COMPONENT 1

Water quality analysis

1. **Specify the accuracy you achieved across 3 architectural modifications (e.g., different numbers of layers, different hyperparameters, etc.)**

Modification	Activation function	Accuracy (validation)	Loss (Validation)	Accuracy (train)	Loss (train)	Epoch	Drop out	Hidden layers	Time (sec)
Before Tunning	Relu	0.88	1.76	0.85	2.2	10	0.9 & 0.2	2	2.8
Model 1	Sigmoid	0.93	0.19	0.93	0.22	10	0.7 & 0.2	3	3.91
Model 2	Relu	0.95	0.26	0.93	0.29	15	0.2	2	4.87
Model 3	Tanh	0.88	0.21	0.88	0.23	20	0.5	2	5.8

2. Why do you think your accuracy is not higher / lower?

Accuracy couldn't go higher than the observed figures recorded because we are working with a dataset that has few entries as a result the model does not learn enough for it to achieve a higher accuracy. Another reason accuracy couldn't go higher is as a result of low dimensionality. To achieve an accuracy of 98% we will need to normalize the dataset, finetune other hyperparameters such as adding more hidden layers or specify the learning rate in the optimizer. Also, we can achieve a higher accuracy by applying dimensionality reduction methods such as PCA but that is not covered in the scope of this project. In spite of all these, higher accuracy might not be achieved due to dimensionality reduction.

3. **What effect does the optimisation function have on network performance?**

Optimization function	Accuracy (Validation)	Loss (Validation)	Time (sec)
RMSprop	0.96	0.12	2.7
Adam	0.96	0.14	2.8
SGD	0.88	nan	2.5
Adadelta	0.90	0.68	3.1

Optimizers are algorithms that change the weights and learning rate of a neural network thereby minimizing the loss and provide the most accurate results possible. “The goal is to hit the sweet spot of maximum value optimization, where foolish risk is balanced against excessive caution” (Steven, 2017 p. 69). Changing the learning rates of a neural network minimises the loss on the model thereby giving you a higher accuracy for the trained data. From the above table, RMSprop and Adam optimizers gave the highest accuracy with RMSprop giving the lowest possible loss which implies that it has the best learning rate.

4. **What happens if you include more than 4 (hidden) layers?**

Several neurons in the hidden layers can lead to overfitting. This occurs when the network has learned a lot about the training data and as such negatively impacts the accuracy of the test data set. This also increases the training time. When 5 hidden layers was used to train the model, the accuracy dropped and it took a longer time to run. Notice how the model overfits the validation loss in the diagram below.

