RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030 PROGRAMME DE RECHERCHE NUMÉRIQUE POUR L'EXASCALE

Optimization and AI

WP2 & WP5

Presented by

Prof. El-Ghazali Talbi & Emmanuel Franck

Sommaire

1. Optimization for AI

1. AI and ML: AutoML

2. Optimization for DNN

1. Optimization problems
2. AI for optimization
3. Parallel algorithmic solutions

3. Optimization for LLM

1. Optimization problems
2. AI for optimization
3. Parallel algorithmic solution

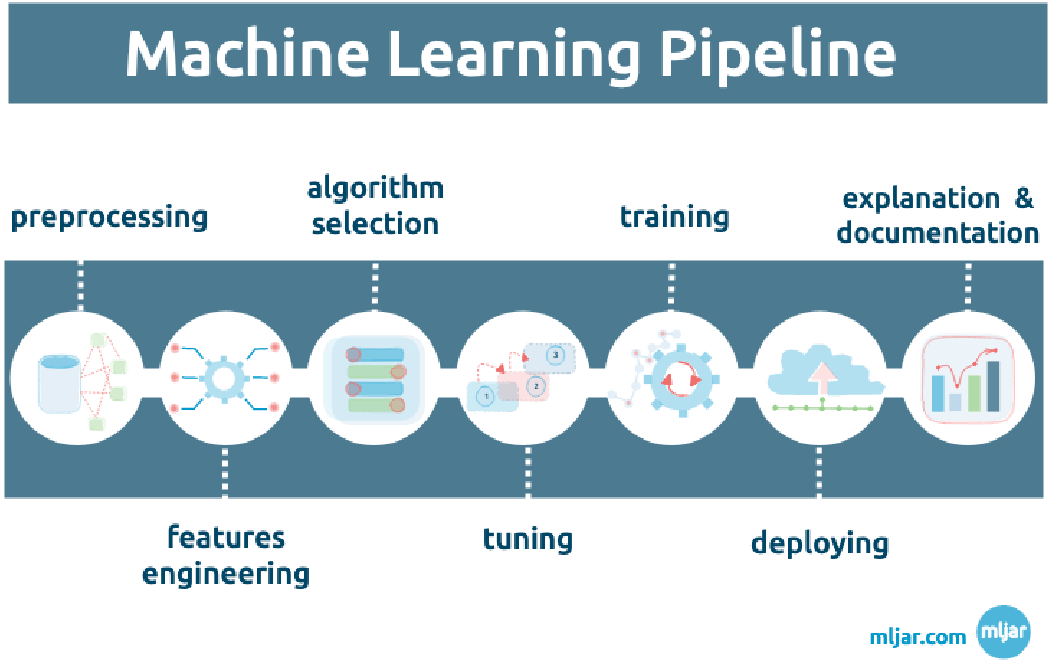
1. Scientific challenges involving both WP2 and WP5
2. Opportunities for the Numpex call HPC and AI.
   - Clusters IA

1. Optimization for AI

AutoML

Sous-titre

- **Machine learning tasks**
  - Supervised learning
  - Unsupervised learning
  - Feature selection
  - Reinforcement learning

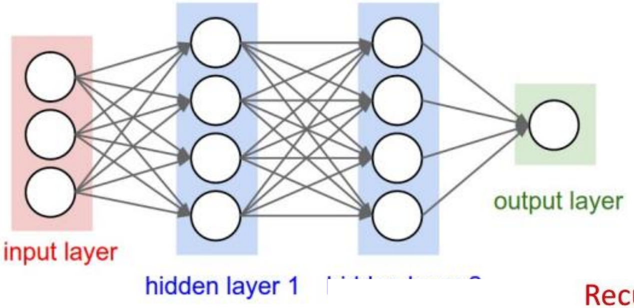


RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

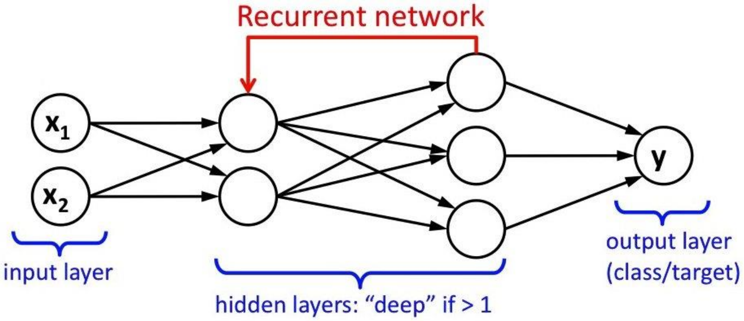
FRANCE 2030 PROGRAMME DE RECHERCHE NUMÉRIQUE POUR L'EXASCALE

2. Optimization and deep neural networks (DNNs)

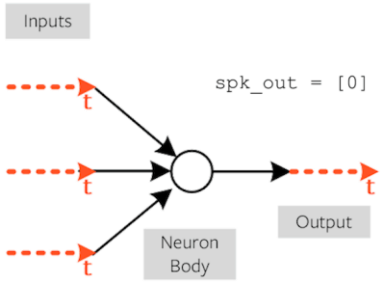
Deep neural networks



- Feed Forward
  - CNN (Convolution Neural Networks)
  - AE (Auto Encoders)
  - Transformers (Attention)
  - GNN (Graph Neural Networks)
  - Generative Adversarial Networks (Game theory)



