# Chair DSAIDIS
# Data Science and Artificial Intelligence *for Digitalized Industry and Services*

Florence d'Alché-Buc

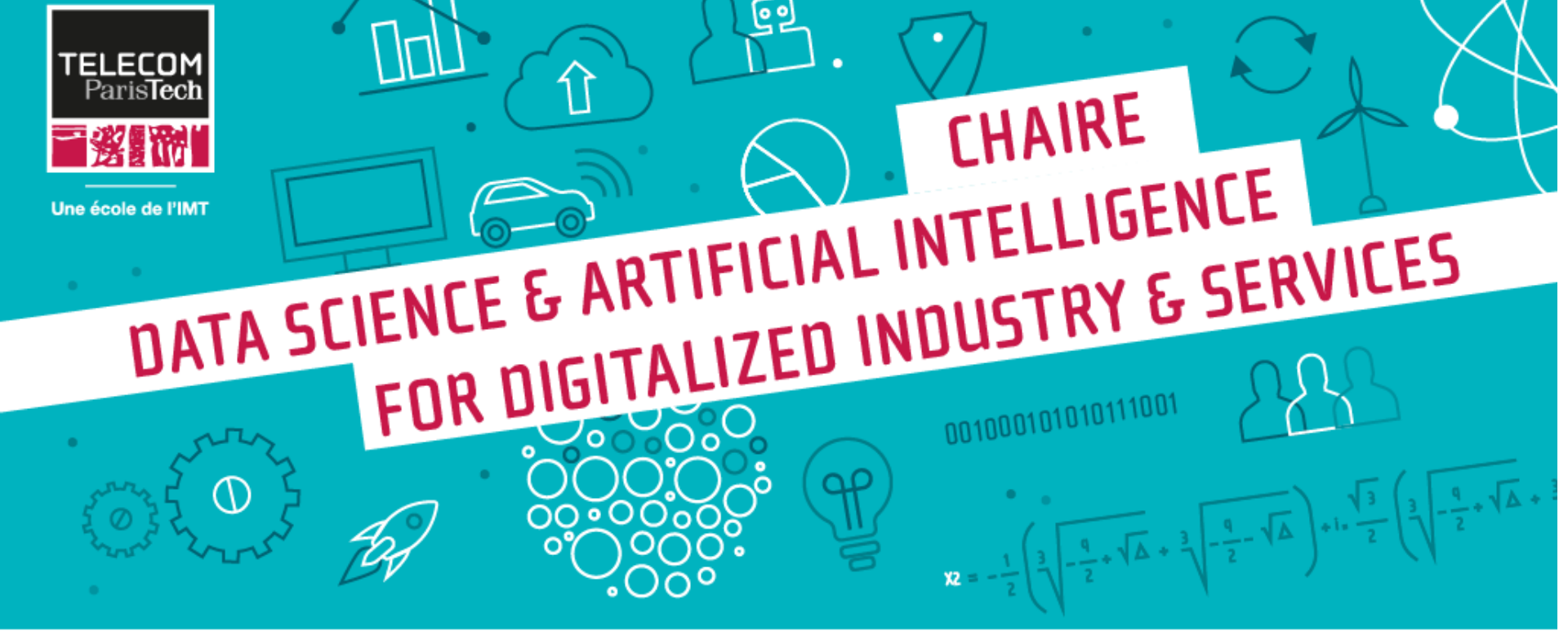
Sept 9, 2020 – DATAIA – Journée des chaires

TELECOM ParisTech

Une école de l'IMT

# CHAIRE

# DATA SCIENCE & ARTIFICIAL INTELLIGENCE FOR DIGITALIZED INDUSTRY & SERVICES




AIRBUS ENGIE IDEMIA augmented identity SAFRAN Valeo SMART TECHNOLOGY FOR SMARTER CARS FONDATION Mines-Télécom La Fondation de l'IMT

TELECOM Paris

IP PARIS

# Goal

- **Support fundamental researches and training on Machine Learning and Artificial Intelligence**
- **5 industrial partners**
- **Academic team: mainly S2A at LTCI, Télécom Paris**
- **A 5-year program started in early 2019**
  - **Research**
  - **Training**

# Academic Team

## REFERENTS FOR THE RESEARCH AXES

**CHLOÉ CLAVEL**
Interaction between humans and machines, analysis of users' socio-emotional behavior and strategies for socio-emotional interaction (axis 4)

