Thesis Project Portfolio

Science Time Series: Deep Learning in Hydrology

(Technical Report)

Environmental Considerations in AI Project Funding: A Government Grant Evaluation

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Junyang He

Spring, 2025

Department of Computer Science

Table of Contents

Executive Summary

As machine learning models grow increasingly powerful driven by big-data and the scaling law, the environmental sustainability of artificial intelligence research becomes an alarming issue. My Capstone Project leverages deep learning techniques to improve rainfall-runoff modeling in hydrology, addressing the challenge of accurately predicting floods and managing water resources. Additionally, my research generalizes a data-driven approach to tackle large-scale scientific challenges where neural networks outperform traditional methods. My STS research investigates how governmental funding decisions in AI research have profound environmental impacts. Despite the enormous energy consumption of computationally intensive models, it is often overlooked in the grant approval process. My Capstone Project provides a basis for quantifying the energy usage of machine learning research, as well as exploring how government funding directly impacts environmental research projects like mine. This study emphasizes the importance of integrating sustainability into AI research at a high level, ensuring that technological advancements are pursued responsibly and ethically.

My Capstone Project, Deep Learning in Hydrology, provides a computational solution to rainfall-runoff modeling by employing the Long Short-Term Memory (LSTM) neural network. Traditional hydrological models, though effective, are computationally expensive and limited by data availability. In this work, we analyzed hydrology time series using the CAMELS and Caravan global datasets. These datasets include up to 6 time series variables and 209 environmental features collected from around 8,000 locations worldwide.

We found that including environmental data in training significantly boosts model accuracy, reducing the error by 40% when tested on the largest dataset. Additionally, including encoding techniques that captures the relationship between catchments and some periodic

hydrological patterns further improves model performance. When compared to state-of-the-art time series foundation models that require huge computational resources to train, our domain-specific LSTM model outperforms all of them in a benchmark experiment. These results advocate for the use of domain-specific knowledge in large-scale scientific time series research to improve efficiency and reduce environmental footprint.

My STS research paper examines how government funding impacts AI development and environmental sustainability. My research question explores the effects government funding has on academic research topics and the extent to which agencies like National Science Foundation (NSF) and the Department of Energy (DOE) incorporate environmental considerations into AI project evaluations. Utilizing a systematic literature review and ethical frameworks, this study highlights the critical oversight of energy efficiency and environmental impact within governmental AI funding strategies.

The analysis reveals a significant gap: governmental funding processes prioritize immediate economic and societal benefits, often neglecting long-term environmental consequences. Case studies of large-scale AI projects illustrate substantial environmental costs due to high GPU usage and energy consumption. The conclusion calls for integrating sustainability metrics into government funding criteria, advocating a balanced approach guided by both utilitarian and environmental ethics to incentivize energy-efficient AI innovations.

Science Time Series: Deep Learning in Hydrology

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Junyang He

Spring, 2025

Technical Project Team Members

Ying-Jung Chen

Anushka Idamekorala

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Geoffrey Fox, Department of Computer Science

Science Time Series: Deep Learning in Hydrology

Junyang He
Computer Science
University of Virginia
Charlottesville, VA, USA
fay5du@virginia.edu

Ying-Jung Chen
College of Computing
Georgia Institute of Technology
Atlanta, GA, USA
yingjungcd@gmail.com

Anushka Idamekorala
Computer Science
University of Virginia
Charlottesville, VA, USA
anb5km@virginia.edu

Geoffrey Fox
Biocomplexity Institute
University of Virginia
Charlottesville, VA, USA
vxj6mb@virginia.edu

***Abstract*—This research is part of a systematic study of scientific time series. In the last three years, hundreds of papers and over fifty new deep-learning models have been described for time series models. These mainly focus on the key aspect of time dependence, whereas in some scientific time series, the situation is more complex with multiple locations, each location having multiple observed and target time-dependent streams and multiple exogenous (known) properties that are either constant or time-dependent. Here, we analyze the hydrology time series using the CAMELS and Caravan global datasets on catchment rainfall and runoff. Together, these have up to 6 observed streams and up to 209 static parameters defined at each of about 8000 locations. This analysis is fully open source with a Jupyter Notebook running on Google Colab for both an LSTM-based analysis and the data engineering preprocessing. Our goal is to investigate the importance of exogenous data, which we look at using eight different choices on representative hydrology tasks. Increasing the exogenous information significantly improves the data representation, with the mean square error decreasing to 60% of its initial value in the largest dataset examined. We present the initial results of studies of other deep-learning neural network architectures where the approaches that can use the full observed and exogenous observations outperform less flexible methods, including Foundation models. Using the natural annual periodic exogenous time series produces the largest impact, but the static and other periodic exogenous streams are also important. Our analysis is intended to be valuable as an educational resource and benchmark.**

I. Introduction

A. Spatio-temporal Series Datasets

Scientific data is frequently represented as spatio-temporal series, where time series data are often influenced by geographical factors. The language of spatio-temporal series is used as a common application type, where the "series" can refer to any ordered sequential data points. These sequences can belong to any collection (bag), not restricted to Euclidean space-time, as long as sequences are labeled in some way and have properties that are consequent to the label. In the case of COVID-19 data [1], [2], daily case / death statistics are grouped by location (e.g. city, county, country) and influenced by demographic characteristics of these locations [3]. In the case of earthquake data [4], the earthquakes are grouped by 11 km x 11 km regions [5]. Similarly, in the case of hydrology data, daily precipitation and streamflow are grouped by catchments and are affected by environmental attributes and locations of these sites.

Typically, the data in these contexts are recorded as space-time-stamped events. Fig 1. However, data can be converted into spatio-temporal series by binning in space and time. Comparing deep learning for time series with coupled ordinary differential equations for multi-particle systems motivates the use of an evolution operator to describe the time dependence of complex systems. Our research views deep learning applied to spatio-temporal series as a method for identifying the time evolution operator governing the behavior of complex systems. Metaphorically, the training process uncovers hidden variables representing the system's underlying theory, similar to Newton's laws. Previous studies on COVID-19 [3] and earthquakes [5] show neural networks' ability to learn spatiotemporal dependencies in spatio-temporal series data. This work extends this approach to hydrology, demonstrating deep learning's ability to model the rainfall-runoff process.

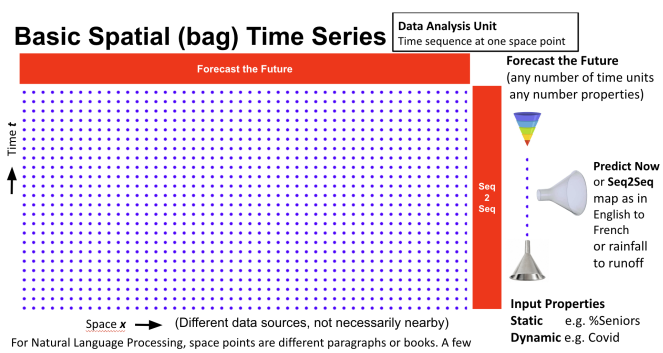

Fig. 1. Spatio-temporal series layout.

B. Rainfall-Runoff Problem

Rainfall-runoff modeling, a key challenge in hydrology, aims to model the physical process by which water on land surface (precipitation or snowmelt) moves to streams [6]. As part of the hydrologic cycle, precipitation on land either evaporates, transpires, infiltrates to recharge groundwater, or becomes surface runoff entering a catchment [6]. A catchment, or watershed, is an area where all precipitation collects and drains into a common outlet, such as a river, lake, or reservoir [6]. Surface runoff contributes to streamflow, which is defined as the volumetric discharge that takes place in a stream or channel. Runoff RO can be estimated with Equation. 1, where Q is streamflow, and GW_{out} is ground-water outflow. Stream outflow Q can be estimated with Equation. 2, where P is precipitation, GW_{in} is ground-water inflow, ET is evapotranspiration, ΔS is the change in liquid and solid forms of storage. Eventually, this streamflow and ground water outflow reaches the ocean, where it evaporates, condenses into clouds, and returns as rainfall on land, completing the hydrologic cycle [6].

$$RO = Q + GW_{out} \tag{1}$$

$$Q = P + GW_{in} - GW_{out} - ET - \Delta S \tag{2}$$

Rainfall, like many natural phenomena, is periodic, with daily patterns that follow consistent seasonal cycles. While streamflow result from precipitation, it is also influenced by static environmental factors such as soil type, land cover, slope, etc. Based on this, we hypothesize that daily streamflow in a catchment can be predicted using a combination of daily meteorological forcing data, and the spatially and temporally distributed hydrologic, climatologic, geologic, pedologic, and land-use data [6]. Additionally, a neural network can learn the seasonal patterns of these hydrological processes to produce accurate forecasts.

C. Related Work

Traditional rainfall-runoff modeling has typically focused on individual catchments. The first documented model, introduced in 1851, used linear regression to predict discharge from precipitation intensity and runoff [7]. Since then, scientific advancements have led to more sophisticated models based on mathematical formulas and physical laws. The advent of computers brought digital hydrological models. At a time when computers are expensive, slow, and low in memory, the Stanford Watershed Model [8] was proposed. It was seen as one of the first and most successful digital computer models [9]. As computers become more powerful, distributed models [10], [11] emerged, allowing for hydrological models to closely couple to geographical information systems for the input data [9]. Taking advantage of the number of parameters offered, these physical-based distributed models perform exceptionally well. However, the high computational cost to calibrate these parameters, and the limited availability of data hinder their use in large-scale forecasting applications [12].

Groundbreaking advancements in deep learning models and the publication of structured large-sample Hydrology dataset have overcome this limitation, enabling the study of nation or global scale rainfall-runoff modeling [13], [14]. In the 1990s, Artificial Neural Network (ANN) based rainfall-runoff model was proposed [15]–[17]. Although scientists were initially hesitant to embrace this novel "black box" approach due to the lack of extensive studies, deep learning based Hydrology models proved successful [17]. Its exceptional capability in simulating complex non-linear systems is particularly advantageous for Hydrology modeling. In 2018, the focus of the field shifted towards Long Short-Term Memory (LSTM) based models, which excelled in learning sequential dependencies within time series data [13] [18]. These models have shown great success in large-scale hydrological time series predictions. Further studies explored the interpretability of such LSTM models within physics context [19]. Since then, an open source library for the LSTM based rainfall-runoff model was published [20].

II. Data And Methods

A. Datasets selection

Hydrology data comprises of both time series and static exogenous features. It falls within the scope of spatio-temporal series since all time series properties (eg. mean temperature, streamflow) are collected and organized by catchment. Hydrology data is collected by gauges, which are stations that collect measurements at each catchment. Static attributes for each catchment refers to environmental conditions (eg. dominant land cover, soil aridity), as well as spatial extent and locations (eg. coordinates).

Recent deep learning studies on Hydrology have been driven by the advent of CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) datasets, which established a standard for organizing big Hydrology data at different nations across the globe. The first CAMELS dataset [21], initially proposed with 671 catchments in the U.S., benchmarked the types of static and time series properties necessary for large sample hydrology datasets containing hundreds of catchments or more. The high dimensionality of static data along with the 20-year duration of daily time series data made it suitable for nation-scale Hydrology modeling using neural network models. Since then, CAMELS-standard datasets have been published for countries including the United Kingdom [22], Chile [23], Australia [24], Brazil [25], Switzerland [26], Sweden [27], France [28], etc.

B. Three-Nation Combined CAMELS Data

Similarities shared by CAMELS-standard national datasets allow for the combination of national datasets into a large global dataset, which can be used for global-scale training. In this study, the experimental dataset was produced by combining CAMELS data from three nations: the United States [21], United Kingdom [22], and Chile. These datasets were selected as they provide the earliest available CAMELS-structured data during the data preprocessing phase of this study. We select static features and time series targets that are shared across the three datasets prior to combination.

