

# Unraveling LLM Efficacy: From Alpha to Effective Rank and Beyond

Iqbal Singh  
220231857  
Haim Dubossarsky  
Msc.AI

**Abstract**— Evaluating the performance of pretrained deep learning models, particularly when devoid of training and testing data, remains an intricate challenge in the realm of machine learning. A noteworthy approach by Martin, Peng, and Mahoney titled "Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data" employed weight matrix properties and Power Law (PL) fits using the parameter alpha to predict the quality of these models. Their exploration, however, spanned a limited set of architectures, prompting a broader investigation. In this study, we extend the discourse to 39 state-of-the-art Large Language Models (LLMs) and introduce one major advancement: the reciprocal of the effective rank as a performance prediction metric. This innovation is designed to offer superior accuracy and robustness in model evaluation. Our extensive evaluation reveals that this approach not only outperforms the alpha-based PL predictor but also demonstrates reliability across diverse architectures. This contribution is paramount in advancing our comprehension of the inherent properties governing LLM performance and offers a pathway for the effective deployment of pretrained models in scenarios devoid of training or testing data.

**Keywords**—LLM, power-law, effective rank

## I. INTRODUCTION

The utilization of pretrained Large Language Models (LLMs) has become a widespread practice in various applications. These pretrained LLMs offer tremendous benefits in terms of time and resource efficiency. However, a-priori choosing the right model for a particular task is very challenging because evaluating their expected performance without access to the specific training data, loss functions, or hyperparameter values can be a complex task. How does one gauge the quality of a model when the details of its expected performance are unknown?

LLMs contain numerous parameters which determine how input data is transformed and processed through the neural network to generate respective outputs (e.g., how pixels of an image are processed to determine if it is a cat in the image). These parameters are learnt during the training phase, where the model refines its parameters to enhance performance. Analyzing the inherent properties of these parameters can provide insights into the model's workings and potential efficacy. Importantly, these parameters are organized and stored as numerical matrices, which lend themselves easily into functional analysis using mathematical tools. It is therefore assumed that the composition and structure of these matrices would reflect key traits that the network has learnt and picked up from the training data.

One of the foundational analyses of matrices is Principal Component Analysis (PCA). PCA is a technique used to emphasize variation and capture strong patterns in a

data set. It essentially distills the data into its principal components, which can then be analyzed to better understand the underlying patterns and relationships. Critical to the process of PCA are eigenvalues, special numerical values that indicate the amount of variance captured by each principal component. In layman's terms, eigenvalues serve as indicators of the "importance" of different features in the data set. When applied to the weight matrices of neural networks, the analysis of eigenvalues through PCA can potentially unveil patterns or structures that are pivotal to the model's performance.

Building upon this concept of analysing numerical matrices, Mahoney et al.[2021]<sup>1</sup> introduced an innovative approach. They focused on the distribution of eigenvalues within these matrices and applied what is known as a "power-law" to this distribution. In simple terms, a power-law is a relationship between two quantities such that one is a constant multiple of the other raised to a fixed power known as alpha. Mahoney et al found that "alpha," emerging from this power-law relationship, could act as a strong predictor of a neural network's performance. While "alpha" offers one insightful metric, there remain other unexplored matrix properties that could provide valuable information, and our research aims to delve into these areas.

It's crucial to note that Mahoney's research reported that the alpha measure is the best predictor for a model's performance. Their analysis was also predicated on a single performance measure for a model. However, in the case of LLM, a single model can handle multiple tasks like summarization, logical reasoning, and inference. It's important to consider all these aspects when evaluating them.

Our research continues their investigation by trying to understand the nuances of LLMs via their matrix properties that aren't explored yet and that can outperform the measure as suggested by previous study. Our hypothesis was that there could be better ways to describe how the network behaves, all while still focusing on the analysis of the weight matrices. Specifically, we questioned whether there could be a matrix property that is more telling than the "alpha" used by Mahoney.

