

# EVALUATING THE ROBUSTNESS OF CHINCHILLA COMPUTE-OPTIMAL SCALING

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Hoffmann et al. (2022)’s Chinchilla paper introduced the principle of compute-optimal scaling, laying a foundation for future scaling of language models. In the years since, however, valid concerns about Chinchilla have been raised: wide confidence intervals, discrepancies between its three approaches, and incongruities with other scaling laws. This raises a critical question for the field: Can practitioners still rely on Chinchilla’s prescriptions? Our work demonstrates the answer is yes. We begin by uncovering that the model parameters central to Chinchilla’s analyses were ambiguous: three interpretations are possible, with relative differences between different interpretations of model parameters as high as 15.2%. We find that, perhaps surprisingly, which model parameters are used for the analyses do not meaningfully affect key results: the scaling law estimates and the compute-optimal tokens-to-parameter ratio. Indeed, under one interpretation, the tokens-to-parameter ratio becomes more constant with the target compute budget. We then ask how distorted the Chinchilla model parameters *could* have been without meaningfully affecting the key results. By deliberately perturbing model parameters in four structured ways, we find that key Chinchilla results are most sensitive to additive or systematic errors, which can alter the otherwise flat trend of the optimal tokens-to-parameter ratio, but overall, Chinchilla’s key results withstand sizable perturbations. Altogether, our findings offer the field renewed confidence in Chinchilla as a durable guide for scaling language models.

## 1 INTRODUCTION

The study of neural scaling laws, which predictably map training resources to model performance, is a cornerstone of modern language modeling research and engineering. Hestness et al. (2017) and soon after Kaplan et al. (2020) laid the foundation by demonstrating that pretraining losses scale as power laws with the number of data points, model parameters and pretraining compute. This was followed by significant additional work in the ensuing years (Sec. 4 Related Work). Such discoveries led to the modern paradigm of training extremely large language models (OpenAI et al., 2024; DeepMind et al., 2025; DeepSeek-AI et al., 2025; Qwen et al., 2025; Kimi et al., 2025; Zhipu-AI et al., 2025).

The field’s understanding of scaling models was later altered by the seminal work of Hoffmann et al. (2022), which introduced the concept of compute-optimal scaling. By training over 400 models ranging from 44M to 16B parameters on 5B to 500B tokens, Hoffmann et al. (2022) discovered that producing the best performing model with respect to a fixed pretraining compute budget (“compute-optimal”) was achieved by linearly scaling model parameters and pretraining data together. Chinchilla established the influential “20-to-1” heuristic: that the compute-optimal amount of training data is approximately 20 tokens per model parameter (Appendix C). Their 70B Chinchilla outperformed larger models (Rae et al., 2022), cementing the methodology as a guiding principle for the field.

In the years since, several contributions have closely scrutinized Chinchilla, raising a number of concerns: Zhang (2023) called attention to Chinchilla’s wide confidence intervals and questioned whether such uncertain estimates can provide practical guidance. Besiroglu et al. (2024) investigated why some of Chinchilla’s approaches yielded inconsistent results. Lastly, Porian et al. (2024) and Pearce & Song (2024) examined why Chinchilla makes different predictions than Kaplan et al. (2020)’s earlier scaling work. While these works are clear contributions to the science of scaling, the field has been left uncertain: can practitioners still confidently rely on Chinchilla’s prescriptions?

In this work, we aim to answer this question. As a warm up, we uncover that the model parameters central to the Chinchilla analyses were ambiguous, with three different possible interpretations as to which model parameters were used: (1) the model parameters reported in Hoffmann et al. (2022)’s Table A9, (2) the model parameters calculated from the reported model architectural hyperparameters (layers, dimensions, number of heads, etc.) using a “standard” formula, and (3) the model parameters calculated from a “best-fit” formula. Although the relative error among these three sets of model parameters rises as high as 15.2%, we show that key Chinchilla results – the estimated scaling law parameters and the compute-optimal tokens-per-parameter ratio – do not meaningfully change. In fact, the only potential consequence is that the compute-optimal tokens-per-parameter ratio becomes *more* constant with respect to the target compute budget, strengthening Chinchilla’s finding.

To more generally assess the robustness of compute-optimal scaling, we then study how distorted the model parameters *could* have been without changing Chinchilla’s key results. We perform a sensitivity analysis by perturbing model parameters in four structured ways. Our analyses reveals that the robustness depends on the nature of the perturbations: while multiplicative perturbations and random noise have limited effects, additive constants or systematic biases can qualitatively change the compute-optimal scaling strategy by altering the trend of the optimal tokens-to-parameter ratio. However, overall, all four sensitivity analyses demonstrate that Chinchilla’s key results withstand sizable perturbations. Our results reveal a clear picture: Chinchilla’s compute-optimal prescription remains robust, further justifying its widespread use as a practical scaling blueprint for practitioners.

## 2 KEY CHINCHILLA RESULTS ARE ROBUST TO THREE INTERPRETATIONS OF CHINCHILLA MODEL PARAMETERS

One of the fundamental inputs to the Chinchilla analyses are the number of parameters per model. However, an ambiguity exists as to which exact model parameters were used, with three different possible interpretations differing by as much as 15.2%. We uncovered this by closely examining Chinchilla’s Table A9, which reports the number of model parameters for each model alongside key architectural hyperparameters, e.g., `d_model`, `ffw_size`, `kv_size`, `n_heads` and `n_layers`. We call the model parameters reported in Chinchilla’s Table A9 the **reported model parameters**. We include a brief snippet in our main text (Table 1) and the full table in our Appendix B.

However, a second interpretation of the model parameters arises from the provided architecture hyperparameters; assuming the embedding and unembedding weights are tied (Press & Wolf, 2017) and no gating is present, a standard formula for the number of model parameters is:

$$\begin{aligned} \text{Standard Formula Model Params} &\approx \text{Embedding Params} + \text{Attn Params} + \text{FFN Params} \\ \text{Embedding Params} &= \text{Vocab Size} \cdot d_{\text{model}} \\ \text{Attn Params} &= n_{\text{layers}} \cdot (4 \cdot d_{\text{model}} \cdot kv_{\text{size}} \cdot n_{\text{heads}}) \\ \text{FFN Params} &= n_{\text{layers}} \cdot (2 \cdot d_{\text{model}} \cdot ffw_{\text{size}}) \end{aligned} \quad (1)$$

Comparing the **standard formula model parameters** with the reported model parameters reveals a mismatch for *every* model, with an average relative error of 7.4% but reaching as high as 15.2% and no less than 3.6% (Fig. 1, left). We calculate relative error as:

$$\text{Relative Error (\%)} = 100 \cdot \frac{\text{Reported Model Params} - \text{Standard Formula Model Params}}{\text{Reported Model Params}}. \quad (2)$$

In an attempt to reconcile the two interpretations of model parameters, we determined a third interpretation based on a “best fit” formula that nearly matches the reported model parameters:

$$\begin{aligned} \text{Best Fit Formula Model Params} &\approx \text{Embedding Params} + \text{Attn Params} + \text{FFN Params} \\ \text{Embedding Params} &= \text{Vocab Size} \cdot d_{\text{model}} \\ \text{Attn Params} &= n_{\text{layers}} \cdot (5 \cdot d_{\text{model}} \cdot kv_{\text{size}} \cdot n_{\text{heads}}) \\ \text{FFN Params} &= n_{\text{layers}} \cdot (2 \cdot d_{\text{model}} \cdot ffw_{\text{size}}) \end{aligned} \quad (3)$$

Switching from the standard formula model parameters to the **best fit model parameters** reduced the number of discrepancies with the reported model parameters from 50/50 models to 6/50 models, and reduced the largest relative error from 15.2% to 8.7% (Fig. 1, right).

Table A9 from Hoffmann et al. (2022)						Our Contribution		
d_model	ffw_size	kv_size	n_heads	n_layers	n_vocab	Chinchilla's Reported Model Parameters (M)	Best Fit Formula's Model Parameters (M)	Standard Formula's Model Parameters (M)
512	2048	64	8	8	32168	44	44	42
576	2304	64	9	9	32168	57	57	54
640	2560	64	10	10	32168	74	74	70
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4864	19456	128	36	47	32168	13775	14319	13266
4992	19968	128	32	49	32168	14940	14939	13937
5120	20480	128	40	47	32168	16183	16182	14950

Table 1: **Three Interpretations of Chinchilla’s Model Parameters.** Hoffmann et al. (2022)’s Table A9 provides the architectural hyperparameters of all models used in the Chinchilla analyses, along with the reported model parameters (specified in millions). However, two alternative interpretations of model parameters are possible: model parameters calculated from architectural hyperparameters using a “standard” formula (Eqn. 1) and model parameters calculated from architectural hyperparameters using a “best fit” formula (Eqn. 3). For the complete table, see Appendix B.

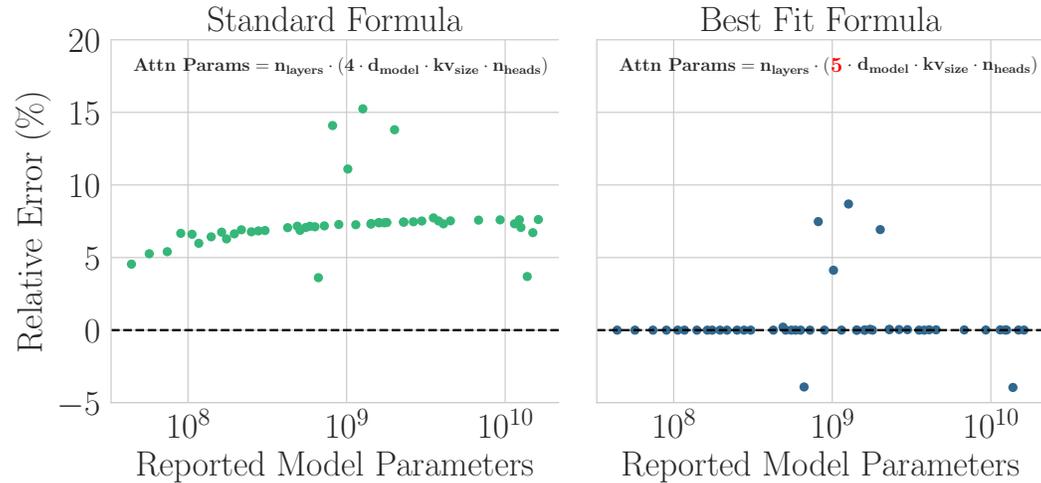


Figure 1: **Disagreement Between Three Interpretations of Chinchilla’s Model Parameters.** Each point is one of the 50 models in Hoffmann et al. (2022)’s Table A9. **Left:** Calculating model parameters from the provided architectural hyperparameters using a *standard formula* (Eqn. 1; attention parameters =  $n\_layers \cdot 4 \cdot d\_model \cdot kv\_size \cdot n\_heads$ ) disagrees with the reported model parameters for 50/50 models, with relative errors averaging 7.388% and rising as high as 15.2%. **Right:** Calculating model parameters using a *best fit formula* (Eqn. 3; replace 4 with 5) matches 44/50 of the reported model parameters, and reduces the largest relative error to 8.7%.

Having identified these three different notions of model parameters, we tested how Chinchilla’s key results might change depending on which model parameters are used for fitting. We focus on two key results in particular: First, Chinchilla fit a neural scaling law to the pretraining loss  $L$  as a function of model parameters  $N$  and pretraining data  $D$ :

$$L(N, D) = E + \frac{A}{N^\alpha} + \frac{B}{D^\beta}, \tag{4}$$

where  $E$  is the irreducible error,  $A$  is the parameter prefactor,  $\alpha$  is the parameter exponent,  $B$  is the data prefactor and  $\beta$  is the data exponent. Second, Chinchilla derived from the estimated scaling law parameters a “20-to-1” heuristic for the compute-optimal ratio of tokens-per-parameters:

$$\text{Compute-Optimal Tokens-per-Parameter Ratio} \approx 20. \tag{5}$$

162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215

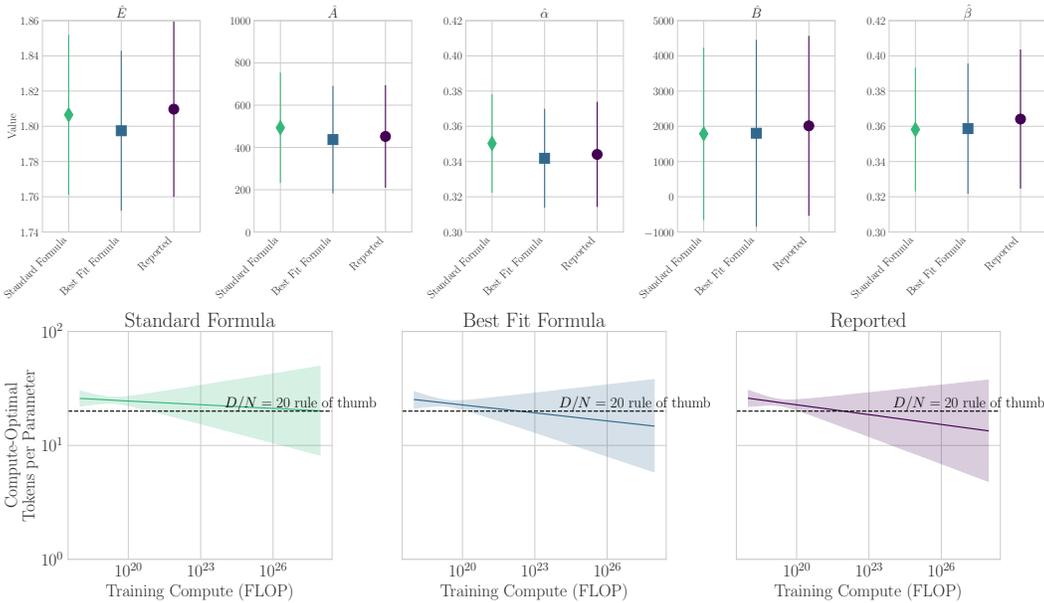


Figure 2: **Key Chinchilla Results Are Robust to All Three Interpretations of Model Parameters.** Hoffmann et al. (2022) fit a neural scaling law  $L(N, D) = E + A \cdot N^{-\alpha} + B \cdot D^{-\beta}$ , where  $N$  is the number of model parameters and  $D$  is the number of data (Eqn. 4). **Top:** The fit parameters ( $\hat{E}$ ,  $\hat{A}$ ,  $\hat{\alpha}$ ,  $\hat{B}$ ,  $\hat{\beta}$ ) do not meaningfully change, regardless of which model parameters are used for fitting. Error bars are standard errors from 4000 bootstrapped samples. **Bottom:** The compute-optimal tokens-per-parameter ratio remains constant at  $\approx 20$ , regardless of which notion of model parameters are used in the fitting process. The slope is flattest with the standard formula model parameters ( $-0.572$  per decade; best fit:  $-1.049$ ; reported:  $-1.248$ ). Error bars are 80% confidence intervals. Fitting and visualization were conducted using Besiroglu et al. (2024)’s code.

We tested how these two Chinchilla headline results change depending on which of the three interpretations is used in the fits: (1) Reported Model Parameters, (2) Standard Formula Model Parameters, or (3) Best Fit Formula Model Parameters. To conduct the Chinchilla fits and visualize the results, we used Besiroglu et al. (2024)’s Chinchilla fitting code that was previously used to align Chinchilla’s three approaches.

Perhaps surprisingly, we found that none of the five fit parameters  $\hat{E}$ ,  $\hat{A}$ ,  $\hat{\alpha}$ ,  $\hat{B}$ ,  $\hat{\beta}$  differed significantly depending on which of our three notions of model parameters were used in fitting (Fig. 2, top). We similarly found the compute-optimal tokens-per-parameter ratio remains constant around 20 tokens per parameter (Fig. 2, bottom). Arguably, the standard formula model parameters yield a *flatter* trend with increasing training compute: the slope for the standard formula model parameters is  $-0.572$  for each 10x increase in compute, which decreased to  $-1.049$  for the best fit formula model parameters and decreased further to  $-1.248$  for the reported model parameters. However, uncertainty makes drawing strong conclusions difficult. These results demonstrate that **key Chinchilla results are robust to whichever of our three notions of model parameters is used in the fitting process.**

### 3 ROBUSTNESS OF CHINCHILLA HEADLINE RESULTS DEPENDS ON TYPE OF PERTURBATION TO MODEL PARAMETERS

Given that the key Chinchilla results did not meaningfully change even when model parameters differed by as much as 15.2%, we next asked:

*How distorted could the model parameters have been without meaningfully affecting Chinchilla’s headline results?*

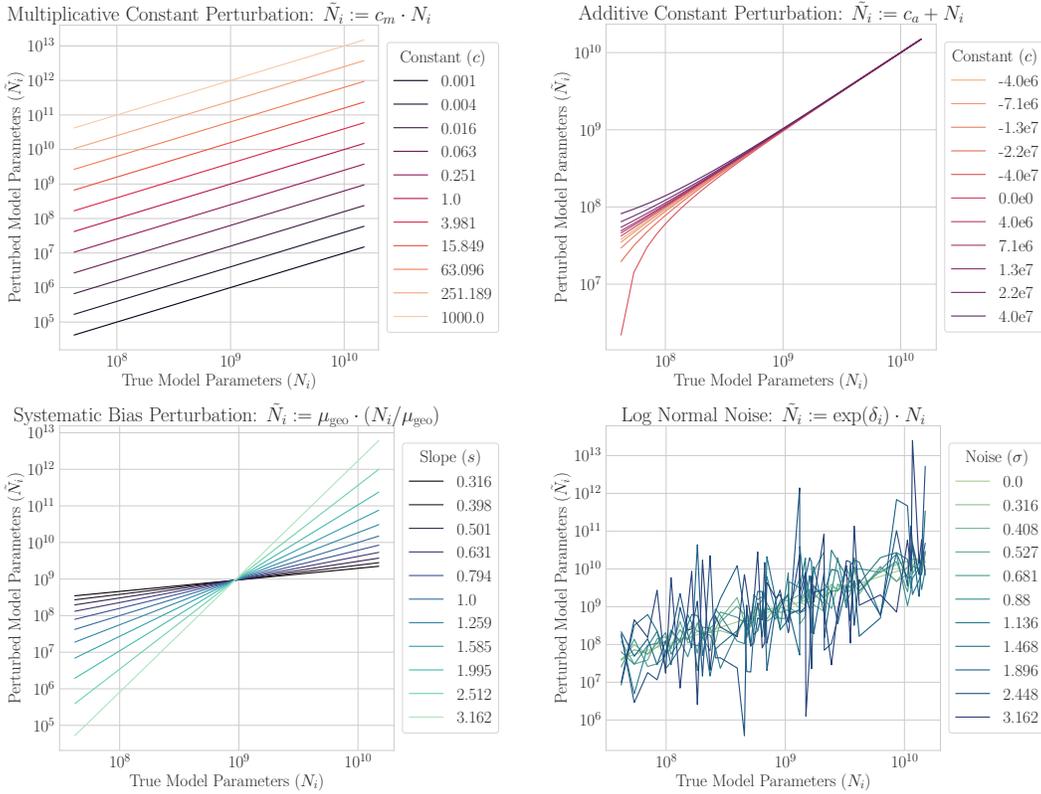


Figure 3: **Evaluating the Robustness of Chinchilla via Four Model Parameter Perturbations.** We study how robust key Chinchilla results are to structured perturbations of models’ parameters. **Top Left:** In the first perturbation, motivated by Sec. 2, we perturb model parameters with a multiplicative constant  $c_m$ . **Top Right:** In the second perturbation, we perturb model parameters with an additive constant  $c_a$ , that could perhaps arise due to embedding parameters being included/excluded. **Bottom Left:** In the third perturbation, we perturb model parameters with a systematic bias: either smaller models’ parameters are larger and larger models’ parameters are smaller, or smaller models’ parameters are smaller and larger models’ parameters are larger; the systematic bias is controlled by slope  $s$ . **Bottom Right:** In the fourth perturbation, we assume the relationship of the loss with model parameters is perhaps noisy, e.g., (Frankle & Carbin, 2019), with noise strength parameterized by  $\sigma$ .

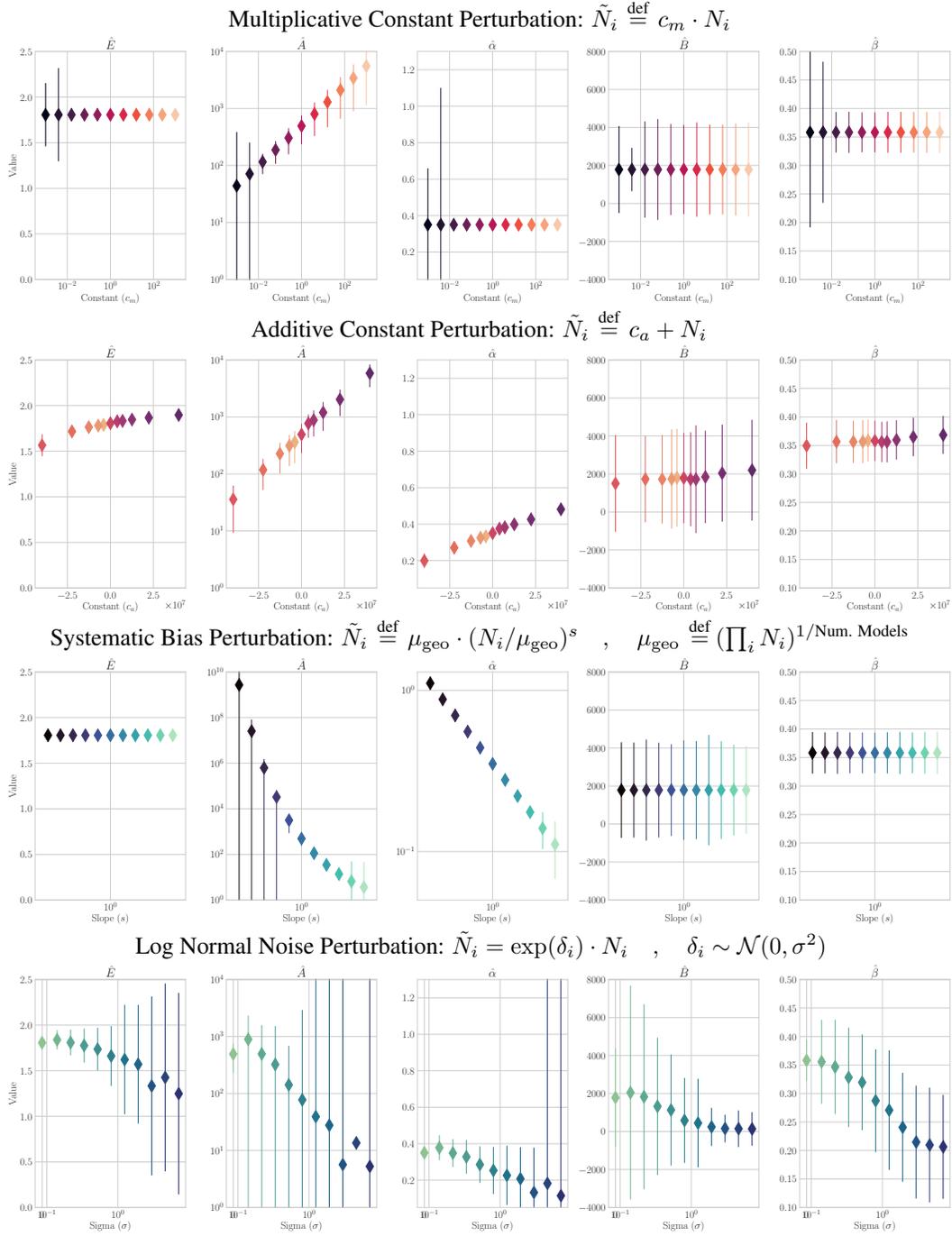
To answer this question, we intentionally perturbed the standard formula model parameters in four structured ways: multiplicative constant, additive constant, systematic bias and log normal noise. We then reran the Chinchilla fitting processes using the perturbed model parameters to see what effect each type of perturbation has on the estimated scaling law parameters and compute-optimal tokens-per-parameter. We offer visual intuition for each of the four types of perturbations in Fig. 3, and mathematically describe each type of perturbation along with its consequences below. For our theoretical analyses, see Appendix C.