TABLE I
Three-Nation Combined CAMELS Data Source

	CAMELS-US	**CAMELS-GB**	**CAMELS-CL**
Forcing [a]	Maurer [29]	CEH-GEAR [30] CHESS-met [32]	CR2MET [31]
Streamflow	USGS [33]	NRFA [34]	CR2 [35]

[a] Time series targets such as precipitation and temperature.

Although CAMELS-US, CAMELS-GB, CAMELS-CL follow the same standards, the exact choice of static attributes they include vary slightly, prohibiting simple concatenation of data. For instance, the “Land Cover” section of CAMELS-US only contains “% cover of forest” data for each catchment, while that section of CAMELS-GB contains the percent cover of all plantation types like woodland, crops, shrublands, etc. Without further details on the data measurement process for these properties, arithmetic manipulations to forge “percent cover of forest” for the CAMELS-GB dataset from the given “percent cover of woodland, crops, shrublands, etc” cannot be performed. To address this issue, the three-nation combined CAMELS dataset used in this study only contains static properties shared across the US, GB, and CL datasets. The processed dataset includes 1858 catchments, 3 dynamic, and 29 static variables.

C. CAMELS Preprocessing

For dynamic variables, we selected the common interval of 7,031 days, spanning from October 2, 1989, to December 31, 2008. The start date of a water year, as defined by the U.S. Geological Survey, is October 1st [6]. To account for time differences between the nations, we shifted the training data by one day, beginning on October 2nd. An analysis of NaNs in data revealed no NaN values in time series data, yet some exists in the static features. Since missing data constitutes only a small percentage of total static data and occurs miscellaneously, it is filled with the mean value of that attribute from all catchments. Categorical variables are only present in the form of months from January to December. We encode them with ordinal encoding between 0 and 1 to preserve the natural order of months.

D. Caravan

Published in 2023, the Caravan dataset [36], consists of seven preprocessed CAMELS-standard national datasets that contain identical static exogenous features and time series properties. Caravan aggregates data from 6,830 catchments across 16 nations spanning four continents, making it ideal for global rainfall-runoff modeling with large-scale hydrological data.

TABLE II
CARAVAN SUB-DATASETS

Sub-dataset	Catchments
CAMELS (US) [21]	482
CAMELS-AUS [24]	150
CAMELS-BR [25]	376
CAMELS-CL [23]	314
CAMELS-GB [22]	408
HYSETS (North America) [37]	4621
LamaH-CE [38]	479

The Caravan dataset is not constructed by identifying common static and time series properties shared by previously published versions of each sub-dataset. Instead, it is derived from data sources that differ significantly from those used in the original CAMELS datasets for various nations. While the catchments included in the Caravan sub-datasets are the same as those in the original versions, the actual data is sourced from global hydrological sources rather than the local sources used in the original publications. This shift to global sources enables the collection of more standardized and abundant data, making it suitable for global-scale comparative hydrological studies—one of the key motivations behind the creation of Caravan. However, it is important to note that data from global sources may vary slightly from local sources, which are often more precise. For example, in the CAMELS-US sub-dataset within Caravan, all forcing data is sourced from ERA5-Land [39], streamflow data from GSIM [40] [41], and static properties from HydroATLAS [42]. In contrast, the original CAMELS-US [21] dataset included static and time series data from three sources: NLDAS [43], Maurer [29], and DayMet [44], with streamflow data from USGS [33]. After comparison, only the streamflow data remains consistent between CAMELS-US in the original publication and in Caravan. The correlation coefficient for precipitation time series data between different data sources is presented in III. Nevertheless, the Caravan paper [36] addresses this concern, noting that the correlations between Caravan and each of the three CAMELS-US data products are not consistently lower than the correlations within the individual CAMELS-US data products.

TABLE III
CAMELS-US CARAVAN COMPARISON

	Maurer	**Daymet**	**NLDAS**
Caravan	0.71532	0.60720	0.75441

III. METHODS

A. Long Short-Term Memory

The Long Short-Term Memory (LSTM) network [45] is a specialized variant of a Recurrent Neural Network (RNN) designed to address the vanishing gradient problem through its unique memory cell structure. In an LSTM block (as shown in the figure below), the cell state (denoted by C) serves as long-term memory. Minimal weight updates to the cell state during backpropagation effectively mitigate the vanishing gradient problem, making LSTM particularly well-suited for time series tasks involving long sequences of data. Additionally, the hidden state (denoted by h) serves as short-term memory. Weights (denoted by W) and biases (denoted by b) are applied to the input and passed through sigmoid and tanh activation functions. The output from the hidden state is then used to update the cell state, while the updated cell state, in turn, informs the hidden state during the final stage of computation within the LSTM block.

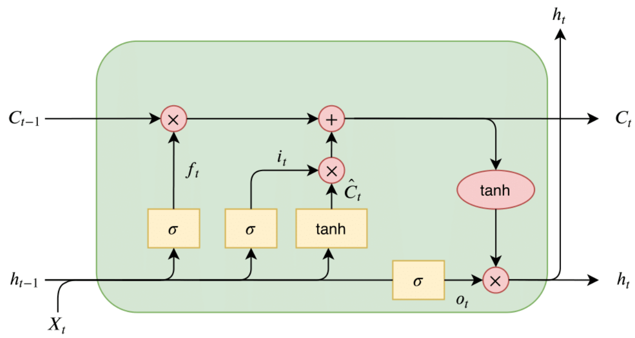

Fig. 2. LSTM network.

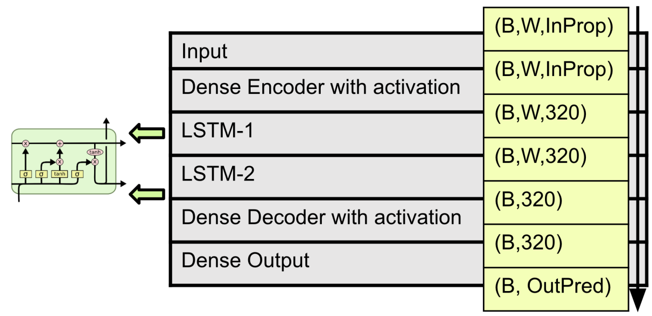

Fig. 3. CAMELS-US model architecture and layer size.

Forget gate. Initially, the input X_t is combined with the previous hidden state h_{t-1} through the forget gate, which determines how much of the long-term memory is preserved. As shown in Equation 3, f_t represents the forget gate value, W_f denotes the input weights for the forget gate, U_f refers to the recurrent weights for the forget gate, and b_f is the bias term for the forget gate.

Input gate. Next, the input X_t passes through the input gate, which determines the specific information to be incorporated into the long-term memory. As shown in Equation 4, i_t represents the input gate value, $\hat{C}_t$ denotes the temporary cell state value, W_i represents the input weights for the input gate, U_i refers to the recurrent weights for the input gate, and b_i is the bias term for the input gate.

Output gate. Lastly, the input X_t passes through the output gate, which updates the short-term memory. As shown in Equation 6, O_t represents the output gate value, W_o is the input weights for the output gate, U_o is the recurrent weights for the output gate, and b_o is the bias term for the output gate.

Cell state and hidden state. The outputs from the forget gate, input gate, and output gate are applied to the previous cell state and hidden state to calculate the new "long-term" and "short-term" memory values. As shown in Equations 7 and 8, C_t represents the updated cell state (long-term memory), and h_t represents the updated hidden state (short-term memory).

$$f_t = sigmoid(W_f \times x_t + U_f \times h_{t-1} + b_f) \tag{3}$$

$$i_t = sigmoid(W_i \times x_t + U_i \times h_{t-1} + b_i) \tag{4}$$

$$\hat{C}_t = tanh(W_i \times x_t + U_i \times h_{t-1} + b_i) \tag{5}$$

$$O_t = sigmoid(W_o \times x_t + U_o \times h_{t-1} + b_o) \tag{6}$$

$$C_t = C_{t-1} \times f_t + i_t \times \hat{C}_t \tag{7}$$

$$h_t = O_t \times tanh(C_t) \tag{8}$$

B. Model Setup

The LSTM model used in this study is implemented using the Tensorflow framework, which allows for customization of layer parameters. The baseline model architecture is shown in Fig. 3. Detailed activation function setup is shown in Tab. IV. To prevent overfitting, a dropout rate of 20% is applied to the layers.

TABLE IV
MODEL ACTIVATION SETUP

Dense Encoder Activation	SELU
LSTM Recurrent Activation	Sigmoid
LSTM Layer Activation	SELU
Dense Decoder Activation	SELU

C. Model Training and Evaluation

Inputs to the model can be classified into known inputs and observed inputs [5]. Known inputs refer to features that are known in both the past and future. In this hydrology study, static known inputs are the exogenous features such as climatic signatures, hydrologic signatures, and catchment topography. Additionally, time series known inputs exist in the form of seasonal patterns hidden in all time series features. Observed inputs are features known only for the past time periods but unknown in the future. In hydrology, those features include precipitation, temperature, and streamflow. Model output, known as targets, are the time-dependent predicted properties. In this study, predicted targets are precipitation, temperature, and streamflow.

The input data is divided into batches with a sequence length of 21 days, selected after testing various other lengths, including 7, 14, and 365 days. The 3-week sequence length was chosen to effectively capture subtle hydrological patterns. The total number of batches processed during one epoch is calculated using Equation 9, while the batch size is determined by Equation 10. In these equations, Day_{total} represents the total duration of the input data in days, l_{seq} denotes the sequence length, $\#Gauges$ refers to the total number of gauges in the input data, and $\#Input\ properties$ indicates the total number of input features, including both static and time series properties.

$$\#batch/epoch = Day_{total} - l_{seq} + 1 \tag{9}$$

$$S_{batch} = l_{seq} \times \#Gauges \times \#Input\ properties \tag{10}$$

Symbolic Window. A key challenge encountered during LSTM model training was the space complexity associated with handling the training data. Under the previously mentioned batch size configuration, storing all batches prior to training requires RAM space of $O(S_{batch} \times \#batch/epoch)$. To address this issue, we leverage a symbolic window to dynamically generate batches during each epoch of training. Instead of storing the entire training set, including all batches, before training, we only store the original input data with a space complexity of $O(Day_{total} \times \#Gauges \times \#Input\ properties)$. During each epoch, we track the start index i for the current batch and extract the corresponding data for the batch from the input data using $Data[day_i : day_{i+l_{seq}}]$. This batch is then trained, and the index i is incremented before proceeding to extract the next batch. This process is repeated $\#batch/epoch$ times to complete a full epoch of training. Note: This technique is environment-dependent and may not be applicable to all frameworks.

Spatial and Temporal Encodings. Like many fields of science, hydrology time series data exhibit strong seasonal patterns. It can reasonably be assumed that, at a specific gauge location, the precipitation and streamflow in October of one year will be similar to those recorded in October of the previous year. Furthermore, research has identified the approximate water residence times in various reservoirs, such as rivers, lakes, soil, and the atmosphere [46]. To effectively capture these known dependencies within the time series properties, spatial and temporal encodings are incorporated into the model during training.

1) **Linear Space:** a linear function with length equaling total number of catchments in input data.
2) **Linear Time:** a linear function with length equaling total number of days in input time series.
3) **Annual Fourier Time:** a basic sine and cosine function with period equaling one year.
4) **Extra Fourier Time:** basic sine and cosine functions with period equaling 8, 16, 32, 64, 128 days.
5) **Legendre Time:** Legendre functions of degree 2, 3, and 4 with range equaling total number of days in input time series.

Spatial Validation. We employ location-based rather than temporal-based validation as the catchments in the CAMELS and Caravan datasets are uncorrelated. The training and validation datasets are randomly selected on an 8:2 ratio by location, given the extensive number of catchments included in the datasets used in this study. For example, out of the 671 catchments included in CAMELS-US dataset, 537 catchments are used for training and 134 catchments are used for validation.