**STEPHAN CLÉMENÇON**
Machine learning, stochastic processes, nonparametric statistics, stochastic modeling and applied statistics (axis 1)

**PAVLO MOZHAROVSKYI**
Statistical data depth function, classification, computational statistics, robust statistics, missing values, and data envelopment analysis (axis 2)

**FRANÇOIS PORTIER**
Sequential Monte-Carlo methods, machine learning for censored and dependent data and parsimonious predictive models (axis 3)

**FRANÇOIS ROUEFF**
Statistical signal processing and the analysis, random modeling of time series and statistics for stochastic processes (axis 1)

**ANNE SABOURIN**
Multivariate extreme value theory, dependence between rare events and dimension reduction in extreme regions (axis 3)

**GIOVANNA VARNI**
Analysis of non-verbal multimodal signals in human-human interaction and human-machine interaction (axis 4)

## PERMANENT MEMBERS

**ROLAND BADEAU**
Statistical modeling of non-stationary signals, with applications to audio and music.

**PASCAL BIANCHI**
Stochastic optimization, probability and signal processing.

**ALBERT BIFET**
Online learning from evolving data streams.

**THOMAS BONALD**
Analysis of network-structured data, multivariate time series and natural language processing.

**SLIM ESSID**
Temporal data analysis, multiview learning, structured prediction, representation learning and data decomposition.

**OLIVIER FERCOQ**
Optimization algorithms for high-dimensional problems.

**ONS JELASSI**
Real-time, distributed and scalable machine learning algorithms.

**LAURENCE LIKFORMAN**
Recurrent and convolutional neural networks in the recognition and analysis of handwriting.

**ROBERT M. GOWER**
Stochastic algorithms for big data problems in machine learning and scientific computing.

**GEOFFROY PEETERS**
Signal processing, machine learning and deep learning applied to audio and music data analysis.

**GAËL RICHARD**
Machine learning applied to speech, audio and music signals

**UMUT SIMSEKLI**
Scalable Bayesian machine learning, audio and music processing and recommendation systems.

**FABIAN M. SUCHANEK**
Semantic web, information extraction, automatic reasoning and knowledge bases.

# Research program

# A Double Motivation

**How can ML and AI tools be really useful in industry ?**
**How can industrials help us to raise and address crucial issues in ML/AI ?**

- data collected all along the life of a product

- noisy data, contaminated data

- new services: user / product / exogeneous data

- use of ML in critical environments

# Selected research topics

- realistic scenarii in Machine Learning: non iid data, extreme values, contaminated data

- new angles to old data: functional data analysis, point processes, survival analysis

- Robustness, Fairness, Reliability, Explainability

- Sustainability of tools: self-adaptation, re-use, knowledge distillation

# Research Axes

## Axis 1: Building predictive analytics on time series and data streams

Stephan Clémençon

François Roueff

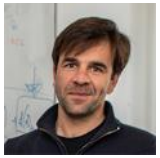
## Axis 2: Exploiting Large Scale, Heterogeneous, Partially Labeled Data

Florence d'Alché-Buc

Pavlo Mozharovskyi

## Axis 3: Machine Learning for trusted and robust decision

Anne Sabourin



François Portier

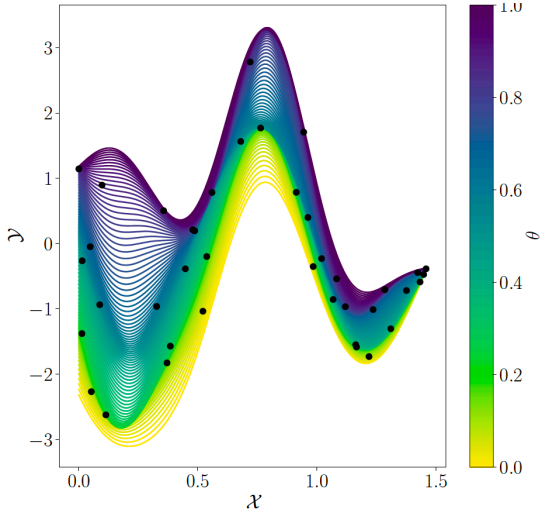
## Axis 4: Learning through interactions with environment

Chloé Clavel

Giovanna Varni

# 1 Predictive analytics and time series

- ***Focus on* functional data analysis**




  - Anomaly detection in functional data

(Staerman et al. 2019)

  - Functional Output Regression

(Lambert et al. 2020, Bouche et al. 2020)

