Energy-Efficient Deep Learning for Finance

by

Tom Maxwell Potter

July 16, 2022

A dissertation submitted in part fulfilment
of the requirements for the degree of

Master of Science

of

University College London

Scientific and Data Intensive Computing
Department of Physics and Astronomy

Declaration

I, Tom Maxwell Potter, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the dissertation.

Abstract

This thesis investigates the use of energy-efficient methods for deep learning-based financial volatility forecasting, aiming to reduce the energetic cost of such models and demonstrate how the sustainability of deep learning for finance can be improved.

Context/background: The financial sector has long been associated with largely negative environmental, social, and governance (ESG) impacts, including being a major contributor to global carbon emissions. Despite the attempts by some to prioritise *sustainable finance*, the recent expansion of financial technology—incorporating new, expensive methods such as *deep learning* (DL)—has only worsened the energy consumption attributed to this industry, accelerating its carbon emissions.

Aims: In an attempt to address these negative impacts of financial technology, this project aims to develop an energy-efficient DL system for financial modelling. This research will explore *Green AI* methods that attempt to reduce the energy expended training DL models and apply these for the first time to models used in finance. To exemplify the benefits of these methods, a performant financial volatility model will be developed that not only produces accurate results but prioritises generating this performance in an efficient manner, minimising the energy and data resources required during training. This system aims to demonstrate that the principles of Green AI are applicable within the financial sector, furthering the scope of sustainable finance by improving the sustainability of deep learning for finance and, hence, minimising the ESG impacts of the financial sector.

Method: This research will commence with an analysis of the resource requirements of typical systems in the field of deep learning for finance. A particular focus will be given to the domain of financial volatility modelling, as this is a major application of deep learning in finance, and the *long short-term memory* (LSTM) networks typically exploited for such tasks. Energy and data-efficient training methods will be explored, developing a deep model that consumes less energy and requires less data to train, but maintains accurate performance. Specifically, methods such as *active learning*, *progressive training*, and *mixed-precision* will be explored that reduce the resource requirements of training, proving the feasibility of efficient models within this field.

Contributions to science:

1. *Expanding the applications of Green AI.* The first application of Green AI to the finance sector, further demonstrating the utility and importance of Green AI in lowering the environmental cost of deep learning.

2. *Reducing the environmental impact of financial technology.* The improvement of sustainable finance to include the new research domain of *sustainable deep learning for sustainable finance*.

3. *Improving the inclusivity of finance.* Lowering the bar-to-entry to engage in deep learning for finance, allowing more individuals to leverage financial technology and analytics.

Outline of research:

- Chapter 3.1: *Baseline finanical volatility model.* An initial deep model will be implemented, using a traditional training process and LSTM architecture, to act as an exemplar of the resource requirements of this domain.

- Chapter 3.2: *Energy-efficient training extensions.* Several adaptations to the model training process will be made that prioritise reducing the energy consumed by the system.

- Chapter 3.3: *Data-efficient training extensions.* Additional adaptations will be made that reduce the necessary amount of training data, further lowering resource requirements.

- Chapter 4: *Discussion & evaluation.* An analysis will be made between the baseline and extended models, comparing the performance and efficiency of each, and discussing their success in reducing resource requirements.

Keywords: Green AI, Green Deep Learning, Energy Efficiency, Data Efficiency, Sustainable Finance, Financial Technology, Financial Volatility Modelling, Long Short-Term Memory

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Topic & Background

Many industries have recently been under increased pressure to monitor and rectify their environmental impact. This pressure is typically directed towards the perceived high carbon industries that constitute the major pollutant sectors of the economy, such as transport, energy supply, and agriculture. For example, recent estimates suggest that of the 33.5 billion tons of carbon dioxide emissions generated globally in 2018, 8 billion tons could be attributed to the transport sector (IEA, 2022), and 6 billion tons to farming and livestock (Ahmad et al., 2022). These concerning figures have rightly sparked increased international discussion surrounding global carbon emissions and sustainability, such as the 2021 *United Nations Climate Change Conference* (COP26).

1.1.1 Sustainable Finance

The finance sector has long been closely associated with sustainability concerns such as those discussed at COP26, being a major contributor to global carbon emissions both directly and indirectly. The most visible environmental impact of the financial industry is its direct emissions from business practices such as the distribution of cash through the economy (e.g. cash transport and ATM power consumption), card payment processing centres, and everyday operational costs such as heating office buildings (Hanegraaf et al., 2018). However, indirect emissions—attributable to services such as investing and lending—have been estimated to contribute over 700 times more to the carbon footprint of the financial industry than all direct emissions (Power et al., 2020). This form of carbon emissions, entitled *financed emissions* by Power et al. (2020), includes practices such as financing fossil fuel companies—who have received \$3.8 trillion in funding from global banks since the *Paris Agreement* was signed in 2016 (Kirsch et al., 2022). In their survey of 700 global financial institutions, Power et al. (2020) estimated that the production of over 1.04 billion tons of carbon dioxide was attributable to financed emissions in 2020 (approximately 3.1% of global emissions). However, they note this figure is likely to significantly understate the total global financed emissions, as of the 700 contacted institutions only 332 responded, and only 25% of those reported financed emissions (typically on less than 50% of their portfolios). Furthermore, a recent report by *Greenpeace* and the *WWF* concluded that

the combined carbon emissions of the largest banks and investors in the UK in 2019 totalled 805 million tons, which (if consolidated into its own country) would rank 9th in the global list of total emissions per country (Greenpeace, 2021). This figure is 1.8 times higher than the total emissions of the UK (455 million tons), and almost 90% of the global emissions from commercial aviation (918 million tons) in the same year (Graver et al., 2020).

The increasing awareness of the negative *environmental, social, and governance* (ESG) impacts of the finance industry, highlighted by studies such as those of Power et al. (2020) and Dennis et al. (2021), has led researchers to investigate methods that prioritise the *sustainable development goals* (SDGs) within the financial sector. Towards this objective, the field of *sustainable finance* has emerged, which aims to consider ESG impacts and SDGs in financial decisions (such as investment and lending activities) to improve the sustainability of finance. Namely, despite the lack of a rigorous consensus on what constitutes sustainable finance, recent reviews—such as those of Cunha et al. (2021) and Kumar et al. (2022)— typically use the term to refer to research into financial activities, resources, and investments that prioritise long-term sustainability. In particular, a focus is given to those practices that produce a measurable positive improvement to the social and environmental impact of the financial industry, global economy, and wider society.

This discussion around sustainable finance largely began with Ferris and Rykaczewski's examination of the benefits of investing pension funds in a socially responsible way. Following this, early research mainly focussed on *socially responsible investing*, where investments are made that not only prioritise profits but further current positive social movements and mitigate societal concerns (Cunha et al., 2021). During the 2000s, research began to exhibit a new focus on environmental sustainability, considering factors such as climate change and renewable energy (Van Der Laan and Lansbury, 2004). Later research further pushed the scope and impact of environmentally-focused practices with the development of new domains such as *climate finance* (Hogarth, 2012), where the mitigation of climate change is prioritised through investment and financing, and *carbon finance* (Aglietta et al., 2015), which focusses on investments that seek to lower or offset carbon emissions. Recent research within sustainable finance has pushed this environmental focus further, aiming to put into practice the sustainability goals set out by the Paris Agreement, ESG factors, and SDGs. Specifically, a new interest has been taken in sustainable investment fields such as *impact investing* (Agrawal and Hockerts, 2021) and *ESG investing* (Alessandrini and Jondeau, 2020), where investments are made that produce measurable improvements to environmental issues (according to criteria such as their ESG impacts). This recent focus has significantly increased the prominence and influence of sustainable finance. In their review of 936 research papers, Kumar et al. (2022) found that almost 70% of sustainable finance research had been published between 2015–2020, and an exponential trend was exhibited in the increase in papers being published each year; additionally, they found that the top three most cited papers all conducted research in the field of impact investing. Furthermore, Kumar et al. assert that in 2020, $400 billion of new sustainability funds were raised on capital markets. Hence, it is clear the scope and impact of sustainable finance is currently on the rise, predominantly driven

by a renewed focus on the ESG impacts of financial practices, resources, and investments.

1.1.2 Financial Technology and the Issues with Sustainable Finance

Whilst the \$400 billion raised in sustainability funds seems impressive, in the same year the total US equity market value was over \$40 trillion (Ltd, 2022), meaning globally only 0.98% of the value of the US equity market alone was raised. Furthermore, recent research has uncovered the prevalence of investment *greenwashing* (Popescu et al., 2021), where institutions misleadingly classify their practices and investments as sustainable without credible data to back up their claims (and often excluding data that would suggest the opposite). In their review, Cunha et al. (2021) raised similar concerns, asserting that research into sustainable finance is currently "excessively fragmented". These issues indicate that whilst attention is growing around the sustainability of financial practices, this domain is still not widely recognised, and further work and research are still necessary to increase the adoption of sustainable methods and tools within finance.