NNSE. Model fit is quantified using normalized Nash-Sutcliffe Efficiency (NNSE) scores [47], calculated using Equation 11, where T represents the total number of days, Q_m^t represents the modeled discharge on day t, Q_o^t represents the observed discharge on day t, and $\overline{Q}_o$ is the mean observed discharge over T days. NNSE values range from 0 (poor performance) to 1 (perfect fit), with a score of 0.5 indicating that the model's predictions are equivalent to the time-averaged mean of the observations. In this study, the NNSE value is calculated for each gauge over time T, and the average NNSE across all gauges in the input data is reported.

$$NNSE = \frac{\Sigma_{t=1}^{T}(Q_m^t - \overline{Q}_o)^2}{\Sigma_{t=1}^{t}(Q_o^t - Q_m^t)^2 + (Q_o^t - \overline{Q}_o)^2} \tag{11}$$

IV. LSTM Benchmark Runs

A. Experiment Setup

All experiments are conducted using the model architecture shown in Fig. 3 and the activation functions listed in Table IV. The benchmark runs utilize **Linear Space**, **Linear Time**, and **Annual Fourier Time** encodings. Streamflow time series data is not trained but predicted as target.

B. Three-Nation Combined Benchmark Run

This run evaluates model performance on the combined CAMELS dataset from three nations: the US, UK, and Chile. The input data include static properties that are common across all three regions, whereas predicted targets consist of precipitation, mean temperature, and streamflow. Results are demonstrated in Table V.

TABLE V
CAMELS Three-Nations Combined Benchmark Run

		MSE	NNSE
Precipitation	Train	0.002764	0.847
	Val	0.003200	0.836
Mean Temperature	Train	0.000212	0.933
	Val	0.000356	0.933
Streamflow	Train	0.000432	0.697
	Val	0.000613	0.654
Total	Train	0.003438	-
	Val	0.004325	-

C. CAMELS Caravan US Benchmark Runs

The runs compare model performance between the original CAMELS-US dataset and the US sub-dataset within the Caravan dataset. The CAMELS-US model is trained using selected static and time series data from the original CAMELS-US dataset, while the Caravan-US model is trained using selected static and time series data from the US sub-dataset within Caravan. For both runs, we utilize the same time series features—precipitation and mean temperature—for training and predict the same targets: precipitation, mean temperature, and streamflow. Results are presented in Table VI.

D. Caravan PCA Runs

This experiment explores the use of Principal Component Analysis (PCA) [48] to reduce the dimensionality of static properties in the Caravan input data. The Caravan sub-datasets contain over 200 static properties, nearly seven times more

TABLE VI
CAMELS CARAVAN US COMPARISON

		CAMELS-US		Caravan US	
		MSE	NNSE	MSE	NNSE
Precipitation	Train	0.003508	0.820	0.002920	0.851
	Val	0.003585	0.819	0.004307	0.801
Mean Temperature	Train	0.000276	0.961	0.000468	0.967
	Val	0.000283	0.960	0.000573	0.963
Streamflow	Train	0.000287	0.806	0.000525	0.814
	Val	0.000296	0.812	0.000955	0.703
Total	Train	0.004111	-	0.003982	-
	Val	0.004203	-	0.005987	-

TABLE VII
CARAVAN US PCA EXPERIMENT

		Original Static		PCA Static	
		MSE	NNSE	MSE	NNSE
Precipitation	Train	0.002920	0.851	0.002994	0.848
	Val	0.004307	0.801	0.004264	0.800
Mean Temperature	Train	0.000468	0.967	0.000468	0.967
	Val	0.000573	0.963	0.000548	0.965
Streamflow	Train	0.000525	0.814	0.000569	0.799
	Val	0.000955	0.703	0.000989	0.703
Total	Train	0.003982	-	0.004075	-
	Val	0.005987	-	0.005871	-

than the number of static input properties used in the original CAMELS studies. While this extensive range of static properties enhances model training, it significantly increases computational demands and GPU usage. To address this, we apply PCA, a widely adopted dimensionality reduction technique, to reduce the number of static input features to a level comparable with that in the CAMELS studies. In this experiment, we set the explained variance threshold to 90%, resulting in the reduction of static input features to approximately 30.

The first part of the experiment assesses the effect of PCA on model trained on US sub-dataset within the Caravan dataset, representing a small-scale input. The second part of the experiment examines the effect of PCA on model trained on North America regional data (HYSETS) [37] within the Caravan dataset, representing a large-scale input.

Results, shown in Table VII and Table VIII, indicate that the models trained with static properties obtained from PCA perform comparably to the models trained with original static properties, demonstrating that the reduction in input static dimensionality does not significantly compromise model accuracy. These findings validate PCA as an effective approach for train hydrology time series models with high static dimensionality.

TABLE VIII
CARAVAN HYSETS PCA EXPERIMENT

		Original Static		PCA Static	
		MSE	NNSE	MSE	NNSE
Precipitation	Train	0.002714	0.835	0.002809	0.830
	Val	0.002960	0.826	0.003055	0.821
Mean Temperature	Train	0.000372	0.968	0.000386	0.967
	Val	0.000385	0.966	0.000397	0.965
Streamflow	Train	0.000614	0.825	0.000699	0.813
	Val	0.000876	0.798	0.000397	0.799
Total	Train	0.003470	-	0.003611	-
	Val	0.003897	-	0.003926	-

E. Caravan Global Benchmark Run

This study provides insights into the model's applicability to large-scale global hydrology datasets. The input data is compiled by concatenating all seven Caravan sub-datasets, which encompass catchments from four continents. As shown in Table IX, accuracy decreases slightly compared to small-scale, individual nation fits (Table VI), yet overall model performance remains relatively high, demonstrating its robustness on global datasets.

TABLE IX
CARAVAN GLOBAL FIT

		MSE	NNSE
Precipitation	Train	0.003115	0.809
	Val	0.003223	0.808
Mean Temperature	Train	0.000357	0.953
	Val	0.000357	0.952
Streamflow	Train	0.000675	0.781
	Val	0.000740	0.768
Total	Train	0.003965	-
	Val	0.004133	-

*Model trained with PCA static features.

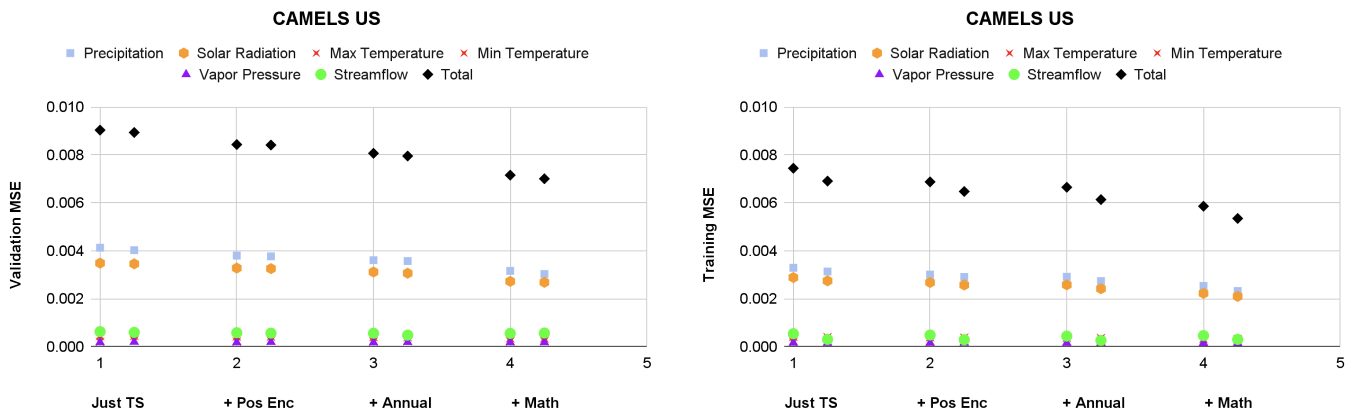

Fig. 4. CAMELS US experiment (MSE loss). Each run configuration is utilized in 2 runs (left: static features not trained, right: static features trained).

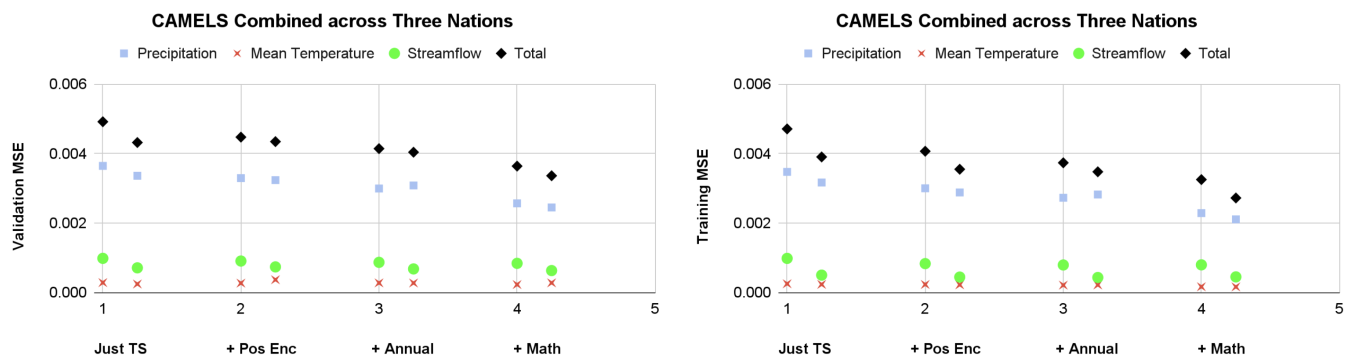

Fig. 5. CAMELS three-nations combined experiment (MSE loss). Each run configuration is utilized in 2 runs (left: static features not trained, right: static features trained).

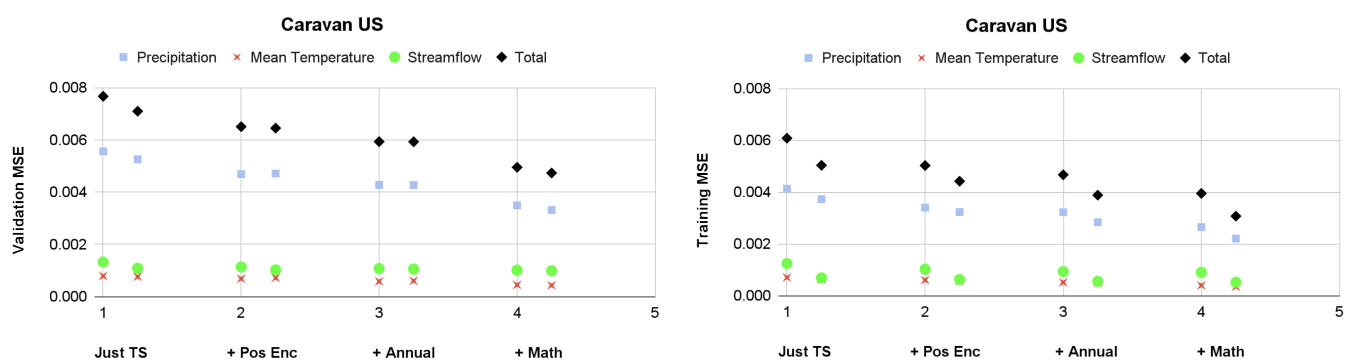

Fig. 6. Caravan US experiment (MSE loss). Each run configuration is utilized in 2 runs (left: static features not trained, right: static features trained).