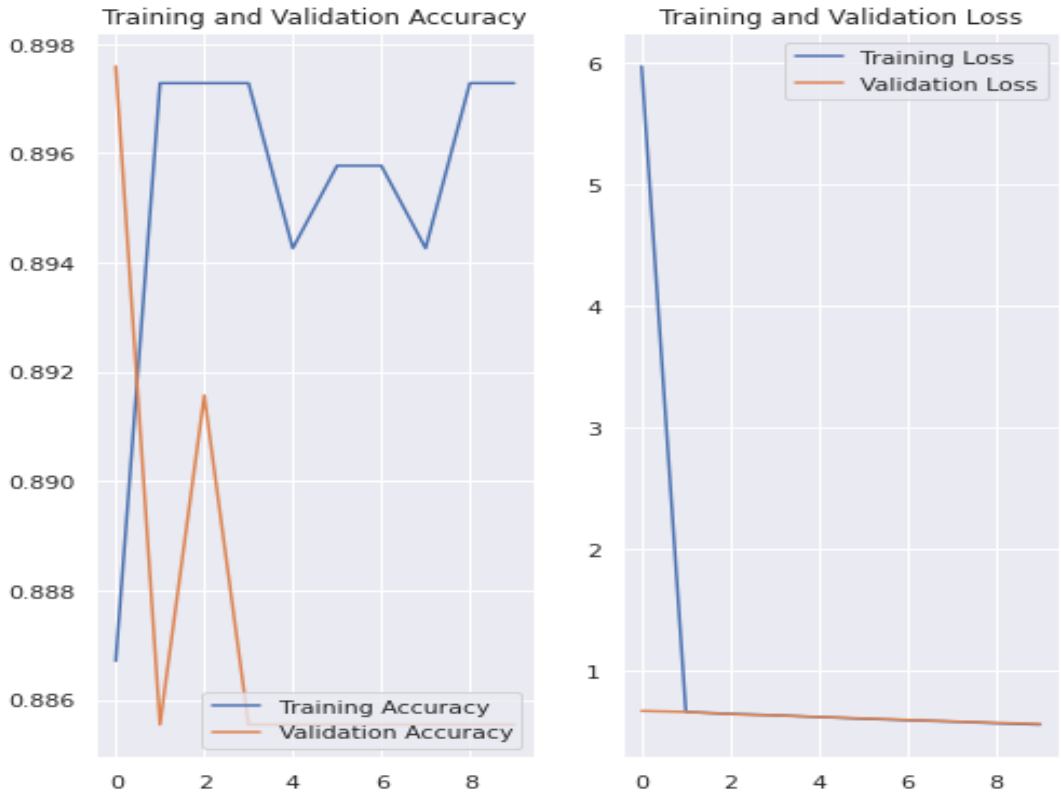

Figure 1: Plot of more than 4 hidden layers

5. What is the effect of the data size on your accuracy?

Small dataset or not enough different image set can result in our model memorising instead learning the dataset.

