On some statistical and machine learning problems in research and engineering

Part 1: mixtures-of-experts for heterogenous and high-dimensional data
Part 2: challenges and industrial applications in Hybrid & Trustworthy AI @ SystemX

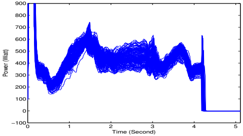
Faïcel Chamroukhi

Talk at:

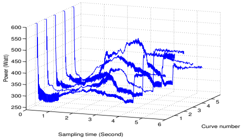
Real-world data are complex

- Heterogenous, Multimodal, High-Dimensional, Unlabeled, Possibly Massive ...
- Need for adapted analysis tools

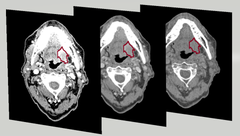
Transport : Railway switch curves diagnostic



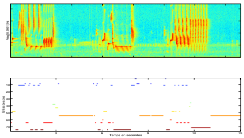
Predictive Maintenance

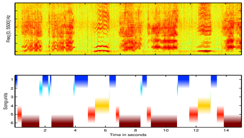


Health : Medical images

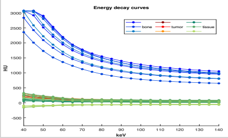


Acoustics : scene listening (marine, terrestrial)

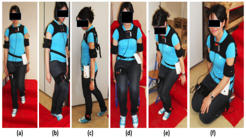




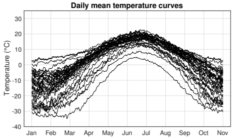
Dual-energy computed tomography

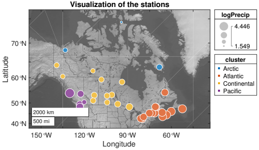


Health & Well Being : Activity recog.



Climate/Environment : meteorological data





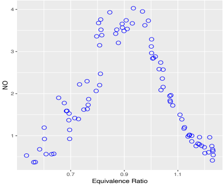
- Establish well-principled (with statistical guarantees) predictions in heterogeneous and high-dimensional situations,
- Construct efficient algorithms that operate in unsupervised way and provide interpretable solutions with computational guarantees.

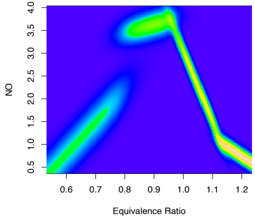
Modeling framework

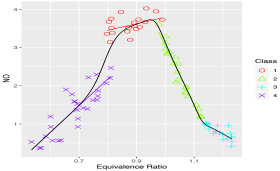
$\hookrightarrow$ Latent variable **models** : $f(x|\boldsymbol{\theta}) = \int_{\mathcal{Z}} f(x, z|\boldsymbol{\theta})\mathrm{d}z$

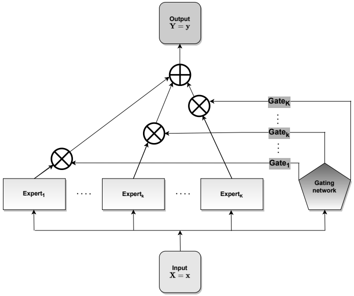
$\hookrightarrow$ **Learning**, **representation** and model **selection** in high-dimension

Mixtures-of-Experts as good candidates to model a response Y given predictor.s X governed by a hidden structure accounting for heterogeneity



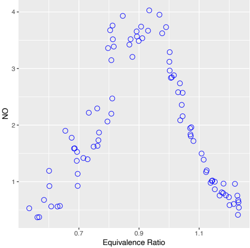




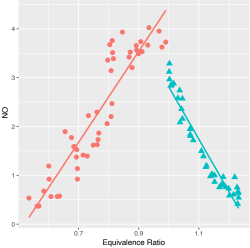


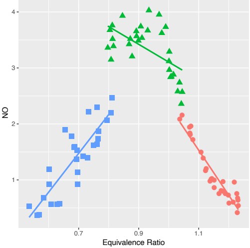
Model estimation and selection

(a) Raw Ethanol data set

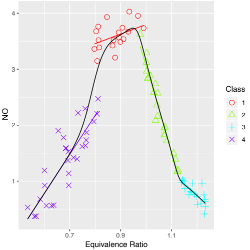


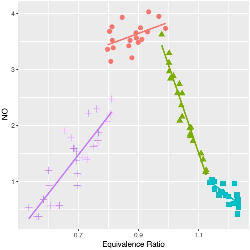
Collection of MoE models with linear mean functions characterized by 2-5 clusters

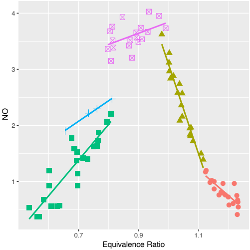




(b) Our best data-driven MoE model







Approximation capabilities of finite mixture distributions

Density approximation in Unsupervised Learning

- **Data** : observations $\{\boldsymbol{x}_i\}$ from $\boldsymbol{X} \in \mathbb{X} \subset \mathbb{R}^d$ of density (multimodal) $f \in \mathcal{F}$
- **Objective** : approximate the density f (and represent the data, e.g. *clustering*)
- **Solution :** Approximate f within the class $\mathcal{H}^{\varphi} = \bigcup_{K \in \mathbb{N}^{\star}} \mathcal{H}_K^{\varphi}$ of finite location-scale mixture h_K^{φ} (of K-components) of density φ (e.g., Gaussian), where

$$\mathcal{H}_K^{\varphi} = \left\{ \boxed{h_K^{\varphi}(\boldsymbol{x}) := \sum_{k=1}^{K} \pi_k \frac{1}{\sigma_k^d} \varphi\left(\frac{\boldsymbol{x} - \boldsymbol{\mu}_k}{\sigma_k}\right)}, \boldsymbol{\mu}_k \in \mathbb{R}^d, \sigma_k \in \mathbb{R}_+, \pi_k > 0 \, \forall k \in [K], \sum_{k=1}^{K} \pi_k = 1 \right\}$$

Theorem : Universal approximation of finite location-scale mixtures

(a) Given any p.d.f $f, \varphi \in \mathcal{C}$ and a compact set $\mathbb{X} \subset \mathbb{R}^d$, there exists a sequence $(h_K^{\varphi}) \subset \mathcal{H}^{\varphi}$, such that $\lim_{K \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X}} |f(\boldsymbol{x}) - h_K^{\varphi}(\boldsymbol{x})| = 0$.

(b) For $p \in [1, \infty)$, if $f \in \mathcal{L}_p$ (Lebesgue p.d.f) and $\varphi \in \mathcal{L}_{\infty}$ (essentially bounded p.d.f), there exists a sequence $(h_K^{\varphi}) \subset \mathcal{H}^{\varphi}$, such that $\lim_{K \to \infty} \|f - h_K^{\varphi}\|_{\mathcal{L}_p} = 0$.

[J. Communications in Statistics - Theory and Methods, 2022] [PhD, TT Nguyen 2021]

Learning with mixtures-of-experts (ME)

- **Context :** n observations $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ from a pair $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{X} \times \mathbb{Y}$ with unknown conditional p.d.f $f \in \mathcal{F} = \left\{f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}_+ \mid \int_{\mathbb{Y}} f(\boldsymbol{y}|\boldsymbol{x})\, \mathrm{d}\lambda(\boldsymbol{y}) = 1, \forall \boldsymbol{x} \in \mathbb{X}\right\}$
- **High-dimensional setting** : $\mathbb{X} \subseteq \mathbb{R}^d$, $\mathbb{Y} \subseteq \mathbb{R}^q$, with $d, q \gg n$ and **heterogeneous**.
- **Objectives :** Regression ; Clustering ; Model selection
- **Solution :** Approximate f within the class of **mixtures-of-experts** :

Let φ be a p.d.f (compactly supported on $\mathbb{Y} \subseteq \mathbb{R}^q$), we define the functional classes :

- Location-scale family : $\mathcal{E}_\varphi = \left\{\phi_q(\boldsymbol{y}; \boldsymbol{\mu}, \sigma) := \frac{1}{\sigma^q} \varphi\left(\frac{\boldsymbol{y}-\boldsymbol{\mu}}{\sigma}\right); \boldsymbol{\mu} \in \mathbb{Y}, \sigma \in \mathbb{R}_+\right\}$.
- Mixture of location-scale experts with softmax activation network : SGaME :

$$\mathcal{H}_S^\varphi = \left\{\boxed{h_K^\varphi(\boldsymbol{y}|\boldsymbol{x}) := \sum_{k=1}^{K} g_k(\boldsymbol{x}; \boldsymbol{\gamma}) \phi_q(\boldsymbol{y}; \boldsymbol{\mu}_k, \sigma_k)}; \quad \phi_q \in \mathcal{E}_\varphi \cap \mathcal{L}_\infty, g_k(\cdot; \boldsymbol{\gamma}) \in \{\text{softmax}\}\right\}$$

Theorem : Approximation capabilities of isotropic mixtures-of-experts SGaME

(a) For $p \in [1, \infty)$, $f \in \mathcal{F}_p \cap \mathcal{C}$, $\varphi \in \mathcal{F} \cap \mathcal{C}$, $\mathbb{X} = [0,1]^d$, there exists a sequence $(h_K^\varphi) \subset \mathcal{H}_S^\varphi$ such that $\lim_{K\to\infty} \left\|f - h_K^\varphi\right\|_{\mathcal{L}_p} = 0$.

(b) For $f \in \mathcal{F} \cap \mathcal{C}$, if $\varphi \in \mathcal{F} \cap \mathcal{C}$, $d = 1$, there exists a sequence $(h_K^\varphi) \subset \mathcal{H}_S^\varphi$ such that $\lim_{K\to\infty} h_K^\varphi = f$ almost uniformly.

[PhD TT. Nguyen, 2021] [Journal of Stat. Distributions and Applicat., 2021] [Neurocomputing, 2019] [WIREs DMKD 2018]

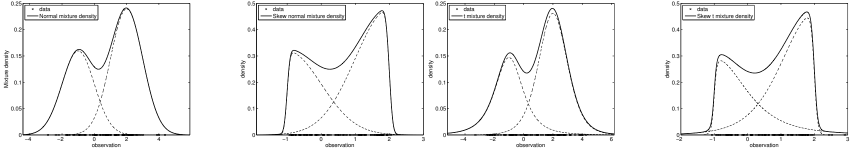
Principled robustness in learning with MoE

Principled robustness in regression and clustering

- **Questionings :** Prediction (non-linear regr., classification) & clustering in presence of **Outliers**, with potentially **skewed**, **heavy-tailed** distributions
- **Answering :** Robust MoE that accommodate asymmetry, heavy tails, and outliers

$$m(y|\boldsymbol{r},\boldsymbol{x};\boldsymbol{\theta}) = \sum_{k=1}^{K} \underbrace{g_k(\boldsymbol{r};\boldsymbol{\alpha})}_{\text{Softmax Gating Network}} \underbrace{\mathcal{ST}(y;\mu(\boldsymbol{x};\boldsymbol{\beta}_k),\sigma_k,\lambda_k,\nu_k)}_{\text{Skew-}t\text{ Expert Network}}$$

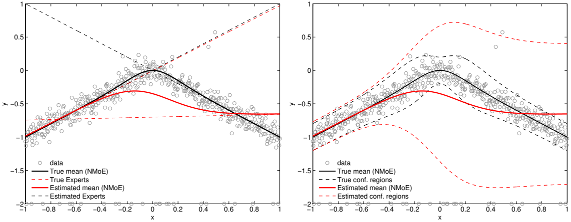
kth expert : has a skew t distribution [Azzalini and Capitanio 2003]



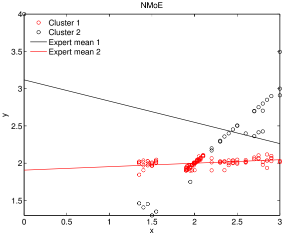
$\pi_k = [0.4, 0.6]$, $\mu_k = [-1, 2]$; $\sigma_k = [1, 1]$; $\nu_k = [3, 7]$; $\lambda_k = [14, -12]$;

Flexible and robust generalization of the standard MoE models

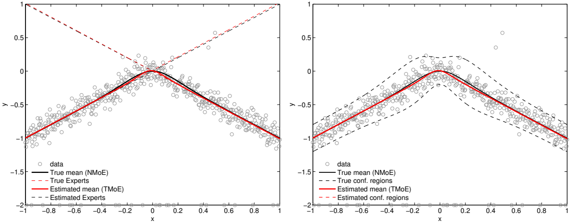
For $\{\nu_k\} \to \infty$, STMoE reduces to SNMoE ; For $\{\lambda_k\} \to 0$, STMoE reduces to TMoE.
For $\{\nu_k\} \to \infty$ and $\{\lambda_k\} \to 0$, StMoE approaches the NMoE.



