., Rains, D., Hulsman, P., Poyatos, R. & Miralles, D. G. A deep learning-based hybrid model of global terrestrial evaporation. *Nat. Commun.* **13**, 1912 (2022).
21. Espeholt, L. et al. Deep learning for twelve hour precipitation forecasts. *Nat. Commun.* **13**, 5145 (2022).
22. Muste, M., Kim, D. & Kim, K. A flood-crest forecast prototype for river floods using only in-stream measurements. *Commun. Earth Environ.* **3**, 78 (2022).
23. Vereecken, H. et al. Soil hydrology in the Earth system. *Nat. Rev. Earth Environ.* **3**, 573–587 (2022).
24. Shen, L. Q., Amatulli, G., Sethi, T., Raymond, P. & Domisch, S. Estimating nitrogen and phosphorus concentrations in streams and rivers, within a machine learning framework. *Sci. Data* **7**, 161 (2020).
25. Podgorski, J. & Berg, M. Global analysis and prediction of fluoride in groundwater. *Nat. Commun.* **13**, 4232 (2022).
26. Sharafati, A., Asadollah, S. B. H. S. & Neshat, A. A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. *J. Hydrol.* **591**, 125468 (2020).
27. Zaniolo, M., Giuliani, M., Sinclair, S., Burlando, P. & Castelletti, A. When timing matters—misdesigned dam filling impacts hydropower sustainability. *Nat. Commun.* **12**, 3056 (2021).
28. Chakkaravarthy, G. V. & Lavanya, R. in *Integrating AI in IoT Analytics on the Cloud for Healthcare Applications* (ed. Jeya Mala, D.) 57–66 (IGI Global, 2022).
29. Mozo, A. et al. Chlorophyll soft-sensor based on machine learning models for algal bloom predictions. *Sci Rep.* **12**, 13529 (2022).
30. Massarelli, C., Campanale, C. & Uricchio, V. F. in *IoT Applications Computing* (eds Singh, I. et al.) Ch. 10 (IntechOpen, 2021).
31. Zarei, M. et al. Machine-learning algorithms for forecast-informed reservoir operation (FIRO) to reduce flood damages. *Sci. Rep.* **11**, 24295 (2021).
32. Riga, R. Brisbane 2011 flood victims win $440 million in class action partial settlement over operation of Wivenhoe Dam. *ABC News* https://web.archive.org/web/20130127022836/http://www.australiangeographic.com.au/journal/the-worst-floods-in-australian-history.htm (2021).
33. Oksen, P. & Favre, L. *Innovative Technology in the Water, Sanitation and Hygiene (WASH) Sector* (WIPO, 2020); https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gc_20_1.pdf
34. Fankhauser, K. et al. Estimating groundwater use and demand in arid Kenya through assimilation of satellite data and in-situ sensors with machine learning toward drought early action. *Sci. Total Environ.* **831**, 154453 (2022).
35. Bauer, P. et al. The digital revolution of Earth-system science. *Nat. Comput. Sci.* **1**, 104–113 (2021).
36. Irrgang, C. et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **3**, 667–674 (2021).
37. Hess, P., Drüke, M., Petri, S., Strnad, F. M. & Boers, N. Physically constrained generative adversarial networks for improving precipitation fields from Earth system models. *Nat. Mach. Intell.* **4**, 828–839 (2022).
38. Fecht, S. Using artificial intelligence to locate risky dams. *Columbia Climate School News* https://news.climate.columbia.edu/2018/08/23/artificial-intelligence-find-risky-dams/ (2018).
39. van Vliet, M. T. H. et al. Global water scarcity including surface water quality and expansions of clean water technologies. *Environ. Res. Lett.* **16**, 24020 (2021).
40. Hosseiny, H., Nazari, F., Smith, V. & Nataraj, C. A framework for modeling flood depth using a hybrid of hydraulics and machine learning. *Sci. Rep.* **10**, 8222 (2020).
41. Perera, D., Smakhtin, V., Williams, S., North, T. & Curry, A. *Ageing Water Storage Infrastructure: An Emerging Global Risk* (UNU-INWEH, 2021); https://inweh.unu.edu/wp-content/uploads/2021/01/Ageing-Water-Storage-Infrastructure-An-Emerging-Global-Risk_web-version.pdf
42. Maleki, R. et al. Materials discovery of ion-selective membranes using artificial intelligence. *Commun. Chem.* **5**, 132 (2022).
43. Seo, D. H. et al. Anti-fouling graphene-based membranes for effective water desalination. *Nat. Commun.* **9**, 683 (2018).
44. Kim, H. et al. Adsorption-based atmospheric water harvesting device for arid climates. *Nat. Commun.* **9**, 1191 (2018).
45. Pham Vu Hong, S. & Nguyen Thanh, V. Application of artificial intelligence algorithm to optimize the design of water distribution system. *Int. J. Constr. Manage.* https://doi.org/10.1080/15623599.2022.2101593 (2022).
46. Tzachor, A., Sabri, S., Richards, C. E., Acuto, M. & Rajabifard, A. Potential and limitations of digital twins to achieve the Sustainable Development Goals. *Nat. Sustain.* **5**, 822–829 (2022).
47. Wong, T. H. F., Rogers, B. C. & Brown, R. R. Transforming cities through water-sensitive principles and practices. *One Earth* **3**, 436–447 (2020).
48. Li, L., Rong, S., Wang, R. & Yu, S. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: a review. *Chem. Eng. J.* **405**, 126673 (2021).
49. Nguyen, X. C. et al. in *Current Developments in Biotechnology and Bioengineering: Advances in Biological Wastewater Treatment Systems* (eds Bui, X.-T. et al.) 587–608 (Elsevier, 2022).
50. Sakiewicz, P., Piotrowski, K., Ober, J. & Karwot, J. Innovative artificial neural network approach for integrated biogas—wastewater treatment system modelling: effect of plant operating parameters on process intensification. *Renew. Sustain. Energy Rev.* **124**, 109784 (2020).
51. Rani, A., Snyder, S. W., Kim, H., Lei, Z. & Pan, S.-Y. Pathways to a net-zero-carbon water sector through energy-extracting wastewater technologies. *npj Clean Water* **5**, 49 (2022).
52. How AI is taking SCADA systems to the next level. *Veolia* https://blog.veolianorthamerica.com/how-ai-taking-scada-systems-to-next-level (2022).
53. Increasing water pump efficiency using artificial intelligence. *Melbourne Water* https://www.melbournewater.com.au/water-data-and-education/news/research-and-innovation/increasing-water-pump-efficiency-using (2021).
54. Balla, K. M., Bendtsen, J. D., Schou, C., Kallesøe, C. S. & Ocampo-Martinez, C. A learning-based approach towards the data-driven predictive control of combined wastewater networks—an experimental study. *Water Res.* **221**, 118782 (2022).