An additional concern is that the tools used to conduct financial practices are becoming increasingly resource-hungry at a pace exceeding the adoption rate of sustainable finance. A clear exemple of this is the increased adoption of technology throughout the finance industry. Recently, a surge of developments in financial technology (*Fintech*) has revolutionalised the methods and practices used across the field of finance, from the large financial institutions and increasing number of Fintech startups, to groups of academic researchers. This Fintech revolution has transformed many aspects of finance, promising to enhance and automate existing financial services, and deliver new, innovative financial products. In their exploration of the evolution of Fintech, Palmié et al. (2020) assert that this adoption emerged in three waves. They suggest that it began with the utilisation of electronic payments and online banking, digitalising the world of finance; the second wave then came with the emergence of blockchain technology and cryptocurrencies, which further disrupted the way currency is stored and transacted. The third and most recent Fintech wave, Palmié et al. (2020) claim, is the current upwards trend in financial institutions' reliance upon *artificial intelligence* (AI). Driven by the promise of increased automation and computing power, the utilisation of AI within the financial sector has been rapidly expanding over recent years, becoming a core component of many of the financial products and services used today: from accurate real-time financial fraud detection (Sadgali et al., 2019) to automated analysis of financial statements (Amel-Zadeh et al., 2020).

1.1.3 Deep Learning for Finance

Because of their power and potential, AI methods have been used to produce state-of-the-art results over a plethora of scientific problems and research fields: from DeepMind's *AlphaFold* (Jumper et al., 2021)—the first programmatic solution to the age-old protein folding problem in Biology—to Google's *PaLM* (Chowdhery et al., 2022)—a cutting-edge human language model that delivers breakthrough results in multi-step arithmetic and common-sense reasoning (a major step to-

wards *artificial general intelligence*). This research typically revolves around the use of *machine learning* (ML), where large collections of data are used to train computational models how to perform certain tasks independently (Samuel, 1959). Recent innovations in ML have increasingly taken advantage of *deep learning* (DL) methods, which use large, complex models to produce state-of-the-art performance (Witten et al., 2017).

The recent success in utilising DL—such as that of AlphaFold and PaLM—has driven an increased adoption of AI and DL further afield, such as within the finance industry. In fact, global spending on AI is predicted to double in value by 2024, from \$50 billion in 2020 to an estimated \$110 billion (Nassr et al., 2021). Specifically, the global AI Fintech market was estimated as being worth \$7.91 billion in 2020 and is forecast to grow to \$27 billion by 2026 (Intelligence, 2021). Furthermore, in a survey of 206 executives from US financial service companies, Gokhale et al. (2019) found that 70% used ML within their financial institutions for practices such as detecting irregular patterns in transitions, and building advanced credit models. In a similar survey, Staff (2019) found ML to be a core component in current financial technology, revolutionising data processing and modelling practices.

In their survey of financial professionals, Staff (2019) discovered that 44% of respondents cited "greater accuracy of process and analysis" as a key motivation behind their adoption of AI methods. This superior accuracy provided by ML and DL methods—publicised through models like PaLM pushing the boundaries of computational accuracy—provides a compelling alternative to traditional statistical models. Hence, the promise of increased accuracy of performance is a major driving factor behind recent ML adoption within Fintech. For example, recent DL-based financial systems have been shown to predict borrower defaults with greater accuracy than achievable with traditional approaches (Albanesi and Vamossy, 2019).

1.1.4 The Issues with Deep Learning

Whilst cutting-edge DL models push the boundaries of computational accuracy, few of these systems prioritise the efficient use of energy and data. This has led to DL inflicting a great cost upon the environment, as the energy-intensive algorithms, long training phases, and power-hungry data centres they utilise inflict a high carbon footprint (Lacoste et al., 2019). Schwartz et al. (2019) label these accurate but energy-intensive DL models as *Red AI*, which they define as "research that seeks to improve accuracy through the use of massive computational power while disregarding the cost". Schwartz et al. explain how these systems generate their performance gains majoritively through the use of extensive computational resources, such as complex models with vast parameter sets, large collections of data, and power-hungry computer hardware. Bender et al. (2021) illustrate this trend through the progression of recent language models: whilst the 2019's state-of-the-art model *BERT* (Devlin et al., 2018) used 340 million parameters and a 16GB dataset, the leading models of 2020 (*GPT-3* by Brown et al. (2020)) and 2021 (*Switch-C* by Fedus et al. (2021)) utilised 175 billion and 1.57 trillion parameters respectively, and data sets of size 570GB and 745GB. In fact, between 2012 and 2018 the computational resources used to train cutting edge models increased by

a factor of $300,000$, outpacing *Moore's Law* (Amodei and Hernandez, 2021).

The intense computational load of these DL models does not come for free; the large parameter and data sets mean training, storing, and computing with these models draws a significant amount of energy—referred to by Bietti and Vatanparast (2019) as *data waste*. Partly due to this inefficiency, the data centres at which DL models rely upon for storage and cloud computing become a significant hidden contributor to carbon emissions (Al-Jarrah et al., 2015). Studies such as Masanet et al. (2020) and Malmodin and Lundén (2018), 2020 have estimated that processing at these data centres consumes around $200--250$TWh of electricity a year, with the cost of data transmission exceeding this at $260--340$TWh per year (IEA, 2022). These estimates suggest that global data centres use more electricity than the majority of countries in the world, ranking above both Australia and Spain (EIA, 2019), and account for around 1% of global electricity consumption (rising to 2.4% when including transmission costs). Furthermore, this energy is likely not entirely carbon-neutral; Cook et al. (2017) showed that of their total electricity demand, *Amazon Web Services* only powered 12% through renewable sources, and *Google Cloud* 56% (with the latter figure ranging between 4% and 94% depending on location). This means that these large energy budgets generate considerable carbon emissions, resulting in the use of data centres coming alongside significant environmental detriment. Specifically, research suggests that in 2018, cloud computing at data centres generated the equivalent of 31 million tons of carbon dioxide (Hockstad and Hanel, 2018) in the US alone, equaling the total emissions generated by electric power in the state of California (IEA, 2022). Furthermore, the digital technology sector as a whole is estimated to be responsible for 4% of global carbon emissions, with this figure forecast to double by 2025 (Bietti and Vatanparast, 2019).

Beyond these general figures, Strubell et al. (2019) showcased the carbon emissions specifically produced by training DL models. They found that training the language model BERT, which utilised 110M parameters and trained for 96 hours over 16 TPUs, produced the equivalent of 1438 lbs of CO_2—the same as a trans-American flight. Strubell et al. also found that whilst the *Evolved Transformer* of So et al. (2019) improves state-of-the-art accuracy in English-German translation by 0.1 *BLEU* (a common metric of translated text quality), if implemented on GPU hardware this model could cost \$3.2 million to train (or \$147,000 if using TPUs), and generate $626,155$lbs of CO_2 (almost five times the lifetime emissions of an average car).

It is important to note that the large financial cost associated with these intensive DL models also inflicts a great social cost. Namely, as the systems used at the forefront of DL research get larger and more complex (such as 175 billion parameters and 745GB dataset of Switch-C), the price of storing the model and its training data, as well as the cost of running its training process on the specialist hardware this would require, becomes prohibitively high (Schwartz et al., 2019). This financial barrier restricts who can engage in cutting-edge research to only those with the backing of a large institution. Thus, this lack of accessibility drives a *"rich get richer"* cycle of research funding (Strubell et al., 2019) where only research driections within the interest of these institutions receive enough funding. This not only stifles creativity, but leaves the allocation of who benefits from the development of these systems, and who bears the negative side effects, to a handful

of large corporations. In particular, Bender et al. (2021) highlight this disconnect between the benefits of energy-intensive DL research and the environmental consequences it inflicts (using the example of language models (LMs)): "is it fair or just, for example, that the residents of the Maldives [...] pay the environmental price of training and deploying ever larger English LMs".

Hence, whilst DL been shown to provide state-of-the-art computational accuracy across a range of fields, its environmental impact cannot be ignored. Therefore, the accelerating reliance on ML and DL within Fintech poses issues for its sustainability, as this the models and methods contribute further to the negative ESG impacts of the financial sector. This issue generates a clear conflict between the growing use of DL in Fintech, and the growing need for sustainable finance. Both of these fields provide great utility to the finance sector: Fintech (including DL) provides innovative services to consumers and accurate tools for institutions, and sustainable finance ensures the industry has minimal ESG problems. Moreover, DL has been shown to have utility for work in sustainability further afield, such as improving the efficiency of exploiting renewable energy sources (Daniel et al., 2021), and even within sustainable finance itself, for example using ML to analyse the ESG factors of potential investments (Mehra et al., 2022). For these reasons, a compromise between the use of DL Fintech and the prioritisation of sustainable finance must be reached.