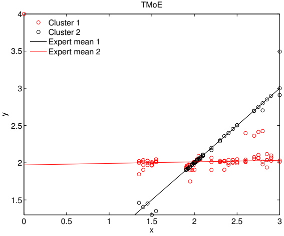
$n = 500$ observations with 5% of outliers $(x; y = -2)$: Normal fit



Tone data with 10 outliers $(0, 4)$: Normal fit



$n = 500$ observations with 5% of outliers $(x; y = -2)$: **Robust** fit



Tone data with 10 outliers $(0, 4)$: **Robust** fit

Learning via the EM algorithm

SaMUraiS : open source software for statistical time-series analysis

SaMUraiS : **St**A**tistical** **M**odels for the **U**nsupe**R**vised segment**A**t**I**on of time-**S**eries

Github | CRAN | Matlab software

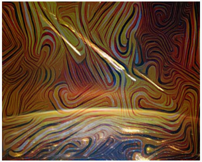
Available algorithms and Packages

RHLP : Regression with Hidden Logistic Process	R software	Matlab software
HMMR : Hidden Markov Model Regression	R software	Matlab software
PWR : Piece-Wise Regression	R software	Matlab software
MRHLP : Multivariate RHLP	R software	Matlab software
MHMMR : Multivariate HMMR	R software	Matlab software
MPWR : Multivariate PWR	R software	Matlab software

Include estimation, segmentation, approximation, model selection, and sampling

Open-Source Toolkit

MEteorits : open-source soft. Robust learning with mixtures-of-experts models



MEteorits : **M**ixtures-of-**E**xper**T**s mod**E**ling for c**O**mplex and non-no**R**mal d**I**s**T**ribution**S**

▸ Github ▸ CRAN ▸ Matlab software

Available algorithms and Packages

`NMoE` : Normal Mixture-of-Experts	▸ R software	▸ Matlab software
`SNMoE` : Skew-Normal Mixture-of-Experts	▸ R software	▸ Matlab software
`tMoE` : Robust MoE using the t-distribution	▸ R software	▸ Matlab software
`StMoE` : Skew-t Mixture-of-Experts	▸ R software	▸ Matlab software

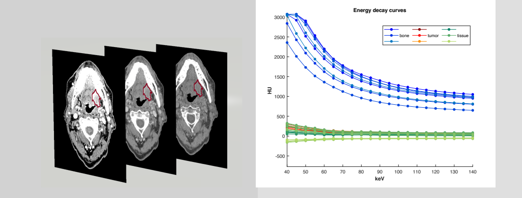
- Meteorits include sampling, fitting, prediction, clustering with each MoE model
- Non-normal mixtures (and MoE) is a very recent topic in the field

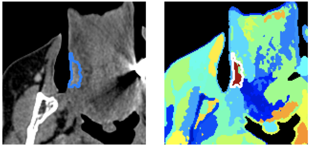
Dual-energy computed tomography (DECT) image Clustering

- Learning from Multimodal information in Healthcare/Radiology
- Cancer detection in Radiology : DECT clustering [Diagnostics (AI in medicine), 2022]

  Spatial mixture of functional regressions for dual-energy CT images

  $m(\boldsymbol{y}|\boldsymbol{x}, v; \boldsymbol{\theta}) = \sum_{k=1}^{K} \alpha_k(v; \boldsymbol{\alpha}) f_k(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}_k)$ where $\alpha_k(\boldsymbol{v}; \boldsymbol{\alpha}) = \frac{w_k \phi_3(\boldsymbol{v}; \boldsymbol{\mu}_k, \mathbf{R}_k)}{\sum_{\ell=1}^{K} w_\ell \phi_3(\boldsymbol{v}; \boldsymbol{\mu}_\ell, \mathbf{R}_\ell)}$

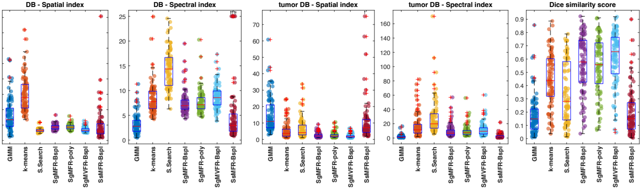




(a) Original slice

(b) Dice=0.84, DB=1.64/6.92

DECT multimodal Data : 3D voxels & energy levels Expert Annotation Automatic Annotation



▸ Codes available on Github

Questioning : Prediction (non-linear regr., classification) & clustering in presence of

[1.] **High-dimensional** predictors : $\boldsymbol{X}_i \in \mathbb{R}^p$ with $p \gg n$

[2.] **Functional** predictors : $X_i(t)$, $t \in \mathcal{T} \subseteq \mathbb{R}$ {eg. continuously recorded variables}

$\hookrightarrow$ Look for parsimonious and interpretable methods

[1.] HDME : High-Dimensional Mixtures-of-Experts

- Learning : PMLE $\widehat{\boldsymbol{\theta}}_n \in \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log h_K^{\varphi}(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta}) - \text{pen}(\boldsymbol{\theta})$
- $\hookrightarrow$ LASSO penalty : $\text{Pen}_\lambda(\boldsymbol{\theta}) = \underbrace{\sum_{k=1}^{K} \lambda_k \|\boldsymbol{\beta}_k\|_1}_{\text{Experts Net.}} + \underbrace{\sum_{k=1}^{K-1} \gamma_k \|\boldsymbol{w}_k\|_1}_{\text{Gating Net.}}$

$\hookrightarrow$ encourages sparse solutions & performs estimation and feature selection

$\hookrightarrow$ computationally attractive (Avoid matrix inversion ; univariate updates)

▸ Software Toolbox HDME on Github (GaussRMoE, LogisticRMoE, PoissonRMoE)

[PhD] Bao Tuyen Huynh. *Estimation and Feature Selection in High-Dimensional Mixtures-of-Experts Modesls* . PhD Thesis, Normandie Université, 2019.

[J] Chamroukhi &Huynh. Regularized Maximum Likelihood Estimation and Feature Selection in Mixtures-of-Experts Models. Journal de la Société Francaise de Statistique, Vol. 160(1), pp :57–85, 2019

[J] Huynh & C. Estimation and Feature Selection in Mixtures of Generalized Linear Experts Models. arXiv :1810.12161, 2019

Questioning : Prediction (non-linear regr., classification) & clustering in presence of **High-dimensional** predictors : Data $\mathcal{D}_n = (\boldsymbol{X}_i, Y_i)_{i=1}^n$ where $\boldsymbol{X}_i \in \mathbb{R}^p$ with $p \gg n$

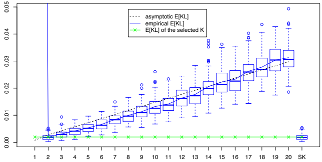
HDME : High-Dimensional MoE : PMLE $\widehat{\boldsymbol{\theta}}_n \in \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^n \log h_K^{\varphi}(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta}) - \mathrm{pen}(\boldsymbol{\theta})$

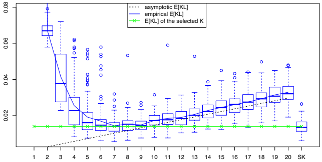
Theorem : Non-asymptotic oracle inequality for collection of MoE models

Result : $\exists$ constants C et $\kappa(\rho, C_1) > 0$ ($C_1 > 1$) s.that whenever for $\mathbf{m} \in \mathcal{M}$, $\mathrm{pen}(\mathbf{m}) \geq \kappa(\rho, C_1)\left[(C + \ln n)\dim(\mathcal{H}_{\mathbf{m}}) + z_{\mathbf{m}}\right]$, the estimator PMLE $\widehat{h}_{\widehat{\mathbf{m}}}$ satisfies

$$\mathbb{E}\left[\mathrm{JKL}_{\rho}^{\otimes n}\left(f, \widehat{h}_{\widehat{\mathbf{m}}}\right)\right] \leq C_1 \inf_{\mathbf{m}\in\mathcal{M}}\left(\inf_{h_{\mathbf{m}}\in\mathcal{H}_{\mathbf{m}}} \mathrm{KL}^{\otimes n}(f, h_{\mathbf{m}}) + \frac{\mathrm{pen}(\mathbf{m})}{n}\right) + \frac{\kappa(\rho, C_1) C_1 \xi}{n} + \frac{\eta + \eta'}{n}.$$

- A non-asymptotic result. If $\mathrm{pen}(\mathbf{m})$ is well chosen, then our PMLE behaves in a comparable manner compared to **the best (oracle) model** $\mathcal{H}_{\mathbf{m}^\star}$ in the collection, minimizing the risk : $\inf_{\mathbf{m}\in\mathcal{M}}\left(\inf_{h_{\mathbf{m}}\in\mathcal{H}_{\mathbf{m}}} \mathrm{KL}^{\otimes n}(f, h_{\mathbf{m}}) + \frac{\mathrm{pen}(\mathbf{m})}{n}\right)$ (f is unknown).





[PhD, Trung-Tin Nguyen, 2021.] [*EJS*, 2022] [Australasian Joint Conference on Artificial Intelligence 2024.]

FLaMingoS : open source software for learning from functions

FLaMingoS : **F**unctional **L**atent dat**A** **M**odels for cluster**ING** heterogene**O**us time-**S**eries

▸ Github ▸ CRAN ▸ Matlab software

Available algorithms and Packages

`mixRHLP` : Mixture of Regressions with HLPs	▸ R	▸ Matlab
`mixHMM` : Mixture of Hidden Markov Models (HMMs)	▸ R	▸ Matlab
`mixHMMR` : Mixture of HMM Regressions	▸ R	▸ Matlab
`PWRM` : Piece-Wise Regression Mixture	▸ R	▸ Matlab
`uReMix` : Unsupervised Regression Mixtures	▸ R	▸ Matlab

$\hookrightarrow$ A flexible full generative modeling for FDA

$\hookrightarrow$ Could be extended to the multivariate case without a major effort

[2.] Learning with functional predictors

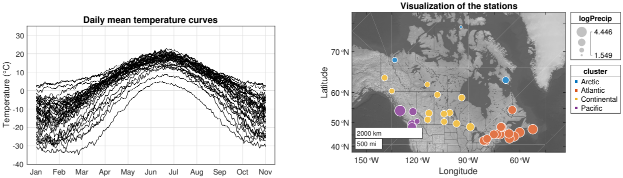


FIGURE – $n = 35$ daily mean temperature measurement curves (X_i's) in different stations (Left) and the $\log$ of precipitation values (Y_i's) visualized with the climate regions (Z_i's) (Right).

