sustainability

MDPI

Article

Data Science and AI for Sustainable Futures: Opportunities and Challenges

Gavin Shaddick [1,*], David Topping [2], Tristram C. Hales [3], Usama Kadri [4], Joanne Patterson [5], John Pickett [6], Ioan Petri [7], Stuart Taylor [6], Peiyuan Li [8], Ashish Sharma [8], Venkat Venkatkrishnan [8], Abhinav Wadhwa [8], Jennifer Ding [9], Ruth Bowyer [10] and Omer Rana [11,*]

1 College of Physical Sciences and Engineering, Cardiff University, Cardiff CF10 3AT, UK
2 School of Earth and Environmental Science, University of Manchester, Manchester M13 9PL, UK; david.topping@manchester.ac.uk
3 School of Earth and Environmental Sciences, Cardiff University, Cardiff CF10 3AT, UK; halest@cardiff.ac.uk
4 School of Mathematics, Cardiff University, Cardiff CF24 4AG, UK; kadriu@cardiff.ac.uk
5 Welsh School of Architecture, Cardiff University, Cardiff CF10 3BN, UK; patterson@cardiff.ac.uk
6 School of Chemistry, Cardiff University, Cardiff CF10 3AT, UK; pickettj4@cardiff.ac.uk (J.P.); taylorsh@cardiff.ac.uk (S.T.)
7 School of Engineering, Cardiff University, Cardiff CF24 3AA, UK; petrii@cardiff.ac.uk
8 Discovery Partners Institute, University of Illinois System, Chicago, IL 60606, USA; peiyuanl@uillinois.edu (P.L.); sharmaa@uillinois.edu (A.S.); venkat@uillinois.edu (V.V.); awadhwa@uillinois.edu (A.W.)
9 The Alan Turing Institute, London NW1 2DB, UK; jding@turing.ac.uk
10 Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK; ruth.c.bowyer@kcl.ac.uk
11 School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, UK
* Correspondence: shaddickg@cardiff.ac.uk (G.S.); ranaof@cardiff.ac.uk (O.R.)

check for updates

Academic Editor: Jun (Justin) Li

Received: 10 December 2024
Revised: 17 January 2025
Accepted: 20 February 2025
Published: 26 February 2025

Citation: Shaddick, G.; Topping, D.; Hales, T.C.; Kadri, U.; Patterson, J.; Pickett, J.; Petri, I.; Taylor, S.; Li, P.; Sharma, A.; et al. Data Science and AI for Sustainable Futures: Opportunities and Challenges. *Sustainability* **2025**, *17*, 2019. https://doi.org/10.3390/su17052019

Abstract: Advances in data science and artificial intelligence (AI) offer unprecedented opportunities to provide actionable insights, drive innovative solutions, and create long-term strategies for sustainable development in response to the triple existential crises facing humanity: climate change, pollution, and biodiversity loss. The rapid development of AI models has been the subject of extensive debate and is high on the political agenda, but at present the vast potential for AI to contribute positively to informed decision making, improved environmental risk management, and the development of technological solutions to sustainability challenges remains underdeveloped. In this paper, we consider four inter-dependent areas in which data science and AI can make a substantial contribution to developing sustainable future interactions with the environment: (i) quantification and tracking progress towards the United Nations Sustainable Development Goals; (ii) embedding AI technologies to reduce emissions at source; (iii) developing systems to increase our resilience to natural hazards; (iv) Net Zero and the built environment. We also consider the wider challenges associated with the widespread use of AI, including data access and discoverability, trust and regulation, inference and decision making, and the sustainable use of AI.

Keywords: artificial intelligence; data science; sustainability; climate change; air pollution; Sustainable Development Goals; natural hazards; Net Zero; built environment; digital twins

1. Introduction

Climate change will bring fundamental changes to our environment—changes that have the potential to pose significant threats to all aspects of society: our health, wealth, wellbeing, critical infrastructure, security, and future prosperity. These impacts are occurring, and increasing in severity, at a greater rate than was previously expected, and there will be severe challenges in implementing adaptation strategies at the pace that is required [1,2]. Developing new technologies, together with successful policies and interventions, will be crucial if we are to respond to these threats and ensure that we develop sustainable interactions with the natural environment [3].

Advances in data science and artificial intelligence (AI) offer unprecedented opportunities to address these challenges and pave the way for sustainable futures. Environmental challenges are inherently global and the complexity of integrated environmental–social–economic systems [4], together with the sheer volume of data that need to be assessed, means that human decision-making abilities need to be augmented with AI, capitalising on the combination of algorithms, machine learning (ML), and high-performance computing to deliver the information necessary for evidence-based decision making [5,6]. By harnessing the vast amounts of data generated across multiple sectors and applying cutting-edge AI algorithms, these technologies can provide actionable insights, drive innovative solutions, and create long-term strategies for sustainable development in response to the triple existential crises facing humanity: climate change, pollution, and biodiversity loss [7]. Within this, AI has an important role to play in improving our ability to forecast and communicate environmental risks which will be critical to enable resilience of individuals, communities, businesses, and governments [8,9] and to improve preparedness and response to natural hazards [8–10].

The rapid development of AI models has recently been the subject of a great deal of debate and is high on the political agenda. Much of the focus has been on the use of large language models (LLMs) that utilise the increasing availability and quantity of text data, including on-line text, websites, and social media posts to train models that can generate natural language responses in a variety of contexts [11]. LLMs have multiple applications, including as virtual assistants, chatbots, or text generators, such as ChatGPT, but the potential uses of AI go far beyond these applications and the ability to generate new insights through identifying patterns within multiple, diverse, data sources means that AI has huge potential to contribute positively to informed decision making, improved risk management, and technological innovation. In addition to recent developments in AI technologies, there has been an explosion in the quantity and complexity of data related to the environment, including those from environmental monitoring and satellite remote sensing/Earth observation large-scale numerical modelling of the climate and other environmental systems.

The UK Government's National AI Strategy [12] highlights the need for digital technologies in supporting the move towards Net Zero targets, and the AI Council's roadmap advocates the use of AI in developing innovative solutions to climate change and increasing the pace of decarbonisation across the most impactful sectors [13]. The AI Strategy gives a number of examples of climate change and mitigation challenges where AI might have a role to play, including: (i) using ML vision to monitor the environment; (ii) using ML to forecast electricity generation and demand and control its distribution around the network; (iii) using data analysis to find efficiencies in emissions-heavy industries; and (iv) using AI to model complex systems, like the Earth's own climate, so we can better prepare for future changes. The AI Strategy notes that, although AI applications in energy and climate are being developed, they are predominantly outliers and there are many applications across different sectors that are yet to be realised.

The aim of this paper is to explore the ways in which data science and AI have the potential to make a substantial contribution in addressing the challenges associated with environmental and climatic change and helping develop sustainable future interactions with the environment. To illustrate this, we present a series of examples of current and possible future applications, based upon the following application areas:

(1) Quantification and tracking progress towards the UN Sustainable Development Goals (SDGs);
(2) Embedding AI technologies to reduce emissions at source;
(3) Developing systems to increase our resilience to natural hazards;
(4) Supporting a transition to Net Zero and a sustainable built environment.

In drawing together the examples presented in the paper, we asked domain experts: (i) what they consider to be the most important environmentally related challenges in their areas; and (ii) how AI can help in addressing those challenges. Some of the examples reflect well-developed use cases, whilst others are more speculative in nature. We also consider some of the challenges associated with the widespread use of AI, including data access and discoverability, trust and regulation, inference and decision making, and the sustainable use of AI.

2. Quantification and Tracking Progress

The UN's 2030 Agenda for Sustainable Development calls for a plan of action for people, planet, and prosperity, aiming to take the bold and transformative steps that are urgently needed to shift the world onto a sustainable and resilient path. A robust follow-up and review mechanism for the implementation of the 2030 Agenda requires a solid framework of indicators and statistical data to monitor progress, inform policy, and ensure accountability of all stakeholders. This requires high-quality, accessible, timely, and reliable disaggregated data with comprehensive global coverage to ensure that indicators are comparable over both space and time, but in many countries such data are not routinely available. The UN Environment Programme (UNEP) estimates that of the 91% of the environmentally related SDG indicators, there are suitable data for tracking progress for less than a third [14].

SDG indicators are, by their very nature, global representations of the state of the environment, climate, and health and calculating many of them is not straightforward, as they often do not align with routinely recorded data. This poses a number of challenges, including: (i) lack of data with comprehensive global coverage; (ii) substantial variation in the availability, quality, and accuracy of data over time and space and in the availability of knowledge about how data were collected, e.g., detailed information on survey design, calibration of sensors used to capture the data, or modelling assumptions; (iii) lack of consistency between datasets from different sources and measured at different scales, for example, whether tropical forests sequester (based on remote sensing) or emit carbon (the latter based on national inventories) [15].

