

GAMIFICATION OF LLMS AND ITS' CONSEQUENCES

By Andrew Repine
and Nhan Tran

INTRODUCTION TO LLAMA



- Collection of Models
- Smaller models
- Public Data
- Research First

LLAMA TRAINING

- Trained using AdamW optimizer
- Reward model via collected human preference data as binary ranking labels.
- Trained over 1 epoch, over training led to overfitting.

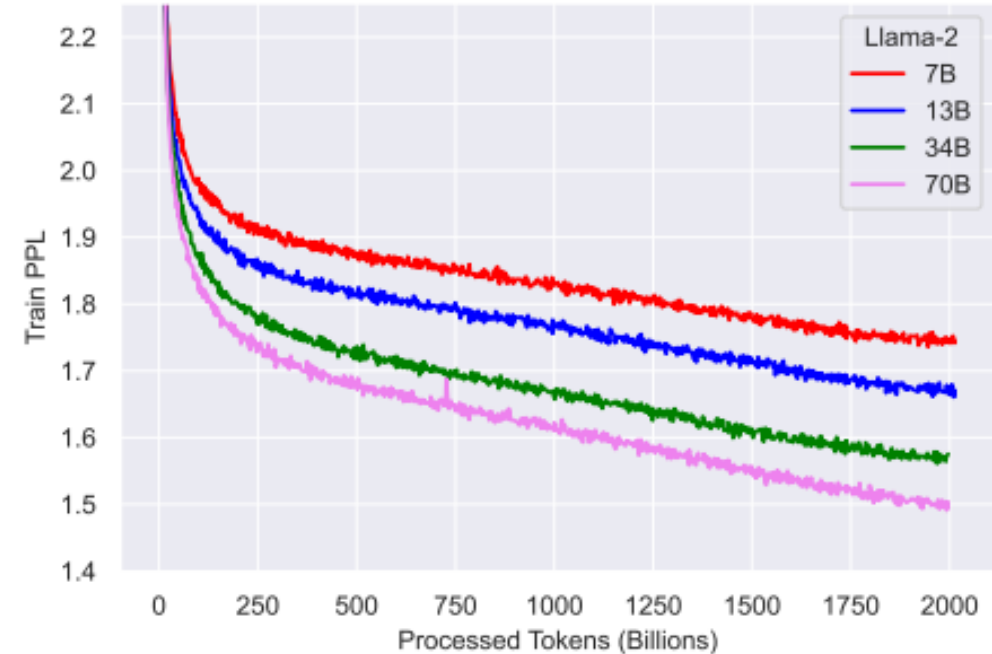
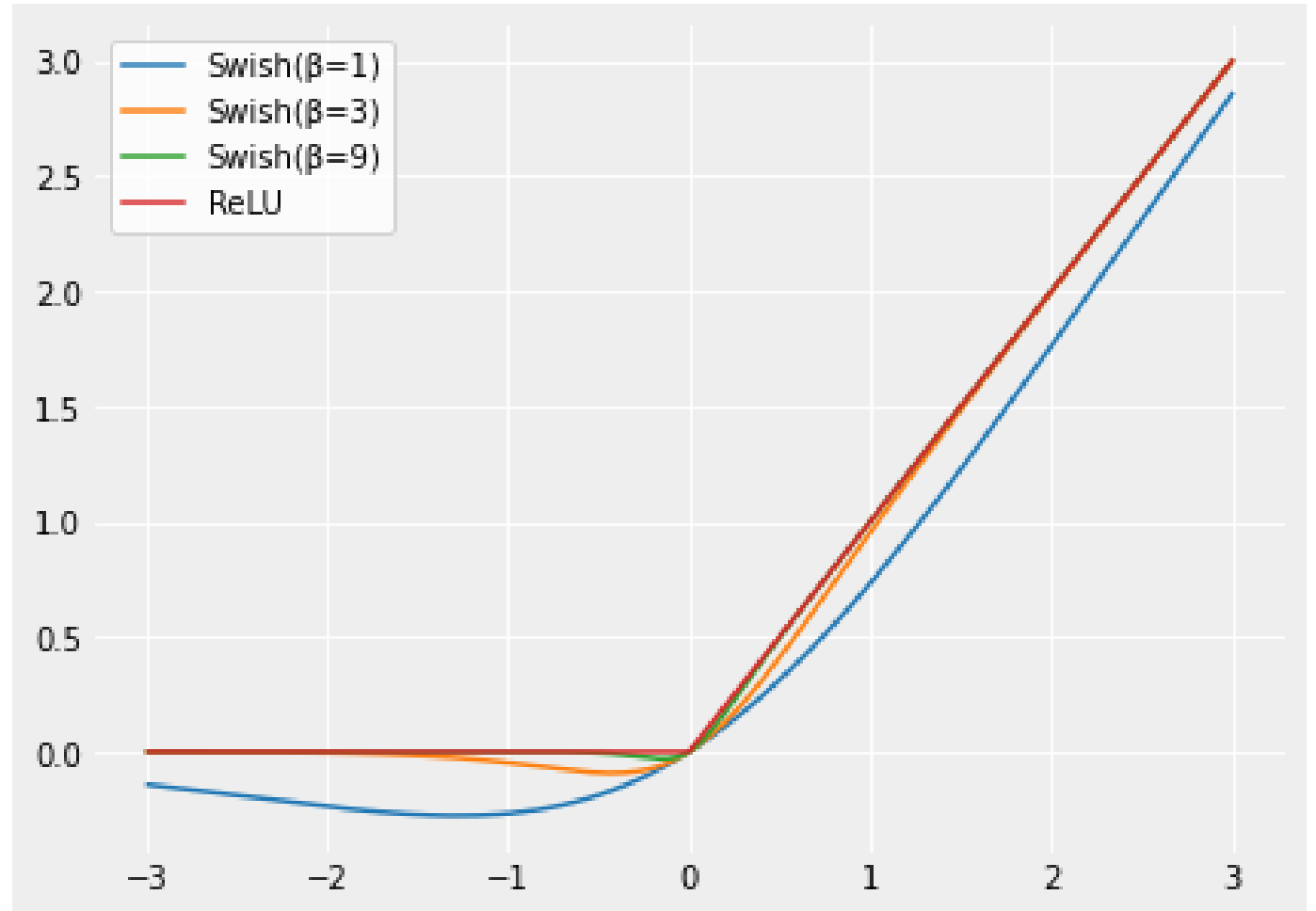