Specifically, we replicated Mahoney's method of studying LLMs (alpha) and discovered a new and improved measurement of LLMs. We focused on 39 LLMs sourced from the Hugging Face open LLM dashboard.

In this study, we analysed performance on four key tasks:

1. AI2 Reasoning Challenge: This gauges a model's ability in answering elementary-school science questions, probing its scientific knowledge base.
2. HellaSwag: A distinct test of commonsense inference. While it's relatively straightforward for

humans, achieving around 95% accuracy, it proves to be a challenge for even the state-of-the-art models.

3. MMLU: A comprehensive assessment, this test evaluates a text model's multitask accuracy across a broad spectrum of 57 tasks, ranging from elementary mathematics and US history to computer science and law.
4. TruthfulQA: A unique benchmark, its objective is to measure a model's inclination to parrot falsehoods prevalent on the internet.

One significant observation was the evident superiority of the Effective Rank (ER) based metrics over the traditionally used metric - Alpha as proposed by Mahoney. The ER-based metrics often displayed stronger correlations and higher R-square values in regression analyses with task performance than Alpha. This emphasizes the potential of ER as a predictor, which might be tapping into an inherent property of the neural network's weight matrices that directly affects model performance.

## II. RELATED WORK

There's interest in judging a model's quality based on its features, without specific tests. However, not much work has been done in this field, especially for LLMs. Most of the research is still in its early stages, apart from a few studies like the one by Mohny.

1. Martin et al. (2021) took a deep dive into this issue, offering a comprehensive meta-analysis spanning numerous pretrained models accessible to the public. Their study brought to the forefront not just norm-based capacity control metrics but also showcased metrics rooted in the Theory of Heavy-Tailed Self Regularization, with a power law focus. Their groundbreaking revelation suggested that the 'alpha' parameter might hold the key to foreseeing a model's generalization capabilities.

2. Venturing into the territory of convolutional neural networks (ConvNets), Tan's work in 2019, titled "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," unpacks the intricate bond between a model's parameter count and its efficacy. This exploration underscored the notion that judiciously ramping up the parameter count could elevate a Neural Network's accuracy benchmarks.

## III. METHODOLOGY

### i. Model Selection:

For our initial study, we used 14 models from the Weighwatcher website<sup>3</sup> by Mohny. Later, we expanded our research. For this expanded study, we adopted a stratified random sampling approach to ensure a comprehensive representation of LLMs. Our selected models span a wide parameter range, with the largest boasting 33 billion parameters and the smallest having 760 million.

To guarantee diverse performance benchmarks, we chose models from the top, mid, and bottom performance tiers. This stratification ensures our analysis encapsulates the holistic performance spectrum of LLMs. It's essential to note that our analysis was inclusive; we incorporated models regardless of their training methods, be it pretraining, fine-tuning for tasks

other than the primary four under study, instruction-based, or RL-tuned.

### ii. Matrices analysis and Model attributes

#### a. Matrices Analysis

In understanding the intricacies of LLMs, it's crucial to break down their foundational structures. LLMs are built on a sequence of layers, with each layer composed of a set of weights, often termed as 'parameters' and is defined as a matrix. These parameters play a pivotal role in guiding the LLM's decision-making processes, here terms parameters is analogues to weight matrix of neural nets.

Given that these parameters are structured as matrices, we can use common matrix analysis methods to explore them. By using these methods, we can better understand the characteristics of the LLM and uncover subtle details for every weight matrix in each layer of the given model, we calculated statistics to gain insight into the model's inherent characteristics. We derived these statistics from the LLM using Mahoney et al.'s WW tool<sup>4</sup>.

