Scaling Laws of Synthetic Images for Model Training … for Now

Lijie Fan, Kaifeng Chen, Dilip Krishnan, Dina Katabi, Phillip Isola, Yonglong Tian

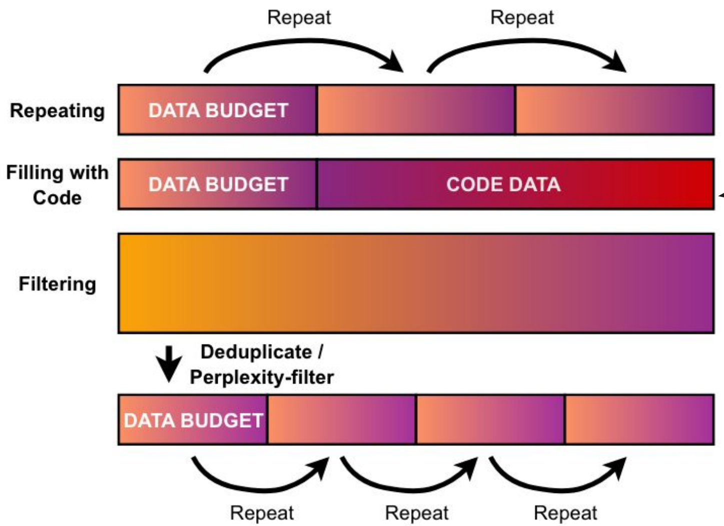
Presented By: Ben Newman

Why use *Synthetic Data*?

- Real data is **expensive**
- Running out of web data (Muennighoff et al., 2023)

Why use *Synthetic Data*?

- Real data is **expensive**
- Running out of web data (Muennighoff et al., 2023)



We can also use ***synthetic data***!

What scaling laws do we see when using *synthetic data*?

- Reminder: Scaling laws answer the question:
  - **How do we partition a fixed amount of compute between model parameters and training data to achieve optimal performance (min loss, max accuracy, etc.)?**

$$\operatorname*{argmin}_{N,D} L(N, D) \text{ s.t. } \mathrm{FLOPs}(N, D) = C$$

  - Then we fit some curves:

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

(Muennighoff et al., 2023)

What scaling laws do we see when using *synthetic data*?

- In this paper, it's different:
  - **Model parameters** are basically **constant** (they use ViT-B)
  - **Data is varied; Compute varies with data**
- The question they ask is:

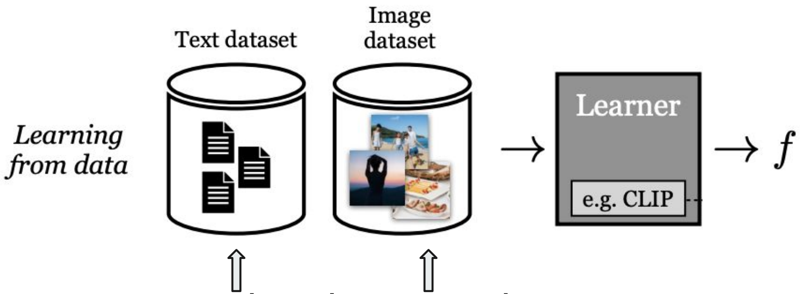
How much synthetic data is needed for desired performance?

How much synthetic data is needed for desired performance?

1. For **image representation learning** (text-image contrastive learning): **synthetic data** can perform **equivalently** to **real data**

How much synthetic data is needed for desired performance?

1. For **image representation learning** (text-image contrastive learning): **synthetic data** can perform **equivalently** to **real data**



Nguyen et al., 2023: synthetic real

Tian et al., Oct. 2023: real synthetic

Tian et al.. Dec. 2023 synthetic synthetic

(Tian et al.. Dec. 2023)

How much synthetic data is needed for desired performance?

1. For **image representation learning** (text-image contrastive learning): **synthetic data** can perform **equivalently** to **real data**

2. For **image classification**, **synthetic data** ***underperforms*** **real data** (Sariyildiz et al., 2023)

How much synthetic data is needed for desired performance?

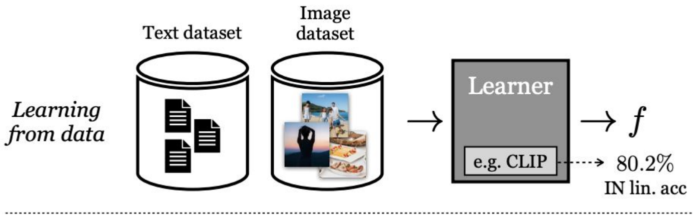
1. For **image representation learning** (text-image contrastive learning): **synthetic data** can perform **equivalently** to **real data**

2. For **image classification**, **synthetic data** ***underperforms*** **real data** (Sariyildiz et al., 2023)

3. For **image classification**, **synthetic data + real data can *outperform* real data** (Aziz et al., 2023; Yu et al., 2023)

What's the background of this paper?

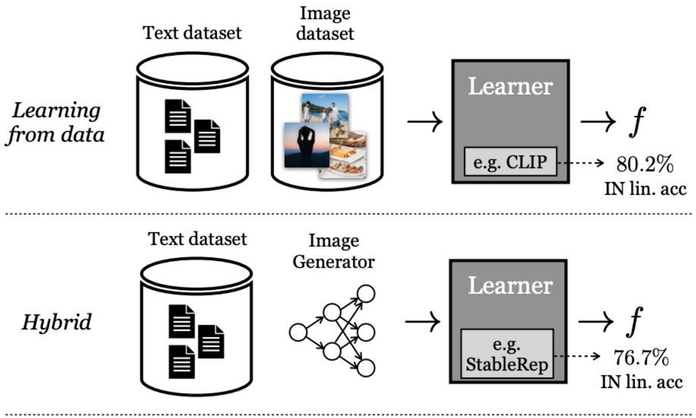
One of three from a collab btw Lijie Fan at MIT and Yonglong Tian at Google Research around synthetic data for representation learning



What's the background of this paper?

One of three from a collab btw Lijie Fan at MIT and Yonglong Tian at Google Research around synthetic data for representation learning

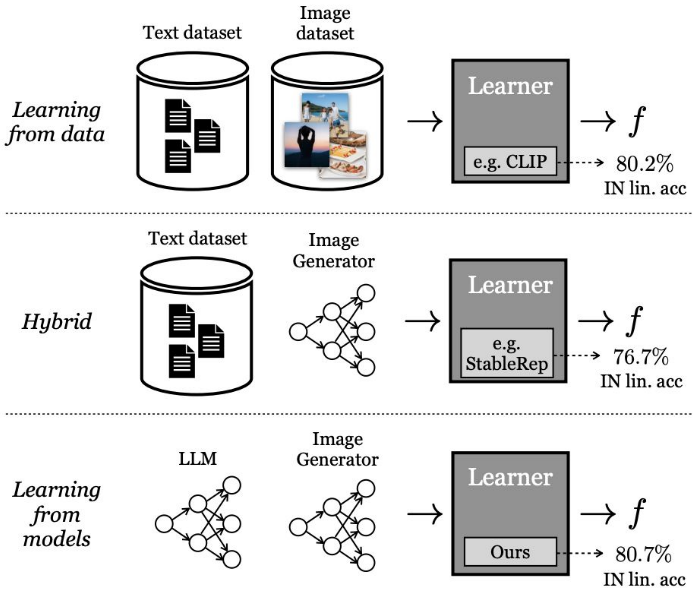
1. Real captions & synthetic images representation learning (Tian et al., Oct, 2023)



What's the background of this paper?

One of three from a collab btw Lijie Fan at MIT and Yonglong Tian at Google Research around synthetic data for representation learning

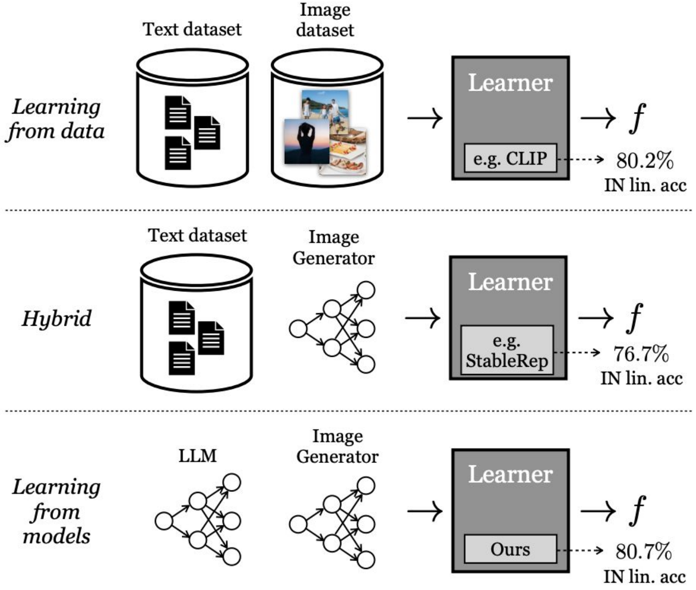
1. Real captions & synthetic images for representation learning (Tian et al., Oct, 2023)
2. Synthetic captions & synthetic images for representation learning (Tian et al., Dec 28, 2023)



What's the background of this paper?