### 3.1 MULTIPLICATIVE CONSTANT PERTURBATION INCREASES $\hat{A}$ EXPONENTIALLY

Motivated by Sec. 2, for our first perturbation, we assume model parameters are systematically under/overestimated by approximately the same percentage. To model this, we multiplied all true model parameters  $\{N_i\}_i$  by constant multiplier  $c_m$  to produce perturbed model parameters  $\{\tilde{N}_i\}_i$ :

$$\tilde{N}_i \stackrel{\text{def}}{=} c_m \cdot N_i. \quad (6)$$

We swept  $c_m$  in  $\text{logspace}(-3, 3, \text{num}=11)$ . For visual intuition, see Fig. 3, top left.



314  
315  
316  
317  
318  
319  
320  
321  
322  
323

**Figure 4: Robustness of Fit Neural Scaling Law Parameters Under Four Types of Model Parameter Perturbations.** Each row visualizes the effect of a different perturbation on the five fit parameters of the Chinchilla scaling law ( $L(N, D) = E + A \cdot N^{-\alpha} + B \cdot D^{-\beta}$ ). **Row 1:** A multiplicative constant perturbation ( $c$ ) increases the model parameter prefactor ( $\hat{A}$ ) exponentially, while other fit parameters remain stable. **Row 2:** An additive constant perturbation ( $c$ ) linearly increases the model parameter exponent ( $\hat{\alpha}$ ) and exponentially increases its prefactor ( $\hat{A}$ ), with only a gentle rise in the irreducible loss ( $\hat{E}$ ). **Row 3:** A systematic bias perturbation ( $s$ ) causes the model parameter exponent ( $\hat{\alpha}$ ) to decay as a power law ( $s^{-1}$ ) and the prefactor ( $\hat{A}$ ) to decline sub-polynomially. **Row 4:** Adding log-normal noise ( $\sigma$ ) primarily increases the uncertainty of all fit parameters, while weakly decreasing the model parameter exponent ( $\hat{\alpha}$ ) logarithmically and the prefactor ( $\hat{A}$ ) polynomially. Error bars are standard errors obtained by 4000 bootstrapped samples.

As we derive in Appendix C.2.1 and show empirically in Fig. 4’s first row, the fitting process compensates for this multiplicative error primarily by adjusting the model size prefactor,  $\hat{A}$ , to approximately  $\hat{A}c_m^\alpha$ , while the scaling exponent,  $\hat{\alpha}$ , remains largely unchanged, i.e.,  $\hat{\alpha} \approx \alpha$ . This makes sense for moderate  $c_m$ : if  $\hat{A}$  is the best fit for the true model parameters, then replacing the true parameters with the perturbed parameters  $N_i \rightarrow \tilde{N}_i$  and rescaling  $\hat{A} \rightarrow \tilde{\hat{A}} = (c_m)^\alpha \hat{A}$  produces approximately the same fit. As a consequence, Fig. 5 Top Left shows the compute-optimal tokens per parameter remains constant with pretraining compute, but the precise constant grows as a power of  $c_m$  if true model parameters are underestimated ( $c_m < 1$ ) and shrinks if true model parameters are overestimated ( $c_m > 1$ ). The only exceptions are for the two smallest multiplicative constants (0.001 and 0.004), where uncertainty in  $\hat{A}$  and  $\hat{\alpha}$  produced NaNs.

### 3.2 ADDITIVE CONSTANT PERTURBATION INCREASES $\hat{\alpha}$ LINEARLY AND $\hat{A}$ EXPONENTIALLY

In our second perturbation, we assume model parameters have an additive term. For example, embedding parameters may be included or excluded, a key detail in previous scaling law studies (Kaplan et al., 2020; Hoffmann et al., 2022) that is partially responsible for discrepancies between estimated scaling laws’ parameters (Pearce & Song, 2024; Porian et al., 2024)). To model this, we added constant  $c_a$  to all model parameters:

$$\tilde{N}_i \stackrel{\text{def}}{=} c_a + N_i. \quad (7)$$

We swept  $c_a$  in  $-\log\text{space}(6.6, 7.6, \text{num}=5) \cup \{0\} \cup \log\text{space}(6.6, 7.6, \text{num}=5)$ . For visual intuition, see Fig. 3 Top Right. For additional context, the smallest Chinchilla model has  $42 \times 10^6$  parameters.

Fig. 4’s second row shows the effects: (i) the irreducible loss  $\hat{E}$  rises only gently from 1.565 to 1.897 ( $\approx 21\%$ ). (ii) the model parameter prefactor  $\hat{A}$  grows exponentially in  $c_a$ , increasing by  $\sim 2.5x$  from the most negative to the most positive constant (iii) the model parameter exponent  $\hat{\alpha}$  increases linearly with  $c_a$  from 0.199 to 0.481 and (iv) both the data prefactor  $\hat{B}$  and data exponent  $\hat{\beta}$  fluctuate only within their bootstrap error bars and show no systematic trend. As a consequence, Fig. 5 Top Right shows the compute-optimal tokens per parameter becomes less constant with the training compute: a larger positive  $c_a$  means larger target training horizons require more tokens per parameter, whereas a larger negative  $c_a$  means larger target training horizons require fewer tokens per parameter.

In Appendix C.2.2, we analytically explain these trends: the most critical parameter in a power law is its exponent, which corresponds to its slope in log-log space. However, for the perturbed function, the slope is now no longer constant and depends on  $N$  as  $N/(N + c_a)$ . Thus, the fitting procedure must select a single exponent that best represents the varying slope over the range of data. When  $c_a > 0$ , the factor  $N/(N + c_a) < 1$ , and the fitting process must select an exponent  $\hat{\alpha} > \alpha$ ; and when  $c_a < 0$ , the factor  $N/(N + c_a) > 1$ , and to compensate, the fitting process must select an exponent  $\hat{\alpha} < \alpha$ .

For comparison, Porian et al. (2024) found that including the model’s head parameters increased the fit model parameter scaling exponent  $\hat{\alpha}$  by 0.080 (0.072  $\rightarrow$  0.152), and Pearce & Song (2024) found that including embedding parameters increased  $\hat{\alpha}$  by 0.231 (0.135  $\rightarrow$  0.366). Although assuming an additive constant is a simplification of both analyses, all three results are quantitatively similar.

### 3.3 SYSTEMATIC BIAS PERTURBATION DECREASES $\hat{\alpha}$ POLYNOMIALLY AND $\hat{A}$ SUB-POLYNOMIALLY

In our third perturbation, we assume the presence of a systematic bias in reported models’ parameters: either the smaller models’ parameters are truly larger and the larger models’ parameters truly smaller, or vice versa. To model this, we define the perturbed parameters as

$$\tilde{N}_i \stackrel{\text{def}}{=} \mu_{\text{geo}} \cdot (N_i / \mu_{\text{geo}})^s, \quad (8)$$

where  $\mu_{\text{geo}} \stackrel{\text{def}}{=} (\prod_i N_i)^{1/\text{Num. Models}}$  is the geometric mean of the model parameters and  $s$  is the systematic bias parameter:  $s < 1$  shrinks large models and inflates small ones, whereas  $s > 1$  does the reverse. We swept  $s$  in  $\log\text{space}(-0.5, 0.5, 11)$ . For visual intuition, see Fig. 3 Bottom Left.

The third row of Fig. 4 illustrates three main effects: (i) The model parameter exponent  $\hat{\alpha}$  decays according to the power-law relationship  $\hat{\alpha} = 10^{-0.46} \cdot s^{-1}$ , which is a nearly perfect fit to the data

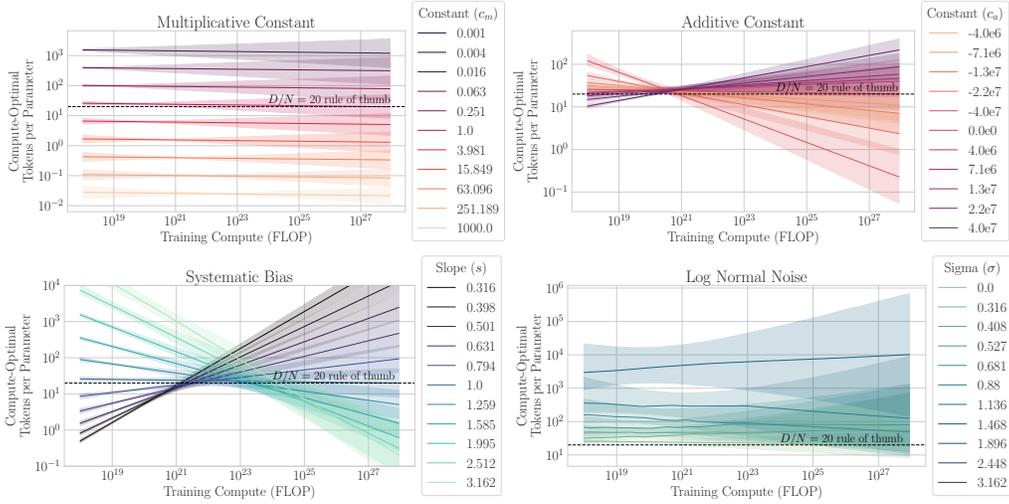


Figure 5: **Robustness of Compute-Optimal Tokens-Per-Parameter Under Four Types of Model Parameter Perturbations.** Shaded regions represent 80% confidence intervals. **Top Left:** A multiplicative constant perturbation by  $c$  shifts the compute-optimal ratio by  $c^\alpha$  but keeps the trend flat with respect to training compute. **Top Right:** An additive constant perturbation by  $c$  makes the compute-optimal ratio less constant across the target training compute horizon. A positive slope means more tokens are needed per parameter for larger compute budgets, while a negative slope means fewer are required. **Bottom Left:** A systematic bias also makes the ratio less constant. A larger bias ( $s > 1$ ) leads to fewer optimal tokens per parameter for larger models, whereas a smaller bias ( $s < 1$ ) requires more. **Bottom Right:** Adding log-normal noise to model parameters increases the uncertainty and the overall magnitude of the compute-optimal tokens per parameter ratio.

( $R^2 > 0.999$ ,  $p \approx 5.9 \times 10^{-90}$ ). (ii) The model parameter prefactor  $\hat{A}$  declines sub-polynomially. (iii) The irreducible loss,  $\hat{E}$ , and the data parameters,  $\hat{B}$  and  $\hat{\beta}$ , show no systematic trend, with fluctuations remaining within their bootstrap error bars. Similar to the Additive Constant perturbation, Fig. 5 Bottom Left shows that the two trends of  $\hat{A}$  and  $\hat{\alpha}$  together make the compute-optimal tokens per parameter ratio less constant with the target training horizon: a larger systematic bias  $s$  means larger target training horizons require fewer tokens per parameter, whereas a smaller systematic bias  $s$  means larger target training horizons require more tokens per parameter.

In Appendix C.2.3, we mathematically derive that under the systematic bias, the model parameter exponent is multiplied by  $s^{-1}$  and the model prefactor is multiplied by  $\mu_{\text{geo}}^{\alpha(1-s)/s}$ , making the exponent in the compute-optimal ratio  $(\alpha/s - \beta)/(\alpha/s + \beta)$ . Thus, if  $s < 1$ , then the exponent on  $C$  becomes positive and compute-optimal ratio of tokens-per-parameter increases with compute, whereas if  $s > 1$ , then the exponent on  $C$  becomes negative and the compute-optimal ratio of tokens-per-parameter decreases with compute.

### 3.4 LOG NORMAL NOISE PERTURBATION INCREASES UNCERTAINTY AND DECREASES $\hat{\alpha}$ LOGARITHMICALLY AND $\hat{A}$ POLYNOMIALLY

In our fourth perturbation, we assume the “value” of model parameters is noisily measured, perhaps due to model initializations. To model this, we added log-normal noise to the number of parameters. Specifically, for each model’s parameter count  $N_i$ , we sampled a new parameter count as:

$$\tilde{N}_i \stackrel{\text{def}}{=} \exp(\delta_i) \cdot N_i, \quad \delta_i \sim \mathcal{N}(0, \sigma^2). \quad (9)$$

We swept  $\sigma$  from  $1 \times 10^{-2}$  to  $1 \times 10^2$ . For visual intuition, see Fig. 3 Bottom Right.

The fourth row of Fig. 4 illustrates three main effects: (i) Nearly all fit parameters have significantly larger confidence intervals, especially as the noise standard deviation  $\sigma$  increases; for the highest value of 3.162,  $\hat{A}$  and  $\hat{\alpha}$  are nearly unidentifiable. (ii) To the extent that trends can be identified, the irreducible error  $\hat{E}$  trends down weakly, while the model parameters prefactor  $\hat{A}$  falls roughly

432 polynomially and the model parameters exponent falls roughly logarithmically with the noise standard  
433 deviation  $\sigma$ . Fig. 5 Bottom Right demonstrates the consequences of the noise: fits with too high  
434 of noise create NaNs, while noise drives up the compute-optimal tokens per parameter and also  
435 increases the width of the 80% confidence intervals by  $\sim 1$  order of magnitude, although the inferred  
436 values are roughly constant with target training compute.

#### 437 438 4 RELATED WORK 439

440 Due to space constraints, we defer most Related Work to Appendix D and focus here on prior  
441 research most relevant to our contribution. The precise details of Hoffmann et al. (2022) have recently  
442 come under scrutiny, leading to a number of important replication and re-evaluation studies. For  
443 instance, Chinchilla used three different approaches, two of which agreed with each other, but the  
444 third did not; Besiroglu et al. (2024) conducted a detailed investigation of this third analysis and  
445 found that it could be made consistent with the first two analyses by fixing optimizer issues and  
446 not rounding reported fit parameters. In a similar vein, Porian et al. (2024) and Pearce & Song  
447 (2024) sought to resolve a discrepancy between Hoffmann et al. (2022) and Kaplan et al. (2020) on  
448 how to scale data and parameters to produce the best performing model; Porian et al. (2024) found  
449 that the discrepancy could be resolved by three differences (last layer computational cost, warmup  
450 duration, and scale-dependent optimizer tuning) while Pearce & Song (2024) found that much of the  
451 discrepancy could be attributed to Kaplan et al. (2020) counting only non-embedding parameters.

452 Like Besiroglu et al. (2024) and Porian et al. (2024), our work scrutinizes the seminal work of Chin-  
453 chilla. However, our analyses focuses specifically on how robust the original Chinchilla methodology  
454 and results are to different perturbations. Our contribution concludes with a direct confirmation of  
455 the original findings, providing evidence that Chinchilla’s compute-optimal guidance is robust.

#### 456 457 5 DISCUSSION 458

459 This work began with a perhaps surprising result: three different interpretations of Chinchilla’s model  
460 parameters are possible, with discrepancies as high as 15.2%, but all three support (or strengthen) key  
461 Chinchilla results. Neither the estimated scaling law parameters nor the widely adopted “20-to-1”  
462 compute-optimal tokens-to-parameter ratio changed meaningfully. Indeed, our refitting using the  
463 standard formula model parameters suggests an even more stable relationship, with the token-to-  
464 parameter ratio varying even less across different pretraining compute budgets.

465 To understand this robustness more deeply, we systematically investigated how various hypothetical  
466 perturbations would affect key Chinchilla results. We perturbed the parameter counts in four structured  
467 ways and re-ran the fitting analysis for each. This stress test revealed the specific ways in which  
468 different types of errors impact the scaling law parameters. A simple multiplicative error, for example,  
469 exponentially shifts the constant in the optimal tokens-per-parameter ratio, while an additive error or  
470 a systematic bias can more dramatically alter its trend with respect to the target training compute  
471 budget.

472 Ultimately, our findings serve as both a critical re-examination and a powerful confirmation of  
473 the original Chinchilla results. Our subsequent analyses should give practitioners even greater  
474 confidence in Chinchilla’s compute-optimal prescription. Its guidance withstands not only the  
475 specific interpretation used, but also a range of other potential perturbations, reinforcing its value as a  
476 durable and practical blueprint for the field.

477 **Future Directions** One obvious next step is to evaluate the robustness of more recent scaling results  
478 with additional considerations such as inference constraints (Sardana et al., 2024), data constraints  
479 (Muennighoff et al., 2023) and overtraining (Gadre et al., 2024).

## REFERENCES

- 486  
487  
488 Samira Abnar, Harshay Shah, Dan Busbridge, Alaaeldin Mohamed Elnouby Ali, Josh Susskind,  
489 and Vimal Thilak. Parameters vs flops: Scaling laws for optimal sparsity for mixture-of-experts  
490 language models. *arXiv preprint arXiv:2501.12370*, 2025.
- 491 Rishabh Agarwal, Avi Singh, Lei M Zhang, Bernd Bohnet, Luis Rosias, Stephanie C.Y. Chan,  
492 Biao Zhang, Ankesh Anand, Zaheer Abbas, Azade Nova, John D Co-Reyes, Eric Chu, Feryal  
493 Behbahani, Aleksandra Faust, and Hugo Larochelle. Many-shot in-context learning. In *The  
494 Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=AB6XpMzvqH>.
- 495  
496 Armen Aghajanyan, Lili Yu, Alexis Conneau, Wei-Ning Hsu, Karen Hambardzumyan, Susan Zhang,  
497 Stephen Roller, Naman Goyal, Omer Levy, and Luke Zettlemoyer. Scaling laws for generative  
498 mixed-modal language models. In *International Conference on Machine Learning*, pp. 265–279.  
499 PMLR, 2023.
- 500 Ibrahim M Alabdulmohsin, Behnam Neyshabur, and Xiaohua Zhai. Revisiting neural scaling laws in  
501 language and vision. *Advances in Neural Information Processing Systems*, 35:22300–22312, 2022.
- 502  
503 Cem Anil, Esin DURMUS, Nina Rimskey, Mrinank Sharma, Joe Benton, Sandipan Kundu, Joshua  
504 Batson, Meg Tong, Jesse Mu, Daniel J Ford, Francesco Mosconi, Rajashree Agrawal, Rylan  
505 Schaeffer, Naomi Bashkansky, Samuel Svenningsen, Mike Lambert, Ansh Radhakrishnan, Carson  
506 Denison, Evan J Hubinger, Yuntao Bai, Trenton Bricken, Timothy Maxwell, Nicholas Schiefer,  
507 James Sully, Alex Tamkin, Tamera Lanham, Karina Nguyen, Tomasz Korbak, Jared Kaplan, Deep  
508 Ganguli, Samuel R. Bowman, Ethan Perez, Roger Baker Grosse, and David Duvenaud. Many-shot  
509 jailbreaking. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*,  
510 2024. URL <https://openreview.net/forum?id=cw5mgd71jW>.
- 511 Aryaman Arora, Dan Jurafsky, Christopher Potts, and Noah D. Goodman. Bayesian scaling laws for  
512 in-context learning, 2024. URL <https://arxiv.org/abs/2410.16531>.
- 513 Alexander Atanasov, Jacob A Zavatore-Veth, and Cengiz Pehlevan. Scaling and renormalization in  
514 high-dimensional regression. *arXiv preprint arXiv:2405.00592*, 2024.
- 515  
516 Gregor Bachmann, Sotiris Anagnostidis, and Thomas Hofmann. Scaling mlps: A tale of inductive  
517 bias, 2023. URL <https://arxiv.org/abs/2306.13575>.
- 518 Yasaman Bahri, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma. Explaining neural  
519 scaling laws. *Proceedings of the National Academy of Sciences*, 121(27):e2311878121, 2024.
- 520  
521 N Barkai, Hyunjune Sebastian Seung, and Haim Sompolinsky. Scaling laws in learning of classifica-  
522 tion tasks. *Physical review letters*, 70(20):3167, 1993.
- 523  
524 Matthew Barnett. An empirical study of scaling laws for transfer, 2024. URL <https://arxiv.org/abs/2408.16947>.
- 525  
526 Tamay Besiroglu, Ege Erdil, Matthew Barnett, and Josh You. Chinchilla scaling: A replication  
527 attempt, 2024. URL <https://arxiv.org/abs/2404.10102>.
- 528  
529 Song Bian, Minghao Yan, and Shivaram Venkataraman. Scaling inference-efficient language models.  
530 In *International Conference on Machine Learning*, 2025.
- 531  
532 Blake Bordelon, Alexander Atanasov, and Cengiz Pehlevan. A dynamical model of neural scaling  
533 laws. *arXiv preprint arXiv:2402.01092*, 2024a.
- 534  
535 Blake Bordelon, Alexander Atanasov, and Cengiz Pehlevan. How feature learning can improve neural  
536 scaling laws. *arXiv preprint arXiv:2409.17858*, 2024b.
- 537 Olivier Bousquet, Steve Hanneke, Shay Moran, Ramon van Handel, and Amir Yehudayoff. A theory  
538 of universal learning, 2020. URL <https://arxiv.org/abs/2011.04483>.
- 539 Ari Brill. Neural scaling laws rooted in the data distribution. *arXiv preprint arXiv:2412.07942*, 2024.

