Chatgpt prompt: Please summarize these meeting notes from the NSTX-U Monday physics meeting on December 2, 2024 where we held a brainstorming session on AI-ML for NSTX-U:

First, Jack Berkery presented: What AI-ML has been done for NSTX/-U already?

1. Profile prediction
   1. SXR-ML for Te profile (Clayton, 2013)
   2. MSE-ML for reversed magnetic shear (Uzun-Kaymak, 2024)
   3. NN for ne and pe profiles (Boyer, 2021)
2. Database / experimental measurements
   1. Alfvenic chirping (Woods, 2020)
   2. ELM patterns (BES) (Smith, 2016)
   3. ST pedestal prediction (Parisi, in progress)
3. Transport
   1. MTM reduced model (Curie, unfinished)
   2. NN for electron thermal transport (Chung, in progress)
   3. NN surrogate model of Multi-Mode-Model (Leard, 2024) (this also belongs in the next section)
4. Code surrogate model / speed-up
   1. Heat flux with HEAT for NSTX-U (Looby, 2020)
   2. ICRF power absorption with TORIC (Sanchez-Villar, 2024)
5. Operations / equilibrium / real-time
   1. ML scenario planning (Boyer, 2020)
   2. NN for EFIT for NSTX-U (Wai, 2022)
   3. NUBEAM-Net (Boyer, 2019) (\*also code surrogate/speedup)
6. Stability / Disruption
   1. Ideal MHD (Piccione, 2020)
   2. Resistive wall mode (Piccione, 2022)
   3. Vertical stability (Tobin, 2024)

What should we do for AI-ML for NSTX-U, going forward?

In the past there were Many individual efforts, not coordinated, many not sustained. Collaborators, students, people have left PPPL, etc…

Revive some previous efforts?

There are, of course, other tools developed on other machines that could be brought to NSTX-U, like Ege’s group’s tools on DIII-D, etc…

An example of this is RTCAKENN, an AI-driven algorithm for real-time plasma profile reconstruction, that has been implemented by Ricardo Shousha on the DIII-D PCS and could be brought to NSTX-U when it operates.

Ideas:

1. NSTX database curated and made public?

Who’s going to do it?

Some databases already exist: Ege’s, Steve Sabbagh’s DECAF…

IAEA “AI for Fusion” effort

MAST database made public (<https://opendata.ukaea.uk/mast-data/>). Later follow-up by Matt Parsons indicates that this isn’t all it’s cracked up to be. Seems to be more of a list of available data, then he requested something and someone had to find it and email it to him.

A set of EFIT equilibrium reconstructions in the database?

Michael Churchill - MDSplus is not a good data format for AI

Sterling Smith (GA) (later, email follow-up) = One difficulty we have faced at DIII-D is when users try to use MDSplus carelessly to gather lots of data for a ML database.  That can crash our MDSplus server.  Is that a problem you have faced?  We implemented a new toksearch copy of the data stored on BGFS with associated software for rapid access to many shots' data in MDSplus (and PTDATA - how we store our raw data). We have also started building an LLM - fusion sage - with various capabilities.  At the Supercomputing conference a few weeks ago I found out that separately at ALCF, they are building Aurora GPT, which effort we in fusion should latch onto.

Dave Smith – should consider making a curated, vetted public database and making a Kaggle competition to predict something specific with it. This would increase public visibility of NSTX-U as well.

Adam McLean -   
General Fusion has done a few 'crowdsourcing' open source competitions making their data available, going back to 2015 (<https://generalfusion.com/post/crowdsourcing-fusion-data-driven-prediction-of-plasma-performance/>). GF offered a $20k first prize to the best algorithm to predict performance. Notably, GF said "We’re not looking for participants who are plasma physicists." - i.e., those people could be particularly unknowingly impacted by the existence of bad data in the database. It may be interesting to reach out to them to see what came of it - was it helpful/useful

Ricardo Shusha – Use a database generator instead of creating one big database?

Matt Parsons – Should look at what DIII-D is doing.  
  