Using data science and AI to bring together disparate sources of data in a coherent fashion is essential in understanding and tracking changes in the complex interactions between the climate, natural ecosystems, human social and economic systems, the built environment, and health. This understanding will be crucial in ensuring that SDG targets and indicators are true and accurate representations of the world that the SDGs aim to protect and are available in a consistent manner.

Consider the need to estimate exposures to air pollution required for SDG 11.6.2 ('country-level population-weighted annual average exposure to particulate matter in cities') and SDG 3.9.1 ('mortality rate attributable to household and ambient air pollution'). Air pollution is one of the greatest threats to global health and economic development and it is estimated that 4.2 million deaths annually can be attributed to ambient (outdoor) air pollution [16]. Producing global assessments of the effects of air pollution on health requires an assessment of exposures in every country. However, although monitoring of air quality is increasing around the world there are still many countries and regions in which it remains sparse and cannot provide the information that is required.

Alongside monitoring there are other methods that can be used to assess levels of air pollution that do provide comprehensive spatial coverage, including estimating aerosol optical depth using remote sensing and land-use information. However, there are a number of challenges in making full use of these data in this setting: (i) they are often not available at high enough resolutions and therefore do not capture finer-scale variation in air pollution; and (ii) they are fundamentally different quantities in monitoring, with each having different biases and uncertainties.

The Data Integration Model for Air Quality (DIMAQ) was developed by the WHO to address these challenges and provide a method for 'joining together' data that are available at different spatial and temporal scales. DIMAQ is based upon a Bayesian hierarchical modelling framework and combines information from ground monitoring, remote sensing satellites, chemical transport models, land use, population density, topography, and other sources. The result is high-resolution estimates of air pollution for every country in the world, together with associated measures of uncertainty. DIMAQ can also produce exceedance probabilities, e.g., the probability that levels of pollution in any area exceeded WHO Air Quality Guidelines (AQGs) [16].

The results can be seen in Figure 1, which show how data from ground monitoring, which are sparse in many regions, can be supplemented with those from other sources to produce high-resolution estimates of PM2.5 in the form that can be used to calculate air-pollution-related SDG indicators 11.6.2, country-level average exposures, and 3.9.1, mortality rate attributed to household and ambient air pollution. In addition to SDG indicators, DIMAQ provides essential information that enables governments to understand the seriousness of air pollution in their countries and provides them with an evidence base with which to promote action [17,18]

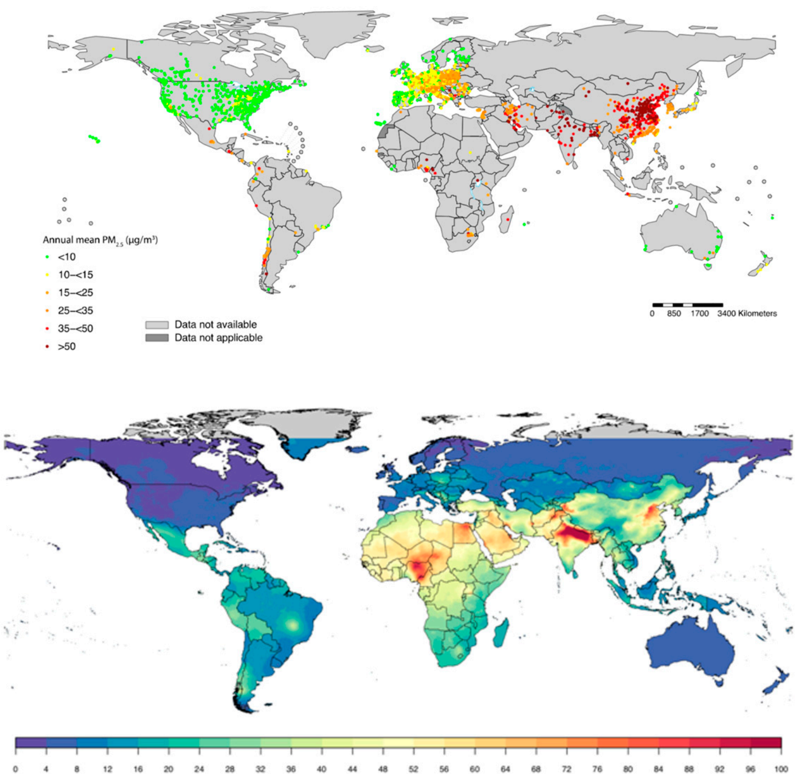

Figure 1. An example of the integration of different sources of information to produce new data products. Here, measurements from ground-based monitoring of fine particulate matter air pollution

(PM2.5) are integrated with estimates from remote sensing (based on aerosol optical depth), land use, and other sources of information related to air pollution. The image at the top shows the locations of PM2.5 monitors within the World Health Organization's Ambient Air Quality Database and the image on the bottom shows the results of the DIMAQ (see text for details) which produces a comprehensive set of estimates at a 0.1° resolution across the globe.

3. Embedding AI Technologies to Reduce Emissions at Source

3.1. Chemical Catalysis

Catalysis is a vitally important enabling technology that will play a critical role in achieving sustainability development targets. Over 90% of chemicals require a catalyst in their manufacture, and at least 80% of all goods use a catalyst somewhere in the manufacturing chain. Hence the impact is broad, contributing to fuels, medicines, food, consumer products, energy, and many other sectors. Improved catalysts are required for many transformations, but one increasingly important example is the development of new catalysts for the conversion of carbon dioxide (CO_2) to valuable chemicals and fuels, which can be achieved by reaction with hydrogen derived from renewable routes. This aligns with sustainable development by converting CO_2 into useful products, reducing greenhouse gases (GHGs), which contribute to mitigating climate change. It also delivers resource efficiency, as using CO_2 as a valuable feedstock reduces our reliance on fossil fuel resources and promotes the concept of the circular economy. This application contributes to the following SDGs: (i) SDG 7: Affordable and clean energy; by manufacturing clean fuels; (ii) SDG 9: Building resilient infrastructure, promoting inclusive and sustainable industrialisation, and fostering innovation; (iii) SDG 12: Ensuring sustainable consumption and production patterns; (iv) SDG 13: Take urgent action to combat climate change and its impacts.

The potential uses for chemicals that can be produced from CO_2 as a consequence of creating better catalysts are wide ranging. A variety of chemical intermediates can be prepared, and the most flexible of these is most likely methanol, which can be used inherently as a fuel, converted to other products used in the manufacture of many other important commodities, or converted into liquid fuels using known conventional technology. Furthermore, CO_2 can be converted directly into synthetic liquid transport fuels, avoiding methanol as an intermediate, and using CO_2 as a replacement feedstock for fossil fuel resources will benefit manufacturers of chemicals, impacting on all of our lives. Examples of potential products are pharmaceuticals, building materials, and consumer products. Other beneficiaries are the energy sector, wider industries, society, and the environment, through the reduction of emissions and more efficient use of resources.

Research into the discovery and development of new and more effective catalysts, such as those for CO_2 conversion, generates data from a range of experimental and computer simulation approaches. The data generated are batch data and include metrics on catalyst performance, such as activity and relative quantities of products generated, and are logged with process conditions, like temperature, flow rate, and pressure. Batch data are also generated from a wide range of analytical techniques that are used to identify the structure, composition, and chemical properties of the catalysts, and typically five or six different analytical techniques are employed for each catalyst prepared, including high-resolution electron microscopy which generates atomic-scale images of the surfaces of active catalysts. Data processing typically takes place manually from spreadsheets of data for the catalyst performance and the spectra, profiles, and images from catalyst analysis are typically processed individually for each experiment. However, AI is already playing a role in the processing of images from electron microscopy, using automated recognition of catalyst particles and atomic features [19].

Research data are typically at the point scale, originating from individual experiments. If the technology extends beyond the laboratory and is successfully commercialised, then the scale expands considerably. For commercial operations, data will move to real time, and the opportunity for data processing could shift to more complex models capable of learning from historical trends and adapting to changing operating conditions. This approach would be required to optimise commercial production.

There are several areas that algorithms impact in catalyst design. A detailed understanding of the kinetics (rate) of a chemical process is important in unravelling the mechanism and critical in scaling up a reaction. Reaction kinetics modelling is approached using AI and ML and it helps to predict the influence of operating parameters, like temperature and pressure, reaction rates, and selectivity. ML is now starting to be used in the process of catalyst discovery and development [20] and identifying novel catalysts with desired properties for specific reactions [21]. There are also approaches used for reinforcement learning for process control, with the aim to optimise chemical reactor parameters over time to improve the desired product yields and reduce energy consumption. Computational modelling of catalysis is making more rapid use of AI compared to experimental studies [22]. The importance of using AI for combined computational and experimental catalysis is now emerging, and it will accelerate to become increasingly important to uncover correlations between catalyst properties and catalytic activity [23]. This is the ultimate goal of much research in catalysis, as knowledge of links between material properties and activity can speed catalyst discovery through directed design, creating improved catalysts that underpin many sustainability goals.

