Green AI

[1] Chirag Bhatheja,
[1] Department of computer science and IT, Aktu university ,
[1] chiragbhatheja@gmail.com

Abstract— The use of technologies has dramatically transformed various domains, but it has also raised issues regarding the use and effects of AI on the environment. This paper aims to highlight the specific issues associated with artificial intelligence, and in particular, green AI, or environmentally conscious AI. AI technologies, when put to use in a number of industries like health care, technology, and climate science, show immense possibilities. However, the challenge remains in the energy cost needed to train vast models. These are some issues that we elucidate upon in more detail below: the sustainability barriers of their systems, particularly focusing on new strategies that adopt Tensor Networks to optimize AI models. We seek new approaches for reducing the workload involved in AI energy demanding tasks, analyze the performance versus energy metrics, and accentuate incorporating intelligent design to make it more sustainable. In an indirect way, these transformations enable us to disclose the issues related to the environment in the context of AI and emphasize the importance of the sustainable direction for further developments in this industry. One of the most critical issues we recognize calls for a compromise between the pursuit of better AI systems and the strive for greener options.

I. INTRODUCTION

AI, as well as ML, has altered several industries from healthcare to self-driving cars to financial modeling to enviromental monitoring. However, these changes have raised a new concern regarding AI, which is environmental changes as a consequence of these technologies. Indeed, as AI systems become more complicated, not only do they consumer more energy, but they also need more data. The demand for computation power increases, and, unfortunately, the environment pays the price in the form of greenhouse gas emissions, excessive energy usage, increased carbon footprint, resource depletion, and even ecosystem collapse.

AI is no different; its environmental cost won't be evident immediately, particularly when set alongside other sectors such as manufacturing or transport. Large models in fact require heavy computation resources, and AI's ecological footprint becomes heavier as they grow. For example, AI models such as GPT-3 use natural language processing and require advanced machine deep learning. Cloud-based resources, when compared to other standard computing options, are far more efficient, however, they still need drastic amounts of energy to function. And when combined with the deep learning models, it equals out to extensive greenhouse emissions. Indeed, AI makes use of millions of petaflop/s becoming a top-tier model.

The search for solutions underlined the need for AI frameworks to be revised with sustainability in mind, seeking to limit the extent of environmental degradation caused by AI systems. These movements define Green AI.

AI models tend to require big data to deliver their intended purpose. By definition, Green AI means cutting down energy use when creating, training, and using machine learning models. This speaks to AI's potential to transform societies and businesses for the better and tackle some of the world's toughest challenges. These alterations include: enhancing the efficiency of AI algorithms, implementing a more low-powered computing infrastructure, switching to greener practices in data centers, and using AI to make other industries more sustainable.
Increasing development of AI solutions has created an even larger hunger for computational power. This increase has furthered the concerns regarding AI's environmental impact. Building sophisticated Business Intelligence systems powered by artificial intelligence technology, also known as AI, depends on enormous quantities of energy, mostly sourced from fossil fuels. The tremendous electricity needed to train sophisticated AI models results in increased emission levels of greenhouse gases. These issues complicate the quest for sustainable global development that focuses on cutting carbon emissions and addressing climate change.

But, the escalating power and complexity of AI may also be part of the fix. With the application of sophisticated algorithms, machine learning approaches can aid in decreasing energy use and improving the sustainability of technological infrastructure within AI systems, across many sectors.

If we focus on Green AI, the integration of sustainability principles into the practice and the research of AI is still marginal. The metrics that are available are not standardized for the evaluation of AI models' environmental impact and there are not enough tools to lessen their ecological footprint. A big number of researchers and practitioners are still unaware of the environmental footprint of AI and its scale, so green technologies are not common. However, this narrative is shifting towards a more Green AI friendly camp, advocating for new energetic efficient algorithms, hardware, and quantifying and mitigating AI's carbon emissions.

This paper describes an operational framework for identifying the weaknesses of current AI practices and aims to address the gaps to obtain their final outcomes by employing a life cycle oriented workflow system design approach that takes into account sustainably the deployment of models.

This paper also addresses the topic which concerns Tensor Networks (TNs) aids the AI models by enhancing their core efficiency and enabling a more sustainable operation. Using deep learning techniques for Green AI implies the use of TNs which enables the high degree reduction of energy required to train and run AI models. Moreover, a review of the current literature examines the position of TNs as an enabler of Green AI's computational efficiency for AI models.

Last but not the least, the environmental interactions of the AI practitioners, the hardware makers, and the data center managers which they happen to be a part of require greater concern. At each stage of AI development from research to deployment, there is a need to devise effective practical measures that can be utilized in the entire life cycle of AI. As such, this paper approaches the issue of AI and sustainability AI in broader senses, which are to cast attention on the environmental aspects and calls for the use of the technology without harming the earth.

Policymakers in different parts of the world share in the concern to enhance long-term and AI-based innovation strategies so that the transition of industry is done in a responsible way.