One of three from a collab btw Lijie Fan at MIT and Yonglong Tian at Google Research around synthetic data for representation learning

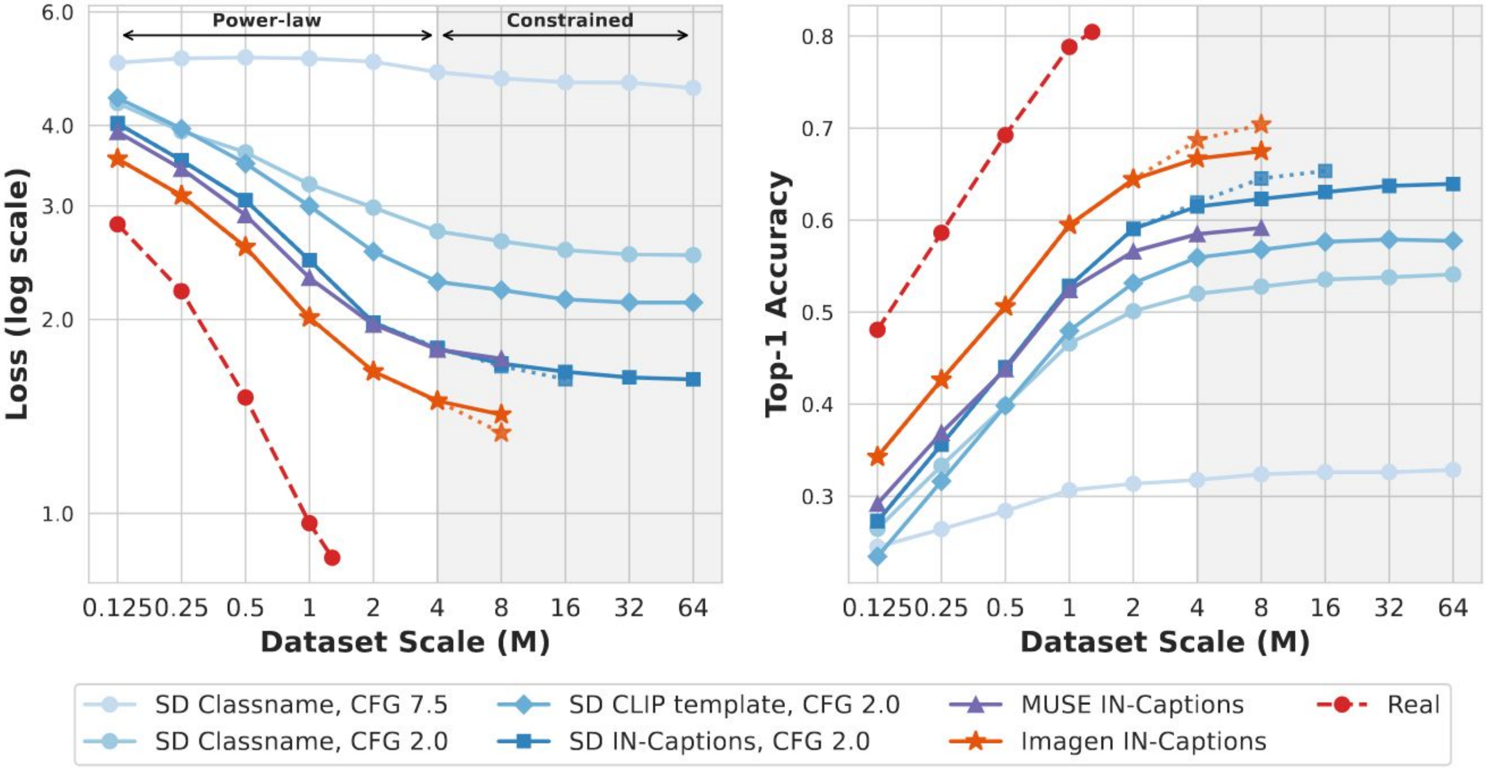
1. Real captions & synthetic images for representation learning (Tian et al., Oct, 2023)
2. Synthetic captions & synthetic images for representation learning (Tian et al., Dec 28, 2023)
3. This paper, scaling laws for synthetic data generation



Three main claims in this paper

1. **Synthetic data scaling for image classification:**
    a. **worse than real data (in-domain),**
    b. better than real data (out-of-domain).
2. Class-based scaling
3. Synthetic data does scale well in CLIP model training

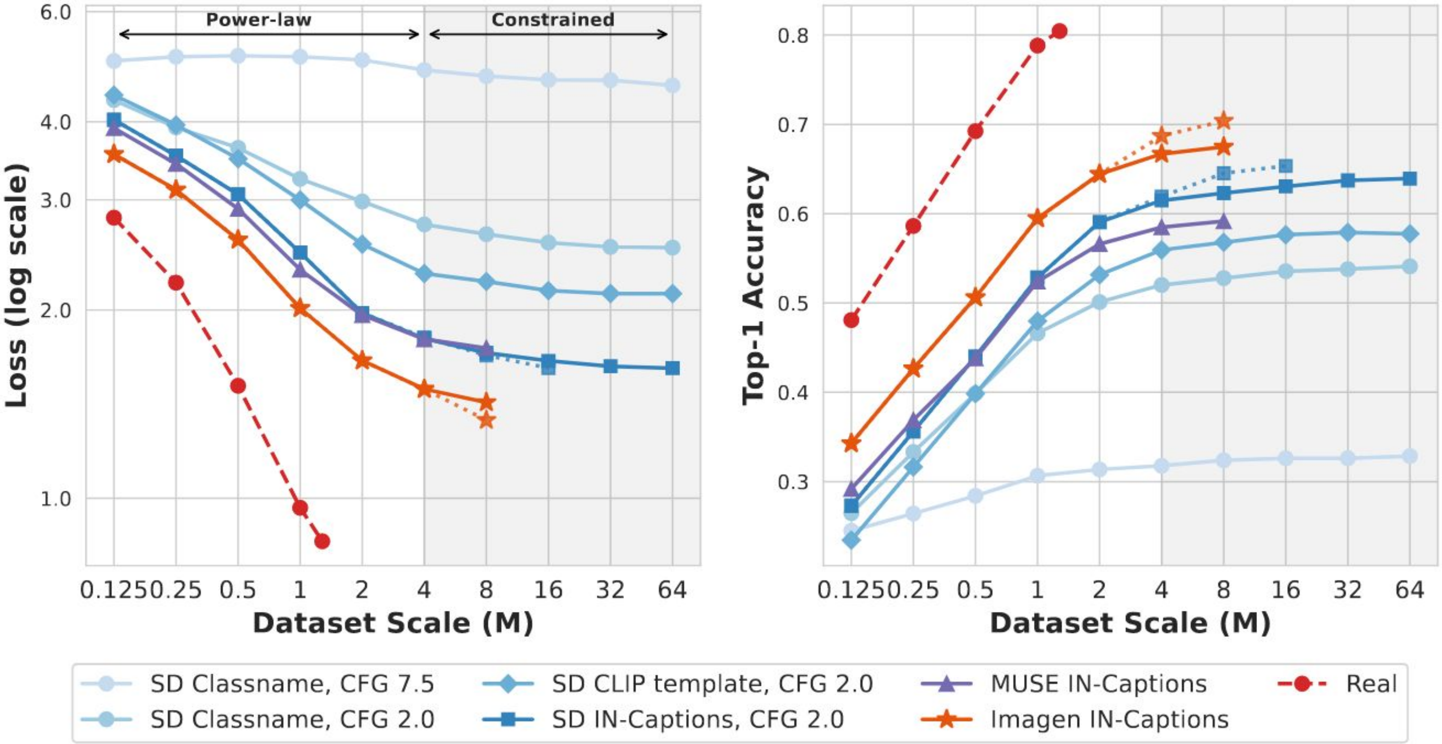
How does synthetic data scale?



How does synthetic data scale?

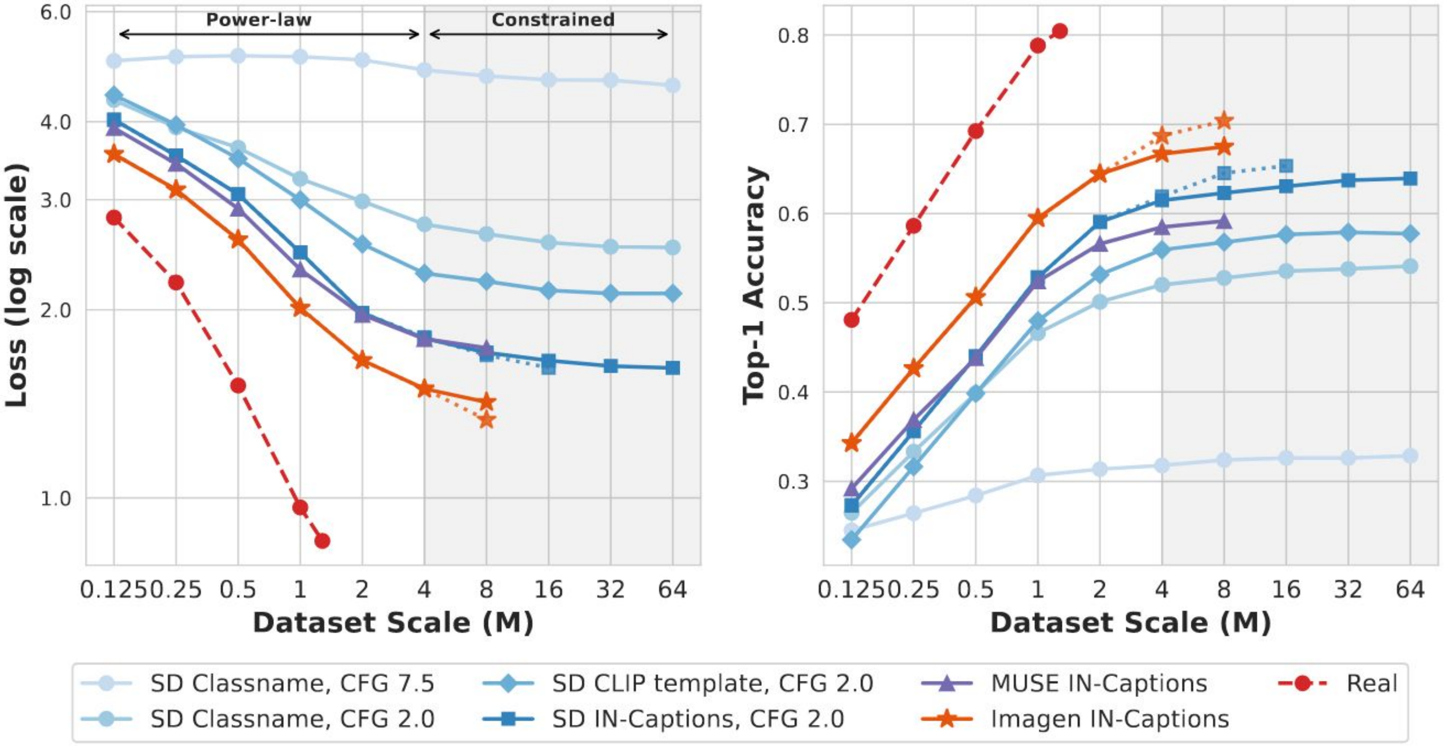
Models (Dosovitskiy et al., 2023):

