SPARK+AI SUMMIT EUROPE

AI that cares about your broadband connection

Spark+AI Summit Europe – Oct. 2018

Nicolas DUPUIS

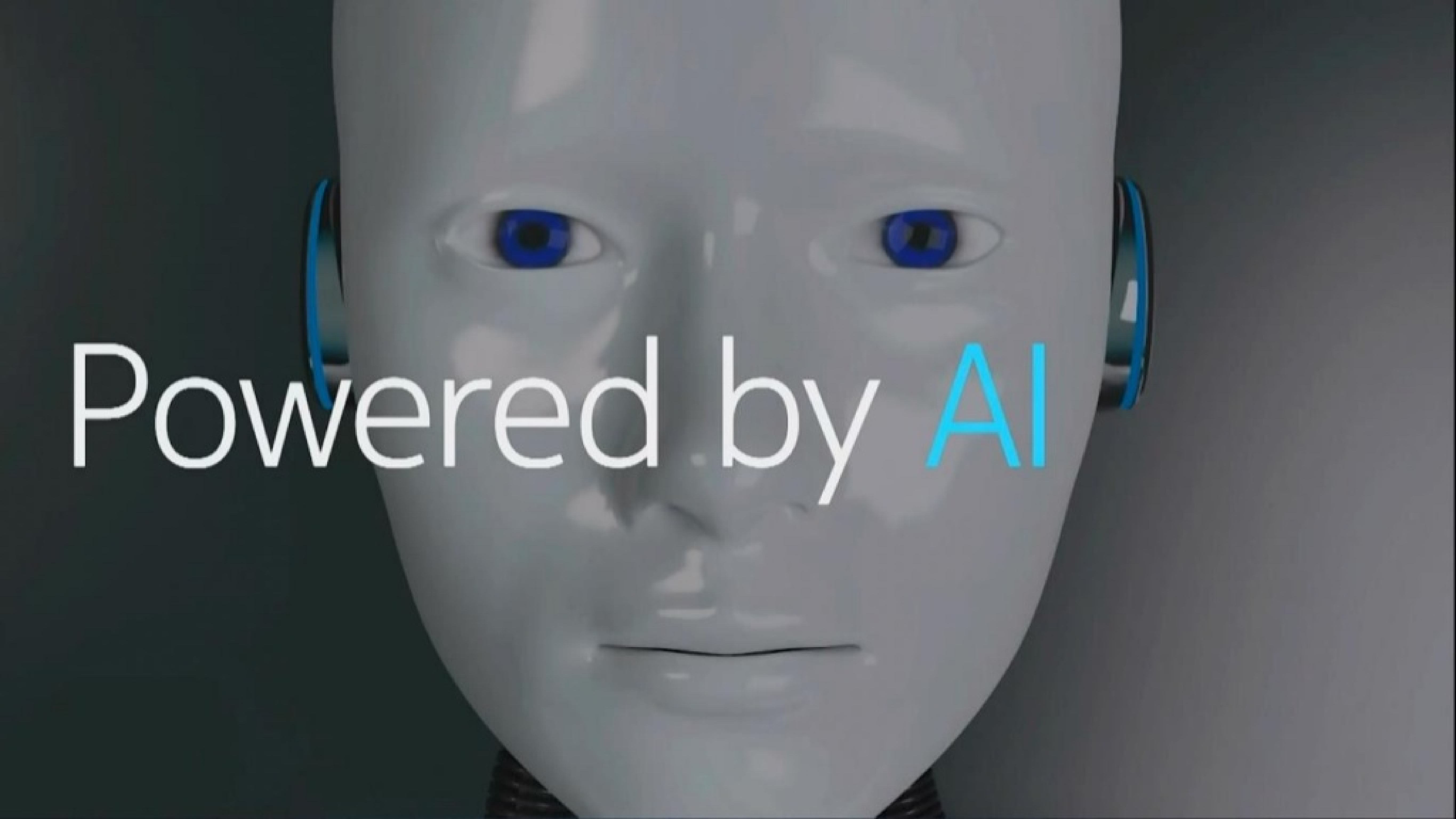
Head of Data Science, DMTS
Digital Experience – Analytics
Nokia Software