The Environmental Effect of AI Technology Artificial Intelligence (AI) has a very great environmental effect which is worrying due to the rapid growth of the sector. Especially AI systems that employ deep learning and large machine learning models are very resource-consuming. These systems need immense amounts of energy, and energy resources that are available are often strained. When complex AI models like those used in Natural Language Processing (NLP) and Computer Vision are trained, they have to sift through wide-ranging data sets over distributed computing. This not only releases a lot of greenhouse gases but also requires lots of energy. The energy consumption associated with AI models is driven by the computational resources necessary to exploit, execute, and train the model. For example, significant AI models require training on even more powerful AI models. GPT-3, for example, is a large AI model condensation that contains over 175 billion parameters. Training it consumes the equivalent of tens of thousands of petaflop/s-day computations.This results in significant energy use which increases the carbon footprint of AI significantly.

A majority of AI architectures are constructed within gigantic data hubs that, in most cases, operate on power drawn from fossil fuels. These centers are made up of thousands of servers that have to work for long durations, even 24 hours a day. For such centers, power consumption is on the higher side which has proven to be the case when examined closely over time, even for data centers built for the sole purpose of aiding AI models. In a study carried out in 2019, it was estimated that one Large Deep Learning Model would result in the total carbon emissions equal to that of five places of combustible automobile engines in their entire lifespan. Additionally, having advanced AI models and their uses commercially available is quite powerful, but so is the energy requisite demand. This further escalates the exponential increase in the amount of computational resources demanded for training, driving deep intel deficits within the current AI working framework. Data hubs have made immense progress in being run on solar or other eco-friendly energy sources, but the energy used in deploying AI models is still a prominent issue, especially for places that run using energy sources that aren't renewable.

II. CARBON FOOTPRINT OF AI MODELS

The carbon footprint of an AI model is a significant part of its environmental impact. Carbon footprint refers to the total amount of carbon dioxide (CO2) and other greenhouse gases. emitted during the production and use of AI systems. This includes not only the energy used during model training. But it also includes the environmental costs of building the hardware used in data centers, such as processors, memory. Storage device -f device Recent studies have attempted to quantify the carbon footprint of AI models, with results showing that large-scale AI is responsible for the greenhouse gas emissions that training is responsible for, such as studying the environmental impact of training. Training large artificial neural networks It was found that carbon emissions from training Weighing 284 metric tons, these models can be as high in CO2 as five cars over their lifetime. The resulting greenhouse gas emissions are comparable. This highlights the urgent need for solutions that reduce the carbon footprint of AI systems.

The carbon footprint of AI is specifically concerning due to the fact it is tied to the power fed on all through model training. If the electricity is derived from fossil fuels, the carbon emissions associated with AI models may be even higher. This is particularly actual for regions in which coal, herbal gas, or oil stay the dominant sources of power. As AI keeps to become greater ubiquitous, addressing its carbon footprint might be essential to reaching worldwide sustainability dreams, such as the Paris Agreement's goal of proscribing global warming to one.Five tiers Celsius.

In addition to the strength consumed at some point of AI model training, the environmental effect of the hardware used to run AI models is another critical consideration. AI research and deployment require specialised hardware, consisting of Graphics Processing Units (GPUs), Tensor Processing Units (TPUs), and other high-overall performance computing devices. The manufacturing of those components involves the extraction of uncooked materials, the producing of semiconductor gadgets, and the assembly of complicated digital systems, all of that have environmental prices. Mining for metals including silicon, copper, and uncommon earth elements utilized in hardware production can result in habitat destruction, water pollution, and high power consumption.Furthermore, the lifecycle of hardware includes strength consumption now not only throughout its manufacturing however also for the duration of its operational lifespan. Data centers that house AI fashions are powered by means of a significant quantity of power, and their cooling structures consume extra power to maintain top of the line temperatures. As AI models develop large and more complex, the demand for hardware resources will continue to boom, contributing to further environmental degradation. To address those challenges, it's miles crucial to explore extra sustainable options for hardware and cooling structures, inclusive of using power-efficient substances, adopting low-strength chips, and transitioning to liquid cooling solutions.

The growing call for for AI infrastructure also places extra stress on information centers, which can be already beneath strain to satisfy the needs of different industries. As extra AI-pushed programs emerge, the need for green, sustainable information facilities will become even extra vital. Innovations consisting of facet computing, in which information processing is moved closer to the supply of records technology, may additionally help to reduce the strength demands of centralized information centers. Additionally, the usage of renewable power sources to strength information facilities, including solar and wind, is a promising road to lessen the environmental effect of AI.

III. STRATEGIES FOR REDUCING ENVIROMENTAL IMPACT OF AI

There are ways and means to address the issue of carbon emissions in the AI sector like enhancing productivity and improving the efficiency of AI systems. One vital strategy is the modification of the algorithms employed by AI models so that they are less computationally intensive. This can be done by pruning, quantization, and knowledge distillation which tweaks the AI model's complexity and performance to make it smaller and more effective. The energy needs of the ai infrastructure can be greatly reduced, as well as the environmental impact while increasing the efficiency of the models.

