Frontier Builders: Open Research in Action

Ion Stocia:

- To build a model you need 3 things
    - Compute
    - People
    - Data
- Look at OS Research, people thought OS research was dead, Ion Stocia talked about Microsoft's dominance. https://t.co/beEoVYj84B
- A ton of people believe in 2 years AGI will be achieved and everything will be solved
- How can we standardize shared stacks for full AI open source?

Bryan Catanzaro

- Open AI can be safe, we don't need AI priests to keep it safe
- The enlightenment showed us it's fine to explore
- Metaphor about how: People just invented electricity and we are trying to figure out how it's going to kill people
- A good narrative is that AGI will be as diverse as the people using rather then AGI will kill all fields
- Nemotron doesn't necessarily get a ton of people joining it
- Chinese open source models in Qwen, Kimi, GLM, etc just decided to be more open for their country. How can we learn from that?
- Nemotron can run 1000 ablation experiments but a lot of people will be like, why did you do that? There needs to be a change in the ML community specifically
- Researchers don't want to document anything typically, we have to change that culture.
- OpenMDW License is fire, just Apache 2.0 with AI

Percy Liang:

- We are very early
- Must build Open AI that is amazing for people to use
- If there was a clear alternative, a ton of students would pursue that
    - Not everyone agreed here
- Speed over good documentation and sharing leads to incomplete open source
    - The way they minimize that is just using Discord and GitHub for all their experiments

  You can release data, weights, and experiments
- Only thing that's super easy (except for ego) is experiments
- ML vs Software
    - ML → Paper must be correct, very toxic at times

    - Software → If it's buggy that's fine, more organic iterative process, buggy culture is healthy

Nathan Lambert:

- Speed over good documentation and sharing leads to incomplete open source
- More people should bring experiment chains together. So that each checkpoint can be recreated
- Researchers (esp pretraining) don't want to document anything, though post-training teams do a little more. This is a culture issue.
- Build a culture where a company/project does their best but then OSS takes care of the rest.

Andy Konwinski:

- Comp is bottlenecking PhD students, he talked about this in the accesspod too. PhD students getting crazy million dollar offers and those offers going even higher due to equity.
- Open doesn't have to be completely open

Resourcing an Open Research Ecosystem

Remzi Arpaci-Dusseau

- THE GOAT NO INTRO NEEDED
- The DB research and distributed computing research at Wisconsin
    - There was more than just Professors and PhD students that were building everything out
- 99% of the population don't necessarily like AI
    - Fear
    - Distrust
    - Notion of not the best people leading these things

David Siegel

- Co-Founder of Two Sigma and Philanthropist
- The MIT Schwarzman advisory board would help at looking at projects at MIT and Beyond
- Wants to build out the idea of higher technical people joining research
    - OpenAthena
- Philanthropy can serve as venture capital like a Series A for results in projects
- If a PI hires a SWE, they're hired by the lab, the SWE can't go to another lab very easily. So this turns off a lot of SWEs. And now there's a career for SWE in Research

- Results: Marin has 6-7 engineers
- There's urgency, the changes are happening very quickly
- Theory: Large level of secrecy, because of that there's a lot of distrust. Openness leads to more trust

Jennifer Chayes

- Been at Berkeley for 6.5 years, Microsoft for 23 years.
- Very strong zero sum game mentality, had to start raising philanthropy very quickly
    - If we had more students and faculty doing computing
        - There would be less money for others
- Raised 750 million in private philanthropy
- She really liked the culture of open source
- Dean vs University President
    - She wants to be close enough to the research level
    - She raises a lot of money for professorship that may not get resourced through the normal processes.
- What's the next Jupyter?
    - Some things in materials, biomedical, etc
- In a lot of their newest gifts they're getting research scientists for the next few years. To keep the pressure off.
- How to enable coordination of the west building the open frontier models
- Close sourced model companies use the fear to pump the stock basically lol
- Marketing campaigns → I personally don't think this works, you need boots on the ground
- Why do we not have something in AI that's like cryptography? Models that we can build that have something like a one way function in cryptography.
    - Something analogous to one way keys?
- Get something like OpenKey or Cryptography for AI models would be a good Turing award accomplishment
- We're all building on Qwen and Chinese models Erwin! LOCK IN!!

