AI task force – 3nd meeting

Wouter Verkerke

Agenda

1. Review summary of AI use and future plans in the various groups
2. Next steps & agenda items for meetings

Current AI - production use

(1.1)	Object reco	Object ID	Calibration	Sig/Bkg discrimination	Anomaly/ Glitch detection	Stat. analysis/ param est
Alice	Tracking ITS & MFT (cellular automaton)	Flavor tagging with DNN	Jet energy with DNN	Simple models	Unsupervised, reinforcement learning	
ATLAS	Boosted objects with GNN	Flavor tagging with GNN	Mass regression of heavy objects	Wide range, including DNNs,	Variety of methods	LR of MVA score NSBI methods high-dim NN (SUSY)
LHCb	Small/fast track reco models	Simple classifiers	Simple classifiers			
GW					DeepExtractor	FlowMC, JIM, Jester (JAX), Flex. WV: NSBI used?
DM	NN event reco (lightweight to GNN)	NN classification (lightweight to GNN)				Neural posterior estimation
Pierre Auger	DL shower reco (RNN/GNN)	-	-	-	-	-
Neutrino	-	-	-	-	-	-
eEDM	-	-	-	-	-	-
Theory	Jet substructure	Jet tagging				SMEFT, WEFT, PDFs etc with (G)NN

Current AI – Nikhef in AI application R&D

	Nikhef involvement in current AI application R&D (2.1)
Alice	Combining tools (transformer + OneClass SVDD) – WV: incomplete answer
ATLAS	Track reconstruction, jet flavor tagging, boosted object mass regression, NSBI, foundation models, anomaly detection, physics language models
LHCb	ML ops in trigger, topics in IRIS-HEP, PINNs for track extrapolation, anomaly detection (with Maastricht)
GW	Parameter estimation and glitch characterization
DM	Event classification mostly WV: inferred answer, probably incomplete
Pierre Auger	- (end user only)
Neutrino	MSc projects on shower/particle ID
eEDM	-
Theory	SMEFT/WEFT/PDF fits, jet substructure

Current AI – Nikhef experts

	Staff (2.2)	PhD/PD (2.2)
Alice	A. Grelli (DNN on FPGA), M. Verweij (LSTM… transformers)	"Plenty"
ATLAS	S. Caron (broad), C. Nellist (selection) S. Rettie (in tracking) (W. Verkerke / L. Brenner – interface stats & ML), applications: F. Filthaut, T. du Pree, M. Vreeswijk	Z. Wolfs, P. Moskvitina, M. Silva, K. de Vries, O. Karkout, A. Sidley, W. Leinonen, L. Franco(PD)
LHCb	To some extent: R. Aaij, M. van Veghel, J. de Vries	X. Chiotopoulos
GW	A. Samajdar	M. Lopez (PD), P. Pang (PD), T. Dooney, L. Negri, T. Wouters, H. Narola
DM	J. Aalberts (DL/SBI). A. Colijn (AI-accelerated coding)	J. Mead (PD), M. v Nuland, M. Flierman, K. Weerman
Pierre Auger	-	T. Bister (PD)
Neutrino	R. Bruin (somewhat)	
eEDM	-	-
Theory	J. Rojo	

Current AI – BSc/Msc project | coding tools

	Master and Bachelor student projects with AI? (2.3)	Use of AI coding tools? (2.4)
Alice	About 70% of student projects have AI use, collaboration with UU math	Some members using it
ATLAS	Many (probably almost all), also have few CS AI MSc students in group	Limited (mostly Sascha)
LHCb	Very positive experience. Multiple CS AI students as well	Sparse use of co-pilot, various levels of ChatGPT
GW	So far about 10 MSc/Bsc theses with AI use	Co-pilot and ChatGPT used
DM	Some Msc projects with NL	Co-pilot and ChatGPT used by most
Pierre Auger	Sporadic use	Co-pilot and ChatGPT used
Neutrino	Three Msc projects	n/a
eEDM	-	WV: answer missing
Theory	A handful of BSc/MSc exploring ML applications	Co-pilot and ChatGpt use starting, want support and data protection