3.2. Agriculture and Food Production

Chemistry has a reputation for producing the causes of climate change. It also has, by definition, the means of helping to mitigate and eliminate these causes, for example, the move towards a completely non-carbon-based fertiliser system for soil fertility and health and other rhizosphere-related resources, e.g., carbon sequestration, the elimination of broad-spectrum eradicant pesticides, only grass-fed cattle and other farmed animal systems, and the eventual perennialization of all arable food cropping [24]. This application contributes to (i) SDG 1: End poverty in all its forms everywhere; (ii) SDG 2: Ending hunger, improving food security and nutrition, and promoting sustainable agriculture; (iii) SDG 13: Take urgent action to combat climate change and its impacts.

Approaches to making a globally sustainable agricultural food production system, in the short term, can sometimes be overlooked in favour of biological engineering research, e.g., vertical farming, fake meat and dairy products, and robotic weed and pest control using mechanical, heating, or other physical processes. These are intensive in raw materials, energy use, and storage and cannot be delivered by even the most optimistic of current projections for the scale of food production needed using technologies currently available.

Even with these potential developments combined with new business plans and policies favouring sustainable agriculture, we will still have an inherently non-sustainable food production system. Apart from the harvesting of food-crop and animal products, all current inputs to agriculture are nominally replaceable, seasonal, and calendar based and involve high energy demands and disruption of what could be an enormously more intrinsically productive system. Furthermore, the problems of annual cropping are demonstratively obviated in the production of food from grass-fed animal husbandry because the crops, grass and its companion plants, are all perennial. This must now change by the perennialization of all our cropping systems. The technology is building very rapidly for wheat [25] in the US and (non-irrigated) rice [26] in China and we now have a new opportunity to direct perennialisation technology development—particularly to demonstrate its necessity and inevitable adoption by exploitation of AI based modelling.

The key challenge is the perennialisation of arable food production and the need for new technologies for crop production and protection. A number of automation approaches already exist, land robotics, unmanned aerial vehicles (drone technologies), and remote detection of key chemical markers for crop and animal production and health. These technologies will need to be developed for purpose but will be able to follow the enormous datasets already accumulated by traditional means and held by organisations such as Defra and the USDA. AI systems that can combine data from all of these sources, together with climate projections, crop yield forecasting, and disease models, could open novel ways to obviate even the least sustainable harvesting systems. One particular challenge where AI has huge potential to contribute is the 'scale of application' from regional all the way to global, involving predicting the perennial crop plants themselves and providing an entirely new crop production and protection system, including the further elaborated sentinel plant technology [27]. Such AI systems could be used to determine what new products and services are required for the perennialisation of arable cropping.

4. Developing Systems to Increase Our Resilience to Natural Hazards

AI has demonstrated significant potential to address resilience goals that form the basis of major international agreements such as the Sendai Framework [28] and the Resilient Cities Network (formerly the Rockefeller 100 Resilient Cities) [29], as well as contribute to (i) SDG 9: Building resilient infrastructure, promoting inclusive and sustainable industrialisation, and fostering innovation; (ii) SDG 11: Making cities and human settlements inclusive, safe, resilient, and sustainable; and (iii) SDG 13: Take urgent action to combat climate change and its impacts. AI has shown particular promise in improving risk management of multi-hazards [30], where two or more concurrent hazards interact to create a greater impact, and cascading hazard scenarios [31], where a major hazard triggers further hazards that continue through time, as well as contributing to improved forecasting and early warning systems [32].

4.1. Earthquakes

Large continental earthquakes pose a classic example of a cascading hazard; earthquake shaking leads to landslide generation followed by impacts from debris flows and flooding that can last for decades after the event [31]. These cascading events can reach volumes greater than 1 million m^3 of sediment and result in thousands of fatalities. Post-earthquake debris flows pose two particular problems in our ability to predict their hazards: (i) triggering events depend on random spatial and temporal factors such as rainfall intensity and duration, subsurface hydrology, ground stress state, and sediment thickness; (ii) key properties of sediment that govern both the triggering and bulking of debris flows change through time by erosion and as vegetation cover increases. The adaptive nature of ML models provides more flexible estimates of susceptibility and hazards when compared to more traditional deterministic and empirical–statistical methods [33,34]. In these rapidly changing contexts, the heavy parameterisation of deterministic debris flow models and requirement for long records of debris flow data in empirical–statistical methods limit their applicability for post-earthquake hazard prediction.

A spatial hazard model for post-earthquake debris flows was developed using a number of geospatial datasets (topography, earthquake intensity, coseismic debris volume) and a post-earthquake debris flow dataset [34]. By integrating a range of different ML methods, it was possible to predict post-earthquake debris flow potential at the catchment scale (<10 km^2) for the Longmen Mountains, China. While there is the possibility to include a temporal element to aid real-time prediction, the datasets available were not of high enough spatial or temporal accuracy to allow that. As with many of the cases considered within this paper, although the model proved to work well, the limitation was (and still is) data.

4.2. Tsunamis

Another example is the classification of submarine earthquakes and determining the risk of tsunami events. Tsunamis are a serious threat to coastal areas, impacting the lives of over 700 million people, leading to an urgent need for effective early warning systems. Since the 1940s, tsunami warning technology has improved with seismic networks, Deep-ocean Assessment and Reporting of Tsunamis (DART) buoys, and GPS buoys providing real-time data which, together with advances in numerical models, have helped reduce false alarms [35–38]. However, warning systems still face challenges including high false alarm rates and unreliability. UNESCO highlights problems in issuing timely warnings for local tsunamis, especially when the tsunami's source and impact area are in close proximity [39]. Dependence on earthquake data often leads to precautionary alerts, later cancelled when sea-level data show no danger. While this approach prioritises safety, it damages the credibility of warning centres and leads to public scepticism. Since the 1950s, 75% of tsunami warnings leading to evacuations were false [35], like the 1986 Honolulu evacuation, which caused over USD 30 million in losses. Improving detection and public awareness is needed to address this credibility and economic issues and aligns with the UNESCO-led intergovernmental priority of SDG 11, promoting inclusive, safe, resilient, and sustainable cities, and in accordance with Target 5 of the Sendai Framework [28]. Digital signal-processing techniques and low-frequency sound recordings from earthquakes, known as acoustic-gravity waves, can be used to train AI algorithms to classify earthquake type and magnitude [34]. Acoustic-gravity waves are generated alongside the tsunami, though travelling at a much higher speed—at the speed of sound in water—across thousands of kilometres while carrying vital information about the earthquake's source. This is a significant step for a reliable early tsunami warning system since the type of earthquake can dictate if a tsunami will be generated at all. For the training of the model, data of 200 earthquakes were collected by hydrophones (underwater microphones) in the Pacific and Indian Oceans. Moreover, a newer generation of the AI model was trained with acoustic signals from 1400 earthquakes alongside the water elevation measured by all available DART buoys (ca. 15–20 per earthquake) [40]. This version is now capable not only of determining if an earthquake is tsunamigenic but can also calculate the size of the tsunami in coastal areas, globally, within a few seconds. An example of the outputs from the resulting Global Real-time Early Assessment of Tsunamis software can be seen in Figure 2. This technology comprises a collection of models that have been integrated into software with the goal to make it operational to complement efforts by warning centres and provide a more reliable assessment, globally.

4.3. Flooding

Understanding flood propagation is challenging due to high computational costs associated with running complex hydrological models. Incorporating dynamic changes in topography and geomorphology in physical-model-based prediction systems is highly complex whereas AI-based predictions can easily accommodate these dynamic changes [41]. With their ability to represent complex systems and integrate diverse inputs, AI models can provide faster predictions and data-driven solutions can help to balance accuracy with efficiency while addressing parameter uncertainties and resource constraints [42]. Artificial neural networks (ANNs) and long short-term memory (LSTM) deep learning have shown promising results for hydrological and hydraulic prediction and forecasting in natural environments at a large geographical scale [43]. Emerging deep learning techniques have proved useful in identifying stage–discharge relationships, rainfall–runoff, sediment transport, flood prediction [44], and sustainable solutions [45]. With the availability of such technologies, it becomes easier to increase the efficiency of the flood prediction processes by minimising human involvement [46].