1.1.5 Green AI

This apparent gap between the utility of DL systems—provided by their superior accuracy over traditional methods—and their environmental consequences has recently started to gather attention from ML researchers. In what has become known as *Green AI* (Schwartz et al., 2019), new research has begun to consider how to mitigate the negative ESG impacts of ML by improving the efficiency of DL models. These efficient models reduce the energy required for training and deployment, minimising the carbon emissions they contribute towards.

Alongside *mobile computing*, which prioritises energy-efficient methods due to the hardware constraints of mobile devices, the research domains of *Natural Language Processing* (NLP) and *Computer Vision* (CV) are currently the predominat fields focussing on Green AI, as both these areas exploit highly complex models, and require efficient real-time processing when deployed. NLP focuses on the processing and understanding of language (e.g. using language models to make predictions about text sequences), typically using large, complex DL models in an attempt to match and exceed human accuracy in language modelling (e.g. Chowdhery et al. (2022)). CV typically uses complex models with expensive methods—such as the *convolution operation* (Dumoulin and Visin, 2016)—to analyse, classify, and map visual environments. Within these fields, Green AI developers have been working to promote the utilisation of efficient methods that reduce the energy, time, and data requirements of model training by using parameters, data, and operations more intelligently (without producing a significant drop in accuracy). However, the major limitation of Green AI is that these methods have yet to garner significant attention outside of the specific research domains of NLP, CV, and mobile computing. Therefore the utility of these methods to improving the sustainability of ML has largely not been demonstrated further afield, meaning their

potential benefits to reduce the energy consumption and carbon emissions of ML have yet to be seen on a wide scale.

1.2 Research Motivations

Due to the energy-intensive nature of high-performance DL systems, it is clear their usage comes alongside several environmental and social issues, including their significant carbon footprint and financial cost. For this reason, the accelerating adoption of ML and DL methods within Fintech raises concerns about increasing the negative ESG impacts of the financial industry. Namely, the side effects of these complex DL models produced by their large parameter and data sets and long training phases (and hence high energy budgets) are in direct conflict with the general global effort to reduce carbon emissions, and the specific goal of sustainable finance to reduce the negative ESG impacts of the financial sector. However, the benefits of DL to finance (and indeed sustainable finance) have been shown, such as improved fraud detection (Sadgali et al., 2019) and ESG data analysis (Mehra et al., 2022), making their continued adoption inevitable.

Therefore, to minimise the negative ESG impacts of the financial sector in light of the recent Fintech revolution (and solidify the benefits of sustainable finance), the adoption of Green AI principles and methods is paramount. In the first study of its kind, this thesis investigates how the energy-efficient Green AI methods can be adapted for DL in finance. This research aims to demonstrate how the Fintech revolution (in particular the expanding use of DL) does not have to come with significant environmental and social costs, and how the models and methods used in this field can coincide with the SDGs of sustainable finance. To exemplify the promise of introducing Green AI to Fintech and sustainable finance, energy-efficient methods will be applied to one of the most popular applications of DL in finance: *financial risk modelling*. Specifically, a popular and important area of financial risk analysis known as *volatility forecasting* will be explored, developing a DL-based model that accurately predicts the future volatility (a statistical measure of dispersion) of the *S&P 500* market, but takes minimal energy and data to train.

Volatility forecasting is commonly used to give insight into the risk of a financial market or asset (French et al., 1987); it is also been extensively explored by researchers to showcase the prediction capabilities of DL: for example, Xiong et al. (2015) combine S&P 500 and Google domestic trends data to model stock volatility through DL, and Zhang et al. (2022) use DL for forecasting intraday volatility. Hence, as financial risk modelling and analysis is often cited as a core application of DL in finance (such as in the reviews of Berat Sezer et al. (2019), Ozbayoglu et al. (2020), and Thakkar and Chaudhari (2021), volatility forecasting has been chosen as the specific application within Fintech to demonstrate the promise of applying Green AI methods to the finance industry.

By showing the utility of Green AI within Fintech, this thesis aims to advance the field of sustainable finance as a whole, creating a new field of *sustainable deep learning for sustainable finance*. This work is the first study to address the conflict of interest between sustainable finance and the growing reliance of Fintech on DL systems with high energy consumption and concerning ESG impacts. Hence, the following research aims to demonstrate how the energy-efficient methods proposed by Green AI research (in the fields of NLP, CV, and mobile computing) can be

exploited within the DL systems used in finance, increasing the scope and impact of sustainable finance by allowing it to use helpful DL systems without compromising SDGs, and reducing the emissions of Fintech in general.

1.3 Contributions to Science

This research contributes to scientific literature and the finance industry in a number of ways:

1. *Expanding the applications of Green AI.* This thesis is the first application of Green AI principles and methodology to the field of finance. This will further demonstrate the utility and promise of Green AI by proving that such systems can provide compelling performance with a lower environmental cost in new domains outside of the existing research focus on NLP, CV, and mobile computing, expanding the scope of this field.

2. *Reducing the environmental impact of financial technology.* The following work combines the research fields of FinTech, deep learning for finance, sustainable finance, and Green AI, to create the new research domain of *sustainable deep learning for sustainable finance.* This further improves the ESG impact of the financial sector beyond previous work in sustainable finance by mitigating the conflict between the utility of DL models for sustainability modelling and the intrinsic carbon footprint of these energy-intensive systems.

3. *Improving the inclusivity of finance.* The reduction in the financial, environmental, and social cost of DL for finance also increases the inclusivity of this field. Improving the energy efficiency of these models creates a lower bar-to-entry to using DL in finance, allowing more industry players, developers, and individual traders to utilise the advantages brought by Fintech and DL (for example, accessing improved analytics that allow more informed financial decisions).

1.4 Research Objectives & Structure

The initial objective of this research is to analyse the efficiency of a baseline DL model used for financial volatility forecasting. In particular, the resource requirements in terms of training time and training dataset size are inspected, from which an estimate can be produced of the carbon emissions associated with training such a model. A particular focus is given to the use of *long short-term memory* (LSTM) networks, as these are the typically used DL model for sequence forecasting tasks in general, and recent research into volatility forecasting with DL (for example by Xiong et al. (2015)). This network and training process is implemented in *Python* with the popular DL framework *TensorFlow* (Abadi et al., 2016).

The baseline model is then extended with energy-efficient methods proposed by the field of Green AI. These methods adapt the way in which the LSTM-based model is trained, minimising its resource requirements. An analysis similar to that conducted upon the baseline model is then conducted, quantifying the reduction

in time and energy (and hence carbon emissions) made by these energy-efficient extensions. Specifically, methods such as *progressive training* and *mixed-precision* training are explored, and their benefits shown. Beyond these initial adaptations, data-efficient extensions to the training process are additionally be explored, such as the use of *active learning*. These methods, also adapted from Green AI research, further reduce the resource requirements of a DL model, reducing the amount of training data necessary and thus also reducing the memory and energy required by the DL training process.

Given these energy and data-efficient model extensions, an extensive analysis of the benefit of Green AI methods to this domain is made. This compares both the accuracy of the final model and each energy and data-efficient extension to the original baseline model. The resource requirements (in terms of time, data, and energy) required for training will then be compared, to evaluate any accuracy-efficiency compromises made. A general conclusion is then settled upon as to the utility and viability of Green AI methods to Fintech, sustainable finance, and the finance industry in general.

The aforementioned research will be organised as follows:

- Chapter 2: *Literature review*. The relevant literature to this research is first explored to give an overview of the key concepts and methodologies utilised in the following experimentation. This begins with a general exploration of machine learning, which then leads to a more refined discussion of the use of DL in finance, and the specific DL methods and concepts focussed upon within this thesis, giving a clear outline of the domain of financial volatility forecasting. Following this, the review explores the research domain of Green AI, with a focus on the specific methods and practices used in Green AI research which will go on to be a core aspect of this thesis.

- Chapter 3.1: *Baseline finanical volatility model*. An explanation and analysis of the baseline volatility model implemented, exemplifying the typical models, methods, and resource requirements of this domain.

- Chapter 3.2: *Energy-efficient training extensions*. Once the baseline model has been implemented, several energy-efficient adaptations to its training process shall be explored. This section gives a complete overview of these methods, explaining how they work, the motivations behind their use, and their utility in decreasing the energy consumption of training.