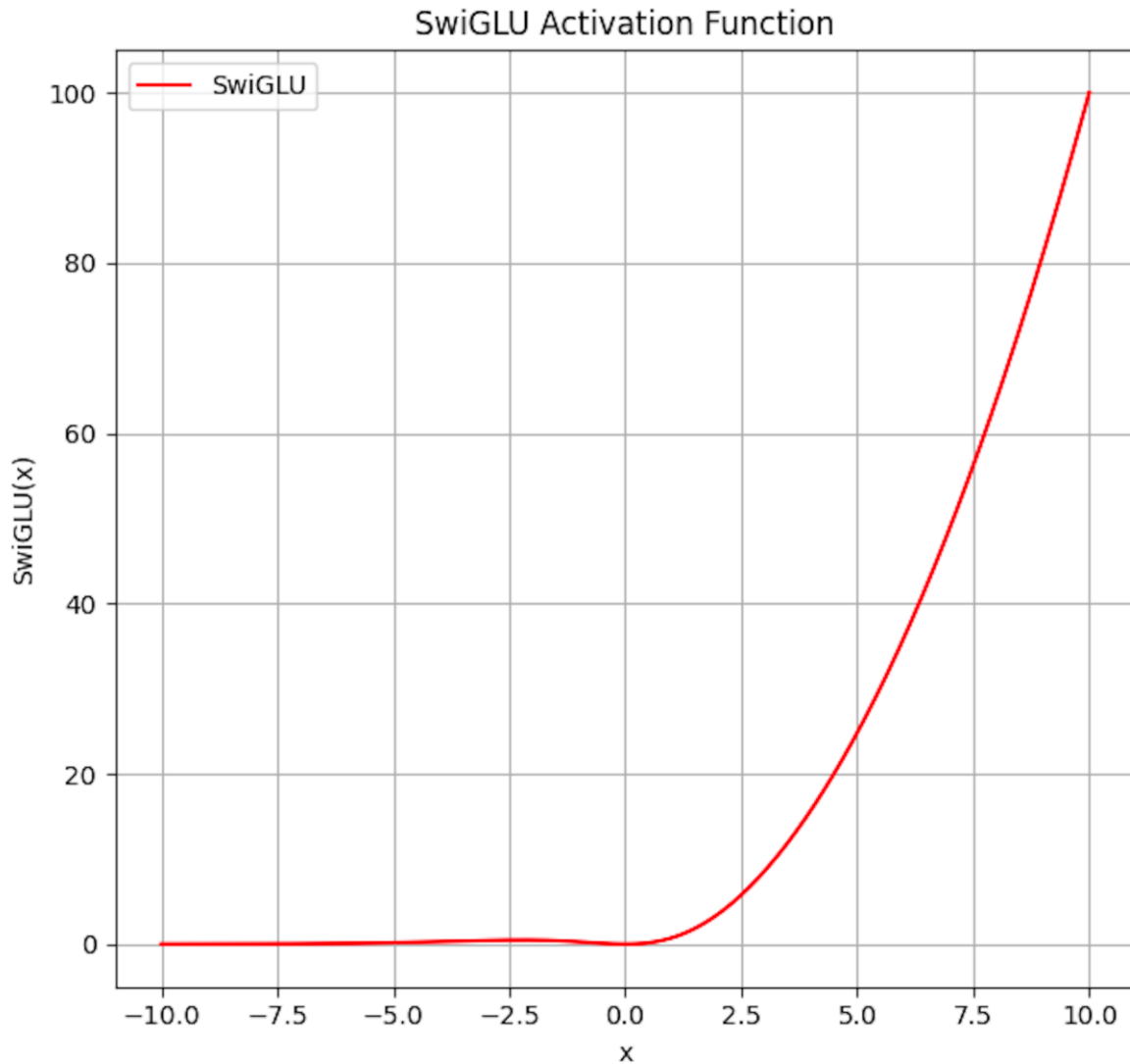
Figure 5: Training Loss for LLAMA 2 models. We compare the training loss of the LLAMA 2 family of models. We observe that after pretraining on 2T Tokens, the models still did not show any sign of saturation.