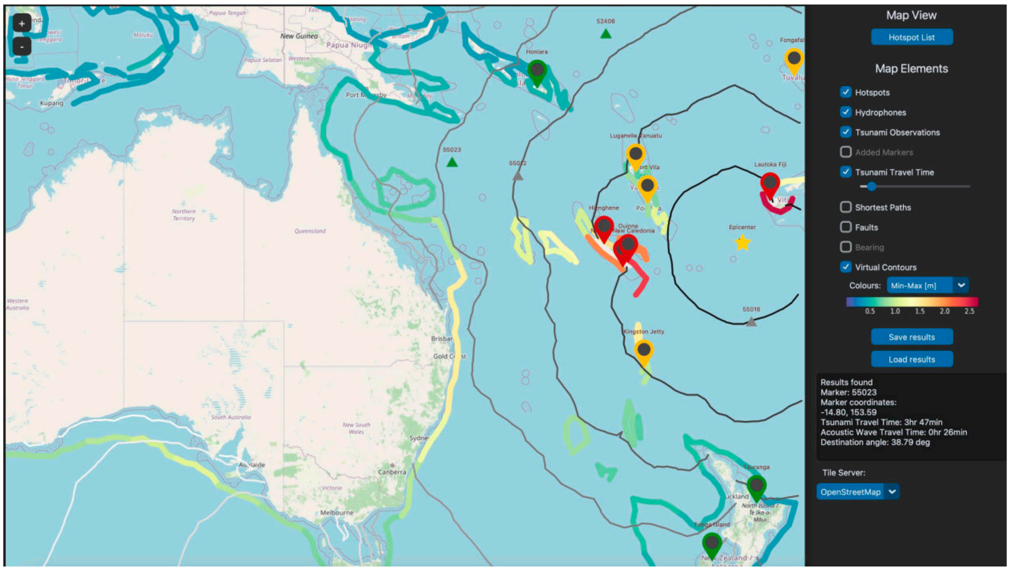

Figure 2. Outputs from GREAT for test case 2022 Hunga Tonga-Hunga Ha'apai tsunami. Yellow star: volcanic eruption epicentre. Green triangles: the location of current DART buoys. Hotspots: user-defined points of interest (red for high risk, yellow for middle risk, green for normal). A snapshot from the software showing tsunami arrival times (black/grey contours) and size (coloured contours) at 50 m depth.

Urban systems are highly complex and current approaches use features such as elevation, slope, aspect, curvatures, topographic wetness index, and hourly rainfall, however, they do not currently incorporate features that are closely related to fooding in the urban environment, e.g., fluvial infrastructure and impervious location within the contributing area [47]. Leveraging data-driven approaches alongside geographic and urban features would significantly mitigate the complexities of urban flooding, providing a more practical and cost-effective method for flood quantification while reducing the need for extensive computational resources.

4.4. Local-Scale Weather Forecasts

Current AI models for weather forecasting, such as GraphCast [48], Pangu-Weather [49], and FourCastNet [50], significantly reduce computation time compared to traditional physical models. However, despite being faster and more efficient, these AI models have not surpassed the comprehensive functionality that is available when using traditional models, which remain essential for capturing the full range of atmospheric processes and long-term climate dynamics. In addition, most of the training data for the AI models are the outputs from physical models, which means that functionally they act as surrogates of the physical models. This limits their training domain within the physics and processes that have been prescribed into the model code that can potentially lead to misrepresentation of spatio-temporal patterns in local environments as physical models deliberately simplify or neglect secondary processes in order to reduce the computational burden [51]. Although these simplifications may not have a significant impact on simulation results for global- or continental-scale dynamics, they can lead to inaccuracies when the interest is in highly localised effects. In addition, there is often a lack of data with which to parameterise these additional processes. Though computation is no longer a major concern with AI approaches, this issue prevents current AI approaches from becoming truly revolutionary successors to traditional methods.

Moving on from the current approach of surrogate modelling, the next generation of AI models represent a more integrative approach, blending AI's ability to learn from complex, real-world, real-time observational datasets with the physical laws governing large-scale weather dynamics [52]. This structured hybrid approach aims to close the gaps in local simulations by prioritising microprocesses and real-world data, directly addressing the two major challenges mentioned above. One focus of this approach is to integrate more observational datasets into the modelling framework through advanced data acquisition methods for detailed geospatial and environmental variables. Examples include, but are not limited to, the use of low-orbit CubeSats for land surface monitoring high-precision light detection and ranging (LiDAR) for accurate urban morphology [52], surveillance cameras for traffic intensity [53,54], crowdsourced weather conditions from citizen weather stations, and IoT-based sensor networks [55]. These measurements generate massive data streams, providing firsthand climate information at microscales. AI plays a twofold role: extracting complex patterns from noisy observational data [56,57] and, more importantly, training on a diverse suite of climate data to learn how geospatial features drive specific environmental outcomes [58].

Intra-urban environmental variabilities lead to disparities in the magnitude of potential hazards and thus vulnerabilities within different areas of urban environments, highlighting the need for detailed weather simulations down to the neighbourhood scale. The next generation of climate models is envisioned to be fully scalable in both space and time and capable of operating at much higher resolutions in near real time [59,60]. These advancements are anticipated to provide special focus on the most vulnerable regions, such as populous urban areas, where two-thirds of the global population is expected to reside in the near future [45].

The ability of AI models to identify critical processes governing hyperlocal environmental dynamics, which ultimately contribute to intra-urban variability, will be revolutionary in this regard. At this fine granularity, diverse human activities and localised urban features, such as transportation patterns, building densities, and green spaces, play a significant role in shaping the environment. These factors influence not only perceptible dynamics like sunshine, temperature, precipitation, wind, and air quality, but also visible yet crucial elements such as GHG emissions. However, emission processes are intricately tied to both human activities and environmental factors [61], making them highly complex and difficult to model through purely process-based approaches [62]. Collectively, the two key challenges, lack of detailed physics and insufficient data to parameterise existing processes, are magnified at this scale, driving the recent emergence of a new genre of climate AI endeavours.

5. Net Zero and the Built Environment

If countries are to meet their Net Zero targets, one of the most important requirements will be to develop a low-carbon built environment. The UK's built environment is responsible for 25% of total UK GHG emissions, with the 29 million homes in the UK accounting for approximately 14%. The UK Green Building Council have indicated that most reductions in operational emissions achieved in the past two decades are a result of the decarbonisation of the electricity grid rather than improvements in the energy efficiency of buildings. All stages of delivering a low-carbon built environment need to be informed as much as possible by data that are as accurate as possible within the financial, skill set, and time constraints that exist. However, data are often treated poorly in the process, without careful consideration and identification of issues such as bias and ethics, from data cleaning to interpretation. These issues are often dismissed as trivial, when in fact they are essential to generate data that are of real value and that can lead to the high-quality outcomes that we need.

5.1. Retrofitting Buildings

Retrofitting buildings will be essential to ensure they are energy efficient and provide healthy environments for residents to live and work. A well-designed whole-building energy-systems-based retrofit approach can make a building more energy efficient, reduce carbon emissions and energy bills, and improve living conditions. An example of how principled data collection has been used to inform retrofitting buildings can be found in Swansea where, working with the Welsh School of Architecture (WSA), the city council have changed the way that they retrofit and build their new homes as a result of the evidence gained through large-scale data collection from residents, the individual technologies, and the system as a whole in homes that have been retrofitted along with new builds [63].

In developing approaches to Net Zero, including low-carbon built environments and energy-efficient buildings, it is crucial that changes to buildings are fit not just for today but for the future climate. The UK Climate Projections (UKCP) indicate a greater chance of warmer, wetter winters and hotter, drier summers, along with an increase in the frequency and intensity of extremes. Bringing together information from climate projections such as UKCP and detailed data and modelling on energy use within buildings using AI will be a key tool in developing 'virtual scenarios' to be run, thus supporting decision making at the design stage and providing the evidence to demonstrate the benefits of implementing low-carbon solutions into the built environment which go beyond energy savings and include improvements in the quality of the built environment, health and wellbeing, and comfort.

5.2. Digital Twins for Energy Management

Another example of the use of data and digital technologies in addressing the energy efficiency of buildings is the Computational Urban Sustainability Platform (CUSP), a digital twin platform that delivers advanced predictive and (near) real-time reactive control for energy optimisation in buildings and connects with a building energy management system. CUSP can provide power and thermal demand simulations, as well as estimate renewable energy generation potential (e.g., using locally hosted photovoltaic panels) to deliver a better understanding of the impact and interactions across different climatic, socio-economic, and technological drivers of power or/and thermal energy demand. An example of the outputs of CUSP and the user interface can be seen in Figure 3. CUSP makes use of a range of data across the building, district, and city energy networks. The gathered information sources take the form of dynamic energy network models, generated from existing 'as-built' information conveyed through building information modelling (BIM), together with current and historical data originating from smart sensor networks. CUSP makes use of a variety of AI algorithms, combined with a semantics-based energy optimisation capability, utilising transfer learning [64].

CUSP aims to achieve long-term sustainability by proposing a paradigm shift in the way buildings are managed—from 'passive' to 'active' assets which can be actively operated and which respond to various stimuli. This transformation requires the consideration of a wide range of factors, including environmental conditions, behavioural changes, economic considerations, and technical advancements. Potential end users and adopters of CUSP include engineers, architects, facilities management companies, contractors, digital twin/BIM technology providers, energy assessors, and building energy management system providers. CUSP has generated savings of 15–45% in previous projects and case studies [65].