- Chapter 3.3: *Data-efficient training extensions*. The utilised methods to reduce the data requirements of training are then be explored, discussing their origins within Green AI, explaining how they work, and evaluating their effectiveness.

- Chapter 4: *Discussion & evaluation*. Once the detail of the utilised model and extensions has been thoroughly explained, the results of the experimentation are discussed, evaluating the accuracy and efficiency of each method to deduce the success of the application of Green AI to finance.

Chapter 2

Background & Literature Review

To understand how Fintech can be aligned with the goals of sustainable finance, first, a detailed understanding of ML and DL must be conveyed, and then the avenues used in Green AI to improve the efficiency of these methods can be effectively conveyed. This chapter achieves this by initially introducing the relevant background literature within the field of ML, centring the research of this thesis within the context of existing scientific literature. Specifically, this begins with a general summary of the components and characteristics of ML and DL methods, before focussing on the specific models and structures used within the chosen application of financial volatility modelling. The second part of the chapter outlines the field of Green AI, its motivations, applications, and the relevant tools and practices used in this domain to improve the energy efficiency of DL. These overviews of DL and Green AI act as the basis for the experimentation and analysis that follow within this thesis. Hence, the chapter concludes by drawing together these concepts to evaluate the current state of this research domain and identify the research gaps focussed upon by this thesis.

2.1 Machine Learning

The field of machine learning, first discussed by Arthur Samuel in his 1953 exploration of "mechanical brains" (Samuel, 1959), is a subset of artificial intelligence concerned with using data to allow computer programs to independently learn how to complete a given task without explicit information about the rules of such task (Samuel, 1959). This learning typically involves a *training* procedure, where the program is supplied with instances of experience from the training dataset describing the problem to be solved. The program learns from dataset D how to perform the desired task T by taking input of a data instance and outputting the result it believes is correct in this scenario. This result is then evaluatied through some performance metric P (typically computed through a *loss function*). For example, in a *prediction* task, the program takes input of a sequence of values relating to some variable and outputs what it predicts to be the next value in the sequence: e.g. Xiong et al. (2015) use a dataset of S&P 500 values and Google domestic trends to predict future stock market volatility. Alternatively, in a *classification* task, a data point is input into the program to be assigned to a distinct class: e.g. Sadgali et al. (2019) use financial transaction data to classify whether a partiuclar transacition is fradulent or not. Such tasks are typically trained through

supervised learning, where the correct result (known as the *label*) is provided to the program after it has produced its output, and the correctness of its solution is evaluated through the given metric P (producing an *accuracy* value), and its bahaviour adjusted to maximise performance according to P.

2.1.1 Neural Networks

Since the conception of machine learning, a core goal has been the development of algorithms that can accurately mimic the processing mechanisms of the human brain. These algorithms, known as *artificial neural networks* (ANN), typically centre around neural models that recreate a version of the complex system of communications between neurons in the brain. The first implementation of such a model was Frank Rosenblatt's *Mark I Perceptron* (Rosenblatt, 1958), which attempted to perform binary classification of images. Each pixel in the input image was represented as a single value in a 2-dimensional matrix of input neurons known as the *input layer*. These values were then passed to the single internal neuron through *weighted channels* to compute the weighted sum of all input values, the result of which was passed through an *activation function* to normalise the result to zero or one—representing an output classification of 'class zero' or 'class one'.

2.1.2 Deep Learning

Whilst Rosenblatt's model produced underwhelming results due to its simplicity, its design became a fundamental building block of the larger, more complex ANNs used today. Namely, modern networks build on this approach by having wider and deeper *model architectures* consisting of more internal neurons arranged in multiple layers. This approach of using *multi-layer perceptrons* was popularised by Rumelhart et al. (1986); however, recent work has pushed the boundaries of neural network performance by designing deeper and deeper architectures—known as *deep neural networks* (DNNs)—to establish the field of deep learning. These deep networks perform computations in a similar vein to Rosenblatt's perceptron. Information is passed between the input layer $l^{(0)}$ and output layer $l^{(L-1)}$ along weighted channels between neurons, with the *activation value* $a_n^{(i)}$ assigned each neuron n in layer i being determined through the dot product between the weights $W^{(i-1,i)}$ of channels entering n and the values $x^{(i-1)}$ of neurons in the preceding layer $i-1$ connected through those channels (Witten et al., 2017). This product is then summed with a *bias* term $b^{(i)}$, and passed through a chosen activation function α (Equation 2.1).

$$a_n^{(i)} = \alpha(W^{(i-1,i)} \cdot x^{(i-1)} + b^{(i)}) \tag{2.1}$$

This computation is performed over all layers of the network, starting with the input layer—whose neurons take the values of the input vector—and propagating through to the output layer—the values of which are the final network output. During training, the values of the network's parameters (i.e. the elements of weight matrix W and bias vector b) are adjusted as the model learns, manipulating its outputs to more closely align with the true labels. *Backpropagation* can be used to implement this tuning process, where the gradient of loss function with respect to the parameters is determined by propagating the error of the network

on the current input backwards through all its layers ([Zaras et al., 2022](#)). These parameters can then be adjusted in the direction of the negative gradient through *gradient descent* (towards the minimum of the loss function).

2.1.3   Sequence Modelling Problems

In machine learning, a *sequence* is an ordered set of data elements, each of uniform type and dimension: for example, a stream of text can be seen as a sequence of strings (i.e. words), or a video can be modelled as a sequence of pixel matrices (i.e. frames of the video). Mathematically, temporal sequences—typically referred to as *time series*—can be denoted by their data elements and time indices:

$$seq = \langle x_1, t_1 \rangle, \ldots, \langle x_T, t_T \rangle. \tag{2.2}$$

These sequences are commonly abbreviated to simply denote their data elements x, indexed over the time t that they were observed:

$$seq = x_1, x_2, \ldots, x_t, \ldots, x_T. \tag{2.3}$$

Sequence modelling (also known as *sequence learning*) involves using a machine learning model, known as a sequence model, to analyse and generate sequences. The problem is common in fields such as NLP, due to data being naturally ordered in sequences (e.g. sentences of text): for example, Google's *Universal Sentence Encoder* ([Cer et al., 2018](#)) is a state-of-the-art DNN for text embedding (the task of encoding sentences as vectors for performing further NLP tasks such as text classification).

Sequence modelling typically takes three main avenues. Firstly, the *sequence-to-vector* problem takes input of a sequence of data elements $x_1, x_2, \ldots, x_T$ up to timestep T (one element at a time) and outputs a single value, or vector of values (either at each timestep or delayed to after the end of the sequence). For example, in *part-of-speech tagging*—a ML application in NLP extensively reviewed by [Chiche and Yitagesu](#) ([2022](#))—a model reads elements of a text sequence, and uses the current instance (a particular word in the sequence) and its context (the previously seen words) to classify elements of the sequence grammatically as adjectives, nouns, verbs, etc. In this case, whee each grammatical tag would be allocated a numerical label, and the output would be a vector of probabilities over all tags from which we could select the most probable predicted tag.

Secondly, the *vector-to-sequence* problem completes the opposite task, taking input of a vector of values and generating a corresponding output sequence. Artificial text generators (e.g. OpenAI's GPT-3 ([Brown et al., 2020](#)) and Google's PaLM ([Chowdhery et al., 2022](#))) are a representative example of this problem, where an output sequence of text is artificially created based upon a vector of input parameters specifying the desired properties of the text.

The final, and most common, form of sequence modelling is the *sequence-to-sequence* problem, where both the inputs and outputs of the model are a sequence of elements. This task typically involves either *translation*, where the M elements in an original sequence are converted into a new sequence of length N, or *prediction*, where the original sequence of elements $x_1, \ldots, x_M$ is used to predict a subsequent sequence $x_{M+1}, \ldots, x_{M+N}$. *Machine translation* is one such application, in which

a model reads a sequence of text written in one language, then translates this text into a different language, output as a new sequence: for example, Meta AI's *M2M-100* (Fan et al., 2020), the first language model that can directly translate between any pair of 100 different languages. Prediction tasks are typically used for *time series forecasting*, where the value of data elements is predicted over N future timesteps, in such applications as *stock price prediction* (extensively surveyed by Berat Sezer et al. (2019)). In this case, the output time series $y = x_{T+1}, \ldots, x_{T+N}$ is computed based upon the context of the past time series $x_0, \ldots, x_T$. Specifically, a model M_θ (parameterised by θ) is build that approximates the mapping $y = f(x_0, \ldots, x_T | \theta)$ between past and future time series.