The Statistics /Attributes used in this study were:

- **Alpha ( $\alpha$ ):** Derived from the Theory of Heavy-Tailed Self Regularization, alpha offers insights into the weight matrices of neural networks. Specifically,  $\alpha$  is deduced from the eigenvalue distribution of these weight matrices. When eigenvalues are plotted on a log-log scale, the slope corresponds to  $\alpha$ . A steeper slope typically signifies a higher alpha value, which can offer insights into the network's characteristics.

When plotting the eigenvalues, say  $x$ , against their frequency,  $y$ , on a log-log scale, they often show a linear relationship, indicative of a power-law distribution:

$$y = ax^\alpha$$

Taking the logarithm of both sides:

$$\log(y) = \log(a) + \alpha \log(x)$$

Here, when we plot  $\log(x)$  (log of eigenvalues) against  $\log(y)$  (log of their frequency), the slope of the resulting line represents  $\alpha$ . The steeper the slope, the higher the  $\alpha$  value, which gives insights into the network's characteristics.

- **Condition Number:** This is defined as the ratio of the largest to the smallest eigenvalue of a matrix. It provides an indication of the matrix's sensitivity to changes or errors.

The conditional number  $\kappa$  of a matrix  $A$  is given by:

$$\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$$

Where:

$\lambda_{\max}$  is the largest eigenvalue of the matrix  $A$

$\lambda_{\min}$  is the smallest eigenvalue of the matrix  $A$

When the condition number is large, the matrix is considered to be ill-conditioned, implying it may be sensitive to small changes or errors.

- **Entropy:** It is a concept that originates from information theory and is used to quantify the amount of unpredictability or randomness in a set of data. In the context of matrices, we're interested in how the singular values are distributed. When we decompose a matrix  $A$  using singular value decomposition, we get singular values that are essentially the diagonal elements of one of the resulting matrices. These singular values can be thought of as the "weights" or "importance" of certain features captured by the matrix  $A$ .

To understand the distribution of these singular values, we first normalize them. Given singular values  $\sigma_1, \sigma_2, \dots, \sigma_n$ , the normalized value for  $\sigma_i$  (denoted as  $p_i$ ) is found by dividing  $\sigma_i$  by the total sum of all singular values, as per the formula:

$$p_i = \frac{\sigma_i}{\sum_j \sigma_j}$$

This way, each  $p_i$  represents the proportion of  $\sigma_i$  in relation to the sum of all singular values. These normalized values lie between 0 and 1, and their total sum equals 1.

To measure the randomness or unpredictability in the distribution of these normalized singular values, we compute the **entropy** using the Shannon entropy formula:

$$H = - \sum_i p_i \log(p_i)$$

A higher entropy indicates a more uniform distribution of the singular values, implying that the matrix has a balanced "importance" across its features. In contrast, a lower entropy suggests that only a few singular values dominate, indicating certain features in the matrix have significantly higher "weights" or "importance" than others.

- **Effective rank:** calculated by exponentiating the entropy of the normalized singular values of the matrix.  

$$E_{\text{rank}}(X) = \exp(H)$$

The effective rank gives us a sense of how many "important" features or patterns a matrix has. Effective rank provides a way to measure the effective dimensionality or information content of a matrix. It's different from the mathematical rank, which simply counts the number of non-zero singular values. Instead, the effective rank, by taking into account the entropy of normalized singular values, provides an indication of how much each singular value contributes to the matrix. This becomes crucial for understanding matrices in LLMs because it can hint at the redundancy or information diversity in the weight matrices.

- **Modified Conditional Number:** Traditional condition number,  $\kappa(A)$ , of a matrix  $A$  gives an indication of how the output (solution) changes with respect to small changes in the input. Mathematically, it's the ratio of the largest eigenvalue to the smallest eigenvalue.

However, when considering a "Modified Conditional Number", it takes into account the "effective rank" of the matrix. The effective rank provides a measure that takes into account the distribution of the singular values themselves. This measure gives an intuition about the "effective dimensionality" of the data in the matrix.