  - Spatio-temporal series: modeling and forecasting functional time-series

# 2 Exploiting large scale, heterogeneous, partially labeled data

- **Ranking Preferences & label ranking (Vogel et al. 2020)**
- **Infinite task learning:** multi-task seen as functional regression (Brault et al. 2019)
- **Structured prediction:** hybrid architecture= kernel learning by neural networks (Motte et  al. 2020, El Ahmad et al. 2020)

# 3 Robust and reliable Machine Learning

- **Fairness :** re-weighting samples (Vogel et al. 2020)
- **Robustness :** robustness to contaminated data (Staerman et al. 2020), decision in presence of outliers (Jalalzai et al. 2019)
- **Reliability:** learning with abstention (Garcia et al. 2018)
- **Interpretability:** learning with interpretation, working group **Operational AI ethics @ Télécom Paris**

# 4 Learning through interactions with the environment

- **Self-adaptation** (Atamna et al. 2020, Ben Yousef et al. 2020)
- **Reinforcement learning and Bandits: profitable bandits** (Achab et al. 2019), logistic bandits (
- **Learning on a budget:** on-going work

# Transversal axis: optimization

- **Primal-dual algorithms: *Tran-Dinh et al. 2019, 2020***
- **Sketching and randomized subspaces: Gower et al. 2019**
- **Analysis of stochastic gradient descent algorithms: Barakat & Bianchi, 2019, Şimşekli et al. 2019**

# PhD students directly funded by the chair

Dimitri Bouche

Functional regression
and spatio-temporal processes

Jayneel Parekh

Learning interpretable
neural networks

Yousef Taheri Sojasi

Weak signals in NLP

# Postdocs

Asma Atamna
Learning for multimodal human-robot interaction

Sanjeel Parekh
Active learning, infinite task learning

Benoit Fuentes
Deep Tensor factorization

# Collaborations

- Aurélien Bellet, INRIA Lille
- Patrice Bertrail, Université Nanterre
- Marianne Clausel, Université de Lorraine
- Aurélien Garivier, ENS Lyon
- Alessandro Rudi, INRIA Paris

To name a few….

International collaborations:

Aalto, EPFL, Oxford U., NYU,KLEUVEN, Laval U.,…

# Scientific Animation

# International Workshop on Machine Learning and AI

- Two editions in 2018 and 2019
- Public Scientific Event
- Academic & industrial  audience

INSTITUTE DATAIA
Data Science, Intelligence & Society


# Events for our partners

**Day of the chair:** presentation of the main results, demo

**Thematic workshops:**

- tutorial on a topic
- talks about Methods and Applications
- Round table and open questions

**Themes:**

Time series modeling, Bandit algorithms, XAI, learning under uncertainty…

# Training

# Professional training

- MS Big Data (septembre)
- MS AI (septembre)
- CES Data Scientist
- CES AI

- Data Challenge: face recognition, semi-supervised classification…

- 6-month projects (« fil rouge ») :

Adversarial training, default detection, sales forecasting, autoML …

# DSAIDIS

**Website: https://datascienceandai.wp.imt.fr/**

**Collaboration, joint efforts on one of the topics: contact us**

**Internship, PhD, postdoc: call to come in november**

# Some recent publications


*Ahmet Alacaoglu, Olivier Fercoq et Volkan Cevher,* **Random extrapolation for primal-dual coordinate descent ,** *ICML 2020.*

*Louis Faury, Marc Abeille, Clément Calauzènes, Olivier Fercoq,* **Improved Optimistic Algorithms for Logistic Bandits** *ICML 2020.*

*Pierre Laforgue, Alex Lambert, Luc Brogat-Motte, Florence d'Alché-Buc,* **Duality in RKHSs with Infinite Dimensional Outputs: Application to Robust Losses,** *ICML 2020.*

Umut Simsekli, Lingjiong Zhu (FSU), Yee Whye Teh (Oxford and DeepMind), Mert Gurbuzbalaban (Rutgers University), **Fractional Underdamped Langevin Dynamics: Retargeting SGD with Momentum under Heavy-Tailed Gradient Noise,** ICML 2020.

Guillaume Staerman, Pavlo Mozharovskyi, Stephan Clémençon , **The Area of the Convex Hull of Sampled Curves: a Robust Functional Statistical Depth measure**, AISTATS 2020.

Robin Vogel, Stephan Clémençon, **A Multiclass Classification Approach to Label Ranking**, AISTATS 2020.