1. ViT-Base (86M params)
2. ViT-Large (307M params)



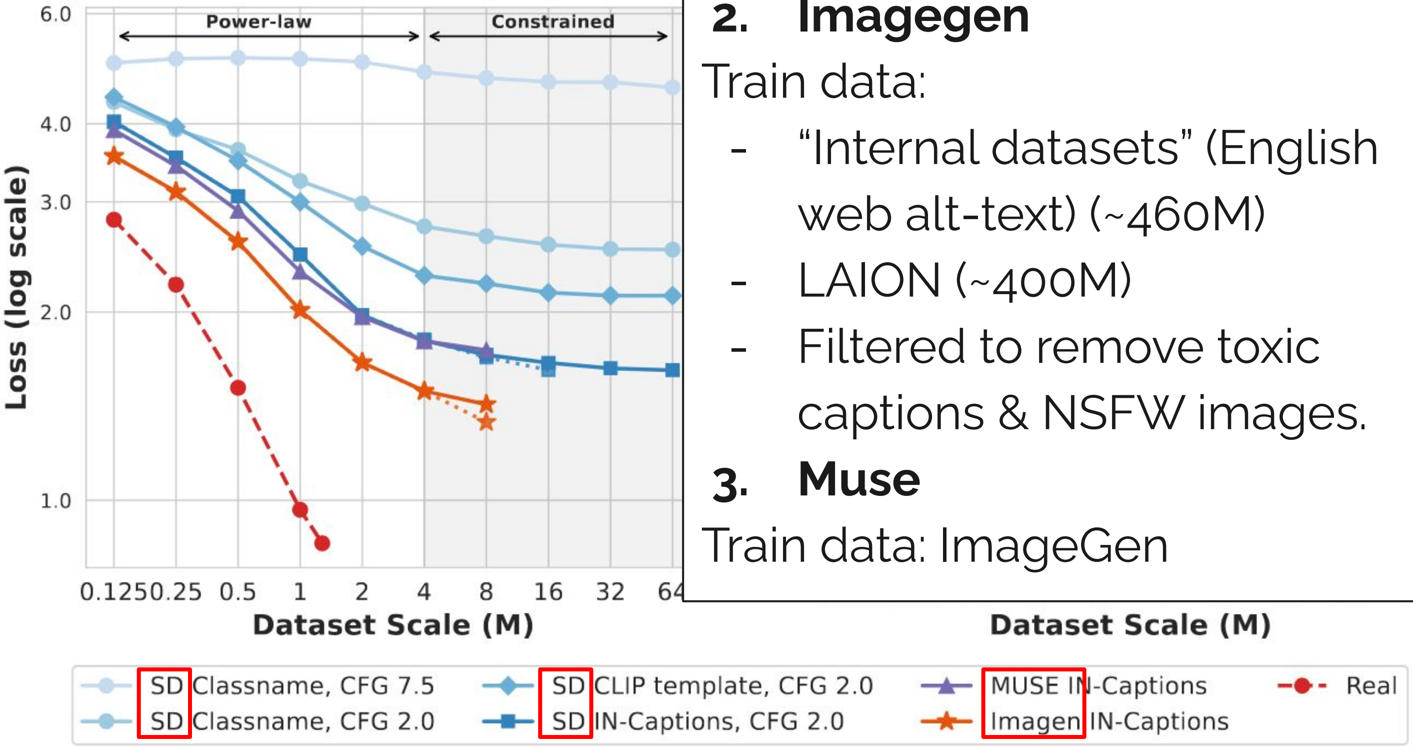
How does synthetic data scale?

1. Image Generation Models
2. Prompts to Generation Models
3. Classifier-Free Guidance



How does synthetic data scale?

1. **Image Generation Models**
2. Prompts to Generation Models
3. Classifier-Free Guidance



1. **Stable Diffusion (SD)**

Train data: LAION-5B (subsets)

2. **Imagegen**

Train data:

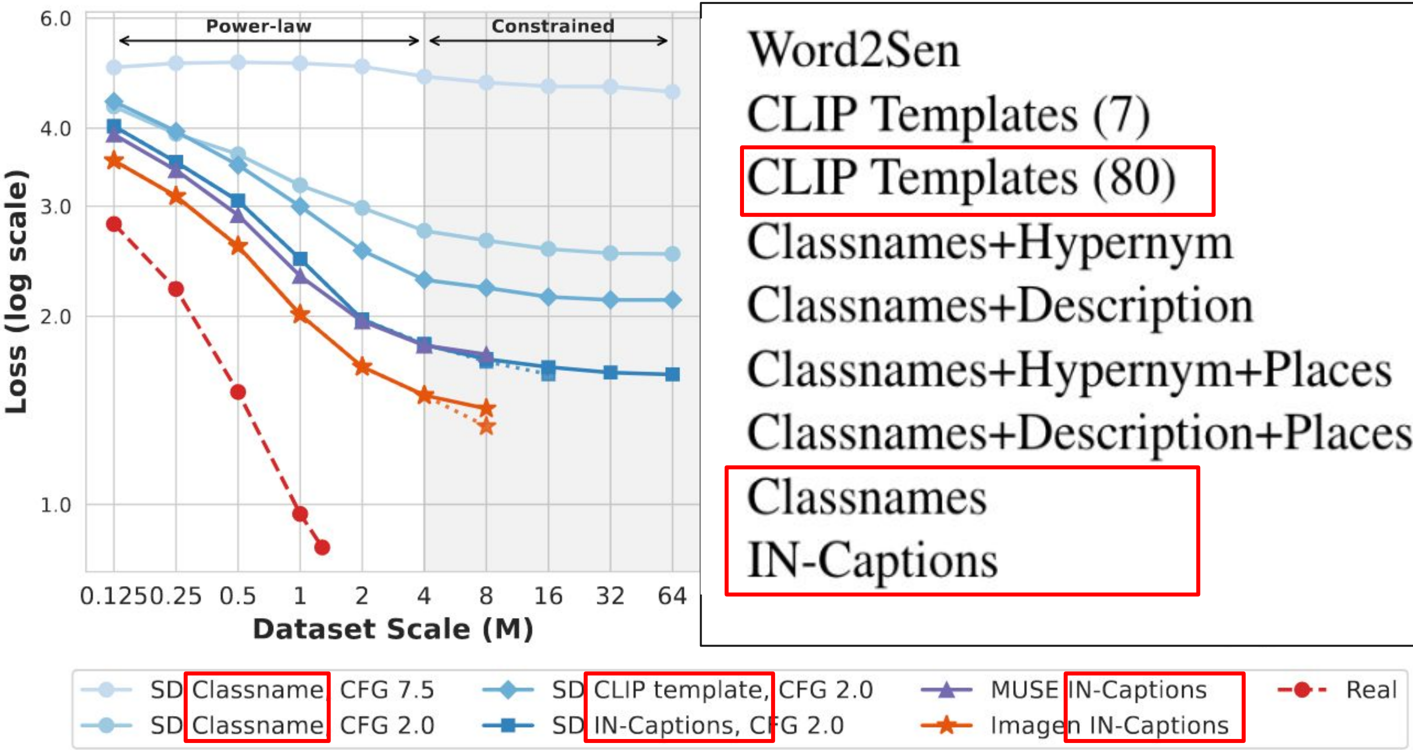
- "Internal datasets" (English web alt-text) (~460M)
- LAION (~400M)
- Filtered to remove toxic captions & NSFW images.

3. **Muse**

Train data: ImageGen

How does synthetic data scale?