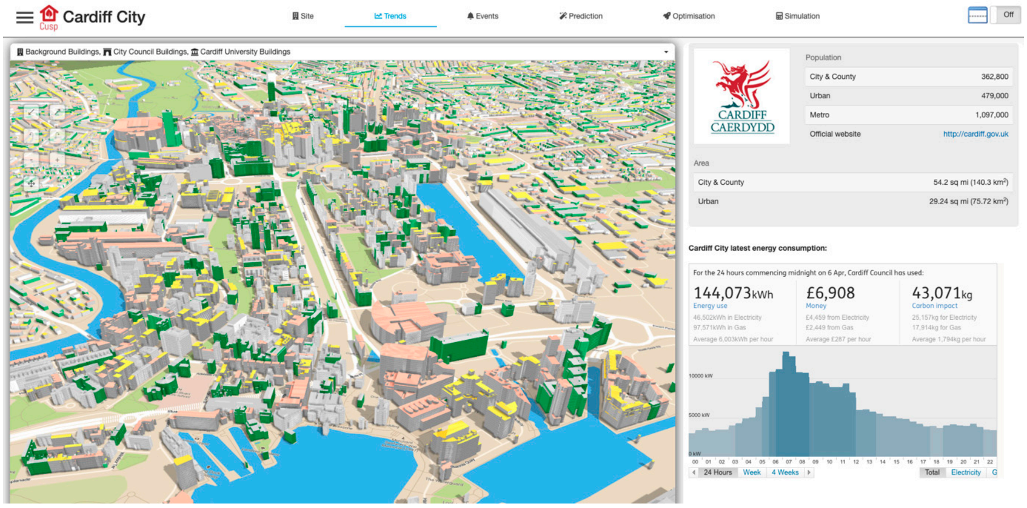

Figure 3. An example of the use of the CUSP digital twin for performance management and energy analytics in smart cities, showing energy use, associated costs, and carbon impacts for a 24 h period in Cardiff, UK.

Buildings are unique ecosystems that require customised interventions, making it challenging to apply general-purpose AI models to manage buildings effectively. CUSP leverages Transformers as a means of understanding patterns in sensor data reading, based on the concept of neural networks where context is inferred from the sensed data analysed within the platform, informed by the semantics of the domain, and interpreted by generative AI algorithms. CUSP has been trialled in stadiums (such as those used to host the Qatar football world cup in Doha), sports facilities, cargo seaports, and rail stations (a digital twin has been developed of Reading railway station in the UK).

6. Discussion

The aim of this paper was to explore the enormous potential for data science and AI technologies to develop sustainable solutions to the challenges associated with environmental and climatic change. Through a series of case studies, we have considered some of the opportunities and challenges associated with a selection of important areas related to sustainability.

In addition to identifying potential AI applications, it is essential to consider the supporting processes, governance, and digital infrastructure on which the ultimate success of any implementation strategy and technology rests. We now consider a number of wider issues related to data science and AI reaching their full potential in this area. This includes issues including data access and provenance of any AI-enabled decisions made but also those related to ensuring sustainable use of technologies and appropriate access to a wide range of stakeholders. Also, with the increased use of AI in addressing climatic and environmental challenges, it will become increasingly important to account for its energy footprint [66]. The significant resources required in AI development, particularly in the development of foundational models and 'flagship' frameworks, is well documented [67–70].

One of the common themes throughout this paper, as mentioned in many of the examples, is that access (or lack of access) to large datasets that accurately represent a specific application on which to train AI models is often a barrier to the use of AI technologies in specific use cases. There has been considerable progress in making data openly accessible, largely driven by mandates that require publicly funded research data to be made open access after a specified embargo period. While this practice is not guaranteed for data from privately managed networks, there has been a noticeable cultural shift toward embracing open science. This shift includes the development of open-source software and the release of open data, supported by various licensing options. The FAIR data principles—findability, accessibility, interoperability, and reusability—have been established to guide these efforts, with similar principles proposed for software [71]. These initiatives are not only pivotal for fostering open collaboration but also crucial for building trust in the use of data and software. Making data and software openly accessible is a crucial first step toward aligning with emerging AI standards and regulations. The field of AI has faced a 'reproducibility crisis', partly due to insufficient data and software sharing [72]. However, simply providing open data and software does not ensure that they are easily discoverable or usable by others [73].

Another aspect of data availability is cases where it is too expensive and/or simply too difficult to obtain robust controlled datasets with which to build new models. A specific example is the ability to accurately identify ice crystal habits in high-altitude conditions, an important feature that affects the radiative properties of clouds and thus the Earth's radiative budget [74]. Replicating environmentally relevant conditions in a controlled laboratory setting to build a robust training dataset is likely to remain a challenge, but transfer learning, a process by which industry standard algorithms can be fine-tuned to the limited data available, presents a sustainable and scalable solution [75]. This presents an opportunity for research and technology partnerships to identify opportunities for knowledge exchange, replicating success across multiple domains.

One of the key aspects of data science and AI is the ability to integrate data from multiple sources to provide new information and insights. However, this means that data are increasingly being used for decision making in areas other than that for which they were collected. The differences between using data that have been collected as part of a carefully controlled experiment, such as a randomised controlled trial, requires a fundamentally different approach that acknowledges all aspects of the 'data journey', including methods of collection, biases, governance, and the use of derived datasets. However, methods for performing inference and decision making have not kept up with the rapid development of AI methods and the growth in the quantity and variety of data sources together with the quality of the data that they produce. The effects that this can have on interpretation, communication, and ultimately decision making will be one of the most important aspects of using AI and data science in applied settings and will require the development of methodological approaches that acknowledge the challenges of bringing together data from multiple sources and allow for uncertainty to be propagated between different models. Another challenge is the ability to identify where errors matter the most. In tasks where mistakes are more tolerable, such as web searches and presenting advertisements, decisions are based upon global error rates, but in 'high-stakes' applications, such as those related to early warning systems or linking environmental stressors to personalised health outcomes, every prediction matters, and more sophisticated measures of success need to be developed and adopted. As we move towards the large-scale use of AI in decision making, there is an urgent need for a deeper understanding of the potential effects of these to ensure that we can build trust across a wide variety of stakeholders.

This highlights the need for well-maintained data platforms and comprehensive metadata to help users understand the scope and utility of the available resources. As an example, consider data on climate projections which are core to many of the applications in this paper, for example, retrofitting buildings in Section 5.1, and are crucial for all aspects of climate adaptation, mitigation, and planning for Net Zero. This information is available from many sources, including, for example, the UK Climate Projections (UKCP) from the

UK Met Office. UKCP provide valuable information on the UK's climate change, but it requires technical expertise to download, manage, and process. This is particularly important for stakeholders like local councils and public health officials who may not have the necessary technical expertise. Bias correction, or recalibration, is a crucial step in adjusting climate projections at local levels to align with actual weather measurements. However, this process is typically performed ad hoc by climate scientists for specific areas and time periods, rather than at scale due to the computational requirements and domain expertise needed. Outputs might also be required in different forms, such as spatial formats and coordinate reference systems. For example, data for decision processes may not naturally align with the gridded output and may be required in different forms. This is an example of where data science and AI can democratise climate data by developing and implementing data and analysis pipelines. The CLIM_RECAL project [76] is an example of this, providing a resource designed to tackle systematic errors or biases in regional climate models (RCMs). The project includes custom scripts for downloading data, pre-processing, applying bias correction, and assessing debiased data. This results in accessible information on bias adjustment methods for non-climate scientists and lay-audience stakeholders, including details of different correction methods and resources for technically applying them and producing data in easily usable formats.

In response to the need to increase the accessibility and utility of environmental data, the UK's Natural Environment Research Council (NERC) has invested GBP £8M in building a new digital platform to provide improved access to the NERC's environmental data holdings, archived across five different data centres straddling subsurface to atmospheric domains [77]. As part of this investment, user research was commissioned that delved into the processes, workflows, and workarounds of people as they tried to use environmental data and sought to understand the concerns, challenges, and barriers that they faced. Covering workshops across the UK, this included engaging with a wide range of public sector organisations. Analysis revealed several common behaviours and traits and these were distilled into multiple user archetypes. The archetypes then revealed a set of attributable stories that map out a journey to achieving end user goals when using environmental data. As distilled in the report [78], several common challenges emerged, including:

- A lack of suitable data to satisfy the task in hand, with many datasets being inaccessible due to lack of discoverability, paywalls, the proliferation and confusing nature of different platforms that required bespoke access approaches, inchoate data formats and methods of retrieval, and unclear provenance.
- Having to work with locked down systems due to security or data protection requirements.
- A lack of locally available high-performance computing resources to undertake computationally intensive processes.
- A lack of a coherent and centralised data management infrastructure, prevalence of legacy datasets, and difficulty in getting people to openly share data.

