

Exploratory meeting on enabling AI in HEP: Skills/Training, Capacity Building and Knowledge Exchange

Andrew Blake (Lancaster), **Dom Brailsford (Lancaster)**, **Tej Kanwar (Edinburgh)**, Eduardo Rodrigues (Liverpool), Tim Scanlon (UCL)

Overview

Focus on issues, barriers and opportunities for AI in the areas of:

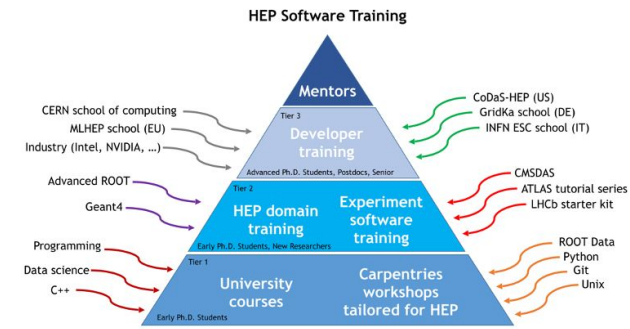
- Skills/Training
 - *Building skills for project members at all stages*
- Capacity Building
 - *Developing and maintaining **personnel**, processes, and resources*
- Knowledge Exchange
 - *Exchanging ideas, expertise and evidence*

Skills/Training: Overview

- Introduction
 - AI will become critical to every programme we undertake
 - Essential we have:
 - Well trained researchers in AI
 - Training and expertise on latest developments
 - Ability to identify and exploit latest developments
- A training programme that provides these skills is vital
 - Currently lacking a coherent and concrete approach to AI training in HEP across the UK

Skills/Training: Background

- Online courses and material
 - A wide array of online courses available
 - Can provide a good starting point, but can lack depth or necessary HEP focus
 - Little (so far) from the HSF on AI
- University courses and 1st year PhD courses
 - Wide range of high-quality courses, but generally linked to institutes
 - Can provide a good starting point, but can lack depth or necessary HEP focus
- STFC's DIS Summer School (also international HEP-ML summer schools)
 - Held at different university every year
 - Curriculum and focus changes every year
- STFC's Centres for Doctoral Training (CDT) in Data Intensive Science (DIS)
 - Front-loaded courses in AI augmented with projects undertaken with industry
 - Extremely successful in creating expert AI practitioners with HEP domain knowledge
 - Equally comfortable in AI and HEP research domains
 - Powerful conduit for bringing latest/greatest developments into our field



Skills/Training: Discussion

- Does anyone disagree that we need more and better AI training in HEP?
- Do we need everyone in the field to have a baseline training?
 - Augmented with more in-depth training for a fraction?
- What form could the training take and how could it be provided?
 - Available for PhD, PDRAs, academics, technical staff?
 - Expand CDT programme and make sure benefits are available across the field
 - In-depth courses available to PhD students at beginning of PhD?
 - Should we collaborate internationally and if so what is the best conduit (HSF)?
 - Hold in person training in software carpentry style to build national cohorts?
 - Make DIS summer school into something more solid?
 - Augment training with interactions with foundational/industry experts?
- How to decide upon the necessary content/focus/reach of the training?
- How do we make this happen in practice, at the least cost?

Capacity Building: Overview

- Capacity building: once we've trained world-leading AI practitioners....
 - Skillset that is in extremely high-demand in industry other research areas
 - A proportion will leave for other areas
 - Some leaving field good for wider socioeconomic impact
- How do we ensure enough AI experts stay in the field?
 - Career progression and job security
 - Training (interactions with industry/foundational-AI experts)
 - Ability to stay at the cutting-edge (software/hardware)
 - Ensuring interesting work environment
- Very similar issues surrounding Research Technical Professionals
 - PPTAP discussion (plus college of RSE White Paper on this):
 - https://indico.stfc.ac.uk/event/331/contributions/2199/attachments/676/1184/PPTAP_RTPs_200721.pdf

Capacity Building: Discussion

- Is there any strategic determination of what capacity is needed? If not, how do we build one?
 - Areas of expertise, levels of expertise and numbers?
- Do you need some dedicated AI experts, general uplifting of AI skills or a good mix of both? What is the right balance?
- How do we ensure we retain enough AI expert practitioners:
 - Career paths, opportunities and job security?
 - See RSE pools e.g. ARC at UCL? Are there similar AI expert pools yet?
 - Based at institutes/labs, centrally or a mix, what is the right balance?
 - Career progression at institutes becomes less research-focussed 'up the ladder'
 - Staying at the cutting-edge and providing compelling opportunities?
 - Create posts split 60/40 with industry or in a pool working across areas
 - Providing sufficient hardware/software tools
 - Are there any other factors important to expert retention?