Other strategies include dynamic and static energy efficient hardware model development such as specially designed chips and processors for AI workloads that are faster and more efficient. Alternative semiconductor materials are also being investigated by researchers to produce better and more efficient devices. More energy efficient cooling systems, the use of renewable energy, and the employment of green building techniques can all help improve the sustainability of data centers.

At last, it is now being acknowledged how AI affects the environment at the data governance step. The processes of data filtering, saving, and transmitting can use a lot of energy, particularly when large volumes of information are being processed. More energy-efficient data centers can be created by adopting more effective data management methods and limiting the volume of data transported to and from the devices.

The issue has been raised.

It is worth noting that the impact AI pose to the environment is an issue that should be taken seriously by academics, practitioners, and legislators. Suffice to say, it should not be ignored as AI is becoming more pervasive in various sectors. Its negative effects on the environment, if any, must however be weighed against the positive benefits it affords. The optimal operation of AI models, the design of low energy consuming processors, and the use of sustainable data center measures can all help to mitigate the problem and enable the world to achieve a truly sustainable future. These methods, consolidated and known as Green AI, form the basis from which it is possible to achieve the desired objectives without impeding on the world sustainability assumptions.

IV. GREEN AI AND ITS PRINCIPLES

A new approach in the field space of AI has emerged, Green AI, that aims to address the growing concern of the environmental effects of AI. Green AI is a concept that looks to mitigate the environment impacts of the development and application of artificial intelligence systems such as energy usage, carbon dioxide emissions, and resource usage. Unlike traditional AI research that tries to make models more accurate and perform better, Green AI focuses on sustainability and efficiency. AI systems should be developed and operated in a manner that is less impactful on the environment.

Understanding Green AI

Green AI can be understood as an approach that focuses on the impacts of information and computing technologies in the efforts to build, maintain, and enhance AI systems. This includes but not limited to the age of energy-efficient algorithms. The primary benchmarks of Green AI are the same as those of traditional AI, albeit with a shift towards improving operational efficiency within ecological boundaries of AI technology. Societal impacts warranting concern include energy spent, emissions rates, the extent of environmental damages caused by AI technology. Green AI is a philosophy that emanates from concern over high-energy AI systems. Sophisticated computer systems that support sophisticated models tend to utilize complex algorithms. Such algorithms and computer systems create sophisticated problems for the operation environment.

There are certain considerations which guide the definition of Green AI, so as to achieve its goals in more effective and sustainable manners. These strategic guidelines can be defined at a more generic level into three groups: energy savings, resource-use optimization, and mitigation of environmental impacts.

1. Energy Efficiency

Perhaps one of the most important aims of Green AI is increasing the energy efficiency of AI systems. It is a common underlying trend that as AI models become more and more complex, the energy resources needed also increase. In order to solve these problems, Green AI focuses on energy neutral design of algorithms and models that are computationally inexpensive with an equivalent or even enhanced level of performance. AI models can be energy efficient due to techniques, including but not limited to, model pruning, quantization, and distillation; all of which minimize a model's size while maintaining the same level of accuracy, thus allowing the hardware to more efficiently consume energy.

Likewise, Green AI encourages the adoption of dedicated hardware that focuses on the specific multilayer tasks of AI like Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), which are capable of AI computation at a much higher rate than traditional processors. These specialized chips are capable of decreasing energy usage during model training and inference, which can minimize the environmental effects. The consumption of renewable energy resources such as solar and wind can also improve the sustainability of AI systems by powering data centers and AI infrastructure.

2. Resource Optimization

Another fundamental principal of Green AI is the optimization of resources like computational resources and data. AI systems are often built using Stored Value Units of Data to train the Machine Learning models and these datasets are not created, stored, and processed in the most efficient way. Green AI promotes the use of data reduction strategies that include but aren’t limited to data compression and efficient data preprocessing techniques.

The promotion of reuse of previously developed models and datasets is another enhancement offered by Green AI which subsequently cuts down unwanted calculations and data collection. For example, models utilizing transfer learning are able to draw upon frameworks from other models to lower the high expenses associated with constructing the particular model from scratch. This not only saves on the cost of building competent frameworks but also increases the pace at which new robust AI frameworks are put into operation by utilizing the previously available accurate frameworks.

3. Reducing The Environmental Impact

Apart from the consumption of energy and the optimazation of resources, Green AI also tends to other issues such as minimization of the carbon footprint of the AI systems both during and after the system has been built. This particular issue also incorporates the effort meant to curb the e-waste originated during the life cycles of outdated or non-functional AI systems. Construction of the AI infrastructure using sustainable materials and sources of energy is being advocated by Green AI as a way of addressing this e-waste issue.