- 540 Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V. Le, Christopher Ré, and  
541 Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling,  
542 2024. URL <https://arxiv.org/abs/2407.21787>.
- 543  
544 Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal,  
545 Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are  
546 few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- 547 Dan Busbridge, Amitis Shidani, Floris Weers, Jason Ramapuram, Etai Littwin, and Russell Webb.  
548 Distillation scaling laws. In *Forty-second International Conference on Machine Learning*, 2025.
- 549  
550 Yanxi Chen, Xuchen Pan, Yaliang Li, Bolin Ding, and Jingren Zhou. A simple and provable scaling  
551 law for the test-time compute of large language models, 2024. URL <https://arxiv.org/abs/2411.19477>.
- 552  
553 Zhengyu Chen, Siqi Wang, Teng Xiao, Yudong Wang, Shiqi Chen, Xunliang Cai, Junxian He, and  
554 Jingang Wang. Revisiting scaling laws for language models: The role of data quality and training  
555 strategies. In *Proceedings of the 63rd Annual Meeting of the Association for Computational*  
556 *Linguistics (Volume 1: Long Papers)*, pp. 23881–23899, 2025.
- 557  
558 Mehdi Cherti, Romain Beaumont, Ross Wightman, Mitchell Wortsman, Gabriel Ilharco, Cade  
559 Gordon, Christoph Schuhmann, Ludwig Schmidt, and Jenia Jitsev. Reproducible scaling laws for  
560 contrastive language-image learning. In *Proceedings of the IEEE/CVF Conference on Computer*  
561 *Vision and Pattern Recognition*, pp. 2818–2829, 2023.
- 562  
563 Aidan Clark, Diego de Las Casas, Aurelia Guy, Arthur Mensch, Michela Paganini, Jordan Hoffmann,  
564 Bogdan Damoc, Blake Hechtman, Trevor Cai, Sebastian Borgeaud, et al. Unified scaling laws for  
565 routed language models. In *International conference on machine learning*, pp. 4057–4086. PMLR,  
566 2022.
- 567  
568 Google DeepMind, Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen  
569 Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, Luke Marris,  
570 Sam Petulla, Colin Gaffney, Asaf Aharoni, Nathan Lintz, Tiago Cardal Pais, Henrik Jacobsson,  
571 Idan Szpektor, Nan-Jiang Jiang, Krishna Haridasan, Ahmed Omran, Nikunj Saunshi, Dara Bahri,  
572 Gaurav Mishra, Eric Chu, Toby Boyd, Brad Hekman, Aaron Parisi, Chaoyi Zhang, Kornaphop  
573 Kawintiranon, Tania Bedrax-Weiss, Oliver Wang, Ya Xu, Ollie Purkiss, Uri Mendlovic, Ilai Deutel,  
574 Nam Nguyen, Adam Langley, Flip Korn, Lucia Rossazza, Alexandre Ramé, Sagar Waghmare,  
575 Helen Miller, Nathan Byrd, Ashrith Sheshan, Raia Hadsell Sangnie Bhardwaj, Pawel Janus, Tero  
576 Rissa, Dan Horgan, Sharon Silver, Ayzaan Wahid, Sergey Brin, Yves Raimond, Klemen Kloboves,  
577 Cindy Wang, Nitesh Bharadwaj Gundavarapu, Iliia Shumailov, Bo Wang, Mantas Pajarskas, Joe  
578 Heyward, Martin Nikolchev, Maciej Kula, Hao Zhou, Zachary Garrett, Sushant Kafle, Sercan Arik,  
579 Ankita Goel, Mingyao Yang, Jiho Park, Koji Kojima, Parsa Mahmoudieh, Koray Kavukcuoglu,  
580 Grace Chen, Doug Fritz, Anton Bulyenov, Sudeshna Roy, Dimitris Paparas, Hadar Shemtov,  
581 Bo-Juen Chen, Robin Strudel, David Reitter, Aurko Roy, Andrey Vlasov, Changwan Ryu, Chas  
582 Leichner, Haichuan Yang, Zelda Mariet, Denis Vnukov, Tim Sohn, Amy Stuart, Wei Liang, Minmin  
583 Chen, Praynaa Rawlani, Christy Koh, JD Co-Reyes, Guangda Lai, Praseem Banzal, Dimitrios  
584 Vytiniotis, Jieru Mei, Mu Cai, Mohammed Badawi, Corey Fry, Ale Hartman, Daniel Zheng, Eric  
585 Jia, James Keeling, Annie Louis, Ying Chen, Efren Robles, Wei-Chih Hung, Howard Zhou, Nikita  
586 Saxena, Sonam Goenka, Olivia Ma, Zach Fisher, Mor Hazan Taeye, Emily Graves, David Steiner,  
587 Yujia Li, Sarah Nguyen, Rahul Sukthankar, Joe Stanton, Ali Eslami, Gloria Shen, Berkin Akin,  
588 Alexey Guseynov, Yiqian Zhou, Jean-Baptiste Alayrac, Armand Joulin, Efrat Farkash, Ashish  
589 Thapliyal, Stephen Roller, Noam Shazeer, Todor Davchev, Terry Koo, Hannah Forbes-Pollard,  
590 Kartik Audhkhasi, Greg Farquhar, Adi Mayrav Gilady, Maggie Song, John Aslanides, Piermaria  
591 Mendolicchio, Alicia Parrish, John Blitzer, Pramod Gupta, Xiaoen Ju, Xiaochen Yang, Puranjay  
592 Datta, Andrea Tacchetti, Sanket Vaibhav Mehta, Gregory Dobb, Shubham Gupta, Federico Piccinini,  
593 Raia Hadsell, Sujeer Rajayogam, Jiepu Jiang, Patrick Griffin, Patrik Sundberg, Jamie Hayes, Alexey  
Frolov, Tian Xie, Adam Zhang, Kingshuk Dasgupta, Uday Kalra, Lior Shani, Klaus Macherey,  
Tzu-Kuo Huang, Liam MacDermed, Karthik Duddu, Paulo Zacchello, Zi Yang, Jessica Lo, Kai  
Hui, Matej Kastelic, Derek Gasaway, Qijun Tan, Summer Yue, Pablo Barrio, John Wieting, Weel  
Yang, Andrew Nystrom, Solomon Demmessie, Anselm Levskaya, Fabio Viola, Chetan Tekur,  
Greg Billock, George Necula, Mandar Joshi, Rylan Schaeffer, Swachhand Lokhande, Christina

594 Sorokin, Pradeep Shenoy, Mia Chen, Mark Collier, Hongji Li, Taylor Bos, Nevan Wichers,  
595 Sun Jae Lee, Angéline Pouget, Santhosh Thangaraj, Kyriakos Axiotis, Phil Crone, Rachel Sterneck,  
596 Nikolai Chinaev, Victoria Krakovna, Oleksandr Ferludin, Ian Gemp, Stephanie Winkler, Dan  
597 Goldberg, Ivan Korotkov, Kefan Xiao, Malika Mehrotra, Sandeep Mariserla, Vihari Piratla, Terry  
598 Thurk, Khiem Pham, Hongxu Ma, Alexandre Senges, Ravi Kumar, Clemens Meyer, Ellie Talius,  
599 Nuo Wang Pierser, Ballie Sandhu, Horia Toma, Kuo Lin, Swaroop Nath, Tom Stone, Dorsa Sadigh,  
600 Nikita Gupta, Arthur Guez, Avi Singh, Matt Thomas, Tom Duerig, Yuan Gong, Richard Tanburn,  
601 Lydia Lihui Zhang, Phuong Dao, Mohamed Hammad, Sirui Xie, Shruti Rijhwani, Ben Murdoch,  
602 Duhyeon Kim, Will Thompson, Heng-Tze Cheng, Daniel Sohn, Pablo Sprechmann, Qiantong  
603 Xu, Srinivas Tadepalli, Peter Young, Ye Zhang, Hansa Srinivasan, Miranda Aperghis, Aditya  
604 Ayyar, Hen Fitoussi, Ryan Burnell, David Madras, Mike Dusenberry, Xi Xiong, Tayo Oguntebi,  
605 Ben Albrecht, Jörg Bornschein, Jovana Mitrović, Mason Dimarco, Bhargav Kanagal Shamanna,  
606 Premal Shah, Eren Sezener, Shyam Upadhyay, Dave Lacey, Craig Schiff, Sebastien Baur, Sanjay  
607 Ganapathy, Eva Schneider, Mateo Wirth, Connor Schenck, Andrey Simanovsky, Yi-Xuan Tan,  
608 Philipp Fränken, Dennis Duan, Bharath Mankalale, Nikhil Dhawan, Kevin Sequeira, Zichuan  
609 Wei, Shivanker Goel, Caglar Unlu, Yukun Zhu, Haitian Sun, Ananth Balashankar, Kurt Shuster,  
610 Megh Umekar, Mahmoud Alnahlawi, Aäron van den Oord, Kelly Chen, Yuexiang Zhai, Zihang  
611 Dai, Kuang-Huei Lee, Eric Doi, Lukas Zilka, Rohith Vallu, Disha Shrivastava, Jason Lee, Hisham  
612 Husain, Honglei Zhuang, Vincent Cohen-Addad, Jarred Barber, James Atwood, Adam Sadovsky,  
613 Quentin Wellens, Steven Hand, Arunkumar Rajendran, Aybuke Turker, CJ Carey, Yuanzhong Xu,  
614 Hagen Soltau, Zefei Li, Xinying Song, Conglong Li, Iurii Kemaev, Sasha Brown, Andrea Burns,  
615 Viorica Patraucean, Piotr Stanczyk, Renga Aravamudhan, Mathieu Blondel, Hila Noga, Lorenzo  
616 Blanco, Will Song, Michael Isard, Mandar Sharma, Reid Hayes, Dalia El Badawy, Avery Lamp,  
617 Itay Laish, Olga Kozlova, Kelvin Chan, Sahil Singla, Srinivas Sunkara, Mayank Upadhyay, Chang  
618 Liu, Aijun Bai, Jarek Wilkiewicz, Martin Zlocha, Jeremiah Liu, Zhuowan Li, Haiguang Li, Omer  
619 Barak, Ganna Raboshchuk, Jiho Choi, Fangyu Liu, Erik Jue, Mohit Sharma, Andreea Marzoca,  
620 Robert Busa-Fekete, Anna Korsun, Andre Elisseeff, Zhe Shen, Sara Mc Carthy, Kay Lamerigts,  
621 Anahita Hosseini, Hanzhao Lin, Charlie Chen, Fan Yang, Kushal Chauhan, Mark Omernick,  
622 Dawei Jia, Karina Zainullina, Demis Hassabis, Danny Vainstein, Ehsan Amid, Xiang Zhou, Ronny  
623 Votol, Eszter Vértés, Xinjian Li, Zongwei Zhou, Angeliki Lazaridou, Brendan McMahan, Arjun  
624 Narayanan, Hubert Soyer, Sujoy Basu, Kayi Lee, Bryan Perozzi, Qin Cao, Leonard Berrada, Rahul  
625 Arya, Ke Chen, Katrina, Xu, Matthias Lochbrunner, Alex Hofer, Sahand Sharifzadeh, Renjie  
626 Wu, Sally Goldman, Pranjal Awasthi, Xuezhi Wang, Yan Wu, Claire Sha, Biao Zhang, Maciej  
627 Mikula, Filippo Graziano, Siobhan Mcloughlin, Irene Giannoumis, Youhei Namiki, Chase Malik,  
628 Carey Radebaugh, Jamie Hall, Ramiro Leal-Cavazos, Jianmin Chen, Vikas Sindhwani, David Kao,  
629 David Greene, Jordan Griffith, Chris Welty, Ceslee Montgomery, Toshihiro Yoshino, Liangzhe  
630 Yuan, Noah Goodman, Assaf Hurwitz Michaely, Kevin Lee, KP Sawhney, Wei Chen, Zheng  
631 Zheng, Megan Shum, Nikolay Savinov, Etienne Pot, Alex Pak, Morteza Zadimoghaddam, Sijal  
632 Bhatnagar, Yoad Lewenberg, Blair Kutzman, Ji Liu, Lesley Katzen, Jeremy Selier, Josip Djolonga,  
633 Dmitry Lepikhin, Kelvin Xu, Jacky Liang, Jiewen Tan, Benoit Schillings, Muge Ersoy, Pete  
634 Blois, Bernd Bandemer, Abhimanyu Singh, Sergei Lebedev, Pankaj Joshi, Adam R. Brown, Evan  
635 Palmer, Shreya Pathak, Komal Jalan, Fedir Zubach, Shuba Lall, Randall Parker, Alok Gunjan,  
636 Sergey Rogulenko, Sumit Sanghai, Zhaoqi Leng, Zoltan Egyed, Shixin Li, Maria Ivanova, Kostas  
637 Andriopoulos, Jin Xie, Elan Rosenfeld, Auriel Wright, Ankur Sharma, Xinyang Geng, Yicheng  
638 Wang, Sam Kwei, Renke Pan, Yujing Zhang, Gabby Wang, Xi Liu, Chak Yeung, Elizabeth  
639 Cole, Aviv Rosenberg, Zhen Yang, Phil Chen, George Polovets, Pranav Nair, Rohun Saxena,  
640 Josh Smith, Shuo yin Chang, Aroma Mahendru, Svetlana Grant, Anand Iyer, Irene Cai, Jed  
641 McGiffin, Jiaming Shen, Alanna Walton, Antonious Girgis, Oliver Woodman, Rosemary Ke, Mike  
642 Kwong, Louis Rouillard, Jinneng Rao, Zhihao Li, Yuntao Xu, Flavien Prost, Chi Zou, Ziwei Ji,  
643 Alberto Magni, Tyler Liechty, Dan A. Calian, Deepak Ramachandran, Igor Krivokon, Hui Huang,  
644 Terry Chen, Anja Hauth, Anastasija Ilić, Weijuan Xi, Hyeontaek Lim, Vlad-Doru Ion, Pooya  
645 Moradi, Metin Toksoz-Exley, Kalesha Bullard, Miltos Allamanis, Xiaomeng Yang, Sophie Wang,  
646 Zhi Hong, Anita Gergely, Cheng Li, Bhavishya Mittal, Vitaly Kovalev, Victor Ungureanu, Jane  
647 Labanowski, Jan Wassenberg, Nicolas Lacasse, Geoffrey Cideron, Petar Dević, Annie Marsden,  
Lynn Nguyen, Michael Fink, Yin Zhong, Tatsuya Kiyono, Desi Ivanov, Sally Ma, Max Bain,  
Kiran Yalasangi, Jennifer She, Anastasia Petrushkina, Mayank Lunayach, Carla Bromberg, Sarah  
Hodkinson, Vilobh Meshram, Daniel Vlasic, Austin Kyker, Steve Xu, Jeff Stanway, Zuguang Yang,  
Kai Zhao, Matthew Tung, Seth Odoom, Yasuhisa Fujii, Justin Gilmer, Eunyoung Kim, Felix Halim,  
Quoc Le, Bernd Bohnet, Seliem El-Sayed, Behnam Neyshabur, Malcolm Reynolds, Dean Reich,

648 Yang Xu, Erica Moreira, Anuj Sharma, Zeyu Liu, Mohammad Javad Hosseini, Naina Raisinghani,  
649 Yi Su, Ni Lao, Daniel Formoso, Marco Gelmi, Almog Gueta, Tapomay Dey, Elena Gribovskaya,  
650 Domagoj Čević, Sidharth Mudgal, Garrett Bingham, Jianling Wang, Anurag Kumar, Alex Cullum,  
651 Feng Han, Konstantinos Bousmalis, Diego Cedillo, Grace Chu, Vladimir Magay, Paul Michel,  
652 Ester Hlavnova, Daniele Calandriello, Setareh Ariaifar, Kaisheng Yao, Vikash Sehwal, Arpi Vezer,  
653 Agustin Dal Lago, Zhenkai Zhu, Paul Kishan Rubenstein, Allen Porter, Anirudh Baddepudi, Oriana  
654 Riva, Mihai Dorin Istin, Chih-Kuan Yeh, Zhi Li, Andrew Howard, Nilpa Jha, Jeremy Chen, Raoul  
655 de Liedekerke, Zafarali Ahmed, Mikel Rodriguez, Tanuj Bhatia, Bangju Wang, Ali Elqursh, David  
656 Klinghoffer, Peter Chen, Pushmeet Kohli, Te I, Weiyang Zhang, Zack Nado, Jilin Chen, Maxwell  
657 Chen, George Zhang, Aayush Singh, Adam Hillier, Federico Lebron, Yiqing Tao, Ting Liu, Gabriel  
658 Dulac-Arnold, Jingwei Zhang, Shashi Narayan, Buhuang Liu, Orhan Firat, Abhishek Bhowmick,  
659 Bingyuan Liu, Hao Zhang, Zizhao Zhang, Georges Rotival, Nathan Howard, Anu Sinha, Alexander  
660 Grushetsky, Benjamin Beyret, Keerthana Gopalakrishnan, James Zhao, Kyle He, Szabolcs Payrits,  
661 Zaid Nabulsi, Zhaoyi Zhang, Weijie Chen, Edward Lee, Nova Fallen, Sreenivas Gollapudi, Aurick  
662 Zhou, Filip Pavetić, Thomas Köppe, Shiyu Huang, Rama Pasumarthi, Nick Fernando, Felix  
663 Fischer, Daria Ćurko, Yang Gao, James Svensson, Austin Stone, Haroon Qureshi, Abhishek  
664 Sinha, Apoorv Kulshreshtha, Martin Matysiak, Jieming Mao, Carl Saroufim, Aleksandra Faust,  
665 Qingnan Duan, Gil Fidel, Kaan Katircioglu, Raphaël Lopez Kaufman, Dhruv Shah, Weize Kong,  
666 Abhishek Bapna, Gellért Weisz, Emma Dunleavy, Praneet Dutta, Tianqi Liu, Rahma Chaabouni,  
667 Carolina Parada, Marcus Wu, Alexandra Belias, Alessandro Bissacco, Stanislav Fort, Li Xiao,  
668 Fantine Huot, Chris Knutsen, Yochai Blau, Gang Li, Jennifer Prendki, Juliette Love, Yinlam  
669 Chow, Pichi Charoenpanit, Hidetoshi Shimokawa, Vincent Coriou, Karol Gregor, Tomas Izo, Arjun  
670 Akula, Mario Pinto, Chris Hahn, Dominik Paulus, Jiaxian Guo, Neha Sharma, Cho-Jui Hsieh,  
671 Adaeze Chukwuka, Kazuma Hashimoto, Nathalie Rauschmayr, Ling Wu, Christof Angermueller,  
672 Yulong Wang, Sebastian Gerlach, Michael Pliskin, Daniil Mirylenka, Min Ma, Lexi Baugher,  
673 Bryan Gale, Shaan Bijwadia, Nemanja Rakićević, David Wood, Jane Park, Chung-Ching Chang,  
674 Babi Seal, Chris Tar, Kacper Krasowiak, Yiwen Song, Georgi Stephanov, Gary Wang, Marcello  
675 Maggioni, Stein Xudong Lin, Felix Wu, Shachi Paul, Zixuan Jiang, Shubham Agrawal, Bilal Piot,  
676 Alex Feng, Cheolmin Kim, Tulsee Doshi, Jonathan Lai, Chuqiao Xu, Sharad Vikram, Ciprian  
677 Chelba, Sebastian Krause, Vincent Zhuang, Jack Rae, Timo Denk, Adrian Collister, Lotte Weerts,  
678 Xianghong Luo, Yifeng Lu, Håvard Garnes, Nitish Gupta, Terry Spitz, Avinatan Hassidim, Lihao  
679 Liang, Izhak Shafran, Peter Humphreys, Kenny Vassigh, Phil Wallis, Virat Shejwalkar, Nicolas  
680 Perez-Nieves, Rachel Hornung, Melissa Tan, Beka Westberg, Andy Ly, Richard Zhang, Brian  
681 Farris, Jongbin Park, Alec Kosik, Zeynep Cankara, Andrii Maksai, Yunhan Xu, Albin Cassirer,  
682 Sergi Caelles, Abbas Abdolmaleki, Mencher Chiang, Alex Fabrikant, Shravya Shetty, Luheng  
683 He, Mai Giménez, Hadi Hashemi, Sheena Panthaplackel, Yana Kulizhskaya, Salil Deshmukh,  
684 Daniele Pighin, Robin Alazard, Disha Jindal, Seb Noury, Pradeep Kumar S, Siyang Qin, Xerxes  
685 Dotiwalla, Stephen Spencer, Mohammad Babaeizadeh, Blake JianHang Chen, Vaibhav Mehta,  
686 Jennie Lees, Andrew Leach, Penporn Koanantakool, Iliia Akolzin, Ramona Comanescu, Junwhan  
687 Ahn, Alexey Svyatkovskiy, Basil Mustafa, David D'Ambrosio, Shiva Mohan Reddy Garlapati,  
688 Pascal Lamblin, Alekh Agarwal, Shuang Song, Pier Giuseppe Sessa, Pauline Coquinot, John  
689 Maggs, Hussain Masoom, Divya Pitta, Yaqing Wang, Patrick Morris-Suzuki, Billy Porter, Johnson  
690 Jia, Jeffrey Dudek, Raghavender R, Cosmin Paduraru, Alan Ansell, Tolga Bolukbasi, Tony Lu,  
691 Ramya Ganeshan, Zi Wang, Henry Griffiths, Rodrigo Benenson, Yifan He, James Swirhun, George  
692 Papamakarios, Aditya Chawla, Kuntal Sengupta, Yan Wang, Vedrana Milutinovic, Igor Mordatch,  
693 Zhipeng Jia, Jamie Smith, Will Ng, Shitij Nigam, Matt Young, Eugen Vušak, Blake Hechtman,  
694 Sheela Goenka, Avital Zipori, Kareem Ayoub, Ashok Popat, Trilok Acharya, Luo Yu, Dawn  
695 Bloxwich, Hugo Song, Paul Roit, Haiqiong Li, Aviel Boag, Nigamaa Nayakanti, Bilva Chandra,  
696 Tianli Ding, Aahil Mehta, Cath Hope, Jiageng Zhang, Idan Heimlich Shtacher, Kartikeya Badola,  
697 Ryo Nakashima, Andrei Sozanschi, Iulia Comşa, Ante Žužul, Emily Caveness, Julian Odell,  
698 Matthew Watson, Dario de Cesare, Phillip Lippe, Derek Lockhart, Siddharth Verma, Huizhong  
699 Chen, Sean Sun, Lin Zhuo, Aditya Shah, Prakhar Gupta, Alex Muzio, Ning Niu, Amir Zait,  
700 Abhinav Singh, Meenu Gaba, Fan Ye, Prajit Ramachandran, Mohammad Saleh, Raluca Ada Popa,  
701 Ayush Dubey, Frederick Liu, Sara Javanmardi, Mark Epstein, Ross Hemsley, Richard Green,  
Nishant Ranka, Eden Cohen, Chuyuan Kelly Fu, Sanjay Ghemawat, Jed Borovik, James Martens,  
Anthony Chen, Pranav Shyam, André Susano Pinto, Ming-Hsuan Yang, Alexandru Țifrea, David  
Du, Boqing Gong, Ayushi Agarwal, Seungyeon Kim, Christian Frank, Saloni Shah, Xiaodan Song,  
Zhiwei Deng, Ales Mikhalap, Kleopatra Chatziprimou, Timothy Chung, Toni Creswell, Susan