2.2 Recurrent Neural Networks

2.2.1 Neural Networks for Sequence Modelling

The task of forecasting future values of a sequence typically poses a challenge for traditional ANN and DNN architectures. A major reason for this is the fixed input structure of a typical ANN, meaning all elements of an input sequences would have to be fed into the network at the same time, with one element per neuron in the input layer. Hence, the input would contain no intrinsic sense of order, severely limiting the amount of analysis possible by the network, as the ordering of sequences such as time series is crucial to understanding and forecasting them (Tsantekidis et al., 2022). The chosen size of the input layer would also pose a challenge for a network with a fixed input structure, as the number of input neurons would have to be made adequately large to accommodate the possible variations in sequence lengths (e.g. the length of a time series of previous values grows as the current timestep being considered moves forward in time). Furthermore, to learn a sequence's characteristic behaviour requires persistence of information independent from its location within the sequence. Namely, the network must contextually use previous understandings to inform the current instance, wheverver they appeared in the preceding sequence. This is not possible with a traditional ANN as the parameters of channels connected to different input neurons are not shared. Therefore, when the parameters of a specific neuron and its channels are adjusted after seeing an important pattern in part of a sequence, this information will only prove useful if that patten appears again in the exact same location on the input neurons, as if the pattern occurs at a different position it will be input through different neurons with different parameters that have not been previously learned to deal with this subsequence.

2.2.2 Recurrent Networks

To overcome these issues with traditional ANNs, *recurrent neural networks* (RNNs) were conceived that utilise recurrent loops to allow varying input lengths and the persistence of learned sequential information (Sharma et al., 2022). These networks take input of a single sequence element x_t at each timestep, using this and the context of previously observed sequence elements to inform the internal state h_t of the network which acts as a form of short-term memory. This state is updated sequentially at each timestep t through the function g, parameterised

by $\theta^{(g)} = \{W^{(g)}, U^{(g)}, b^{(g)}\}$, where $W^{(g)}$ and $U^{(g)}$ are weight matrices, and $b^{(g)}$ a bias vector. Namely, the function updates the previous state h_{t-1} with new useful information learned from the input x_t (Sharma et al., 2022):

$$h_t = g(h_{t-1}, x_t | \theta^{(g)}) \tag{2.4}$$
$$= W^{(g)} \cdot h_{t-1} + U^{(g)} \cdot x_t + b^{(g)}. \tag{2.5}$$

The state h_t is then output by the network along a recurrent loop that connects the output of the network at time $t - 1$ to the input of the network at the next timestep t. This allows information to be passed between successive versions of the same neural network at different points in time; namely, the network itself stays the same at each computational step, it is only the internal state that changes. The concept can be demonstrated by unrolling the network in time, shown in Figure 2.2.2. Tsantekidis et al. (2022) note that to avoid confusion is important to remember that the sequence of networks depicted in illustrations such as Figure BELOW are in fact just the same network only at different points in time, with the connecting arrows between each signifying that the output at time $t - 1$ is being passed on to the same network at the subsequent timestep t.
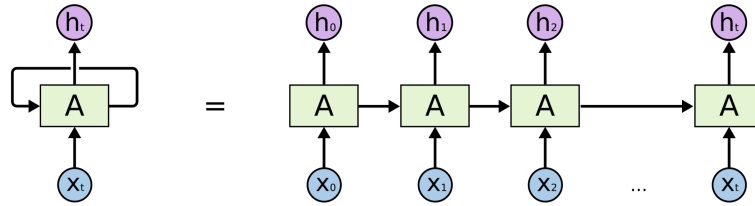

Figure 2.1: Diagram of a RNN at a single timestep t (left) and unrolled in time between timesteps 0 and t (right)

To capture more information from the input sequences, RNNs can consist of multiple internal layers (shown in Figure BELOW), allowing them to learn higher-dimensional internal representations (Bengio, 2009). In this case, each layer i retains its own internal state $h_t^{(i)}$, which is passed on to both the next layer of the network to compute its state $h_t^{(i+1)}$ (at the current input timestep) and reccurred back into itself at the subsequent timestep to compute the state $h_{t+1}^{(i)}$ (Zhang et al., 2021).
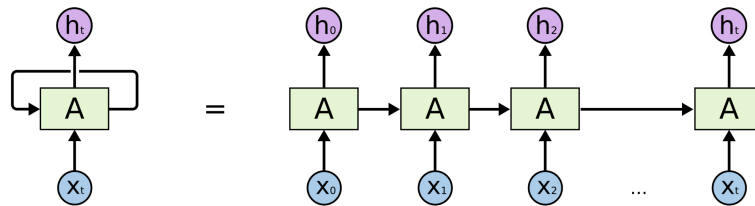

Figure 2.2: Diagram of a multi-layer RNN (showing layers 1 to L) with each layer unrolled in time between timesteps 1 and T

2.2.3   Training

Training of RNNs can be conducted through a similar backpropagation process to that used on traditional DNNs. This is done through computing the final output of the network on a given sequence (i.e. the output state at the final timestep) and evaluating the loss function between this and the true result, finding the total error of the RNN. The error is then propagated backwards through both the internal layers of the network and through time to the network state at each timestep of the input sequence—and hence is referred to as *backpropagation through time* (Zhang et al., 2021). This allows the training procedure to adjust both the network parameters (i.e. weights W and biases b) and the parameters $\theta^{(g)}$ of the state function g determining how exactly the internal state $h_t^{(i)}$ is updated at each timestep.

2.2.4   Applications

Much recent research into sequence learning has exploited recurrent networks to conduct accurate modelling. In their review of the applications of RNNs within this domain, Lipton et al. (2015) explain how NLP tasks and time-series forecasting are the most common avenues explored by researchers utilising RNNs. These two areas have been extensively surveyed in literature; in their review of DL for NLP, described the rapidly increasing popularity of RNNs in recent years, highlighting language modelling, machine translation, speech recognition, and image captioning as major areas of recent process. They highlight the RNN model of Karpathy and Fei-Fei (2014) for generating image descriptions—which significantly outperformed existing baselines—as a illustrative example of the benefits provided by this architecture. Additionally, Hewamalage et al. (2021) survey the current and future applications of RNNs for time series forecasting, highlighting the recent success of this model architecture at forecasting competitions, such as an RNN-based model by Smyl (2020) winning The M4 Competition in 2019 with cutting-edge accuracy nearly 10% greater than the utilised baseline (Makridakis et al., 2020).

2.3   Long Short-Term Memory

2.3.1   The Issue with Recurrent Networks

The traditional RNN described above is incredibly effective at learning short-term dependencies within an input sequence, and exploiting these to make accurate predictions of subsequent values. In many cases, however, to understand the full context of a sequence longer-term dependencies need to be taken into account. Unfortunately, research into the effectiveness of recurrent networks by Hochreiter (1991) and Bengio et al. (1994) has shown that RNNs cannot accurately capture or exploit long-term dependencies within sequences, as previous elements that are contextually relevant but were observed many timesteps ago can be forgotten prematurely. This issue is known as the short-term memory problem of RNNs, and is caused by the *vanishing gradient problem* (Hochreiter, 1991). This issue is inherent to the backpropagation through time algorithm and the nature of recurrent variables. As demonstrated in Equation 2.4, to calculate the value of a single re-

currence involves multiplying by the weight matrix $W^{(g)}$: ignoring the bias and input element terms (as these aren't recurrent), we have $h_t = W^{(g)} \cdot h_{t-1}$. This poses a challenge when a recurrence is made over many steps, which is the case when evaluating the recurrent state function over many timesteps in a long sequence, as the same weight matrix is used over all recurrent state computations (shown in Equation 2.6 for a length T sequence).

$$h_T = W^{(g)} \cdot W^{(g)} \cdot \ldots \cdot W^{(g)} \cdot h_0 \tag{2.6}$$

$$= (W^{(g)})^T \cdot h_0 \tag{2.7}$$

Hence, the calculation is dependent on the computation of the weight matrix product $(W^{(g)})^T$; this means that if the parameters in the weight matrix are below one (i.e. we have $W^{(g)} < 1$) then the value of $(W^{(g)})^T$ will tend to zero with increasing sequence length T. This causes the gradient of the loss function with respect to the parameters $W^{(g)}$ to also tend to zero during backpropagation, meaning no significant updates are made to the values of $W^{(g)}$ when training on long sequences. Therefore, the gradient is said to *vanish*, causing an inability of the network to learn dependencies over long sequences (Bengio et al., 1994).