Near future AI –use, applications, resources

	Expected future AI use (3.1)	Novel AI applications (3.2)	Resource requirements (3.3)
Alice	Inference and explainability, generally much more broad use	Use in FPGA trigger, transformers in tracking, generative AI simulation	No clear, biggest reqs likely for AI tracking of PbPb collisions
ATLAS	Increased use in all areas, move to larger AI models (foundation models), strong increase of use in stat analysis NSBI & unfolding,	Already used almost everywhere ML on FPGA in TDAQ/trigger	Large GPUs(H100 etc) for tracking & foundation models, otherwise mostly modest use of inference models (L40S etc)
LHCb	Inclusive ftag with transformers, Generative models for simulation, more extensive use in online and trigger	Anomaly detection, reco of complex signatures, full event interpretation	GPU in trigger farms and datacenters
GW	Replace classical parameter estimation and glitch detection tools with AI-based tools,		GPUs recently acquired as part of grant (L40S), access to Snellius
DM	Possibly training on 'all data' (but gains hard to predict now)		GPUs for student model training, large models may require investments?
Pierre Auger	Improved classification & reco with GNNs and LLMs (gains unclear now)	Possibly in radio trigger and arrival detection interpretation	unclear
Neutrino	Improve classification & reco (in particular hadronic FS in Orca),	determine composition of CRs, use mixed LL/AI models (WV: this is really NSBI)	Not yet known
eEDM	AI(NN) for experimental design optimization	stabilizing feedback using FPGA	-
Theory	Simulation based inference (but generally unsure)		-

Near future AI – Nikhef involvement in R&D

	Ambitions (4.1)	Leading institutes intl (4.2)	Comparison with Nikhef (4.3)
Alice	Be part of new developments in analysis and tracking	Scattered over many institutes worldwide	We lack computer scientists, could profit more from collab with math & AI. Lack of PhDs. So far compute cluster was sufficient
ATLAS	Sascha: AL/ML leader in HEP worldwide. ATLAS group: strong AI/ML involvement in all areas in diHiggs (selection, calibration, regression, inference). Tracking: strong involvement in HL-LHC AI tracking (Sebastien, Sascha), R&D on statistical properties of ML methods (Wouter & Lydia)	Paris (Rousseau), Germany (Heinrich, Plehn), Geneva (Golling), US (Cranmer, Whiteson, Nachman, Brehmer). With exception of Cranmer, mostly 1-person efforts	Staff size & expertise comparable to world leaders, but much more limited in PhD volume and in computing resources
LHCb	Jet reconstruction / substructure, isolation at high PU, ftag with transformer models, AI tracking, online use of AI/ML	MIT (multiple areas), Coruna (ftag), soon Spain with supercomputers and funds	A bit behind, but expecting to catch up. Less dedicated focus on AI, but excellent infrastructure. Leading online
GW	Replace classical parameter estimation and glitch detection tools with AI-based tools (WV – replicated from 3.1)	Cambridge, Glasgow, Johns Hopkins	We are among leading institutes
DM	Possibly training on 'all data' (but gains hard to predict now) (WV – replicated from 3.1)	-	-
Pierre Auger	Still in exploratory phase. Likely more involvement in data analysis and reconstruction	KIT, Aachen, Erlangen	Nikhef has little expertise and person power. Not enough info on infrastructure.
Neutrino	Should get into business, but first need to educate selves	Ecap, Erlangen	n/a
eEDM	Optimize experimental design and use feedback loops with AI/ML (WV - replicated from 3.1)	-	-
Theory	-	-	-

Near future AI – Nikhef involvement in R&D

	National and int'l partners (4.4)	What is missing to realize ambition (4.5)	Expectations coding assistance(4.6)
Alice	Changes per project	PhDs & PDs, computer scientist	Importance will grow in time
ATLAS	Sascha: Uva(C. Weniger), RU(AI dept), SURF, eScience Center, Lydia: groups in Heidelberg, Munich, US. Ftag/reco: largely through ATLAS collaboration	GPU infrastructure, broader/visible AI/ML group at Nikhef, in-house AI expertise. Own inference machine at Nikhef / participation in European effort	In-house coding/knowledge tools that are safe for use.
LHCb	SURF, DACS (Maastricht) ; Coruna, LHCb collaboration	Person power to the work (PhD+PD), ML + statistics expertise with focus on math	Experiment-specific and/or private ChatGPT (concern about review code of AI generated code)
GW	UC Louvain, Potsdam	In-house AI experts	Importance will grow in time. (Our in-house expertise is not sufficient to match ambitions.)
DM	Rice U (C. Tunnel) & UCSD (A. Li)	(WV - answer missing)	Research and coding assistant (LLM) trained on private data (pilots in Xenon)
Pierre Auger	KIT and Aachen	Education, training, documentation, acces to GPUs	Requires institutional support for optimal use
Neutrino	n/a	Self-education	No formal plans, but expect that importance increases
eEDM	-	Expertise & expertise and information, opportunities for joint projects	-
Theory	-	Support in form of joint PHD to explore options	-

