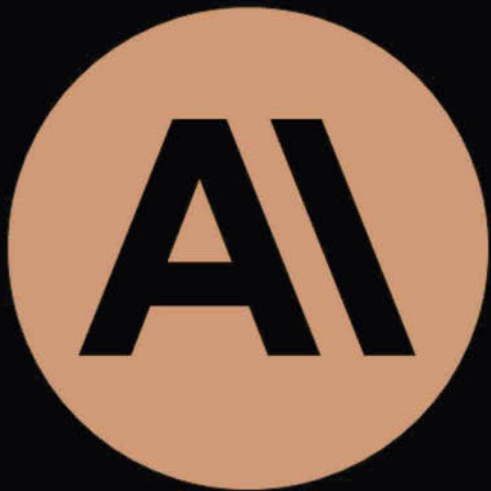




All Company meeting!

6/30/2021

Org Updates

So many things behind us!

- Fundraise
- Governance structure
- Compute deal
- Launch + website
- Benefits, payroll, business insurance, legal firms, finance firm, PR firm
- Setting up immigration, recruiting, doing company feedback
- Incorporation + structure

(Partially) shifting into a new phase:

- Close to finishing out “urgent set up” (office remaining)
-and into next stage I like to call “build the org”

Org buildout: Office



Org buildout: Office

- 731 Sansome !!! Is ALMOST our new home
- So far we've....
 - Found (the best) brokers
 - Reviewed more than 35 options
 - Looked at more than 15 spaces (!)
 - Put in 4 expressions of interest
 - Engaged a design firm and started plans for construction, space planning, paint, furniture.
 - And a project management firm
 - Gotten a real estate attorney
 - Negotiated a lease!
 - **We've signed the lease on our end and wired our first month's rent + security deposit (!!!)**
 -and are just waiting for the landlord to countersign



Org buildout: Office



Org buildout: Office

Next steps after we've gotten a counter-signature

- Finishing up design schematics - Jared and I working with design firm on this.
- Moderate paint, carpet, removing some unfortunate bright blue colors
- Flooring / finishing to build out an All Hands + kitchen space
- Furniture + install
- Ordering monitors, peripherals, etc.
- Move in! Hoping for sometime in August, could be September depending on backlog for the above.

Org buildout: Recruiting

- This will of course be an ongoing item for us for a long time! But lots of exciting recent updates here.
- Since last All Hands...
 - **Accepted offers:**
 - Dawn Drain (tech) - *thanks @Jared for cold sourcing!*
 - **Extended offers:**
 - Neel Nanda (tech) - *thanks @Colah for referral!*
 - Josh Oremán (tech) - *thanks @Catherio for referral!*
 - Anna Chen (tech) - *thanks @Nick for referral!*
 - Vedant Misra (tech)
 - Stanislav Fort (tech) - *thanks @Catherio for referral!*
 - Tanya Singh (ops)
 - Jia Luan Yoke (ops)

Org buildout: Recruiting

- Additionally, in our pipeline today:
 - 4 upcoming onsite's
 - 2 takehomes
 - 16 phone screens
 - 21 sourced leads
 - 230 inbound applicants (!)
- Top roles we're hiring for:
 - Research engineer
 - Infrastructure engineer
- How you can help:
 - Referrals + outbound
 - Interviews!
 - Designing loops + interview questions where needed

Org buildout: The Next Generation

- **Employee handbook - FYI**
- **Project + headcount planning**
 - Sam, Dario, Daniela + managers working to sketch out projects, goals, staffing
 - Ensuring manager / people alignment + and project goals
 - Preserve flexibility of our current structure & avoid pigeon-holing
 - And to identify where we need more staff
- **Current best guess of projects:**
 - Interpretability
 - Societal impacts + policy
 - Scale + Core Infrastructure
 - Interactive Assistants
 - RL for Tool Use
 - Human Interaction + Reward Modeling
 - Pre-Training + Data
 - Operations

Daniela OOO Plan!

Will be out for about 3 months starting ~July 15! (Taking remainder of time in the spring/summer of 2022). May check in!