Designing AI mlodels that can be easily decomposed or transformed into other functional components once decommissioned is another crucial way to mitigate the carbon footprints. For instance, AI models data centers could be built with efficient cooling solutions and their construction materials could be eco-friendly. Moreover, sourcing of components for AI systems peripherals can be subsituted by components and minimume useage of rare earth metals to provide a wider sustainability focus.
Attainable Steps Towards Green AI

These steps include the practice of using data to enhance deployment efficiency and hardware management alongside synthesis and training of models and data.

1. Increasing Efficiency in AI Model Development

Adele Lele, Program on Computational Social Science, Cambridge University AI has a great risk to the environment and one of its measure involves reducing the size of the AI models alongside their operational costs, especially in deep learning. Deep learning networks have entirely been referred to as Green AI. Deep computing encourages applying strategies within these models that are less demanding on the resources available.

• Pruning: In AI, pruning describes the efficient model creation period that acts towards the elimination of unnecessary parameters or connections from a neural network. By eliminating redundant weights and neurons, pruning can significantly reduce sizeable models that are overly large, which would result in high energy usage throughout training and inference periods.

• Quantization: Quantization allows for fewer numbers to be processed during a model’s computations. This results in easier data types that can be processed much faster. By using lower precision for computations, quantization can reduce the computational load and energy consumption of AI models.

• Knowledge Distillation: This technique normally surrounds the process of the transferring specific information from the teacher type of model to the student type of model. In other words, knowledge distillation encompasses transferring knowledge from unique broad, complex model into an already existing simpler one. This allows the smaller model to perform at its peak while being less demanding on resources.

2. Applying AI for environmental Issues Mitigation

Green AI not only seeks to examine how AI systems can be made to have minimum environmental footprints but also how AI can be brought to bear on environmental problems. This is the branching out of AI into areas like AI4GreenAI where AI solves the green issues like climate change, energy crises, and resource allocation.

AI can be used to enhance energy efficiency in buildings, transportation, and industry, and in doing so, diminish energy consumption and carbon emissions. Similarly, AI can be used to monitor and anticipate future environmental scenarios, making it possible to designate conservation resources where they will have maximum impact. Machine learning based models can also be used to provide improved climate data analytics, weather forecasting, and renewable energy technologies deployment and implementation.

3. Environmentally Friendly Hardware Creation

There is a growing focus on creating AI systems hardware that is in line with Green AI, which has sustainability as one of its key pillars. Custom made processors like GPUs and TPUs could be fabricated to be more power friendly, thus minimizing consumption when AI tasks are being processed. There is also room for innovation to develop new hardware materials or methods that have higher performance and lower environmental impacts.

Improvements are also being made aimed at lessening the environmental costs caused by data centers which have the components required for the AI systems. New technologies like liquid cooling systems and powering the data centers using renewable sources of energy can help considerably in curtailing the carbon emissions associated with infrastructure for AI.

Future Challenges and Directions

Even though the concepts of Green AI seem appealing, the major issue is how to implement them on a wider perspective. One of the key challenges is the balancing act between model accuracy and efficiency. With the continuous advancements at the core of AI, the terrain of technology grows richer in detail, but so does the demand for computing resources. Striking this particular balance in the field remains a major challenge.

In addition to that, better tools and methodologies are required to quantify and assess the structural environment of impact AI systems. Some efforts have manufactured energy consumption, carbon footprint and other sustainable indicators for AI models; however, there is no universally accepted standard for doing so.

In the future, it's probable that the domain of Green AI will still develop further as scientists develop methods for boosting AI effectiveness, improving hardware, and limiting the carbon footprint left behind by AI systems. It is very likely that Green AI can play a massive role in achieving the global goals for sustainability by developing sustainable practices in artificial intelligence and employing AI as a tool for environmental conservation.

V. EFFECTS OF TENSOR FLOW ON GREEN AI

With regard to the impact of tensor networks on the green AI, it must be said that in the past few years, one has witnessed the development of Tensor Networks (TNs) as a powerful tool for machine learning and artificial intelligence, which can lead to enhanced application of efficient computation. Their use in Green AI is encouraging because they may help solve some of the major problems of energy consumption and resource usage in AI systems. There is great hope and expectation that as understanding of the structure of TNs deepens, it will be possible to build efficient and more sustainable AI systems, which is at the heart of the Green AI initiative.

Get to Know Tensor Networks (TNs)

At the most fundamental level, tensor networks are a mathematical model used to store and process complicated multidimensional data in an efficient and compact form. A tensor is multdimentional in the way that it extends the concept of matrices and vectors. In machine learning and AI in general, tensors are multi-dimensional arrays of data that can be used to denote inputs and their corresponding weights and outputs of AI models.

Tensors take advantage of the structural aspects found deep within the high-dimensional tensors and compress them into a more manageable form. This process can minimize the number of parameters needed to represent a model while still retaining its ability to approximate relationships of higher order in the data. This feature of compression offered by TNs enables AI models to work under lower computational costs which translates into less energy and smaller carbon footprints which is one of the aims of Green AI.