702 Zhang, Yennie Jun, Carl Lebsack, Will Truong, Slavica Andačić, Itay Yona, Marco Fornoni, Rong  
703 Rong, Serge Toropov, Afzal Shama Soudagar, Andrew Audibert, Salah Zaiem, Zaheer Abbas,  
704 Andrei Rusu, Sahitya Potluri, Shitao Weng, Anastasios Kementsietsidis, Anton Tsitsulin, Daiyi  
705 Peng, Natalie Ha, Sanil Jain, Tejasi Latkar, Simeon Ivanov, Cory McLean, Anirudh GP, Rajesh  
706 Venkataraman, Canoe Liu, Dilip Krishnan, Joel D'sa, Roey Yogev, Paul Collins, Benjamin Lee,  
707 Lewis Ho, Carl Doersch, Gal Yona, Shawn Gao, Felipe Tiengo Ferreira, Adnan Ozturel, Hannah  
708 Muckenhirn, Ce Zheng, Gargi Balasubramaniam, Mudit Bansal, George van den Driessche, Sivan  
709 Eiger, Salem Haykal, Vedant Misra, Abhimanyu Goyal, Danilo Martins, Gary Leung, Jonas  
710 Valfridsson, Four Flynn, Will Bishop, Chenxi Pang, Yoni Halpern, Honglin Yu, Lawrence Moore,  
711 Yuvein, Zhu, Sridhar Thiagarajan, Yoel Drori, Zhisheng Xiao, Lucio Dery, Rolf Jagerman, Jing  
712 Lu, Eric Ge, Vaibhav Aggarwal, Arjun Khare, Vinh Tran, Oded Elyada, Ferran Alet, James Rubin,  
713 Ian Chou, David Tian, Libin Bai, Lawrence Chan, Lukasz Lew, Karolis Misiunas, Taylan Bilal,  
714 Aniket Ray, Sindhu Raghuram, Alex Castro-Ros, Viral Carpenter, CJ Zheng, Michael Kilgore,  
715 Josef Broder, Emily Xue, Praveen Kallakuri, Dheeru Dua, Nancy Yuen, Steve Chien, John Schultz,  
716 Saurabh Agrawal, Reut Tsarfaty, Jingcao Hu, Ajay Kannan, Dror Marcus, Nisarg Kothari, Baochen  
717 Sun, Ben Horn, Matko Bošnjak, Ferjad Naeem, Dean Hirsch, Lewis Chiang, Boya Fang, Jie Han,  
718 Qifei Wang, Ben Hora, Antoine He, Mario Lučić, Beer Changpinyo, Anshuman Tripathi, John  
719 Youssef, Chester Kwak, Philippe Schlattner, Cat Graves, Rémi Leblond, Wenjun Zeng, Anders  
720 Andreassen, Gabriel Rasskin, Yue Song, Eddie Cao, Junhyuk Oh, Matt Hoffman, Wojtek Skut,  
721 Yichi Zhang, Jon Stritar, Xingyu Cai, Saarthak Khanna, Kathie Wang, Shriya Sharma, Christian  
722 Reisswig, Younghoon Jun, Aman Prasad, Tatiana Sholokhova, Preeti Singh, Adi Gerzi Rosenthal,  
723 Anian Ruoss, Françoise Beaufays, Sean Kirmani, Dongkai Chen, Johan Schalkwyk, Jonathan  
724 Herzig, Been Kim, Josh Jacob, Damien Vincent, Adrian N Reyes, Ivana Balazevic, Léonard  
725 Hussenot, Jon Schneider, Parker Barnes, Luis Castro, Spandana Raj Babbula, Simon Green,  
726 Serkan Cabi, Nico Duduta, Danny Driess, Rich Galt, Noam Velan, Junjie Wang, Hongyang Jiao,  
727 Matthew Mauger, Du Phan, Miteyan Patel, Vlado Galić, Jerry Chang, Eyal Marcus, Matt Harvey,  
728 Julian Salazar, Elahe Dabir, Suraj Satishkumar Sheth, Amol Mandhane, Hanie Sedghi, Jeremiah  
729 Willcock, Amir Zandieh, Shruthi Prabhakara, Aida Amini, Antoine Miech, Victor Stone, Massimo  
730 Nicosia, Paul Niemczyk, Ying Xiao, Lucy Kim, Sławek Kwasiborski, Vikas Verma, Ada Maksutaj  
731 Oflazer, Christoph Hirsenschall, Peter Sung, Lu Liu, Richard Everett, Michiel Bakker, Ágoston  
732 Weisz, Yufe Wang, Vivek Sampathkumar, Uri Shaham, Bibo Xu, Yasemin Altun, Mingqiu Wang,  
733 Takaaki Saeki, Guanjie Chen, Emanuel Taropa, Shanthal Vasanth, Sophia Austin, Lu Huang,  
734 Goran Petrovic, Qingyun Dou, Daniel Golovin, Grigory Rozhdestvenskiy, Allie Culp, Will Wu,  
735 Motoki Sano, Divya Jain, Julia Proskurnia, Sébastien Cevey, Alejandro Cruzado Ruiz, Piyush  
736 Patil, Mahdi Mirzazadeh, Eric Ni, Javier Snaider, Lijie Fan, Alexandre Fréchet, AJ Piergiovanni,  
737 Shariq Iqbal, Kenton Lee, Claudio Fantacci, Jinwei Xing, Lisa Wang, Alex Irpan, David Raposo,  
738 Yi Luan, Zhuoyuan Chen, Harish Ganapathy, Kevin Hui, Jiazhong Nie, Isabelle Guyon, Heming  
739 Ge, Roopali Vij, Hui Zheng, Dayeong Lee, Alfonso Castaño, Khuslen Baatarsukh, Gabriel  
740 Ibagon, Alexandra Chronopoulou, Nicholas FitzGerald, Shashank Viswanadha, Safeen Huda,  
741 Rivka Moroshko, Georgi Stoyanov, Prateek Kolhar, Alain Vaucher, Ishaan Watts, Adhi Kuncoro,  
742 Henryk Michalewski, Satish Kambala, Bat-Orgil Batsaikhan, Alek Andreev, Irina Jurenka, Maigo  
743 Le, Qihang Chen, Wael Al Jishi, Sarah Chakera, Zhe Chen, Aditya Kini, Vikas Yadav, Aditya  
744 Siddhant, Iliia Labzovsky, Balaji Lakshminarayanan, Carrie Grimes Bostock, Pankil Botadra,  
745 Ankesh Anand, Colton Bishop, Sam Conway-Rahman, Mohit Agarwal, Yani Donchev, Achintya  
746 Singhal, Félix de Chaumont Quitry, Natalia Ponomareva, Nishant Agrawal, Bin Ni, Kalpesh  
747 Krishna, Masha Samsikova, John Karro, Yilun Du, Tamara von Glehn, Caden Lu, Christopher A.  
748 Choquette-Choo, Zhen Qin, Tingnan Zhang, Sicheng Li, Divya Tyam, Swaroop Mishra, Wing  
749 Lowe, Colin Ji, Weiyi Wang, Manaal Faruqui, Ambrose Slone, Valentin Dalibard, Arunachalam  
750 Narayanaswamy, John Lambert, Pierre-Antoine Manzagol, Dan Karliner, Andrew Bolt, Ivan  
751 Lobov, Aditya Kusupati, Chang Ye, Xuan Yang, Heiga Zen, Nelson George, Mukul Bhutani,  
752 Olivier Lacombe, Robert Riachi, Gagan Bansal, Rachel Soh, Yue Gao, Yang Yu, Adams Yu,  
753 Emily Nottage, Tania Rojas-Esponda, James Noraky, Manish Gupta, Ragha Kotikalapudi, Jichuan  
754 Chang, Sanja Deur, Dan Graur, Alex Mossin, Erin Farnese, Ricardo Figueira, Alexandre Moufared,  
755 Austin Huang, Patrik Zochbauer, Ben Ingram, Tongzhou Chen, Zelin Wu, Adria Puigdomènech,  
Leland Rechis, Da Yu, Sri Gayatri Sundara Padmanabhan, Rui Zhu, Chu ling Ko, Andrea Banino,  
Samira Daruki, Aarush Selvan, Dhruva Bhaswar, Daniel Hernandez Diaz, Chen Su, Salvatore  
Scellato, Jennifer Brennan, Woohyun Han, Grace Chung, Priyanka Agrawal, Urvashi Khandelwal,  
Khe Chai Sim, Morgane Lustman, Sam Ritter, Kelvin Guu, Jiawei Xia, Prateek Jain, Emma Wang,  
Tyrone Hill, Mirko Rossini, Marija Kostelac, Tautvydas Misiunas, Amit Sabne, Kyuyeun Kim,

756 Ahmet Iscen, Congchao Wang, José Leal, Ashwin Sreevatsa, Utku Evci, Manfred Warmuth, Saket  
757 Joshi, Daniel Suo, James Lottes, Garrett Honke, Brendan Jou, Stefani Karp, Jieru Hu, Himanshu  
758 Sahni, Adrien Ali Taïga, William Kong, Samrat Ghosh, Renshen Wang, Jay Pavagadhi, Natalie  
759 Axelsson, Nikolai Grigorev, Patrick Siegler, Rebecca Lin, Guohui Wang, Emilio Parisotto, Sharath  
760 Maddineni, Krishan Subudhi, Eyal Ben-David, Elena Pochernina, Orgad Keller, Thi Avrahami,  
761 Zhe Yuan, Pulkit Mehta, Jialu Liu, Sherry Yang, Wendy Kan, Katherine Lee, Tom Funkhouser,  
762 Derek Cheng, Hongzhi Shi, Archit Sharma, Joe Kelley, Matan Eyal, Yury Malkov, Corentin Tallec,  
763 Yuval Bahat, Shen Yan, Xintian, Wu, David Lindner, Chengda Wu, Avi Caciularu, Xiyang Luo,  
764 Rodolphe Jenatton, Tim Zaman, Yingying Bi, Ilya Kornakov, Ganesh Mallya, Daisuke Ikeda, Itay  
765 Karo, Anima Singh, Colin Evans, Praneeth Netrapalli, Vincent Nallatamby, Isaac Tian, Yannis  
766 Assael, Vikas Raunak, Victor Carbune, Ioana Bica, Lior Madmoni, Dee Cattle, Snchit Grover,  
767 Krishna Somandepalli, Sid Lall, Amelio Vázquez-Reina, Riccardo Patana, Jiaqi Mu, Pranav Talluri,  
768 Maggie Tran, Rajeev Aggarwal, RJ Skerry-Ryan, Jun Xu, Mike Burrows, Xiaoyue Pan, Edouard  
769 Yvinec, Di Lu, Zhiying Zhang, Duc Dung Nguyen, Hairong Mu, Gabriel Barcik, Helen Ran,  
770 Lauren Beltrone, Krzysztof Choromanski, Dia Kharrat, Samuel Albanie, Sean Purser-haskell,  
771 David Bieber, Carrie Zhang, Jing Wang, Tom Hudson, Zhiyuan Zhang, Han Fu, Johannes Mauerer,  
772 Mohammad Hossein Bateni, AJ Maschinot, Bing Wang, Muye Zhu, Arjun Pillai, Tobias Weyand,  
773 Shuang Liu, Oscar Akerlund, Fred Bertsch, Vittal Premachandran, Alicia Jin, Vincent Roulet,  
774 Peter de Boursac, Shubham Mittal, Ndaba Ndebele, Georgi Karadzhov, Sahra Ghalebikesabi,  
775 Ricky Liang, Allen Wu, Yale Cong, Nimesh Ghelani, Sumeet Singh, Bahar Fatemi, Warren, Chen,  
776 Charles Kwong, Alexey Kolganov, Steve Li, Richard Song, Chenkai Kuang, Sobhan Miryoosefi,  
777 Dale Webster, James Wendt, Arkadiusz Socala, Guolong Su, Artur Mendonça, Abhinav Gupta,  
778 Xiaowei Li, Tomy Tsai, Qiong, Hu, Kai Kang, Angie Chen, Sertan Girgin, Yongqin Xian, Andrew  
779 Lee, Nolan Ramsden, Leslie Baker, Madeleine Clare Elish, Varvara Krayvanova, Rishabh Joshi,  
780 Jiri Simsa, Yao-Yuan Yang, Piotr Ambroszczyk, Dipankar Ghosh, Arjun Kar, Yuan Shangguan,  
781 Yumeya Yamamori, Yaroslav Akulov, Andy Brock, Haotian Tang, Siddharth Vashishtha, Rich  
782 Munoz, Andreas Steiner, Kalyan Andra, Daniel Eppens, Qixuan Feng, Hayato Kobayashi, Sasha  
783 Goldshtein, Mona El Mahdy, Xin Wang, Jilei, Wang, Richard Killam, Tom Kwiatkowski, Kavya  
784 Kopparapu, Serena Zhan, Chao Jia, Alexei Bendebury, Sheryl Luo, Adrià Recasens, Timothy  
785 Knight, Jing Chen, Mohak Patel, YaGuang Li, Ben Withbroe, Dean Weesner, Kush Bhatia, Jie  
786 Ren, Danielle Eisenbud, Ebrahim Songhori, Yanhua Sun, Travis Choma, Tasos Kementsietsidis,  
787 Lucas Manning, Brian Roark, Wael Farhan, Jie Feng, Susheel Tatineni, James Cobon-Kerr, Yunjie  
788 Li, Lisa Anne Hendricks, Isaac Noble, Chris Breaux, Nate Kushman, Liqian Peng, Fuzhao Xue,  
789 Taylor Tobin, Jamie Rogers, Josh Lipschultz, Chris Alberti, Alexey Vlaskin, Mostafa Dehghani,  
790 Roshan Sharma, Tris Warkentin, Chen-Yu Lee, Benigno Uria, Da-Cheng Juan, Angad Chandorkar,  
791 Hila Sheftel, Ruibo Liu, Elnaz Davoodi, Borja De Balle Pigem, Kedar Dhamdhere, David Ross,  
792 Jonathan Hoeh, Mahdis Mahdieh, Li Liu, Qiuqia Li, Liam McCafferty, Chenxi Liu, Markus  
793 Mircea, Yunting Song, Omkar Savant, Alaa Saade, Colin Cherry, Vincent Hellendoorn, Siddharth  
794 Goyal, Paul Pucciarelli, David Vilar Torres, Zohar Yahav, Hyo Lee, Lars Lowe Sjoesund, Christo  
795 Kirov, Bo Chang, Deepanway Ghoshal, Lu Li, Gilles Baechler, Sébastien Pereira, Tara Sainath,  
796 Anudhyan Boral, Dominik Grewe, Afief Halumi, Nguyet Minh Phu, Tianxiao Shen, Marco Tulio  
797 Ribeiro, Dhriti Varma, Alex Kaskasoli, Vlad Feinberg, Navneet Potti, Jarrod Kahn, Matheus  
798 Wisniewski, Shakir Mohamed, Arnar Mar Hrafnkelsson, Bobak Shahriari, Jean-Baptiste Lespiau,  
799 Lisa Patel, Legg Yeung, Tom Paine, Lantao Mei, Alex Ramirez, Rakesh Shivanna, Li Zhong, Josh  
800 Woodward, Guilherme Tubone, Samira Khan, Heng Chen, Elizabeth Nielsen, Catalin Ionescu,  
801 Utsav Prabhu, Mingcen Gao, Qingze Wang, Sean Augenstein, Neesha Subramaniam, Jason Chang,  
802 Fotis Iliopoulos, Jiaming Luo, Myriam Khan, Weicheng Kuo, Denis Teplyashin, Florence Perot,  
803 Logan Kilpatrick, Amir Globerson, Hongkun Yu, Anfal Siddiqui, Nick Sukhanov, Arun Kandoor,  
804 Umang Gupta, Marco Andreetto, Moran Ambar, Donnie Kim, Paweł Wośowski, Sarah Perrin,  
805 Ben Limonchik, Wei Fan, Jim Stephan, Ian Stewart-Binks, Ryan Kappedal, Tong He, Sarah Cogan,  
806 Romina Datta, Tong Zhou, Jiayu Ye, Leandro Kieliger, Ana Ramalho, Kyle Kastner, Fabian  
807 Mentzer, Wei-Jen Ko, Arun Suggala, Tianhao Zhou, Shiraz Butt, Hana Strejček, Lior Belenki,  
808 Raghavini Venugopalan, Mingyang Ling, Evgenii Eltyshv, Yunxiao Deng, Geza Kovacs, Mukund  
809 Raghavachari, Hanjun Dai, Tal Schuster, Steven Schwarcz, Richard Nguyen, Arthur Nguyen, Gavin  
Buttimore, Shrestha Basu Mallick, Sudeep Gandhe, Seth Benjamin, Michal Jastrzebski, Le Yan,  
Sugato Basu, Chris Apps, Isabel Edkins, James Allingham, Immanuel Odisho, Tomas Kocisky,  
Jewel Zhao, Linting Xue, Apoorv Reddy, Chrysovalantis Anastasiou, Aviel Atias, Sam Redmond,  
Kieran Milan, Nicolas Heess, Herman Schmit, Allan Dafoe, Daniel Andor, Tynan Gangwani,  
Anca Dragan, Sheng Zhang, Ashyana Kachra, Gang Wu, Siyang Xue, Kevin Aydin, Siqi Liu,

810 Yuxiang Zhou, Mahan Malihi, Austin Wu, Siddharth Gopal, Candice Schumann, Peter Stys,  
811 Alek Wang, Mirek Olšák, Dangyi Liu, Christian Schallhart, Yiran Mao, Demetra Brady, Hao  
812 Xu, Tomas Mery, Chawin Sitawarin, Siva Velusamy, Tom Cobley, Alex Zhai, Christian Walder,  
813 Nitzan Katz, Ganesh Jawahar, Chinmay Kulkarni, Antoine Yang, Adam Paszke, Yinan Wang,  
814 Bogdan Damoc, Zalán Borsos, Ray Smith, Jinning Li, Mansi Gupta, Andrei Kapishnikov, Sushant  
815 Prakash, Florian Luisier, Rishabh Agarwal, Will Grathwohl, Kuangyuan Chen, Kehang Han,  
816 Nikhil Mehta, Andrew Over, Shekoofeh Azizi, Lei Meng, Niccolò Dal Santo, Kelvin Zheng, Jane  
817 Shapiro, Igor Petrovski, Jeffrey Hui, Amin Ghafouri, Jasper Snoek, James Qin, Mandy Jordan,  
818 Caitlin Sikora, Jonathan Malmaud, Yuheng Kuang, Aga Świetlik, Ruoxin Sang, Chongyang Shi,  
819 Leon Li, Andrew Rosenberg, Shubin Zhao, Andy Crawford, Jan-Thorsten Peter, Yun Lei, Xavier  
820 Garcia, Long Le, Todd Wang, Julien Amelot, Dave Orr, Praneeth Kacham, Dana Alon, Gladys  
821 Tyen, Abhinav Arora, James Lyon, Alex Kurakin, Mimi Ly, Theo Guidroz, Zhipeng Yan, Rina  
822 Panigrahy, Pingmei Xu, Thais Kagohara, Yong Cheng, Eric Noland, Jinhyuk Lee, Jonathan Lee,  
823 Cathy Yip, Maria Wang, Efrat Nehoran, Alexander Bykovsky, Zhihao Shan, Ankit Bhagatwala,  
824 Chaochao Yan, Jie Tan, Guillermo Garrido, Dan Ethier, Nate Hurley, Grace Vesom, Xu Chen,  
825 Siyuan Qiao, Abhishek Nayyar, Julian Walker, Paramjit Sandhu, Mihaela Rosca, Danny Swisher,  
826 Mikhail Dektiarev, Josh Dillon, George-Cristian Muraru, Manuel Tragut, Artiom Myaskovsky,  
827 David Reid, Marko Velic, Owen Xiao, Jasmine George, Mark Brand, Jing Li, Wenhao Yu, Shane  
828 Gu, Xiang Deng, François-Xavier Aubet, Soheil Hassas Yeganeh, Fred Alcober, Celine Smith,  
829 Trevor Cohn, Kay McKinney, Michael Tschannen, Ramesh Sampath, Gowoon Cheon, Liangchen  
830 Luo, Luyang Liu, Jordi Orbay, Hui Peng, Gabriela Botea, Xiaofan Zhang, Charles Yoon, Cesar  
831 Magalhaes, Paweł Stradomski, Ian Mackinnon, Steven Hemingray, Kumaran Venkatesan, Rhys  
832 May, Jaeyoun Kim, Alex Druinsky, Jingchen Ye, Zheng Xu, Terry Huang, Jad Al Abdallah, Adil  
833 Dostmohamed, Rachana Fellingner, Tsendsuren Munkhdalai, Akanksha Maurya, Peter Garst, Yin  
834 Zhang, Maxim Krikun, Simon Bucher, Aditya Srikanth Veerubhotla, Yaxin Liu, Sheng Li, Nishesh  
835 Gupta, Jakub Adamek, Hanwen Chen, Bernett Orlando, Aleksandr Zaks, Joost van Amersfoort,  
836 Josh Camp, Hui Wan, HyunJeong Choe, Zhichun Wu, Kate Olszewska, Weiren Yu, Archita Vadali,  
837 Martin Scholz, Daniel De Freitas, Jason Lin, Amy Hua, Xin Liu, Frank Ding, Yichao Zhou, Boone  
838 Severson, Katerina Tsihlias, Samuel Yang, Tammo Spalink, Varun Yerram, Helena Pankov, Rory  
839 Blevins, Ben Vargas, Sarthak Jauhari, Matt Miecnikowski, Ming Zhang, Sandeep Kumar, Clement  
840 Farabet, Charline Le Lan, Sebastian Flennerhag, Yonatan Bitton, Ada Ma, Arthur Brażinskas,  
841 Eli Collins, Niharika Ahuja, Sneha Kudugunta, Anna Bortsova, Minh Giang, Wanzheng Zhu,  
842 Ed Chi, Scott Lundberg, Alexey Stern, Subha Puttagunta, Jing Xiong, Xiao Wu, Yash Pande,  
843 Amit Jhinal, Daniel Murphy, Jon Clark, Marc Brockschmidt, Maxine Deines, Kevin R. McKee,  
844 Dan Bahir, Jiajun Shen, Minh Truong, Daniel McDuff, Andrea Gesmundo, Edouard Rosseel,  
845 Bowen Liang, Ken Caluwaerts, Jessica Hamrick, Joseph Kready, Mary Cassin, Rishikesh Ingale,  
846 Li Lao, Scott Pollom, Yifan Ding, Wei He, Lizzetth Bellot, Joana Iljazi, Ramya Sree Boppana,  
847 Shan Han, Tara Thompson, Amr Khalifa, Anna Bulanova, Blagoj Mitrevski, Bo Pang, Emma  
848 Cooney, Tian Shi, Rey Coaguila, Tamar Yakar, Marc’ aurelio Ranzato, Nikola Momchev, Chris  
849 Rawles, Zachary Charles, Young Maeng, Yuan Zhang, Rishabh Bansal, Xiaokai Zhao, Brian  
850 Albert, Yuan Yuan, Sudheendra Vijayanarasimhan, Roy Hirsch, Vinay Ramasesh, Kiran Vodrahalli,  
851 Xingyu Wang, Arushi Gupta, DJ Strouse, Jianmo Ni, Roma Patel, Gabe Taubman, Zhouyuan  
852 Huo, Dero Gharibian, Marianne Monteiro, Hoi Lam, Shobha Vasudevan, Aditi Chaudhary, Isabela  
853 Albuquerque, Kilol Gupta, Sebastian Riedel, Chaitra Hegde, Avraham Ruderman, András György,  
854 Marcus Wainwright, Ashwin Chaugule, Burcu Karagol Ayan, Tomer Levinboim, Sam Shleifer,  
855 Yogesh Kalley, Vahab Mirrokni, Abhishek Rao, Prabakar Radhakrishnan, Jay Hartford, Jialin  
856 Wu, Zhenhai Zhu, Francesco Bertolini, Hao Xiong, Nicolas Serrano, Hamish Tomlinson, Myle  
857 Ott, Yifan Chang, Mark Graham, Jian Li, Marco Liang, Xiangzhu Long, Sebastian Borgeaud,  
858 Yanif Ahmad, Alex Grills, Diana Mincu, Martin Izzard, Yuan Liu, Jinyu Xie, Louis O’ Bryan,  
859 Sameera Ponda, Simon Tong, Michelle Liu, Dan Malkin, Khalid Salama, Yuankai Chen, Rohan  
860 Anil, Anand Rao, Rigel Swavely, Misha Bilenko, Nina Anderson, Tat Tan, Jing Xie, Xing Wu,  
861 Lijun Yu, Oriol Vinyals, Andrey Ryabtsev, Rumens Dangovski, Kate Baumli, Daniel Keysers,  
862 Christian Wright, Zoe Ashwood, Betty Chan, Artem Shtefan, Yaohui Guo, Ankur Bapna, Radu  
863 Soricut, Steven Pecht, Sabela Ramos, Rui Wang, Jiahao Cai, Trieu Trinh, Paul Barham, Linda  
864 Friso, Eli Stickgold, Xiangzhuo Ding, Siamak Shakeri, Diego Ardila, Eleftheria Briakou, Phil  
865 Culliton, Adam Raveret, Jingyu Cui, David Saxton, Subhrajit Roy, Javad Azizi, Pengcheng Yin,  
866 Lucia Loher, Andrew Bunner, Min Choi, Faruk Ahmed, Eric Li, Yin Li, Shengyang Dai, Michael  
867 Elabd, Sriram Ganapathy, Shivani Agrawal, Yiqing Hua, Paige Kunkle, Sujevan Rajayogam, Arun  
868 Ahuja, Arthur Conmy, Alex Vasiloff, Parker Beak, Christopher Yew, Jayaram Mudigonda, Bartek