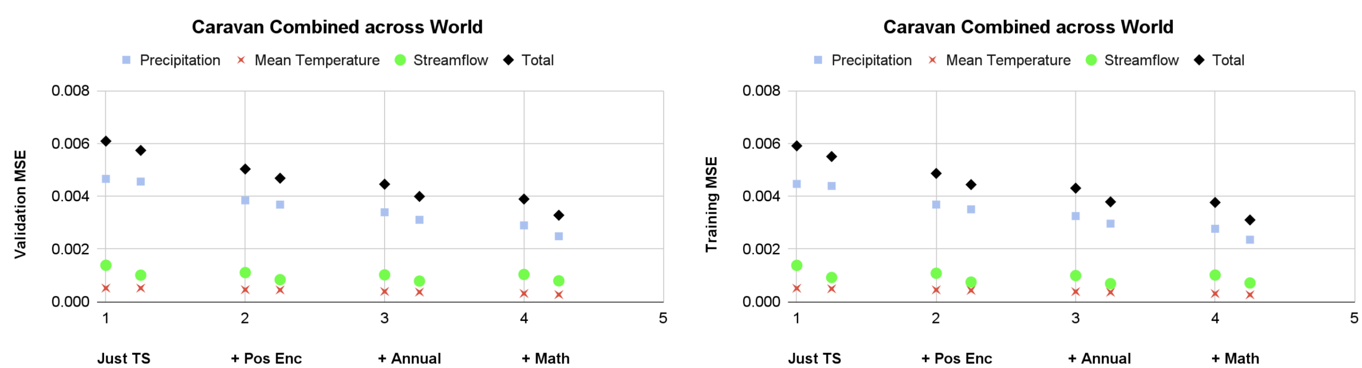

Fig. 7. Caravan combined experiment (MSE loss). Each run configuration is utilized in 2 runs (left: static features not trained, right: static features trained).

V. Static Properties and Spatial Temporal Encodings Experiment

A. Experiment Setup

This set of experiments examines the influence of static properties and spatial-temporal encodings on rainfall-runoff modeling accuracy, demonstrated through eight run configurations on four datasets. The first experiment uses the original CAMELS-US dataset [21], representing a traditional, locally-sourced small-scale dataset. Results are presented in Fig. 4. The second experiment utilizes the Three-Nation Combined CAMELS dataset [21]–[23], representing a locally-sourced mid-scale dataset. Results are shown in Fig. 5. The third experiment is conducted on the US sub-dataset of the Caravan dataset [36], representing a globally-sourced small-scale dataset. Results are shown in Fig. 6. The fourth experiment uses the full Caravan concatenated dataset [36], representing a globally-sourced large-scale dataset. Results are shown in Fig. 7. All four spatial-temporal encoding configurations tested are presented in Table X. For each configuration, we conduct runs with and without static features as input to highlight the impact of static features. The model architecture remained consistent across all runs.

TABLE X
STATIC PROPERTIES AND ENCODINGS EXPERIMENT CONFIGURATIONS

	Config 1	Config 2	Config 3	Config 4
Time Series Input	✓	✓	✓	✓
Linear Space		✓	✓	✓
Linear Time		✓	✓	✓
Annual Fourier Time			✓	✓
Extra Fourier Time				✓
Legendre Time				✓

B. Experiment Findings

Results suggest that while the addition of static features in training has a marginal effect, the LSTM network generally benefits from their inclusion. This is demonstrated by the slightly lower training and validation losses in the runs conducted with static input features for each encoding configuration, as shown in Fig. 4, Fig. 5, Fig. 6, and Fig. 7. Furthermore, the figures indicate that static input features have a greater impact on large-scale datasets compared to small-scale ones, and they appear to be more beneficial for models trained on the Caravan dataset than those trained on the CAMELS datasets. We hypothesize that this is due to the greater number of static properties used in the Caravan runs—approximately five times more than in the CAMELS runs—and that larger datasets, covering a greater number of catchments, introduce more complexity.

This study further demonstrates that spatial and temporal encoding are crucial for effectively training time series data that follow seasonal patterns. The incorporation of linear spatial-temporal encoding, as evidenced by the drop in loss from run configuration 1 to run configuration 2 in the plots, captures both the spatial relationships of catchments and the time dependence of the data. The inclusion of annual Fourier temporal encoding, shown by the drop in loss from run configuration 2 to run configuration 3, captures the yearly seasonality of hydrological data. The addition of extra Fourier and Legendre temporal encoding, indicated by the drop in loss from run configuration 3 to run configuration 4, captures both known and unknown hydrological patterns of varying lengths, thereby enhancing model performance.

In hydrology, it is reasonable to assume that interrelated time series targets are highly dependent on static properties, such as land characteristics at individual gauges. Additionally, the water cycle is governed by processes with both known and unknown periodicities, ranging from atmospheric to micro scales. However, deep learning-based time series studies often overlook the significance of these characteristics, as they are typically application-specific. Notably, recent studies tend to focus on developing state-of-the-art time series forecasting models that rely solely on raw time series data. While these foundational models allow for broad applicability across various fields, they may compromise prediction accuracy in specialized downstream applications if static features or known patterns are not incorporated into the training process.

VI. Comparison with Foundation Models

Recent studies in time series analysis have focused on a foundation model approach, inspired by the success of large language models [49]. Since 2022, there has been a spike in publications on pre-trained time series foundation models, with the majority built on Transformers [50] and multi-layer perceptrons (MLPs) [51] architectures. Some models support the use of static exogenous variables as input [52], [53], while others only accept time series input [54].

In this experiment, we present a performance comparison between our LSTM-based rainfall-runoff model and the TSMixer foundation model [52] on CAMELS-US [21] data. We leverage the Nixtla Neuralforecast framework [55] to train the TSMixerx model, a variant of the TSMixer model that allows static exogenous features along with multivariate time series as input. The results, shown in Fig. 8, Fig. 9, Fig. 10, and Fig. 11, demonstrate that the LSTM model outperforms the TSMixerx model. In future work, we will expand this by looking at other time series models and a new foundation model MultiFoundationPattern discussed for earthquakes [56].

VII. Conclusion and Future Work

Often, one thinks about deep learning models discovering hidden variables that control observables that one wishes to predict. In this paper, we explored CAMELS and Caravan, which together offer large datasets covering many (about 8000) spatial locations and a long time period (around 30 years). Further, there are multiple (6 in CAMELS and 39 in Caravan) observed dynamic streams and up to 209 static (exogenous) parameters for each location. Some of these exogenous and

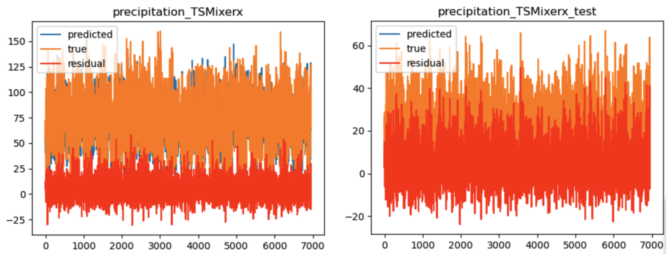

Fig. 8. TSMixer precipitation fit (left: training set, right: validation set).

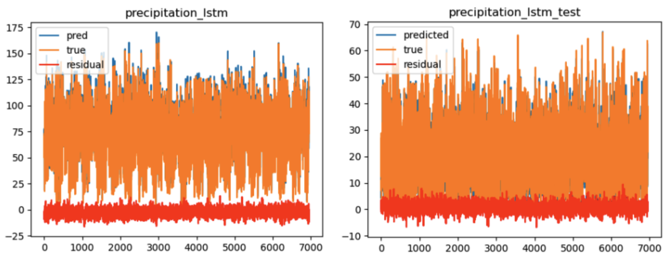

Fig. 9. LSTM precipitation fit (left: training set, right: validation set).

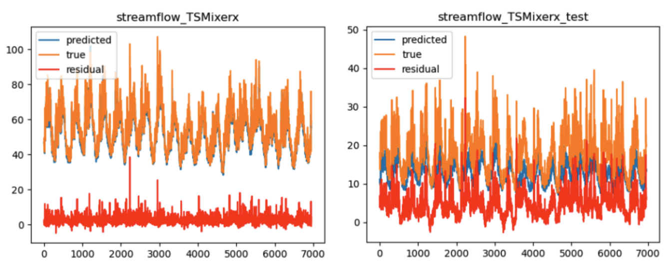

Fig. 10. TSMixer streamflow fit (left: training set, right: validation set).

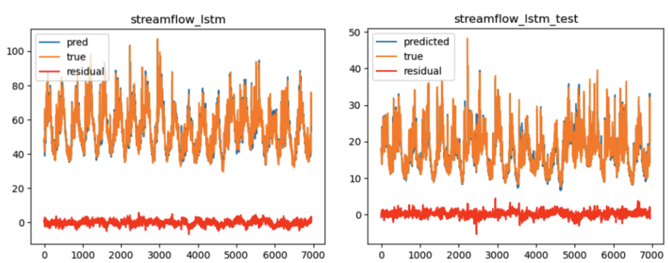

Fig. 11. LSTM streamflow fit (left: training set, right: validation set).

observed data would be natural inputs into a physics model for the catchments. A priori, it is not obvious which exogenous or even endogenous data should be used as they could implied by the large datasets used in the training and implicitly learned by the training. However, most likely, it is a mixed situation where the combination of observed and exogenous data will give the best results, with the exogenous data declining in importance as the observed dataset grows in size. We clearly show in this paper with the CAMELS and Caravan data that using exogenous data makes significant improvements in the model fits where both static physical quantities and dynamical mathematical functions have value as exogenous information. In future papers, we hope to understand the trade-off better as it varies in the nature and size of exogenous data. We also will present further study of different deep learning models using the compilation of hundreds of papers and around 100 models in [57]–[61]. We aim to study a range of time series problems to build a taxonomy so that benchmark sets can be built covering the essentially different cases [62]. Another interesting feature of science time series is that they naturally vary in magnitude by large factors, and this seems not to be very consistent with neural network activation functions at fixed values independent of the stream. We will present a study of new activation functions in the LSTM using a PRELU-like (Parametric Rectified Linear Unit, or PReLU) structure, which naturally deals with magnitude changes in stream values.

ACKNOWLEDGMENT

We would like to thank Gregor von Laszewski, Niranda Perera, and Alireza Jafari for their contributions and guidance. We gratefully acknowledge the partial support of DE-SC0023452: FAIR Surrogate Benchmarks Supporting AI and Simulation Research and the Biocomplexity Institute at the University of Virginia.