Hiro Josep Farre Kaga   - I agree with Ricardo. Our Egemen group has a database generating script for DIII-D, which looks at and processes mds+ and other DIII-D databases given what we’re looking for.  
  
Jalal Butt  - also agreed with Ricardo, Steve Sabbagh’s group also converged to a similar approach

Nate Ferraro said that if you want to project to future FPP devices, then you aren’t going to use experimental data, but rather simulations in a database.

Theory department wants to start a project on ST turbulence and FPP design

Fatima Ebrahimi said that on her scidac for edge stability they are developing a hybrid database with experiments and simulations.

Joe Schmidt pointed out that they are already compiling a multi device database of gyrokinetic simulations.

1. Engineering: digital twin?

Nate pointed out that usually people think of digital twin as all encompassing, but we could start with something smaller, concrete.

Laura said there is a Widow of opportunity. Difference between CAD + real. Matt - measurements aren't made unless requested. Laura - now is time to have a list

1. Diagnostics: data reduction and analysis programs make use of AI-ML strategies?

Brent Stratton was not in attendance, but will start to think about this topic.

Mate Lampert (email afterwards) - One idea I had about today’s meeting is that we could use AI/ML for diagnostic fault detection. It occurs quite often that there is a problem with the diagnostic but we only realize days or weeks later that it occurred and we lose a lot of good data. If a tool could be developed to detect faults in diagnostics, that would help a lot with the experiments.

Dave Smith - Faulty sensor AI - AI anomaly detector that flags off-normal sensors/signals.

Dave Smith – diagnostics, real-time: high throughput engines can analyze data streams, then be available for real-time control. This aligns with the high energy physics area, they use these engines. Could align with the CHIPS act? New processors, for basic science?

1. We should consider what is most aligned with the new NSTX-U five year plan

Dave: fast ion physics and energetic particle modes is an NSTX-U specialty. We could do real-time FIDA measurements and simulations, and connect to the plasma control system and actuators.

Phil Bonofiglio: Bill Heidbrink has started some of this. Mode identification. We’d need a labeled dataset for NSTX-U and then could identify modes in real time. His Irvine group has recent APS presentations on this: FIDA giving fast ion distributions.

1. AI for plasma operations?

Other ideas: use AI to do other cognitive actions.

Dave Smith - A machine learning auto logbook would be something new. Could interface with reasoning agents. Between-shot AI parsing of previous discharge evolution to tag events and phases.

Michael Churchill passed along a draft of a paper he wrote “AI foundation models for experimental fusion tasks”, which gives a perspective on AI foundation models and how they can be used on fusion experiments. This is the type of work where you would want to create a database more easily useable for AI (i.e. not in MDSplus), and how this can more easily enable automated logbooks, a stepping stone for the type of AI reasoning agents driving scenario planning, etc.

Dave Smith - One additional layer may be to investigate how a foundation model trained on an existing device (NSTX) adjusts to a new device (NSTX-U). This could be analogous to transfer learning from existing devices to ITER.

Determine why a plasma fizzled. Plan your way ahead for the next discharge.

Use AI for scheduling of experiments (run coordination)

Experimental planning: GSevolve tool. Dan Boyer’s idea was to do this at lower fidelity but faster.

This all fits in with enhancing the user experience – if NSTX-U is going to be operating as a “user facility”.

TAE has done something like this, called the optometrist algorithm, we should check with them if it was useful.

Jason Parisi suggested eventually maybe AI could run operations at NSTX-U  
  
Dom Power :   
There is a growing field of experimental design which looks at optimising experiments in order to quantitatively maximise information gained about some model, see e.g. https://projecteuclid.org/journals/statistical-science/volume-39/issue-1/Modern-Bayesian-Experimental-Design/10.1214/23-STS915.short  
so perhaps Jason’s idea is not so radical 🙂

Laura said we should Document trials + tribulations on standing up these efforts. Give failures and background info. Would build community trust

1. Resources:

https://drive.google.com/file/d/1jPcVdxMtGxrRKvCzYIRPsS-D9Rn\_R6Qy/view