The foremost feature of appeal for TNs lies in their ability to decompose high-dimensional data into factors with lower dimensionality in a more efficient way. This is accomplished by decomposing large tensors into lower-rank components which are called modes or bonds. These modes can be used to manipulate data while maintaining important relationships. The structure of TNs, more specifically, the way they decompose complex data into lower dimensional components is crucial for obtaining lower computational complexity and energy expenses in machine learning models.

Tensor Networks and Their Computational Efficiency

The ability of tensor networks (TNs) to enhance computational effectiveness is one of the major advantages within Green AI. The conventional machine learning approaches, notably deep learning models, have a very high requirement on computing resources needed to conduct training and inference. This is mainly because of the high parameter counts and the complexity of the architectural models. To illustrate, deep neural networks (DNNs) are designed with millions, or billions of parameters, which place a significant load on the hardware for both training and deployments phases.

One possible means of alleviating this burden is the compression of model parameters. With the help of the internal structure of data, TNs are capable of xapproximating the same functions with fewer parameters. This results in a lower number of operations during model training along with lower energy consumption during model deployment. Furthermore, by modeling the data intelligently, TNs allow for compression of larger amount of information into smaller, more efficient format, which reduces the requirements for memory resources significantly.

For example, in deep learning, TNs have been employed in the creation of Tensorized Neural Networks (TNN) - more advanced and powerful forms of neural networks. TNNs achieve this by employing tensor networks that limit the amount of parameters assigned to the network making it more efficient without sacrificing precision. In contrast to the prevailing techniques in deep neural networks, TNNs are more energy efficient because they require less computational power and memory.

In addition, different forms of TNs can be employed in several parts of the machine learning process such as data understanding, data modeling, and even modeling deployment. Since TNs assist in optimizing the entire machine learning process, they aid in the minimizing of energy resources that are often more than the combatable environmental impact.

TNs provide efficiency gains beyond just computational resources. In addition to this, TNs allow for greater model performance, scalability, and accuracy. One of the toughest issues in machine learning remains the trade-off between model complexity and accuracy. Yes, there is such a thing as too much computation. Larger models require more of it though, and that is a problem on its own. However, with TNs, there is an option to have models that are devoid of complexity but still maintain high accuracy.

When considering Green AI, the power of TNs to obtain a coding sugar without the crumple of model performance is vital. Due to tensor decomposition, the TN structures can lead to appropriate simplification which makes the model to capture the key aspects but generalize fairly without the need of large amounts of parameters. This creates models that greatly reduce the resources needed while maintaining accuracy, thus lessening the damage to the environment.

TNs, in addition, have the capacity of being widely scalable. With the growth of AI complexity, there is also growth in the demand for computational resources. But, with TNs, there is great flexibility that allows to scale to larger datasets and more complex models without the increase in computation needs. This is crucial within the scope of Green AI, developing complex AI systems with no damage to the ecosystem.

The Diversity in AI Applications of Tensor Networks

AI domains have Widely adopted the use of Tensor Networks which include Deep Learning, Computer Vision, Reinforced Learning and even Natural Language Processing. They are appropriate for almost any work that involves heavy lifting in processing or multidimensional calculations since they compress data while boosting effectiveness.

1. Deep Learning: TNs are being incorporated into Neural Networks with the aim of higher performance in deep learning. For instance, the Tensorized Convolutional Neural Networks (TCNNs) which are Tensor Neural Networks, utilize advanced tensor decomposition methods to enhance the speed of training and reduce energy spending. This is done by decreasing the number of parameters in the convolutional layers. Likewise, sequence processing tasks including language modeling and time-series forecasting can be done using Tensorized recurrent neural networks (TRNNs) with less calculations and energy expenditure.

2. Reinforcement Learning: Complex state-action value functions can be accurately modeled with TNs in reinforcement learning, all while using fewer parameters. Making use of tensor decomposition in TNs can approximate the value function with greater efficiency, using less effort as TNs with low power are easier to train. This makes Reinforcement Learning algorithms more powerful and scalable. This ability is vital for environments that have a vast state environment where classic means are too energy expensive.

3. Natural Language Processing (NLP): TNs have also been applied to tasks such as language and machine translation modeling. By utilizing tensor models, research can be carried out more efficiently, which helps to lower the model parameters in NLP without compromising the accuracy in these NLP models, such as sentiment analysis, text generation, and even named entity recognition. This promotes the effective processing of the large text corpuses, which alleviates the burden on computation during the training and inference stages of the NLP applications.

4. Computer Vision: In most computer vision tasks, tensor networks are able to lower the model parameters within the convolutional neural networks (CNNs) which perform object recognition, image segmentation, and high efficiency face recognition within the CNNs. These tensor based models also help to increase the processing speed for images that are in higher dimensions, resulting in higher growth rates in training speed and lower energy expenditure.

Challenges and Limitations of Tensor Networks in Green AI

Despite providing a wonderful opportunity for improving energy consumption and potentially increased computational effectiveness, including additional challenges and limitations for the use of Tensor Networks in Green AI.