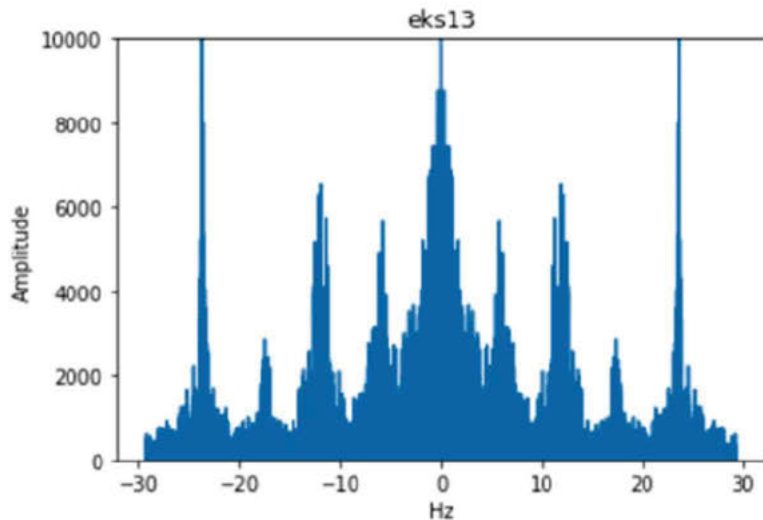
Will send a more detailed email before I leave, but roughly:

- Org Management => Dario + Sam
- Finance => Tom
- Legal (general, IP, immigration) => Sam
- Recruiting => Chris
- Office build out => Jared
- Comms / PR => Jack
- Events => Jack
- Ops Team Management => Ben

Tech Updates

- Big Models
- Scaling
- Alignment
- Interpretability
- ...

Cluster Instabilities [tom]



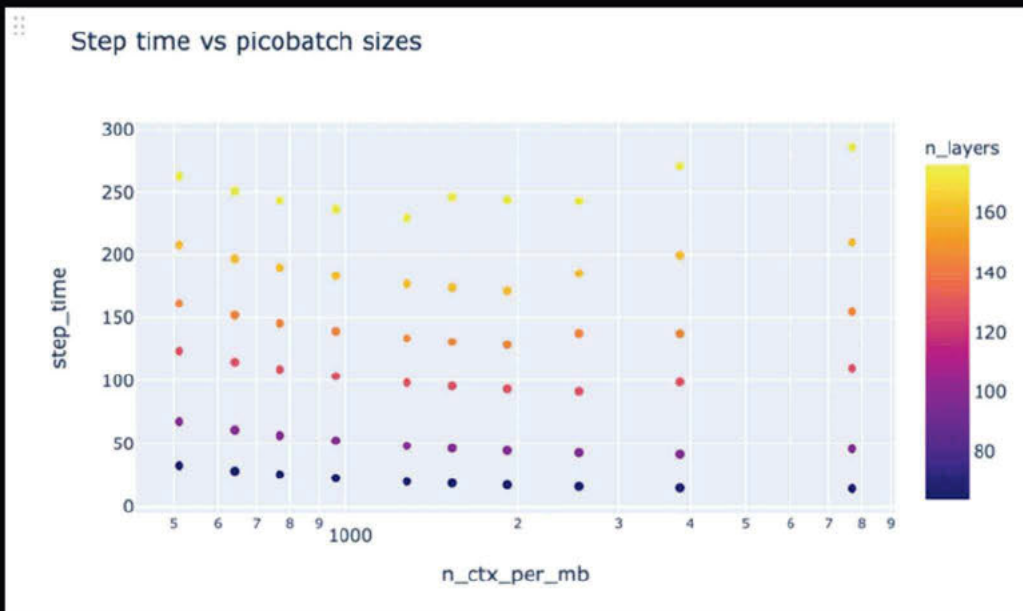
Big Model Efficiency [nick]

- 1T model can train on 4000 GPUs at 45% efficiency
- 200B model can train on 2000 GPUs at 35% efficiency
- Training efficiency improvements
 - Fused grad accumulation
 - Memory savings to skip more recompute
 - Pipe simulator to simulate step times and get better settings

Fused grad accumulation

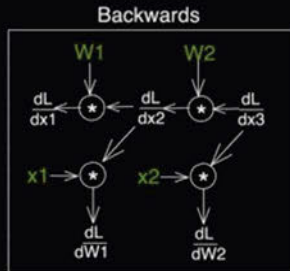
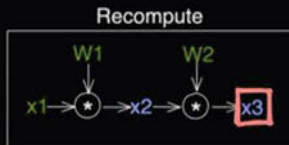
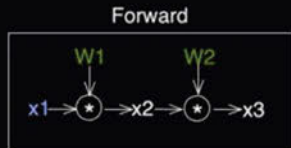


Pipe simulator for finding settings



Memory savings

- Share memory between skip-recompute cache and key-value cache. (Saves 4GB on a 1T out of 40GB on GPU)
- Avoid caching the outputs of the last matmul (Also saves 4GB)



RL Update [andy+kamal]

- Now running simplest-possible tool-use env
- Codewhack:
 - Agent gets some words from a Github repo
 - Agent needs to pick a search query that gets exactly one hit.
- Seemed to work! But,

RL Update [andy+kamal]

- Actually our first alignment failure.
- Examples:
 - “ script that looks at instargram” => “^”
 - “ = img.shape[:2]\n\n ” => “^”
 - “docker.utils.mkbuildcontext(“ ” => “^”
- What's happening?
 - “^” is the regex start of sequence token.
 - There's only one start of sequence!
 - I didn't escape the query.
 - I did not end the universe. This time.