Mathematically, if we define:

- $\lambda_{\max}$  as the largest eigenvalue of matrix  $A$ ,
- ER as the effective rank of matrix  $A$ ,

The formula for the Modified Conditional Number ' $\kappa'$ ' is:

$$\kappa'(A) = \frac{\lambda_{\max}}{\text{ER}}$$

This modified measure provides a normalized way to understand the matrix's sensitivity to changes. Instead of just considering the range of the eigenvalues (like in the traditional condition number), it looks at the sensitivity in the context of the effective dimensionality of the matrix.

- **Spectral Norm:** For a given matrix  $A$ , the spectral norm, denoted  $\|A\|_2$ , is defined as the largest singular value of  $A$ .

$$\|A\|_2 = \sigma_{\max(A)}$$

#### b. Model attributes

For every layer of the LLM, statistics mentioned above were computed. These statistics were then aggregated to characterize the model on a holistic level. In total, 44 model-level statistics were derived, with a detailed breakdown available in the appendix. These attributes can be classified into four main classes:

- Alpha-based
- Spectral Norm-based
- Effective Rank-based
- Modified Conditional Number-based
- Number of model's parameters

For each of the above attributes descriptive statistics such as maximum, minimum, range, standard deviation, average, log-transformed values were calculated to describe a model. This approach ensured a comprehensive representation of the model's intrinsic characteristics.

#### iii. Correlation Analysis

After deriving the model attributes, we tested the relationship between these attributes and the task performance of each model. This process was important in uncovering which intrinsic properties of the LLMs bore direct impact on their performance outcomes. By employing statistical measures like the Pearson correlation coefficient, we quantified the strength and direction of linear relationships between our derived attributes and the observed model performance. Key insights from this analysis, showcasing which intrinsic characteristics profoundly influence LLM performance, are elaborated in the ensuing sections.

#### iv. Regression Analysis

To understand how each feature affects model performance, we did a regression analysis for all models on our four tasks. In this analysis, we always included 'no. params' (as a control

variable )because it has a big effect on how models perform. We then added each feature one by one with 'no. params' to see how much more they can explain about performance. By doing this, we could see the separate impact of each feature when considered along with the 'no. params'. We used  $R^2$  to see how well our features matched with the actual performance across different tasks.

#### v. Pairwise difference analysis

To delve deeper into the intricacies between model attributes and performance, we executed a pairwise difference analysis. In this approach, for given model attribute measures and their respective performance measures, we calculated the simple difference for all possible combinations of data points. This meant, for example, if two models displayed a particular difference in an attribute like their effective ranks, we would investigate: Does this attribute difference correspond to a noticeable difference in their actual task performance?

The merits of this approach are twofold:

1. **Enhancement of Data Points:** By calculating differences across all model pairs, our data points expanded almost tenfold, substantially bolstering the robustness of our analysis.
2. **Analysis of Derivative Correlations:** Studying differences is akin to exploring the correlation of the derivative. While the original data's analysis helps establish connections between variables, inspecting the correlation of differences lets us infer links between shifts in variables and associated performance changes.

In our study, we closely examined 741 unique model pairs for each of the four tasks, resulting in 2,964 pairwise evaluations. To further validate the consistency and strength of relationships between the model ranks, based on attributes and performance, we utilized Kendall's Tau, Spearman Rank Correlation, and R-Squared.

- **Kendall's Tau:** This metric is particularly well-suited for our study as it is designed to measure the correlation between rankings, making it perfect for comparing our model ranks.
- **Spearman's Rank Correlation:** Unlike Kendall's Tau, Spearman is less sensitive to outliers. This offers another robust avenue to validate our findings by focusing on the strength and direction of the rank-order relationship between model attributes and performance.
- **R-Squared:** The R-Squared value adds another dimension to our analysis by quantifying how well differences in our model attributes can explain differences in performance. Essentially, a high R-Squared value would mean that most of the variability in performance can be attributed to variability in the measured attributes.