REFERENCES

[1] CDC, "National Center for Health Statistics — cdc.gov," https://www.cdc.gov/nchs/index.html, [Accessed 04-10-2024].
[2] "Behavioral Risk Factor Surveillance System — cdc.gov," https://www.cdc.gov/brfss/index.html, [Accessed 04-10-2024].
[3] G. C. Fox, G. von Laszewski, F. Wang, and S. Pyne, "Aicov: An integrative deep learning framework for covid-19 forecasting with population covariates," *Journal of Data Science*, vol. 19, no. 2, pp. 293–313, 2021.
[4] "Earthquake hazards program of united states geological survey," https://earthquake.usgs.gov/earthquakes/search/, [Accessed 04-10-2024].
[5] G. C. Fox, J. B. Rundle, A. Donnellan, and B. Feng, "Earthquake nowcasting with deep learning," *GeoHazards*, vol. 3, no. 2, pp. 199–226, 2022. [Online]. Available: https://www.mdpi.com/2624-795X/3/2/11
[6] S. L. Dingman, *Physical hydrology*. Waveland press, 2015.
[7] M. T. J., "On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment," *Proceedings of the Institution of Civil Engineers of Ireland*, vol. 4, pp. 19–31, 1851. [Online]. Available: https://cir.nii.ac.jp/crid/1573105975719082240
[8] N. H. Crawford, *Digital simulation in hydrology : Stanford watershed model IV / by N.H. Crawford and Ray K. Linsley.*, ser. Stanford University Department of Civil Engineering Technical report ;, . Linsley, Ray K. (Ray Keyes), Ed. Stanford University. Dept of Civil Engineering, 1966.
[9] K. J. Beven, Ebook Central - Academic Complete, Wiley Online Library UBCM Earth & Environmental, Wiley Online Library UBCM German Language, and Wiley Online Library UBCM All Obooks, *Rainfall-runoff Modelling: The Primer*. Chichester, West Sussex, Hoboken, NJ: Wiley-Blackwell, 2012.
[10] M. Abbott, J. Bathurst, J. Cunge, P. O'Connell, and J. Rasmussen, "An introduction to the european hydrological system — systeme hydrologique europeen, "she", 1: History and philosophy of a physically-based, distributed modelling system," *Journal of Hydrology*, vol. 87, no. 1, pp. 45–59, 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022169486901149
[11] R. J. Moore and R. T. Clarke, "A distribution function approach to rainfall runoff modeling," *Water Resources Research*, vol. 17, no. 5, pp. 1367–1382, 1981.
[12] J. Sitterson, R. P. Chris Knightes, K. Wolfe, M. Muche, and B. Avant, *An Overview of Rainfall-Runoff Model Types*. U.S. Environmental Protection Agency, Washington, DC, 2017.
[13] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (lstm) networks," *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 6005–6022, 2018. [Online]. Available: https://hess.copernicus.org/articles/22/6005/2018/

[14] C. Shen, “A transdisciplinary review of deep learning research and its relevance for water resources scientists,” *Water Resources Research*, vol. 54, no. 11, pp. 8558–8593, 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022643

[15] K.-l. Hsu, H. V. Gupta, and S. Sorooshian, “Artificial neural network modeling of the rainfall-runoff process,” *Water Resources Research*, vol. 31, no. 10, pp. 2517–2530, 1995. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95WR01955

[16] Tokar A. Sezin and Johnson Peggy A., “Rainfall-Runoff Modeling Using Artificial Neural Networks,” *Journal of Hydrologic Engineering*, vol. 4, no. 3, pp. 232–239, Jul. 1999, publisher: American Society of Civil Engineers. [Online]. Available: https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(232)

[17] R. S. Govindaraju and A. R. Rao, *Artificial Neural Networks in Hydrology*. Springer Science & Business Media, Mar. 2013.

[18] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, “Deep learning with a long short-term memory networks approach for rainfall-runoff simulation,” *Water*, vol. 10, no. 11, 2018. [Online]. Available: https://www.mdpi.com/2073-4441/10/11/1543

[19] F. Kratzert, M. Herrnegger, D. Klotz, S. Hochreiter, and G. Klambauer, *NeuralHydrology – Interpreting LSTMs in Hydrology*. Cham: Springer International Publishing, 2019, pp. 347–362. [Online]. Available: https://doi.org/10.1007/978-3-030-28954-6_19

[20] F. Kratzert, M. Gauch, G. Nearing, and D. Klotz, “NeuralHydrology — A Python library for Deep Learning research in hydrology,” *Journal of Open Source Software*, vol. 7, no. 71, p. 4050, Mar. 2022. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.04050

[21] N. Addor, A. J. Newman, N. Mizukami, and M. P. Clark, “The camels data set: catchment attributes and meteorology for large-sample studies,” *Hydrology and Earth System Sciences*, vol. 21, no. 10, pp. 5293–5313, 2017. [Online]. Available: https://hess.copernicus.org/articles/21/5293/2017/

[22] G. Coxon, N. Addor, J. P. Bloomfield, J. Freer, M. Fry, J. Hannaford, N. J. K. Howden, R. Lane, M. Lewis, E. L. Robinson, T. Wagener, and R. Woods, “Camels-gb: hydrometeorological time series and landscape attributes for 671 catchments in great britain,” *Earth System Science Data*, vol. 12, no. 4, pp. 2459–2483, 2020. [Online]. Available: https://essd.copernicus.org/articles/12/2459/2020/

[23] C. Alvarez-Garreton, P. A. Mendoza, J. P. Boisier, N. Addor, M. Galleguillos, M. Zambrano-Bigiarini, A. Lara, C. Puelma, G. Cortes, R. Garreaud, J. McPhee, and A. Ayala, “The camels-cl dataset: catchment attributes and meteorology for large sample studies – chile dataset,” *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 5817–5846, 2018. [Online]. Available: https://hess.copernicus.org/articles/22/5817/2018/

[24] K. J. A. Fowler, S. C. Acharya, N. Addor, C. Chou, and M. C. Peel, “Camels-aus: hydrometeorological time series and landscape attributes for 222 catchments in australia,” *Earth System Science Data*, vol. 13, no. 8, pp. 3847–3867, 2021. [Online]. Available: https://essd.copernicus.org/articles/13/3847/2021/

[25] V. B. P. Chagas, P. L. B. Chaffe, N. Addor, F. M. Fan, A. S. Fleischmann, R. C. D. Paiva, and V. A. Siqueira, “Camels-br: hydrometeorological time series and landscape attributes for 897 catchments in brazil,” *Earth System Science Data*, vol. 12, no. 3, pp. 2075–2096, 2020. [Online]. Available: https://essd.copernicus.org/articles/12/2075/2020/

[26] M. Höge, M. Kauzlaric, R. Siber, U. Schönenberger, P. Horton, J. Schwanbeck, M. G. Floriancic, D. Viviroli, S. Wilhelm, A. E. Sikorska-Senoner, N. Addor, M. Brunner, S. Pool, M. Zappa, and F. Fenicia, “Camels-ch: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic switzerland,” *Earth System Science Data*, vol. 15, no. 12, pp. 5755–5784, 2023. [Online]. Available: https://essd.copernicus.org/articles/15/5755/2023/

[27] C. Teutschbein, “Camels-se : Long-term hydroclimatic observations (1961–2020) across 50 catchments in sweden as a resource for modelling, education, and collaboration,” *Geoscience Data Journal*, 2024.

[28] O. Delaigue, P. Brigode, V. Andréassian, C. Perrin, P. Etchevers, J.-M. Soubeyroux, B. Janet, and N. Addor, “Camels-fr: A large sample hydroclimatic dataset for france to explore hydrological diversity and support model benchmarking,” in *IAHS-2022 Scientific Assembly*, Montpellier, France, May 2022. [Online]. Available: https://hal.inrae.fr/hal-03687235

[29] B. Livneh, E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, “A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states: Update and extensions,” *Journal of Climate*, vol. 26, no. 23, pp. 9384 – 9392, 2013. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/26/23/jcli-d-12-00508.1.xml

[30] V. D. J. Keller, M. Tanguy, I. Prosdocimi, J. A. Terry, O. Hitt, S. J. Cole, M. Fry, D. G. Morris, and H. Dixon, “Ceh-gear: 1 km resolution daily and monthly areal rainfall estimates for the uk for hydrological and other applications,” *Earth System Science Data*, vol. 7, no. 1, pp. 143–155, 2015. [Online]. Available: https://essd.copernicus.org/articles/7/143/2015/

[31] J. P. Boisier, “Cr2met: A high-resolution precipitation and temperature dataset for the period 1960-2021 in continental chile.” Jan. 2023.

[32] E. Robinson, E. Blyth, D. Clark, E. Comyn-Platt, J. Finch, and A. Rudd, “Climate hydrology and ecology research support system meteorology dataset for great britain (1961-2015) [chess-met] v1.2,” 2017. [Online]. Available: https://doi.org/10.5285/b745e7b1-626c-4ccc-ac27-56582e77b900

[33] H. ”Lins, “”usgs hydro-climatic data network 2009 (hcdn–2009)”,” *”U.S. Geological Survey Fact Sheet 2012–3047”*, 2012. [Online]. Available: ”https://pubs.usgs.gov/fs/2012/3047/”

[34] J. Hannaford, “Development of a strategic data management system for a national hydrological database, the uk national river flow archive,” *Hydroinformatics*, pp. 637–644, 2004.

[35] C. for Climate and R. Research, “Streamflow.” [Online]. Available: https://www.cr2.cl/datos-de-caudales/

[36] F. Kratzert, G. Nearing, N. Addor, T. Erickson, M. Gauch, O. Gilon, L. Gudmundsson, A. Hassidim, D. Klotz, S. Nevo, G. Shalev, and Y. Matias, “Caravan - a global community dataset for large-sample hydrology,” *Scientific Data*, vol. 10, no. 1, p. 61, Jan. 2023. [Online]. Available: https://doi.org/10.1038/s41597-023-01975-w

[37] R. Arsenault, F. Brissette, J.-L. Martel, M. Troin, G. Lévesque, J. Davidson-Chaput, M. C. Gonzalez, A. Ameli, and A. Poulin, “A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds,” *Scientific Data*, vol. 7, no. 1, p. 243, Jul. 2020. [Online]. Available: https://doi.org/10.1038/s41597-020-00583-2

[38] C. Klingler, K. Schulz, and M. Herrnegger, “Lamah-ce: Large-sample data for hydrology and environmental sciences for central europe,” *Earth System Science Data*, vol. 13, no. 9, pp. 4529–4565, 2021. [Online]. Available: https://essd.copernicus.org/articles/13/4529/2021/

[39] J. Muñoz Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, “Era5-land: a state-of-the-art global reanalysis dataset for land applications,” *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, 2021. [Online]. Available: https://essd.copernicus.org/articles/13/4349/2021/

[40] H. X. Do, L. Gudmundsson, M. Leonard, and S. Westra, “The global streamflow indices and metadata archive (gsim) – part 1: The production of a daily streamflow archive and metadata,” *Earth System Science Data*, vol. 10, no. 2, pp. 765–785, 2018. [Online]. Available: https://essd.copernicus.org/articles/10/765/2018/

[41] L. Gudmundsson, H. X. Do, M. Leonard, and S. Westra, “The global streamflow indices and metadata archive (gsim) – part 2: Quality control, time-series indices and homogeneity assessment,” *Earth System Science Data*, vol. 10, no. 2, pp. 787–804, 2018. [Online]. Available: https://essd.copernicus.org/articles/10/787/2018/

[42] S. Linke, B. Lehner, C. Ouellet Dallaire, J. Ariwi, G. Grill, M. Anand, P. Beames, V. Burchard-Levine, S. Maxwell, H. Moidu, F. Tan, and M. Thieme, “Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution,” *Scientific Data*, vol. 6, no. 1, p. 283, Dec. 2019. [Online]. Available: https://doi.org/10.1038/s41597-019-0300-6

[43] Y. Xia, K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko, “Continental-scale water and energy flux analysis and validation for the north american land data assimilation system project phase 2 (nldas-2): 1. intercomparison and application of model products,” *Journal of Geophysical Research Atmospheres*, vol. 117, no. 3, 2012, copyright: Copyright 2018 Elsevier B.V., All rights reserved.

[44] P. THORNTON, M. THORNTON, B. MAYER, N. WILHELMI, Y. WEI, R. DEVARAKONDA, and R. COOK, “Daymet: Daily surface

weather data on a 1-km grid for north america, version 2," 2014. [Online]. Available: http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1219

[45] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[46] M. Pidwirny, "The hydrologic cycle," 2006. [Online]. Available: http://www.physicalgeography.net/fundamentals/8b.html

[47] J. Nossent and W. Bauwens, "Application of a normalized Nash-Sutcliffe efficiency to improve the accuracy of the Sobol' sensitivity analysis of a hydrological model," in *EGU General Assembly Conference Abstracts*, ser. EGU General Assembly Conference Abstracts, Apr. 2012, p. 237.