- Recurrent neural networks (RNN)
  - LSTM (Long Short-Term Memory networks)
  - GRU (Gated Recurrent Units)
  - LLM (Large Langage Models)



- Neuromorphic networks
  - Collaboration with PEPR IA (EMERGENCES project)

Optimization Problems

- Neural architectures search (NAS)
  - Search the optimal DNN topology (e.g., number of layers, types of operations, connections between operations)
  - Hyperparameters are supposed to be a priori fixed
- Hyperparameter optimization (HPO)
  - Requires an *a priori* definition of the DNN architecture
  - Optimize the hyperparameters of the DNN
  - Two types of hyperparameters
    - Operation hyperparameters: features associated to operations
    - Global hyperparameters: optimization features of DNN
- Joint optimization (NAS+HPO)

1. Global optimization: optimizing all levels at the same time
2. Nested optimization: optimizing the different levels in a hierarchical way.
3. Sequential optimization: NAS problem is solved first. Then, the hyperparameters for the obtained final solution are optimized.

E-G. Talbi, Automated design of deep neural networks, ACM Computing Surveys, 2022.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030 PROGRAMME DE RECHERCHE NUMÉRIQUE POUR L'EXASCALE

Characteristics of the Optimization Problems

- Large-scale optimization problem
  - High number of decision variables.
- Mixed optimization problem
  - Continuous: learning rate, momentum, …
  - Discrete ordinal (i.e., quantitative): size of the filter, stride in CNN pooling operations
  - Discrete categorical (i.e., qualitative): type of operations, training optimizer
- Variable-size design space
  - Search space varies dynamically as a function of some variables values
  - Decision variable is relevant only if another variable takes a certain value.
- Extremely expensive black-box objective function(s)
  - Training the whole DNN (e.g. loss function).
  - Might take several hours, days or even months
- Noisy objective function
- Multi-objective optimization problem
  - Various conflicting objectives

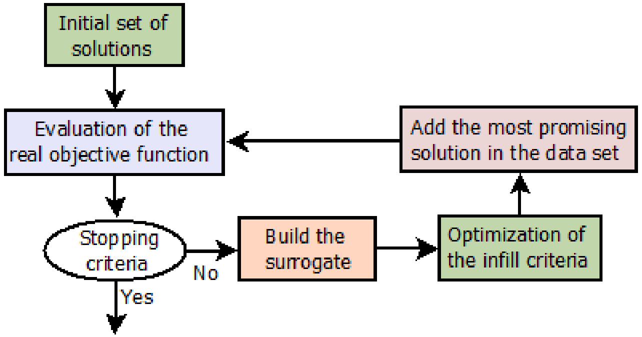
Ouertatani, H., Maxim, C., Niar, S., & Talbi, E. G. (2024, September). Accelerated NAS via pretrained ensembles and multi-fidelity Bayesian Optimization. *Int. Conf. on Artificial Neural Networks ICANN'*2024

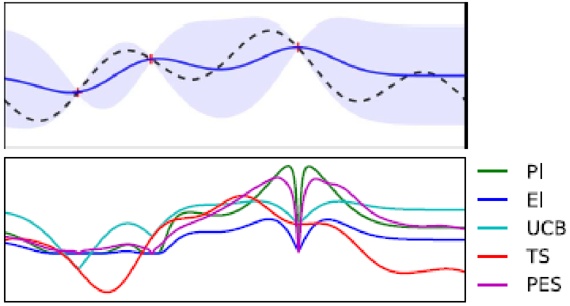
Multiple objectives

- Energy consumption
  - Low-power mobile and embedded areas → Energy consumption (i.e. power)
- Inference speed
  - Real-time applications (e.g. video analysis)
- Computational and memory cost
  - Number of floating-point operations (FLOPs), Memory usage
  - Can concern both training and inference
- Hardware cost
  - Hardware for training and/or inference
- Number of parameters
- Diversity
  - Ensemble models using diverse DNNs tends to achieve better generalization
  - Diversity measures the discrepancy between the output of a DNN and the outputs of other DNNs

AI for Optimization

- Bayesian optimization & Surrogate optimization
  - Multi-fidelity models
  - Coupling of surrogates, optimization and sampling
- Construction of surrogates (i.e. reduced models)
  - Deep neural networks
    - PINNs, …
    - Opérateurs neuronaux pour EDP (Transformers, LLM, …)
    - Composition de réseaux, …
    - Problématique optimisation (hyperparametres, entrainemeent, …) de ces grands réseaux
- WP2 –WP5 (2 demi-thèses)
  - Une thèse dans chaque WP
  - Collaboration à travers les doctorants, un ingénieur, …?