**Note:** The values used in our analysis for pairwise difference have been log-transformed for clarity and ease of interpretation.

By employing these statistical methods, we not only affirm the robustness of our pairwise difference analysis but also offer an insight into the relationships between model attributes and performance.

## V. RESULTS

Results highlights:

1. **Validation of Previous Findings:** We successfully replicated the alpha scores from earlier studies, ensuring our methods align with established benchmarks.
2. **Performance Tied to Parameter Count:** A strong correlation was identified between the number of parameters a model has and performance. Notably, this factor consistently outperformed the alpha scores in predicting model efficiency.
3. **Introduction of New Predictors:** Our research introduces fresh predictors of task performance derived from varied matrix analysis techniques.
4. **Superiority Over Alpha:** These newly proposed predictors consistently outshine the alpha scores. Especially worth mentioning the Effective Rank measure and its derivatives, which often surpasses even the number of parameters in predictive accuracy.

With these highlights in mind, we'll now delve deeper into the empirical evidence supporting these findings.

#### a. Simple correlation analysis

The table shows the correlation coefficients between various model attributes and their respective performances across four distinct tasks: ARC, HELLASWAG, MMLU, and TRUTHFULQA.

TABLE I. CORRELATION ANALYSIS RESULT

	Performance Tasks			
	ARC	HELLASWAG	MMLU	TRUTHFULQA
No.of Parameters	0.75	0.82	0.61	0.06
ALPHA	0.23	0.19	0.34	0.23
1/ER_AVG	0.57	0.57	0.57	0.24
1/ER_MED	0.58	0.54	0.63	0.33
1/ER_MIN	0.78	0.78	0.67	0.22
1/ER_MAX	-0.25	-0.02	-0.35	-0.4
ER_Range	0.43	0.23	0.48	0.38

Fig. 1. Correlation Coefficients of Model Attributes with Performance Across Four Tasks

Upon analysis, the reciprocal values of the effective rank-based attributes clearly outperform ALPHA, especially in the case of the median reciprocal. This highlights the potential superiority of effective rank metrics in predicting model performance. The 'No. of Parameters' exhibits a universally positive correlation across tasks, underscoring its substantial influence on performance.

In summation, effective rank attributes, especially the median reciprocal, emerge as strongest predictors of task performance, overshadowing the ALPHA metric.

#### b. Regression Analysis

The tables illustrate the  $R^2$  values derived from regression analyses involving 'no. of param' combined with each attribute, taken one at a time, when benchmarked against four distinct performance measures: ARC, HELLASWAG, MMLU, and TRUTHFULQA. This  $R^2$  value indicates how much of the variability in our dependent variable (performance measure) can be explained by the combined effect of 'no. of param' and the individual attribute. By incorporating 'no. of param' in every analysis, we account for its foundational influence, allowing us to evaluate the unique contribution of each attribute in predicting task performance. For a comprehensive understanding, the  $R^2$  value of 'no. of param' in isolation is also presented to highlight its individual predictive power.

	ARC	HELLASWAG	MMLU	TRUTHFULQA
No. of Params(standalone)	0.55	0.67	0.35	0.03
1/ER_MED	0.64	0.69	0.49	0.13
1/ER_MIN	0.63	0.68	0.45	0.09
1/ER_MAX	0.63	0.69	0.50	0.17
1/ER_AVG	0.61	0.70	0.45	0.05
Alpha_mean	0.59	0.69	0.45	0.05

Fig 2: R-Squared of Model Attributes Across Four Tasks

A close examination reveals several compelling trends:

- Effective Rank (ER) Measures:** Across various tasks, the attributes based on effective rank, such as 1/ER\_MAX, 1/ER\_AVG, 1/ER\_MED, and others, show substantial R-Squared values. This highlights the potency of effective rank measures in explaining the performance variance, outperforming the traditional alpha metric.
- Alpha:** While Alpha\_mean is considered, it ranks relatively lower compared to ER-based measures, especially in the TRUTHFULQA task, where it demonstrates a particularly weak correlation.