#SAISEnt6

NOKIA

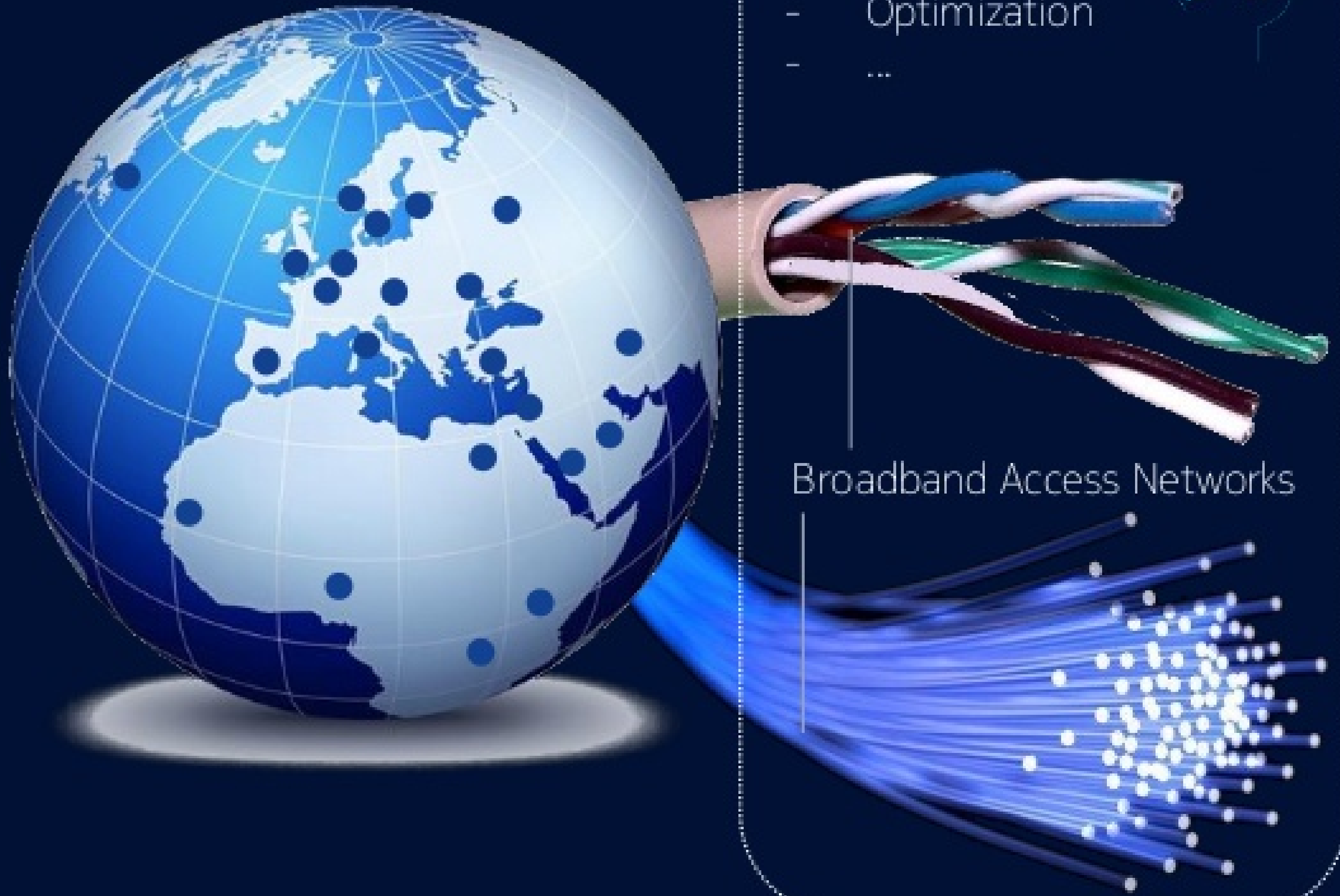
Nokia

2000


Powered by AI

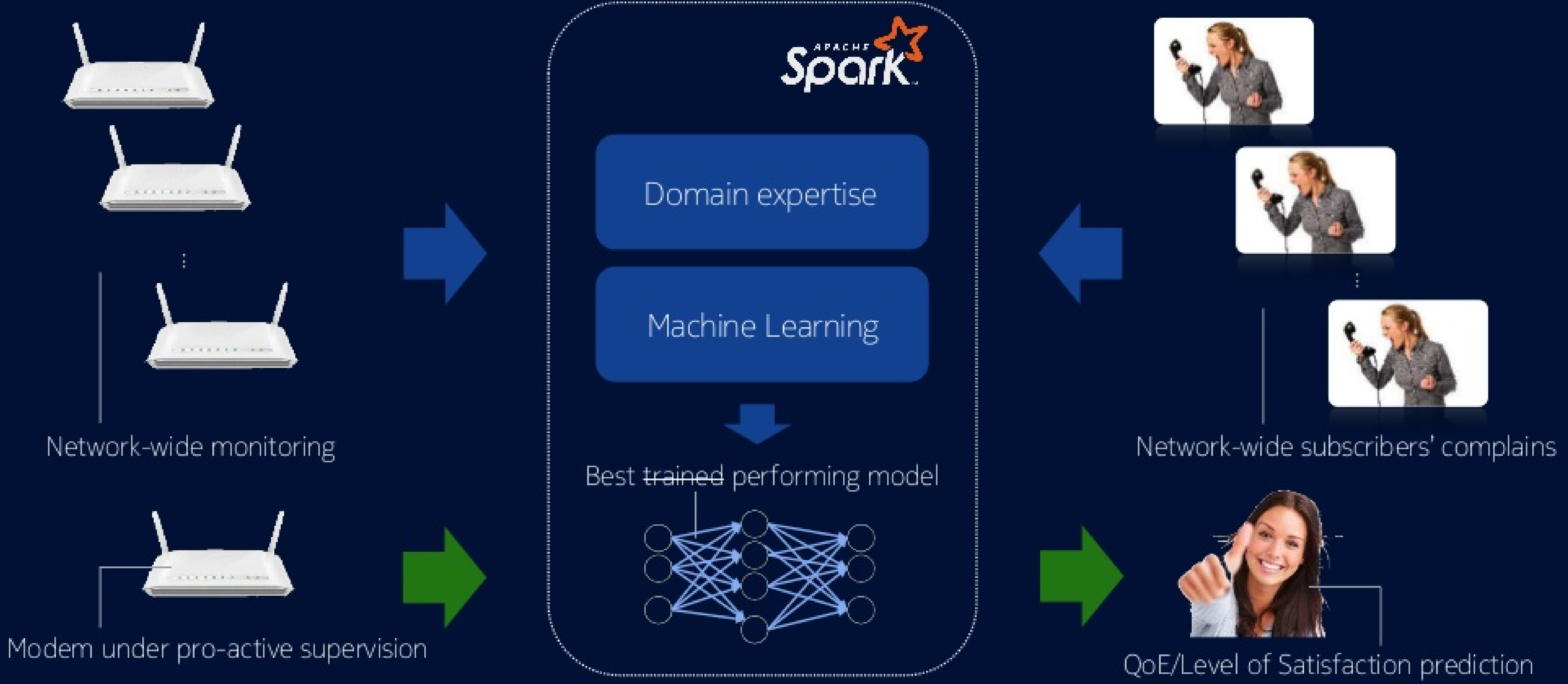
Care Analytics

- Prediction
- Diagnosis
- Optimization
- ...



Digital Experience

Predicting the Quality of Experience



Predicting the Quality of Experience

Domain expertise

|Scope/requirements
- What is needed/feasible
- What input data needed
- How to obtain unbiased data
- How to measure performances
- ML vs. DL?
- ...

- Selection of relevant input data
- Producing KPI's combining various input data
- Space transformation

|Data transformation
|KPI's computation

|Feature extraction
- Scaling
- Time-series FE
- Ranking
- Balancing

|Performance validation
- Real-world validation
- Interpretation
- Adjustements

|Pattern recognition

Because of:
- Not applicable targets
- Mislabelling
- Lack of information

a resampling step, based on pattern recognition, has been added to the process.