55. Machine learning can help protect urban water. Here's how. *World Economic Forum* https://www.weforum.org/agenda/2022/04/how-to-prevent-urban-water-stress-through-machine-learning/ (2022).
56. Vanijjirattikhan, R. et al. AI-based acoustic leak detection in water distribution systems. *Results Eng.* **15**, 100557 (2022).
57. Dawood, T., Elwakil, E., Novoa, H. M. & Delgado, J. F. G. Artificial intelligence for the modeling of water pipes deterioration mechanisms. *Autom. Constr.* **120**, 103398 (2020).
58. Zhou, B., Lau, V. & Wang, X. Machine-learning-based leakage-event identification for smart water supply systems. *IEEE Internet Things J.* **7**, 2277–2292 (2020).
59. Fu, G., Jin, Y., Sun, S., Yuan, Z. & Butler, D. The role of deep learning in urban water management: a critical review. *Water Res.* **223**, 118973 (2022).
60. *Harnessing the Fourth Industrial Revolution for Water* (World Economic Forum, 2018).
61. Stańczyk, J., Kajewska-Szkudlarek, J., Lipiński, P. & Rychlikowski, P. Improving short-term water demand forecasting using evolutionary algorithms. *Sci. Rep.* **12**, 13522 (2022).
62. Food and Agriculture Organization. Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal). *World Bank Data* https://data.worldbank.org/indicator/ER.H2O.FWAG.ZS (2018).
63. Tzachor, A., Richards, C. E. & Jeen, S. Transforming agrifood production systems and supply chains with digital twins. *npj Sci. Food* **6**, 47 (2022).
64. Dehghanisanij, H., Emami, H., Emami, S. & Rezaverdinejad, V. A hybrid machine learning approach for estimating the water-use efficiency and yield in agriculture. *Sci. Rep.* **12**, 6728 (2022).
65. Bauer, A. et al. Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: a case study of lettuce production. *Hortic. Res.* **6**, 70 (2019).
66. García, L., Rodríguez, J. D., Wijnen, M. & Pakulski, I. *Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector* (World Bank, 2016).
67. Cominola, A. et al. Long-term water conservation is fostered by smart meter-based feedback and digital user engagement. *npj Clean Water* **4**, 29 (2021).
68. *Harnessing Artificial Intelligence for the Earth* (World Economic Forum, 2018).
69. Bedell, E., Harmon, O., Fankhauser, K., Shivers, Z. & Thomas, E. A continuous, in-situ, near-time fluorescence sensor coupled with a machine learning model for detection of fecal contamination risk in drinking water: design, characterization and field validation. *Water Res.* **220**, 118644 (2022).
70. Andres, L., Boateng, K., Borja-Vega, C. & Thomas, E. A review of in-situ and remote sensing technologies to monitor water and sanitation interventions. *Water* https://doi.org/10.3390/w10060756 (2018).
71. Swan, A., Cooper, N., Gamble, W. & Pritchard, M. Using smart pumps to help deliver universal access to safe and affordable drinking water. *Proc. Inst. Civ. Eng. Eng. Sustain.* **171**, 277–285 (2017).
72. Aiken, E., Bellue, S., Karlan, D., Udry, C. & Blumenstock, J. E. Machine learning and phone data can improve targeting of humanitarian aid. *Nature* **603**, 864–870 (2022).
73. Pandey, R. et al. A machine learning application for raising WASH awareness in the times of COVID-19 pandemic. *Sci. Rep.* **12**, 810 (2022).
74. De Santi, M. et al. Modelling point-of-consumption residual chlorine in humanitarian response: can cost-sensitive learning improve probabilistic forecasts? *PLoS Water* **1**, e0000040 (2022).
75. *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years Into the SDGs* (World Health Organization and United Nations Children's Fund, 2021).
76. Guenat, S. et al. Meeting sustainable development goals via robotics and autonomous systems. *Nat. Commun.* **13**, 3559 (2022).
77. Slow pace of innovation continues to frustrate supply chain. *British Water* https://www.britishwater.co.uk/news/602860/Slow-pace-of-innovation-continues-to-frustrate-supply-chain-.htm (2022).
78. Without universal AI literacy, AI will fail us. *World Economic Forum* https://www.weforum.org/agenda/2022/03/without-universal-ai-literacy-ai-will-fail-us/ (2022).