864 Wydrowski, Jon Blanton, Zhengdong Wang, Yann Dauphin, Zhuo Xu, Martin Polacek, Xi Chen,  
865 Hexiang Hu, Pauline Sho, Markus Kunesch, Mehdi Hafezi Manshadi, Eliza Rutherford, Bo Li,  
866 Sissie Hsiao, Iain Barr, Alex Tudor, Matija Kecman, Arsha Nagrani, Vladimir Pchelin, Martin  
867 Sundermeyer, Aishwarya P S, Abhijit Karmarkar, Yi Gao, Grishma Chole, Olivier Bachem, Isabel  
868 Gao, Arturo BC, Matt Dobb, Mauro Verzetti, Felix Hernandez-Campos, Yana Lunts, Matthew  
869 Johnson, Julia Di Trapani, Raphael Koster, Idan Brusilovsky, Binbin Xiong, Megha Mohabey, Han  
870 Ke, Joe Zou, Tea Sabolić, Víctor Campos, John Palowitch, Alex Morris, Linhai Qiu, Pranavraj  
871 Ponnuramu, Fangtao Li, Vivek Sharma, Kiranbir Sodhia, Kaan Tekelioglu, Aleksandr Chuklin,  
872 Madhavi Yenugula, Erika Gemzer, Theofilos Strinopoulos, Sam El-Husseini, Huiyu Wang, Yan  
873 Zhong, Edouard Leurent, Paul Natsev, Weijun Wang, Dre Mahaarachchi, Tao Zhu, Songyou Peng,  
874 Sami Alabed, Cheng-Chun Lee, Anthony Brohan, Arthur Szlam, GS Oh, Anton Kovsharov, Jenny  
875 Lee, Renee Wong, Megan Barnes, Gregory Thornton, Felix Gimeno, Omer Levy, Martin Sevenich,  
876 Melvin Johnson, Jonathan Mallinson, Robert Dadashi, Ziyue Wang, Qingchun Ren, Preethi Lahoti,  
877 Arka Dhar, Josh Feldman, Dan Zheng, Thatcher Ulrich, Liviu Panait, Michiel Blokzijl, Cip  
878 Baetu, Josip Matak, Jitendra Harlalka, Maulik Shah, Tal Marian, Daniel von Dincklage, Cosmo  
879 Du, Ruy Ley-Wild, Bethanie Brownfield, Max Schumacher, Yury Stuken, Shadi Noghbi, Sonal  
880 Gupta, Xiaoqi Ren, Eric Malmi, Felix Weissenberger, Blanca Huergo, Maria Bauza, Thomas  
881 Lampe, Arthur Douillard, Mojtaba Seyedhosseini, Roy Frostig, Zoubin Ghahramani, Kelvin  
882 Nguyen, Kashyap Krishnakumar, Chengxi Ye, Rahul Gupta, Alireza Nazari, Robert Geirhos, Pete  
883 Shaw, Ahmed Eleryan, Dima Damen, Jennimaria Palomaki, Ted Xiao, Qiyin Wu, Quan Yuan,  
884 Phoenix Meadowlark, Matthew Bilotti, Raymond Lin, Mukund Sridhar, Yannick Schroecker,  
885 Da-Woon Chung, Jincheng Luo, Trevor Strohman, Tianlin Liu, Anne Zheng, Jesse Emond, Wei  
886 Wang, Andrew Lampinen, Toshiyuki Fukuzawa, Folawiyi Campbell-Ajala, Monica Roy, James  
887 Lee-Thorp, Lily Wang, Iftekhar Naim, Tony, Nguy ên, Guy Bensusky, Aditya Gupta, Dominika  
888 Rogozińska, Justin Fu, Thanumalayan Sankaranarayanan Pillai, Petar Veličković, Shahar Drath,  
889 Philipp Neubeck, Vaibhav Tulsyan, Arseniy Klimovskiy, Don Metzler, Sage Stevens, Angel  
890 Yeh, Junwei Yuan, Tianhe Yu, Kelvin Zhang, Alec Go, Vincent Tsang, Ying Xu, Andy Wan,  
891 Isaac Galatzer-Levy, Sam Sobell, Abodunrinwa Toki, Elizabeth Salesky, Wenlei Zhou, Diego  
892 Antognini, Sholto Douglas, Shimu Wu, Adam Lelkes, Frank Kim, Paul Cavallaro, Ana Salazar,  
893 Yuchi Liu, James Besley, Tiziana Refice, Yiling Jia, Zhang Li, Michal Sokolik, Arvind Kannan,  
894 Jon Simon, Jo Chick, Avia Aharon, Meet Gandhi, Mayank Daswani, Keyvan Amiri, Vighnesh  
895 Birodkar, Abe Ittycheriah, Peter Grabowski, Oscar Chang, Charles Sutton, Zhixin, Lai, Umesh  
896 Telang, Susie Sargsyan, Tao Jiang, Raphael Hoffmann, Nicole Brichtova, Matteo Hessel, Jonathan  
897 Halcrow, Sammy Jerome, Geoff Brown, Alex Tomala, Elena Buchatskaya, Dian Yu, Sachit  
898 Menon, Pol Moreno, Yuguo Liao, Vicky Zayats, Luming Tang, SQ Mah, Ashish Shenoy, Alex  
899 Siegman, Majid Hadian, Okwan Kwon, Tao Tu, Nima Khajehnouri, Ryan Foley, Parisa Haghani,  
900 Zhongru Wu, Vaishakh Keshava, Khyatti Gupta, Tony Bruguier, Rui Yao, Danny Karmon, Luisa  
901 Zintgraf, Zhicheng Wang, Enrique Piqueras, Junehyuk Jung, Jenny Brennan, Diego Machado,  
902 Marissa Giustina, MH Tessler, Kamyu Lee, Qiao Zhang, Joss Moore, Kaspar Daugaard, Alexander  
903 Frömmgen, Jennifer Beattie, Fred Zhang, Daniel Kasenberg, Ty Geri, Danfeng Qin, Gaurav Singh  
904 Tomar, Tom Ouyang, Tianli Yu, Luowei Zhou, Rajiv Mathews, Andy Davis, Yaoyiran Li, Jai  
905 Gupta, Damion Yates, Linda Deng, Elizabeth Kemp, Ga-Young Joong, Sergei Vassilvitskii, Mandy  
906 Guo, Pallavi LV, Dave Dopson, Sami Lachgar, Lara McConnaughey, Himadri Choudhury, Dragos  
907 Dena, Aaron Cohen, Joshua Ainslie, Sergey Levi, Parthasarathy Gopavarapu, Polina Zablotskaia,  
908 Hugo Vallet, Sanaz Bahargam, Xiaodan Tang, Nenad Tomasev, Ethan Dyer, Daniel Balle, Hongrae  
909 Lee, William Bono, Jorge Gonzalez Mendez, Vadim Zubov, Shentao Yang, Ivor Rendulic, Yanyan  
910 Zheng, Andrew Hogue, Golan Pundak, Ralph Leith, Avishkar Bhoopchand, Michael Han, Mislav  
911 Žanić, Tom Schaul, Manolis Delakis, Tejas Iyer, Guanyu Wang, Harman Singh, Abdelrahman  
912 Abdelhamed, Tara Thomas, Siddhartha Brahma, Hilal Dib, Naveen Kumar, Wenxuan Zhou, Liang  
913 Bai, Pushkar Mishra, Jiao Sun, Valentin Anklin, Roykronk Sukkerd, Lauren Agubuzu, Anton  
914 Briukhov, Anmol Gulati, Maximilian Sieb, Fabio Pardo, Sara Nasso, Junquan Chen, Kexin Zhu,  
915 Tiberiu Sosea, Alex Goldin, Keith Rush, Spurthi Amba Hombaiah, Andreas Noever, Allan Zhou,  
916 Sam Haves, Mary Phuong, Jake Ades, Yi ting Chen, Lin Yang, Joseph Pagadora, Stan Bileschi,  
917 Victor Cotruta, Rachel Saputro, Arijit Pramanik, Sean Ammirati, Dan Garrette, Kevin Vilella, Tim  
Blyth, Canfer Akbulut, Neha Jha, Alban Rustemi, Arissa Wongpanich, Chirag Nagpal, Yonghui  
Wu, Morgane Rivière, Sergey Kishchenko, Pranesh Srinivasan, Alice Chen, Animesh Sinha, Trang  
Pham, Bill Jia, Tom Hennigan, Anton Bakalov, Nithya Attaluri, Drew Garmon, Daniel Rodriguez,  
Dawid Wegner, Wenhao Jia, Evan Senter, Noah Fiedel, Denis Petek, Yuchuan Liu, Cassidy Hardin,  
Harshal Tushar Lehri, Joao Carreira, Sara Smoot, Marcel Prasetya, Nami Akazawa, Anca Stefanoiu,

918 Chia-Hua Ho, Anelia Angelova, Kate Lin, Min Kim, Charles Chen, Marcin Sieniek, Alice Li,  
919 Tongfei Guo, Sorin Baltateanu, Pouya Tafti, Michael Wunder, Nadav Olmert, Divyansh Shukla,  
920 Jingwei Shen, Neel Kovelamudi, Balaji Venkatraman, Seth Neel, Romal Thoppilan, Jerome Connor,  
921 Frederik Benzing, Axel Stjerngren, Golnaz Ghiasi, Alex Polozov, Joshua Howland, Theophane  
922 Weber, Justin Chiu, Ganesh Poomal Girirajan, Andreas Terzis, Pidong Wang, Fangda Li, Yoav Ben  
923 Shalom, Dinesh Tewari, Matthew Denton, Roece Aharoni, Norbert Kalb, Heri Zhao, Junlin Zhang,  
924 Angelos Filos, Matthew Rahtz, Lalit Jain, Connie Fan, Vitor Rodrigues, Ruth Wang, Richard  
925 Shin, Jacob Austin, Roman Ring, Mariella Sanchez-Vargas, Mehadi Hassen, Ido Kessler, Uri Alon,  
926 Gufeng Zhang, Wenhui Chen, Yenai Ma, Xiance Si, Le Hou, Azalia Mirhoseini, Marc Wilson,  
927 Geoff Bacon, Becca Roelofs, Lei Shu, Gautam Vasudevan, Jonas Adler, Artur Dwornik, Tayfun  
928 Terzi, Matt Lawlor, Harry Askham, Mike Bernico, Xuanyi Dong, Chris Hidey, Kevin Kilgour,  
929 Gaël Liu, Surya Bhupatiraju, Luke Leonhard, Siqi Zuo, Partha Talukdar, Qing Wei, Aliaksei  
930 Severyn, Vít Listík, Jong Lee, Aditya Tripathi, SK Park, Yossi Matias, Hao Liu, Alex Ruiz, Rajesh  
931 Jayaram, Jackson Tolins, Pierre Marcenac, Yiming Wang, Bryan Seybold, Henry Prior, Deepak  
932 Sharma, Jack Weber, Mikhail Sirotenko, Yunhsuan Sung, Dayou Du, Ellie Pavlick, Stefan Zinke,  
933 Markus Freitag, Max Dylla, Montse Gonzalez Arenas, Natan Potikha, Omer Goldman, Connie  
934 Tao, Rachita Chhaparia, Maria Voitovich, Pawan Dogra, Andrija Ražnatović, Zak Tsai, Chong  
935 You, Olesar Johnson, George Tucker, Chenjie Gu, Jae Yoo, Maryam Majzoubi, Valentin Gabeur,  
936 Bahram Raad, Rocky Rhodes, Kashyap Kolipaka, Heidi Howard, Geta Sampemane, Benny Li,  
937 Chulayuth Asawaroengchai, Duy Nguyen, Chiyuan Zhang, Timothee Cour, Xinxin Yu, Zhao Fu,  
938 Joe Jiang, Po-Sen Huang, Gabriela Surita, Iñaki Iurrate, Yael Karov, Michael Collins, Martin  
939 Baeuml, Fabian Fuchs, Shilpa Shetty, Swaroop Ramaswamy, Sayna Ebrahimi, Qiuchen Guo,  
940 Jeremy Shar, Gabe Barth-Maron, Sravanti Addepalli, Bryan Richter, Chin-Yi Cheng, Eugénie  
941 Rives, Fei Zheng, Johannes Griesser, Nishanth Dikkala, Yoel Zeldes, Ilkin Safarli, Dipanjan Das,  
942 Himanshu Srivastava, Sadh MNM Khan, Xin Li, Aditya Pandey, Larisa Markeeva, Dan Belov, Qiqi  
943 Yan, Mikolaj Rybiński, Tao Chen, Megha Nawhal, Michael Quinn, Vineetha Govindaraj, Sarah  
944 York, Reed Roberts, Roopal Garg, Namrata Godbole, Jake Abernethy, Anil Das, Lam Nguyen  
945 Thiet, Jonathan Tompson, John Nham, Neera Vats, Ben Caine, Wesley Helmholtz, Francesco  
946 Pongetti, Yeongil Ko, James An, Clara Huiyi Hu, Yu-Cheng Ling, Julia Pawar, Robert Leland,  
947 Keisuke Nishinaka, Waleed Khawaja, Marco Selvi, Eugene Ie, Danila Sinopalnikov, Lev Proleev,  
948 Nilesh Tripuraneni, Michele Bevilacqua, Seungji Lee, Clayton Sanford, Dan Suh, Dustin Tran,  
949 Jeff Dean, Simon Baumgartner, Jens Heitkaemper, Sagar Gubbi, Kristina Toutanova, Yichong Xu,  
950 Chandu Thekkath, Keran Rong, Palak Jain, Annie Xie, Yan Virin, Yang Li, Lubo Litchev, Richard  
951 Powell, Tarun Bharti, Adam Kraft, Nan Hua, Marissa Ikonomidis, Ayal Hitron, Sanjiv Kumar,  
952 Loic Matthey, Sophie Bridgers, Lauren Lax, Ishaan Malhi, Ondrej Skopek, Ashish Gupta, Jiawei  
953 Cao, Mitchell Rasquinha, Siim Pöder, Wojciech Stokowiec, Nicholas Roth, Guowang Li, Michaël  
954 Sander, Joshua Kessinger, Vihan Jain, Edward Loper, Wonpyo Park, Michal Yarom, Liqun Cheng,  
955 Guru Guruganesh, Kanishka Rao, Yan Li, Catarina Barros, Mikhail Sushkov, Chun-Sung Ferng,  
956 Rohin Shah, Ophir Aharoni, Ravin Kumar, Tim McConnell, Peiran Li, Chen Wang, Fernando  
957 Pereira, Craig Swanson, Fayaz Jamil, Yan Xiong, Anitha Vijayakumar, Prakash Shroff, Kedar  
958 Soparkar, Jindong Gu, Livio Baldini Soares, Eric Wang, Kushal Majmundar, Aurora Wei, Kai  
959 Bailey, Nora Kassner, Chizu Kawamoto, Goran Žužić, Victor Gomes, Abhirut Gupta, Michael  
960 Guzman, Ishita Dasgupta, Xinyi Bai, Zhufeng Pan, Francesco Piccinno, Hadas Natalie Vogel,  
961 Octavio Ponce, Adrian Hutter, Paul Chang, Pan-Pan Jiang, Ionel Gog, Vlad Ionescu, James  
962 Manyika, Fabian Pedregosa, Harry Ragan, Zach Behrman, Ryan Mullins, Coline Devin, Aroonlok  
963 Pyne, Swapnil Gawde, Martin Chadwick, Yiming Gu, Sasan Tavakkol, Andy Twigg, Naman  
964 Goyal, Ndidi Elue, Anna Goldie, Srinivasan Venkatachary, Hongliang Fei, Ziqiang Feng, Marvin  
965 Ritter, Isabel Leal, Sudeep Dasari, Pei Sun, Alif Raditya Rochman, Brendan O'Donoghue, Yuchen  
966 Liu, Jim Sproch, Kai Chen, Natalie Clay, Slav Petrov, Sailesh Sidhwani, Ioana Mihailescu, Alex  
967 Panagopoulos, AJ Piergiovanni, Yunfei Bai, George Powell, Deep Karkhanis, Trevor Yacovone,  
968 Petr Mitrichev, Joe Kovac, Dave Uthus, Amir Yazdanbakhsh, David Amos, Steven Zheng, Bing  
969 Zhang, Jin Miao, Bhuvana Ramabhadran, Soroush Radpour, Shantanu Thakoor, Josh Newlan, Oran  
970 Lang, Orion Jankowski, Shikhar Bharadwaj, Jean-Michel Sarr, Shereen Ashraf, Sneha Mondal, Jun  
971 Yan, Ankit Singh Rawat, Sarmishta Velury, Greg Kochanski, Tom Eccles, Franz Och, Abhanshu  
Sharma, Ethan Mahintorabi, Alex Gurney, Carrie Muir, Vered Cohen, Saksham Thakur, Adam  
Bloniarz, Asier Mujika, Alexander Pritzel, Paul Caron, Altaf Rahman, Fiona Lang, Yasumasa Onoe,  
Petar Sirkovic, Jay Hoover, Ying Jian, Pablo Duque, Arun Narayanan, David Soergel, Alex Haig,  
Loren Maggiore, Shyamal Buch, Josef Dean, Ilya Figotin, Igor Karpov, Shaleen Gupta, Denny  
Zhou, Muhuan Huang, Ashwin Vaswani, Christopher Semturs, Kaushik Shivakumar, Yu Watanabe,

972 Vinodh Kumar Rajendran, Eva Lu, Yanhan Hou, Wenting Ye, Shikhar Vashishth, Nana Nti, Vytenis  
 973 Sakenas, Darren Ni, Doug DeCarlo, Michael Bendersky, Sumit Bagri, Nacho Cano, Elijah Peake,  
 974 Simon Tokumine, Varun Godbole, Carlos Guía, Tanya Lando, Vittorio Selo, Seher Ellis, Danny  
 975 Tarlow, Daniel Gillick, Alessandro Epasto, Siddhartha Reddy Jonnalagadda, Meng Wei, Meiyang  
 976 Xie, Ankur Taly, Michela Paganini, Mukund Sundararajan, Daniel Toyama, Ting Yu, Dessie  
 977 Petrova, Aneesh Pappu, Rohan Agrawal, Senaka Buthpitiya, Justin Frye, Thomas Buschmann,  
 978 Remi Crocker, Marco Tagliasacchi, Mengchao Wang, Da Huang, Sagi Perel, Brian Wieder, Hideto  
 979 Kazawa, Weiyue Wang, Jeremy Cole, Himanshu Gupta, Ben Golan, Seojin Bang, Nitish Kulkarni,  
 980 Ken Franko, Casper Liu, Doug Reid, Sid Dalmia, Jay Whang, Kevin Cen, Prasha Sundaram, Johan  
 981 Ferret, Berivan Isik, Lucian Ionita, Guan Sun, Anna Shekhawat, Muqthar Mohammad, Philip  
 982 Pham, Ronny Huang, Karthik Raman, Xingyi Zhou, Ross Mcilroy, Austin Myers, Sheng Peng,  
 983 Jacob Scott, Paul Covington, Sofia Erell, Pratik Joshi, João Gabriel Oliveira, Natasha Noy, Tajwar  
 984 Nasir, Jake Walker, Vera Axelrod, Tim Dozat, Pu Han, Chun-Te Chu, Eugene Weinstein, Anand  
 985 Shukla, Shreyas Chandrakaladharan, Petra Poklukar, Bonnie Li, Ye Jin, Prem Eruvbetine, Steven  
 986 Hansen, Avigail Dabush, Alon Jacovi, Samrat Phatale, Chen Zhu, Steven Baker, Mo Shomrat, Yang  
 987 Xiao, Jean Pouget-Abadie, Mingyang Zhang, Fanny Wei, Yang Song, Helen King, Yiling Huang,  
 988 Yun Zhu, Ruoxi Sun, Juliana Vicente Franco, Chu-Cheng Lin, Sho Arora, Hui, Li, Vivian Xia,  
 989 Luke Vilnis, Mariano Schain, Kaiz Alarakyia, Laurel Prince, Aaron Phillips, Caleb Habtegebriel,  
 990 Luyao Xu, Huan Gui, Santiago Ontanon, Lora Aroyo, Karan Gill, Peggy Lu, Yash Katariya,  
 991 Dhruv Madeka, Shankar Krishnan, Shubha Srinivas Raghvendra, James Freedman, Yi Tay, Gaurav  
 992 Menghani, Peter Choy, Nishita Shetty, Dan Abolafia, Doron Kukliansky, Edward Chou, Jared  
 993 Lichtarge, Ken Burke, Ben Coleman, Dee Guo, Larry Jin, Indro Bhattacharya, Victoria Langston,  
 994 Yiming Li, Suyog Kotecha, Alex Yakubovich, Xinyun Chen, Petre Petrov, Tolly Powell, Yanzhang  
 995 He, Corbin Quick, Kanav Garg, Dawsen Hwang, Yang Lu, Srinadh Bhojanapalli, Kristian Kjems,  
 996 Ramin Mehran, Aaron Archer, Hado van Hasselt, Ashwin Balakrishna, JK Kearns, Meiqi Guo,  
 997 Jason Riesa, Mikita Sazanovich, Xu Gao, Chris Sauer, Chengrun Yang, XiangHai Sheng, Thomas  
 998 Jimma, Wouter Van Gansbeke, Vitaly Nikolaev, Wei Wei, Katie Millican, Ruizhe Zhao, Justin  
 999 Snyder, Levent Bolelli, Maura O'Brien, Shawn Xu, Fei Xia, Wentao Yuan, Arvind Neelakantan,  
 1000 David Barker, Sachin Yadav, Hannah Kirkwood, Farooq Ahmad, Joel Wee, Jordan Grimstad, Boyu  
 1001 Wang, Matthew Wiethoff, Shane Settle, Miaosen Wang, Charles Blundell, Jingjing Chen, Chris  
 1002 Duvarney, Grace Hu, Olaf Ronneberger, Alex Lee, Yuanzhen Li, Abhishek Chakladar, Alena  
 1003 Butryna, Georgios Evangelopoulos, Guillaume Desjardins, Jonni Kanerva, Henry Wang, Averi  
 1004 Nowak, Nick Li, Alyssa Loo, Art Khurshudov, Laurent El Shafey, Nagabhushan Baddi, Karel Lenc,  
 1005 Yasaman Razeghi, Tom Lieber, Amer Sinha, Xiao Ma, Yao Su, James Huang, Asahi Ushio, Hanna  
 1006 Klimczak-Plucińska, Kareem Mohamed, JD Chen, Simon Osindero, Stav Ginzburg, Lampros  
 1007 Lamprou, Vasilisa Bashlovkina, Duc-Hieu Tran, Ali Khodaei, Ankit Anand, Yixian Di, Ramy  
 1008 Eskander, Manish Reddy Vuyyuru, Jasmine Liu, Aishwarya Kamath, Roman Goldenberg, Mathias  
 1009 Bellaiche, Juliette Pluto, Bill Rosgen, Hassan Mansoor, William Wong, Suhas Ganesh, Eric Bailey,  
 1010 Scott Baird, Dan Deutsch, Jinoo Baek, Xuhui Jia, Chansoo Lee, Abe Friesen, Nathaniel Braun, Kate  
 1011 Lee, Amayika Panda, Steven M. Hernandez, Duncan Williams, Jianqiao Liu, Ethan Liang, Arnaud  
 1012 Autef, Emily Pitler, Deepali Jain, Phoebe Kirk, Oskar Bunyan, Jaume Sanchez Elias, Tongxin Yin,  
 1013 Machel Reid, Aedan Pope, Nikita Putikhin, Bidisha Samanta, Sergio Guadarrama, Dahun Kim,  
 1014 Simon Rowe, Marcella Valentine, Geng Yan, Alex Salcianu, David Silver, Gan Song, Richa Singh,  
 1015 Shuai Ye, Hannah DeBalsi, Majd Al Mery, Eran Ofek, Albert Webson, Shibl Mourad, Ashwin  
 1016 Kakarla, Silvio Lattanzi, Nick Roy, Evgeny Sluzhaev, Christina Butterfield, Alessio Tonioni,  
 1017 Nathan Waters, Sudhindra Kopalle, Jason Chase, James Cohan, Girish Ramchandra Rao, Robert  
 1018 Berry, Michael Voznesensky, Shuguang Hu, Kristen Chiafullo, Sharat Chikkerur, George Scrivener,  
 1019 Ivy Zheng, Jeremy Wiesner, Wolfgang Macherey, Timothy Lillicrap, Fei Liu, Brian Walker, David  
 1020 Welling, Elinor Davies, Yangsibo Huang, Lijie Ren, Nir Shabat, Alessandro Agostini, Mariko  
 1021 Iinuma, Dustin Zelle, Rohit Sathyanarayana, Andrea D'olimpio, Morgan Redshaw, Matt Ginsberg,  
 1022 Ashwin Murthy, Mark Geller, Tatiana Matejovicova, Ayan Chakrabarti, Ryan Julian, Christine  
 1023 Chan, Qiong Hu, Daniel Jarrett, Manu Agarwal, Jeshwanth Challagundla, Tao Li, Sandeep Tata,  
 1024 Wen Ding, Maya Meng, Zhuyun Dai, Giulia Vezzani, Shefali Garg, Jannis Bulian, Mary Jasarevic,  
 1025 Honglong Cai, Harish Rajamani, Adam Santoro, Florian Hartmann, Chen Liang, Bartek Perz,  
 Apoorv Jindal, Fan Bu, Sungyong Seo, Ryan Poplin, Adrian Goedeckemeyer, Badih Ghazi, Nikhil  
 Khadke, Leon Liu, Kevin Mather, Mingda Zhang, Ali Shah, Alex Chen, Jinliang Wei, Keshav  
 Shivam, Yuan Cao, Donghyun Cho, Angelo Scorza Scarpato, Michael Moffitt, Clara Barbu, Ivan  
 Jurin, Ming-Wei Chang, Hongbin Liu, Hao Zheng, Shachi Dave, Christine Kaeser-Chen, Xiaobin  
 Yu, Alvin Abdagic, Lucas Gonzalez, Yanping Huang, Peilin Zhong, Cordelia Schmid, Bryce