|Model selection
|Modelling
- Cross-validation
- Classificator training

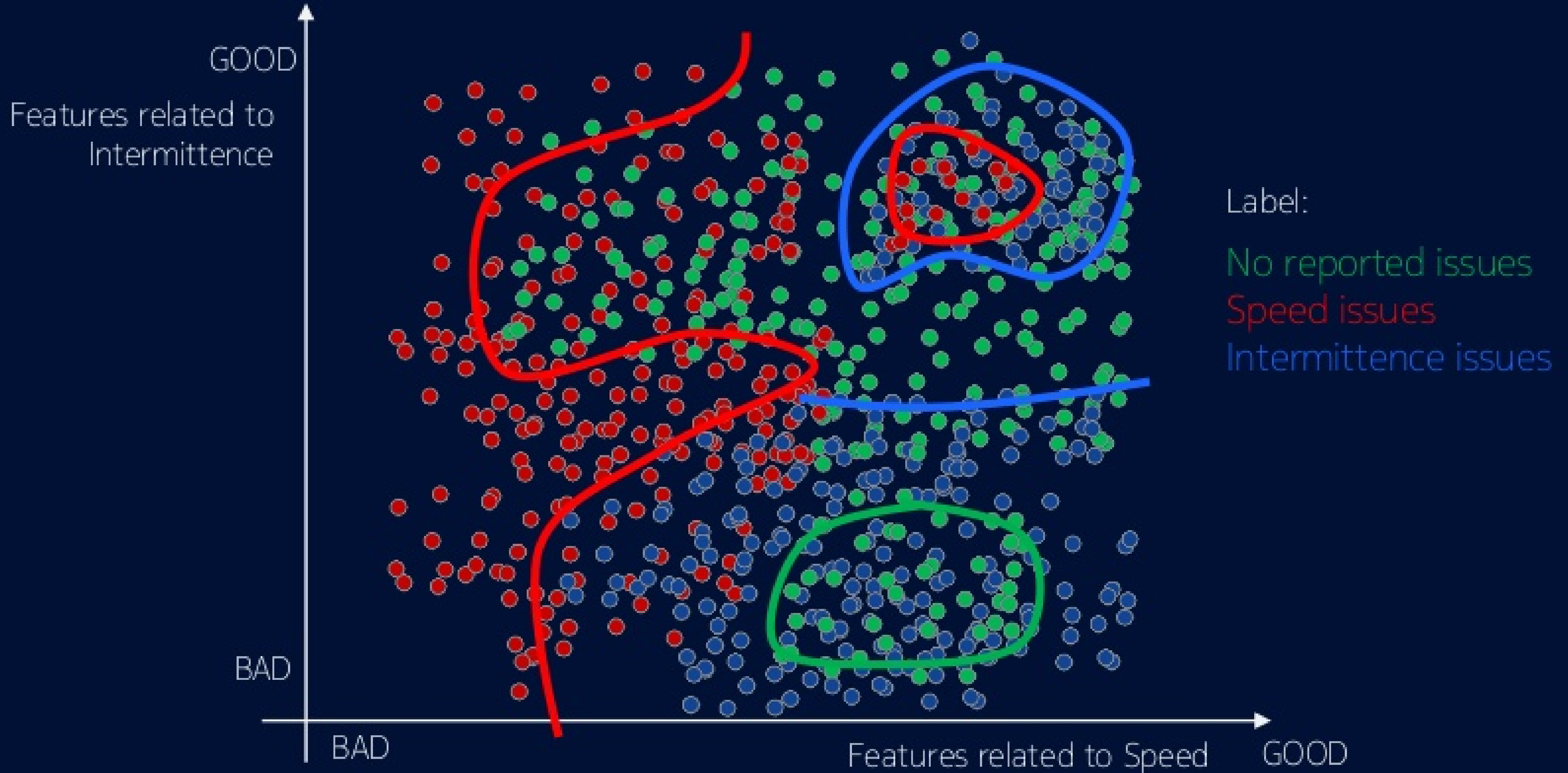
2

4

Machine Learning

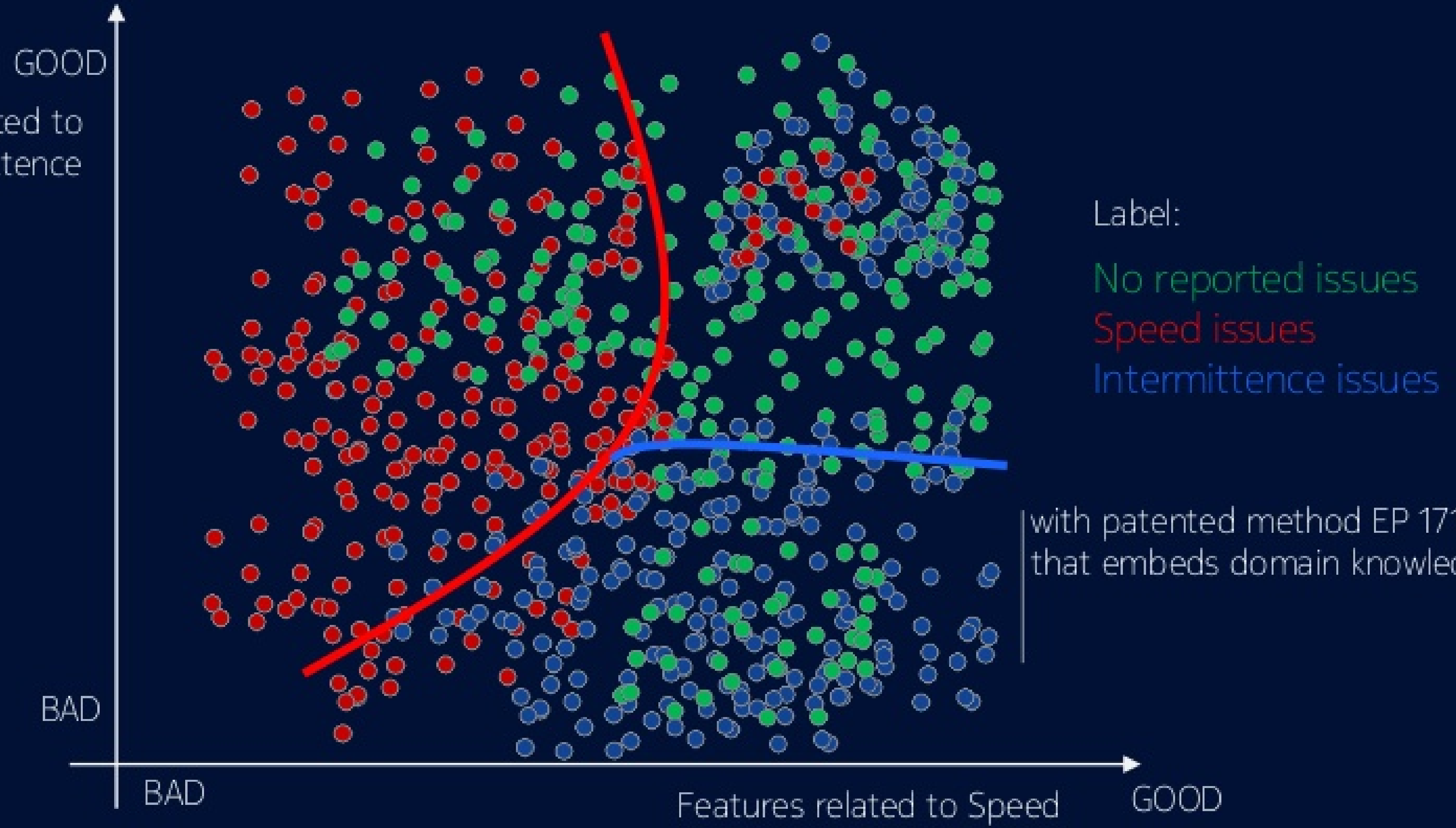
Predicting the Quality of Experience

Pattern recognition



Predicting the Quality of Experience

Pattern recognition



with patented method EP 17189765.5 that embeds domain knowledge.

Predicting the Quality of Experience

- ETL
- Windowing/Pre-processing over >100M data rows
- Spark MLlib
- Efficient/distributed learning
- Execution

- Complete Spark MLlib API
- Java world
- Ability to produce JAR files

- Executed as Spark jobs
- Compiled code

Predicting the Quality of Experience

	Empirical algorithms	Machine Learning		Machine Learning + domain knowledge
Prediction LIFT (gain)	10x	80x		80x
Field performances	☆	☆	➡	☆☆☆

Predictions much valuable for the business.