2.3.2 Long Short-Term Memory Networks

Whilst several solutions to the vanishing gradient problem have been proposed— such as *skip connections* and *leaky recurrent units* (Pascanu et al., 2012)—by far the most widespread approach is the use of *Long Short-Term Memory* (LSTM), presented by Hochreiter and Schmidhuber (1997). An LSTM network is a RNN that uses multiple internal states and parameter matrices to mitigate the vanishing gradient problem and model both short-term and long-term dependencies within sequences. Similarly to an RNN, this architecture contains a core network consistent of one or more layers—where each layer is a distinct LSTM cell—and information is passed both between these cells (from the input to the output layer of the network) and between cell and and itself at the next timestep (reccuring the output at time $t-1$ to be input at time t). However, instead of the single recurrent state h_t of the simple RNN, the LSTM contains two: a short-term memory state h_t (analagous to that of a RNN) and long-term memory state c_t. This long-term state c_t is used to retain information from far back in the input sequence, adding and forgetting relevant knowledge as the context of the sequence changes.

At each timestep, the preceding long-term state c_{t-1} is input into the LSTM cell (alongside the current sequence element x_t) to compute the new state c_t. These updates are achieved through several logical *gates* within the network (between each cell and timestep). Firstly, the *forget gate* is used to decide what information to discard from the previous state c_{t-1}. The previous short-term state h_{t-1} and input element x_t are multiplied with the forget gate's weight matrices $W^{(F)}$ and $U^{(F)}$ and added to its bias $b^{(F)}$, the result of which is passed through the *sigmoid* activation function to produce the intermediate state value $\tilde{c}_t^{(F)}$, shown in Equation 2.8 (Zhang et al., 2021).

$$\tilde{c}_t^{(F)} = \sigma(W^{(F)} \cdot h_{t-1} + U^{(F)} \cdot x_t + b^{(F)}) \tag{2.8}$$

Secondly, when new relevant information is seen in the input sequence, the *input gate* is used to add it to the long-term memory state c_t. This process similarly uses the short-term state h_{t-1} and current input x_t, but evaluates both a sigmoid and *hyperbolic tangent* activation function in parallel on these inputs to decide both which values of c_{t-1} must be updated (through the sigmoid evaluation) and by how much (the tangent evaluation). The result of these two operations are then multiplied together to give the second intermediate state value $\tilde{c}_t^{(I)}$, shown in Equation 2.9 (Zhang et al., 2021).

$$(\tilde{c}_t^{(I)})_\sigma = \sigma(W^{(I,0)} \cdot h_{t-1} + U^{(I,0)} \cdot x_t + b^{(I,0)}) \tag{2.9}$$

$$(\tilde{c}_t^{(I)})_{\tanh} = \tanh(W^{(I,1)} \cdot h_{t-1} + U^{(I,1)} \cdot x_t + b^{(I,1)}) \tag{2.10}$$

$$\tilde{c}_t^{(I)} = (\tilde{c}_t^{(I)})_\sigma \cdot (\tilde{c}_t^{(I)})_{\tanh} \tag{2.11}$$

Both intermediate state values $\tilde{c}_t^{(F)}$ and $\tilde{c}_t^{(I)}$ are combined to give the new updated long-term state c_t (Equation 2.12).

$$c_t = \tilde{c}_t^{(F)} \cdot c_{t-1} + \tilde{c}_t^{(I)} \tag{2.12}$$

The third and final gate used by LSTMs is the *output gate*, which decides what each cell outputs. This is used as both the short-term state h_t and the final network output y. The output gate uses the newly computed long-term state c_t, input x_t, and previous short-term state $ht-1$; a biased weighted sum of h^{t-1} and x_t is computed and fed into the sigmoid function, which is then multiplied by the hyperbolic tangent function evaluated on c_t (Equation 2.13). Similarly to within the input gate, the sigmoid evaluation determines the values of h_{t-1} to be updated, and the tangent evaluation determines by how much (Zhang et al., 2021).

$$(h_t)_\sigma = \sigma(W^{(O)} \cdot h_{t-1} + U^{(O)} \cdot x_t + b^{(O)}) \tag{2.13}$$

$$(h_t)_{\tanh} = \tanh(c_t) \tag{2.14}$$

$$h_t = (h_t)_\sigma \cdot (h_t)_{\tanh} \tag{2.15}$$

This system of gates occurs in every LSTM cell, passing information from the input to output layer of the network at every timestep in the form of the short and long-term states. When the network reaches the final timestep, the short-term state is output as the final network output y; note, if the sequence modelling problem involves outputting a generated sequence of length N (e.g. forecasting multiple future values of a time series), this is produced one element at a time over N additional timesteps that take no input and output the value $y_i = h_{T+i}$.

2.3.3   Applications

Whilst reviewing modern RNN architectures, Lipton et al. (2015) note that most state-of-the-art applications within the field of sequence learning use a LSTM-based model. Similarly, in their review of the applications of this architecture and its variants Yu et al. (2019) assert that "almost all" important recent advancements in this domain have been facilitated by LSTMs. This is primarily due to the

architecture's ability to accurately model both long and short-term dependencies in sequences; hence, such networks are largely used for the same sequence modelling problems as traditional RNNs.

Due to the superior accuracy provided by this architecture, one of the most common use cases of LSTMs for sequence learning is NLP. For example, in their exhaustive study of language modelling, Jozefowicz et al. (2016) found an LSTM-based network to provide the most competitive results in this field. Their work evaluated several ML models such as *convolutional neural networks* and RNNs, finding that a large scale LSTM language model produced the best results, significantly improving state-of-the-art *perplexity* (a commonly used NLP metric of prediction error) from 51.3 to 30.0 on the commonly used *One Billion Word Benchmark* dataset (Chelba et al., 2013). Recently, LSTM models have been applied to even more complex tasks within the NLP space due to the ability to capture high-fidelity dependencies. This includes the work of Saleh et al. (2021) who used an LSTM-based DNN to detect hate speech in online content, producing an impressive *F1-score* (the weighted average of a model's *precision* and *recall*) of 93% on a composite dataset from multiple online hate speech repositories.

Like other RNNs, LSTMs have also been widely used for time series forecasting. Due to the advantages provided by these networks' ability to model long-term dependencies, the benefits of LSTMs have been explored across an extensive spectrum of different time series applications. For example, Shi et al. (2022) recently compared the performance of a number of networks (including both traditional RNNs and LSTMs) for predicting Bejing air quality. They concluded that LSTMs generated more accurate predictions than simpler RNNs, and demonstrated their improved long-term memory by showing LSTMs outperform other networks even when the context window available to the model is small. Another popular area of time series forecasting in which LSTM networks are increasingly being used is in the prediction of financial market variables, such as commodity prices (Ly et al., 2021), stock market indices like the S&P 500 (Fjellström, 2022), currency pairs on the Foreign Exchange (Qi et al., 2021), and the fluctuations of financial market risk (Du et al., 2019).

Chapter 3

Experiments

3.1   Baseline Financial Volatility Model

3.2   Energy-Efficient Training Extensions

3.3   Data-Efficient Training Extensions

Chapter 4

Discussion & Evaluation

Chapter 5

Conclusions & Future Work

Bibliography

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv e-prints*, art. arXiv:1603.04467, March 2016.

Michel Aglietta, Jean-Charles Hourcade, Carlo Jaeger, and Baptiste Perrissin Fabert. Financing transition in an adverse context: climate finance beyond carbon finance. *International Environmental Agreements: Politics, Law and Economics*, 15(4):403–420, Nov 2015. ISSN 1573-1553. doi: 10.1007/s10784-015-9298-1. URL https://doi.org/10.1007/s10784-015-9298-1.

Anirudh Agrawal and Kai Hockerts. Impact investing: review and research agenda. *Journal of Small Business & Entrepreneurship*, 33(2):153–181, 2021. doi: 10.1080/08276331.2018.1551457. URL https://doi.org/10.1080/08276331.2018.1551457.

Faheem Ahmad, Qamar Saeed, Syed Muhammad Usman Shah, Muhammad Asif Gondal, and Saqib Mumtaz. Chapter 11 - Environmental sustainability: Challenges and approaches. In Manoj Kumar Jhariya, Ram Swaroop Meena, Arnab Banerjee, and Surya Nandan Meena, editors, *Natural Resources Conservation and Advances for Sustainability*, pages 243–270. Elsevier, 2022. ISBN 978-0-12-822976-7. doi: https://doi.org/10.1016/B978-0-12-822976-7.00019-3. URL https://www.sciencedirect.com/science/article/pii/B9780128229767000193.

Omar Y. Al-Jarrah, Paul D. Yoo, Sami Muhaidat, George K. Karagiannidis, and Kamal Taha. Efficient Machine Learning for Big Data: A Review. *Big Data Research*, 2(3):87–93, 2015. ISSN 2214-5796. doi: https://doi.org/10.1016/j.bdr.2015.04.001. URL https://www.sciencedirect.com/science/article/pii/S2214579615000271. Big Data, Analytics, and High-Performance Computing.