- 1026 Petrini, Alex Wertheim, Jifan Zhu, Hoang Nguyen, Kaiyang Ji, Yanqi Zhou, Tao Zhou, Fangxiaoyu  
 1027 Feng, Regev Cohen, David Rim, Shubham Milind Phal, Petko Georgiev, Ariel Brand, Yue Ma,  
 1028 Wei Li, Somit Gupta, Chao Wang, Pavel Dubov, Jean Tarbouriech, Kingshuk Majumder, Huijian  
 1029 Li, Norman Rink, Apurv Suman, Yang Guo, Yinghao Sun, Arun Nair, Xiaowei Xu, Mohamed  
 1030 Elhawaty, Rodrigo Cabrera, Guangxing Han, Julian Eisenschlos, Junwen Bai, Yuqi Li, Yamini  
 1031 Bansal, Thibault Sellam, Mina Khan, Hung Nguyen, Justin Mao-Jones, Nikos Parotsidis, Jake  
 1032 Marcus, Cindy Fan, Roland Zimmermann, Yony Kochinski, Laura Graesser, Feryal Behbahani,  
 1033 Alvaro Caceres, Michael Riley, Patrick Kane, Sandra Lefdal, Rob Willoughby, Paul Vicol, Lun  
 1034 Wang, Shujian Zhang, Ashleah Gill, Yu Liang, Gautam Prasad, Soroosh Mariooryad, Mehran  
 1035 Kazemi, Zifeng Wang, Kritika Muralidharan, Paul Voigtlaender, Jeffrey Zhao, Huanjie Zhou,  
 1036 Nina D’Souza, Aditi Mavalankar, Séb Arnold, Nick Young, Obaid Sarvana, Chace Lee, Milad  
 1037 Nasr, Tingting Zou, Seokhwan Kim, Lukas Haas, Kaushal Patel, Neslihan Bulut, David Parkinson,  
 1038 Courtney Biles, Dmitry Kalashnikov, Chi Ming To, Aviral Kumar, Jessica Austin, Alex Greve,  
 1039 Lei Zhang, Megha Goel, Yeqing Li, Sergey Yaroshenko, Max Chang, Abhishek Jindal, Geoff  
 1040 Clark, Hagai Taitelbaum, Dale Johnson, Ofir Roval, Jeongwoo Ko, Anhad Mohanane, Christian  
 1041 Schuler, Shenil Dodhia, Ruichao Li, Kazuki Osawa, Claire Cui, Peng Xu, Rushin Shah, Tao Huang,  
 1042 Ela Gruzewska, Nathan Clement, Mudit Verma, Olcan Sercinoglu, Hai Qian, Viral Shah, Masa  
 1043 Yamaguchi, Abhinit Modi, Takahiro Kosakai, Thomas Strohmman, Junhao Zeng, Beliz Gunel, Jun  
 1044 Qian, Austin Tarango, Krzysztof Jastrzebski, Robert David, Jyn Shan, Parker Schuh, Kunal Lad,  
 1045 Willi Gierke, Mukundan Madhavan, Xinyi Chen, Mark Kurzeja, Rebeca Santamaria-Fernandez,  
 1046 Dawn Chen, Alexandra Cordell, Yuri Chervonyi, Frankie Garcia, Nithish Kannen, Vincent Perot,  
 1047 Nan Ding, Shlomi Cohen-Ganor, Victor Lavrenko, Junru Wu, Georgie Evans, Cicero Nogueira dos  
 1048 Santos, Madhavi Sewak, Ashley Brown, Andrew Hard, Joan Puigcerver, Zeyu Zheng, Yizhong  
 1049 Liang, Evgeny Gladchenko, Reeve Ingle, Uri First, Pierre Sermanet, Charlotte Magister, Mihajlo  
 1050 Velimirović, Sashank Reddi, Susanna Ricco, Eirikur Agustsson, Hartwig Adam, Nir Levine, David  
 1051 Gaddy, Dan Holtmann-Rice, Xuanhui Wang, Ashutosh Sathe, Abhijit Guha Roy, Blaž Bratanič,  
 1052 Alen Carin, Harsh Mehta, Silvano Bonacina, Nicola De Cao, Mara Finkelstein, Verena Rieser,  
 1053 Xinyi Wu, Florent Alché, Dylan Scandinaro, Li Li, Nino Vieillard, Nikhil Sethi, Garrett Tanzer,  
 1054 Zhi Xing, Shibo Wang, Parul Bhatia, Gui Citovsky, Thomas Anthony, Sharon Lin, Tianze Shi,  
 1055 Shoshana Jakobovits, Gena Gibson, Raj Apte, Lisa Lee, Mingqing Chen, Arunkumar Byravan,  
 1056 Petros Maniatis, Kellie Webster, Andrew Dai, Pu-Chin Chen, Jiaqi Pan, Asya Fadeeva, Zach  
 1057 Gleicher, Thang Luong, and Niket Kumar Bhumihar. Gemini 2.5: Pushing the frontier with  
 1058 advanced reasoning, multimodality, long context, and next generation agentic capabilities, 2025.  
 1059 URL <https://arxiv.org/abs/2507.06261>.
- 1058 DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang  
 1059 Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli  
 1060 Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen,  
 1061 Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Haowei Zhang, Honghui Ding,  
 1062 Huajian Xin, Huazuo Gao, Hui Li, Hui Qu, J. L. Cai, Jian Liang, Jianzhong Guo, Jiaqi Ni, Jiashi  
 1063 Li, Jiawei Wang, Jin Chen, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, Junxiao Song,  
 1064 Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang,  
 1065 Lei Xu, Leyi Xia, Liang Zhao, Litong Wang, Liyue Zhang, Meng Li, Miaojun Wang, Mingchuan  
 1066 Zhang, Minghua Zhang, Minghui Tang, Mingming Li, Ning Tian, Panpan Huang, Peiyi Wang,  
 1067 Peng Zhang, Qiancheng Wang, Qihao Zhu, Qinyu Chen, Qiushi Du, R. J. Chen, R. L. Jin, Ruiqi  
 1068 Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, Runxin Xu, Ruoyu Zhang, Ruyi Chen, S. S. Li,  
 1069 Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shaoqing Wu, Shengfeng Ye, Shengfeng Ye,  
 1070 Shirong Ma, Shiyu Wang, Shuang Zhou, Shuiping Yu, Shunfeng Zhou, Shuting Pan, T. Wang,  
 1071 Tao Yun, Tian Pei, Tianyu Sun, W. L. Xiao, Wangding Zeng, Wanbiao Zhao, Wei An, Wen Liu,  
 1072 Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, X. Q. Li, Xiangyue Jin, Xianzu Wang,  
 1073 Xiao Bi, Xiaodong Liu, Xiaohan Wang, Xiaojin Shen, Xiaokang Chen, Xiaokang Zhang, Xiaosha  
 1074 Chen, Xiaotao Nie, Xiaowen Sun, Xiaoxiang Wang, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu,  
 1075 Xingkai Yu, Xinnan Song, Xinxia Shan, Xinyi Zhou, Xinyu Yang, Xinyuan Li, Xuecheng Su,  
 1076 Xuheng Lin, Y. K. Li, Y. Q. Wang, Y. X. Wei, Y. X. Zhu, Yang Zhang, Yanhong Xu, Yanhong  
 1077 Xu, Yanping Huang, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Li, Yaohui Wang, Yi Yu, Yi Zheng,  
 1078 Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Ying Tang, Yishi Piao, Yisong Wang, Yixuan  
 1079 Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yu Wu, Yuan Ou, Yuchen Zhu, Yudian Wang, Yue  
 Gong, Yuheng Zou, Yujia He, Yukun Zha, Yunfan Xiong, Yunxian Ma, Yuting Yan, Yuxiang Luo,  
 Yuxiang You, Yuxuan Liu, Yuyang Zhou, Z. F. Wu, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu,

- 1080 Zhean Xu, Zhen Huang, Zhen Zhang, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhibin Gou,  
1081 Zhicheng Ma, Zhigang Yan, Zhihong Shao, Zhipeng Xu, Zhiyu Wu, Zhongyu Zhang, Zhuoshu  
1082 Li, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Ziyi Gao, and Zizheng Pan.  
1083 Deepseek-v3 technical report, 2025. URL <https://arxiv.org/abs/2412.19437>.
- 1084 Mostafa Dehghani, Josip Djolonga, Basil Mustafa, Piotr Padlewski, Jonathan Heek, Justin Gilmer,  
1085 Andreas Peter Steiner, Mathilde Caron, Robert Geirhos, Ibrahim Alabdulmohsin, et al. Scaling  
1086 vision transformers to 22 billion parameters. In *International Conference on Machine Learning*,  
1087 pp. 7480–7512. PMLR, 2023.
- 1088 Tim Dettmers and Luke Zettlemoyer. The case for 4-bit precision: k-bit inference scaling laws. In  
1089 *International Conference on Machine Learning*, pp. 7750–7774. PMLR, 2023.
- 1090 Elvis Dohmatob, Yunzhen Feng, Pu Yang, Francois Charton, and Julia Kempe. A tale of tails: Model  
1091 collapse as a change of scaling laws, 2024. URL <https://arxiv.org/abs/2402.07043>.
- 1092 Ricardo Dominguez-Olmedo, Florian E. Dorner, and Moritz Hardt. Training on the test task confounds  
1093 evaluation and emergence, 2024. URL <https://arxiv.org/abs/2407.07890>.
- 1094 Katie E Everett, Lechao Xiao, Mitchell Wortsman, Alexander A Alemi, Roman Novak, Peter J Liu,  
1095 Izzeddin Gur, Jascha Sohl-Dickstein, Leslie Pack Kaelbling, Jaehoon Lee, et al. Scaling exponents  
1096 across parameterizations and optimizers. In *International Conference on Machine Learning*, pp.  
1097 12666–12700. PMLR, 2024.
- 1098 Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural  
1099 networks. In *International Conference on Learning Representations*, 2019.
- 1100 Samir Yitzhak Gadre, Georgios Smyrnis, Vaishaal Shankar, Suchin Gururangan, Mitchell Wortsman,  
1101 Rulin Shao, Jean Mercat, Alex Fang, Jeffrey Li, Sedrick Keh, et al. Language models scale reliably  
1102 with over-training and on downstream tasks. *arXiv preprint arXiv:2403.08540*, 2024.
- 1103 Aryo Pradipta Gema, Alexander Hägele, Runjin Chen, Andy Arditi, Jacob Goldman-Wetzler, Kit  
1104 Fraser-Taliente, Henry Sleight, Linda Petrini, Julian Michael, Beatrice Alex, Pasquale Minervini,  
1105 Yanda Chen, Joe Benton, and Ethan Perez. Inverse scaling in test-time compute, 2025. URL  
1106 <https://arxiv.org/abs/2507.14417>.
- 1107 Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes,  
1108 Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang,  
1109 David L. Donoho, and Sanmi Koyejo. Is model collapse inevitable? breaking the curse of  
1110 recursion by accumulating real and synthetic data, 2024. URL <https://arxiv.org/abs/2404.01413>.
- 1111 Behrooz Ghorbani, Orhan Firat, Markus Freitag, Ankur Bapna, Maxim Krikun, Xavier Garcia,  
1112 Ciprian Chelba, and Colin Cherry. Scaling laws for neural machine translation. In *International  
1113 Conference on Learning Representations*, 2021.
- 1114 Mitchell A Gordon, Kevin Duh, and Jared Kaplan. Data and parameter scaling laws for neural  
1115 machine translation. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-  
1116 tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language  
1117 Processing*, pp. 5915–5922, Online and Punta Cana, Dominican Republic, November 2021.  
1118 Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.478. URL <https://aclanthology.org/2021.emnlp-main.478>.
- 1119 Tom Henighan, Jared Kaplan, Mor Katz, Mark Chen, Christopher Hesse, Jacob Jackson, Heewoo  
1120 Jun, Tom B Brown, Prafulla Dhariwal, Scott Gray, et al. Scaling laws for autoregressive generative  
1121 modeling. *arXiv preprint arXiv:2010.14701*, 2020.
- 1122 Danny Hernandez, Jared Kaplan, Tom Henighan, and Sam McCandlish. Scaling laws for transfer,  
1123 2021. URL <https://arxiv.org/abs/2102.01293>.
- 1124 Danny Hernandez, Tom Brown, Tom Conerly, Nova DasSarma, Dawn Drain, Sheer El-Showk, Nelson  
1125 Elhage, Zac Hatfield-Dodds, Tom Henighan, Tristan Hume, et al. Scaling laws and interpretability  
1126 of learning from repeated data. *arXiv preprint arXiv:2205.10487*, 2022.

- 1134 Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad,  
1135 Md. Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou. Deep learning scaling is predictable,  
1136 empirically, 2017. URL <https://arxiv.org/abs/1712.00409>.
- 1137  
1138 Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza  
1139 Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Thomas  
1140 Hennigan, Eric Noland, Katherine Millican, George van den Driessche, Bogdan Damoc, Au-  
1141 relia Guy, Simon Osindero, Karén Simonyan, Erich Elsen, Oriol Vinyals, Jack Rae, and  
1142 Laurent Sifre. An empirical analysis of compute-optimal large language model training.  
1143 In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Ad-  
1144 vances in Neural Information Processing Systems*, volume 35, pp. 30016–30030. Curran Asso-  
1145 ciates, Inc., 2022. URL [https://proceedings.neurips.cc/paper\\_files/paper/  
2022/file/c1e2faff6f588870935f114ebe04a3e5-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/c1e2faff6f588870935f114ebe04a3e5-Paper-Conference.pdf).
- 1146  
1147 Nikolaus H. R. Howe, Ian R. McKenzie, Oskar John Hollinsworth, Michał Zając, Tom Tseng,  
1148 Aaron David Tucker, Pierre-Luc Bacon, and Adam Gleave. Scaling trends in language model  
1149 robustness. In *Forty-second International Conference on Machine Learning*, 2025. URL [https://  
1150 openreview.net/forum?id=tNGdLEL4R0](https://openreview.net/forum?id=tNGdLEL4R0).
- 1151  
1152 Shengding Hu, Xin Liu, Xu Han, Xinrong Zhang, Chaoqun He, Weilin Zhao, Yankai Lin, Ning Ding,  
1153 Zebin Ou, Guoyang Zeng, Zhiyuan Liu, and Maosong Sun. Predicting emergent abilities with  
1154 infinite resolution evaluation, 2024. URL <https://arxiv.org/abs/2310.03262>.
- 1155  
1156 John Hughes, Sara Price, Aengus Lynch, Rylan Schaeffer, Fazl Barez, Sanmi Koyejo, Henry Sleight,  
1157 Erik Jones, Ethan Perez, and Mrinank Sharma. Best-of-n jailbreaking, 2024. URL [https://  
1158 arxiv.org/abs/2412.03556](https://arxiv.org/abs/2412.03556).
- 1159  
1160 Marcus Hutter. Learning curve theory, 2021. URL <https://arxiv.org/abs/2102.04074>.
- 1161  
1162 Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec  
1163 Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv preprint  
1164 arXiv:2412.16720*, 2024.
- 1165  
1166 Minhao Jiang, Ken Ziyu Liu, Ming Zhong, Rylan Schaeffer, Siru Ouyang, Jiawei Han, and Sanmi  
1167 Koyejo. Investigating data contamination for pre-training language models, 2024. URL [https://  
1168 arxiv.org/abs/2401.06059](https://arxiv.org/abs/2401.06059).
- 1169  
1170 Damjan Kalajdzievski. Scaling laws for forgetting when fine-tuning large language models, 2024.  
1171 URL <https://arxiv.org/abs/2401.05605>.
- 1172  
1173 Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott  
1174 Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models.  
1175 *arXiv preprint arXiv:2001.08361*, 2020.
- 1176  
1177 Joshua Kazdan, Rylan Schaeffer, Apratim Dey, Matthias Gerstgrasser, Rafael Rafailov, David L.  
1178 Donoho, and Sanmi Koyejo. Collapse or thrive? perils and promises of synthetic data in a  
1179 self-generating world, 2024. URL <https://arxiv.org/abs/2410.16713>.
- 1180  
1181 Kimi, Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen, Yanru Chen,  
1182 Yuankun Chen, Yutian Chen, Zhuofu Chen, Jialei Cui, Hao Ding, Mengnan Dong, Angang Du,  
1183 Chenzhuang Du, Dikang Du, Yulun Du, Yu Fan, Yichen Feng, Kelin Fu, Bofei Gao, Hongcheng  
1184 Gao, Peizhong Gao, Tong Gao, Xinran Gu, Longyu Guan, Haiqing Guo, Jianhang Guo, Hao  
1185 Hu, Xiaoru Hao, Tianhong He, Weiran He, Wenyang He, Chao Hong, Yangyang Hu, Zhenxing  
1186 Hu, Weixiao Huang, Zhiqi Huang, Zihao Huang, Tao Jiang, Zhejun Jiang, Xinyi Jin, Yongsheng  
1187 Kang, Guokun Lai, Cheng Li, Fang Li, Haoyang Li, Ming Li, Wentao Li, Yanhao Li, Yiwei  
1188 Li, Zhaowei Li, Zheming Li, Hongzhan Lin, Xiaohan Lin, Zongyu Lin, Chengyin Liu, Chenyu  
1189 Liu, Hongzhang Liu, Jingyuan Liu, Junqi Liu, Liang Liu, Shaowei Liu, T. Y. Liu, Tianwei Liu,  
1190 Weizhou Liu, Yangyang Liu, Yibo Liu, Yiping Liu, Yue Liu, Zhengying Liu, Enzhe Lu, Lijun  
1191 Lu, Shengling Ma, Xinyu Ma, Yingwei Ma, Shaoguang Mao, Jie Mei, Xin Men, Yibo Miao,  
1192 Siyuan Pan, Yebo Peng, Ruoyu Qin, Bowen Qu, Zeyu Shang, Lidong Shi, Shengyuan Shi, Feifan  
1193 Song, Jianlin Su, Zhengyuan Su, Xinjie Sun, Flood Sung, Heyi Tang, Jiawen Tao, Qifeng Teng,  
1194 Chensi Wang, Dinglu Wang, Feng Wang, Haiming Wang, Jianzhou Wang, Jiaying Wang, Jinhong

- 1188 Wang, Shengjie Wang, Shuyi Wang, Yao Wang, Yejie Wang, Yiqin Wang, Yuxin Wang, Yuzhi  
1189 Wang, Zhaoji Wang, Zhengtao Wang, Zhexu Wang, Chu Wei, Qianqian Wei, Wenhao Wu, Xingzhe  
1190 Wu, Yuxin Wu, Chenjun Xiao, Xiaotong Xie, Weimin Xiong, Boyu Xu, Jing Xu, Jinjing Xu,  
1191 L. H. Xu, Lin Xu, Suting Xu, Weixin Xu, Xinran Xu, Yangchuan Xu, Ziyao Xu, Junjie Yan,  
1192 Yuzi Yan, Xiaofei Yang, Ying Yang, Zhen Yang, Zhilin Yang, Zonghan Yang, Haotian Yao,  
1193 Xingcheng Yao, Wenjie Ye, Zhuorui Ye, Bohong Yin, Longhui Yu, Enming Yuan, Hongbang Yuan,  
1194 Mengjie Yuan, Haobing Zhan, Dehao Zhang, Hao Zhang, Wanlu Zhang, Xiaobin Zhang, Yangkun  
1195 Zhang, Yizhi Zhang, Yongting Zhang, Yu Zhang, Yutao Zhang, Yutong Zhang, Zheng Zhang,  
1196 Haotian Zhao, Yikai Zhao, Huabin Zheng, Shaojie Zheng, Jianren Zhou, Xinyu Zhou, Zaida Zhou,  
1197 Zhen Zhu, Weiyu Zhuang, and Xinxing Zu. Kimi k2: Open agentic intelligence, 2025. URL  
1198 <https://arxiv.org/abs/2507.20534>.
- 1199 Sneha Kudugunta, Aditya Kusupati, Tim Dettmers, Kaifeng Chen, Inderjit Dhillon, Yulia Tsvetkov,  
1200 Hannaneh Hajishirzi, Sham Kakade, Ali Farhadi, Prateek Jain, et al. Matformer: Nested transformer  
1201 for elastic inference. *arXiv preprint arXiv:2310.07707*, 2023.
- 1202 Tanishq Kumar, Zachary Ankner, Benjamin F Spector, Blake Bordelon, Niklas Muennighoff, Man-  
1203 sheej Paul, Cengiz Pehlevan, Christopher Ré, and Aditi Raghunathan. Scaling laws for precision.  
1204 *arXiv preprint arXiv:2411.04330*, 2024.
- 1205 Jacky Kwok, Christopher Agia, Rohan Sinha, Matt Foutter, Shulu Li, Ion Stoica, Azalia Mirhoseini,  
1206 and Marco Pavone. Robomoney: Scaling test-time sampling and verification for vision-language-  
1207 action models. *arXiv preprint arXiv:2506.17811*, 2025.
- 1208 Noam Itzhak Levi. A simple model of inference scaling laws. In *Forty-second International  
1209 Conference on Machine Learning*, 2025.
- 1210 Seng Pei Liew, Takuya Kato, and Sho Takase. Scaling laws for upcycling mixture-of-experts language  
1211 models. *arXiv preprint arXiv:2502.03009*, 2025.
- 1212 Licong Lin, Jingfeng Wu, Sham M Kakade, Peter L Bartlett, and Jason D Lee. Scaling laws in linear  
1213 regression: Compute, parameters, and data. *arXiv preprint arXiv:2406.08466*, 2024.
- 1214 Jan Ludziejewski, Maciej Pióro, Jakub Krajewski, Maciej Stefaniak, Michał Krutul, Jan Małaśnicki,  
1215 Marek Cygan, Piotr Sankowski, Kamil Adamczewski, Piotr Miłoś, et al. Joint moe scaling laws:  
1216 Mixture of experts can be memory efficient. *arXiv preprint arXiv:2502.05172*, 2025.
- 1217 Alexander Maloney, Daniel A Roberts, and James Sully. A solvable model of neural scaling laws.  
1218 *arXiv preprint arXiv:2210.16859*, 2022.
- 1219 Ryan McKenna, Yangsibo Huang, Amer Sinha, Borja Balle, Zachary Charles, Christopher A  
1220 Choquette-Choo, Badih Ghazi, Georgios Kaissis, Ravi Kumar, Ruibo Liu, et al. Scaling laws for  
1221 differentially private language models. In *Forty-second International Conference on Machine  
1222 Learning*, 2025.
- 1223 Ian R. McKenzie, Alexander Lyzhov, Michael Pieler, Alicia Parrish, Aaron Mueller, Ameya Prabhu,  
1224 Euan McLean, Aaron Kirtland, Alexis Ross, Alisa Liu, Andrew Gritsevskiy, Daniel Wurgaft, Derik  
1225 Kauffman, Gabriel Recchia, Jiacheng Liu, Joe Cavanagh, Max Weiss, Sicong Huang, The Floating  
1226 Droid, Tom Tseng, Tomasz Korbak, Xudong Shen, Yuhui Zhang, Zhengping Zhou, Najoung Kim,  
1227 Samuel R. Bowman, and Ethan Perez. Inverse scaling: When bigger isn't better, 2024. URL  
1228 <https://arxiv.org/abs/2306.09479>.
- 1229 Hrushikesh N Mhaskar. Neural networks for optimal approximation of smooth and analytic functions.  
1230 *Neural computation*, 8(1):164–177, 1996.
- 1231 Niklas Muennighoff, Alexander Rush, Boaz Barak, Teven Le Scao, Nouamane Tazi, Aleksandra  
1232 Piktus, Sampo Pyysalo, Thomas Wolf, and Colin A Raffel. Scaling data-constrained language  
1233 models. *Advances in Neural Information Processing Systems*, 36:50358–50376, 2023.
- 1234 OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni  
1235 Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor  
1236 Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian,  
1237 Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny  
1238