LLAMA TRAINING

- Swiss Function & Gated Linear Function
 - Swiss: $f(x) = x * \text{sigmoid}(\beta x)$
 - Non-monotonic activation function. Performs better than ReLU with better optimization and faster convergence.
 - GLU: $\text{GLU}(x) = x * \text{sigmoid}(Wx + b)$
 - Gated by sigmoid activation function.



LLAMA TRAINING



- SwiGLU Function

- $SwiGLU(x) = x * \text{sigmoid}(\beta x) + (1 - \text{sigmoid}(\beta x)) * (Wx + b)$

LLAMA TRAINING

- Training data used
 - Common crawl is a database of the downloaded internet that is often used in LLM training
 - C4 is a preprocessed part of common crawl
 - From Github only Apache, BSD, and MIT licensed projects are used, filtered to remove poor projects
 - The rest of the data is meant to compile a comprehensive subset of the internet and human language that allows the LLaMa models to be so effective

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

LLAMA SIZE

- LLaMa, despite being smaller by around 3 times comparing the 175 billion parameter GPT-3 model to the 65 billion parameter LLaMa model, easily outperforms GPT-3
- This proves the idea that smaller models trained on more data can be more effective
- This also allows the model to be available to more people since it requires less architecture to run and use the model

		0-shot	1-shot	5-shot	64-shot
GPT-3	175B	14.6	23.0	-	29.9
Gopher	280B	10.1	-	24.5	28.2
Chinchilla	70B	16.6	-	31.5	35.5
	8B	8.4	10.6	-	14.6
PaLM	62B	18.1	26.5	-	27.6
	540B	21.2	29.3	-	39.6
	7B	16.8	18.7	22.0	26.1
LLaMA	13B	20.1	23.4	28.1	31.9
	33B	24.9	28.3	32.9	36.0
	65B	23.8	31.0	35.0	39.9