Broadband connection diagnosis with DL

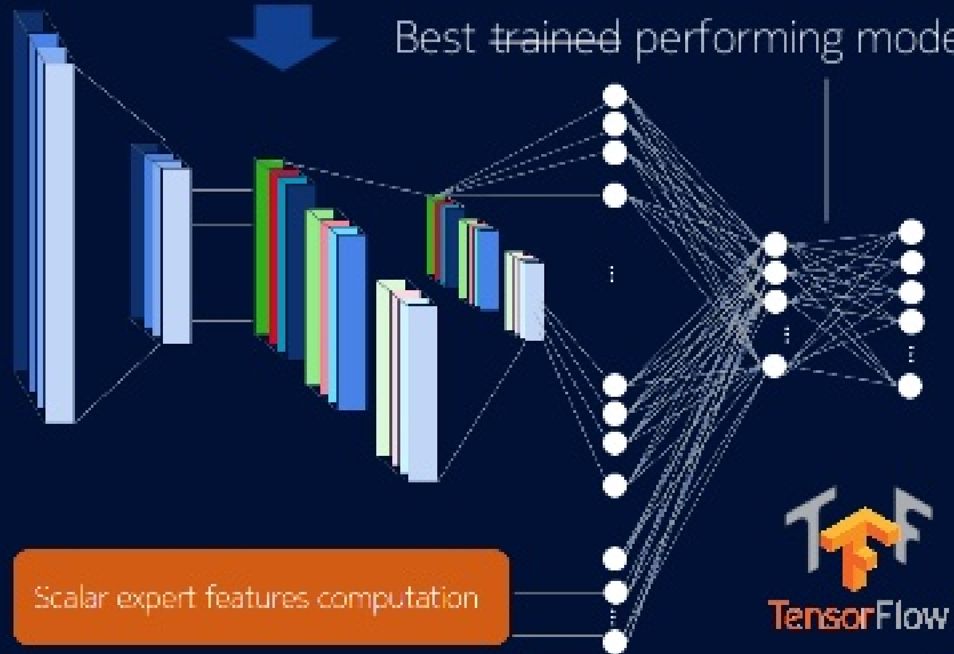
Domain expertise

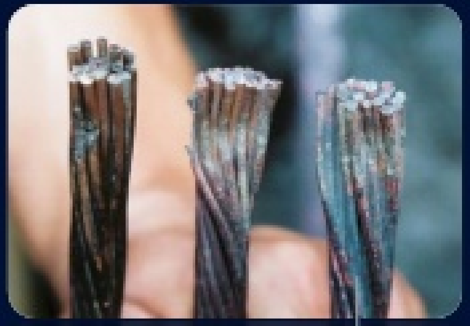
Data Science

Training set creation
Soft-target labeling
Pre-processing
Scalar expert features
CNN design
Validation