1. Image Generation Models
2. **Prompts to Generation Models**
3. Classifier-Free Guidance

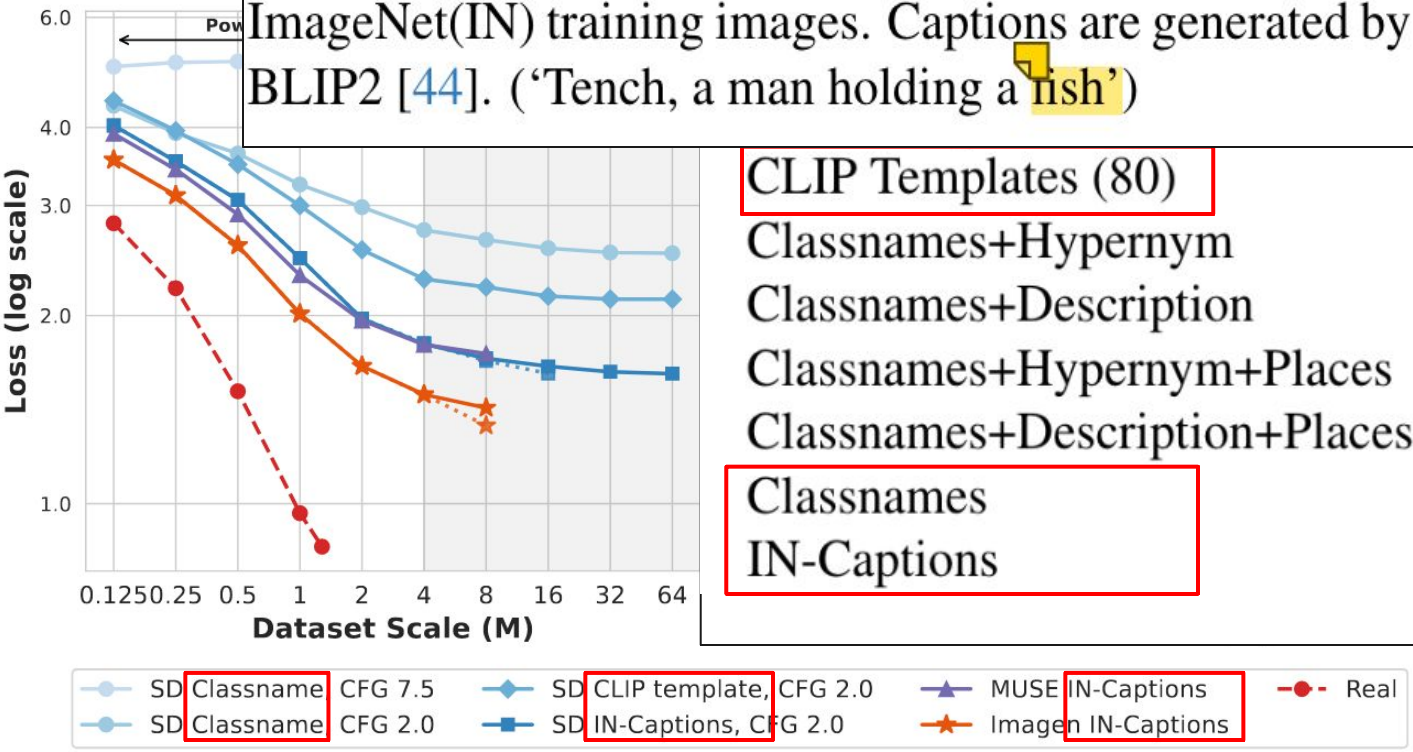


Word2Sen
CLIP Templates (7)
CLIP Templates (80)
Classnames+Hypernym
Classnames+Description
Classnames+Hypernym+Places
Classnames+Description+Places
Classnames
IN-Captions

How does synthetic data scale?

1. Image Generation Models
2. **Prompts to Generation Models**
3. Classifier-Free Guidance

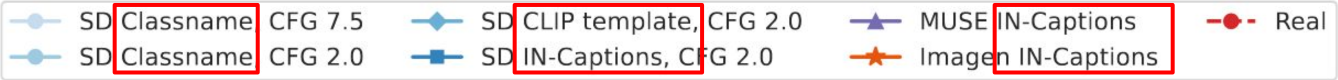
IN-Captions: Combine the class name with captions from ImageNet(IN) training images. Captions are generated by BLIP2 [44]. ('Tench, a man holding a fish')



How does synthetic data scale?

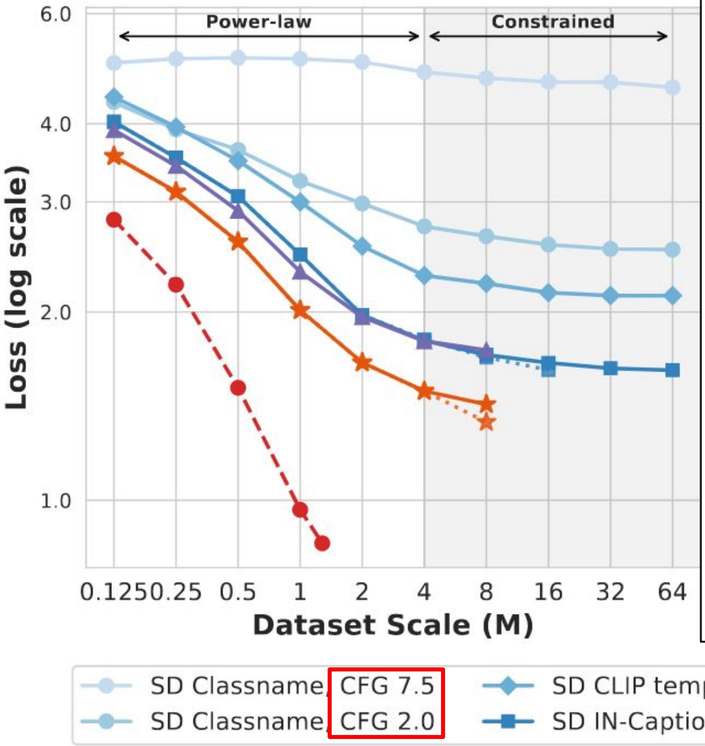
1. Image Generation Models
2. **Prompts to Generation Models**
3. Classifier-Free Guidance

- *Classnames:* Directly use the ImageNet class name. ('Tench')
- *CLIP templates:* Generate either 7 or 80 sentences with the text templates CLIP used for zero-shot classification task. ('a photo of the large tench')
- *IN-Captions:* Combine the class name with captions from ImageNet(IN) training images. Captions are generated by BLIP2 [44]. ('Tench, a man holding a fish')



How does synthetic data scale?

1. Image Generation Models
2. Prompts to Generation Models
3. **Classifier-Free Guidance**



Idea: move logits from the unconditional (ℓ_u) toward conditional (ℓ_c) for high quality

Training:

90% : "Dog" → Gen. → ℓ_c

10% : Ø → Gen. → ℓ_u

Inference:

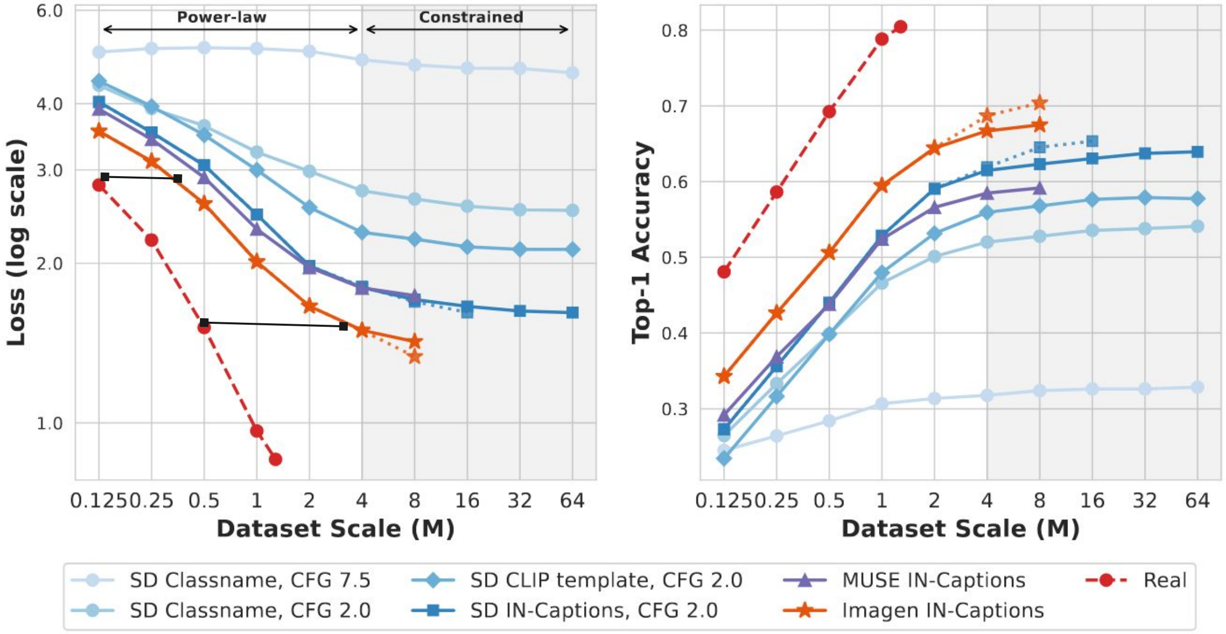
High ℓ = higher quality

Low ℓ = higher diversity

$$\ell_g = (1+t)\ell_c - t\ell_u$$

How does synthetic data scale?