Knowledge Exchange (KE): Background

- Several layers on which this can be considered:
 - Built into HEP - train fully fledged expert AI partitioners or hire AI-expert practitioners directly
 - Effective bridges between the AI and HEP communities (e.g. CDT)
 - Across HEP - Areas enriched in ML expertise, encourage KE:
 - PPTAP identified this as a key area to build upon
 - Foundational AI experts (e.g. Computer Science, Statistics....)
 - Many keen to engage providing a unique perspective
 - Better engage SciML and other expertise in Ada Lovelace Centre (see earlier talk)
 - Industry (see next discussion item)
 - Big possible gains from engaging with industry
 - Training, hardware access, tools/techniques, funding
 - Hartree and CDT programmes plus direct engagement already ongoing
 - Good examples to build upon and grow community-wide
- Effectively engaging KE can have a transformative effect upon our field
 - Can we improve our approach?

Knowledge Exchange (KE): Discussion Points

- Is there a preferred option for enhancing KE to focus on or should we see all four as equally important?
- Imbedding AI expertise in field requires overhaul of training:
 - Rather than training HEP domain experts, can we hire foundational AI experts to work in the field?
- How can we more effectively share ML expertise across the community?
 - AI forums
 - Interdisciplinary PhDs or co-supervision of PhDs (CDTs)
 - Develop common tools and training courses
 - Build stronger cohorts (CDT programmes, common training programmes, summer schools)
 - Mapping expertise across the community, matching need to expertise
 - Other ideas?
- Can we effectively and consistently engage with AI foundational experts?
 - Can be very hit and miss, requiring a lot of time investment
 - Determine when this is essential and if expertise exists within STFC
 - How to build stronger and more concrete links with keen foundational experts
- How do we make this happen in practice, at the least cost?

Backup

Skills/Training: Discussion

- Having access to proper hardware/software to train upon
 - Access to the sufficient hardware and GPU resources to train upon (local resources)
 - Website to inform everybody where resources are
- A training programme that provides these skills is vital
 - Should be a core part of the training we offer to all researchers PhD, PDRA, core staff and academics
- Engage foundational AI experts to help training in the field
 - Engage SciML
- Lack of instructors to teach some of these courses
- Need to develop courses for the community
 - We don't curate these as a community
 - Basic courses online, nothing of the depth and specificity that we need
 - Can do this internationally
 - Develop courses that can be held in person, build cohorts/links, train PhD students
 - Develop modules, build on top of them, save
 - Have different levels of courses
 - Engage people from other disciplines, get them engaged in HEP research
 - Engage IRIS-HEP fellows, Ada Lovelace Centre, SciML
 - Retroactive training for all in field

Knowledge Exchange (KE): Background

- KE beyond STFC
 - Good examples of working with statisticians
 - Statistics in neutrino physics
 - Lead to workshops very useful engagement
 - Statistics
 - Build a network of HEP friendly
- Well trained PhD students
 - Can obtain and deploy models
 - Roadmap of AI ongoing in other areas (maybe a survey)
 - Would enable people to reach out
- Explore common problems
 - Engage STFC SCIML team more in foundational AI problems
 - List of common problems, might be better to solve them internally and to reach out external
 -

Skills/Training: Background

- Skills/training
 - AI is becoming critical to every programme we undertake
 - Essential we have:
 - Well trained researchers in the latest
 - Produce robust, efficient and easily maintainable code/frameworks
- Training and expertise on latest developments
- Exploit latest techniques and architectures
- Train models in a robust manner (ML-Ops)
 - Robust,
- A software training programme that provides these skills is vital
 - Should be a core part of the training we offer to all researchers PhD, PDRA, core staff and academics