Whilst the public narrative, and perhaps investment, tend to focus on AI algorithm development and deployment, these issues clearly need addressing through targeted investment and cultural shifts. In addition, in discussing sustainable futures we often focus the impacts on human health. Health data are, of course, sensitive and require the use of secure/trusted environments. There is a broader national movement in the UK towards the standardisation of trusted research environments (TREs), exemplified by initiatives like SATRE [79], which focus on establishing open standards to assess whether a system qualifies as a TRE. Within a TRE, health data must remain securely contained, and any ancillary data must be integrated into the TRE for further analysis. For instance, environmental data might be incorporated into a TRE to evaluate the impact of residential postcodes on health outcomes. While environmental data alone do not typically require a TRE, future scenarios could arise where environmental data can be used to infer human behaviour.

Moving beyond data access and discoverability, the regulation of AI, and consequently the implementation of ML, is an evolving field both on a national and global scale. In November 2023, the UK hosted an AI Safety Summit, bringing together international governments, AI companies, academics, and civil organisations to discuss the risks associated with AI and strategies for mitigation. In parallel, the dialogue around AI safety and regulation is gaining momentum. The AI Council has emphasised that the UK can fully benefit from AI only if there is widespread public confidence in the underlying science, technologies, and the governance that supports them. While discussions around regulation often raise concerns about stifling innovation, there is an increasing recognition that regulation and innovation can coexist [80]. There is benefit in organisations in joining conversations related to regulation for internal and external use, even if the responsibility for setting regulatory procedures lies at higher levels of government and is based within different contributors to the data lifecycle (e.g., instrument vendors, research modellers, private networks, satellite consortia, etc.). Alongside future regulations that will set mandatory requirements related to the use of AI, such as the EU AI Act [81], and technological standards covering areas such as data interoperability, model evaluations and other aspects will be increasingly important. In addition, non-technical standards are equally important if somewhat less obvious when planning implementation of technical solutions (e.g., IEEE P7000 Series (Ethical Considerations in Autonomous and Intelligent Systems [82]); ISO 37101, Sustainable Development in Communities [83]). This includes both ethical and social–cultural standards [84] to ensure the benefits of AI respect cultural nuances and ensure that potential benefits can be realised, regardless of socio-economic status and cultural backgrounds.

7. Conclusions

In summary, while data science and AI allow us to gain information and insights that are predicated on much wider evidence bases than has traditionally been possible, there are of course barriers associated with widespread implementation: creating coherent and reproducible methods for sourcing and integrating data from multiple sources [85]; developing the accessible, scalable infrastructure that will be required to facilitate critical workflows, data management, and the simplification of extracting knowledge from data [86]; meeting the operational needs of end users, including accessible computational facilities, suitable time lags between data retrieval and processing, and the production of user-defined information outputs that integrate with existing business processes [87]. There are also often disconnects between the communities with domain expertise and those who generate data, those with the skills to implement AI, and those who are making decisions. Addressing these challenges and unlocking the full power of data and AI will have far-reaching impact across a wide spectrum of environmental areas, will raise the bar in data-driven environment research, and will contribute to wider technology developments, applications, and economic benefits. Building upon this, AI has huge potential in enhancing and guiding effective environmental governance by fostering culturally appropriate practices and organisational processes [88]. As such, the true value of AI may be its capacity to drive systemic and impactful environmental [89,90]. Current paradigms that integrate training, education, policies, and cultural factors often lead to decision making dominated by short-term self-interest, which inadequately addresses the complexity of environmental issues such as water, energy, and food supply. As a result, the prevalent reductionist approach tends to produce oversimplified and suboptimal solutions, despite their appearance of rationality, failing to address the intricate trade-offs and long-term sustainability.

Finally, it needs to be recognised that regions most affected by climate change, thus where research could have significant benefits, are less likely to have access to significant computational resources [90,91]. In addition, the global adoption of solutions hinges on culturally specific ethical frameworks: differing views on privacy, equity, and communal vs. individual wellbeing will likely influence acceptance. For example, meaningful, inclusive stakeholder engagement, especially in communities wary of 'digital colonialism', is essential to align AI deployments with local values and ensure fair, effective, and globally sustainable outcomes.

Author Contributions: Conceptualization, G.S., D.T. and O.R.; writing—original draft preparation, G.S., D.T. and O.R.; writing—review and editing, G.S., D.T., T.C.H., U.K., J.P. (John Pickett), J.P. (Joanne Patterson), I.P., S.T., P.L., A.S., V.V., A.W., J.D., R.B. and O.R.; funding acquisition, A.S. and O.R. All authors have read and agreed to the published version of the manuscript.

Funding: Part of this research is supported by NSF Awards # 2139316, 2230772, and 2330565 and UKRI Award No. EP/Y026233/1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data on air quality described in Section 2 can be accessed at https://www.who.int/data/gho/data/themes/air-pollution/who-modelled-estimates-of-air-pollution-from-particulate-matter (accessed on 16 January 2025).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fekete, H.; Kuramochi, T.; Roelfsema, M.; Elzen, M.D.; Forsell, N.; Höhne, N.; Luna, L.; Hans, F.; Sterl, S.; Olivier, J.; et al. A review of successful climate change mitigation policies in major emitting economies and the potential of global replication. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110602. [CrossRef]
2. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2022—Impacts, Adaptation and Vulnerability*; Cambridge University Press: Cambridge, UK, 2023. [CrossRef]
3. Swart, R. Climate change and sustainable development: Expanding the options. *Clim. Policy* **2003**, *3*, S19–S40. [CrossRef]
4. Ostrom, E. A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science* **2009**, *325*, 419–422. [CrossRef] [PubMed]
5. Fox, G.; Glazier, J.A.; Kadupitiya, J.C.S.; Jadhao, V.; Kim, M.; Qiu, J.; Sluka, J.P.; Somogyi, E.; Marathe, M.; Adiga, A.; et al. Learning everywhere: Pervasive machine learning for effective high-performance computation. In Proceedings of the 2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Rio de Janeiro, Brazil, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019.
6. Ihde, N.; Marten, P.; Eleliemy, A.; Poerwawinata, G.; Silva, P.; Tolovski, I.; Ciorba, F.M.; Rabl, T. A survey of big data, high performance computing, and machine learning benchmarks. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2022; pp. 98–118.
7. UN Climate Change. What is the Triple Planetary Crisis? Available online: https://unfccc.int/news/what-is-the-triple-planetary-crisis (accessed on 9 October 2024).
8. Akhyar, A.; Zulkifley, M.A.; Lee, J.; Song, T.; Han, J.; Cho, C.; Hyun, S.; Son, Y.; Hong, B.-W. Deep artificial intelligence applications for natural disaster management systems: A methodological review. *Ecol. Indic.* **2024**, *163*, 112067. [CrossRef]
9. Albahri, A.; Khaleel, Y.L.; Habeeb, M.A.; Ismael, R.D.; Hameed, Q.A.; Deveci, M.; Homod, R.Z.; Albahri, O.; Alamoodi, A.; Alzubaidi, L. A systematic review of trustworthy artificial intelligence applications in natural disasters. *Comput. Electr. Eng.* **2024**, *118*, 109409. [CrossRef]
10. GOA. *Artificial Intelligence in Natural Hazard Modeling*; United States Government Accountability Office: Washington, DC, USA, 2023. Available online: https://www.gao.gov/assets/870/864546.pdf (accessed on 9 October 2024).
11. Minaee, S.; Mikolov, T.; Nikzad, N.; Chenaghlu, M.; Socher, R.; Amatriain, X.; Gao, J. Large Language Models: A survey. *arXiv* **2024**, arXiv:2402.06196.
12. HM Government. National AI Strategy. 2021. Available online: https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (accessed on 30 August 2024).