- Relate functional predictors $\{X(t) \in \mathbb{R}; t \in \mathcal{T} \subset \mathbb{R}\}$ to a scalar response $Y \in \mathcal{Y} \subset \mathbb{R}$
- Regression and classification of heterogeneous responses given functional predictors

  (1) generative functional modeling, sparsity and feature selection (high-dimension)

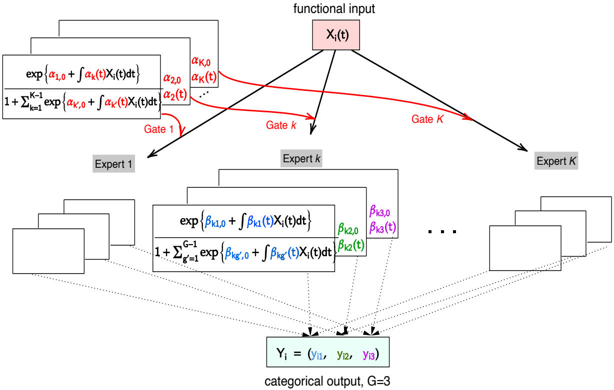
  (2) User guideline : keep an interpretable fit

[2.] Functional Mixtures-of-Experts (and Different Learning strategies, in particular)

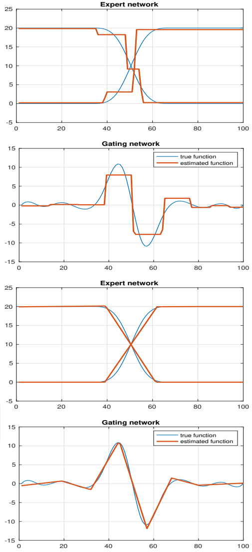
- $Y_i = \beta_{z_i,0} + \int_{\mathcal{T}} X_i(t)\beta_{z_i}(t)dt + \varepsilon_i$ avec $h_z(X_i(.)) = \alpha_{z_i,0} + \int_{\mathcal{T}} X_i(t)\alpha_{z_i}(t)dt$
- Lasso-type Regularized MLE w.r.t the derivatives of the $\alpha(\cdot)$ and $\beta(\cdot)$ functions

Chamroukhi, Pham, Hoang, McLachlan. Functional Mixtures-of-Experts. *Statistics and Computing Springer.*, Vol. 34 (98), 2024

Mixture-of-Experts Architecture



Interpretable fits



$Y_i = \beta_{z_i,0} + \int_{\mathcal{T}} X_i(t)\beta_{z_i}(t)dt + \varepsilon_i$ with $h_z(X_i) = \alpha_{z_i,0} + \int_{\mathcal{T}} X_i(t)\alpha_{z_i}(t)dt$
l_1-Regularized MLE w.r.t the derivatives of the $\alpha(\cdot)$ and $\beta(\cdot)$ functions

$\hookrightarrow$ produces a meaningful sparse estimates for $\beta_{z_i}(t)$ curves :

$\beta^{(0)}_{z_i}(t) = 0$ implies that $X(t)$ has no effect on Y at t

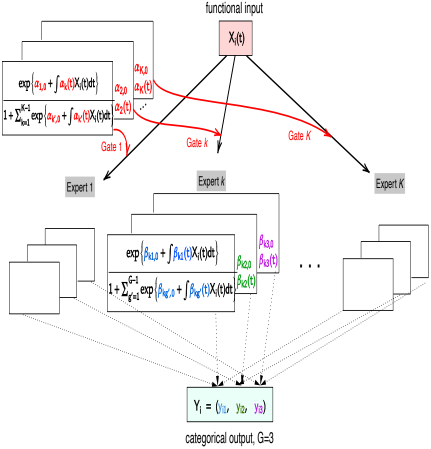
$\beta^{(1)}_{z_i}(t) = 0$ means that $\beta_{z_i}(t)$ is constant at t,

$\beta^{(0)}_{z_i}(t) = 1$ shows that $\beta_{z_i}(t)$ is a linear function of t, etc.

[PhD TN. Pham, 2022] [Functional Mixtures-of-Experts. *Statistics and Computing Springer.*, Vol. 34 (98), 2024]

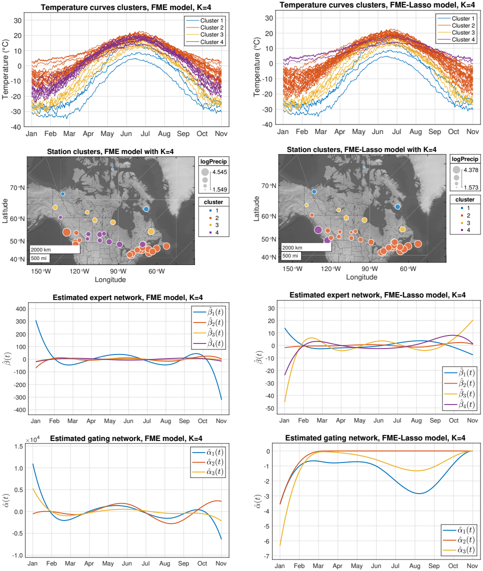
Interpretable learning with time-series inputs

Mixture-of-Experts Architecture



No regularization

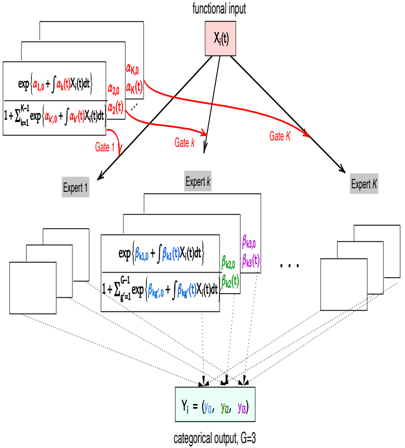
LASSO regularization



[PhD TN. Pham, 2022]

[Functional Mixtures-of-Experts. *Statistics and Computing Springer.*, Vol. 34 (98), 2024]

Mixture-of-Experts Architecture



produces a meaningful sparse estimates for $\beta_{z_i}(t)$ curves :

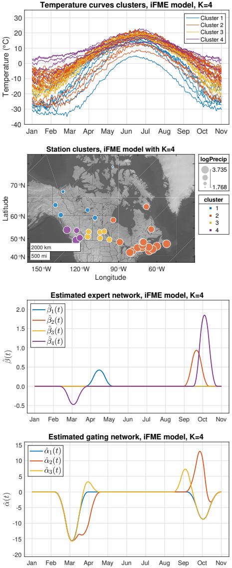
$\beta^{(0)}_{z_i}(t) = 0$ implies that $X(t)$ has no effect on Y at t

$\beta^{(1)}_{z_i}(t) = 0$ means that $\beta_{z_i}(t)$ is constant at t,

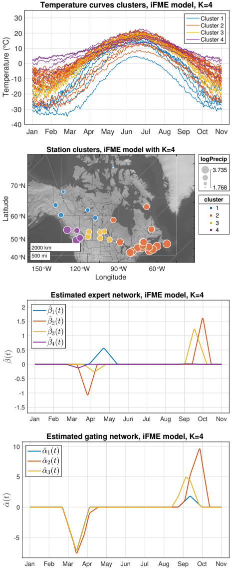
$\beta^{(0)}_{z_i}(t) = 1$ shows that $\beta_{z_i}(t)$ is a linear function of t,

etc.

OUR regularization



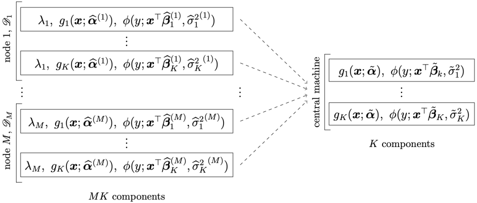
OUR regularization



[PhD TN. Pham, 2022] [Functional Mixtures-of-Experts. *Statistics and Computing Springer.*, Vol. 34 (98), 2024]

Aggregating distributed mixtures-of-experts models (MoE)

- collaborative MoE for distributed (eg. large-scale data) or federated learning



- Local estimators : $\hat{f}_m = f(\cdot|\mathbf{x},\widehat{\boldsymbol{\theta}}_m) = \sum_{k=1}^K g_k(\mathbf{x},\widehat{\boldsymbol{\alpha}}^{(m)})\phi(\cdot;\mathbf{x}^\top\widehat{\boldsymbol{\beta}}_k^{(m)},\widehat{\sigma}_k^{2\,(m)})$,
- weighted average : $\bar{f} = f(y|\mathbf{x};\bar{\boldsymbol{\theta}}) = \sum_{m=1}^M \lambda_m \hat{f}_m$ where $\lambda_m = \frac{N_m}{N}$ the sample proportion. $\bar{f}$ is good but relates MK components so not our direct target.

↪ Reduced estimator : $\bar{f}^R = \underset{h_K\in\mathcal{M}_K}{\arg\inf}\ \rho\left(h_K,\sum_{m=1}^M \lambda_m\hat{f}_m\right)$: we seek for a K-component ME h that is closest to the MK-component ME $\bar{f} = \sum_{m=1}^M \lambda_m\hat{f}_m$ w.r.t a transportation divergence $\rho(\cdot,\cdot)$, e.g. KL.

{PhD, Pham. 2022} [Distributed Learning of Mixtures of Experts. *arxiv 2312.09877*, 2024]

▸ Source codes publicly available on Github.

Federated Learning

Numerical results in Distributed clustering and Prediction

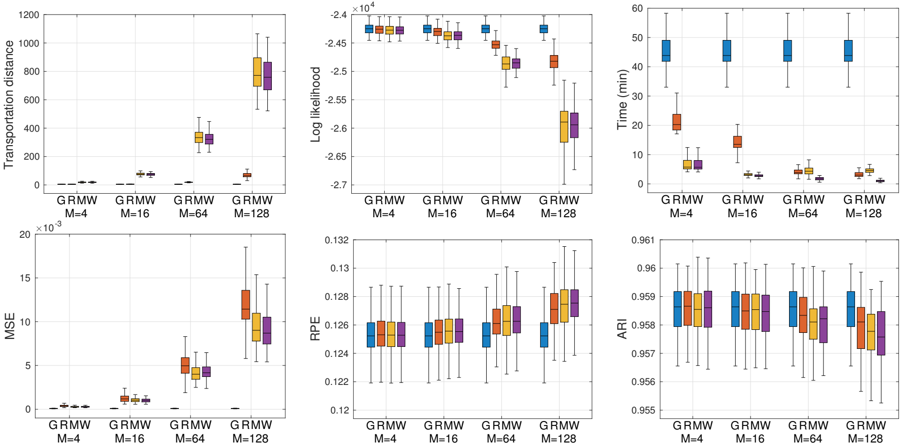


FIGURE – Performance of the Global ME (G), Reduction (R), Middle (M) and Weighted average (W) estimator for sample size $N = 10^6$ and M machines.

{PhD, Pham. 2022} [Distributed Learning of Mixtures of Experts. *arxiv 2312.09877*, 2024] [Under revision at IEEE TNNLS]

▸ Source codes publicly available on Github.

Available (in free access) at
https://chamroukhi.com/publications.php

Systemx INSTITUTE OF RESEARCH AND TECHNOLOGY

Boosting
digital transformation

Challenges and industrial applications in Hybrid AI Trustworthy AI @ SystemX

Faïcel Chamroukhi

Talk at:

KOÇ ÜNİVERSİTESİ

KUIS AI CENTER

Istanbul, 09 may 2024

The Research and Technology Organisation (RTO) of Paris-Saclay

SystemX
INSTITUTE OF RESEARCH AND TECHNOLOGY

Research and Technology Organisation (RTO)
Non-profit Scientific Cooperation Foundation