- 1242 Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks,  
1243 Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea  
1244 Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen,  
1245 Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung,  
1246 Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch,  
1247 Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty  
1248 Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte,  
1249 Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel  
1250 Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua  
1251 Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike  
1252 Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon  
1253 Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne  
1254 Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo  
1255 Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar,  
1256 Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik  
1257 Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich,  
1258 Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy  
1259 Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie  
1260 Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini,  
1261 Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne,  
1262 Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David  
1263 Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie  
1264 Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély,  
1265 Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo  
1266 Noh, Long Ouyang, Cullen O’Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano,  
1267 Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng,  
1268 Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto,  
1269 Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power,  
1270 Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis  
1271 Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted  
1272 Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel  
1273 Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon  
1274 Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky,  
1275 Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie  
1276 Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng,  
1277 Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun  
1278 Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang,  
1279 Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian  
1280 Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren  
1281 Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming  
1282 Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao  
1283 Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. Gpt-4 technical report, 2024. URL  
1284 <https://arxiv.org/abs/2303.08774>.
- 1285 Elliot Paquette, Courtney Paquette, Lechao Xiao, and Jeffrey Pennington. 4+ 3 phases of compute-  
1286 optimal neural scaling laws. *arXiv preprint arXiv:2405.15074*, 2024.
- 1287 Tim Pearce and Jinyeop Song. Reconciling kaplan and chinchilla scaling laws. *Transactions on*  
1288 *Machine Learning Research*, 2024.
- 1289 Tim Pearce, Tabish Rashid, David Bignell, Raluca Georgescu, Sam Devlin, and Katja Hofmann.  
1290 Scaling laws for pre-training agents and world models. In *Forty-second International Conference on*  
1291 *Machine Learning*, 2025. URL <https://openreview.net/forum?id=HHwGfL0Kxq>.
- 1292 Allan Pinkus. Approximation theory of the mlp model in neural networks. *Acta numerica*, 8:143–195,  
1293 1999.
- 1294 Tomer Porian, Mitchell Wortsman, Jenia Jitsev, Ludwig Schmidt, and Yair Carmon. Re-  
1295 solving discrepancies in compute-optimal scaling of language models. In A. Globerson,  
L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances*

- 1296 *in Neural Information Processing Systems*, volume 37, pp. 100535–100570. Curran Asso-  
 1297 ciates, Inc., 2024. URL [https://proceedings.neurips.cc/paper\\_files/paper/](https://proceedings.neurips.cc/paper_files/paper/2024/file/b6341525cd84f3be0ef203e4d7cd8556-Paper-Conference.pdf)  
 1298 [2024/file/b6341525cd84f3be0ef203e4d7cd8556-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2024/file/b6341525cd84f3be0ef203e4d7cd8556-Paper-Conference.pdf).  
 1299
- 1300 Ofir Press and Lior Wolf. Using the output embedding to improve language models. In *Proceedings*  
 1301 *of the 15th Conference of the European Chapter of the Association for Computational Linguistics:*  
 1302 *Volume 2, Short Papers*, pp. 157–163, 2017.
- 1303 Zeyu Qin, Qingxiu Dong, Xingxing Zhang, Li Dong, Xiaolong Huang, Ziyi Yang, Mahmoud  
 1304 Khademi, Dongdong Zhang, Hany Hassan Awadalla, Yi R Fung, et al. Scaling laws of synthetic  
 1305 data for language models. *arXiv preprint arXiv:2503.19551*, 2025.
- 1306 Shikai Qiu, Lechao Xiao, Andrew Gordon Wilson, Jeffrey Pennington, and Atish Agarwala. Scaling  
 1307 collapse reveals universal dynamics in compute-optimally trained neural networks. In *Forty-second*  
 1308 *International Conference on Machine Learning*, 2025.
- 1309 Qwen, An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu,  
 1310 Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang,  
 1311 Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang,  
 1312 Jianxin Yang, Jiayi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin  
 1313 Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin  
 1314 Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin,  
 1315 Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang,  
 1316 Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and  
 1317 Zihan Qiu. Qwen3 technical report, 2025. URL <https://arxiv.org/abs/2505.09388>.
- 1318 Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John  
 1319 Aslanides, Sarah Henderson, Roman Ring, Susannah Young, Eliza Rutherford, Tom Hennigan,  
 1320 Jacob Menick, Albin Cassirer, Richard Powell, George van den Driessche, Lisa Anne Hendricks,  
 1321 Maribeth Rauh, Po-Sen Huang, Amelia Glaese, Johannes Welbl, Sumanth Dathathri, Saffron  
 1322 Huang, Jonathan Uesato, John Mellor, Irina Higgins, Antonia Creswell, Nat McAleese, Amy Wu,  
 1323 Erich Elsen, Siddhant Jayakumar, Elena Buchatskaya, David Budden, Esme Sutherland, Karen  
 1324 Simonyan, Michela Paganini, Laurent Sifre, Lena Martens, Xiang Lorraine Li, Adhiguna Kuncoro,  
 1325 Aida Nematzadeh, Elena Gribovskaya, Domenic Donato, Angeliki Lazaridou, Arthur Mensch,  
 1326 Jean-Baptiste Lespiau, Maria Tsimpoukelli, Nikolai Grigorev, Doug Fritz, Thibault Sottiaux,  
 1327 Mantas Pajarskas, Toby Pohlen, Zhitao Gong, Daniel Toyama, Cyprien de Masson d’Autume,  
 1328 Yujia Li, Tayfun Terzi, Vladimir Mikulik, Igor Babuschkin, Aidan Clark, Diego de Las Casas,  
 1329 Aurelia Guy, Chris Jones, James Bradbury, Matthew Johnson, Blake Hechtman, Laura Weidinger,  
 1330 Iason Gabriel, William Isaac, Ed Lockhart, Simon Osindero, Laura Rimell, Chris Dyer, Oriol  
 1331 Vinyals, Kareem Ayoub, Jeff Stanway, Lorraine Bennett, Demis Hassabis, Koray Kavukcuoglu,  
 1332 and Geoffrey Irving. Scaling language models: Methods, analysis & insights from training gopher,  
 1333 2022. URL <https://arxiv.org/abs/2112.11446>.
- 1334 Daniel A Roberts, Sho Yaida, and Boris Hanin. *The principles of deep learning theory*, volume 46.  
 1335 Cambridge University Press Cambridge, MA, USA, 2022.
- 1336 Jonathan S Rosenfeld, Amir Rosenfeld, Yonatan Belinkov, and Nir Shavit. A constructive prediction  
 1337 of the generalization error across scales. In *International Conference on Learning Representations*,  
 1338 2020.
- 1339 Jonathan S Rosenfeld, Jonathan Frankle, Michael Carbin, and Nir Shavit. On the predictability  
 1340 of pruning across scales. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th*  
 1341 *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning*  
 1342 *Research*, pp. 9075–9083. PMLR, 18–24 Jul 2021. URL [https://proceedings.mlr.](https://proceedings.mlr.press/v139/rosenfeld21a.html)  
 1343 [press/v139/rosenfeld21a.html](https://proceedings.mlr.press/v139/rosenfeld21a.html).
- 1344 Ranajoy Sadhukhan, Zhuoming Chen, Haizhong Zheng, Yang Zhou, Emma Strubell, and Beidi Chen.  
 1345 Kinetics: Rethinking test-time scaling laws, 2025. URL [https://arxiv.org/abs/2506.](https://arxiv.org/abs/2506.05333)  
 1346 [05333](https://arxiv.org/abs/2506.05333).
- 1347  
 1348 Nikhil Sardana, Jacob Portes, Sasha Doubov, and Jonathan Frankle. Beyond chinchilla-optimal:  
 1349 Accounting for inference in language model scaling laws. In *International Conference on Machine*  
*Learning*, pp. 43445–43460. PMLR, 2024.

- 1350 Rylan Schaeffer. Pretraining on the test set is all you need, 2023. URL <https://arxiv.org/abs/2309.08632>.  
1351  
1352
- 1353 Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo. Are emergent abilities of large language  
1354 models a mirage? In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- 1355 Rylan Schaeffer, Hailey Schoelkopf, Brando Miranda, Gabriel Mukobi, Varun Madan, Adam Ibrahim,  
1356 Herbie Bradley, Stella Biderman, and Sanmi Koyejo. Why has predicting downstream capabilities  
1357 of frontier ai models with scale remained elusive?, 2024. URL <https://arxiv.org/abs/2406.04391>.  
1358
- 1359 Rylan Schaeffer, Joshua Kazdan, John Hughes, Jordan Juravsky, Sara Price, Aengus Lynch, Erik  
1360 Jones, Robert Kirk, Azalia Mirhoseini, and Sanmi Koyejo. How do large language monkeys get  
1361 their power (laws)? In *Forty-second International Conference on Machine Learning*, 2025.  
1362
- 1363 Utkarsh Sharma and Jared Kaplan. Scaling laws from the data manifold dimension. *Journal of*  
1364 *Machine Learning Research*, 23(9):1–34, 2022.
- 1365 Mustafa Shukor, Enrico Fini, Victor Guilherme Turrisi da Costa, Matthieu Cord, Joshua Susskind,  
1366 and Alaaeldin El-Nouby. Scaling laws for native multimodal models, 2025. URL <https://arxiv.org/abs/2504.07951>.  
1367
- 1368 Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling llm test-time compute optimally  
1369 can be more effective than scaling model parameters. *arXiv preprint arXiv:2408.03314*, 2024a.  
1370
- 1371 Charlie Snell, Eric Wallace, Dan Klein, and Sergey Levine. Predicting emergent capabilities by  
1372 finetuning, 2024b. URL <https://arxiv.org/abs/2411.16035>.  
1373
- 1374 Stefano Spigler, Mario Geiger, and Matthieu Wyart. Asymptotic learning curves of kernel methods:  
1375 empirical data versus teacher–student paradigm. *Journal of Statistical Mechanics: Theory and*  
1376 *Experiment*, 2020(12):124001, December 2020. ISSN 1742-5468. doi: 10.1088/1742-5468/abc61d.  
1377 URL <http://dx.doi.org/10.1088/1742-5468/abc61d>.
- 1378 Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam  
1379 Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska,  
1380 Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W.  
1381 Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda  
1382 Askeell, Amanda Dsouza, Ambrose Slone, Ameet Rahane, Anantharaman S. Iyer, Anders An-  
1383 dreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew Dai, Andrew La,  
1384 Andrew Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna  
1385 Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum, Arul Menezes,  
1386 Arun Kirubakaran, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut  
1387 Erdem, Ayla Karakaş, B. Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej Bojanowski,  
1388 Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno Stein, Berk  
1389 Ekmekci, Bill Yuchen Lin, Blake Howald, Bryan Orinion, Cameron Diao, Cameron Dour, Cather-  
1390 ine Stinson, Cedrick Argueta, César Ferri Ramírez, Chandan Singh, Charles Rathkopf, Chenlin  
1391 Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Chris Waites, Christian Voigt, Christo-  
1392 pher D. Manning, Christopher Potts, Cindy Ramirez, Clara E. Rivera, Clemencia Siro, Colin Raffel,  
1393 Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman,  
1394 Dan Roth, Daniel Freeman, Daniel Khashabi, Daniel Levy, Daniel Moseguí González, Danielle  
1395 Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David  
1396 Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Deniz  
1397 Yuret, Derek Chen, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho  
1398 Mollo, Diyi Yang, Dong-Ho Lee, Dylan Schrader, Ekaterina Shutova, Ekin Dogus Cubuk, Elad  
1399 Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele Rodola,  
1400 Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan A. Chi, Ethan Dyer, Ethan  
1401 Jerzak, Ethan Kim, Eunice Engefu Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar,  
1402 Fernando Martínez-Plumed, Francesca Happé, Francois Chollet, Frieda Rong, Gaurav Mishra,  
1403 Genta Indra Winata, Gerard de Melo, Germán Kruszewski, Giambattista Parascandolo, Giorgio  
Mariani, Gloria Wang, Gonzalo Jaimovitch-López, Gregor Betz, Guy Gur-Ari, Hana Galijasevic,  
Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar, Henry Shevlin,  
Hinrich Schütze, Hiromu Yakura, Hongming Zhang, Hugh Mee Wong, Ian Ng, Isaac Noble, Jaap

- 1404 Jumelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon Lee, Jaime Fernández Fisac,  
 1405 James B. Simon, James Koppel, James Zheng, James Zou, Jan Kocoń, Jana Thompson, Janelle  
 1406 Wingfield, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason Phang, Jason Wei, Jason  
 1407 Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy Kim, Jeroen Taal, Jesse  
 1408 Engel, Jesujoba Alabi, Jiacheng Xu, Jiaming Song, Jillian Tang, Joan Waweru, John Burden,  
 1409 John Miller, John U. Balis, Jonathan Batchelder, Jonathan Berant, Jörg Frohberg, Jos Rozen,  
 1410 Jose Hernandez-Orallo, Joseph Boudeman, Joseph Guerr, Joseph Jones, Joshua B. Tenenbaum,  
 1411 Joshua S. Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth, Karthik Gopalakr-  
 1412 ishnan, Katerina Ignatyeva, Katja Markert, Kaustubh D. Dhole, Kevin Gimpel, Kevin Omondi,  
 1413 Kory Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar, Kyle McDonell, Kyle  
 1414 Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin, Lidia Contreras-  
 1415 Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble, Ludwig Schmidt,  
 1416 Luheng He, Luis Oliveros Colón, Luke Metz, Lütfi Kerem Şenel, Maarten Bosma, Maarten Sap,  
 1417 Maartje ter Hoeve, Maheen Farooqi, Manaal Faruqui, Mantas Mazeika, Marco Baturan, Marco  
 1418 Marelli, Marco Maru, Maria Jose Ramírez Quintana, Marie Tolkiehn, Mario Giulianelli, Martha  
 1419 Lewis, Martin Potthast, Matthew L. Leavitt, Matthias Hagen, Mátyás Schubert, Medina Orduna  
 1420 Baitemirova, Melody Arnaud, Melvin McElrath, Michael A. Yee, Michael Cohen, Michael Gu,  
 1421 Michael Ivanitskiy, Michael Starritt, Michael Strube, Michał Śwędrowski, Michele Bevilacqua,  
 1422 Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimeo Xu, Mirac Suzgun, Mitch Walker, Mo Tiwari,  
 1423 Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhdeh Gheini, Mukund Varma T, Nanyun Peng,  
 1424 Nathan A. Chi, Nayeon Lee, Neta Gur-Ari Krakover, Nicholas Cameron, Nicholas Roberts, Nick  
 1425 Doiron, Nicole Martinez, Nikita Nangia, Niklas Deckers, Niklas Muennighoff, Nitish Shirish  
 1426 Keskar, Niveditha S. Iyer, Noah Constant, Noah Fiedel, Nuan Wen, Oliver Zhang, Omar Agha,  
 1427 Omar Elbaghdadi, Omer Levy, Owain Evans, Pablo Antonio Moreno Casares, Parth Doshi, Pascale  
 1428 Fung, Paul Pu Liang, Paul Vicol, Pegah Alipoormolabashi, Peiyuan Liao, Percy Liang, Peter Chang,  
 1429 Peter Eckersley, Phu Mon Htut, Pinyu Hwang, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour,  
 1430 Priti Oli, Qiaozhu Mei, Qing Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer  
 1431 Gabriel, Rahel Habacker, Ramon Risco, Raphaël Millière, Rhythm Garg, Richard Barnes, Rif A.  
 1432 Saurous, Riku Arakawa, Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman  
 1433 Sitelew, Ronan LeBras, Rosanne Liu, Rowan Jacobs, Rui Zhang, Ruslan Salakhutdinov, Ryan  
 1434 Chi, Ryan Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif M. Mohammad, Sa-  
 1435 jant Anand, Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel R. Bowman,  
 1436 Samuel S. Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah A. Rous, Sarik Ghazarian, Sayan  
 1437 Ghosh, Sean Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi,  
 1438 Shadi Hamdan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi,  
 1439 Shixiang Shane Gu, Shubh Pachchigar, Shubham Toshniwal, Shyam Upadhyay, Shyamolima,  
 1440 Debnath, Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini,  
 1441 Soo-Hwan Lee, Spencer Torene, Sriharsha Hatwar, Stanislas Dehaene, Stefan Divic, Stefano  
 1442 Ermon, Stella Biderman, Stephanie Lin, Stephen Prasad, Steven T. Piantadosi, Stuart M. Shieber,  
 1443 Summer Mishnerghi, Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li,  
 1444 Tao Yu, Tariq Ali, Tatsuo Hashimoto, Te-Lin Wu, Théo Desbordes, Theodore Rothschild, Thomas  
 1445 Phan, Tianle Wang, Tiberius Nkinyili, Timo Schick, Timofei Kornev, Titus Tunduny, Tobias Ger-  
 1446 stenberg, Trenton Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra,  
 1447 Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Ramasesh, Vinay Uday Prabhu, Vishakh  
 1448 Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen,  
 1449 Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair  
 1450 Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Yiding Hao, Yifu Chen, Yonatan  
 1451 Belinkov, Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J.  
 1452 Wang, Zirui Wang, and Ziyi Wu. Beyond the imitation game: Quantifying and extrapolating the  
 1453 capabilities of language models, 2023. URL <https://arxiv.org/abs/2206.04615>.
- 1454 Xingwu Sun, Shuaipeng Li, Ruobing Xie, Weidong Han, Kan Wu, Zhen Yang, Yixing Li, An Wang,  
 1455 Shuai Li, Jinbao Xue, Yu Cheng, Yangyu Tao, Zhanhui Kang, Chengzhong Xu, Di Wang, and Jie  
 1456 Jiang. Scaling laws for floating point quantization training, 2025. URL <https://arxiv.org/abs/2501.02423>.
- 1457 Chaofan Tao, Qian Liu, Longxu Dou, Niklas Muennighoff, Zhongwei Wan, Ping Luo, Min Lin, and  
 Ngai Wong. Scaling laws with vocabulary: Larger models deserve larger vocabularies. *arXiv preprint arXiv:2407.13623*, 2024.

- 1458 Yi Tay, Mostafa Dehghani, Jinfeng Rao, William Fedus, Samira Abnar, Hyung Won Chung, Sharan  
1459 Narang, Dani Yogatama, Ashish Vaswani, and Donald Metzler. Scale efficiently: Insights from  
1460 pre-training and fine-tuning transformers. *arXiv preprint arXiv:2109.10686*, 2021.  
1461
- 1462 Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama,  
1463 Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals,  
1464 Percy Liang, Jeff Dean, and William Fedus. Emergent abilities of large language models, 2022.  
1465 URL <https://arxiv.org/abs/2206.07682>.
- 1466 Tung-Yu Wu and Pei-Yu Lo. U-shaped and inverted-u scaling behind emergent abilities of large  
1467 language models, 2024. URL <https://arxiv.org/abs/2410.01692>.
- 1468 Yangzhen Wu, Zhiqing Sun, Shanda Li, Sean Welleck, and Yiming Yang. Inference scaling laws: An  
1469 empirical analysis of compute-optimal inference for problem-solving with language models, 2024.  
1470 URL <https://arxiv.org/abs/2408.00724>.
- 1471
- 1472 Wenhan Xiong, Jingyu Liu, Igor Molybog, Hejia Zhang, Prajjwal Bhargava, Rui Hou, Louis Martin,  
1473 Rashi Rungta, Karthik Abinav Sankararaman, Barlas Oguz, Madian Khabsa, Han Fang, Yashar  
1474 Mehdad, Sharan Narang, Kshitiz Malik, Angela Fan, Shruti Bhosale, Sergey Edunov, Mike Lewis,  
1475 Sinong Wang, and Hao Ma. Effective long-context scaling of foundation models, 2023. URL  
1476 <https://arxiv.org/abs/2309.16039>.
- 1477 Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers.  
1478 In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.  
1479 12104–12113, 2022.
- 1480
- 1481 Biao Zhang, Zhongtao Liu, Colin Cherry, and Orhan Firat. When scaling meets llm finetuning: The  
1482 effect of data, model and finetuning method, 2024. URL <https://arxiv.org/abs/2402.17193>.
- 1483
- 1484 Susan Zhang. After ignoring the details in all these "lets-fit-a-cloud-of-points-to-a-single-line" papers  
1485 (all likely wrong when you really extrapolate), @stephenroller finally convinced me to work  
1486 through the math in the chinchilla paper and as expected, this was a doozy., January 2023. URL  
1487 <https://x.com/suchenzhang/status/1616752482226671620>.
- 1488
- 1489 Zhipu-AI, Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin Chen, Chengxing Xie, Cunxiang  
1490 Wang, Da Yin, Hao Zeng, Jiajie Zhang, Kedong Wang, Lucen Zhong, Mingdao Liu, Rui Lu, Shulin  
1491 Cao, Xiaohan Zhang, Xuancheng Huang, Yao Wei, Yean Cheng, Yifan An, Yilin Niu, Yuanhao  
1492 Wen, Yushi Bai, Zhengxiao Du, Zihan Wang, Zilin Zhu, Bohan Zhang, Bosi Wen, Bowen Wu,  
1493 Bowen Xu, Can Huang, Casey Zhao, Changpeng Cai, Chao Yu, Chen Li, Chendi Ge, Chenghua  
1494 Huang, Chenhui Zhang, Chenxi Xu, Chenzheng Zhu, Chuang Li, Congfeng Yin, Daoyan Lin,  
1495 Dayong Yang, Dazhi Jiang, Ding Ai, Erle Zhu, Fei Wang, Gengzheng Pan, Guo Wang, Hailong  
1496 Sun, Haitao Li, Haiyang Li, Haiyi Hu, Hanyu Zhang, Hao Peng, Hao Tai, Haoke Zhang, Haoran  
1497 Wang, Haoyu Yang, He Liu, He Zhao, Hongwei Liu, Hongxi Yan, Huan Liu, Huilong Chen, Ji Li,  
1498 Jiajing Zhao, Jiamin Ren, Jian Jiao, Jiani Zhao, Jianyang Yan, Jiaqi Wang, Jiayi Gui, Jiayue Zhao,  
1499 Jie Liu, Jijie Li, Jing Li, Jing Lu, Jingsen Wang, Jingwei Yuan, Jingxuan Li, Jingzhao Du, Jinhua  
1500 Du, Jinxin Liu, Junkai Zhi, Junli Gao, Ke Wang, Lekang Yang, Liang Xu, Lin Fan, Lindong Wu,  
1501 Lintao Ding, Lu Wang, Man Zhang, Minghao Li, Minghuan Xu, Mingming Zhao, Mingshu Zhai,  
1502 Pengfan Du, Qian Dong, Shangde Lei, Shangqing Tu, Shangtong Yang, Shaoyou Lu, Shijie Li,  
1503 Shuang Li, Shuang-Li, Shuxun Yang, Sibao Yi, Tianshu Yu, Wei Tian, Weihang Wang, Wenbo Yu,  
1504 Weng Lam Tam, Wenjie Liang, Wentao Liu, Xiao Wang, Xiaohan Jia, Xiaotao Gu, Xiaoying Ling,  
1505 Xin Wang, Xing Fan, Xingru Pan, Xinyuan Zhang, Xinze Zhang, Xiuqing Fu, Xunkai Zhang, Yabo  
1506 Xu, Yandong Wu, Yida Lu, Yidong Wang, Yilin Zhou, Yiming Pan, Ying Zhang, Yingli Wang,  
1507 Yingru Li, Yinpei Su, Yipeng Geng, Yitong Zhu, Yongkun Yang, Yuhang Li, Yuhao Wu, Yujiang  
1508 Li, Yunan Liu, Yunqing Wang, Yuntao Li, Yuxuan Zhang, Zezhen Liu, Zhen Yang, Zhengda Zhou,  
1509 Zhongpei Qiao, Zhuoer Feng, Zhuorui Liu, Zichen Zhang, Zihan Wang, Zijun Yao, Zikang Wang,  
1510 Ziqiang Liu, Ziwei Chai, Zixuan Li, Zuodong Zhao, Wenguang Chen, Jidong Zhai, Bin Xu, Minlie  
1511 Huang, Hongning Wang, Juanzi Li, Yuxiao Dong, and Jie Tang. Glm-4.5: Agentic, reasoning, and  
coding (arc) foundation models, 2025. URL <https://arxiv.org/abs/2508.06471>.