1. Data Compatibility: The primary challenge of implementing tensor networks (TNs) in machine learning is data compatibility. Structure-oriented data like images, time series, or even natural language texts are most compatible with TNs. On the contrary, not all data can undergo tensor decomposition. Therefore, the usage of TNs in loosely structured or unstructured data could lead to inadequate results.

2. Overhead Computation: It is true that tensor networks (TNs) minimize the number of parameters and computations in AI models, however, the process of tensor decomposition in itself can add some level of computational overhead. For some operations, this additional overhead might surpass the

efficiency gains, especially when data is not suitable for tensor decomposition.

3. Asension: Although hyper-D TNs have great asension for model dimension, there might be asension issues regarding the volume of data or the complexity of the model. In such situations, the amount of resource consumption that goes to the processing and storage may increase due to the additional intricate details of tensor decomposition.

In Closing Remarks

In the close of my research work, I can state that tensor networks make it possible to improve the efficiency targets of Green AI because Tensor Networks fosters the development of AI technologies with maximized efficiency in model, energy, and resource consumption. Taking advantage of the structures of high-dimensional data makes it possible for TNs to build greater systems with less expense, and more powerful machine learning algorithms. The use of TNs in many areas of AI such as deep learning and even natural language processing reveals its differentiating features useful for the green AIs.

Nevertheless, there are still challenges with regard to data compatibility, computationally intensive, and scaling up. With the ongoing development of Green AI, new innovative approaches to traditional networks as well as their implementation into AI systems will play an important role in striking a balance between achieving accuracy and being environmentally conscious. If TNs are used in the construction of AI models, the AI community can significantly make progress in diminishing the negative consequences of AI on the world while still being useful and efficient.

VI. REAL WORLD USES OF TENSOR NETWORK IN GREEN AI

When it comes to increasing the energy efficiency of AI models, TNs bring invaluable practical uses within the realm of Green AI. With the reduction of self imposed calculative strain, they serve as an efficient means of lessening the carbon footprints left by machine learning and artificial intelligence systems. In this segment, we are going to look into particular ways that different AI types can make use of TNs and how they can solve major problems posed by Green AI.

1. Tensor Networks for Deep Learning Optimization

Artificial intelligence favors deep learning as one of the most energy-demanding areas that requires attention. The adoption of large neural networks for image classification, language translation, or even reinforcement learning like most things in AI, necessitates phenomenal computational power during training and deployment. And with the increase in size comes an increase in carbon footprints. The environmental, social and economic impacts of the power consumption and carbon emissions of AI models have become a topic of attention. Some suggest that training a large model is said to result in as much CO2 emissions as multiple cars fault into one device.

The intriguing thing is that where one AI Tensor Networks can drastically benefit the modern day by solving this problem by reducing the number of parameters required by a deep learning algorithm. Taking the example of a classic convolutional neural network (CNN), each convolution layer utilizes a tremendous quantity of weights and filters. These weights can, however, be made more compact using tensor decomposition techniques without losing accuracy, which further leads to reduction in training times, energy, and carbon emissions.

An excellent illustration of how TNs can increase the efficiency of deep learning models is in Tensorized Convolutional Neural Networks (TCNNs). Researchers have applied tensor decomposition to the convolutional layers of a CNN, reducing the number of parameters and computations which decreases the amount of power needed for training and inference. This makes deep learning models not only more sustainable but also more scalable which is needed in resource-capped settings.

This particular installation will explore Reinforcement Learning (RL) and how it can be pushed to be more sustainable with the use of Tensor Networks. Reinforcement Learning (RL) is a type of machine learning in which an agent learns a task by performing actions and evaluating results. There are many tasks to which RL models are applied, including robotic control, game playing, and autonomous driving. These applications are of high importance as well.

In Reinforced Learning, one of the challenges involves harnessing intense computing power for effective strategy exploration and decision-making optimization. Now Tensor Networks will help alleviate the computing burden associated with RL by compressing state-action value functions representing the anticipated reward received for each action in a state. By means of tensor decomposition techniques, TNs will enable the efficient representation of these functions, thus decreasing the compute power needed to train and deploy the models.

As a result of these developments, the amount of computational power and time required to learn is cut down significantly. In addition, there is a decrease in energy consumption compared to conventional RL algorithms. Furthermore, as a testament to their ability to enhance the green credentials of reinforcement learning systems, TNs can be integrated into numerous types of RL algorithms - model free and model based.

Natural Language Processing (NLP) is another field where Tensor Networks may make a remarkable contribution in energy efficiency. Tasks like machine translation, sentiment analysis and text generation have seen a leap in productivity with modern NSP models, especially those based on large transformers, like GPT and BERT. These models, however, are highly resource-consuming at training and runtime.

Incorporating Tensor Networks into NLP architectures provides researchers the potential to achieve sharp declines in computation time and cost. For instance, in transformer models, the ‘attention’ mechanism that allows the model to focus on various aspects of the input sequence may be made more efficient with tensor decomposition. This, will not only reduce the number of parameters but will also reduce the expenses on training and inference because those computations will be faster.