13. AI Council. AI Roadmap. 2021. Available online: https://assets.publishing.service.gov.uk/media/5ff3bc6e8fa8f53b76ccee23/AI_Council_AI_Roadmap.pdf (accessed on 2 September 2024).
14. UNEP. Sustainable Development Goals Policy Brief 008. 2019. Available online: https://wedocs.unep.org/bitstream/handle/20.500.11822/30672/SDG_Brief_008.pdf?sequence=1&isAllowed=y (accessed on 30 August 2024).
15. Heinrich, V.; House, J.; Gibbs, D.A.; Harris, N.; Herold, M.; Grassi, G.; Cantinho, R.; Rosan, T.M.; Zimbres, B.; Shimbo, J.Z.; et al. Mind the gap: Reconciling tropical forest carbon flux estimates from earth observation and national reporting requires transparency. *Carbon Balance Manag.* **2023**, *18*, 22. [CrossRef] [PubMed]
16. Shaddick, G.; Thomas, M.L.; Green, A.; Brauer, M.; Donkelaar, A.; Burnett, R.; Chang, H.H.; Cohen, A.; Dingenen, R.V.; Dora, C.; et al. Data integration model for air quality: A hierarchical approach to the global estimation of exposures to ambient air pollution. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2018**, *67*, 231–253. [CrossRef]
17. OECD. *Economic Outlook for Southeast Asia, China and India 2019—Update*; OECD: Paris, France, 2019. [CrossRef]
18. United Nations Environment Programme (UNEP); Climate and Clean Air Coalition (CCAC). Integrated Assessment of Short-Lived Climate Pollutants in Latin America and the Caribbean Improving Air Quality While Contributing to Climate Change Mitigation. 2016. Available online: https://www.unep.org/resources/report/integrated-assessment-short-lived-climate-pollutants-latin-america-and-caribbean (accessed on 30 August 2024).
19. Nartova, A.V.; Mashukov, M.Y.; Astakhov, R.R.; Kudinov, V.Y.; Matveev, A.V.; Okunev, A.G. Particle Recognition on Transmission Electron Microscopy Images Using Computer Vision and Deep Learning for Catalytic Applications. *Catalysts* **2022**, *12*, 135. [CrossRef]
20. Li, Z.; Wang, S.; Xin, H. Toward artificial intelligence in catalysis. *Nat. Catal.* **2018**, *1*, 641–642. [CrossRef]
21. Ulissi, Z.W.; Tang, M.T.; Xiao, J.; Liu, X.; Torelli, D.A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N.S.; Jaramillo, T.F.; et al. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO_2 Reduction. *ACS Catal.* **2017**, *7*, 6600–6608. [CrossRef]
22. Lamoureux, P.S.; Winther, K.T.; Torres, J.A.G.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem* **2019**, *11*, 3581–3601. [CrossRef]
23. Miyazaki, R.; Belthle, K.S.; Tüysüz, H.; Foppa, L.; Scheffler, M. Materials Genes of CO_2 Hydrogenation on Supported Cobalt Catalysts: An Artificial Intelligence Approach Integrating Theoretical and Experimental Data. *J. Am. Chem. Soc.* **2024**, *146*, 5433–5444. [CrossRef]
24. Baulcombe, D.; Crute, I.; Davies, B.; Dunwell, J.; Gale, M.; Jones, J.; Pretty, J.; Sutherland, W.; Toulmin, C. *Reaping the Benefits: Science and the Sustainable Intensification of Global Agriculture*; The Royal Society: London, UK, 2009.
25. DeHaan, L.R.; Anderson, J.A.; Bajgain, P.; Basche, A.; Cattani, D.J.; Crain, J.; Crews, T.E.; David, C.; Duchene, O.; Gutknecht, J.; et al. Discussion: Prioritize perennial grain development for sustainable food production and environmental benefits. *Sci. Total. Environ.* **2023**, *895*, 164975. [CrossRef] [PubMed]
26. Zhang, S.; Huang, G.; Zhang, Y.; Lv, X.; Wan, K.; Liang, J.; Feng, Y.; Dao, J.; Wu, S.; Zhang, L.; et al. Sustained productivity and agronomic potential of perennial rice. *Nat. Sustain.* **2022**, *6*, 28–38. [CrossRef]
27. Pickett, J.A.; Khan, Z.R. Plant volatile-mediated signalling and its application in agriculture: Successes and challenges. *New Phytol.* **2016**, *212*, 856–870. [CrossRef]
28. UNDRR. Sendai Framework for Disaster Risk Reduction 2015–2030. 2015. Available online: https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030 (accessed on 30 August 2024).
29. Resilient Cities Network. Resilient Cities Network. Available online: https://resilientcitiesnetwork.org (accessed on 4 September 2024).
30. Gill, J.C.; Malamud, B.D. Reviewing and visualizing the interactions of natural hazards. *Rev. Geophys.* **2014**, *52*, 680–722. [CrossRef]
31. Fan, X.; Scaringi, G.; Korup, O.; West, A.J.; van Westen, C.J.; Tanyas, H.; Hovius, N.; Hales, T.C.; Jibson, R.W.; Allstadt, K.E.; et al. Earthquake-Induced Chains of Geologic Hazards: Patterns, Mechanisms, and Impacts. *Rev. Geophys.* **2019**, *57*, 421–503. [CrossRef]
32. Gomez, B.; Kadri, U. Earthquake source characterization by machine learning algorithms applied to acoustic signals. *Sci. Rep.* **2021**, *11*, 23062. [CrossRef]
33. Ermini, L.; Catani, F.; Casagli, N. Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* **2005**, *66*, 327–343. [CrossRef]
34. Huang, J.; Hales, T.C.; Huang, R.; Ju, N.; Li, Q.; Huang, Y. A hybrid machine-learning model to estimate potential debris-flow volumes. *Geomorphology* **2020**, *367*, 107333. [CrossRef]
35. Bernard, E.; Titov, V. Evolution of tsunami warning systems and products. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *373*, 20140371. [CrossRef] [PubMed]
36. Igarashi, Y.; Kong, L.; Yamamoto, M.; McCreery, C.S. Anatomy of Historical Tsunamis: Lessons Learned for Tsunami Warning. *Pure Appl. Geophys.* **2011**, *168*, 2043–2063. [CrossRef]

37. Kong, L.S.L.; Dunbar, P.K.; Arcos, N. *Pacific Tsunami Warning System: A Half-Century of Protecting the Pacific 1965–2015*; Government Printing Office: Mumbai, India, 2015.
38. Tsuchiya, Y.; Shuto, N. *Tsunami: Progress in Prediction, Disaster Prevention and Warning*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1995; Volume 4.
39. UNESCO. Tsunami Warning System: Preparing for the Unpredictable. Available online: https://www.unesco.org/en/tsunami-warning-system-preparing-unpredictable (accessed on 30 August 2024).
40. Kadri, U.; Abdolali, A.; Filimonov, M. GREAT v1. 0: Global Real-time Early Assessment of Tsunamis. *Geosci. Model Dev. Discuss.* **2024**, *2024*, 1–35.
41. Gonzales-Inca, C.; Calle, M.; Croghan, D.; Torabi Haghighi, A.; Marttila, H.; Silander, J.; Alho, P. Geospatial artificial intelligence (GeoAI) in the integrated hydrological and fluvial systems modeling: Review of current applications and trends. *Water* **2022**, *14*, 2211. [CrossRef]
42. Baghalian, S.; Ghodsian, M. Experimental analysis and prediction of velocity profiles of turbidity current in a channel with abrupt slope using artificial neural network. *J. Braz. Soc. Mech. Sci. Eng.* **2017**, *39*, 4503–4517. [CrossRef]
43. Hauswirth, S.M.; Bierkens, M.F.; Beijk, V.; Wanders, N. The potential of data driven approaches for quantifying hydrological extremes. *Adv. Water Resour.* **2021**, *155*, 104017. [CrossRef]
44. Nearing, G.; Cohen, D.; Dube, V.; Gauch, M.; Gilon, O.; Harrigan, S.; Hassidim, A.; Klotz, D.; Kratzert, F.; Metzger, A.; et al. Global prediction of extreme floods in ungauged watersheds. *Nature* **2024**, *627*, 559–563. [CrossRef] [PubMed]
45. UN-Habitat. Envisaging the Future of Cities. 2022. Available online: https://unhabitat.org/sites/default/files/2022/06/wcr_2022.pdf (accessed on 13 October 2024).
46. Gessang, O.M.; Lasminto, U. The flood prediction model using Artificial Neural Network (ANN) and weather Application Programming Interface (API) as an alternative effort to flood mitigation in the Jenelata Sub-watershed. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *930*, 12080. [CrossRef]
47. Al Mehedi, M.A.; Smith, V.; Hosseiny, H.; Jiao, X. Unraveling the complexities of urban fluvial flood hydraulics through AI. *Sci. Rep.* **2022**, *12*, 18738. [CrossRef]
48. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [CrossRef]
49. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **2023**, *619*, 533–538. [CrossRef]
50. Kurth, T.; Subramanian, S.; Harrington, P.; Pathak, J.; Mardani, M.; Hall, D.; Miele, A.; Kashinath, K.; Anandkumar, A. FourCastNet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators. In Proceedings of the Platform for Advanced Scientific Computing Conference, Davos, Switzerland, 26–28 June 2023; ACM: New York, NY, USA, 2023.
51. Alizadeh, O. Advances and challenges in climate modeling. *Clim. Chang.* **2022**, *170*, 18. [CrossRef]
52. Li, P.; Sharma, A. Detailed height mapping of trees and buildings (HiTAB) in Chicago and its implications to urban climate studies. *Environ. Res. Lett.* **2024**, *19*, 094013. [CrossRef]
53. Rodríguez, L.R.; Delgado, M.G.; Medina, D.C.; Ramos, J.S.; Domínguez, S.Á. Forecasting urban temperatures through crowd-sourced data from Citizen Weather Stations. *Urban Clim.* **2024**, *56*, 102021. [CrossRef]
54. Venter, Z.S.; Chakraborty, T.; Lee, X. Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. *Sci. Adv.* **2021**, *7*, eabb9569. [CrossRef]
55. Middel, A.; Nazarian, N.; Demuzere, M.; Bechtel, B. Urban Climate Informatics: An Emerging Research Field. *Front. Environ. Sci.* **2022**, *10*, 867434. [CrossRef]
56. Dabove, P.; Daud, M.; Olivotto, L. Revolutionizing urban mapping: Deep learning and data fusion strategies for accurate building footprint segmentation. *Sci. Rep.* **2024**, *14*, 13510. [CrossRef] [PubMed]
57. Gerges, F.; Llaguno-Munitxa, M.; Zondlo, M.A.; Boufadel, M.C.; Bou-Zeid, E. Weather and the City: Machine Learning for Predicting and Attributing Fine Scale Air Quality to Meteorological and Urban Determinants. *Environ. Sci. Technol.* **2024**, *58*, 6313–6325. [CrossRef]
58. Yu, Y.; Li, P.; Huang, D.; Sharma, A. Street-level temperature estimation using graph neural networks: Performance, feature embedding and interpretability. *Urban Clim.* **2024**, *56*, 102003. [CrossRef]
59. Betancourt, M. Are we entering the golden age of climate modeling? *Eos* **2022**, *103*. [CrossRef]
60. Wong, C. How AI is improving climate forecasts. *Nature* **2024**, *628*, 710–712. [CrossRef]
61. Li, P.; Sharma, A. Hyper-Local Temperature Prediction Using Detailed Urban Climate Informatics. *J. Adv. Model. Earth Syst.* **2024**, *16*, e2023MS003943. [CrossRef]
62. Li, P.; Xu, T.; Wei, S.; Wang, Z.-H. Multi-objective optimization of urban environmental system design using machine learning. *Comput. Environ. Urban Syst.* **2022**, *94*, 101796. [CrossRef]