Stefania Albanesi and Domonkos F. Vamossy. Predicting Consumer Default: A Deep Learning Approach. *arXiv e-prints*, art. arXiv:1908.11498, August 2019.

Fabio Alessandrini and Eric Jondeau. ESG Investing: From Sin Stocks to Smart Beta. *The Journal of Portfolio Management*, 46(3):75–94, 2020. ISSN 0095-4918. doi: 10.3905/jpm.2020.46.3.075. URL https://jpm.pm-research.com/content/46/3/75.

Amir Amel-Zadeh, Jan-Peter Calliess, Daniel Kaiser, and Stephen Roberts. Machine Learning-based Financial Statement Analysis. *Available at SSRN 3520684*, November 2020. doi: http://dx.doi.org/10.2139/ssrn.3520684.

Dario Amodei and Danny Hernandez. AI and Compute, Jun 2021. URL https://openai.com/blog/ai-and-compute/.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 610–623, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445922. URL https://doi.org/10.1145/3442188.3445922.

Y. Bengio, P. Simard, and P. Frasconi. Learning Long-term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, 5(2): 157–166, 1994. doi: 10.1109/72.279181.

Yoshua Bengio. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.*, 2(1):1–127, jan 2009. ISSN 1935-8237. doi: 10.1561/2200000006. URL https://doi.org/10.1561/2200000006.

Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review: 2005-2019. *arXiv e-prints*, art. arXiv:1911.13288, November 2019.

Elettra Bietti and Roxana Vatanparast. Data Waste. *Harvard International Law Journal*, 61, December 2019. URL https://ssrn.com/abstract=3584251.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *arXiv e-prints*, art. arXiv:2005.14165, May 2020.

Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Universal Sentence Encoder. *arXiv e-prints*, art. arXiv:1803.11175, March 2018.

Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge, Thorsten Brants, Phillipp Koehn, and Tony Robinson. One Billion Word Benchmark for Measuring Progress in Statistical Language Modeling. *arXiv e-prints*, art. arXiv:1312.3005, December 2013.

Alebachew Chiche and Betselot Yitagesu. Part of Speech Tagging: A systematic review of deep learning and machine learning approaches. *Journal of Big Data*, 9(1):10, Jan 2022. ISSN 2196-1115. doi: 10.1186/s40537-022-00561-y. URL https://doi.org/10.1186/s40537-022-00561-y.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. PaLM: Scaling Language Modeling with Pathways. *arXiv e-prints*, art. arXiv:2204.02311, April 2022.

Gary Cook, Jude Lee, Tamina Tsai, Ada Kongn, John Deans, Brian Johnson, and Elizabeth Jardim. Clicking Clean: Who is winning the race to build a green internet? Technical report, Greenpeace, 2017. URL https://www.greenpeace.org/usa/fighting-climate-chaos/click-clean/.

Felipe Arias Fogliano de Souza Cunha, Erick Meira, and Renato J. Orsato. Sustainable Finance and Investment: Review and research agenda. *Business Strategy and the Environment*, 30(8):3821–3838, 2021. doi: https://doi.org/10.1002/bse.2842. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bse.2842.

Chris Daniel, Anoop Kumar Shukla, and Meeta Sharma. Applications of machine learning in harnessing of renewable energy. In Prashant V. Baredar, Srinivas Tangellapalli, and Chetan Singh Solanki, editors, *Advances in Clean Energy Technologies*, pages 177–187, Singapore, 2021. Springer Singapore. ISBN 978-981-16-0235-1.

Jon Dennis, Charlie Kronick, Raymond Dhirani, Anthony Field, Karen Ellis, and Becky Jarvis. The Big Smoke: the global emissions of the UK financial sector. Technical report, Greenpeace UK, WWF UK, and South Pole, May 2021. URL https://www.southpole.com/publications/the-big-smoke-the-global-emissions-of-the-uk-financial-sector.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-prints*, art. arXiv:1810.04805, October 2018.

Zhichao Du, Menghan Wang, and Zhewen Xu. On Estimation of Value-at-Risk with Recurrent Neural Network. In *2019 Second International Conference on*

Artificial Intelligence for Industries (AI4I), pages 103–106, 2019. doi: 10.1109/ AI4I46381.2019.00034.

Vincent Dumoulin and Francesco Visin. A Guide to Convolution Arithmetic for Deep Learning. *arXiv e-prints*, art. arXiv:1603.07285, March 2016.

EIA. Total Energy Production 2019, 2019. Data retrieved from the U.S. Energy Information Administration, https://www.eia.gov/international/overview/world.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. Beyond English-Centric Multilingual Machine Translation. *arXiv e-prints*, art. arXiv:2010.11125, October 2020.

William Fedus, Barret Zoph, and Noam Shazeer. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *arXiv e-prints*, art. arXiv:2101.03961, January 2021.

Stephen P. Ferris and Karl P. Rykaczewski. Social Investing and Portfolio Management. *Business & Society*, 25(1):1–7, 1986. doi: 10.1177/000765038602500101. URL https://doi.org/10.1177/000765038602500101.

Carmina Fjellström. Long Short-Term Memory Neural Network for Financial Time Series. *arXiv e-prints*, art. arXiv:2201.08218, January 2022.

Nikhil Gokhale, Ankur Gajjaria, Rob Kaye, and Dave Kuder. AI Leaders in Financial Services: Common traits of frontrunners in the artificial intelligence race. Technical report, Deloitte Insights, August 2019. URL https://www2.deloitte.com/uk/en/insights/industry/financial-services/artificial-intelligence-ai-financial-services-frontrunners.html.

Brandon Graver, Dan Rutherford, and Sola Zheng. CO2 Emissions from Commercial Aviation: 2013, 2018, and 2019. Technical report, The International Council on Clean Transportation, October 2020. URL https://theicct.org/publication/co2-emissions-from-commercial-aviation-2013-2018-and-2019/.

Randall Hanegraaf, Nicole Jonker, S. Mandley, and Jelle Miedema. Life Cycle Assessment of Cash Payments. *SSRN Electronic Journal*, 01 2018. doi: 10.2139/ssrn.3267868.

Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427, 2021. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2020.06.008. URL https://www.sciencedirect.com/science/article/pii/S0169207020300996.

Sepp Hochreiter. Untersuchungen zu dynamischen neuronalen Netzen. Master's thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München, 1991.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Comput.*, 9(8):1735–1780, nov 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9. 8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

Leif Hockstad and L Hanel. Inventory of US Greenhouse Gas Emissions and Sinks. Technical report, Environmental System Science Data Infrastructure for a Virtual Ecosystem, 2018.

J. Ryan Hogarth. The Role of Climate Finance in Innovation Systems. *Journal of Sustainable Finance & Investment*, 2(3-4):257–274, 2012. doi: 10. 1080/20430795.2012.742637. URL https://www.tandfonline.com/doi/abs/10.1080/20430795.2012.742637.

IEA. Transport sector CO2 emissions by mode in the Sustainable Development Scenario, 2000-2030. Technical report, International Energy Agency, Paris, January 2022. URL https://www.iea.org/data-and-statistics/charts/transport-sector-co2-emissions-by-mode-in-the-sustainable-development-scenario-2

Mordor Intelligence. AI in Fintech Market - Growth, Trends, COVID-19 Impact, and Forecasts (2022 - 2027). Technical report, Mordor Intelligence LLP, March 2021. URL https://www.reportlinker.com/p06039495/AI-in-Fintech-Market-Growth-Trends-COVID-19-Impact-and-Forecasts.html.

Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. Exploring the Limits of Language Modeling. *arXiv e-prints*, art. arXiv:1602.02410, February 2016.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596(7873):583–589, Aug 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

Andrej Karpathy and Li Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions. *arXiv e-prints*, art. arXiv:1412.2306, December 2014.

Alison Kirsch, Grant Marr, Jason Opena Disterhoft, Henrieke Butijn, Johan Frijns, Maaike Beenes, Alberto Saldamando, Mea Johnson, Collin Rees, David Tong, Kyle Gracey, Lorne Stockman, Clement Faul, Maude Lentilhac, Ryan Cooper, Yann Louvel, Adele Shraiman, Ben Cushing, Julia Dubslaff, and Katrin Ganswindt. Banking on Climate Chaos: Fossil Fuel Finance Report 2022. Technical report, Rainforest Action Network, BankTrack, Indigenous Environmental Network, Oil Change International, Reclaim Finance, and Sierra Club, 2022. URL https://www.bankingonclimatechaos.org/.