Long-term future – AI potential

	'Disruptively big' ideas for future AI applications (5.1)
Alice	Lead-Lead track reconstruction (WV ??), detector simulation and design, extract physics from the tails
ATLAS	Large Physics models as natural end point of LLMs and foundation models. AI-assisted optimal detector/sensor design
LHCb	Pat. rec in high pileup, Lattice QCD, generic model-indendent anomaly detection, optimal detector design, global reoptimization of software
GW	Tools for the Einstein telescope – current methods don't work (partly funded in awarded XL)
DM	Redesign of analysis and computing pipelines of next-gen experiments (XLZD etc)
Pierre Auger	Full 3D reco of galactic magnetic field. AI-assisted optimal layout of observatory. Unthinned simulations (w/o energy cutoff)
Neutrino	Signal detection in acoustic neutrino detectors
eEDM	-
Theory	Lattice QCD

AI – Other comments

	Other comments (6.1)
Alice	Alice has a stats & ML forum where we try to steer the various efforts
ATLAS	Happy to help (Sascha) / Need to invest in both hardware and person power
LHCb	Structured training of people beyond superficial use
GW	
DM	Can we have an AI/ML bootcamp – for new people to get started
Pierre Auger	-
Neutrino	Basic knowledge and education is missing, propose education of Nikhef staff
eEDM	
Theory	

Some first observations – AI use & user base

- **Current AI use in production**
  - **Clear 'front-runner' group: ATLAS+Alice+LHCb+GW+DM** where AI is pervasively in use
  - Second group Auger+eEDM+Neutrino+Theory*, with more limited AI use at the moment (but all with interest to catch up)
  - Most AI use currently in object-level reconstruction/identification and event-level discrimination tasks.
  - Limited use of AI in inference (limited use in GW ATLAS at present) and in simulated data generation

- **Level of AI involvement in Nikhef programs divides along same groups**
  - Strong multi-front involvement in front runner group
  - Limited (end-user) use in second group

- **We have a quite sizeable group of scientists involved across programs!**
  - About **15 staff** – with expertise ranging from top-tier (Sascha) to 'limited, but with strong interest'
  - Another **20 PhD/PD** closely involved with AI application, plus a few more on AI method development

- **AI is commonly used in MSc/BSc projects in 'front runner group' (50-70%?)**
  - **Generally (very) positive experiences for those participating**
  - Also embedding of CS/AI students in several programs

- **Use of coding tools (co-pilot, ChatGPT) still quite limited**
  - A handful of users in each program (including groups that otherwise don't use much AI)

Some first observations – AI in near future

- **Expected future use to increase all experiments**
  - Overall use expected in become more pervasive.
  - General move to **larger models** with more **ambitious goals** (transformer models directly using all data) in experiments already deploying AI (→ foundation models)
  - Use of AI for **highly complex data** (HL-LHC tracking, ET GW signals)
  - Production deployment of AI methods in statistical **inference**/parameter estimation (NBSI), and fast event **simulation** (GANs etc), **anomaly detection**
  - Use of AI in low-latency online environment (trigger) → **FPGA** implementation (Alice, ATLAS, eEDM)
  - Interest in AI-assisted experimental **design optimization**

- **Required infrastructure/resources**
  - Most ambitious R&D projects (ATLAS/Alice tracking) require state-of-the art GPUS (nVidia H100/H200), available in Snellius, not locally. LHCb needs GPU in trigger/online farms
  - GW purchased already nVidia L40S for medium-term needs
  - All other Nikhef applications likely (much) less demanding, but not always clear

Some first observations – AI ambitions

- **AI ambitions of Nikhef groups**
  - 'front runner' experiments: to have leading involvement AI application development in many areas (tracking, object identification, regression of object properties, statistical inference, anomaly detection, physics-inspired models)
  - 'second group' experiments: deployment in reconstruction and data analysis

- **Leading institutes in AI in HEP world-wide**
  - Many relatively small groups (1 staff + PhD/PD) at universities in US and Europe.
  - Little overlap between leading institutes per experiment → few seem to on multiple experiments?
  - Main exception – UW Data Science center / Kyle Cranmer – 20 FTE. (KC is also former supervisor many of the current strong senior scientists with groups in ATLAS)

- **Comparison of leading institutes with Nikhef**
  - Often O(1) staff – thus comparable in size and expertise with Nikhef groups (with a bit of spread)
  - Several groups have better connection to non-HEP AI/CS/Math experts
  - Often have more sizeable PhD/PD groups working on AI projects
  - In certain countries access to vastly larger computing resources (notably US, Spain in future?)