In conclusion, the regression analysis firmly corroborates the effectiveness of the ER-based measures as substantial predictors of task performance, outstripping the alpha metric across different tasks.

#### c. Pairwise Difference Analysis

In the pairwise difference analysis, visualizations distinctly illustrate a more pronounced association between Effective Rank (ER) based attributes and task performance, when compared with the traditionally leveraged Alpha. For each task, the scatter plots reveal a clear trend where 1/ER\_MED, as a representative metric, holds consistent predictive potency.

Furthermore, a deeper dive into the  $R^2$ , Kendall, and Spearman correlation values reinforces this observation. While Alpha has its merits, within the context of our research, the ER-based

measures—especially 1/ER\_MED—emerge as more salient predictors of model success. For a more comprehensive view, plots representing other ER-based metrics, such as 1/ER\_MIN, 1/ER\_MAX, and 1/ER\_AVG, have been furnished in the appendix, which echo the same sentiment of ER's superior predictive capability.

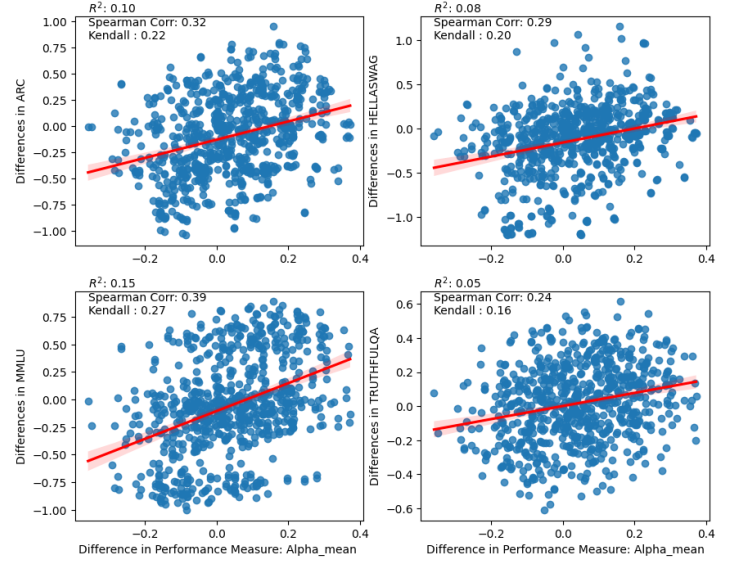


Fig3. Pairwise Difference: Performance Metrics vs. Alpha Descriptor.