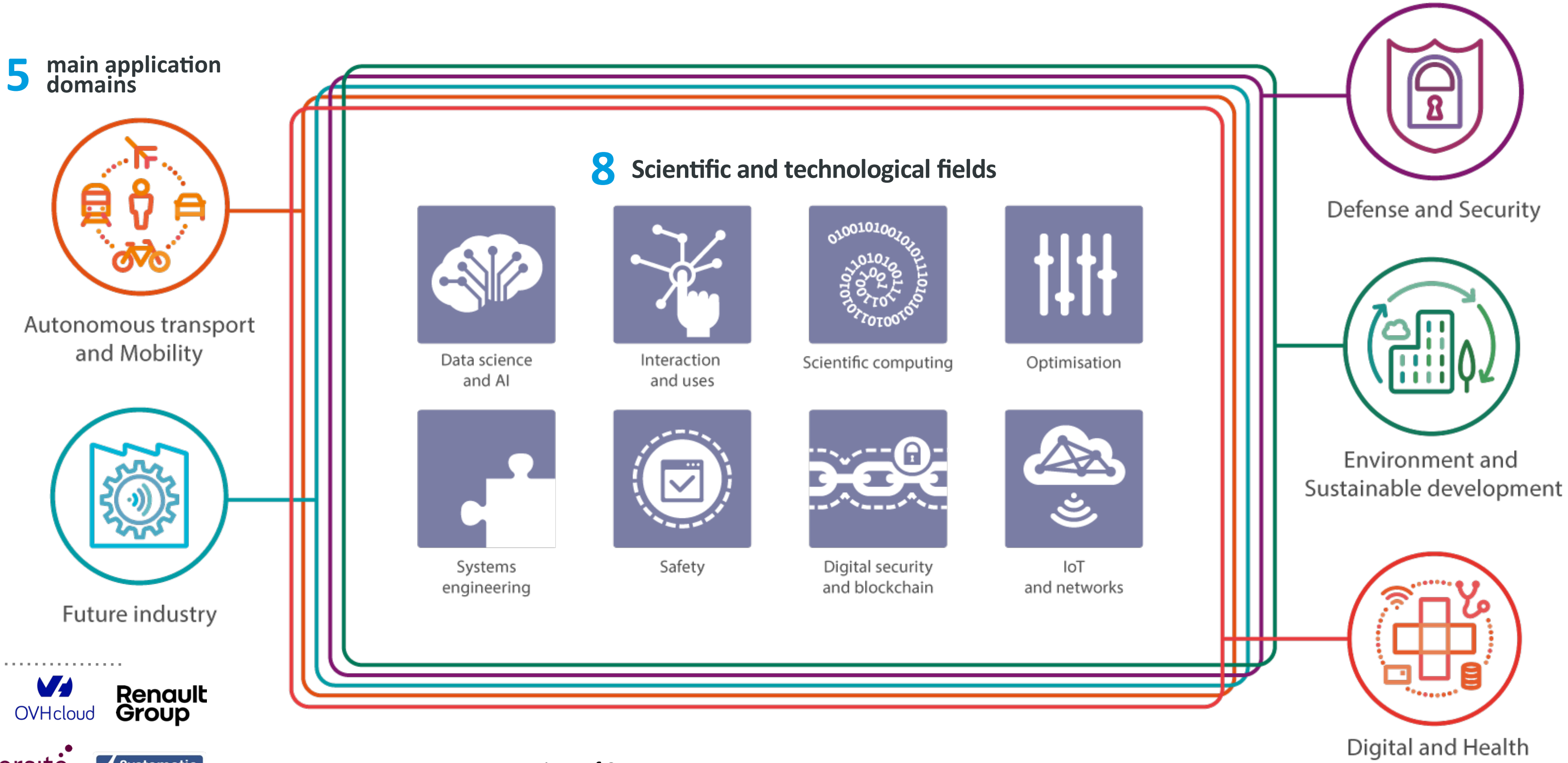
Paris-Saclay

65 Economic partners of which **1/3** are large groups and **2/3** are SMEs

50 Academic partners

Leads market-driven and applied research projects for the digital transformation of industry, services and territories:

1. Expertise: analysis, modeling, simulation and decision management
2. Own skills
3. Own assets: software, cyber-physical and tool-based platforms



Founding members ALSTOM, Bull atos technologies, COSMOTECH, IDnomic, OVHcloud, Renault Group, SAFRAN, SHERPA ENGINEERING, INSTITUT POLYTECHNIQUE DE PARIS, Inria, Institut Mines-Télécom, université PARIS-SACLAY, Systematic Paris Region Digital Ecosystem

An overview of SystemX

Outline

Challenges and industrial applications in

- Hybrid AI
- Trustworthy AI

Hybrid modeling: combining Machine Learning and *Physics*

Hybrid modeling: combining ML and *Physics*

➔ Enables prior scientific knowledge based on physics to be taken into account in data-driven machine learning methods: e.g approcahes include PINNs - Physics-Informed Neural Nets (Raissi's paper in 2019)

➔ Has been successfully and increasingly applied to solve a wide variety of linear and nonlinear problems in physics, covering various fields like mechanics, fluid dynamics, thermodynamics, electromagnetism …

In engineering, it allows

➔ the integration of analytical knowledge from physical laws governing the studied engineering systems

- to augment th statistical knowledge learned from observed/measured data (eg. Information extracted by deep learning from data)

- for reducing the high cost of physical simulation, in particular in the industrial sector

Raissi, M et al. (2019) Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. Journal of Computational Physics. 378. Online

Cuomo, S., et al., (2022). Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, *92*(3), 88. Read Online

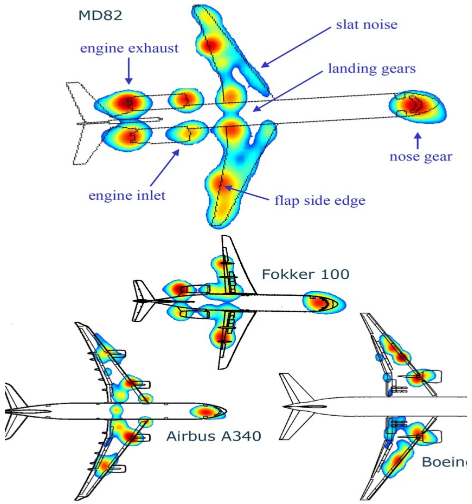
Some physical problems in Industry

- Related to the desing and supervision of complex (physical) systems
- Covering various fields in physics (mechanics, fluid dynamics, aerodynamics, electromagnetism …)
- In a wide variety of Applications in industry, in particular in numerical **simulation**

Electricity (power grids)

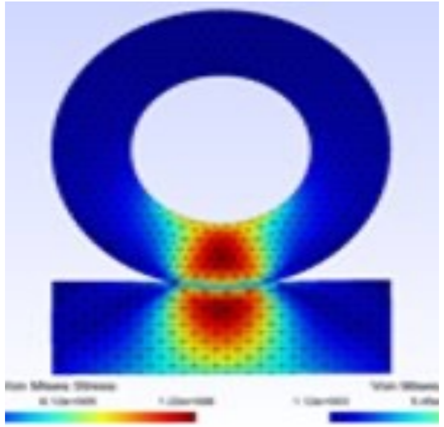
Picture from Marot, A., et al. (2018). Guided machine learning for power grid segmentation. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* (pp. 1-6).

Aerodynamics



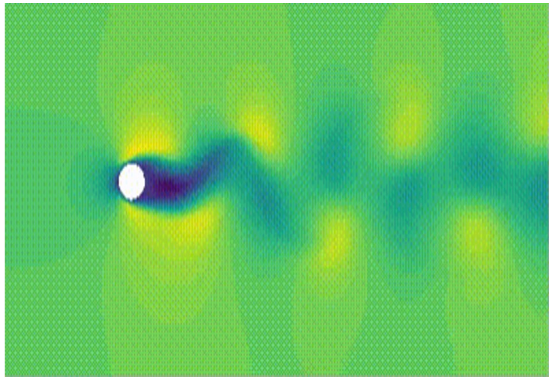
Merino-Martínez *et al. CEAS Aeronautical* Journal (2019).

Solid Mechanics pneumatics



From the internet

Fluid Flows/Dynamics



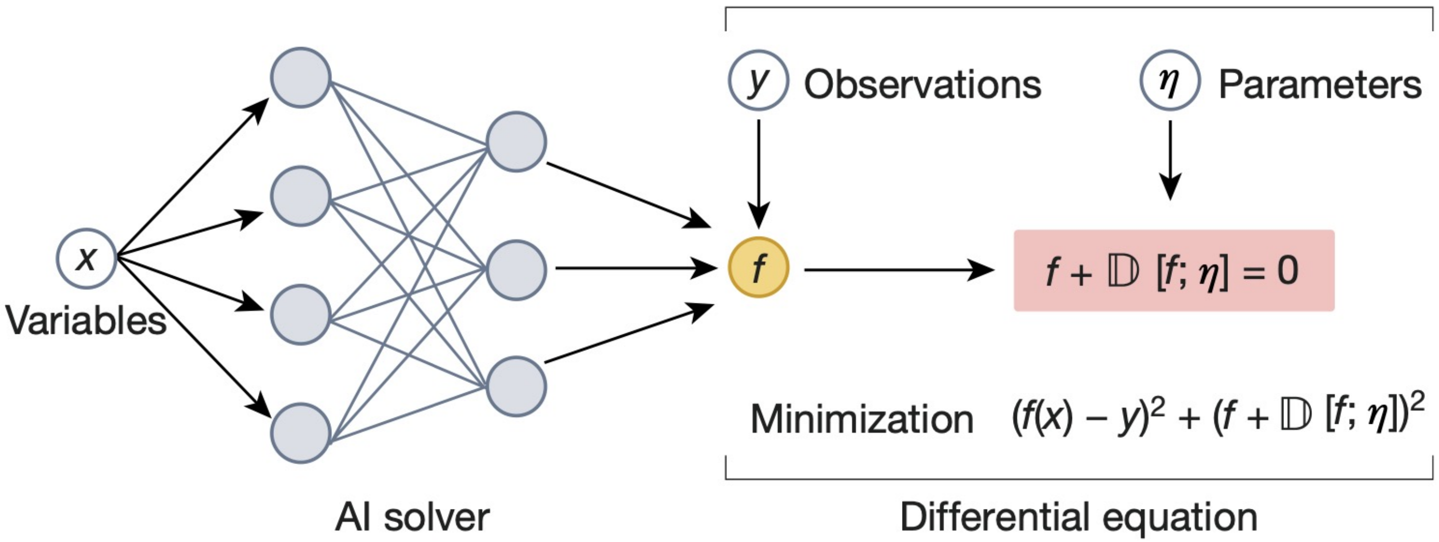
from Emmanuel Menier (PhD, LSIN/SystemX, 2024)

Domain Challenges : Physical systems that are
- Complex to model/solve analytically
- Compuationally expensive to solve numerically

eg. , Computational Fluid Dynamics – CFD, Turbulance, Flows

Scientific Challenges
- Problems highly-nonlinear, high-dimensional, with complex structures (eg. organized in graphs…)
- Need for adapted NN architectures: Graph NNets, Deep AE ..

Hybrid ML modeling for solving Partial Differential Equations



A neural framework for solving PDEs, where

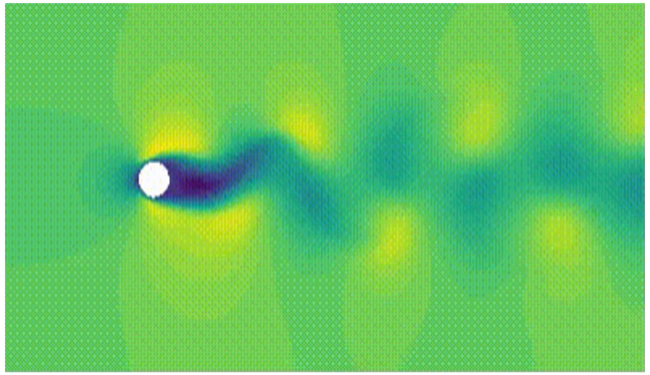
- the AI solver is a PINN trained to estimate target function ***f***.
- The derivative of ***x*** is calculated by automatically differentiating the NN's outputs.
- When the differential equation **D(*f;η*)** is unknown, it can be estimated by solving a loss that optimizes both the functional form of the equation and its fit to observations ***y***.

- Eg. Learning Computational Fluid Dynamics

- Navier-Stokes Equations: fundamental partial differentials equations (**PDE**) that describe the flow of incompressible fluids.

C.L. M. H. Navier, Memoire sur les Lois du Mouvements des Fluides, Mem. de l'Acad. d. Sci.,6, 398 (1822)
C.G. Stokes, On the Theories of the Internal Friction of Fluids in Motion, Trans. Cambridge Phys. Soc., 8, (1845)

- Challenge: High-Dimensional non-linear Physical Equations

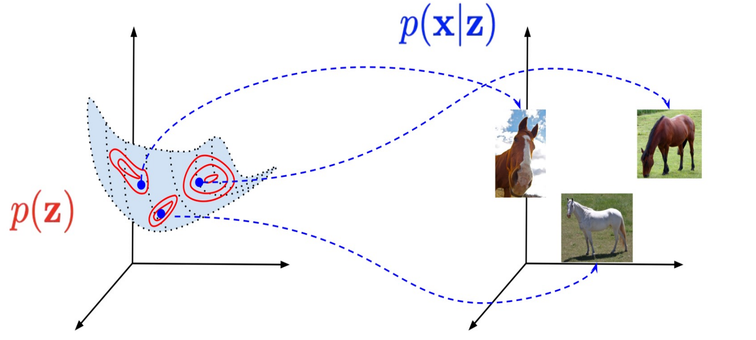


Simulation from Emmanuel Menier

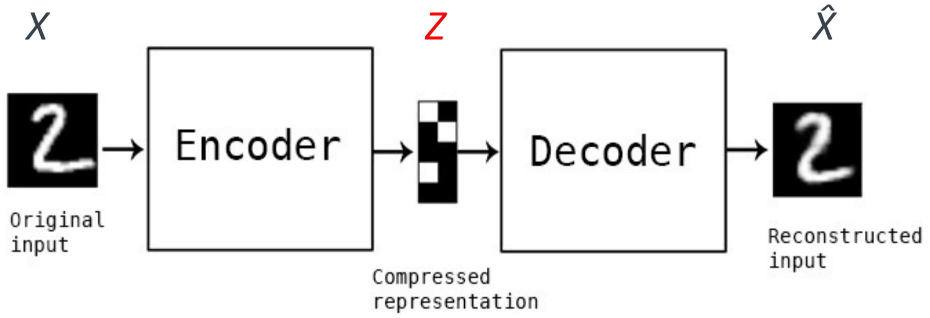
Wang & al. (2023). Scientific discovery in the age of artificial intelligence. Nature, 620. Read Online

Deep NNets for Unsupervised representation Learning

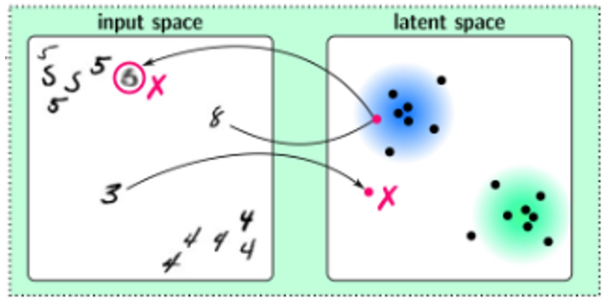
Latent Variable Models: A family of probabilistic models capable of inferring the intrinsic latent structure (of reduced dimension) to the data



- **Auto-Encoders - AE** (LeCun 1987): The **encoder** projects the input ***X*** (of high-dimension dimension) in a compressed latent representation ***Z*** (the **code**) to reconstruct it using the **decoder** with outpu $\hat{X}$
- ➔ Learning by minimizing the reconstruction error between $\hat{X}$ and **X**. The smaller the error, the better the compressed representation ***Z***.