6. Suitable graphical plot of the data.

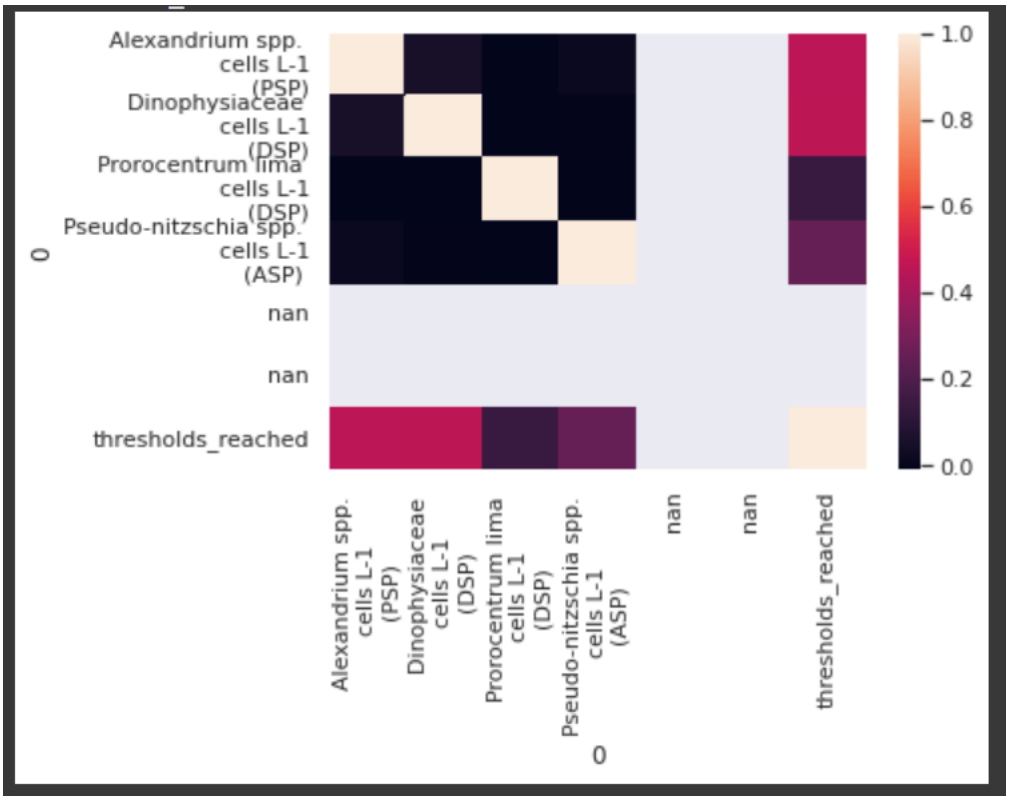

Figure 2: Heat map showing the correlation between the data

COMPONENT 2

Multiple Object Recognizer

Neural Network	Pooling	Accuracy (validation)	Loss (validation)	Accuracy (train)	Loss (train)	Time
Shallow	Max pooling Average pooling	0.56 0.54	1.82 1.72	0.96 0.94	0.11 0.16	35m 11s 33m 16s
Deep	Average pooling	0.60	0.97	0.74	0.62	43m 21s
Augmented	Average pooling	0.81	0.50	0.76	0.63	48m 10s

1. **How long does the network need to train until reaching an accuracy of 95% (or does it not reach this level at all)**
   From the above table, we can see that the highest validation accuracy recorded was 81% after passing the model via 10 epochs. It took the network 58 minutes and 10 seconds to train up to this level. However, the model did train up to 96% accuracy under 45 minutes and 11 seconds using max pooling.

2. **What is the trade-off between using many layers (i.e., having a "deeper" network) and accuracy? And layers and time?**
   The more the layers in a neural network, the more non-linearity it brings to the network. Technically this means, that the network gets powerful enough to learn complex pattern in the data set. As the data is a complex one with a large number of input variables, it is better to increase the number of layers. Conversely, it is advised not to use too many layers in the model as it can consequently lead to bias due to more neurons involved. Bias can occur when there are too many neurons and only a few are associated with heavy weights and biases such that other neurons are never used. This will lead to a low accuracy because the model has not learned properly. The solution to this is to include a drop out layer to allow every neuron have equal weights and biases.
   From the table above, it is observed that having a deeper network increases the time the model will take to train due to the increased number of parameters.

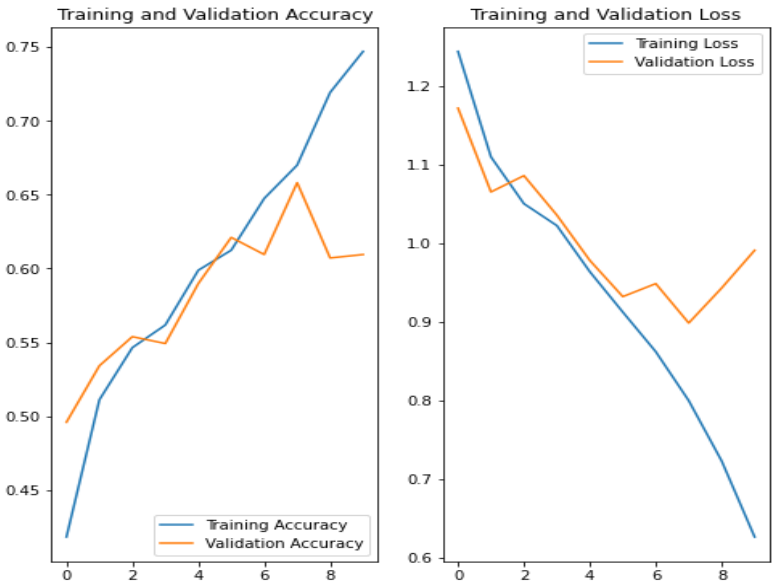

Fig 1: Deeper network

3. **What is the effect of changing the pooling mechanism, e.g., average vs max?**

Pooling is a regularisation technique that improves the performance of CNN while reducing the computation time. A max pooling layer retains the most important features of an image. In order words, it is more focused on the maximum values of the pixels. Average pooling on the other hand retains the average values of features in the image. By changing the pooling method, the image is decreased to a set of features.