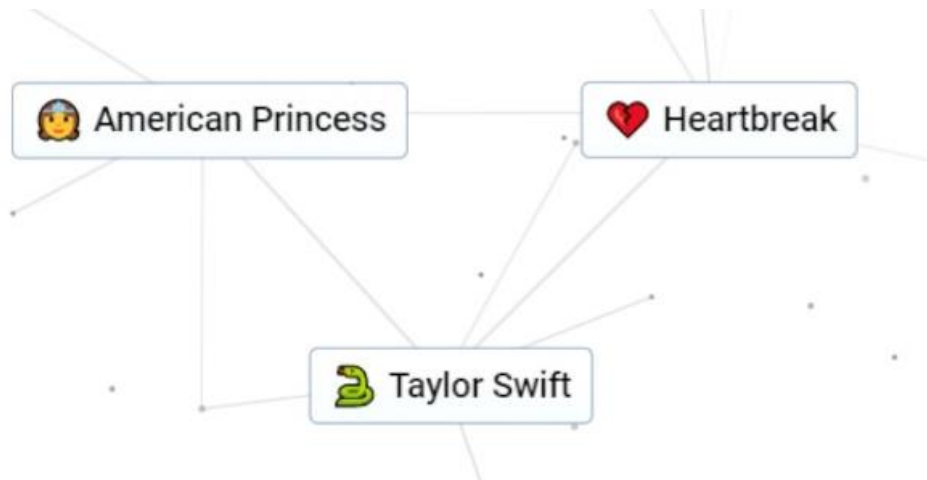
Table 4: NaturalQuestions. Exact match performance.

		RACE-middle	RACE-high
GPT-3	175B	58.4	45.5
	8B	57.9	42.3
PaLM	62B	64.3	47.5
	540B	68.1	49.1
	7B	61.1	46.9
LLaMA	13B	61.6	47.2
	33B	64.1	48.3
	65B	67.9	51.6

Table 6: Reading Comprehension. Zero-shot accuracy.

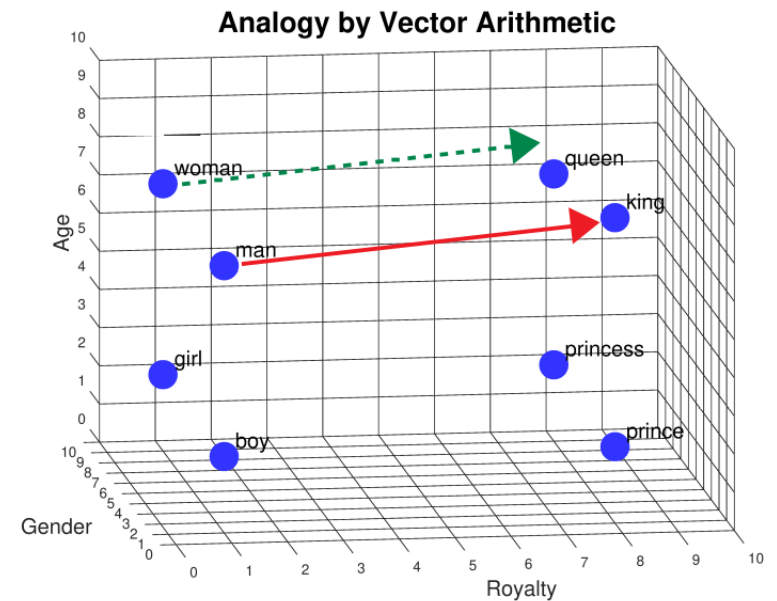
INFINITE CRAFT – A CASE STUDY

- The “recipes” are entirely automated and generated by the AI, leading to funny, but still logical combinations that are entirely infinite
- The LLaMa 2 model behind the program allows for new discoveries to be made when someone creates something never created before



INFINITE CRAFT AND WORD EMBEDDINGS

- Much like the idea that math operations can be done on word embeddings for example, the vector KING – QUEEN is similar to that of MAN – WOMAN, infinite craft lets us peer into infinite addition of word embeddings
- Although the program doesn't truly work like this, the idea remains, and the game gives us tool to play with adding word embeddings



