Enterprise Neurosystem Community: Activity Report 1

Stanford SLAC / Harvard Analytics PoC: We received a readout on the AI model development track from Andrew Sima (Harvard Analytics) and Ryan Coffee (Sr. Scientist, SLAC National Accelerator Lab), and we were excited to observe the collaboration and progress.

Andrew is building a transformer model that can be used to replace and emulate the model in the Stanford LCLS Cookiebox detector. The prior CNN model used for inference on the particle accelerator was effective but a bit too compute intensive. His team has worked on new LSTM and transformer models to achieve similar accuracy while minimizing the hardware resource requirements. He's been guided by Ryan and Dennis Trujillo of Argonne National Lab, and was recently given access to the Stanford compute environment with a Graphcore processor for batch training. He now has one more model to complete before the end of the year. Training has initially been in the 5MB to 50MB range, and will be increased to 100GB (1,000,000 images).

As this new detector architecture acts as a small scale Neurosystem, it becomes an interesting first PoC / use case. Andrew indicated these benchmark models will be contributed for use by the community.

https://www6.slac.stanford.edu/news/2021-02-17-bigger-faster-more-powerful-slacs-new-x-ray-laser-data-system-will-process-million

NAIRR RFI Response: Ryan Coffee invited the community to submit an RFI response for the White House initiative to build a national AI research resource for labs and academia. Five community members wrote the response, with wider review by the extended community prior to submission.

The NAIRR indicated they will use this response and others to guide the development and implementation of this resource. Our thanks to Ryan and Ben Cushing (Federal Health and Science Lead, Red Hat) for their help in preparing the proposal.

https://www.whitehouse.gov/ostp/news-updates/2021/10/26/readout-of-the-third-national-artificial-intelligence-research-resource-nairr-task-force-meeting-and-announcement-of-posting-of-request-for-information-rfi-responses/

https://www.ai.gov/nairrtf/86-fr-39081-responses/

AI@DOE Submission: Ryan also invited us to submit a whitepaper in response to a four part US Department of Energy request on AI initiatives. The community is answering section 3, which is regarding the use of AI to assist with resiliency and recovery from natural and man-made disasters.

Ben Cushing worked with us over the last few weeks to finalize the two page draft, which was submitted on 12/13. If selected for presentation, Ryan will deliver the talk on the community's behalf.

Development Tracks: We've had the central/federated intelligence architecture track underway for a few months, led by Dinesh Verma (IBM Research). We are expanding in the new year with one more architecture development area, and two vertical focus groups for Telco and FSI. These groups have started with smaller gatherings, but we will update the call invitations to include the entire community.

Central Intelligence: Dinesh will prepare federated AI architecture concepts for review in January. We also had a superb overview during the session on 12/13 from Ravi Singha (Reliance Jio) on 5G/6G/ORAN in relation to AI and Edge use cases. In another recent session, Tong Zhang (Intel) presented an excellent overview of recent advances and contributions by Intel in AI R&D and networking.

Secure AI Connectivity Fabric: Led by Sanjay Aiyagari (Red Hat), this team will focus on AI model connectivity and security - enabling the secure transport of data for cross-correlation and storage. This track will result in a reference architecture and PoC.

MLconf Session: Ryan and I gave an overview of the Enterprise Neurosystem and the Cookiebox PoC at the MLconf San Francisco event. This is the primary US event for enterprise data science R&D, so it was a significant opportunity to raise awareness. The video was recently posted for viewing, and our thanks to Jennifer McMillan (Red Hat) and Mark Arthur (Red Hat) for their sponsorship and assistance.

https://www.youtube.com/watch?v=wSPVam1ZFQM

Membership Activity: We'd like to welcome aboard our most recent members from the mobile networks vertical, Ravi Singha (Reliance Jio) and Rashid Siddiqui (Verizon Wireless). We've also held recent talks with Morgan Stanley and SWIFT, and members of both organizations have expressed interest in joining the Neurosystem community. Our thanks to Kelly Switt (Red Hat), Ganesh Harinath (Fiducia AI), John Overton (Kove) and Narendra Narang (Red Hat) for these introductions and others.

University Outreach: We've held recent talks with UC Berkeley's Codebase organization (https://codebase.berkeley.edu/). They have asked us to submit a project outline in the first week of January for consideration. I have also contacted the head of the CS Department at St. Mary's College of California, and Ganesh Harinath (Fiducia AI) will also promote the community in his upcoming meeting with UC Davis.

Governance: Lisa Caywood (Red Hat) has been leading our community governance efforts. Following our discussions during the last two group calls, she will be sending out a vote to

approve the v2 modified governance structure, followed by a vote on various community positions for the coming year.

Logo / Website / Logistics: Lisa Caywood has secured a number of logo and web development resources for the community. We voted on a series of designs by Tuomas Kuosmanen (Red Hat), and we now have an official logo (image attached). We also held a meeting on website design and implementation on 12/13, with next steps to follow. The website will be a key part of communications, including updates on activity and video sessions. We will also post a calendar of meetings on our Github page.

Operate First: Karsten Wade (Red Hat) gave an Operate First overview session that was supported by a Q&A with Daniel Riek (Red Hat). They mentioned the resources and development framework related to this concept, and can arrange a development environment free of charge for non-HPC dev use cases.

https://www.operate-first.cloud/data-science/
https://www.theopensourceway.org/the_open_source_way-guidebook-2.0.html