Since we are more interested in detecting the overall features of the image rather than whether a specific feature appears or not average pooling did a little better than max pooling. This might be because average pooling prevents the network from learning the image structures such as edges and textures.

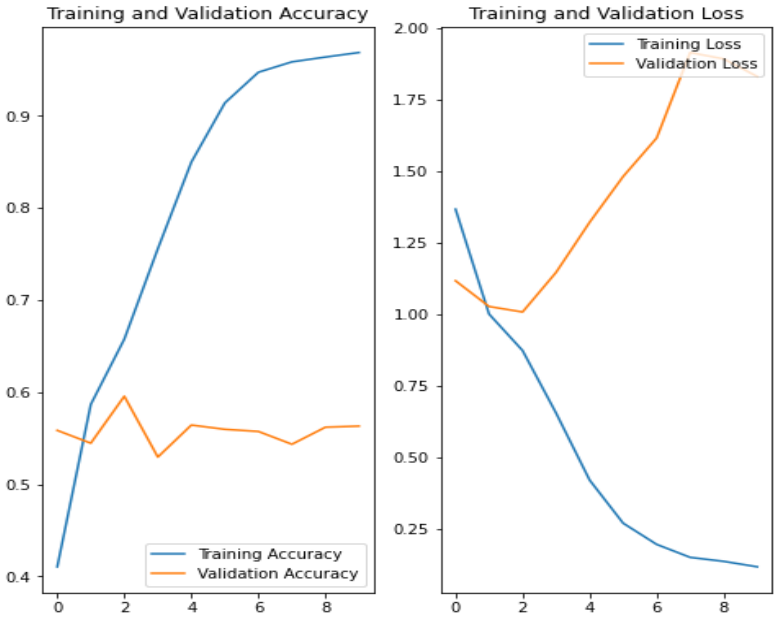

Fig 2: Average pooling

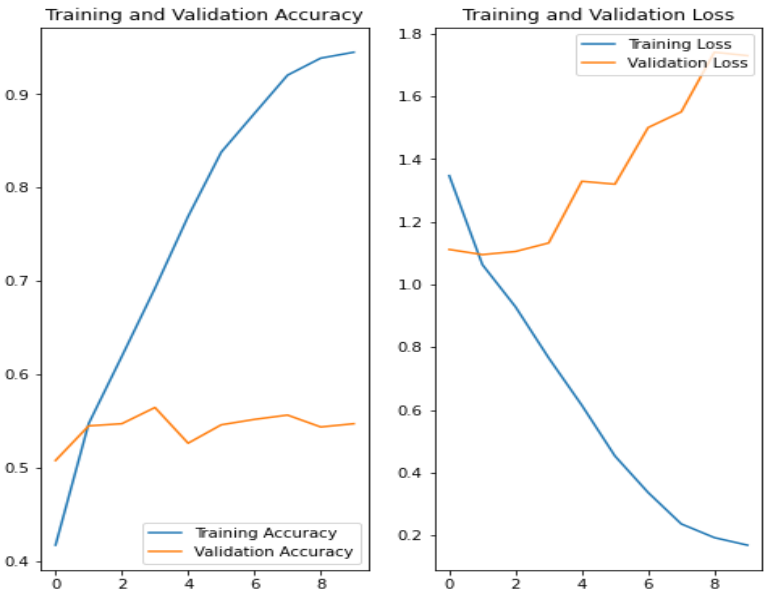

Fig 3: Max pooling

4. **How well does your network do at classifying these images?**

The model did well in classifying the images accurately. Below here is a plot showing how the model did at classifying the images.

