

2022-01-13 - Anuket: Thoth Shared Data Model for Intelligent Networking

Topic Leader(s)

- [Beth Cohen](#)
- [Sridhar Rao](#)

Topic Description

[Beth Cohen Sridhar Rao](#)60m

Discussion of how to create a common data model to support AI/ML for Intelligent Networking

Topic Overview

Data standardization, shared data sets and models have been long-term challenges for the adoption of intelligent networking. A shared understanding of the data models themselves is a basic requirement. For example how jitter is defined can vary wildly across operators or even within a single operator. Even something as seemingly simple as a basic AI algorithm framework is a much need capability to advance the industry. Vendors and operators need to develop common AI models for data, through a mechanism for model and data sharing. An AI/ML and model sharing project would be a good way to promote industry collaboration, promote the sharing of data and models through the joint construction of intelligent networking scenarios.

For more information about what LFN is doing in about Intelligent Networking:

White Paper: Intelligent Networking, AI and Machine Learning: <https://www.lfnetworking.org/publications/2021/11/03/white-paper-intelligent-networking-ai-and-machine-learning/>

Webinar on the same topic

The purpose of this session is to explore some options for creating the shared models.

<https://aiforgood.itu.int/about/aiml-in-5g-challenge/>

Wiki Page to capture the collaboration: [Collaboration - ITU](#)

1. Take up relevant problems (NFV) and start creating models - 2 Problems.
2. Initiate dialogue with TSC-Anuket, LFN and **EUAG** to be one of the Hosts for next round of challenge.
 - a. Tentative dates: Kickstart (February).
 - b. EUAG: Can consider Opendataset and define a novel problem.

Slides & Recording

- Live Interactive Session



Agenda

- Explore some options for creating the shared models.
- Should we leverage the <https://aiforgood.itu.int/about/aiml-in-5g-challenge/> to get there faster
- Are there other open data sets beside the one that Orange has shared?

Minutes

Review on where we are in the development. There are two aspects to intelligent networking. For the Thoth project we are focused on the data and network traffic, not the operational aspects of the issue.

[Vishnu Ram](#) excellent initiative. It is much needed and can push the boundaries. question - are you looking so data model for specific problem? e.g. root cause analysis, fault isolation, or prediction. in other words, in your mind, would the data model be "per usecase"??

[Walter Kozlowski](#) Just to make sure: from ML perspective we are talking here about training data, is this the case? Yes this is for training data. It needs to be anonymized in a way that still returns valid analysis and useful information. Might we look at companies that possibly have a data pool such as Newrelic or another company that is working in data analysis and data collection?

EUAG and Thoth can work together. Big data researchers need to have domain knowledge in WAN Networking and visa versa and that is not the case, so the two groups need to work together to reach our goals. Need to validate that the algorithms are working.

Need a robust set of use cases – i.e. what problems are we trying to solve with AI? One of the goals is to create a model as a service to share the data with all the community – operators and vendors alike. Looking at Cubeflow right now, but others are possible. [Ranny Haiby](#) has a question about the Thoth scope. Need RAN data for this area of research. The list of data sets available: <https://wiki.anuket.io/display/HOME/Collaboration++ITU> today.

[Lei Huang](#) are we focused on research or production. We are starting with research, but the intention is to make it useful for real world production deployments. Question about log generation and analysis of the logs. We need to use real logs and also synthetic generated logs. Are working with some OPenstack projects to get some log data to test things.

Action Items