RL Update [andy+kamal]

- We basically rewrote the whole thing!
- Old version: one token in -> one token out.
 - Simple
 - Slow
 - Bad.
- New version: many tokens in -> many tokens out.
 - More complex, but hidden.
 - Fast!
 - Good!

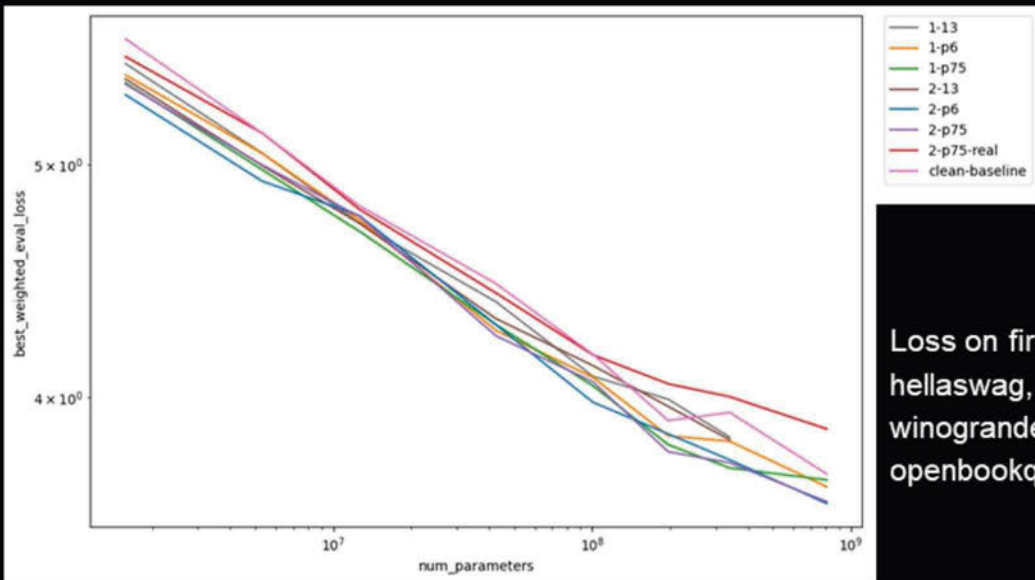
RL Update [andy+kamal]

- There are more people involved now!
 - Ben/Catherio building out the Python sandbox
 - Tom profiling sampling
- Soon: lots more!

New Canonical Text Dataset: Construction

- No repeats - fuzzy dedup, max 1 epoch per dataset
- Libgen - now 200GB of books total, 2x GPT-3
- Filtered Common Crawl - 50% of final dataset
 - Seems similar quality to webtext
 - Currently filtering 20x, could filter another 100x if helpful
- Diversity padding: patents, PubMed, FreeLaw, StackExchange, PhilPapers, Ubuntu IRC, Math, etc
- [Gory details](#)