1. You need **~3-8x** synthetic data than real data to achieve the same loss in **best case**

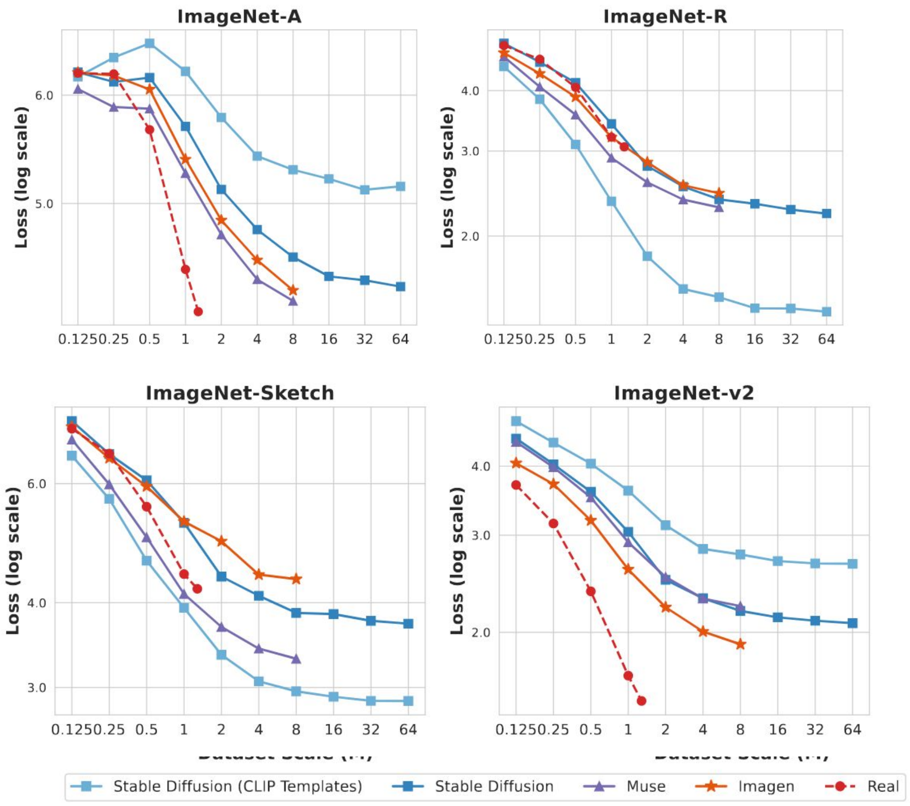


How does synthetic data scale OOD?

Inconsistently!

Real data is better on ImageNet-A and ImageNet-v2

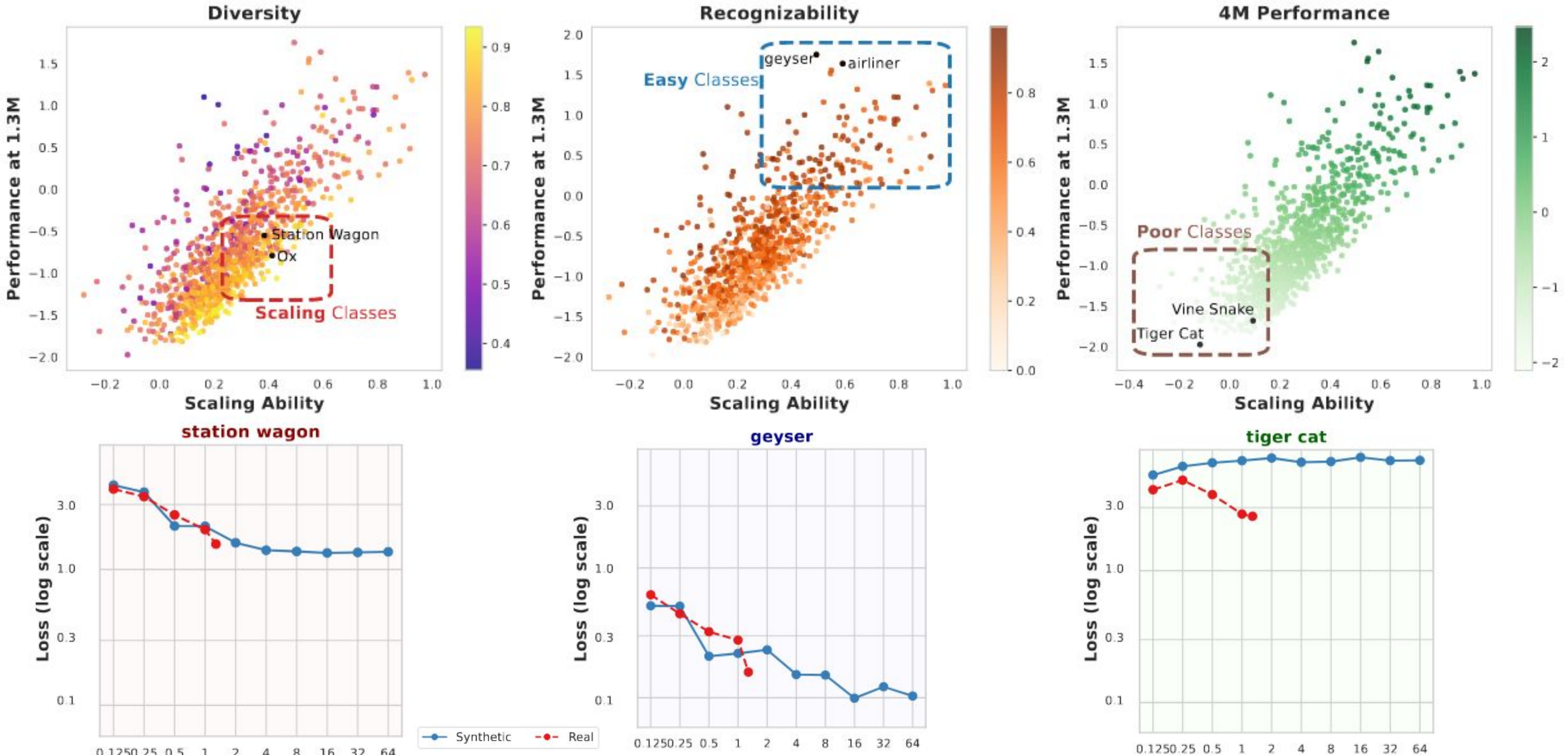
Stable Diffusion (CLIP Templates) is best on Sketch and ImageNet.
Why?



How does synthetic data scale per class?

Deep Dive

Inconsistently! - They point to three classes: “scaling”, “easy”, “poor”



References

Azizi, Shekoofeh, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J. Fleet. 2023. "Synthetic Data from Diffusion Models Improves ImageNet Classification." arXiv. http://arxiv.org/abs/2304.08466.

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, et al. 2021. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv. http://arxiv.org/abs/2010.11929.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. "Scaling Laws for Neural Language Models." arXiv. http://arxiv.org/abs/2001.08361.

Sariyildiz, Mert Bulent, Karteek Alahari, Diane Larlus, and Yannis Kalantidis. 2023. "Fake It till You Make It: Learning Transferable Representations from Synthetic ImageNet Clones." arXiv. http://arxiv.org/abs/2212.08420.

Muennighoff, Niklas, Alexander M. Rush, Boaz Barak, Teven Le Scao, Aleksandra Piktus, Nouamane Tazi, Sampo Pyysalo, Thomas Wolf, and Colin Raffel. 2023. "Scaling Data-Constrained Language Models." arXiv. http://arxiv.org/abs/2305.16264.

Nguyen, Thao, Samir Yitzhak Gadre, Gabriel Ilharco, Sewoong Oh, and Ludwig Schmidt. 2023. "Improving Multimodal Datasets with Image Captioning." arXiv. http://arxiv.org/abs/2307.10350.

Tian, Yonglong, Lijie Fan, Phillip Isola, Huiwen Chang, and Dilip Krishnan. 2023. "StableRep: Synthetic Images from Text-to-Image Models Make Strong Visual Representation Learners." arXiv. http://arxiv.org/abs/2306.00984.

Tian, Yonglong, Lijie Fan, Kaifeng Chen, Dina Katabi, Dilip Krishnan, and Phillip Isola. 2023. "Learning Vision from Models Rivals Learning Vision from Data." arXiv. http://arxiv.org/abs/2312.17742.

- *Classnames:* Directly use the ImageNet class name. ('Tench')
- *Classnames + Description:* Combine class name with its WordNet [50] description. ('tench, freshwater dace-like game fish of Europe and western Asia ...')
- *Classnames + Hypernyms:* Combine ImageNet class name with its Wordnet hypernyms. ('Tench, Tinca tinca, cyprinid, cyprinid fish')
- *Word2Sen:* Use a pre-trained T5 model [60] as used in [24] to convert the ImageNet class name into a sentence. We generate 100 sentences for each class. ('a tench with fish in the distance.')
- *CLIP templates:* Generate either 7 or 80 sentences with the text templates CLIP used for zero-shot classification task. ('a photo of the large tench')
- *IN-Captions:* Combine the class name with captions from ImageNet(IN) training images. Captions are generated by BLIP2 [44]. ('Tench, a man holding a fish')

A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity

Shayne Longpre, Gregory Yauney, Emily Reif, Katherine Lee, Adam Roberts, Barret Zoph, Denny Zhou, Jason Wei, Kevin Robinson, David Mimno, Daphne Ippolito (2023)

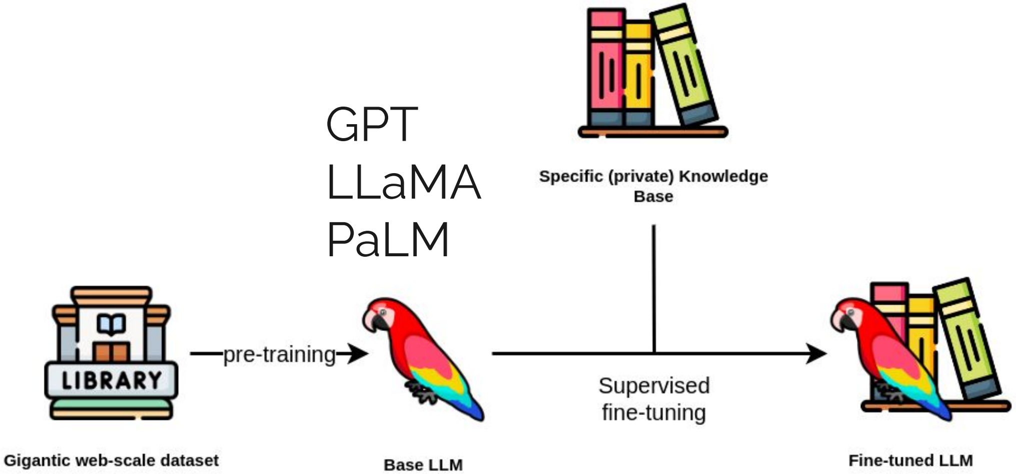
Presented By: Cailin Winston

Pretraining LLMs

- Fine-tuning LLMs **pretrained from large datasets** is the norm.
- **Why** pretraining datasets are curated in a certain way is unclear.