IMPACT OF AI ON THE SPACE

- Flexibility and much faster in content generation.
 - Sora, Infinite Craft, AI Dungeon are a few example
- Chatbot functionality can be used to make much more live-like interaction in games and less structured games.
- As AI models become more advanced, their impact will likely be felt even more on the gaming space, but the issue of that is yet to be seen

ETHICAL CONSIDERATIONS

- Biases
 - LLMs need to be tested and evaluated on biases and output to make sure they are reasonable when it comes to assumptions or errors
 - Two of the most important test are CrowS-Pairs and TruthfulQA which intend to validate bias and prevent hallucinations, respectively
 - The results are show on the right with LLaMa out performing GPT-3

	LLaMA	GPT3	OPT
Gender	70.6	62.6	65.7
Religion	79.0	73.3	68.6
Race/Color	57.0	64.7	68.6
Sexual orientation	81.0	76.2	78.6
Age	70.1	64.4	67.8
Nationality	64.2	61.6	62.9
Disability	66.7	76.7	76.7
Physical appearance	77.8	74.6	76.2
Socioeconomic status	71.5	73.8	76.2
Average	66.6	67.2	69.5

Table 12: **CrowS-Pairs**. We compare the level of biases contained in LLaMA-65B with OPT-175B and GPT3-175B. Higher score indicates higher bias.

		Truthful	Truthful*Inf
GPT-3	1.3B	0.31	0.19
	6B	0.22	0.19
	175B	0.28	0.25
LLaMA	7B	0.33	0.29
	13B	0.47	0.41
	33B	0.52	0.48
	65B	0.57	0.53

Table 14: **TruthfulQA**. We report the fraction of truthful and truthful*informative answers, as scored by specially trained models via the OpenAI API. We follow the QA prompt style used in Ouyang et al. (2022), and report the performance of GPT-3 from the same paper.

ETHICAL CONSIDERATIONS

- Content being reproduced
 - In many cases LLMs are trained on user work that is often reproduced and stolen, one of the making conflicts when it comes to using these for commercial products
- Without filtered responses, these LLMs amplify biases seen in training data, so a proper training technique needs to be followed when training LLMs

ETHICAL CONSIDERATIONS

- Training LLMs takes a large amount of power and hardware that leaves a quite significant carbon footprint
 - A quote taken from the LLaMa paper says, “... developing these models would have cost around 2,638 MWh under our assumptions, and a total emission of 1,015 tCO₂eq.”
 - Similarly, by releasing pre-trained models, the hope is to prevent future impact by only having the models have to be trained once

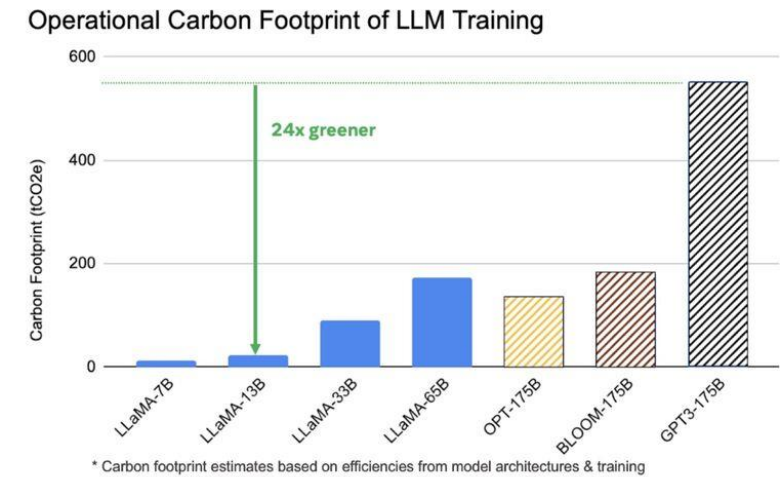


Image credit to Yann LeCun

DISCUSSION QUESTIONS

- In what ways can AI improve the gaming space?
- How should we try to limit the creation of these AI models?
- How can we create accessible AI models that don't place technological or financial barriers on the use of AI?
- Thank you for the time, do you guys have any questions for us?