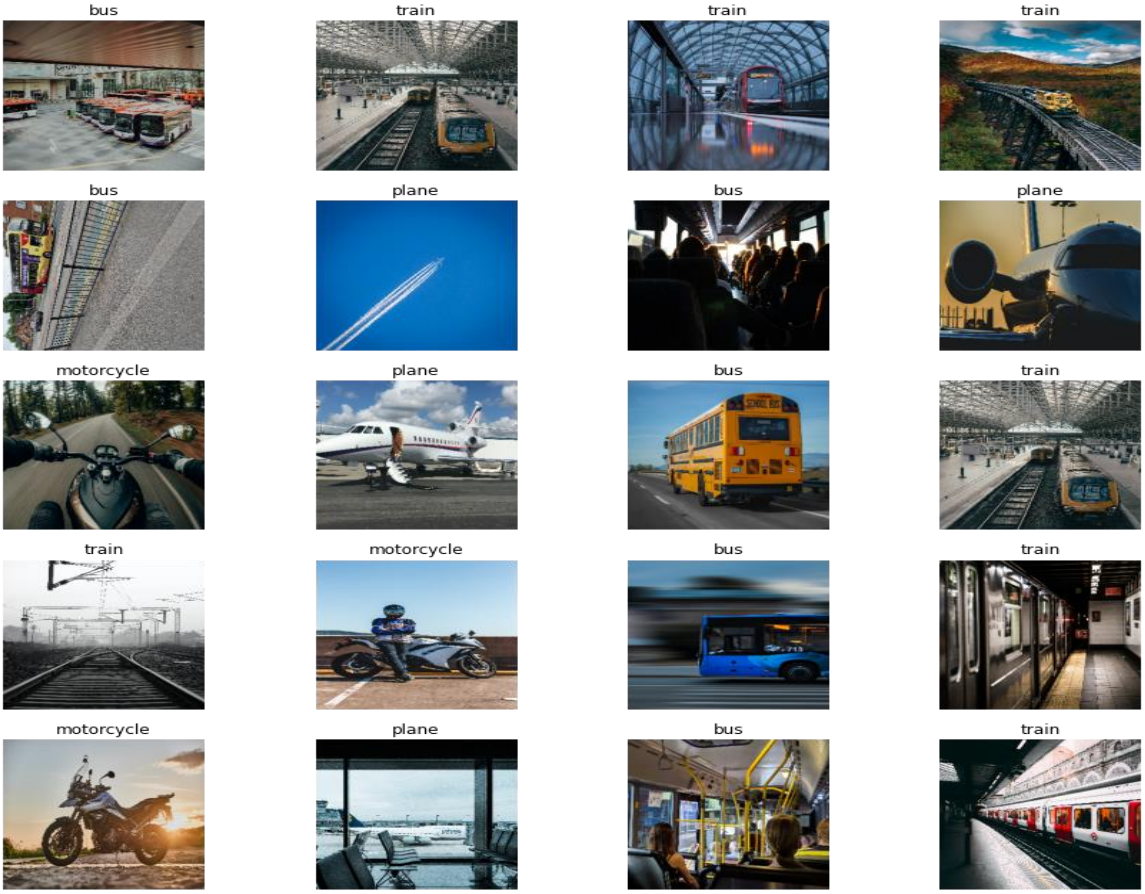

Fig 4: Test images

5. **Does fine-tuning make a difference?**

The model did a nice job at classifying the test images so fine-tuning was not used. Here is a correlation matrix to show how the model did at classifying the images.