Some first observations – AI ambitions

- **Inter(national) partners in AI development**
  - national entities: SURF, eScience center
  - non-HEP groups at NL universities: Grappa@Uva, AI@Radboud, DACS@Maastricht
  - *But mostly through personal contacts of a few individuals or locations (Sascha, Maastricht)*
  - International universities: handful of universities in Europe and US, O(1-2) per group

- **What are we missing to realize our ambitions?**
  - Education/training/expertise (mostly in the '2nd group')
  - Close connections to domain experts in CS/AI, math/statistics
  - Sufficient in-house GPU compute power, access to large-scale NL and EU AI centers
  - PhD/PD person power

- **Expectations on future of coding assistance**
  - Expected to generally grow and importance and use
  - Emphasis/concerns on secure (in-house) services for use on private data and/or information
  - Institutional support

Some first observations – AI long-term ambitions

- **Long-term goals that may be realizeable with AI**
  - Large-scale data reconstruction from low-level data (HI PbPb tracking)
  - AI-assisted (re)design of software structure and analysis pipelines ('end-to-end optimization')
  - AI-assisted design – from sensors to entire detectors / observatories
  - Large Physics Models as evolution of foundation models (trained on low-level data, allows for high-level physics queries)
  - AI Lattice QCD

Many areas of cross-group interest

- **AI classification strategies**
  - Pretty much every group
- **AI track/pattern reconstruction**
  - ATLAS+LHCb+Alice (+Neutrino?)
- **AI anomaly detection**
  - Known background, unknown signal
  - GW, ATLAS, Neutrino, PierreAuger, DM
- **AI physics-inspired model architectures**
  - LHCb, ATLAS, + ??
- **AI statistical inference / param estimation (NSBI etc) / stat properties of ML methods**
  - GW, ATLAS, LHCb, Neutrino, Theory + ??
  - Mostly powerful for physics problems with >>1 parameter of interest
- **AI deployment in FPGAs for trigger, or feedback loops**
  - ATLAS, Alice, eEDM
- **AI assisted optimization of data analysis pipelines and software ('end-to-end optimization')**
  - ATLAS, LHCb, DM + ??
- **AI foundation models / full event interpretation**
  - ATLAS, LHCb, Alice, + ??
- **AI assisted detector/observatory optimization**
  - ATLAS, PierreAuger, eEDM, Alice

Next steps

- **Driving the agenda in the short term is the submission of a grant proposal to the NWO-I 'Strategische Vernieuwingsfonds' by July 8th**
  - Along the lines of abstract submitted earlier: focus on expert person power (staff + 2 PD) + limited about of state-of-the art HW
  - About 4 pages + figures. To be written by VV (with your input), to be submitted by JDH.
  - Durable result (beyond funding period) in align with Nikhef strategic goals
  - As 'kick starter' to setup a strong cross-program AI organization within Nikhef.

- **Important open questions to answer next (in this order)**

1. What **goals/activities** should this new 'cross-program organization' deliver?
2. How to best **structure/organize** this?

Future AI organization – goals & activities

- **What goals/activities should this new 'cross-program organization' deliver?**
  - Organize AI education/training programs for PhD/PD as well as staff
    (lectures series?, schools? interactive-handson?). Education on coding assistance?
  - Involvement in setting up local LLM knowledge services (ChatGPT/DeepSeek)?
  - Organize access to AI infrastructure facilities (at Nikhef and outside [Snellius etc], at EU level?)?
  - Maintain in-house expertise on AI techniques that are of crucial importance for group in the Nikhef portfolio (which ones)?
  - Do R&D on development of AI techniques on HEP applications of importance to Nikhef (which ones)?
  - Community building role (meetings, workshops, etc)?
  - Become central point of contact for AI/CS and AI/Math departments for universities all over NL?
  - Become attractive partner for participation in large-scale AI funding proposals (at national or international level) [nlaic.com]
  - Role in (inter)national strategy discussion on roadmaps for AI?

Next meetings

- **4th Task force meeting – Monday June 2nd 10-11**
  - Discuss goals and proposed structure of 'cross-program AI organization'

- **5th Task force meeting – Thu June 19th 14-15.30**
  - Discuss outline of the proposal text for the 'Strategische Vernieuwings fonds'

- Not sure if we need another meeting before the proposal submission
  - Will circulate draft for comments well in time

- **After summer break – more meetings (to be scheduled)**
  - Inventorize AI landscape at CS/Math departments at Dutch universities, existing connections with Nikhef, opportunities for new connections. [Theory connections]
  - Finalize task force report before October (decision SV fund)