[48] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987, proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0169743987800849

[49] S. Mirchandani, F. Xia, P. Florence, B. Ichter, D. Driess, M. G. Arenas, K. Rao, D. Sadigh, and A. Zeng, "Large language models as general pattern machines," 2023. [Online]. Available: https://arxiv.org/abs/2307.04721

[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: https://arxiv.org/abs/1706.03762

[51] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "Mlp-mixer: An all-mlp architecture for vision," 2021. [Online]. Available: https://arxiv.org/abs/2105.01601

[52] S.-A. Chen, C.-L. Li, N. Yoder, S. O. Arik, and T. Pfister, "Tsmixer: An all-mlp architecture for time series forecasting," 2023. [Online]. Available: https://arxiv.org/abs/2303.06053

[53] A. Das, W. Kong, A. Leach, S. Mathur, R. Sen, and R. Yu, "Long-term forecasting with tide: Time-series dense encoder," 2024. [Online]. Available: https://arxiv.org/abs/2304.08424

[54] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, S. Kapoor, J. Zschiegner, D. C. Maddix, H. Wang, M. W. Mahoney, K. Torkkola, A. G. Wilson, M. Bohlke-Schneider, and Y. Wang, "Chronos: Learning the language of time series," 2024. [Online]. Available: https://arxiv.org/abs/2403.07815

[55] K. G. Olivares, C. Challú, F. Garza, M. M. Canseco, and A. Dubrawski, "Neuralforecast: User friendly state-of-the-art neuralforecasting models." PyCon Salt Lake City, Utah, US 2022, 2022. [Online]. Available: https://github.com/Nixtla/neuralforecast

[56] A. Jafari, G. Fox, J. B. Rundle, A. Donnellan, and L. G. Ludwig, "Time series foundation models and deep learning architectures for earthquake temporal and spatial nowcasting," *arXiv [cs.LG, physics.geo-ph]*, 22 Aug. 2024. [Online]. Available: http://arxiv.org/abs/2408.11990

[57] "ScienceFMHub portal for science foundation model community," http://sciencefmhub.org, 2 Nov. 2023, accessed: 2023-11-3.

[58] "Nixtla: A future for everybody. time series research and deployment," https://www.nixtla.io/, accessed: 2024-3-9.

[59] ddz, "Awesome time series forecasting/prediction papers: a reading list of papers on time series forecasting/prediction (TSF) and spatio-temporal forecasting/prediction (STF)," https://github.com/ddz16/TSFpaper, accessed: 2024-5-16.

[60] M. Jin, Q. Wen, Y. Liang, C. Zhang, S. Xue, X. Wang, J. Zhang, Y. Wang, H. Chen, X. Li, S. Pan, V. S. Tseng, Y. Zheng, L. Chen, and H. Xiong, "Large models for time series and spatio-temporal data: A survey and outlook," *arXiv [cs.LG]*, 16 Oct. 2023. [Online]. Available: http://arxiv.org/abs/2310.10196

[61] Y. Liang, H. Wen, Y. Nie, Y. Jiang, M. Jin, D. Song, S. Pan, and Q. Wen, "Foundation models for time series analysis: A tutorial and survey," *arXiv [cs.LG]*, 21 Mar. 2024. [Online]. Available: http://arxiv.org/abs/2403.14735

[62] X. Huang, G. C. Fox, S. Serebryakov, A. Mohan, P. Morkisz, and D. Dutta, "Benchmarking deep learning for time series: Challenges and directions," in *2019 IEEE International Conference on Big Data (Big Data)*. ieeexplore.ieee.org, Dec. 2019, pp. 5679–5682. [Online]. Available: http://dx.doi.org/10.1109/BigData47090.2019.9005496

Environmental Considerations in AI Project Funding: A Government Grant Evaluation

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Junyang He

Spring 2025

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Pedro A. P. Francisco, Department of Engineering and Society

1. Introduction

The advent of big data and powerful machine learning models has facilitated a data-driven approach to solving complex scientific challenges such as disaster prevention, climate modeling, and weather forecasting (Choudhary et al., 2022; Wang, H. et al., 2023). The success of large language models like GPT has spurred widespread interest in developing large-scale models for global-scale modeling and prediction tasks across various domains.

From tech companies to the research communities in academia, an arms race for graphics processing unit (GPU) and computing power has begun, often rarely accounting for the enormous energy usage and environmental footprint of the systems that power machine learning innovations (Crawford, 2021). If computational optimization techniques fail to lower the energy usage of these models, if a worldwide AI political battle begins, or if clean energy supply cannot meet the energy demand of machine learning innovations, we will soon witness environmental damages that outweigh the benefits of these advanced models.

Regardless, no research or product development can be conducted without adequate funding. The sponsors are responsible for considering the negative consequences before approving grants or fundings. In the industry, the primary source of funding comes from investors and venture capitalists, while in academia, research grants usually come from government agencies (Metha et al., 2021). This research focuses on the academic side and studies the sociotechnical problem of how the environmental impacts of AI projects are considered or overlooked in the grant approval processes of government agencies funding scientific research. Through literature review, this study aims to highlight the potential impact that the government could have on the environment through research funding in the midst of the accelerating AI revolution.

2. Background

Modern machine learning primarily focuses on deep learning, enabled by neural networks, large-scale annotated datasets, and advancements in computing hardware (LeCun et al., 2015). In 2012, Prof. Geoffrey Hinton’s team trained the first image classification model using deep neural networks with GPU (Krizhevsky et al., 2012). The success of this study ignited the deep learning revolution, inspiring the predominant use of deep neural networks to train models with large amount of input data. The popularization of deep learning inspired developments in computer architecture as scientists looked for ways to speed up the training process of these neural networks. They soon discovered that GPUs are well-suited to the highly parallelizable nature of neural network training, which relies on large-scale linear algebra operations (LeCun et al., 2015). As networks and datasets continue to grow in size, the demand for more powerful GPU clusters has increased exponentially. Many research laboratories and technology companies rushed to invest in specialized data centers, high-performance computing (HPC) infrastructure, and GPUs (Reuther et al., 2021).

Two of the most predominant areas of deep learning research that requires the most computation power are natural language processing and computer vision. In natural language processing, the introduction of the transformer architectures (Vaswani, A. et al., 2017) has led to unprecedented progress in language reasoning and generation. This enabled the development of large language models such as GPT-3 (Brown, T. et al., 2020), which demonstrated the ability to perform complex reasoning tasks using text. These large language models (LLMs) often contain millions to billions of parameters, which require enormous computational resources like GPUs for both training and inference (Raffel, C. et al., 2020). Similarly, the computer vision research community has also embraced larger, more complex architectures to approach tasks like generative AI for vision (Wang, W. et al., 2024), object detection (Zou et al., 2023), 3D reconstruction (Wang, S. et al., 2024), etc. Inspired by the success of transformers (Vaswani, A. et al., 2017) in natural language processing, the Vision Transformer (ViT) was invented (Han et al., 2022) to apply transformer-based mechanisms to images. Unlike language models, vision models require more computational power to train due to an image being a larger data modality.

These breakthroughs are reinforced by the scaling law (Kaplan et al., 2020), which states that large language model performance will improve predictably with increased model size, larger datasets, and higher compute budgets. This has incentivized researchers to train larger models, further increasing GPU usage. This trend is evident in the training process behind the state-of-the-art LLMs proposed by companies like OpenAI. Given that performance increases are slower than expected, the continued pursuit of higher intelligence is likely to further drive GPU demands (Kaplan et al., 2020).

Growing awareness of this technological trajectory and potential environmental impact has given rise to the Green AI movement. Green AI, which focuses on data efficiency, carbon footprint, and the ecological footprint of AI, has gained popularity over the past few years as awareness has grown regarding the energy challenges we must address before the next stage of mass AI innovations can proceed (Verdecchia et al., 2023; Budennyy et al., 2022). A literature review revealed that most efforts toward Green AI focus on developing more efficient models and training techniques, while others work on quantifying the energy consumption of training machine learning models (García-Martín et al., 2019). In areas where energy consumption is exploding, such as large language models, scientists have warned and provided recommendations for reducing energy expenditure (McDonald et al., 2022).

In the United States, most academic research in machine learning is funded by agencies like the National Science Foundation (NSF) and the Department of Energy (DOE). While numerous other agencies are increasing their AI research spending, these two are traditionally the main funders of research at universities in the U.S. On the NSF funding website, sustainability is one of the funding areas. However, the broad topic of sustainability, as explained on their website, focuses on sustainable manufacturing and green buildings. Artificial Intelligence is not mentioned under this topic, nor is Green AI (NSF, n.d.). Similarly, the DOE has funded the AI for Energy initiative, supporting research that utilizes AI to fuel the growth of clean energy in all key sectors of renewable energy (Daniel et al., 2024). Although it has funding available for research in this area, they do not emphasize the energy consumption of AI research itself. Not only is the government yet to derive a substantial fund for Green AI research, but it has also failed to allocate sufficient funding for energy research in general compared to other categories, even though energy is one of the most important contributors to the global GDP (Murray, 2017).

Taken together, these strands reveal a tension at the heart of contemporary machine-learning research: every architectural breakthrough and dataset expansion is powered by ever-larger GPU clusters, yet the institutional mechanisms that finance such progress have not evolved at the same pace to safeguard sustainability. The scaling law promises continued accuracy gains, but its unchecked application risks locking the field into an energy-intensive trajectory just as governments struggle to prioritise Green AI. Consequently, the environmental cost of deep learning is no longer a peripheral issue—it is a structural one, rooted in how models are designed, benchmarked, and, crucially, funded. Recognising this systemic interplay sets the stage for the next sections of this paper, which examine how policy levers and ethical frameworks can redirect research incentives toward computational efficiency without stalling scientific momentum.

3. Methodology

To evaluate government impact on recent machine learning developments, this study begins with a systematic literature review of quantitative reports of machine learning energy consumption, as well as government policy and funding initiatives. Quantitative data include estimates of energy usage in various machine learning tasks, particularly the language and vision tasks that involve large models (Strubell et al., 2019). Furthermore, funding reports from federal agencies are analyzed to examine patterns in how governmental resources have been allocated to ML research (Artificial Intelligence, 2025). Primary sources, including policy documents and official legislative records, will be employed to investigate the intentions behind funding decisions.

The potential impact of such government actions will be discussed through the lens of utilitarian ethics, environmental ethics, and the Social Construction of Technology (SCOT) framework. Utilitarian ethics evaluates the morality of an action by its outcome (Quinton, 1973). Environmental ethics underscores humanity's moral responsibility to preserve the natural environment (Rolston, 1988). This study examines the tension between these two ethical frameworks in the context of funding allocation for machine learning research.

Utilitarian ethics is centered around the belief that the moral value of an action is determined by its consequences, specifically its impact on overall happiness or well-being (Mill, 2016). Under this scope, the decision-making process involves careful consideration of how benefits and harms are distributed across the society (Quinton, 1973). In the context of this study, a utilitarian assessment would consider whether investments in high-energy-consuming ML research would lead to breakthroughs in economic growth, healthcare, work efficiency, etc., which ultimately lead to increased social wellbeing. The outcomes will be compared to the outcomes of other possible uses of the same energy resources.

Unlike utilitarian ethics, environmental ethics emphasizes human's moral responsibility to safeguard the nature. Rolston argued that the ecosystems have intrinsic value and urged humans not to make decisions solely based on their values to human well-being (Rolston, 1988). In the context of this study, environmental ethics investigates whether the carbon footprint of large-scale machine learning training are justifiable against the potential harm to the environment. This is the hidden motive behind the development of Green AI.

The SCOT framework argues that technological artifacts are shaped by social and political factors. Human attribute meaning to technologies and influence their uses (Pinch, 2012). From a SCOT perspective, government grants and policies play a crucial role in guiding the development of large machine learning models. By prioritizing funding in powerful general intelligence models that incur high computational demands, policymakers can steer the trajectory of AI development toward models with greater number of parameters. While most researchers examine the exact energy consumption of machine learning training for different tasks, few investigate the topic of Green AI from a higher level: the role of government funding. Academic research cannot proceed without research grants. Therefore, understanding how government agencies consider environmental impacts in their funding decisions is crucial. If the government has not placed significant emphasis on this topic through the grant approval process, few research groups will work towards Green AI.