GPT
LLaMA
PaLM

C4
WebText



https://neo4j.com/developer-blog/fine-tuning-retrieval-augmented-generation/

Pretraining Dataset Curation

Semi-curated, task-specific datasets

Wikipedia, BookCorpus, One Billion Word Benchmark.

(Zhu et al., 2015), (Chelba et al., 2013)

(Yang et al., 2019)

Use of web-scraped data

Use of Common Crawl for XLNet.

C4 dataset

Cleaned, curated version of Common Crawl for T5.

Raffel et al., 2020

Radford et al., 2019

WebText

Websites from highly-ranked Reddit posts for GPT-2.

- Web text, books, news, code, Wikipedia, dialog, multilingual data.

Pretraining Dataset Curation

(1) Some are specialized but most use varied domains.

Model	Represented Domains (%)									Filters		Data	
	Wiki	Web	Books	Dialog	Code	Acad	Pile	C4	M-L	Tox	Qual	Pub	Year
BERT	76		24				✗	✗			H	Part	2018
GPT-2		100					✗	✗			H	Part	2019
RoBERTa	7	90	3				✗	✔			H	Part	2019
XLNet	8	89	3				✗	✔			H	Part	2019
T5	<1	99					✗	✔		H	H	✔	2019
GPT-3	3	82	16				✗	✔	7%		C	✗	2021
GPT-J/Neo	1.5	38	15	4.5	13	28	✔	Part			C	✔	2020
GLaM	6	46	20	28			✗	✔			C	✗	2021
LaMDA	13	24		50	13		✔	✔	10%	C	C	✗	2021
AlphaCode					100		✗	✗			H	✗	2021
CodeGen	1	24	10	3	40	22	✔	Part			H	Part	2020
Chinchilla	1	65	10		4		✔	✔		H	C	✗	2021
Minerva	<1	1.5	<1	2.5	<1	95	✔	✔	<1%		C	✗	2022
BLOOM	5	60	10	5	10	10	✔	✔	71%	H	C	Part	2021
PaLM	4	28	13	50	5		✗	✔	22%		C	✗	2021
Galactica	1	7	1		7	84	✔	Part			H	Part	2022
LLaMA	4.5	82	4.5	2	4.5	2.5	Part	✔	4%		C	Part	2020

Pretraining Dataset Curation

(1) Some are specialized but most use varied domains.

Model	Represented Domains (%)									Filters		Data	
	Wiki	Web	Books	Dialog	Code	Acad	Pile	C4	M-L	Tox	Qual	Pub	Year
BERT	76		24				✗	✗			H	Part	2018
GPT-2		100					✗	✗			H	Part	2019
RoBERTa	7	90	3				✗	✔			H	Part	2019
XLNet	8	89	3				✗	✔			H	Part	2019
T5	<1	99					✗	✔		H	H	✔	2019
GPT-3	3	82	16				✗	✔	7%		C	✗	2021
GPT-J/Neo	1.5	38	15	4.5	13	28	✔	Part			C	✔	2020
GLaM	6	46	20	28			✗	✔			C	✗	2021
LaMDA	13	24		50	13		✔	✔	10%	C	C	✗	2021
AlphaCode					100		✗	✗			H	✗	2021
CodeGen	1	24	10	3	40	22	✔	Part			H	Part	2020
Chinchilla	1	65	10		4		✔	✔		H	C	✗	2021
Minerva	<1	1.5	<1	2.5	<1	95	✔	✔	<1%		C	✗	2022
BLOOM	5	60	10	5	10	10	✔	✔	71%	H	C	Part	2021
PaLM	4	28	13	50	5		✗	✔	22%		C	✗	2021
Galactica	1	7	1		7	84	✔	Part			H	Part	2022
LLaMA	4.5	82	4.5	2	4.5	2.5	Part	✔	4%		C	Part	2020

(2) Frequent use of web data and C4.

Pretraining Dataset Curation

(1) Some are specialized but most use varied domains.

Model	Represented Domains (%)									Filters		Data	
	Wiki	Web	Books	Dialog	Code	Acad	Pile	C4	M-L	Tox	Qual	Pub	Year
BERT	76		24				✗	✗			H	Part	2018
GPT-2		100					✗	✗			H	Part	2019
RoBERTa	7	90	3				✗	✓			H	Part	2019
XLNet	8	89	3				✗	✓			H	Part	2019
T5	<1	99					✗	✓		H	H	✓	2019
GPT-3	3	82	16				✗	✓	7%		C	✗	2021
GPT-J/Neo	1.5	38	15	4.5	13	28	✓	Part			C	✓	2020
GLaM	6	46	20	28			✗	✓			C	✗	2021
LaMDA	13	24		50	13		✓	✓	10%	C	C	✗	2021
AlphaCode					100		✗	✗			H	✗	2021
CodeGen	1	24	10	3	40	22	✓	Part			H	Part	2020
Chinchilla	1	65	10		4		✓	✓		H	C	✗	2021
Minerva	<1	1.5	<1	2.5	<1	95	✓	✓	<1%		C	✗	2022
BLOOM	5	60	10	5	10	10	✓	✓	71%	H	C	Part	2021
PaLM	4	28	13	50	5		✗	✓	22%		C	✗	2021
Galactica	1	7	1		7	84	✓	Part			H	Part	2022
LLAMA	4.5	82	4.5	2	4.5	2.5	Part	✓	4%		C	Part	2020

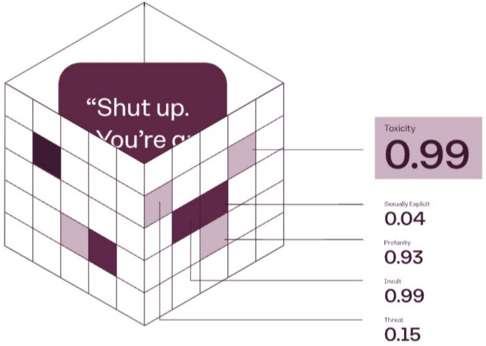
(3) Quality filters are applied.

(2) Frequent use of web data and C4.

Dataset Quality and Toxicity

- Quality filters:
  - Classifier & heuristics
  - −-defined & +-defined
- Toxicity filters:
  - Black-box commercially available APIs

- ***Lack of ground truth for definitions of quality.***

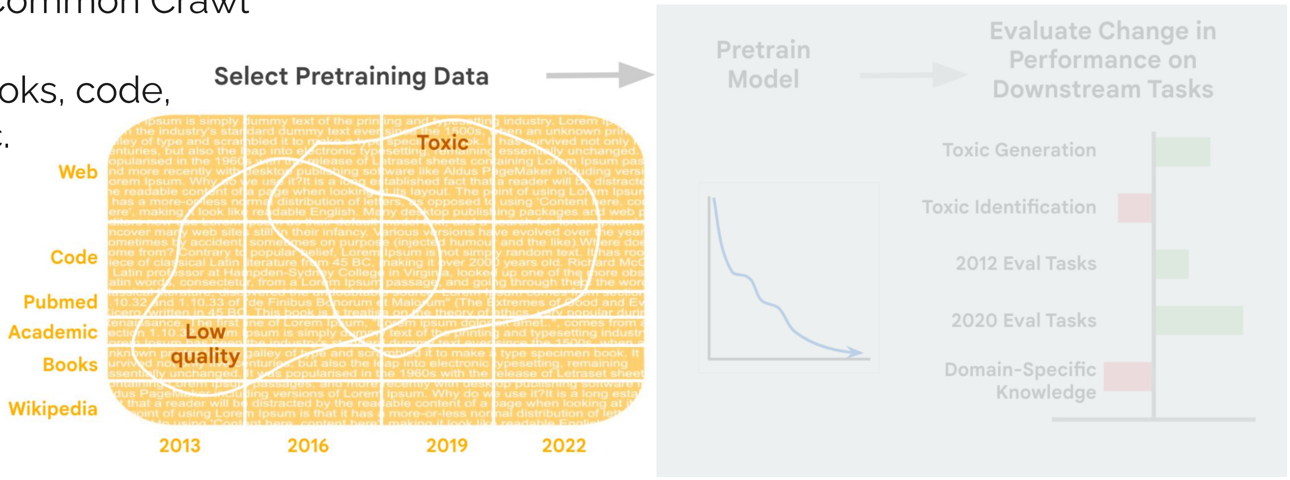


[Scaling Language Models: Methods, Analysis & Insights from Training Gopher (Rae et al., 2022)]
[Perspective API by Jigsaw]

Evaluation of Dataset Curation on Pretrained Models

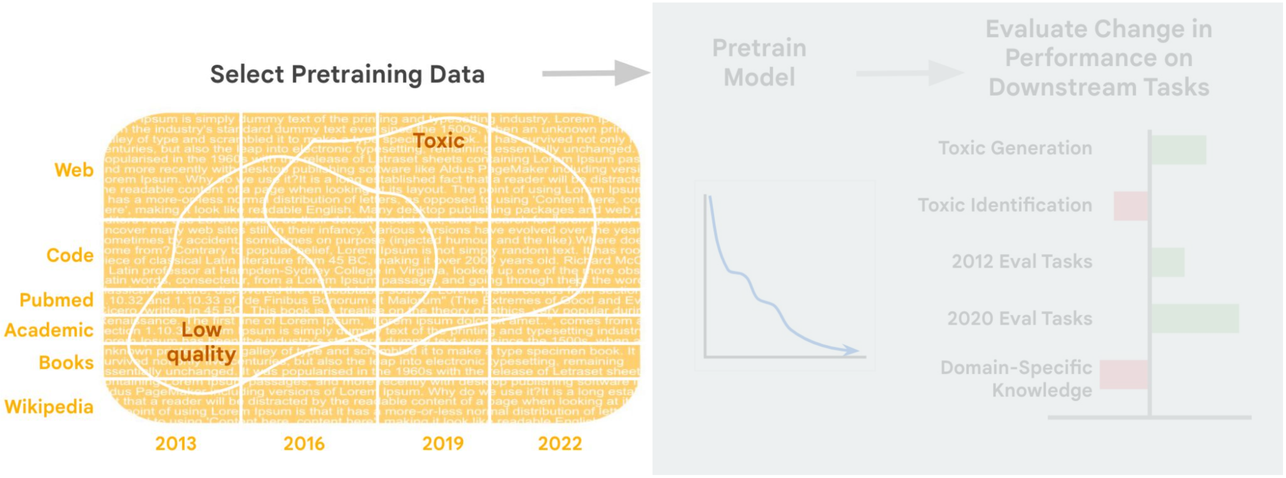
1) C4 (Raffel et al., 2020)
    - Cleaned version of Common Crawl
2) The Pile (Gao et al., 2020)
    - Common Crawl + books, code, various domains, etc.