Moreover, TNs can also be used in other places of NLP models, such as in embedding layers and in recurrent networks, to enhance their efficiency. Utilizing the compression capabilities of TNs allows NLP models to handle larger datasets with smaller environments.

Computer vision tasks, such as object detection, image segmentation, and facial recognition are composed of processing dimensional images. Such tasks are achieved using high accuracy deep learning models comprising high-level parameters and layers. With enhancement in model complexity, energy consumption to train and perform inference on the models also increases.

The use of tensor networks significantly decreases the model’s size and complexity which can be used in computer vision tasks. For example, in a Convolutional Neural Network (CNN) image classifier, TNs can be used to compress the convolutional layers which extract features from the input image. Because of this compression, the number of parameters in the network is reduced, thus increasing the efficiency of the model.

In addition, other areas of computer vision such as object detection and image segmentation can benefit from the application of TNs to further optimally utilize resources. These tasks require high computation which can be mitigated with the usage of tensor networks. Reducing the computational load required to implement these tasks enhance the overall sustainability of computer vision models, enabling more of them to be used in energy deficient places like mobile devices or edge computing systems.

Another relevant use for tensor networks in Green AI is model compression and pruning. Many AI systems (primarily deep learning) have models that can be too large and demand an excessive amount of memory and computation resources. This challenge is even greater when instancing a model in an edge device or any other place where resources are scant.

TNs have a number of advantages for AI since they enable the compression of a model by adjusting the number of parameters which can be employed to represent a model or AI with practically no change in the quality of the AI generated. tensor networks also improve how AI models are stored and executed by lowering the memory requirements.

Aside from segmentation, TNs are also useful for cutting down models. This process is pruning, which incorporates the deletion of unnecessary or repetitive weights in a neural network to cut down on parameters and increase the model’s efficiency. Tensor Networks help remove these irrelevant weights more systematically, leading to better model optimization and reduced energy usage.

In Edge Computing and IoT Uses, Green AI

AI model deployment into devices with power restrictions has led to the emergence of smart phones, smart sensors, and edge servers which are all part of the growing internet of Things (IoT) and edge computing fields. These devices tend to not have powerful AI processors leading to local data processing versus computer cloud resource usage. Therefore, the AI models used within these devices should be energy efficient in nature.

Endowing AI models with edge computing features can strongly be supported by Tensor Networks. TNs allow the deployment of complicated AI systems onto these resourceful limited devices without diminishing their performance, all thanks to the limited parameters and computations within the model. It is particularly beneficial in scenarios such as real time video processing, predictive maintenance, and smart home systems where energy efficiency matters a lot.

Thanks to the compression and efficiency provided by Tensor Networks, AI systems operating on IoT devices can use less power which translates to longer battery life, and a lesser environmental footprint. Powered Networks are also easily scalable, thus, they can be employed in a variety of IoT systems from small devices to large industrial systems.

CONCLUSION

But there is more – Tensor Networks (TNs) are now, more than ever, proving to be a valuable resource in the search for more green and energy efficient AI systems. Because of their ability to enhance compression, it becomes possible to build powerful, but effective and accurate AI models that consume less energy and have a lower environmental impact. Their application in deep learning, reinforcement learning, natural language processing, computer vision, and Internet of Things is only an indication of their wide area of usage along with their potential in advancing Green AI.

It is true that the integration of Tensor Networks comes with challenges like data integration and the processing cost. Nevertheless, their potential in energy optimization makes them a welcomed answer for AI’s green development. With the evolvement of the Green AI field, additional research and fine-tuning of the Tensor Networks will be key in accomplishing the objectives of lowing the environmental cost of AI systems versus their needed performance and scalability.

The Tensor Networks’ incorporation in AI models is a step towards the greener AI technologies. The TNs can facilitate the AI sector in delivering systems that enhance technology while ensuring the planet's sustainable future.

REFERENCES

1. Devlin, H. (2023). AI 'could be as transformative as Industrial Revolution'. *The Guardian, St Martin's Press*.

2. Wang, H., Fu, T., Du, Y., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60.

3. Basu, K., Sinha, R., Ong, A., & Basu, T. (2020). Artificial Intelligence: How is It Changing Medical Sciences and Its Future? *Indian J Dermatol*, 65(5), 365–370.

4. Zhang, B., & Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. Available at SSRN 3312874.

5. Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2021). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

6. Dyer, E.S., Lee, J., Kaplan, J.D., Sharma, U., & Bahri, Y. (2021). Explaining Neural Scaling Laws. *arXiv:2102.06701v1*.

7. Shalf, J. (2020). The future of computing beyond Moore's. *LawPhil. Trans. R. Soc. A*, 37820190061.

8. Bollaart, S. (2023). The Environmental Cost of LLMs: A Call for Efficiency. *Medium*.

9. Huang, Z., Damalapati, P.K., & Wu, E. (2023). Data Ambiguity Strikes Back: How Documentation Improves GPT's Text-to-SQL. *arXiv:2310.18742*.