Erwin Gianchandani

- NSF TIP Directorate
- Background in CS and Biomedical Engineering
- Shift in Fundamental Research
- 15 years ago, CS + X. It was very provocative.
- More about how can we solve problems (use inspired) for real world impact
- Small teams with funding over long periods can make huge breakthroughs (RAID, RISC, etc)

- Don’t settle the government short either
    - The messages from The White House
        - Experimentation is being appreciated
        - Gives a couple examples
- Most folks don’t understand what AI is, we should change that for the vast majority of Americans

Andy Konwinski
- New Paradigm where research has to be done differently
- Research that ships > h.index
- Academia has to go faster
- We can prove concepts with even just $5000 to millions even without giant sovereign wealth funds. And even engineers or just advice etc.

Building Things That Last: Lessons from Computing’s Long Arc

Dave Patterson
- RISC, RAID, Touring Award Winner
- 50th year in the field
- Berkeley cares about impact and that’s all that matters
- Citation count and h-index is horrible, bye bye Google Scholar
- Most people and projects don’t have impact
- They only count winners so 5 year projects are the method
- Fortune favors the bold
- A few years ago Multi-layer perceptions were all the rage at Google, now Transformers have taken over, so just know there will be a lot more to discover.

François Chollet
- Created Keras, center of Tensorflow. Deep Learning with Python Book
- Advice: Find what platforms you can build to have enable other people have impact
- What people are missing: Deep Learning worked because it’s the first thing that worked that can scan, but there’s a lot more problems that still need to be solved. Because deep learning worked we stopped other research, but we should still look at those and more.

Vint Cerf
- TCP/IP Founder
- Was friends with Dave in the 70s
- Competing with open spec systems and closed spec systems
- Advice for : Get help from people who are smarter than you

- Make risks when you're younger
- A lot of the successes from enabling technologies that others can build on
- Wikipedia is wrong, AT&T never wanted to buy the internet and they actually couldn't. Vint gave a story about the demo of Arpanet.
- No one fully understands what's going on in the transformers
- Make modular layers in AI
- The proliferation of agents might lead to the distributed nature of the early internet
- Precision for agent to agent interaction is important, not just English

John Ousterhout
- Built systems his whole life
- TCL/TK
- Market size: How many people will adopt and how hard is it going to adopt
- If you're working in a new area, it'll be hard to gauge if there's a market for it.
- Competition with more open source stuff
- Try and build communities around projects, he didn't do it but he's saying you should
- Speed and Granularity: There's not a lot of room on a tombstone, they're not going to etch 300 citations. You'll be remembered for 1 great thing, and sometimes 2-3 if you're really good
- Historical Fact: The first new tech isn't usually the long term winners
    - Brickland → Lotus 123 → Excel
    - Google wasn't the first search engine

Matei Zaharia
- Created Spark, Databricks etc.
- Faculty at MIT, Stanford, Berkeley
- Recognized by ACM in Computing Prize
- Understand your users and talk to them actively
- A lot of PRs that are AI-generated are ruining OSS projects
- Security is a great area to be in
- Custom agents for specific problems like biology, chemistry, etc.

From Open Research to World-Scale Infrastructure

Ali Ghodsi:
- Databricks Origin Story:
    - They had a ton of time at Berkeley to build their things
    - Mesos was thought to be the commercial product, but really it was Spark
    - Product-market fit discussion, Spark was the real fit for the company.

- Takes a long time for the world for awareness
- Until Nov 2022 no one cared about AI
    - Especially CSuite execs
- DBRX was the best open source model for 3 weeks. Cost 10 million, now top models take 10 billion or so to train.
- Ali thinks that the models and training them will be owned by the governments and ran as supercomputer labs/HPC like back in the day.
- Go from first principles like in academia and he believes it helped him start a company
- Stay off Twitter lol, they engagement bait all the time.
- Everyone is interested in GLM 5.2 and before that Kimi 2.6, so many companies are trying to scale back on the giant AI bill.
    - Databricks is helping folks with that
- AGI is already here
    - Right now the frontier labs are trying to solve Humanity's Last Exam. But we don't need to get that far for most tasks.
- There's a huge context gap. Building out that context gap will take 10-20 years.
- People replaced steam engines with electric engines, and same with typewriters and PCs. We just need to change society and how they use that technology.
- We need decades for society to adopt AI
- Current frontier models trends are getting bigger and bigger clusters and physics and more people.
- Open source will have a huge boom in the future
    - Omnigen for example, took 2 months
- AI Models even prefer open source projects because all the source code is public and they can play with it even with the documentation
- Might be a lot of programming language research because Agents are writing all the code and everything is a lot more robust. Like Rust for example.
- More software written in the next 12 months than all previous mankind.
    - Either way Python and the current languages will be dwarfed so either way might as well do more programming languages.

Thomas Wolf:

- Building community and awareness takes more time than you think
- Agreed a lot with what Ali said