79. Brundage, M. et al. The malicious use of artificial intelligence: forecasting, prevention, and mitigation. *Apollo* https://doi.org/10.17863/CAM.22520 (2018)
80. Tzachor, A., Devare, M., King, B., Avin, S. & Ó hÉigeartaigh, S. Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. *Nat. Mach. Intell.* **4**, 104–109 (2022).
81. Amodei, D. et al. Concrete problems in AI safety. Preprint at https://doi.org/10.48550/arXiv.1606.06565 (2016).
82. Trávníček, P., Junga, P., Kotek, L. & Vítěz, T. Analysis of accidents at municipal wastewater treatment plants in Europe. *J. Loss Prev. Process Ind.* **74**, 104634 (2022).
83. Doorn, N. Artificial intelligence in the water domain: opportunities for responsible use. *Sci. Total Environ.* **755**, 142561 (2021).
84. Rulli, M. C., Saviori, A. & D'Odorico, P. Global land and water grabbing. *Proc. Natl Acad. Sci. USA* **110**, 892–897 (2013).
85. Gallagher, R. UK water supplier hit by 'extremely concerning' cyberattack. *Bloomberg*. https://www.bloomberg.com/news/articles/2022-08-17/uk-water-supplier-hit-by-extremely-concerning-cyberattack (2022).
86. Taddeo, M., McCutcheon, T. & Floridi, L. Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nat. Mach. Intell.* **1**, 557–560 (2019).
87. Milne, S. How water shortages are brewing wars. *BBC Future* https://www.bbc.com/future/article/20210816-how-water-shortages-are-brewing-wars (2021).
88. Omerspahic, M., Al-Jabri, H., Siddiqui, S. A. & Saadaoui, I. Characteristics of desalination brine and its impacts on marine chemistry and health, with emphasis on the Persian/Arabian Gulf: a review. *Front. Mar. Sci.* https://doi.org/10.3389/fmars.2022.845113 (2022).
89. Zaidi, S. M. A. et al. Machine learning for energy–water nexus: challenges and opportunities. *Big Earth Data* **2**, 228–267 (2018).
90. Patterson, D. A. et al. Carbon emissions and large neural network training. Preprint at https://arxiv.org/abs/2104.10350 (2021).
91. Kaack, L. H. et al. Aligning artificial intelligence with climate change mitigation. *Nat. Clim. Change* **12**, 518–527 (2022).
92. Alonso, C., Kothari, S. & Rehman, S. *How Artificial Intelligence Could Widen the Gap between Rich and Poor Nations* (International Monetary Fund, 2020).
93. Sarni, W., White, C., Webb, R., Cross, K. & Glotzbach, R. *Digital Water: Industry lEaders Chart the Transformation Journey* (IWA, 2019); https://iwa-network.org/wp-content/uploads/2019/06/IWA_2019_Digital_Water_Report.pdf
94. Barredo Arrieta, A. et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).
95. Gunning, D. et al. XAI—explainable artificial intelligence. *Sci. Robot.* **4**, eaay7120 (2019).
96. Ethics guidelines for trustworthy AI. *European Commission* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2023).
97. Avin, S. et al. Filling gaps in trustworthy development of AI. *Science* **374**, 1327–1329 (2021).
98. Brundage, M. et al. *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims* (2020).

99. *AI Governance: A Holistic Approach to implement Ethics into AI* (World Economic Forum, 2019).
100. Davenport, T. H. & Ronanki, R. Artificial intelligence for the real world: don't start with moon shots. *Harvard Business Review* https://www.hbsp.harvard.edu/product/R1801H-PDF-ENG (2018).
101. Martineau, P. Toronto tapped artificial intelligence to warn swimmers. The experiment failed. *The Information* https://www.theinformation.com/articles/when-artificial-intelligence-isnt-smarter (2022).

Acknowledgements
This paper was made possible through the support of a grant from Templeton World Charity Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation. We thank K. Atanasova for assistance with production of Figs. 1 and 2.

Author contributions
C.E.R., A.T., S.A. and R.F. jointly developed and contributed to the writing of this paper.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence should be addressed to Catherine E. Richards or Asaf Tzachor.

Peer review information *Nature Water* thanks Evan Thomas, Guangtao Fu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.