- **Variational Auto-encoders - VAE** (Kingma & Welling 2014) improve the representational capabilities of AEs by regularizing the latent space with a Gaussian priori, coupled with a **variational learning**
- **=>** can learn complex distributions.
- **Deep NNets** are excellent candidates



Deep NNets for Unsupervised representation Learning

- Nnets with a hidden layer are universal approximators
- Nnets are capable to recover highly non-linear relationships in the data
- Adapted architectures that work in a low-dimensional (latent) space

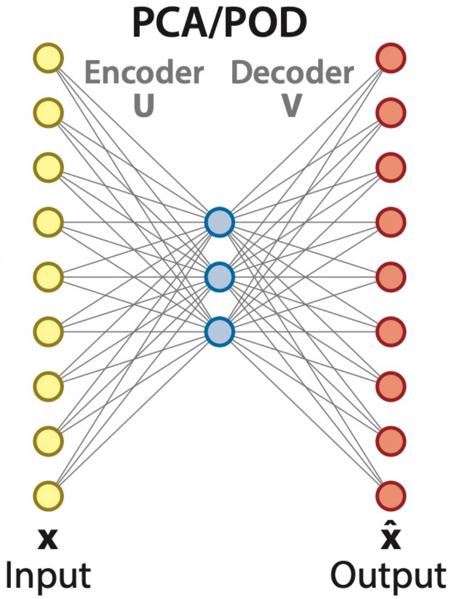
Shallow / Linear

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^{\mathrm{T}}$$

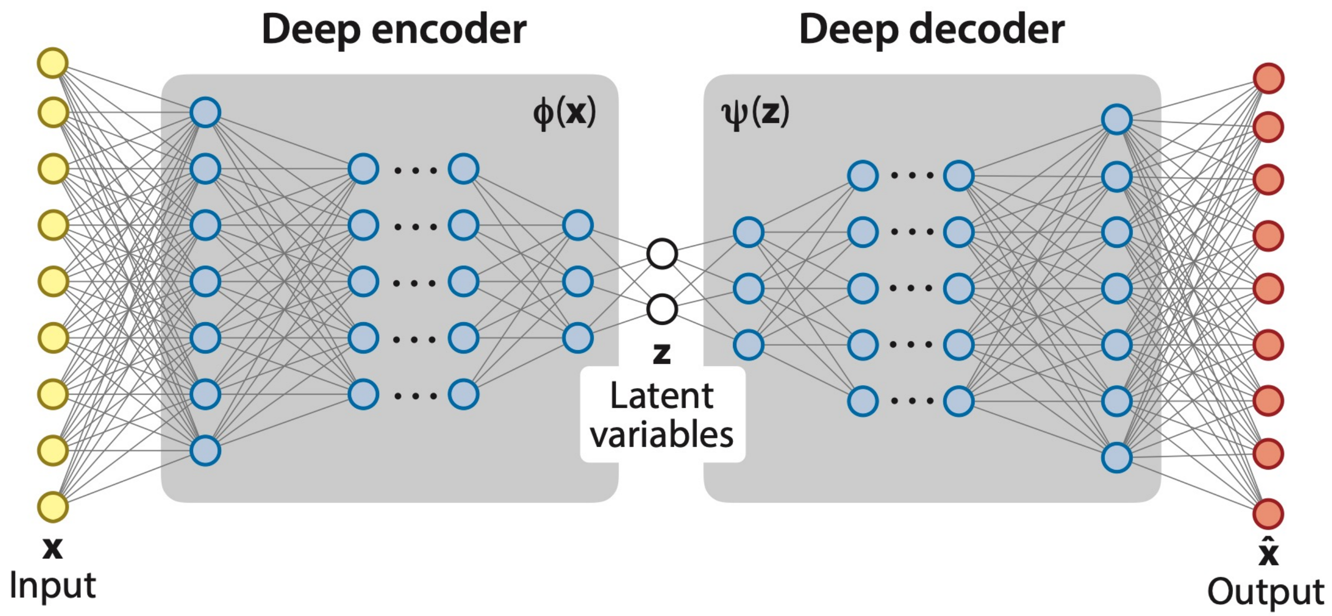
$$\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

Retain $M < D$ eigenvectors



POD: proper orthogonal decomposition
PCA: principal component analysis

Deep / Non-Linear



Auto-Encoding Deep Nets

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual review of fluid mechanics*, *52*, 477-508. Read Online

The Research Program IA2: AI and Augmented Engineering

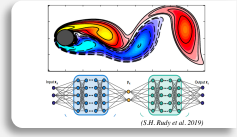
IA² Intelligence artificielle et ingénierie augmentée

Artificial Intelligence
an Augmented Engineering

- a program with **6 R&D collaborative projects** based on concrete **industrial use cases**
- Area: Hybrid AI

Advance project
Thesis / Postdocs / Shared work

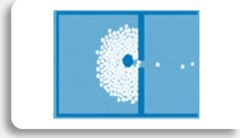
HSA: Simulation/machine learning hybrid modeling
How industrial solvers and learned models can enrich each other ?
01



AFS: Agility and fidelity of simulations
How to imporve agility and fidelity of simulation in complex systems design?
02

S2I: Industrial infrastructure supervision
How to improve decision-making on distubuted industrial systems via machine learning techniques ?
03

SAA: Augmented multi-agent simulation
How can multi-agent models benefit from real data and bring out atypical situations?
04



SMD: Business Semantics for Multi-source Data Mining
How to link heterogeneous data with established practical knowledge?
05

CAB: Cockpit and Bidirectional Assistant
How to develop a virtual assistant that learns from expert and learns the expert
06

Diagram/pictures credit to Mouadh Yagoubi & IA2 Program

HSA Project : Simulation/machine learning hybrid modeling

Challenges and possible solutions (studied as part of the HSA project):

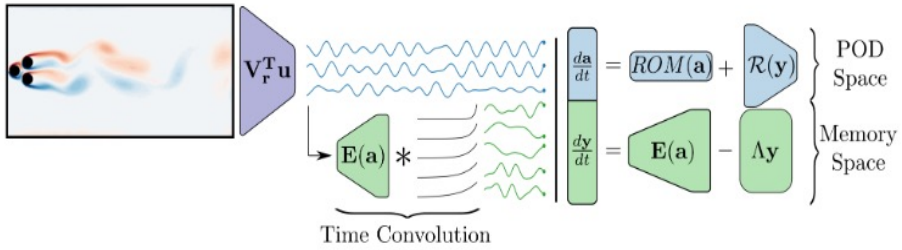
- Augmenting physical solvers with data-driven models that integrate physics constraints
- Building model architecture adapted to the complex physical structures/systems
- Reducing the simulation cost

➔ Hybrid Machine Learninrg as surrogate models for physical simulation, aiming to

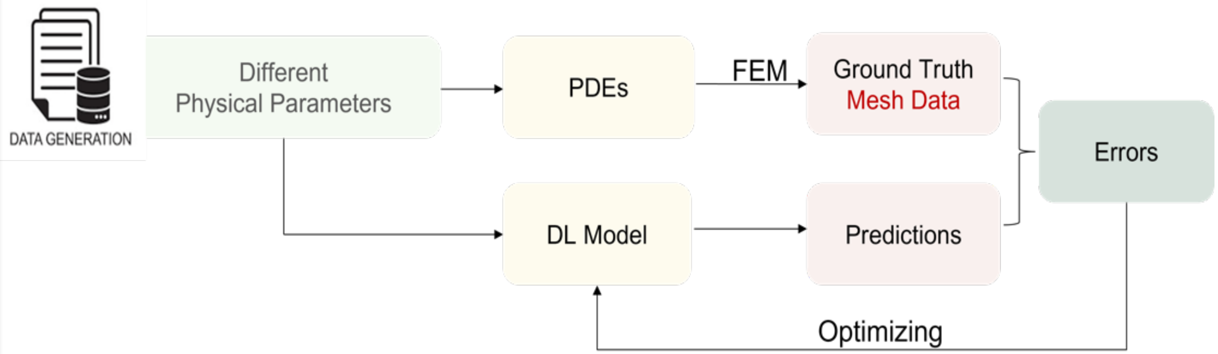
Replace physical solvers with

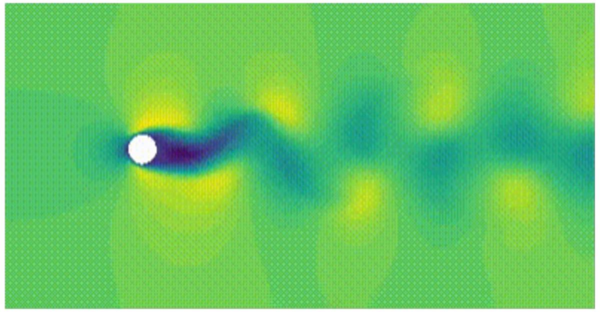
➔ Deep learning intergrating physical constraints (eg. Deep Graph Nets for PDEs)

➔ Deal with high-dimensional, non-linear, and complex structurs (e.g reduced modeling, ..)