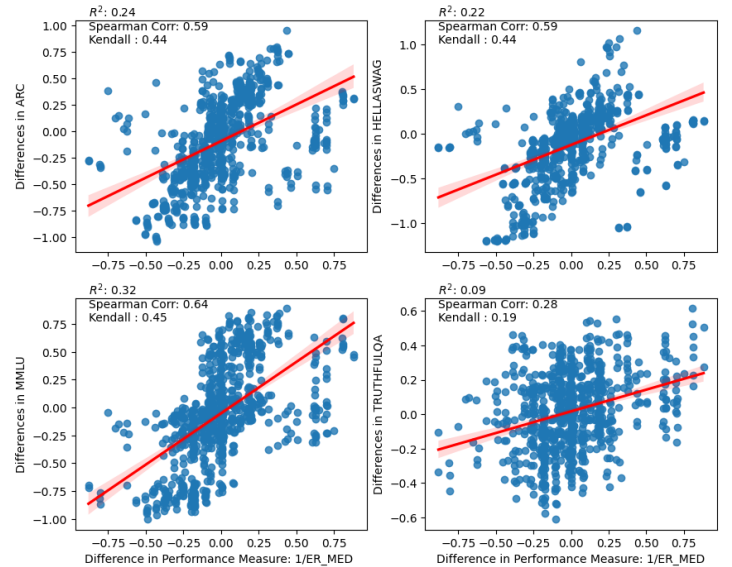


Fig4. Pairwise Difference: Performance Metrics vs. 1/ED\_MED

#### Graph for Alpha: Fig 3

The graph plotting Alpha against evaluation metrics like Kendall's Tau, R-Square, and Spearman shows moderate relationships across all four tasks: ARC, HELLASWAG, MMLU, and TRUTHFULQA. For example, the Kendall's Tau values range from 0.16 to 0.27, suggesting that while Alpha



may offer some predictive value, it is not a strongly dominant factor in task performance.

#### Graph for Effective Rank (ER): Fig4

In contrast, the graph plotting Effective Rank (ER) paints a different picture. Across all tasks, ER demonstrates a stronger relationship with performance. The Kendall's Tau values vary between 0.19 and 0.45, and R-Square values are consistently higher compared to those for Alpha, indicating that Effective Rank is a more robust predictor for model performance in these tasks.

These graphs offer compelling evidence that while Alpha has some predictive value, Effective Rank generally provides a more reliable metric for performance evaluation.

#### IV. DISCUSSION/CONCLUSION

##### Discussion:

The results from our research offer compelling evidence on the potential predictors of neural network performance, especially within the context of specific tasks like ARC, HELLASWAG, MMLU, and TRUTHFULQA.

One significant observation was the evident superiority of the Effective Rank (ER) based metrics, particularly when they were inverted (reciprocal), over the traditionally used metric - Alpha as proposed by Mahoney. The ER-based metrics often displayed stronger correlations and higher R-square values in regression analyses with task performance than Alpha. This emphasizes the potential of ER as a predictor, which might be tapping into an inherent property of the neural network's weight matrices that directly affects model performance.

##### Conclusion:

In the vast field of neural network studies, our research was able to provide compelling evidence of a novel metric/measure/descriptor of LLMs. We've spotlighted features that previous study had missed. Not much work has been done in this direction, apart from some research by Mahoney. We discovered that 'Effective Rank' is crucial, and our unique score offers a fresh perspective on how to judge models. Crucially, our findings are robust as we obtained converging evidence through a series of independent analyses that provided consistent results.

Crucially, our study also taps into the essence of model evaluation that is agnostic to testing data, underscoring the significance of intrinsic metrics over extrinsic performance alone. This orientation provides a refreshing lens through which we can perceive and evaluate the potential of neural architectures.

It is our aspiration that this work not only informs the immediate community but also acts as an inspiration for subsequent explorations. The world of neural networks is vast and varied, and it's through collaborations - both human and machine - that we'll continue to unravel its mysteries.

#### V. FUTURE WORK

Our research has started to scratch the surface in understanding the complexities of LLMs through their matrix properties. While promising, there's still much more to explore. Here are some avenues for future research:

1. Non-linear Combinations: We've focused on straightforward characteristics of the models. Future studies could examine more complex combinations of these attributes, perhaps utilizing polynomial combinations or kernel methods to delve into non-linear correlations between metrics and performance.
2. Specific Model Selection: We included all models available to us, even those not fine-tuned for our tasks. Going forward, research should focus on models tailored to specific tasks, or ones using novel training methods like instruction-based learning or RL tuning.
3. Epoch-by-Epoch Analysis: Understanding how these matrix properties evolve over time during training, on an epoch-by-epoch basis, could provide additional insights into the learning dynamics and their relation to final model performance.
4. Mathematical Proofs: To move from empirical observations to a solid theoretical framework, future work could aim to provide mathematical proofs that strengthen the theoretical foundations of our findings.
5. Transfer Learning Study: An interesting avenue is to investigate how these mathematical properties, such as Effective Rank, behave when a model is fine-tuned for different tasks—a common practice in deploying LLMs.
6. Expanded Dataset and Performance Tasks: Increasing the number and diversity of tasks for performance evaluation can add more robustness and generalizability to our findings.