- Both are deduplicated (Lee et al., 2022).



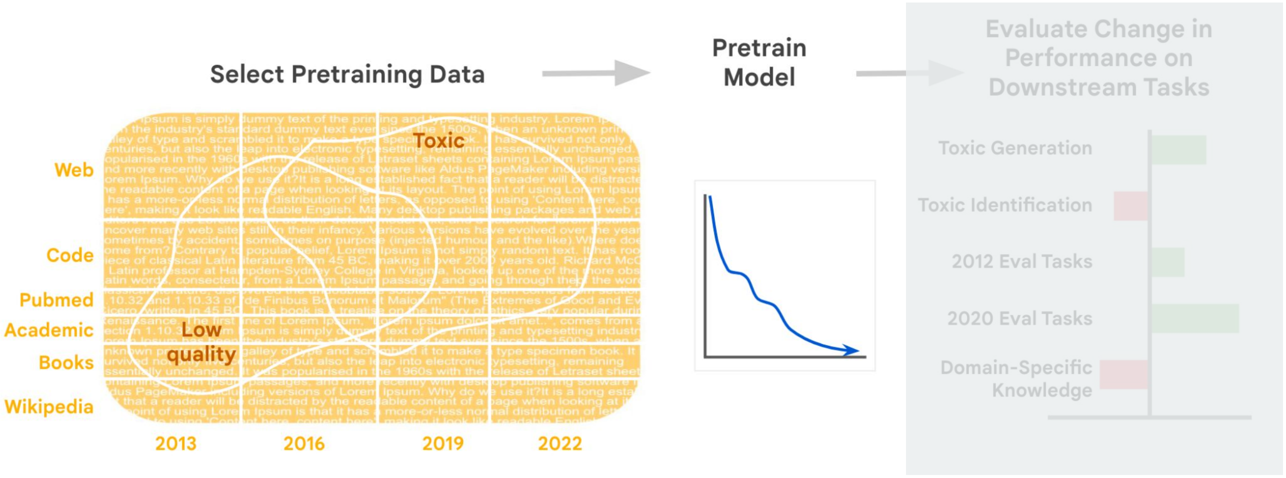
Evaluation of Dataset Curation on Pretrained Models

1) Dataset age
2) Quality and toxicity filtering
3) Domain composition



Evaluation of Dataset Curation on Pretrained Models

1) Dataset age
2) Quality and toxicity filtering
3) Domain composition

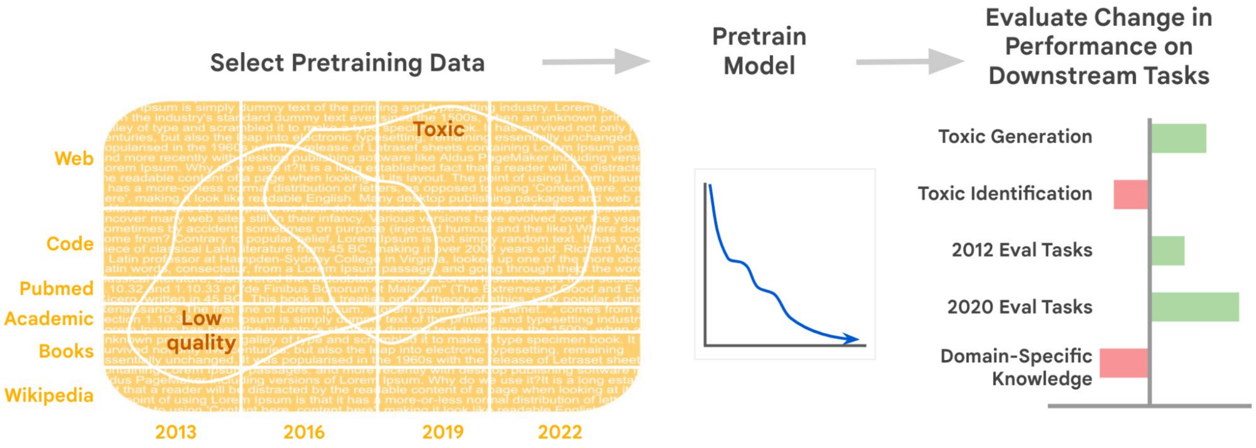


1) **LM-XL (1.5B)**
2) **LM-SMALL (20M)**

Evaluation of Dataset Curation on Pretrained Models

1) **Fine-tune**
2) **Evaluate**

1) Dataset age
2) Quality and toxicity filtering
3) Domain composition



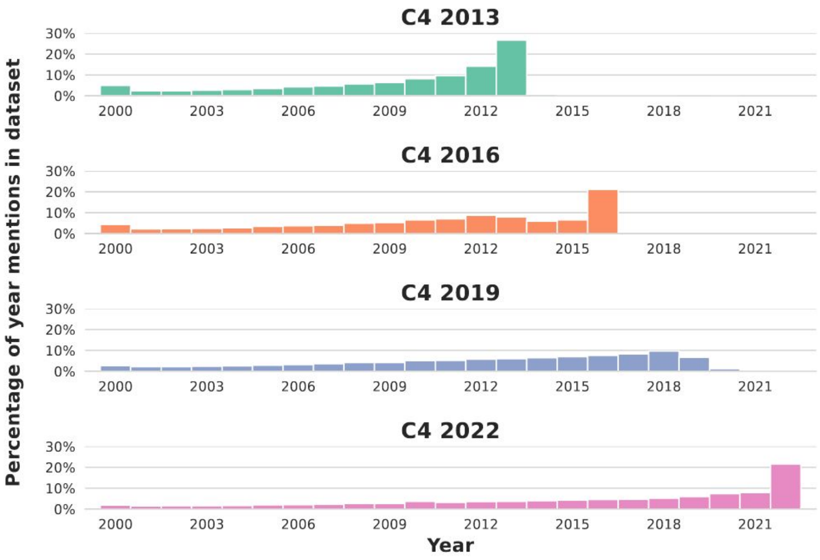
1) LM-XL (1.5B)
2) LM-SMALL (20M)

Impact of Data Curation on Data Characteristics

- Time
  - Percent non-ascii characters increasing.
  - Text quality decreasing.
- Toxicity and quality
  - Inversely correlated (e.g., books).
- Domains
  - Books are high quality but high toxicity.
  - Technical domains have lower quality. → poor filters

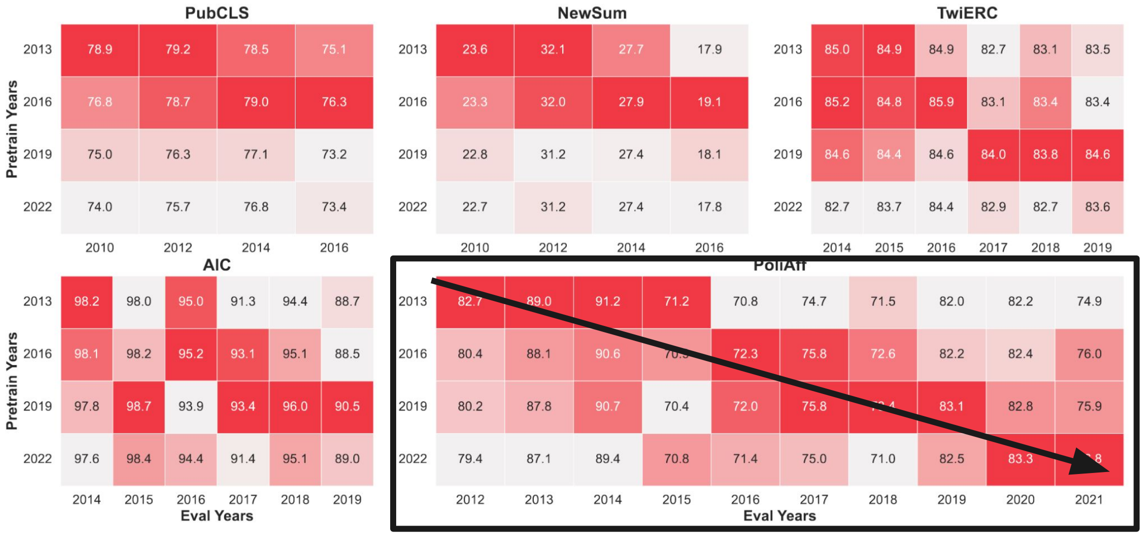
Impact of Dataset Age on Pretrained Models

- Majority of models downloaded on HuggingFace < 2020.
- Create time-snapshots of Common Crawl (C4).