High-Dimensional non-linear Physical Equations





Reduced models and deep learning for PDEs
PhD Thesis of E. Menier, 2024 (LISN, Inria/SystemX)

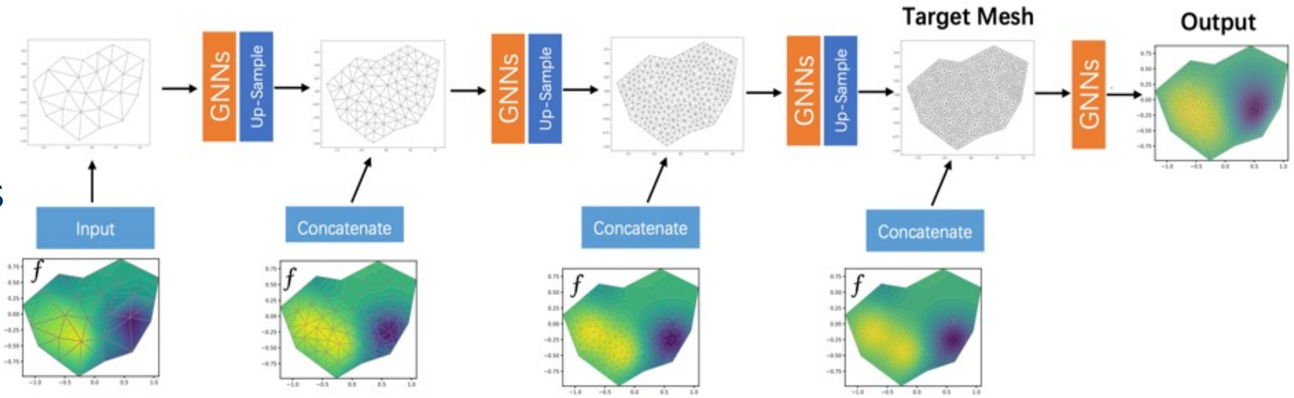
Deep Graph Neural Networks for Numerical Simulation of PDEs. PhD of W. Liu. 2023 (LISN, Inria/SystemX). Read Online

E. Menier et al., 2023. CD-ROM: Complementary Deep-Reduced Order Model. *Computer Methods in Applied Mechanics and Engineering* 410. Read Online

https://www.irt-systemx.fr/projets/HSA/

Project HSA : simulation and deep learning of graphs

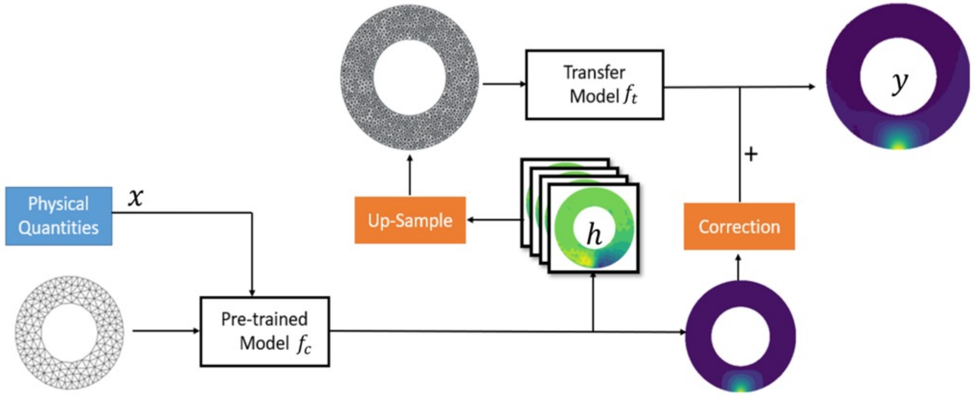
Graph Neural Nets for 3D **meshes**



More suitable, as they operate by construction on graphs

- Generic nature of the learned models
- Transfer learning for improved results
- Prediction can be improved via transfer learning:

from low fidelity (coarse mesh) to high fidelity (finer mesh) models



Wheel contact profile

Physics: contact equations

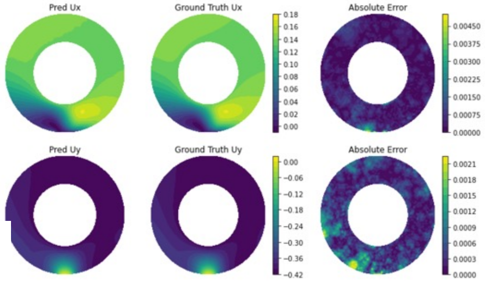
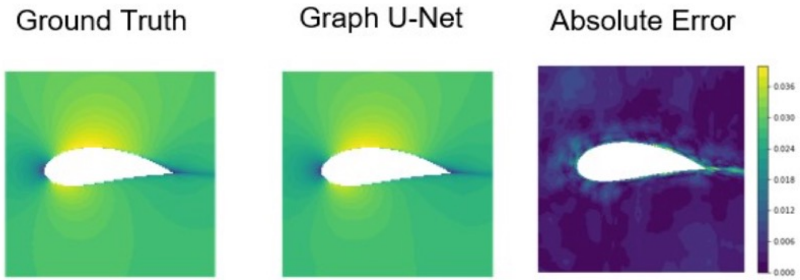


Figure 6.3: An example of wheel contact prediction

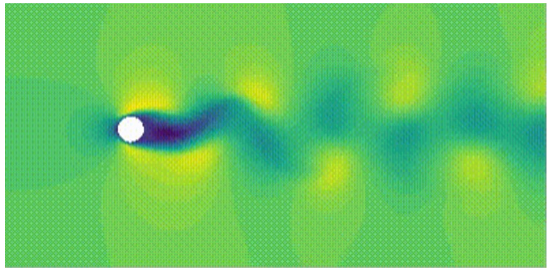
Prediction of the **airflow** profile around an aircraft wing (Air Foil)

Physics: Navier-Stokes equations



PhD theis of W. Liu, 2023 (LISN, Inria/SystemX)

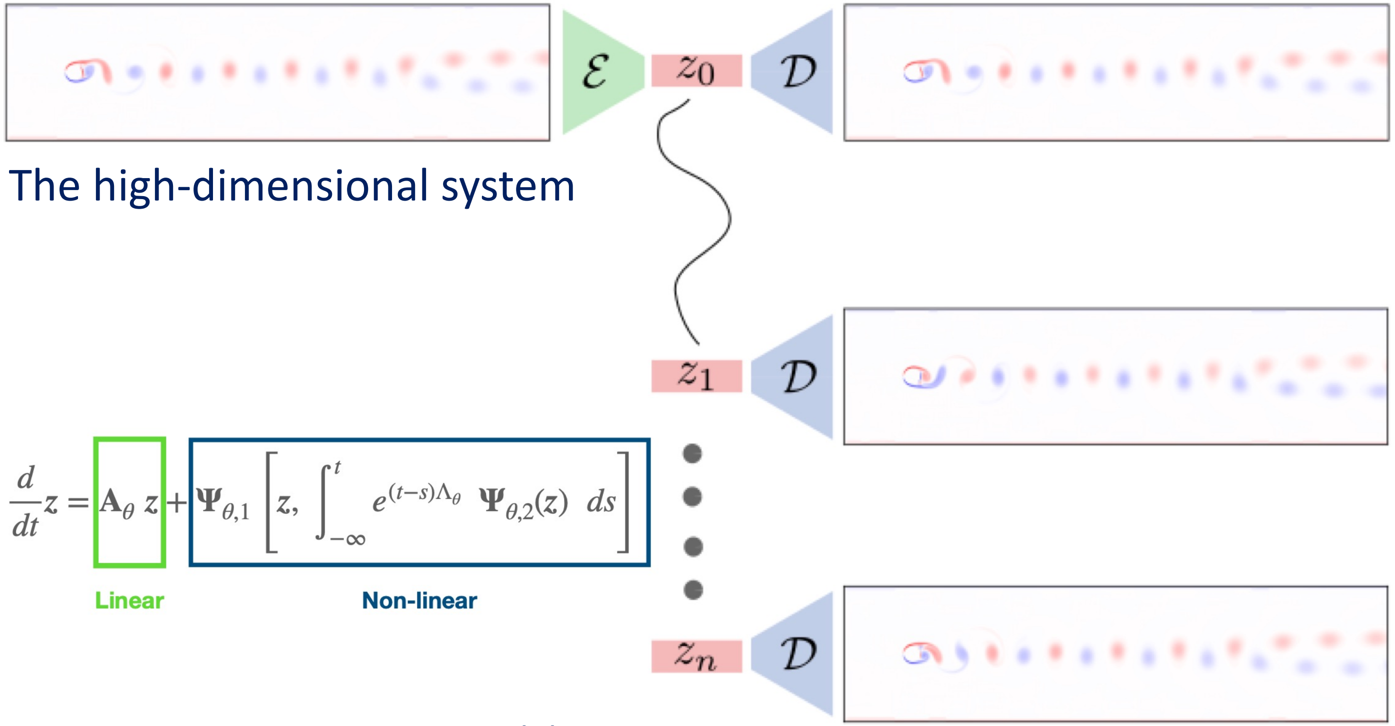
Dynamics: Hybrid ML for HD dynamical physical systems



High-Dimensional non-linear Dymical Systems:

Goals:
Recover the dynamics, non-linearity in a high-dimensitonal setting

Interpretable learning of effective dynamics (ILED) architecture:



The high-dimensional system

$$\frac{d}{dt}z = \underbrace{\mathbf{A}_\theta\, z}_{\text{Linear}} + \underbrace{\mathbf{\Psi}_{\theta,1}\left[z, \int_{-\infty}^{t} e^{(t-s)\Lambda_\theta}\ \mathbf{\Psi}_{\theta,2}(z)\ ds\right]}_{\text{Non-linear}}$$

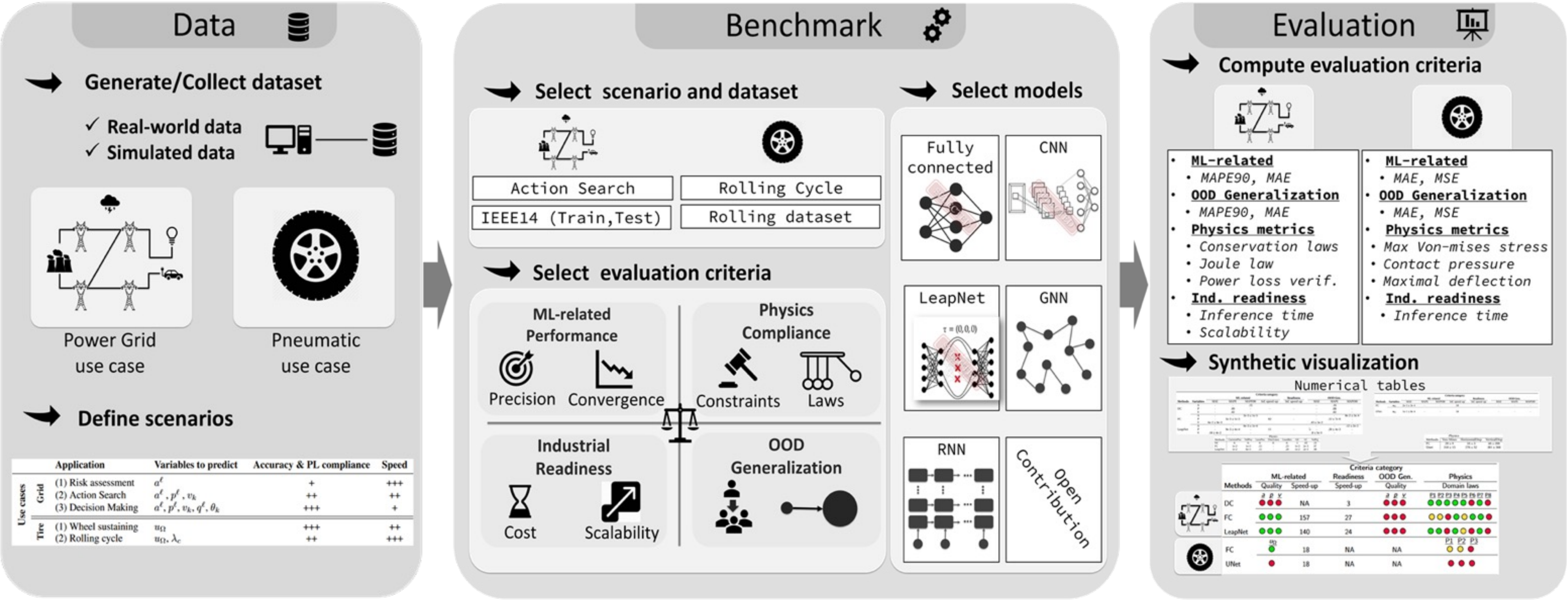
The lower-dimensional representation (**z**) is propagated in time using a linear and a non-linear part based on the Mori-Zwanzig formalism

The decoder D reconstructs the high-dimensional systems.

Menier, E., et al. (2023). Interpretable learning of effective dynamics for multiscale systems. *arXiv preprint arXiv:2309.05812*. Read Online
PhD Thesis of E. Menier, 2024 (LISN, Inria/SystemX), 2024

LIPS: Platform of validation of hybrid AI models

- **LIPS** : Learning Industrial Physical Simulation benchmark suite *(Result of the project HSA-IA2)*
- Evaluation of physical simulator augmented by machine learning
- Open-source Framework https://github.com/IRT-SystemX/LIPS Published at NeurIPS2022
- 1st framework for evaluating augmented physical simulators
- 7 use cases integrated



Competitions on Codabench/codalab

LIPS hosts the two following competitions:

https://www.codabench.org/competitions/1534/ (Closed)

MACHINE LEARNING FOR PHYSICAL SIMULATION CHALLENGE

128 PARTICIPANTS

1165 SUBMISSIONS

€7000 to be shared by the 5 winners (see prizes page)

ORGANIZED BY: Systemx
CURRENT PHASE ENDS: Never
CURRENT SERVER TIME: 8 Mai 2024 À 08:00 UTC+2
Docker image: lipsbenchmark/ml4physim:1.4

Dec 2023 Jan 2024 Feb 2024 Mar 2024

Get Started | Phases | My Submissions | Results | Forum | ?