Best ~~trained~~ performing model

Modem under pro-active supervision

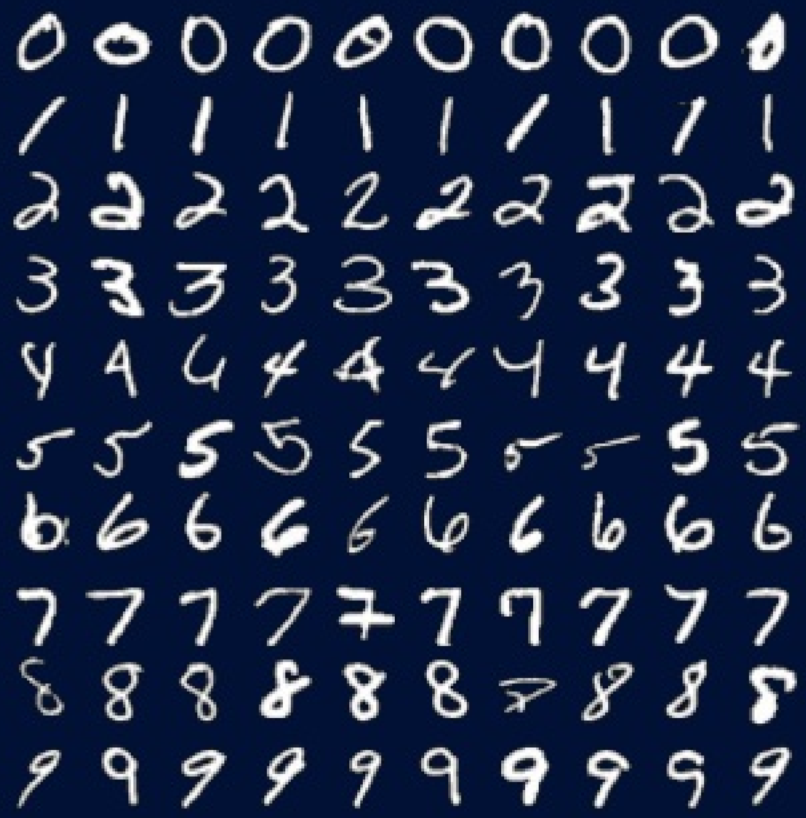




Identification of broadband connection impairments

Broadband connection diagnosis with DL

Soft-target labeling

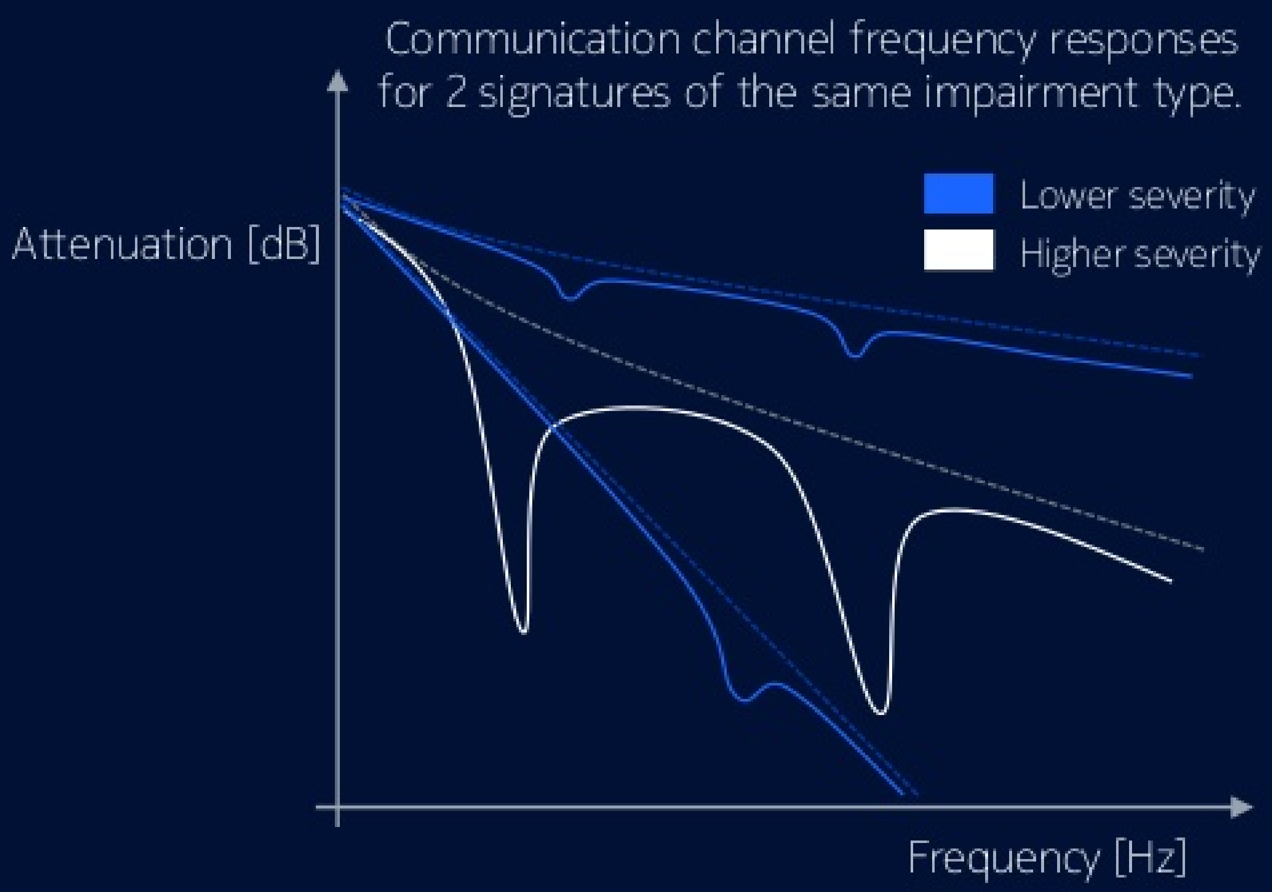


MNIST handwritten digits

- « Experts » (e.g. humans, authors) labelled formally each digit.
- No ambiguity is assumed during the training set creation (hard-labelling).