4. Literature Review

The environmental footprint of contemporary AI research is a rapidly escalating reality. Empirical measurements reported by Strubell et al. show that training a single BERT-Base model on 64 NVIDIA V100 GPUs for 79 hours consumes about 1507 kWh of electricity and releases roughly 719 lb (326 kg) of CO_2, which is comparable to driving an average gasoline car from New York to Chicago (Strubell et al., 2020). Hyper-parameter tuning and iterative experimentation amplify that baseline dramatically. The Linguistically-Informed Self-Attention project required 9998 days of GPU time, equivalent to 60 GPUs running continuously for six months, costing nearly 10,000 kWh and about 9 tons of CO_2 before the model was deployed (Strubell et al., 2020). Strubell's group ends the paper by highlighting the ethical imperative for funding agencies to balance machine learning model performance gains against planetary limits.

Studies have shown that government funded patents have profound impact on AI innovation in the US, with funding on projects existing primarily in the earlier stages of AI innovation to leave room for privately funded research when the field is more mature (Iori et al., 2022). This suggests that the government plays a significant role in initiating and directing the development of AI with its "start-up funds". Scholars have also examined the effect of government funding on scientific research output. According to Goldfarb, government grants are often mission oriented and can lead to scientific research adopting commercial goals (Goldfarb, 2008). According to research on NASA grants and the projects funded by them, research labs do not always use the funding for research that aligns with the goals of the grant (Goldfarb, 2008). Although published in 2007 when AI was still in the early stages, this paper addresses the bias inherent in government funded scientific research and the extent to which researchers' publications are influenced by their sponsors. Besides, surveys into whether researchers are aware of the unintended consequences of their research have also been assessed. Do et al. demonstrate that researchers do not formally consider the unintended societal consequences of their research. This lack of awareness is caused by both an academic practice promoting fast progress and a lack of guidelines for considering those consequences (Do et al., 2023). Academic researchers face substantial pressure from grant conditions, publication quotas, and competitive innovation dynamics that often prioritize rapid development and immediate results over environmental considerations (Do et al., 2023). Although most funding agencies request authors to submit broader impact statements, these statements are mostly used for advertisement rather than reflection on negative consequences (Do et al., 2023). Government grants directly frame the direction of academic research. Its generous support in the early growth of AI comes at the cost of freedom of research. This facilitates the emergence of academic research topics and proposals carefully designed to secure government funding. This situation is further exacerbated by a lack of formal requirement to consider negative AI societal consequences and an academic practice that cherishes fast innovation.

5. Results and Discussion

The current barrier to controlling the environmental impact of academic AI research is rooted in the intense global competition between nations to achieve artificial general intelligence (AGI) and pursue advanced military capabilities. This incentivizes heavy government investments in GPU-intensive AI research, often disregarding environmental considerations (Haner et al., 2019). Moreover, the relentless pursuit of more sophisticated models is deeply embedded in contemporary machine learning research culture, driven by the scaling law. This has created an environment where researchers and companies perceive extensive computational investments as necessary for groundbreaking innovations, thereby escalating energy demands (Kaplan et al., 2020). Labs intend to fulfill this government intention by orientating their research toward these goals to seek funding.

Recent shifts in government policies, notably the executive orders signed by President Trump to reduce federal funding for scientific research, have significantly altered the trajectory of AI development (Witze et al., 2025). Research grants are rescinded, and scientists are laid off in response to the immediate orders which greatly shakes the research landscape across US colleges. These rapid contractions not only stalled ongoing deep learning projects but also discouraged prospective graduate students to pursue research. This underscores the vulnerability of AI research to political fluctuations, where funding strategies are deeply influenced by differing party agendas and priorities. It also highlights the importance of SCOT (Pinch, 2012), where technological breakthroughs are often guided by human intentions. In this case, executive goals facilitate academic research in a mutual beneficial relationship.

The existing approach in government funding frameworks also reveals a lack of considerations for energy efficiency and environmental impacts. Research grants typically prioritize innovations capable of immediate societal or economic returns, with less emphasis on potential environmental consequences. This trend suggests that current funding mechanisms do not adequately incentivize the implementation of environmentally sustainable AI practices (Goldfarb, 2008). Until federal agencies integrate sustainability metrics into peer review and grant approval processes, the research community will continue to sacrifice environmental cost for performance.

On the technical side, researchers have increasingly focused on computational optimization strategies designed to enhance energy efficiency. Recent projects like DeepSeek's LLM (DeepSeek-AI et al., 2025) raised the awareness of research aimed at reducing GPU dependency by employing more efficient training techniques, thereby decreasing energy consumption during training without compromising model accuracy or performance. Similarly, emerging concepts of training a powerful large AI model to instruct smaller specialized AI agents offer a more environmentally friendly solution with lowered computational demands (Hinton et al., 2015).

Integrating ethical frameworks into AI research and funding policies is a crucial step towards sustainable AI. Employing utilitarian ethics, policymakers should systematically evaluate the societal impacts of funded AI projects, balancing immediate benefits against long-term environmental harms, and design grants accordingly to reflect these goals. Conducting rigorous cost-benefit analyses of energy-intensive AI initiatives could reveal scenarios where societal and political benefits, such as improved healthcare or national defense, might justify certain levels of environmental trade-offs (Mill, 2016). In contrast, adopting an environmental ethics approach necessitates that policymakers and researchers acknowledge the inherent value of our nature that outweighs the benefits that we seek from large AI models. Funding agencies could prioritize AI projects aimed at minimizing environmental footprint (Green AI). Under strict governmental funding policies, the academic AI research community will almost certainly shift to emphasize environmental accountability as an essential component of innovation (Rolston, 1988). Funding strategies that combine utilitarian and environmental considerations could guide research towards technological breakthroughs that not only preserve societal interests but also maintain environmental resilience in the long term.

Governments and funding agencies must proactively integrate sustainability criteria into their grant evaluation processes. Funding decisions should explicitly prioritize projects that incorporate energy-efficient AI methods and promote innovations that challenge the existing reliance on the scaling law (Kaplan et al., 2020). Introducing such criteria will incentivize researchers to develop sustainable AI practices systematically (Iori et al., 2022).

6. Conclusion

Government funding profoundly influence which research directions flourish. The funding decisions can either perpetuate energy-intensive AI developments or catalyze a transformation towards energy efficient AI or Green AI. In an age where countries compete for intelligent agents and researchers compete for funding, the responsibility of the government in shaping potential negative effects of new AI innovations is increasingly vital.

For stakeholders inside the government and the academia, the next steps involve more than just model efficiency breakthroughs. Government agencies should specify environmental considerations in their grant calls and research frameworks. Such requirements would encourage universities and labs to pursue alternative approaches, from efficient model architectures to novel high performance computing techniques. The collaborative efforts between AI researchers, policymakers, and environmental experts can ensure that state-of-the-art research incorporate sustainability as a design principle. By relying on the ethical frameworks like Utilitarianism, Environmental ethics, and Social Construction of Technology, government agencies can setup more adequate AI innovation goals. Finally, schools should incorporate ethics courses to make engineers and researchers reflect on the environmental challenges that AI advancements may pose.

Modern AI research requires collaboration and accountability. Government agencies and universities each have a role in forging a sustainable technological development. The future lies in harnessing AI’s transformative potential while withholding our collective responsibility to safeguard our planet.

References

Artificial Intelligence (2025) *NSF*. Available at: https://www.nsf.gov/focus-areas/artificial-intelligence.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Budennyy, S. A., Lazarev, V. D., Zakharenko, N. N., Korovin, A. N., Plosskaya, O. A., Dimitrov, D. V., … Zhukov, L. E. (2022). eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. *Doklady Mathematics*, *106*(1), S118–S128. doi:10.1134/S1064562422060230

Crawford, K. (2021) *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Choudhary, A., Fox, G., & Hey, T. (2023). *Artificial Intelligence for Science*. doi:10.1142/13123

Crawford, N. (1966). *Digital simulation in hydrology : Stanford watershed model IV / by N.H. Crawford and Ray K. Linsley.*. Stanford University. Dept of Civil Engineering.

Daniel, C., Gehin, J. C., Laurin-Kovitz, K., Morreale, B., Stevens, R., & Tumas, W. (n.d.). AI for Energy Report 2024 | Argonne National Laboratory. https://www.anl.gov/ai/reference/ai-for-energy-report-2024

DeepSeek-AI *et al.* (2025) 'DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning', *arXiv [cs.CL]* [Preprint]. Available at: http://arxiv.org/abs/2501.12948.

Do, K., Pang, R. Y., Jiang, J., & Reinecke, K. (2023). "That's important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Presented at the Hamburg, Germany. doi:10.1145/3544548.358134

Dingman, S. (2015). *Physical hydrology*. Waveland press.

Environmental sustainability. NSF. (n.d.). https://new.nsf.gov/funding/opportunities/environmental-sustainability

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, *134*, 75–88. doi:10.1016/j.jpdc.2019.07.007

Goldfarb, B. (2008). The effect of government contracting on academic research: Does the source of funding affect scientific output? *Research Policy*, *37*(1), 41–58. doi:10.1016/j.respol.2007.07.011

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., ... & Tao, D. (2022). A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, *45*(1), 87-110.

Haner, J. and Garcia, D. (2019) 'The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development', *Global Policy*, 10(3), pp. 331–337. Available at: https://doi.org/10.1111/1758-5899.12713.

Hinton, G., Vinyals, O. and Dean, J. (2015) 'Distilling the knowledge in a neural network', *arXiv preprint arXiv:1503. 02531* [Preprint].

Iori, M., Martinelli, A., & Mina, A. (2022, October 3). *The direction of technical change in AI and the trajectory effects of government funding*. Handle Proxy. https://hdl.handle.net/11382/549252

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences, 22(11), 6005–6022.*

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Mehta, D., Desai, D., & Pradeep, J. (2021). Machine learning fund categorizations. *Proceedings of the First ACM International Conference on AI in Finance*. Presented at the New York, New York. doi:10.1145/3383455.3422555

Mill, J. S. (2016). Utilitarianism. In *Seven masterpieces of philosophy* (pp. 329-375). Routledge.

Mulvaney T J. (1851). On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. *Proceedings of the Institution of Civil Engineers of Ireland, 4, 19-31.*

Murray, C. (2017). Public Funding of Energy Research. *Joule*, *1*(2), 204–208. doi:10.1016/j.joule.2017.09.002

Pinch, T. (2012). The social construction of technology: A review. *Technological change*, 17-35.

Quinton, A. (1973). *Utilitarian ethics*. Springer.

Reuther, A., Michaleas, P., Jones, M., Gadepally, V., Samsi, S., & Kepner, J. (2021, September). AI accelerator survey and trends. In *2021 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1-9). IEEE.

Rolston, H. (1988). *Environmental ethics*. Temple University Press.

Sitterson, J., Chris Knightes, R., K. Wolfe, M. Muche, & B. Avant (2017). *An Overview of Rainfall-Runoff Model Types*. U.S. Environmental Protection Agency, Washington, DC.

Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 09, pp. 13693-13696).

Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *WIREs Data Mining and Knowledge Discovery*, *13*(4), e1507. doi:10.1002/widm.1507

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., … Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. doi:10.1038/s41586-023-06221-2

Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., & Revaud, J. (2024). Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20697-20709).

Wang, W., Chen, Z., Chen, X., Wu, J., Zhu, X., Zeng, G., ... & Dai, J. (2024). Visionllm: Large language model is also an open-ended decoder for vision-centric tasks. *Advances in Neural Information Processing Systems*, *36*.