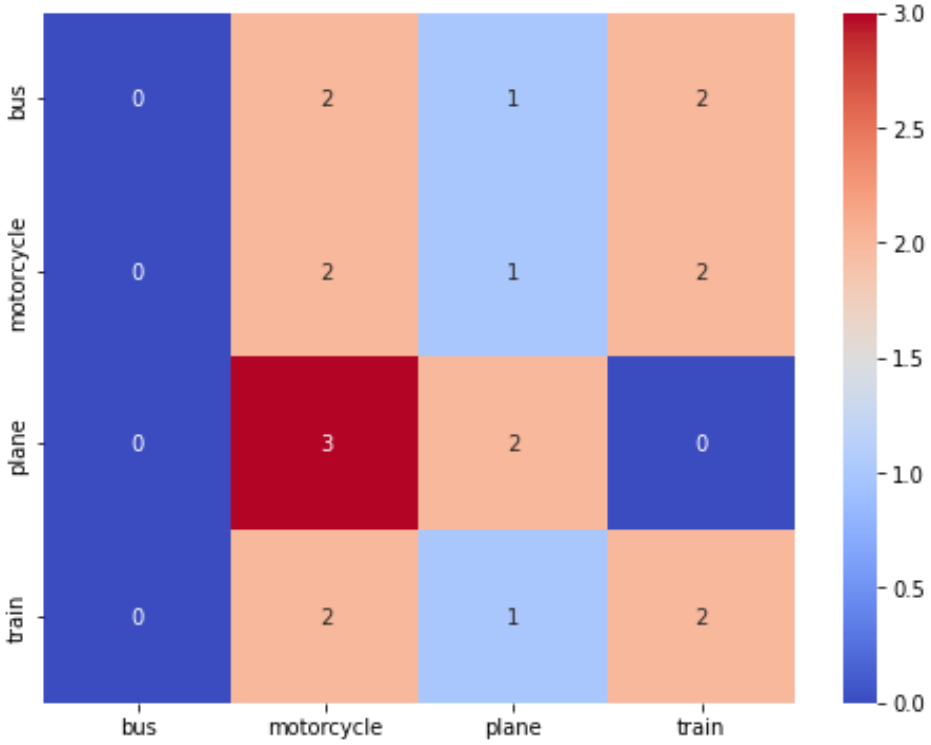

Fig 5: Correlation matrix

Extra Challenge - Model Explainability:

Explainability in machine learning is the process of analysing ML models parameters so as to understand the result produced. This is an important concept with *black box* ML models when dealing with unsupervised learning (Seldon, 2022). The 20 new images collected were passed through tf-explain library to evaluate the interpretability of the model.

The bright yellow spots represent the part of the image that influenced the model's decision to make predictions.

Fig 6: Model Explainability

COMPONENT 3

An overview On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

1. BRIEF HISTORY OF LANGUAGE MODELS (LMS)

Language mode can be defined as a system trained to do string prediction. The statistical approach to string prediction was initially proposed by Dr Claude Shannon in 1949 but has been implemented for speech learning and machine translation in early 80's. In the past, we have seen a pattern of achieving better results through more data and increasing the size of models until scores don't see an improvement. New architectures are implemented that are able to take advantage of the big data we have currently. This has resulted in a change of the type of task these language models are used for.

2. RISKS OF LANGUAGE MODEL

I. Costs

"The average human across the globe responsible for 5tons of CO2 emissions per year" (Strubell et al. 2019). They also looked at the process of training a transformer model and they found out that it would produce 284tons of CO2 emissions. To reduce carbon footprint, the authors gave recommendation to report the time taken to train sensitive hyperparameters and also urged government to invest in cloud computing to provide equitable access to researchers.

Current mitigation efforts

- Renewable energy sources: The world is opting for more cleaner sources of electricity but the disadvantage is that this still incurs a cost on the environment particularly in the form of infrastructure.

- Another mitigation strategy is to prioritize computationally efficient hardware as advocated by organizations such as Sustain NLP workshop. Schwartz et al. (2020) argue for promoting efficiency as the evaluation metric using green AI.

Large language models benefit hegemonic views but marginalised communities are most likely to be negatively impacted by climate change.

II. Large Dataset

Several factors jeopardize internet participation. Think about the following:

Younger people and those from developed countries have more access to the web and thus are contributing more.

Twitter accounts receiving death threats more likely to be suspended due to twitter moderation than those issuing the threats.