- About
- Get Started
- Dataset
- SDK & GPU ressources
- Evaluation
- Prizes
- Organizers
- Terms
- Files

New ML4PhySim challenge : The powergrid usecase

Competition Overview

This competition aims at promoting the use of ML based surrogate models to solve physical problems, through a task addressing a recently published dataset called **AirfRANS** related to **airfoil design (CFD simulation)**.

The competition will address the challenge of improving baseline solutions of the Airfoils design use case by building ML-based surrogate models. The overall aim is to improve the tradeoff between the precision of obtained solutions and the related computational cost.

		Criteria category																			
		ML-related (40%)						Application-based context (30%)										Physics (30%)				Score (100%)
	Baseline	Accuracy					Speed-up	OOD Accuracy									Speed-up	Domain laws				
		$\overline{u}_x$	$\overline{u}_y$	$\overline{p}$	$\overline{\nu}_t$	$\overline{p}_s$		$\overline{u}_x$	$\overline{u}_y$	$\overline{p}$	$\overline{\nu}_t$	$\overline{p}_s$	C_D	C_L	ρ_D	ρ_L		C_D	C_L	ρ_D	ρ_L	
AirFoil	GraphSAGE						1000										1000					55.87
	FC						1300										1300					44.57
	OpenFOAM						1										1					82.5

https://www.codabench.org/competitions/2378/ **Running!**

MACHINE LEARNING FOR PHYSICAL SIMULATION CHALLENGE - POWERGRID USE CASE

31 PARTICIPANTS

0 SUBMISSIONS

ORGANIZED BY: Systemx
CURRENT PHASE ENDS: 14 Mai 2024 À 02:00 UTC+2
CURRENT SERVER TIME: 8 Mai 2024 À 08:04 UTC+2
Docker image: codalab/codalab-legacy:py37

May 2024 Jun 2024 Jul 2024

Get Started | Phases | My Submissions | Results | Forum | ?

- About
- Starting kit
- Evaluation
- Prizes
- SDK & GPU ressources
- Organizers
- Terms
- Files

Prizes

General prizes:

- **1st Prize** : 3000 €
- **2nd Prize** : 2000 €
- **3rd Prize** : 1000 €

Special prizes:

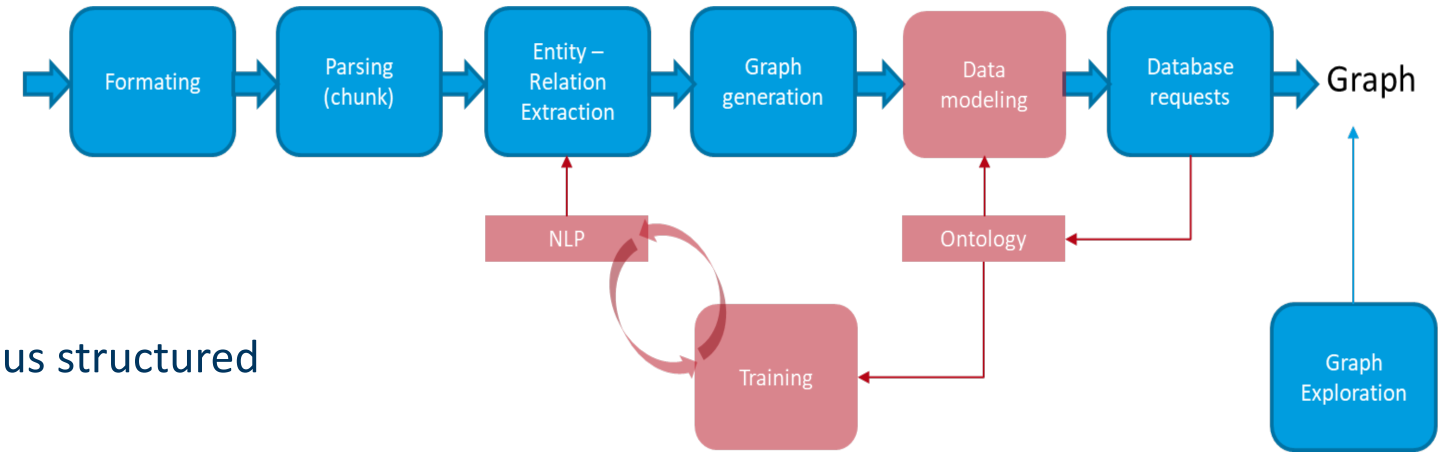
- **Most accurate ML model (without speedup consideration)** : 1000 €
- **Best student solution** : 1000 €

The general and special prizes are not cumulative. Winning one of the general prizes hinder the access to special prizes.

Project SMD : Integrating Expert/Business *semantics* in ML

https://www.irt-systemx.fr/projets/SMD/

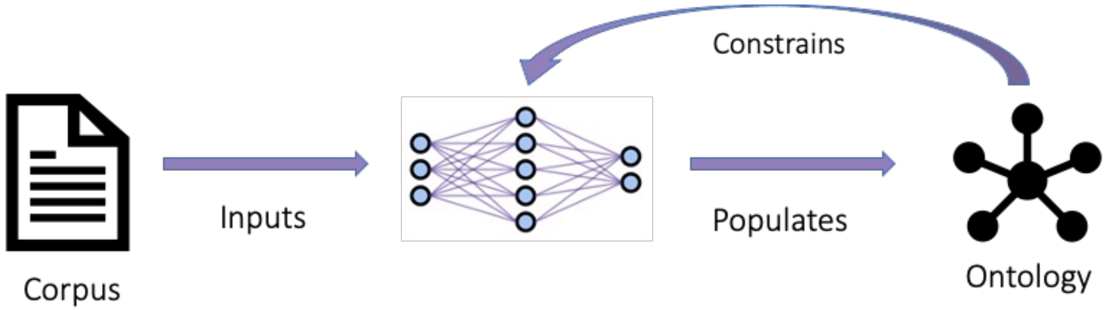
Framework:
Building ML-based on knowledge graphs from expert/business language data



Neuro-symbolic pipeline

Approach

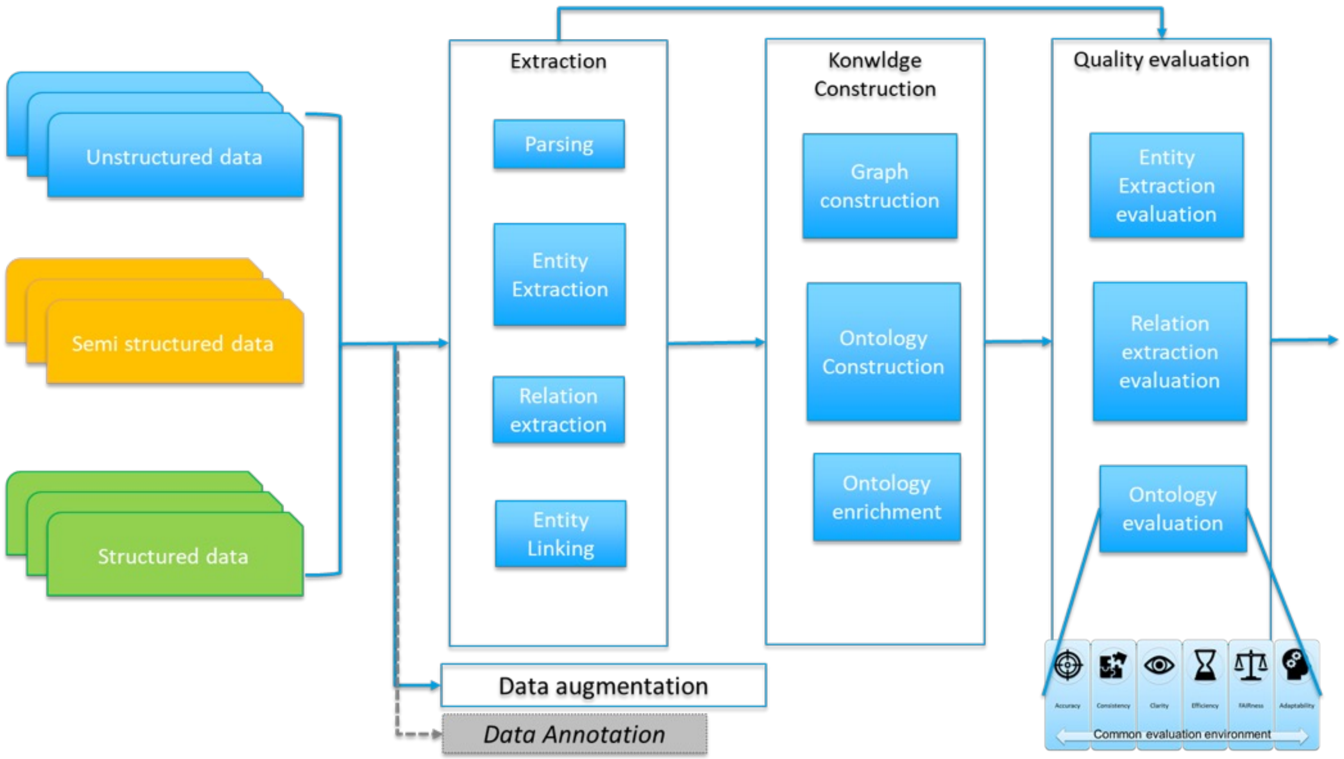
- Information extraction from heterogeneous structured and semi-structured text corpora...:
- NLP approach; Semantic annotation
- Taking into account domain rules/constraints in the numerical-AI based decision

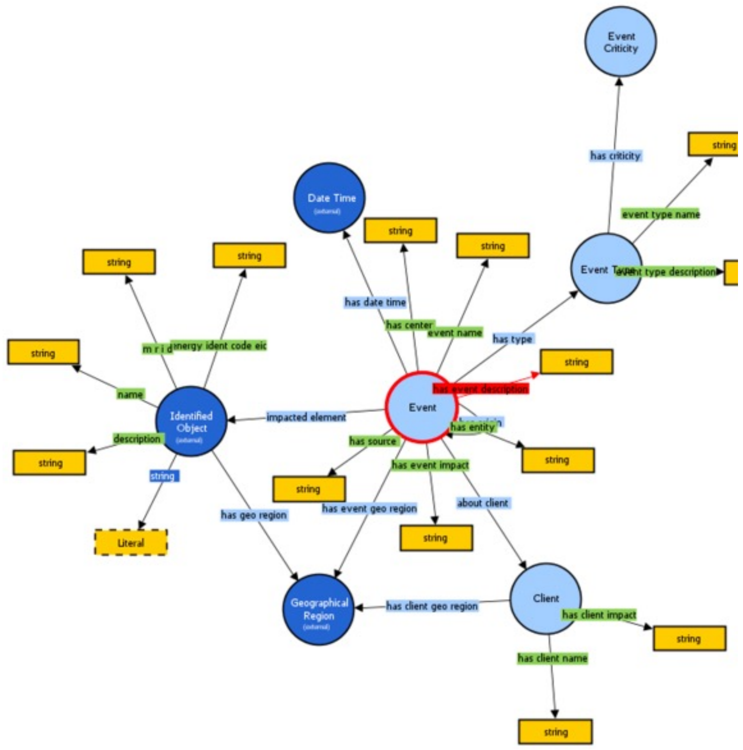


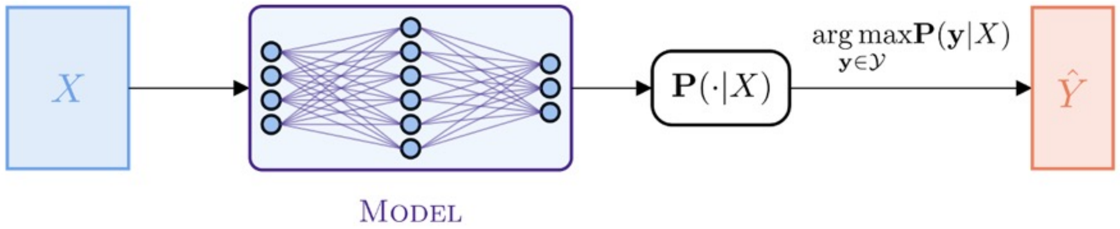
Intelligence artificielle et ingénierie augmentée