Witze, J.T. and Kozlov, M. (2025) 'SCIENCE UNDER TRUMP 2.0: THE LIKELY WINNERS AND LOSERS', *Nature*, 637, p. 533.

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, *111*(3), 257-276.

Science Time Series: Deep Learning in Hydrology

Environmental Considerations in AI Project Funding: A Government Grant Evaluation

A Thesis Prospectus

In STS 4500

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

By

Junyang He

November 8, 2024

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Prof. Pedro Augusto P. Francisco, Department of Engineering and Society

Geoffrey Fox, Department of Computer Science

Introduction

The advent of big data and powerful machine learning models has facilitated a data-driven approach to solving complex scientific challenges such as disaster prevention, climate modeling, and weather forecasting (Choudhary et al., 2022; Wang et al., 2023). The success of large language models like GPT has spurred widespread interest in developing large-scale models for global-scale modeling and prediction tasks across various domains. From tech companies to the supercomputing communities in academia, an arms race for GPUs and computing power has begun, often rarely accounting for the enormous energy usage and environmental footprint of the systems that power machine learning innovations.

Regardless, no research or product development can be conducted without adequate funding. In the industry, the primary source of funding comes from investors and venture capitalists, while in academia, research grants usually come from government agencies (Metha et al., 2021). In this research, I aim to focus on the academic side and study the sociotechnical problem of how the environmental impacts of AI projects are considered or overlooked in the grant approval processes of government agencies funding scientific research. Specifically, I aim to examine the intentions and priorities of stakeholders behind university research, and whether they have thoroughly considered the potential environmental harm that their actions may bring.

My capstone research involves developing a deep learning model to revolutionize the rainfall-runoff modeling problem in hydrological science and studying various methods for increasing prediction accuracy and computational efficiency. While the project contributes to the greater challenge of flood prediction and water resource management, it also provides a basis for quantifying the energy usage of machine learning research, as well as exploring how government funding directly impacts environmental research projects like mine. By connecting the environmental considerations in AI project funding with my technical work, this study emphasizes the importance of integrating sustainability into AI research at a high level, ensuring that technological advancements are pursued responsibly and ethically.

Technical Topic: Deep Learning in Hydrology

My capstone research is part of a broader investigation into scientific time series data and an effort to study data-driven approaches to natural science challenges. The project specifically examines a deep learning approach to the rainfall-runoff problem in hydrology. Time series data is data that is time dependent and varies based on time. Examples of such data include stock market data, product sales data, heartrate data, etc. Unlike other fields, time series data in natural science involve environmental features and data that are collected from multiple locations. In hydrology, time series data include precipitation, temperature, streamflow, etc. Daily values are collected from thousands of gauging stations around the world, with time series data being heavily influenced by geographical factors of these locations. For instance, the relationship between precipitation and runoff (amount of water that flows over land surface as result of precipitation) is highly dependent on vegetation, soil or rock type, and other environmental factors. Therefore, simply inputting time series data into a machine learning model to make predictions might lead to poor outcomes in a field of natural science like hydrology. We claimed that it is crucial to train models with environmental factors to help them grasp the hidden relationships within time series data.

Rainfall-runoff modeling, a key challenge in hydrology, aims to model the physical process by which water on land surface (precipitation or snowmelt) moves to streams (Dingman, 2015). As part of the hydrologic cycle, precipitation on land either evaporates, transpires, infiltrates to recharge groundwater, or becomes surface runoff entering a catchment. A catchment, is an area where all precipitation collects and drains into a common outlet, such as a river, lake, or reservoir (Dingman, 2015). Surface runoff contributes to streamflow, which is defined as the volumetric discharge that takes place in a stream or channel. Eventually, this streamflow and ground water outflow reaches the ocean, where it evaporates, condenses into clouds, and returns as rainfall on land, completing the hydrologic cycle (Dingman, 2015).

Traditional rainfall-runoff modeling has typically focused on individual catchments. The first documented model, introduced in 1851, used linear regression to predict discharge from precipitation intensity and runoff (Mulvaney, 1851). Since then, scientific advancements have led to more sophisticated models based on mathematical formulas and physical laws. The advent of computers brought digital hydrological models (Crawford, 1966). Taking advantage of the number of parameters offered, these physical-based digital models perform exceptionally well. However, the high computational cost to calibrate these parameters, and the limited availability of data hinder their use in large-scale forecasting applications (Sitterson et al, 2017). Groundbreaking advancements in deep learning models and the publication of structured large-sample Hydrology dataset have overcome this limitation, enabling the study of nation or global scale rainfall-runoff modeling. In 2018, the focus of the field shifted towards Long Short-Term Memory (LSTM) based models, which excelled in learning sequential dependencies within time series data (Kratzert et al, 2018). These models have shown great success in large-scale hydrological time series predictions.

In this work, we analyzed hydrology time series using the CAMELS and Caravan global datasets. These datasets include up to 6 time series variables and 209 environmental features collected from around 8,000 locations worldwide. We constructed a model based on the LSTM architecture and tested eight different training configurations. We found that including environmental data in training significantly boosts model accuracy, reducing the error by 40%

when tested on the largest dataset. Additionally, including encoding techniques that captures the relationship between catchments and some yearly periodic We also present initial results from studies on other deep learning neural network architectures. We show that methods trained using environmental features of the locations perform better than less flexible methods, including Foundation models. Our analysis is intended to serve as an educational source and benchmark for future studies.

STS Topic: Environmental Considerations in AI Project Funding

My research aims to understand how funding behind machine learning research in academia relates to the environmental harm caused by AI. This question is important because the significant energy consumption of AI research contributes to environmental degradation, and addressing this issue at the funding level could lead to more sustainable AI practices. Modern machine learning research is usually big data-driven. In the realm of large language models, billions of parameters and vast amounts of text inputs require enormous amounts of energy to train. In scientific domains, global-scale predictions necessitate training on years of data. While simple machine learning models can be run directly on laptops without significant GPU usage, most studies require the use of supercomputers that have access to hundreds of computing units and GPUs, resulting in unprecedented energy usage and environmental impact.

Green AI, which focuses on data efficiency, carbon footprint, and the ecological footprint of AI, has gained popularity over the past few years as awareness has grown regarding the energy challenges we must address before the next stage of mass AI innovations can proceed (Verdecchia et al., 2023; Budennyy et al., 2022). A literature review revealed that most efforts toward Green

AI focus on developing more efficient models and training techniques, while others work on quantifying the energy consumption of training machine learning models (García-Martín et al., 2019). In areas where energy consumption is exploding, such as large language models, scientists have warned and provided recommendations for reducing energy expenditure (McDonald et al., 2022).

While most researchers examine the exact energy consumption of machine learning training for different tasks, few investigate this topic from a higher level: the role of government funding. Academic research cannot proceed without research grants. Therefore, understanding how government agencies consider environmental impacts in their funding decisions is crucial. If the government has not placed significant emphasis on this topic through the grant approval process, few research groups will work towards Green AI.

Most academic research in machine learning is funded by agencies like the National Science Foundation (NSF) and the Department of Energy (DOE). While numerous other agencies are increasing their AI research spending, these two are traditionally the main funders of research at universities in the U.S. On the NSF funding website, sustainability is one of the funding areas. However, the broad topic of sustainability, as explained on their website, focuses on sustainable manufacturing and green buildings. Artificial Intelligence is not mentioned under this topic, nor is Green AI (NSF, n.d.). Similarly, the DOE has funded the AI for Energy initiative, supporting research that utilizes AI to fuel the growth of clean energy in all key sectors of renewable energy (Daniel et al., 2024). Although it has funding available for research in this area, they do not emphasize the energy consumption of AI research itself. Not only is the government yet to derive a substantial fund for Green AI research, but it has also failed to allocate sufficient funding for energy research in general compared to other categories, even though energy is one of the most important contributors to the global GDP (Murray, 2017).

Even if research grants are targeted for Green AI, they might not correspond to direct efforts toward reducing the environmental harm AI brings. According to research on NASA grants and the projects funded by them, research labs do not always use the funding for research that aligns with the goals of the grant (Goldfarb, 2008). Nevertheless, government funding has a profound impact on fueling AI innovation in general, especially in the early stages of development (Iori, 2022). Therefore, these government grants are influential in powering early stages of Green AI research.

This research explores the problem through the lens of utilitarian ethics and environmental ethics. Utilitarian ethics evaluates the morality of an action by its outcome (Quinton, 1973). Environmental ethics underscores humanity's moral responsibility to preserve the natural environment (Rolston, 1988). This study examines the tension between these two ethical frameworks in the context of funding allocation for machine learning research. Evidence will be presented through literature review on journal articles, funding reports, and government policies. Furthermore, a case study on the computational demands of large-scale machine learning models will provide a detailed analysis, quantifying the environmental impacts caused by AI development.

Conclusion

My capstone research on deep learning in hydrology aim to demonstrate how machine learning can be effectively applied to solve scientific problems involving time series data. It serves as a benchmark and an example pipeline for what others should focus on when applying machine learning to solve scientific challenges. my STS research will present the potential environmental harm caused by future large-sample and large-model AI studies. By evaluating the importance of government funding in efforts toward Green AI, this research will analyze how grants can effectively contribute to promoting environmentally sustainable AI practices. The expected outcome is to urge policymakers and funding agencies to allocate more resources to Green AI initiatives, encouraging research groups to prioritize energy efficiency and reduce the ecological footprint of AI developments.

References

Budennyy, S. A., Lazarev, V. D., Zakharenko, N. N., Korovin, A. N., Plosskaya, O. A., Dimitrov, D. V., … Zhukov, L. E. (2022). eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI. *Doklady Mathematics*, *106*(1), S118–S128. doi:10.1134/S1064562422060230

Choudhary, A., Fox, G., & Hey, T. (2023). *Artificial Intelligence for Science*. doi:10.1142/13123

Crawford, N. (1966). *Digital simulation in hydrology : Stanford watershed model IV / by N.H. Crawford and Ray K. Linsley.*. Stanford University. Dept of Civil Engineering.

Daniel, C., Gehin, J. C., Laurin-Kovitz, K., Morreale, B., Stevens, R., & Tumas, W. (n.d.). AI for Energy Report 2024 | Argonne National Laboratory. https://www.anl.gov/ai/reference/ai-for-energy-report-2024

Dingman, S. (2015). *Physical hydrology*. Waveland press.

Environmental sustainability. NSF. (n.d.). https://new.nsf.gov/funding/opportunities/environmental-sustainability

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, *134*, 75–88. doi:10.1016/j.jpdc.2019.07.007

Iori, M., Martinelli, A., & Mina, A. (2022, October 3). *The direction of technical change in AI and the trajectory effects of government funding*. Handle Proxy. https://hdl.handle.net/11382/549252

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks*. Hydrology and Earth System Sciences, 22(11), 6005–6022.*

Mehta, D., Desai, D., & Pradeep, J. (2021). Machine learning fund categorizations. *Proceedings of the First ACM International Conference on AI in Finance*. Presented at the New York, New York. doi:10.1145/3383455.3422555

Mulvaney T J. (1851). On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. *Proceedings of the Institution of Civil Engineers of Ireland, 4, 19-31.*

Murray, C. (2017). Public Funding of Energy Research. *Joule*, *1*(2), 204–208. doi:10.1016/j.joule.2017.09.002

Quinton, A. (1973). *Utilitarian ethics*. Springer.

Rolston, H. (1988). *Environmental ethics*. Temple University Press.

Sitterson, J., Chris Knightes, R., K. Wolfe, M. Muche, & B. Avant (2017). *An Overview of Rainfall-Runoff Model Types*. U.S. Environmental Protection Agency, Washington, DC.

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *WIREs Data Mining and Knowledge Discovery*, *13*(4), e1507. doi:10.1002/widm.1507

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., … Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. doi:10.1038/s41586-023-06221-2