63. Perisoglou, E.; Patterson, J.; Ionas, M. Evaluating the performance of six 1970s off-gas deep retrofit bungalows. *Build. Serv. Eng. Res. Technol.* **2024**, *46*, 9–25. [CrossRef]
64. Petri, I.; Li, H.; Rezgui, Y.; Chunfeng, Y.; Yuce, B.; Jayan, B. A modular optimisation model for reducing energy consumption in large scale building facilities. *Renew. Sustain. Energy Rev.* **2014**, *38*, 990–1002. [CrossRef]
65. CUSP. Computational Urban Sustainability Platform. Available online: https://www.cuspplatform.com (accessed on 30 August 2024).
66. Jay, C.; Yu, Y.; Crawford, I.; Archer-Nicholls, S.; James, P.; Gledson, A.; Shaddick, G.; Haines, R.; Lannelongue, L.; Lines, E.; et al. Prioritize environmental sustainability in use of AI and data science methods. *Nat. Geosci.* **2024**, *17*, 106–108. [CrossRef]
67. World Economic Forum. How to Manage AI's Energy Demand—Today, Tomorrow and in the Future. Available online: https://www.weforum.org/agenda/2024/04/how-to-manage-ais-energy-demand-today-tomorrow-and-in-the-future/ (accessed on 5 September 2024).
68. Lauterbach, A. Artificial intelligence and policy: Quo vadis? *Digit. Policy Regul. Gov.* **2019**, *21*, 238–263. [CrossRef]
69. Zhang, J.; Budhdeo, S.; William, W.; Cerrato, P.; Shuaib, H.; Sood, H.; Ashrafian, H.; Halamka, J.; Teo, J.T. Moving towards vertically integrated artificial intelligence development. *Npj Digit. Med.* **2022**, *5*, 143. [CrossRef] [PubMed]
70. Renz, A.; Hilbig, R. Prerequisites for artificial intelligence in further education: Identification of drivers, barriers, and business models of educational technology companies. *Int. J. Educ. Technol. High. Educ.* **2020**, *17*, 14. [CrossRef]
71. Barker, M.; Hong, N.P.C.; Katz, D.S.; Lamprecht, A.-L.; Martinez-Ortiz, C.; Psomopoulos, F.; Harrow, J.; Castro, L.J.; Gruenpeter, M.; Martinez, P.A.; et al. Introducing the FAIR Principles for research software. *Sci. Data* **2022**, *9*, 622. [CrossRef] [PubMed]
72. Ball, P. Is AI leading to a reproducibility crisis in science? *Nature* **2023**, *624*, 22–25. [CrossRef]
73. Weerakkody, V.; Irani, Z.; Kapoor, K.; Sivarajah, U.; Dwivedi, Y.K. Open data and its usability: An empirical view from the Citizen's perspective. *Inf. Syst. Front.* **2017**, *19*, 285–300. [CrossRef]
74. Xiao, H.; Zhang, F.; He, Q.; Liu, P.; Yan, F.; Miao, L.; Yang, Z. Classification of Ice Crystal Habits Observed From Airborne Cloud Particle Imager by Deep Transfer Learning. *Earth Space Sci.* **2019**, *6*, 1877–1886. [CrossRef]
75. Hosna, A.; Merry, E.; Gyalmo, J.; Alom, Z.; Aung, Z.; Azim, M.A. Transfer learning: A friendly introduction. *J. Big Data* **2022**, *9*, 102. [CrossRef] [PubMed]
76. Alan Turing Institute. CLIM_RECAL. Available online: https://urban-analytics-technology-platform.github.io/projects/clim_recal/ (accessed on 30 August 2024).
77. NERC. NERC Digital Solutions Hub. Available online: https://www.digital-solutions.uk (accessed on 30 August 2024).
78. OpenData Manchester. Digital Solutions Hub: User Research Report. Available online: https://www.digital-solutions.uk/wp-content/uploads/2023/09/NERC-DSH-Report-ODM-FINAL.pdf (accessed on 5 September 2024).
79. Cole, C.; Sood, H.; Oldfield, K.; Craddock, M.; Swanepoel, N.; Coleman, S.; O'Reilly, M.; Kerr, D.; O'Donovan, C.; Hetherington, J.; et al. *SATRE: Standardised Architecture for Trusted Research Environments*; DARE UK: London, UK, 2023.
80. Ada Lovelace Institute. Regulate to Innovate. 2021. Available online: https://www.adalovelaceinstitute.org/report/regulate-innovate/ (accessed on 5 September 2024).
81. EU. EU AI Act: First Regulation on Artificial Intelligence. Available online: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (accessed on 4 September 2024).
82. IEEE. Our Role in Addressing Ethical Considerations of Autonomous and Intelligent Systems (A/IS). 2018. Available online: https://standards.ieee.org/wp-content/uploads/import/documents/other/ethical-considerations-ai-as-29mar2018.pdf (accessed on 4 September 2024).
83. *ISO 37101*; Sustainable Development in Communities. International Organization for Standardization: Geneva, Switzerland, 2016. Available online: https://www.iso.org/files/live/sites/isoorg/files/archive/pdf/en/iso_37101_sustainable_development_in_communities.pdf (accessed on 4 September 2024).
84. Auld, G.; Casovan, A.; Clarke, A.; Faveri, B. Governing AI through ethical standards: Learning from the experiences of other private governance initiatives. *J. Eur. Public Policy* **2022**, *29*, 1822–1844. [CrossRef]
85. Albahri, A.S.; Duhaim, A.M.; Fadhel, M.A.; Alnoor, A.; Baqer, N.S.; Alzubaidi, L.; Albahri, O.S.; Alamoodi, A.H.; Bai, J.; Salhi, A.; et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Inf. Fusion* **2023**, *96*, 156–191. [CrossRef]
86. DeCost, B.; Hattrick-Simpers, J.; Trautt, Z.; Kusne, A.; Campo, E.; Green, M. Scientific AI in materials science: A path to a sustainable and scalable paradigm. *Mach. Learn. Sci. Technol.* **2020**, *1*, 033001. [CrossRef] [PubMed]
87. Sharma, M.; Gupta, R.; Acharya, P.; Jain, K. Systems approach to cloud computing adoption in an emerging economy. *Int. J. Emerg. Mark.* **2023**, *18*, 3283–3308. [CrossRef]
88. Duan, Y.; Edwards, J.S.; Dwivedi, Y.K. Artificial intelligence for decision making in the era of Big Data—Evolution, challenges and research agenda. *Int. J. Inf. Manag.* **2019**, *48*, 63–71.
89. Mikalef, P.; Lemmer, K.; Schaefer, C.; Ylinen, M.; Fjørtoft, S.O.; Torvatn, H.Y.; Gupta, M.; Niehaves, B. Examining how AI capabilities can foster organizational performance in public organizations. *Gov. Inf. Q.* **2023**, *40*, 101797. [CrossRef]

90. Scientific American. The AI Boom Could Use a Shocking Amount of Electricity. Available online: https://www.scientificamerican.com/article/the-ai-boom-could-use-a-shocking-amount-of-electricity/ (accessed on 5 September 2024).
91. Acosta, M.C.; Palomas, S.; Ticco, S.V.P.; Utrera, G.; Biercamp, J.; Bretonniere, P.-A.; Budich, R.; Castrillo, M.; Caubel, A.; Doblas-Reyes, F.; et al. The computational and energy cost of simulation and storage for climate science: Lessons from CMIP6. *Geosci. Model Dev.* **2024**, *17*, 3081–3098. [CrossRef]