10. Ludvigsen, K.G.A. (2023). The carbon footprint of GPT-4. *Towards Data Science*, 19th July.

11. Ludvigsen, K.G.A. (2023). ChatGPT's Electricity Consumption. *Medium*, 5th March.

12. International Energy Agency (IEA). (2024). Electricity 2024 - Analysis and forecast to 2026.

13. Dempsey, M. (2024). Future data centres may have built-in nuclear reactors. *BBC News*.

14. Schwartz, R., Dodge, J., Smith, N.A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.

15. Sarker, I.H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 16.

16. El Naqa, I., & Murphy, M.J. (2015). What Is Machine Learning? *Springer International Publishing*, 3–11.

17. Vasudevan, R.K., Ziatdinov, M., Vlcek, L., & Kalinin, S.V. (2021). Off-the-shelf deep learning is not enough, and requires parsimony, Bayesianity, and causality. *npj Comput Mater*, 7, 16.

18. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

19. Kaack, L.H., Donti, P.L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning Artificial Intelligence with Climate Change Mitigation. *Nature Climate Change*, 12(6), 518–527.

20. Järvenpää, H., Lago, P., Bogner, J., Lewis, G., Muccini, H., Ozkaya, I. (2024). A Synthesis of Green Architectural Tactics for ML-Enabled Systems. Accepted for publication at the 2024 International Conference on Software Engineering - Software Engineering in Society (ICSE-SEIS'2024).

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

22. Haakman, M., Cruz, L., Huijgens, H., & van Deursen, A. (2021). AI lifecycle models need to be revised. *Empirical Software Engineering*, 26(5), 1–29.

23. Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *Journal of Machine Learning Research*, 21, 248.

24. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the*

Association for Computational Linguistics, pp. 3645-3650, Florence, Italy.

25. Proske, M., Sánchez, D., Clemm, C., Baur, S.-J. (2020). Life cycle assessment of the Fairphone 3.

26. Elster, A.C., & Haugdahl, T.A. (2022). Nvidia Hopper GPU and Grace CPU Highlights. *Computing in Science & Engineering*, 24(2), 95-100.

27. Stobbe, L., Richter, N., Quaeck, M., Knüfermann, K., Druschke, J., Fahland, M., et al. (2023). Environmental technology assessment of mobile communication in Germany (UTAMO). *Umweltbundesamt*.

28. Billaud, M., Clemm, C., Sánchez, D. (2023). ICs as drivers of ICT carbon footprint: an approach to more accurate die size assessment. *Going Green CARE INNOVATION*, Vienna, Austria.

29. Intel. (2024). Intel® Xeon® Platinum 8468 Processor. Available at Intel (Accessed: 13 April 2024).

30. Chen, Z., Wu, M., Chan, A., Li, X., & Ong, Y.-S. (2023). Survey on AI Sustainability: Emerging Trends on Learning Algorithms and Research Challenges. *IEEE Computational Intelligence Magazine*, 18(2), 60-77.

31. Rokh, B., Azarpeyvand, A., & Khanteymoori, A. (2023). A Comprehensive Survey on Model Quantization for Deep Neural Networks in Image Classification. *ACM Trans. Intell. Syst. Technol.*, 14(6), Article 97.

32. Hoefler, T., Alistarh, D., Ben-Nun, T., Dryden, N., Peste, A. (2021). Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks. *J. Mach. Learn. Res.*.

33. Gou, J., Yu, B., Maybank, S.J., et al. (2021). Knowledge Distillation: A Survey. *Int J Comput Vis*, 129, 1789–1819.

34. Xiyu, Y., Liu, T., Wang, X., & Tao, D. (2017). On Compressing Deep Models by Low Rank and Sparse Decomposition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 67-76.

35. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. *arXiv*, abs/2106.09685.

36. Zhou, A., Yao, A., Guo, Y., Xu, L., Chen, Y. (2017). Incremental Network Quantization: Towards Lossless CNNs with Low-Precision Weights. *International Conference on Learning Representations*.

37. Salehi, S., & Schmeink, A. (2023). Data-Centric Green Artificial Intelligence: A Survey. *IEEE Transactions on Artificial Intelligence*, 1(01), 1-18.

38. Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: applications to image and text data. *Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '01)*, 245–250.

39. Wang, W., Huang, Y., Wang, Y., & Wang, L. (2014). Generalized Autoencoder: A Neural Network Framework for Dimensionality Reduction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

40. Abdulkadirov, R., Lyakhov, P., & Nagornov, N. (2023). Survey of Optimization Algorithms in Modern Neural Networks. *Mathematics*, 11, 2466.

41. Li, S., Xue, F., Baranwal, C., Li, Y., & You, Y. (2023). Green Artificial Intelligence: A Green Machine Learning and Artificial Intelligence in the Cloud. *Cloud Computing and Big Data Analysis: Concepts, Tools, and Techniques*.

42. Vinay, P. (2023). Climate Change: Green AI solutions for sustainable future. *Science Advances*, 9(31), eabi7077.