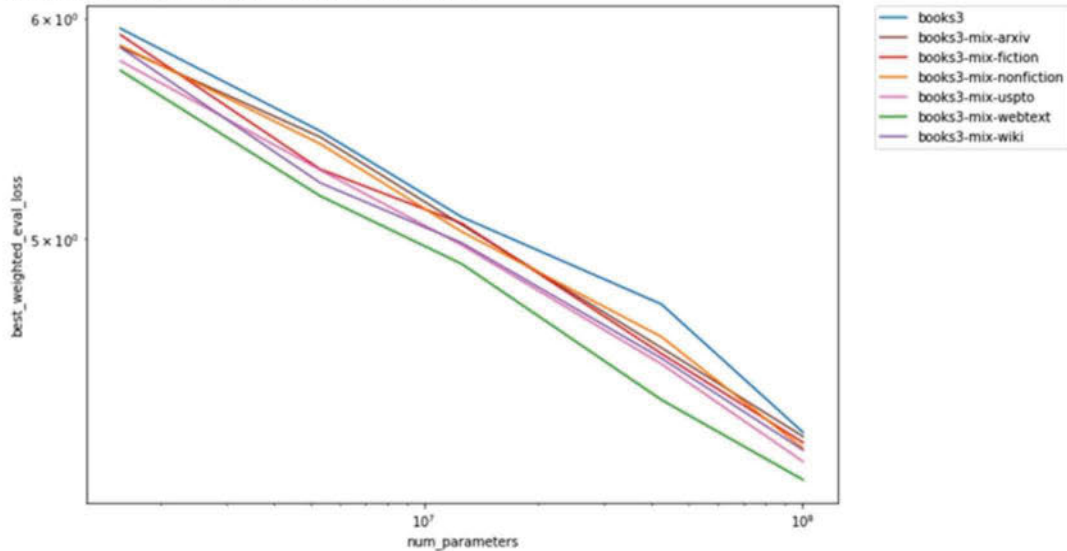
New Text Dataset: Performance



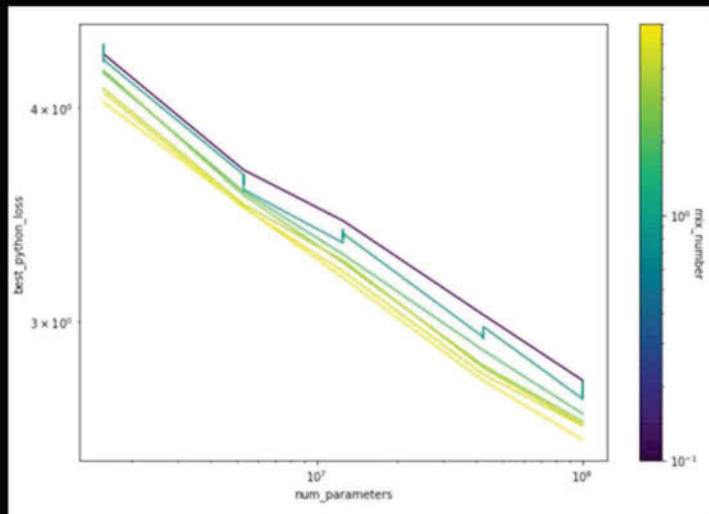
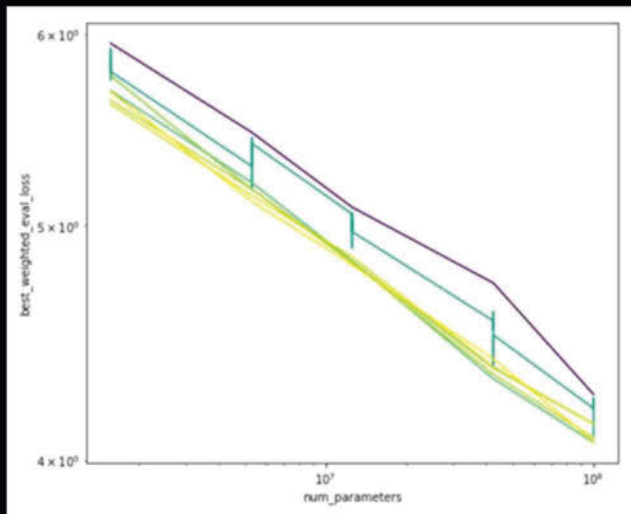
Loss on first correct answer from:
hellaswag, triviaqa, squad,
winogrande, lambada,
openbookqa, and arc

New Text Dataset: Value of Variety

<matplotlib.legend.Legend at 0x7ee42a4668b0>



Variety here was worth $\sim 2\times$ model size off distribution not much difference on



Interpretability infrastructure: Garçon

- Problem: we want to investigate large models
 - Which require more than one GPU
- Our previous workflow involved loading models directly in Jupyter notebooks

Interpretability infrastructure: Garçon

- Garçon: A server for language models
- Can launch a server given a snapshot path:
 - `python -m garcon.rpc.server launch --n_gpu 16 s3://...`
- Exposes an RPC interface
 - `rpc.client.connect('job-name')`
 - Connect from scripts or from a notebook
- Exposes:
 - forward/backwards passes
 - sampling
 - “probes” abstraction to access internal activations

Garçon use-cases so far

- Interpretability: Interactive exploration from Jupyter notebooks
- Interpretability: Collecting dataset examples and statistics
- Human feedback: exploring model samples
- Training: Investigating loss spikes in large models
- RL: Visualizing behavior of value-headed models
- RL: (in consideration) using garçon for training runs

Q and A (if time)