What part of the internet are being included in these large datasets? Reddit-US users are 67 percent men and 64 percent are aged between 18 to 29. Wikipedia only 8.8 to 15 percent are women or girls. But not blog sites with fewer traffic.

Who are being filtered out? Some identities are being filtered primarily on target words referencing sex and also LGBTQ online spaces.

The reason this is a problem is that if we overrepresent a large amount of view point, then we are representing the language of people who are deliberately using their language in ways that are consistent with systems of oppression which can be expressed through racism, ageism, transphobia, etc.

Not only does the training data going to overrepresent hegemonic views, but the language model is going to absorb the biases from those views (Blodgett et al. 2020).

III. Synthetic Data

Stochastic

From linguistics and psychology, we learn that human-human interaction is co-constructed and leads to a shared model of the world (Reddy 1979, Clark 1996). In contrast, a language model is a system for haphazardly stitching together linguistic forms from its vast training data, without any reference to meaning: *a stochastic parrot*.

The problem arises because as humans whenever we encounter synthetic text in a language that we are proficient in we make sense of it.

3. POTENTIAL HARMS

- Can be used for denigration, stereotype threat, hate speech: harms to reader, harms to bystanders.
- People can deliberately use these systems to create cheap synthetic text to do harm in the world like boosting extremist recruiting (McGuffie & Newhouse 2020).
- Language model errors attributed to human author in the other language.
- Language models can be probed to replicate personal identifiable information from the training data. (Carlini et al. 2020).
- Language models as hidden components can influence query expansion & results (Noble 2018).

4. CRITICAL REFLECTION

The paper answered questions surrounding if language models can be too big but does not give us an insight whether ever larger language models is inevitable or necessary?

Furthermore, the paper gives us more insight on the associated cost with this research direction but does not tell us what we should consider before pursuing it?

Lastly, it fails to inform us if NLP needs larger language models? Or how we can pursue mitigate the associated risks in this research direction?

Further research on the ethical and social risks of harm from language models (Weidinger, et al. 2021) throws more light on these questions but cannot be answered within the scope of this essay.

References

Bender, E. M., Gebru, T., McMillan-Major, A., et al. (2021). *On the dangers of stochastic parrots: Can language models be too big?* In Proceedings of FAccT 2021.

Claude Elwood Shannon (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Emma Strubell, Ananya Ganesh, and Andrew McCallum (2019). *Energy and Policy Considerations for Deep Learning in NLP*. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 3645–3650.

Herbert H. Clark (1996). *Using Language*. Cambridge University Press, Cambridge.

Kris McGuffie and Alex Newhouse (2020). *The Radicalization Risks of GPT-3 and Advanced Neural Language Models*. Middlebury Institute of International Studies at Monterrey. https://www.middlebury.edu/institute/sites/www.middlebury.edu.institute/files/2020-09/gpt3-article.pdf.

Nicholas Carlini, Florian Tramer, Eric Wallace, et al. (2020). *Extracting Training Data from Large Language Models*. arXiv:2012.07805 [cs.CR].

Roy Schwartz, Jesse Dodge, Noah A. Smith, et al. (2020). *Green AI*. ACM 63, 12 (Nov. 2020), 54–63. https://doi.org/10.1145/3381831

Safiya Umoja Noble (2018). *Algorithms of Oppression*: How Search Engines Reinforce Racism.

Seldon, (2021). *Explainability in machine learning.* Available at: https://www.seldon.io/explainability-in-machine-learning (Accessed at: 10th May 2022).

Steven J. Bowen (2017). *Total value optimization.* SJDB LLC.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, et al. (2020). *Language (Technology) is Power: A Critical Survey of bias in NLP*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, 5454–5476. https: //doi.org/10.18653/v1/2020.acl-main.485.

Weidinger, L., Mellor, J., Rauh, M., et al. (2021). *Ethical and social risks of harm from language models*. https://arxiv.org/abs/2112.04359.