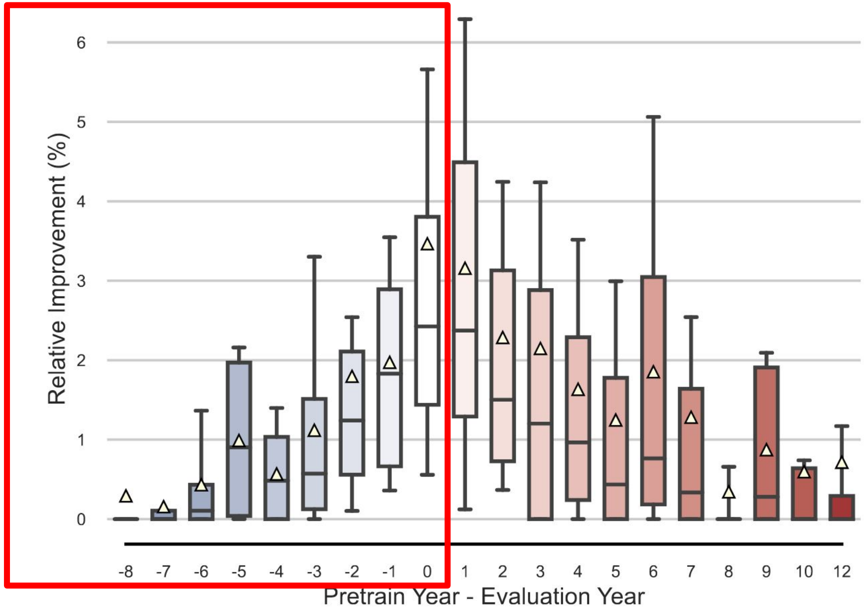
Impact of Dataset Age on Pretrained Models

- Temporal misalignment between pre-training and evaluation datasets impact performance negatively.
- More fine-tuning does not improve performance.



Impact of Dataset Age on Pretrained Models

- Impact is worse when pretraining precedes evaluation years.



Impact of Domain Composition on Pretrained Models

- Common Crawl, OpenWeb and Books have strongest positive effects.
- Domain heterogeneity (multiple sources, domains) is critical.

	Wiki	Web	Books	Biomed	Academic	Common Sense	Contrast Sets	Average
Full Dataset (100%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No Social (99%)	-0.8	-3.7	2.6	0.1	3.5	-3.5	3.5	0.4
No Wiki (98%)	-1.3	-5.3	3.0	0.2	0.9	-4.4	7.2	-0.3
No Books (93%)	-3.5	-6.3	1.0	0.0	-1.6	-6.5	-4.4	-2.7
No OpenWeb (93%)	-2.0	-4.1	0.1	-1.0	0.6	-5.8	-2.9	-1.4
No Legal (91%)	-2.7	-2.9	3.8	0.4	0.8	-2.6	-0.4	-0.6
No Academic (87%)	-0.3	-2.5	0.3	-0.9	2.2	-1.1	4.3	0.2
No Pubmed (85%)	-0.3	-3.0	3.9	-5.8	-1.5	-5.9	3.9	-1.2
No Code (81%)	-0.5	-3.1	2.9	-1.2	1.2	-5.8	4.4	-0.1
No CC (73%)	-3.2	-6.2	-2.9	-4.6	-5.9	-8.0	-5.2	-4.8

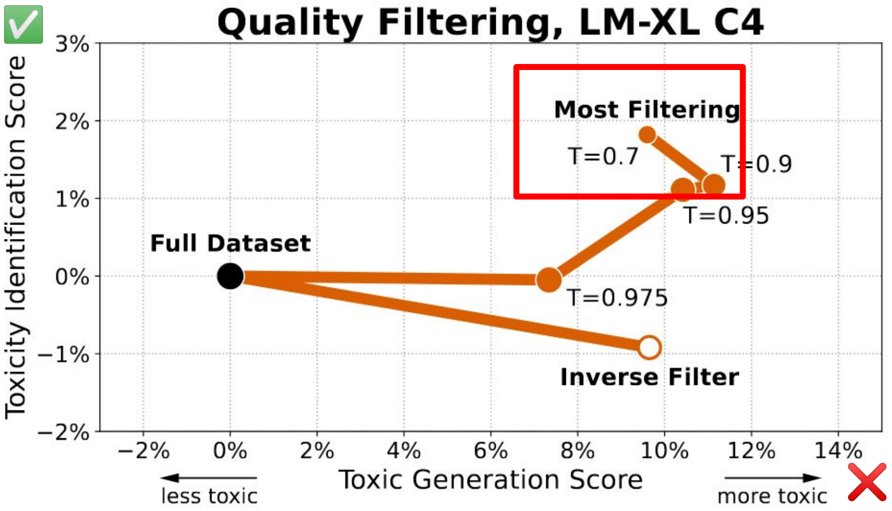
Impact of Quality/Toxicity Filters on Pretrained Models

- **Quality**: Use classifier employed by PaLM and GLaM
  - 0 (high, ✅) -> 1 (low)
  - High quality examples are books, certain webpages (Du et al., 2022)

- **Toxicity**: Jigsaw's Perspective API
  - 0 (low, ✅) -> 1 (high)
  - Has been shown to be unreliable (Pozzobon et al., 2023)

Impact of Quality/Toxicity Filters on Pretrained Models

- Quality filters improve **(1) toxic identification** and (2) QA tasks.



Impact of Quality/Toxicity Filters on Pretrained Models

- Quality filters improve (1) toxic identification and **(2) QA tasks**.

	Wiki	Web	Books	Biomed	Academic	Common Sense	Contrast Sets	Average
Inverse T=0.5 (73%)	-5.0	-4.5	2.1	-2.2	-2.7	1.2	-6.4	-3.1
Full Dataset (100%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T=0.975 (91%)	1.2	0.7	-2.2	6.1	6.4	4.7	6.1	2.5
T=0.95 (84%)	-1.2	1.0	-4.0	3.7	-0.3	3.2	4.9	1.0
T=0.9 (73%)	-0.3	0.8	-3.5	1.8	1.0	1.9	6.8	1.2
T=0.7 (46%)	-1.2	0.8	-6.7	1.7	0.8	2.0	4.2	0.7

Impact of Quality/Toxicity Filters on Pretrained Models

- Toxicity filters worsen downstream task performance.
- Likely due to toxicity and quality tradeoff.

	Wiki	Web	Books	Biomed	Academic	Common Sense	Contrast Sets	Average
Inverse T=0.06 (92%)	0.4	-1.4	3.8	0.7	4.9	4.1	2.7	1.7
Full Dataset (100%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T=0.95 (98%)	-1.0	-0.4	0.2	-0.5	0.6	1.7	1.3	0.2
T=0.9 (95%)	-2.2	-1.1	-0.6	-3.0	0.2	2.9	0.2	-0.7
T=0.7 (86%)	-2.1	-1.4	0.1	-2.9	0.1	-0.9	-0.2	-1.2
T=0.5 (76%)	-4.2	-2.4	-0.9	-3.3	-1.1	-0.3	-0.1	-2.0
T=0.3 (61%)	-3.8	-4.4	-1.4	-2.5	-0.3	-1.3	-3.5	-2.7

References

Shayne Longpre, Gregory Yauney, Emily Reif, Katherine Lee, Adam Roberts, Barret Zoph, Denny Zhou, Jason Wei, Kevin Robinson, David Mimno, Daphne Ippolito. 2023. “A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity.” arXiv. https://arxiv.org/abs/2305.13169.

Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. 2022. “Scaling language models: Methods, analysis & insights from training Gopher.” arXiv. https://arxiv.org/abs/2112.11446.

Thao Nguyen, Gabriel Ilharco, Mitchell Wortsman, Sewoong Oh, Ludwig Schmidt. 2023. “Quality Not Quantity: On the Interaction between Dataset Design and Robustness of CLIP.” arXiv. https://arxiv.org/abs/2208.05516.

Tomas Bratanic. Fine Tuning Retrieval Augmented Generation. https://neo4j.com/developer-blog/fine-tuning-retrieval-augmented-generation/.

Shibani Santurkar, Yann Dubois, Rohan Taori, Percy Liang, Tatsunori Hashimoto. 2022. “Is a Caption Worth a Thousand Images? A Controlled Study for Representation Learning.” arXiv. https://arxiv.org/abs/2207.07635.

Luiza Pozzobon, Beyza Ermis, Patrick Lewis, Sara Hooker. 2023. “On the Challenges of Using Black-Box APIs for Toxicity Evaluation in Research.” arXiv. https://arxiv.org/abs/2304.12397

Du, N., Huang, Y., Dai, A. M., Tong, S., Lepikhin, D., Xu, Y., Krikun, M., Zhou, Y., Yu, A. W., Firat, O., et al. 2021. “GLaM: Efficient scaling of language models with mixture-of-experts.” https://arxiv.org/pdf/2112.06905.

Questions?

Discussion Questions

Discussion Questions

Scaling Laws of Synthetic Images for Model Training ... for Now

1. How can we use the real vs. synthetic set up to learn about the synthetic+real data case?
2. Which do we care more about: Scaling laws for image classifiers or Scaling laws for representation learning algorithms? Why is there a gap between these?
3. Why does Stable Diffusion do so much better on the sketch and artistic rendering OOD imagenet datasets?
4. Why are the "poor" classes so difficult?

A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity

1. How will the intuitions gained from this research guide the way we approach fine-tuning models for new tasks?
2. Will the trends we see here apply to areas of ML/DL outside of LLMs?
3. How can we determine when older datasets will start negatively impacting training tasks?
4. How do we address the tradeoff between quality and toxicity?