These suggested directions not only serve to fortify the credibility and applicability of our work but also invite other researchers to engage in this burgeoning field of study.

#### VI. REFERENCES

- [1] Martin, C. H., Peng, T., & Mahoney, M. W. (2021). Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data. *Nature Communications*, 12(1), 1–13. <https://doi.org/10.1038/s41467-021-24025-8>
- [2] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional Neural Networks. In arXiv [cs.LG]. <http://arxiv.org/abs/1905.11946>
- [3] WeightWatcher: Data-free diagnostics for deep learning (no date) Weightwatcher.ai. Available at <https://weightwatcher.ai/leaderboard.html>
- [4] Calculation Consulting WeightWatcher: Data-free diagnostics for Deep Learning, Weightwatcher.ai. Available at: <https://weightwatcher.ai/>

#### APPENDIX:

Sr.no	Model Attribute	Description
1	Alpha_max	Maximum of the alpha observed across all the layers
2	Alpha_min	Minimum of the alpha observed across all the layers
3	Spectral_Norm_Max	Maximum of the spectral norm observed for output dense layers
4	log_spectral_norm_max	Log of the Spectral_Norm_Max
5	Spectral_Norm_Min	Minimum of the spectral norm observed of all output dense layers
6	Max_Spectral_Range	Maximum of the spectral range observed of all output dense layers
7	Spectral_Range_out	spectral range of the from the max-min observed spectral values.
8	SN_MAX	Maximum of the spectral Norm observed across the layers
9	SN_MIN	Minimum of the spectral Norm observed across the layers
10	SN_range	Range of the spectral norm
11	SN_MEDIAN	Median of spectral norm observed across all the layers
12	SN_Range_Log	Log of the spectral norm range
13	SN_MEDIAN_LOG	Log of the median of spectral norm
14	SR_MEDIAN	Median of spectral range observed across all the layers
15	LOG_SR_MEDIAN	Log of the median value of spectral range
16	Std_Rng_spec	Standard deviation of spectral norm
17	Avg_Cond_Num	Average of the conditional number observed across all the layers
18	STD_Cond_Num	Standard deviation of the conditional number
19	MCN_MAX	Maximum of the modified conditional number
20	MCN_MIN	Minimum of the modified conditional number
21	MCN_AVG	Average of the modified conditional number
22	MCN_RANGE	Range of the modified conditional number
23	MCN_STD	Standard deviation of the modified conditional number
24	MCN_MEDIAN	Median of the modified conditional number
25	ER_MAX	Maximum effective range observed across all the layers
26	ER_MIN	Minimum effective range observed across all the layers
27	ER_RANGE	Range of effective range across all layers(max- min value)
28	ER_AVG	Average of the effective range observed across all the layers
29	ER_STD	Standard deviation of the effective range
30	ER_MEDIAN	Median of the effective range across all the layers
31	1/ER_AVG	Reciprocal of average of effective range
32	1/ER_MED	Reciprocal of median of effective range
33	1/ER_MIN	Reciprocal of minimum of effective range
34	1/ER_MAX	Reciprocal of maximum of effective range
35	LOG_MCN_MAX	Log of MCN_MAX
36	LOG_MCN_MIN	Log of MCN_MIN
37	LOG_MCN_AVG	Log of MCN_AVG
38	LOG_MCN_RANGE	Log of MCN_RANGE
39	LOG_MCN_STD	Log of MCN_STD
40	LOG_MCN_MEDIAN	Log of MCN_MEDIAN
41	LOG_MAX_SPECTRAL_RANGE	Log of Max Spectral Range
42	CN_MEDIAN	Median of Conditional Number
43	RangeCorr_Alpha_Reprtd	Correlation of spectral range and alpha
44	corr_conditional_num	Correlation of conditional number with alpha