Satish Kumar, Dipasha Sharma, Sandeep Rao, Weng Marc Lim, and Sachin Kumar Mangla. Past, Present, and Future of Sustainable Finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, Jan 2022. ISSN 1572-9338. doi: 10.1007/s10479-021-04410-8. URL https://doi.org/10.1007/s10479-021-04410-8.

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the Carbon Emissions of Machine Learning. *arXiv e-prints*, art. arXiv:1910.09700, October 2019.

Zachary C. Lipton, John Berkowitz, and Charles Elkan. A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv e-prints*, art. arXiv:1506.00019, May 2015.

Siblis Research Ltd. Total Market Capitalization of Public U.S. Companies, March 2022. Data retrieved from the Global Equity Valuations of Siblis Research, https://siblisresearch.com/data/us-stock-market-value/.

Racine Ly, Fousseini Traore, and Khadim Dia. Forecasting Commodity Prices Using Long Short-Term Memory Neural Networks. *arXiv e-prints*, art. arXiv:2101.03087, January 2021.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74, 2020. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2019.04.014. URL https://www.sciencedirect.com/science/article/pii/S0169207019301128. M4 Competition.

Jens Malmodin and Dag Lundén. The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015. *Sustainability*, 10(9), 2018. ISSN 2071-1050. doi: 10.3390/su10093027. URL https://www.mdpi.com/2071-1050/10/9/3027.

Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koome. Recalibrating Global Data Center Energy-use Estimates. *Science*, 367(6481):984–986, 2020. doi: 10.1126/science.aba3758. URL https://www.science.org/doi/abs/10.1126/science.aba3758.

Srishti Mehra, Robert Louka, and Yixun Zhang. ESGBERT: Language Model to Help with Classification Tasks Related to Companies Environmental, Social, and Governance Practices. *arXiv e-prints*, art. arXiv:2203.16788, March 2022.

Iota Kaousar Nassr, Robert Patalano, Pamela Duffin, and Ed Smiley. Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges and Implications for Policy Makers. Technical report, OECD, August 2021. URL https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm.

Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer. Deep Learning for Financial Applications : A Survey. *Applied Soft Computing*, 93:106384, 2020. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.

2020.106384. URL https://www.sciencedirect.com/science/article/pii/S1568494620303240.

Maximilian Palmié, Joakim Wincent, Vinit Parida, and Umur Caglar. The Evolution of the Financial Technology Ecosystem: An introduction and agenda for future research on disruptive innovations in ecosystems. *Technological Forecasting and Social Change*, 151:119779, 2020. ISSN 0040-1625. doi: https://doi.org/10.1016/j.techfore.2019.119779. URL https://www.sciencedirect.com/science/article/pii/S0040162519310595.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training Recurrent Neural Networks. *arXiv e-prints*, art. arXiv:1211.5063, November 2012.

Ioana-Stefania Popescu, Claudia Hitaj, and Enrico Benetto. Measuring the Sustainability of Investment Funds: A critical review of methods and frameworks in sustainable finance. *Journal of Cleaner Production*, 314:128016, 2021. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2021.128016. URL https://www.sciencedirect.com/science/article/pii/S0959652621022344.

Joseph Power, Jordan McDonald, So Lefebvre, and Tom Coleman. The Time to Green Finance: CDP Financial Services Disclosure Report 2020. Technical Report DOE-SLC-6903-1, CDP Worldwide, Great Tower Street, London, 2020. URL https://www.cdp.net/en/research/global-reports/financial-services-disclosure-report-2020.

Ling Qi, Matloob Khushi, and Josiah Poon. Event-Driven LSTM For Forex Price Prediction. *arXiv e-prints*, art. arXiv:2102.01499, January 2021.

Frank Rosenblatt. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 6:386–408, 1958.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning Representations by Back-Propagating Errors. *Nature*, 323(6088):533–536, October 1986. doi: 10.1038/323533a0.

I. Sadgali, N. Sael, and F. Benabbou. Performance of Machine Learning Techniques in the Detection of Financial Frauds. *Procedia Computer Science*, 148:45–54, 2019. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2019.01.007. URL https://www.sciencedirect.com/science/article/pii/S1877050919300079. THE SECOND INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING IN DATA SCIENCES, ICDS2018.

Hind Saleh, Areej Alhothali, and Kawthar Moria. Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model. *arXiv e-prints*, art. arXiv:2111.01515, November 2021.

A. L. Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959. doi: 10.1147/rd.33.0210.

Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. Green AI. *arXiv e-prints*, art. arXiv:1907.10597, July 2019.

Deepak Kumar Sharma, Mayukh Chatterjee, Gurmehak Kaur, and Suchitra Vavilala. 3 - Deep Learning Applications for Disease Diagnosis. In Deepak Gupta, Utku Kose, Ashish Khanna, and Valentina Emilia Balas, editors, *Deep Learning for Medical Applications with Unique Data*, pages 31–51. Academic Press, 2022. ISBN 978-0-12-824145-5. doi: https://doi.org/10.1016/B978-0-12-824145-5.00005-8. URL https://www.sciencedirect.com/science/article/pii/B9780128241455000058.

Jimeng Shi, Mahek Jain, and Giri Narasimhan. Time Series Forecasting (TSF) Using Various Deep Learning Models. *arXiv e-prints*, art. arXiv:2204.11115, April 2022.

Slawek Smyl. A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting. *International Journal of Forecasting*, 36(1):75–85, 2020. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2019.03.017. URL https://www.sciencedirect.com/science/article/pii/S0169207019301153. M4 Competition.

David R. So, Chen Liang, and Quoc V. Le. The Evolved Transformer. *arXiv e-prints*, art. arXiv:1901.11117, January 2019.

Chartis Research Staff. AI in RegTech: a quiet upheaval. Technical report, Chartis Research and IBM, February 2019. URL https://www.chartis-research.com/technology/artificial-intelligence-ai/ai-regtech-quiet-upheaval-10726.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and Policy Considerations for Deep Learning in NLP. *arXiv e-prints*, art. arXiv:1906.02243, June 2019.

Ankit Thakkar and Kinjal Chaudhari. A Comprehensive Survey on Deep Neural Networks for Stock Market: The need, challenges, and future directions. *Expert Systems with Applications*, 177:114800, 2021. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2021.114800. URL https://www.sciencedirect.com/science/article/pii/S0957417421002414.

Avraam Tsantekidis, Nikolaos Passalis, and Anastasios Tefas. Chapter 5 - Recurrent Neural Networks. In Alexandros Iosifidis and Anastasios Tefas, editors, *Deep Learning for Robot Perception and Cognition*, pages 101–115. Academic Press, 2022. ISBN 978-0-323-85787-1. doi: https://doi.org/10.1016/B978-0-32-385787-1.00010-5. URL https://www.sciencedirect.com/science/article/pii/B9780323857871000105.

Sandra Van Der Laan and Nina Lansbury. Socially Responsible Investing and Climate Change: Contradictions and Challenges. *Australian Accounting Review*, 14(34):21–30, 2004. doi: https://doi.org/10.1111/j.1835-2561.2004.tb00237.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1835-2561.2004.tb00237.x.

Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. Chapter 10 - Deep Learning. In Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal, editors, *Data Mining (Fourth Edition)*, pages 417–466. Morgan Kaufmann, fourth edition edition, 2017. ISBN 978-0-12-804291-5. doi: https://doi.org/10.1016/B978-0-12-804291-5.00010-6. URL https://www.sciencedirect.com/science/article/pii/B9780128042915000106.

Ruoxuan Xiong, Eric P. Nichols, and Yuan Shen. Deep Learning Stock Volatility with Google Domestic Trends. *arXiv e-prints*, art. arXiv:1512.04916, December 2015.

Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, 31(7):1235–1270, 07 2019. ISSN 0899-7667. doi: 10.1162/neco_a_01199. URL https://doi.org/10.1162/neco_a_01199.

Adamantios Zaras, Nikolaos Passalis, and Anastasios Tefas. Chapter 2 - Neural Networks and Backpropagation. In Alexandros Iosifidis and Anastasios Tefas, editors, *Deep Learning for Robot Perception and Cognition*, pages 17–34. Academic Press, 2022. ISBN 978-0-323-85787-1. doi: https://doi.org/10.1016/B978-0-32-385787-1.00007-5. URL https://www.sciencedirect.com/science/article/pii/B9780323857871000075.

Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into Deep Learning. *arXiv e-prints*, art. arXiv:2106.11342, June 2021.

Chao Zhang, Yihuang Zhang, Mihai Cucuringu, and Zhongmin Qian. Volatility Forecasting with Machine Learning and Intraday Commonality. *arXiv e-prints*, art. arXiv:2202.08962, February 2022.