1512 A LANGUAGE MODEL USAGE

1513

1514 Language models were used by the authors to aid or polish the writing of the paper. Authors take full  
1515 responsibility for the content.

1516

1517

1518

1519

1520

1521

1522

1523

1524

1525

1526

1527

1528

1529

1530

1531

1532

1533

1534

1535

1536

1537

1538

1539

1540

1541

1542

1543

1544

1545

1546

1547

1548

1549

1550

1551

1552

1553

1554

1555

1556

1557

1558

1559

1560

1561

1562

1563

1564

1565

## B CHINCHILLA’S ARCHITECTURAL HYPERPARAMETERS AND MODEL PARAMETERS

Below, we include Hoffmann et al. (2022)’s Table A9 listing model architectural hyperparameters alongside the reported model parameters. We augment the table with the standard formula model parameters (Eqn. 1) and the best fit formula model parameters (Eqn. 3).

Table A9 from Hoffmann et al. (2022)						Our Contribution		
d_model	ffw_size	kv_size	n_heads	n_layers	n_vocab	Chinchilla’s Reported Model Parameters (M)	Best Fit Formula’s Model Parameters (M)	Standard Formula’s Model Parameters (M)
512	2048	64	8	8	32168	44	44	42
576	2304	64	9	9	32168	57	57	54
640	2560	64	10	10	32168	74	74	70
640	2560	64	10	13	32168	90	90	84
640	2560	64	10	16	32168	106	106	99
768	3072	64	12	12	32168	117	117	110
768	3072	64	12	15	32168	140	140	131
768	3072	64	12	18	32168	163	163	152
896	3584	64	14	14	32168	175	175	164
896	3584	64	14	16	32168	196	196	183
896	3584	64	14	18	32168	217	217	202
1024	4096	64	16	16	32168	251	251	234
1024	4096	64	16	18	32168	278	278	259
1024	4096	64	16	20	32168	306	306	285
1280	5120	128	10	18	32168	425	425	395
1280	5120	128	10	21	32168	489	488	454
1408	5632	128	11	18	32168	509	509	474
1280	5120	128	10	24	32168	552	552	513
1408	5632	128	11	21	32168	587	587	545
1536	6144	128	12	19	32168	632	632	587
1408	5632	128	11	25	32168	664	690	640
1536	6144	128	12	22	32168	724	724	672
1536	6144	128	12	23	32168	816	755	701
1792	7168	128	14	20	32168	893	893	828
1792	7168	128	14	22	32168	1018	976	905
1792	7168	128	14	26	32168	1143	1143	1060
2048	8192	128	16	20	32168	1266	1156	1073
2176	8704	128	17	22	32168	1424	1424	1320
2048	8192	128	16	25	32168	1429	1429	1324
2048	8192	128	16	28	32168	1593	1593	1475
2176	8704	128	17	25	32168	1609	1609	1490
2304	9216	128	18	24	32168	1731	1730	1603
2176	8704	128	17	28	32168	1794	1794	1661
2304	9216	128	18	26	32168	2007	1868	1730
2304	9216	128	18	32	32168	2283	2282	2113
2560	10240	128	20	26	32168	2298	2297	2127
2560	10240	128	20	30	32168	2639	2638	2442
2560	10240	128	20	34	32168	2980	2979	2756
2688	10752	128	22	36	32168	3530	3530	3257
2816	11264	128	22	36	32168	3802	3802	3516
2944	11776	128	22	36	32168	4084	4083	3785
3072	12288	128	24	36	32168	4516	4515	4176
3584	14336	128	28	40	32168	6796	6795	6281
4096	16384	128	32	42	32168	9293	9292	8587
4352	17408	128	32	47	32168	11452	11450	10613
4608	18432	128	36	44	32168	12295	12294	11360
4608	18432	128	32	47	32168	12569	12568	11680
4864	19456	128	36	47	32168	13775	14319	13266
4992	19968	128	32	49	32168	14940	14939	13937

1620	5120	20480	128	40	47	32168		16183		16182		14950
------	------	-------	-----	----	----	-------	--	-------	--	-------	--	-------

---

1621

1622

1623

1624

1625

1626

1627

1628

1629

1630

1631

1632

1633

1634

1635

1636

1637

1638

1639

1640

1641

1642

1643

1644

1645

1646

1647

1648

1649

1650

1651

1652

1653

1654

1655

1656

1657

1658

1659

1660

1661

1662

1663

1664

1665

1666

1667

1668

1669

1670

1671

1672

1673

Table 2: **Chinchilla Language Models.** We copy Chinchilla’s Table A9 listing the model parameters and model architectural hyperparameters of all models used in the Chinchilla fitting processes. Parameters specified in millions (1e6).

## 1674 C THEORETICAL ANALYSIS

1675  
1676 Here, we provide a detailed analysis of the empirical results obtained in the main text from a  
1677 theoretical perspective. We begin by repeating the derivation of the baseline compute-optimal scaling  
1678 for the number of tokens as a function of model parameters, and continue to systematically work  
1679 through the perturbations discussed in section 3.

### 1681 C.1 BASELINE COMPUTE-OPTIMAL SCALING DERIVATION

1682  
1683 The Chinchilla scaling law for pretraining loss  $L$  as a function of non-embedding model parameters  
1684  $N$  and number of training tokens  $D$  is given by eq. (4), which we repeat here

$$1685 L(N, D) = E + \frac{A}{N^\alpha} + \frac{B}{D^\beta}, \quad (10)$$

1686 where  $E$  is the irreducible loss, and  $(A, \alpha)$  and  $(B, \beta)$  are parameters for the model size and data  
1687 scaling terms, respectively.

1688  
1689 The training compute budget  $C$  is approximately proportional to the product of model size and  
1690 training data,  $C \approx cND$ , where  $c > 0$  is some constant factor. This allows us to express the number  
1691 of training tokens as a function of compute and model size:  $D = C/(cN)$ . Substituting this into the  
1692 loss function yields the loss for a fixed compute budget

$$1693 L(N, C) = E + AN^{-\alpha} + B \left( \frac{C}{cN} \right)^{-\beta} = E + AN^{-\alpha} + B(c^\beta C^{-\beta})N^\beta. \quad (11)$$

1694  
1695 The optimal model size  $N_{\text{opt}}$  that minimizes the loss for a fixed compute budget  $C$  is simply found by  
1696 differentiating eq. (11) with respect to  $N$  and setting the derivative to zero

$$1697 \frac{\partial L}{\partial N} = -\alpha AN^{-(\alpha+1)} + \beta B(c^\beta C^{-\beta})N^{\beta-1} = 0. \quad (12)$$

1698 Rearranging this equation reveals the optimal trade-off:

$$1699 \alpha AN_{\text{opt}}^{-(\alpha+1)} = \beta B(c^\beta C^{-\beta})N_{\text{opt}}^{\beta-1}. \quad (13)$$

1700 Solving for  $N_{\text{opt}}$ :

$$1701 N_{\text{opt}}^{\alpha+\beta} = \frac{\alpha A}{\beta B c^\beta} C^\beta \implies N_{\text{opt}} = \left( \frac{\alpha A}{\beta B c^\beta} \right)^{\frac{1}{\alpha+\beta}} C^{\frac{\beta}{\alpha+\beta}} \quad (14)$$

1702 The compute-optimal tokens-per-parameter ratio is  $R_{\text{opt}} = D_{\text{opt}}/N_{\text{opt}}$ . Using  $D_{\text{opt}} = C/(cN_{\text{opt}})$ , we  
1703 find  $R_{\text{opt}} = C/(cN_{\text{opt}}^2)$ . Substituting our expression for  $N_{\text{opt}}$  results in

$$1704 R_{\text{opt}} = \frac{C}{c} \left[ \left( \frac{\alpha A}{\beta B c^\beta} \right)^{\frac{1}{\alpha+\beta}} C^{\frac{\beta}{\alpha+\beta}} \right]^{-2} = \frac{C}{c} \left( \frac{\beta B c^\beta}{\alpha A} \right)^{\frac{2}{\alpha+\beta}} C^{\frac{-2\beta}{\alpha+\beta}}. \quad (15)$$

1705 This simplifies to the final form, which shows the ratio's dependence on the compute budget  $C$  as

$$1706 \frac{D_{\text{opt}}}{N_{\text{opt}}} = K \cdot C^{\frac{\alpha-\beta}{\alpha+\beta}}, \quad (16)$$

1707 where  $K = \frac{1}{c} \left( \frac{\beta B c^\beta}{\alpha A} \right)^{\frac{2}{\alpha+\beta}}$  is a constant. The key insight from the Chinchilla works is that empirically,  
1708 one finds that  $\alpha \approx \beta$ , making the exponent on  $C$  approximately zero and the optimal ratio nearly  
1709 constant. In mathematical terms, it leads to the relation

$$1710 \frac{D_{\text{opt}}}{N_{\text{opt}}} \approx K = \left( \frac{B}{A} \right)^{\frac{1}{\alpha}}. \quad (17)$$

1711 If we want to recover the 20 : 1 ratio of training tokens to number of parameters we expect to find  
1712 that  $B \approx 2.85A$  for  $\alpha \approx 0.35$ .

## 1728 C.2 ANALYSIS OF PERTURBATIONS

1729  
1730 We now analyze how systematic errors in model parameter counts affect the fitted scaling param-  
1731 eters  $(\hat{A}, \hat{\alpha})$  and the resulting optimal ratio. Let  $N$  be the true parameter count and  $\tilde{N}$  be the  
1732 perturbed (incorrect) count used for fitting. The fitting process minimizes the error between the  
1733 model  $L(\tilde{N}, D) = \hat{E} + \hat{A}\tilde{N}^{-\hat{\alpha}} + \hat{B}D^{-\hat{\beta}}$  and the observed losses. For most types of perturbations  
1734 we consider, since the data  $D$  is unaffected, we assume  $\hat{B} \approx B$  and  $\hat{\beta} \approx \beta$ . We will break this  
1735 assumption when necessary.

### 1737 C.2.1 MULTIPLICATIVE CONSTANT PERTURBATION

1738  
1739 Assume the reported parameters are a constant multiple of the true parameters:  $\tilde{N} = c_m \cdot N$ . The  
1740 model size term in the loss is modified to

$$1741 \hat{A}\tilde{N}^{-\hat{\alpha}} = \hat{A}(c_m N)^{-\hat{\alpha}} = (\hat{A}c_m^{-\hat{\alpha}})N^{-\hat{\alpha}}. \quad (18)$$

1742  
1743 For eq. (18) to match the true term  $AN^{-\alpha}$ , the fitting procedure will ideally find parameters such  
1744 that:

- 1745 •  $\hat{\alpha} \approx \alpha$
- 1746 •  $\hat{A}c_m^{-\hat{\alpha}} \approx A \implies \hat{A} \approx Ac_m^\alpha$

1747  
1748 The exponent in the optimal ratio (16) becomes  $(\hat{\alpha} - \beta)/(\hat{\alpha} + \beta) \approx (\alpha - \beta)/(\alpha + \beta)$ .

1750 **Conclusion:** A multiplicative error does not change the exponent governing the trend of the optimal  
1751 ratio. The flat relationship with compute budget is preserved. However, the constant prefactor  $K$  is  
1752 shifted by a factor of  $(c_m^\alpha)^{-2/(\alpha+\beta)} = c_m^{-2\alpha/(\alpha+\beta)} \approx c_m^{-1}$ , which shifts the entire line up or down on  
1753 a log-log plot. This is shown in Fig. 4 (top row) and Fig. 5 (top left).

### 1755 C.2.2 ADDITIVE CONSTANT PERTURBATION

1756  
1757 Assume an additive error, e.g., from including/excluding embeddings:  $\tilde{N} = N + c_a$ . The loss term  
1758  $\hat{A}(N + c_a)^{-\hat{\alpha}}$  is no longer a pure power law in  $N$ .

1759 We examine the process of fitting the perturbed model  $g(N; \hat{A}, \hat{\alpha}) = \hat{A}(N + c_a)^{-\hat{\alpha}}$  to the true  
1760 function  $f(N) = AN^{-\alpha}$  by minimizing the Mean Squared Error (MSE) in log-space.

1761  
1762 The objective is to find  $(\hat{A}, \hat{\alpha})$  that minimize  $\sum_i [\log f(N_i) - \log g(N_i)]^2$ . The core of the problem  
1763 lies in approximating the term  $\log(N + c_a)$ . For the regime where  $N \gg |c_a|$ , which applies to the  
1764 larger models in the study, we can analyze the local behavior of the function.

1765  
1766 **Effect on the Scaling Exponent  $\hat{\alpha}$ :** The most critical parameter in a power law is its exponent,  
1767 which corresponds to the slope in a log-log plot. The true slope is constant

$$1768 \frac{d(\log f(N))}{d(\log N)} = -\alpha \quad (19)$$

1770 For the perturbed function, the effective slope is not constant and depends on  $N$  as

$$1771 \text{Effective Slope}(N) = \frac{d(\log g(N))}{d(\log N)} = -\hat{\alpha} \frac{d(\log(N + c_a))}{d(\log N)} = -\hat{\alpha} \left( \frac{N}{N + c_a} \right) \quad (20)$$

$$1772 \implies \hat{\alpha} = \alpha / \left( \frac{N}{N + c_a} \right).$$

1773  
1774  
1775  
1776  
1777 The fitting procedure must select a single exponent  $\hat{\alpha}$  that best represents this varying slope over the  
1778 range of data. To match the true average slope of  $-\alpha$ , the fitted  $\hat{\alpha}$  must compensate for the factor  
1779  $N/(N + c_a)$ :

- 1780 • When  $c_a > 0$ , the factor  $N/(N + c_a) < 1$ . To achieve the target slope, the fitting process  
1781 must select an exponent  $\hat{\alpha} > \alpha$ .

- When  $c_a < 0$ , the factor  $N/(N + c_a) > 1$  (for  $N > |c_a|$ ). To compensate, the fitting process must select an exponent  $\hat{\alpha} < \alpha$ .

This provides a direct analytical explanation for the observed behavior of  $\hat{\alpha}$  in Figure 4, which is smaller than the true  $\alpha$  for  $c_a < 0$  and increases approximately linearly for  $c_a > 0$ .

**Effect on the Prefactor  $\hat{A}$ :** Once the optimal  $\hat{\alpha}$  is determined, the prefactor  $\hat{A}$  is chosen to minimize the remaining offset. We can approximate this by enforcing that the functions match at some effective "pivot" point  $N_0$  that is characteristic of the dataset.

$$f(N_0) \approx g(N_0) \implies AN_0^{-\alpha} \approx \hat{A}(N_0 + c_a)^{-\hat{\alpha}}. \quad (21)$$

Solving for  $\hat{A}$  gives

$$\hat{A} \approx A \cdot N_0^{-\alpha} \cdot (N_0 + c_a)^{\hat{\alpha}} = A \left( \frac{N_0 + c_a}{N_0} \right)^{\hat{\alpha}} (N_0)^{\hat{\alpha} - \alpha}. \quad (22)$$

Assuming for simplicity that the pivot point is chosen such that the  $N_0^{\hat{\alpha} - \alpha}$  term is of order one, we focus on the dominant term

$$\hat{A} \approx A \left( 1 + \frac{c_a}{N_0} \right)^{\hat{\alpha}}. \quad (23)$$

This relationship explains the rapid growth of  $\hat{A}$ . Since we have already established that  $\hat{\alpha}$  itself increases with  $c$ , the prefactor  $\hat{A}$  grows due to two compounding effects: an increase in the base  $(1 + c_a/N_0)$  and an increase in the exponent  $\hat{\alpha}$ . This leads to the exponential-like growth observed empirically in fig. 4.

### C.2.3 SYSTEMATIC BIAS PERTURBATION

Assume a bias where the error itself scales with model size, modeled as  $\tilde{N} = \mu_{\text{geo}}(N/\mu_{\text{geo}})^s$ , where  $\mu_{\text{geo}}$  is the geometric mean of the true parameter counts and  $s$  is the bias factor. The model size term becomes

$$\hat{A}\tilde{N}^{-\hat{\alpha}} = \hat{A}(\mu_{\text{geo}}^{1-s}N^s)^{-\hat{\alpha}} = \left(\hat{A}\mu_{\text{geo}}^{-(1-s)\hat{\alpha}}\right)N^{-s\hat{\alpha}} \quad (24)$$

To match the true term  $AN^{-\alpha}$ , the exponent and the constant term must satisfy the relations

$$\hat{\alpha} = \frac{\alpha}{s}, \quad \hat{A} = \mu_{\text{geo}}^{\frac{\alpha(1-s)}{s}} A. \quad (25)$$

The fitted exponent is inversely proportional to the bias factor  $s$ , which is verified empirically in section 3.3. The exponent in the optimal ratio is now  $(\alpha/s - \beta)/(\alpha/s + \beta)$ .

**Conclusion:** A systematic bias also breaks the  $\hat{\alpha} \approx \beta$  condition, unless  $s = 1$ .

- If  $s < 1$  (inflating larger models relative to smaller ones), then  $\hat{\alpha} > \alpha \approx \beta$ . The exponent on  $C$  becomes positive, and the optimal ratio *increases* with compute.
- If  $s > 1$  (shrinking larger models relative to smaller ones), then  $\hat{\alpha} < \alpha \approx \beta$ . The exponent on  $C$  becomes negative, and the optimal ratio *decreases* with compute.

This perturbation also qualitatively changes the optimal scaling strategy, with the direction of the change depending on the nature of the bias, as seen in fig. 5 (bottom left).

## D RELATED WORK

**Scaling Laws in Neural (Language) Models** While initial research on scaling laws in neural models began decades ago (Barkai et al., 1993; Mhaskar, 1996; Pinkus, 1999), advances in scaling large language models brought such interest into renewed focus (Hestness et al., 2017; Kaplan et al., 2020; Brown et al., 2020), causing an explosion of research. For a non-exhaustive list, theoretical understanding of scaling laws has advanced substantially (Spigler et al., 2020; Bousquet et al., 2020; Hutter, 2021; Sharma & Kaplan, 2022; Maloney et al., 2022; Roberts et al., 2022; Bahri et al., 2024; Paquette et al., 2024; Atanasov et al., 2024; Bordelon et al., 2024a;b; Lin et al., 2024; Brill, 2024), complemented by empirical studies (Rosenfeld et al., 2020; Henighan et al., 2020; Gordon et al., 2021; Tay et al., 2021; Ghorbani et al., 2021; Zhai et al., 2022; Alabdulmohsin et al., 2022; Dehghani et al., 2023; Bachmann et al., 2023; Everett et al., 2024; Qiu et al., 2025).

Additional research has also studied how scaling interacts with specific considerations such as efficient inference (Sardana et al., 2024; Bian et al., 2025), transfer (Hernandez et al., 2021; Barnett, 2024), data quality and diversity (Chen et al., 2025; Hernandez et al., 2022; Muennighoff et al., 2023; Qin et al., 2025; Shukor et al., 2025), overtraining (Gadre et al., 2024), quantization and precision (Dettmers & Zettlemoyer, 2023; Sun et al., 2025; Kumar et al., 2024), differential privacy (McKenna et al., 2025), distillation (Busbridge et al., 2025), model architecture (Clark et al., 2022; Kudugunta et al., 2023; Abnar et al., 2025; Ludziejewski et al., 2025; Liew et al., 2025), context length (Xiong et al., 2023; Agarwal et al., 2024; Arora et al., 2024), vocabulary size (Tao et al., 2024), robustness to jailbreaking (Howe et al., 2025; Anil et al., 2024; Hughes et al., 2024), pruning (Rosenfeld et al., 2021), multimodality (Aghajanyan et al., 2023; Cherti et al., 2023), fine-tuning (Kalajdziewski, 2024; Zhang et al., 2024) and agents and world models (Pearce et al., 2025).

Recent work has also highlighted novel scaling phenomena such as inverse scaling (McKenzie et al., 2024; Gema et al., 2025), emergent capabilities (Srivastava et al., 2023; Wei et al., 2022; Schaeffer et al., 2023; Hu et al., 2024; Schaeffer et al., 2024; Snell et al., 2024b; Wu & Lo, 2024), and critical issues like data contamination (Schaeffer, 2023; Jiang et al., 2024; Dominguez-Olmedo et al., 2024) and model-data feedback loops (Dohmatob et al., 2024; Gerstgrasser et al., 2024; Kazdan et al., 2024). The advent of so-called “thinking” or “reasoning” models (Jaech et al., 2024) has sparked a new wave of interest in scaling inference compute (Brown et al., 2024; Snell et al., 2024a; Wu et al., 2024; Chen et al., 2024; Sadhukhan et al., 2025; Levi, 2025; Schaeffer et al., 2025; Kwok et al., 2025).