SMD Project

Project SMD : Integrating Expert/Business semantics in ML







=> Insertion of logical rules at the inference step :

- Recalculer P(y|X, α) in lieu of the learnt P(y|X)
- α is a rule encoding the validity of the prediction $\hat{y}$

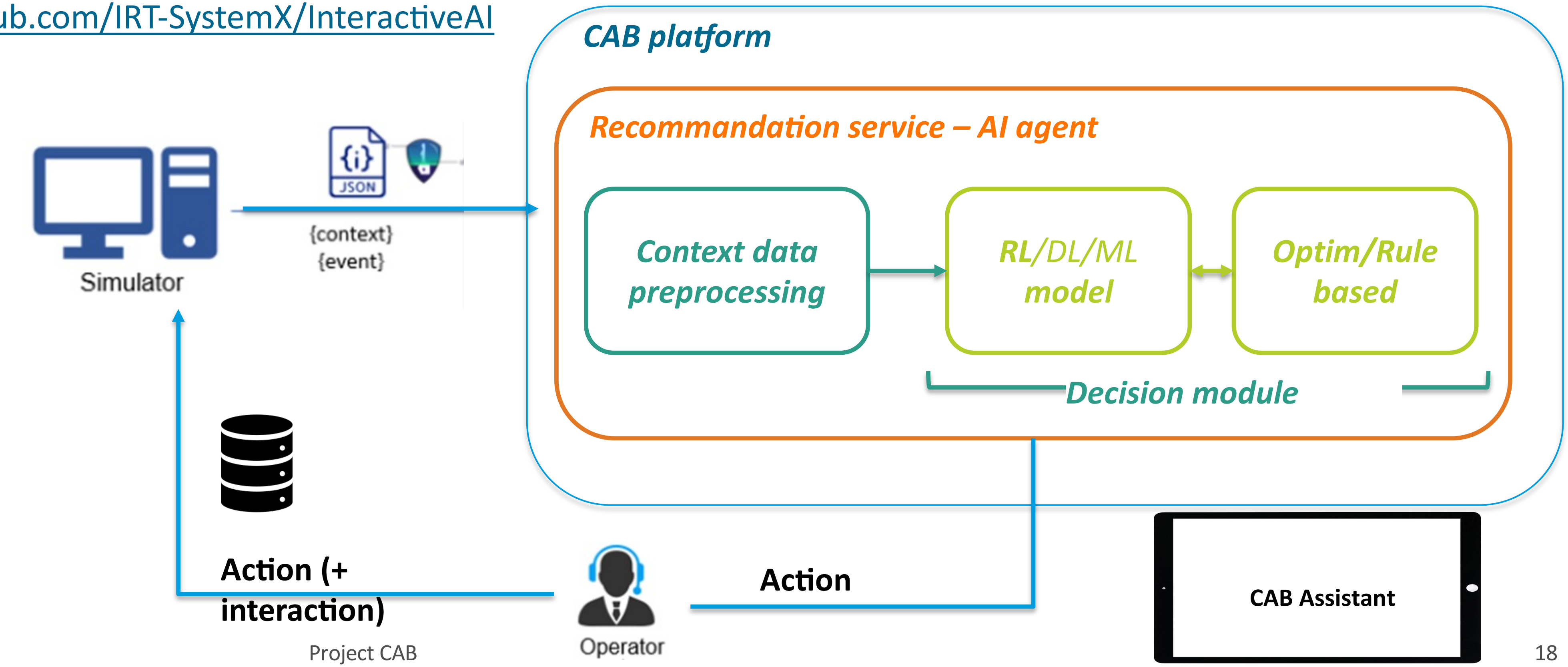
Ontologies and (machine/transfer) learning for multimedia document analysis. **PhD thesis of A. Ledaguenel** (in progress, MICS/SystemX)

Human – AI Interaction: CAB Project (Bi-directional assistant)

Objectives and challenges

- Design and implement a Bidirectional Assistant to support operators in network supervision and aircraft piloting activities
- Bidirectional assistant: The assistant can learn from and to (inform) the operator
- **Platform**: https://github.com/IRT-SystemX/InteractiveAI

Project IA² CAB

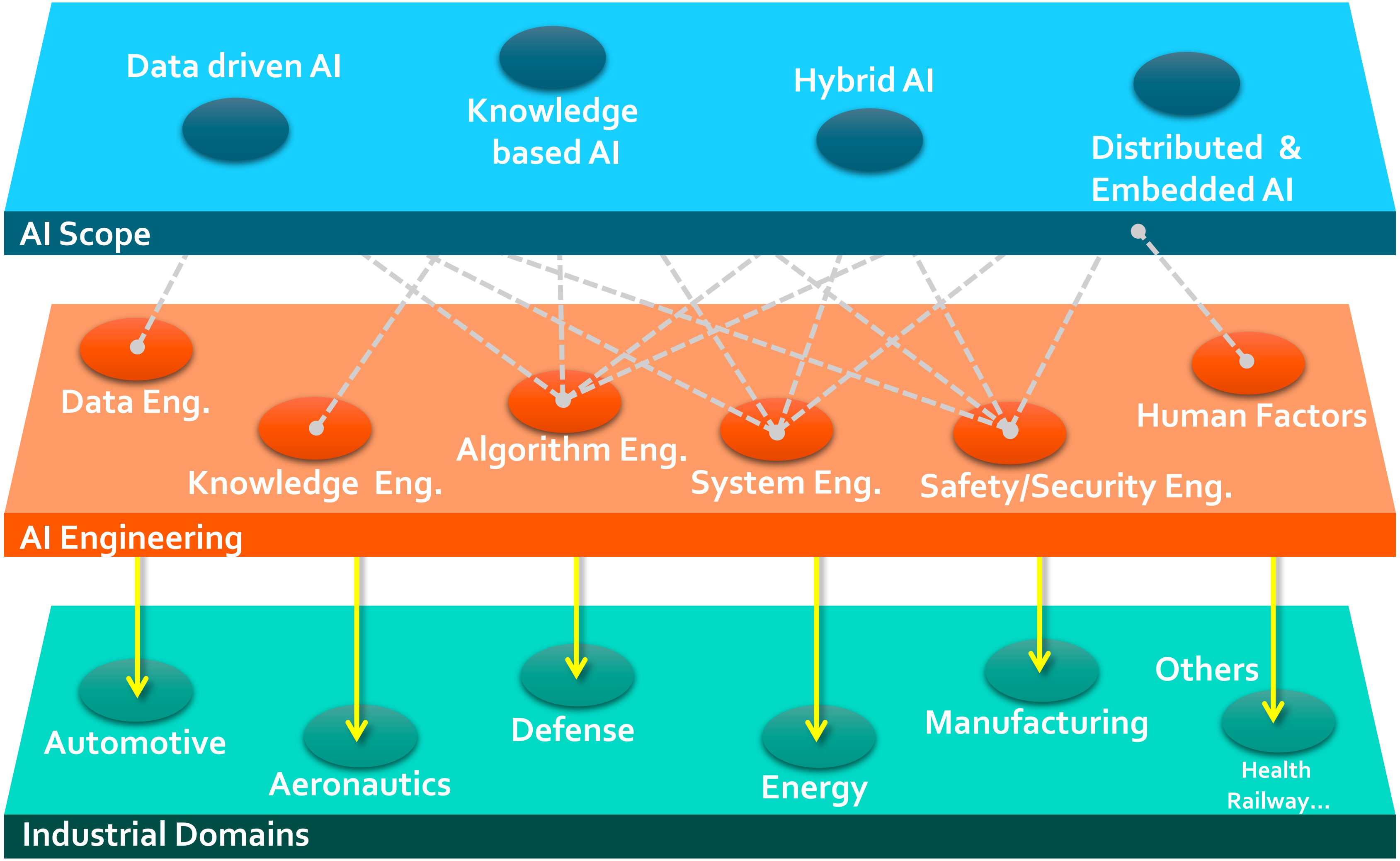


Challenges and industrial applications in Trustworthy AI: Confiance.AI programme

A French unique community to design and industrialise trustworthy AI-based critical systems

Multi-technology, multi-domain, multi-engineering

AIRBUS
Air Liquide
Atos
NAVAL GROUP
RENAULT
SAFRAN
sopra steria
THALES Building a future we can all trust
Valeo



cea
cnrs
Inria
IRT Saint Exupery
SystemX

Key figures

4 YEARS
Duration

9
Large industrial groups

4
Research centers

7
Thematic projects

+300
FTP involved over 4 years

2
Sites: Paris-Saclay and Toulouse

45 M€
Budget

12
Associate partners (laboratories, SMIs, startups)

Scientific challenges of confiance.ai

- **confiance.ai:** methods and tools for trusted AI.
- High expectations for industry
- In parallel of development of tool chain, many scientific challenges remain:
- **3 groups** of **scientific challenges** to cover all aspects of trust

  1) **Trust and learning data**
  2) **Trust and human interaction**
  3) **Trust and AI-based system engineering**

- Organization in **7 projects:**
  - EC1: Integration & **use cases**
  - EC2: Process, **methodology** and guidelines
  - EC3: Characterization & **qualification** of trustworthy AI
  - EC4: **Design** for Trustworthy AI
  - EC5: Data, information and knowledge **engineering** for trusted AI
  - EC6: **IVV&Q** strategy toward homologation / **certification**
  - EC7: Target **Embed**ded AI

PhD theses within confiance.ai

Ongoing PhD theses within confiance.ai about:

- PhD Theis of Adrien Le Coz: Data coverage and operational **domain design** ODD (Computer Vision)
- PhD Thesis of Paul La Barbarie: **Robustness** to ‘patch’ adversarial attacks (Computer Vision)
- PhD Thesis of Lucas Schott: **Reinforcement** learning and human in the loop
- PhD Thesis of Gayane Taturyan: Statistical control of **Fairness**/Bias in Machine Learning (Stat)
- PhD Theis of Houssem Ouertatani : **Optimized** hardware **deployment** (Neural Architecture Search)
- PhD Thesis of Abdelmouaiz Tebjou: Conformal prediction (deployed) AI algorithms **monitoring**

Other topics

- Generative AI (Evaluation, Benchmarking, Multimodalities..)
- Hybrid AI (GenAI, Augmentation, Trust, UQ,..)
- Trustworthy AI (Hybrid AI, Implementation of the AI act, LLMs …)
- Federating Learning …

Some Perspectives

“HORIZON-CL4-2024-**HUMAN-03-01**: **Advancing Large AI Models**: Integration of New Data Modalities and Expansion of Capabilities (AI, Data and Robotics Partnership) (RIA)

Expected Outcome: Projects are expected to contribute to one or more of the following outcomes:

- Enhanced applicability of large AI systems to new domains through the integration of innovative data modalities, such as sensor measurements (e.g. in robotics, IoT) or remote sensing (e.g. earth observation), as input.
- Improvement of current multimodal large AI systems capabilities and expansion of the number of data modalities jointly handed by one AI system, leading to broader application potential and improved AI performance.

Scope: Large artificial intelligence (AI) models refer to a new generation of general-purpose AI models (i.e., generative AI) capable of adapting to diverse domains and tasks without significant modification. Notable examples, such as OpenAI's GPT-4V and META’s Llama 2 or DinoV2, have demonstrated a wide and growing variety of capabilities.”

Some Perspectives

"HORIZON-CL4-2024-**HUMAN-03-02**: **Explainable and Robust AI** (AI Data and Robotics Partnership) (RIA)

Expected Outcome: Projects are expected to contribute to one of the following outcomes:

- Enhanced robustness, performance and reliability of AI systems, including generative AI models, with awareness of the limits of operational robustness of the system.
- Improved explainability and accountability, transparency and autonomy of AI systems, including generative AI models, along with an awareness of the working conditions of the system.

Scope: Trustworthy AI solutions, need to be robust, safe and reliable when operating in real-world conditions, and need to be able to provide adequate, meaningful and complete explanations when relevant, or insights into causality, account for concerns about fairness, be robust when dealing with such issues in real world conditions, while aligned with rights and obligations around the use of AI systems in Europe. Advances across these areas can help create human-centric AI, which reflects the needs and values of European citizens and contribute to an effective governance of AI technologies ."

SystemX INSTITUT DE RECHERCHE TECHNOLOGIQUE

Accélérateur
de la transformation
numérique

FRANCE 2030

THANK YOU FOR YOUR ATTENTION