Broadband connection diagnosis with DL

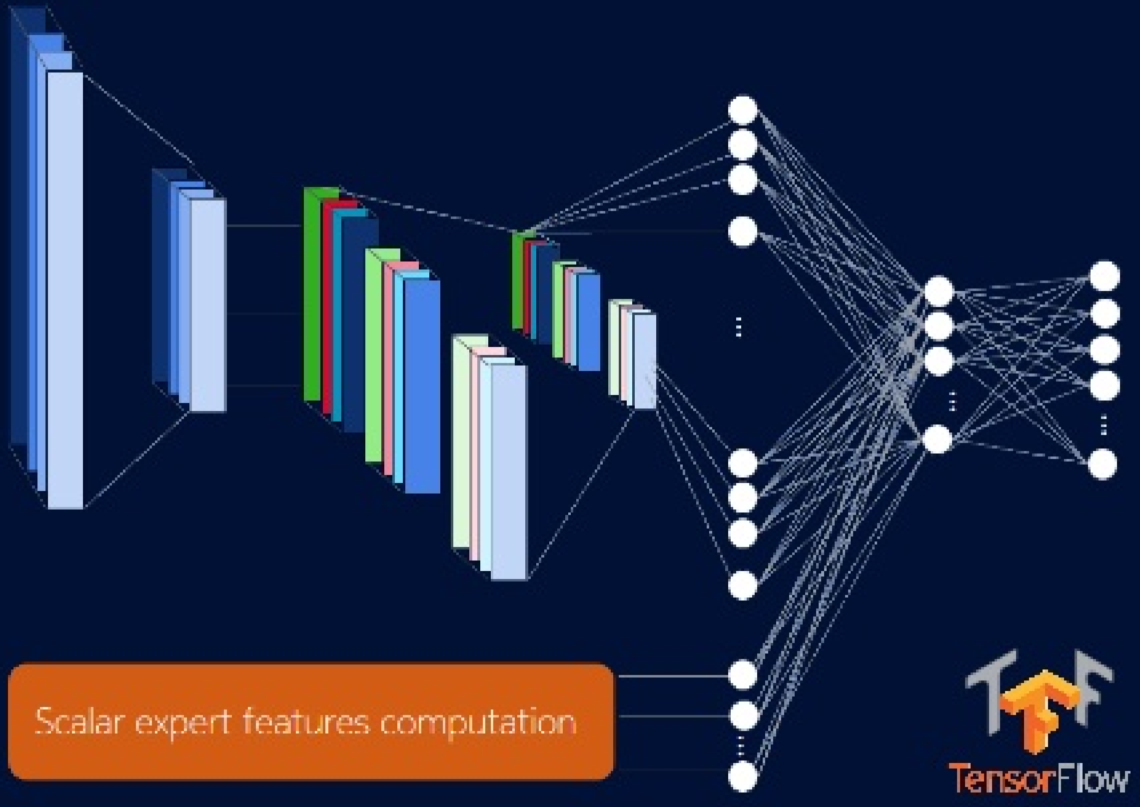
Soft-target labeling



- In problem detection, there is the notion of « severity ».
- This means within the same class of problem, the returned confidence needs also to reflect such severity.
- Solution is to make use of soft-labelling.
- Domain knowledge is required to assist deriving such soft-labels.
- Spark has been used to facilitate such processing over >20M curves.

Broadband connection diagnosis with DL

Expert scalar features addition



- Having the « perfect trained model » would require to build the « perfect training set ».
- Building a large, various and unbiased training set is hard.
- The convolutional layers might get therefore biased, leading to extra sensitivity.
- Adding empirical quantities to the fully-connected layer have helped in gaining in robustness (conservative approach).

Broadband connection diagnosis with DL

TensorFlow

- CNN model design
- Training over >20M samples
- Expert scalar features computation
- Trained model/session storage

python

- Pre-processing
- TensorFlow Python API

Apache Spark

- Soft-target labelling
- Hyper-parameter tuning (« grid search ») distribution
- Distributed execution

Broadband connection diagnosis with DL

	Empirical algorithms	Machine Learning	Deep Learning	Deep Learning + domain knowledge
Top accuracy	<50%*	>85%	>96%	>97%
Foot false positive rate	>20%*	<5%	<1%	<0.1%
Field performances	☆	☆☆	☆☆	☆☆☆☆

Quality of insights much valuable for the business.

* Narrower coverage.

Last advices…

- Data Science gives always better results when you know what is behind the data.
- Increasing your domain knowledge will save you a lot of time (and will make you a better data scientist ☺).
- ML/DL models that perform well in Notebooks may not give expected results in the field. Knowing how to move forward is the key !
- DL models usually performs better when guided with domain expertise.
- Try to get domain experts and data scientists in the same team.


#SAISEnt6
DIGITAL
EXPERIENCE
NOKIA