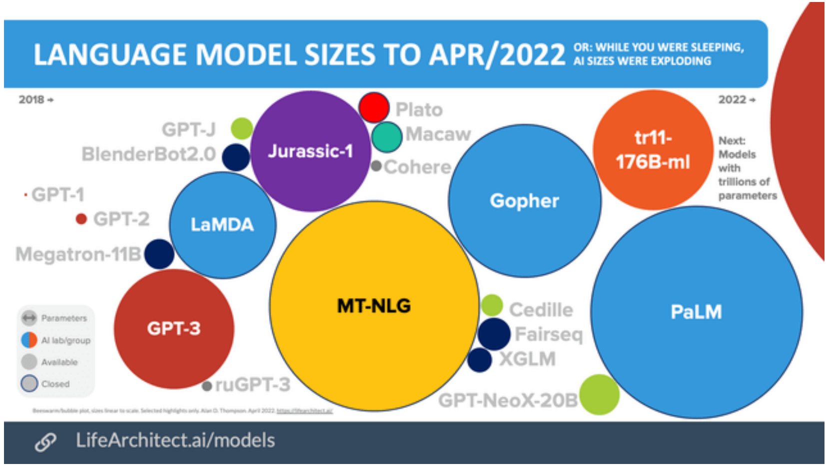
Optimization algorithms

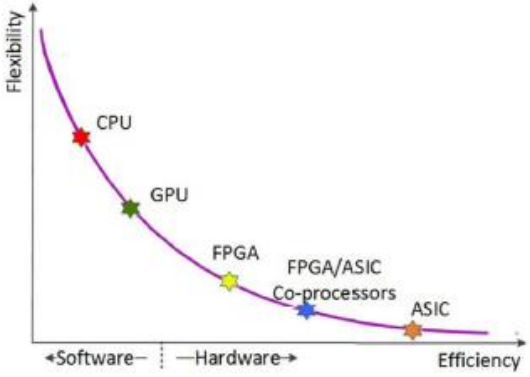
- A wide variety of algorithms have been used
  - Grid search
  - Monte Carlo Tree Search (MCST)
  - Reinforcement learning (RL)
  - Bayesian Optimization
  - METAHEURSTICS
    - Local-search based (eg. Gradient based)
    - Evolutionary algorithms
    - Swarm Intelligence

J. Keisler, E-G. Talbi, A framework for the optimization of deep neural networksarchitectures and hyperparameters
, JMLR Journal of Machine Learning Research, 2024.

Parallel Optimization Algorithms

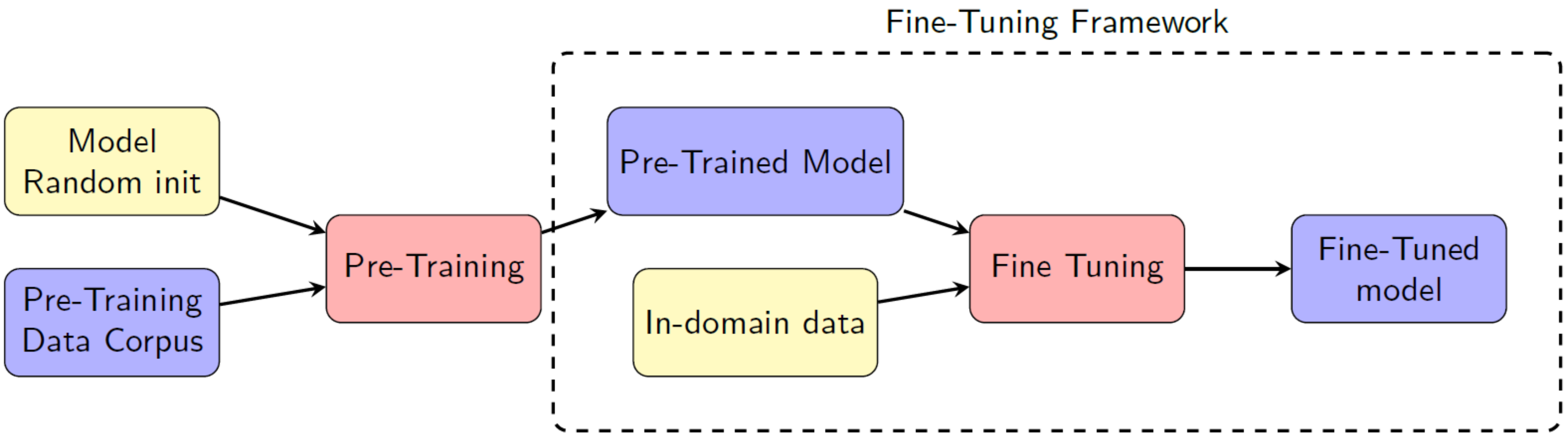
- AutoDNN problems are more and more complex (Generative AI)
  - Dataset, network size
  - LLM: Billions of parameters
  - GPT-4 Trillion parameters
- Rapid development of hardware
  - CPU, GPU, FPGA, ASICS, ...
  - State-of-the-art DNNs required more than 2,000 GPU days.
- Parallel algorithm design
  - Decision space decomposition → Most scalable for Exascle
  - Neighborhood exploration
  - Parallel handling of the population of solutions
  - Parallel handling of the objective function
- Hardware-aware NAS
  - Configuration of hardware : GPUs, ...





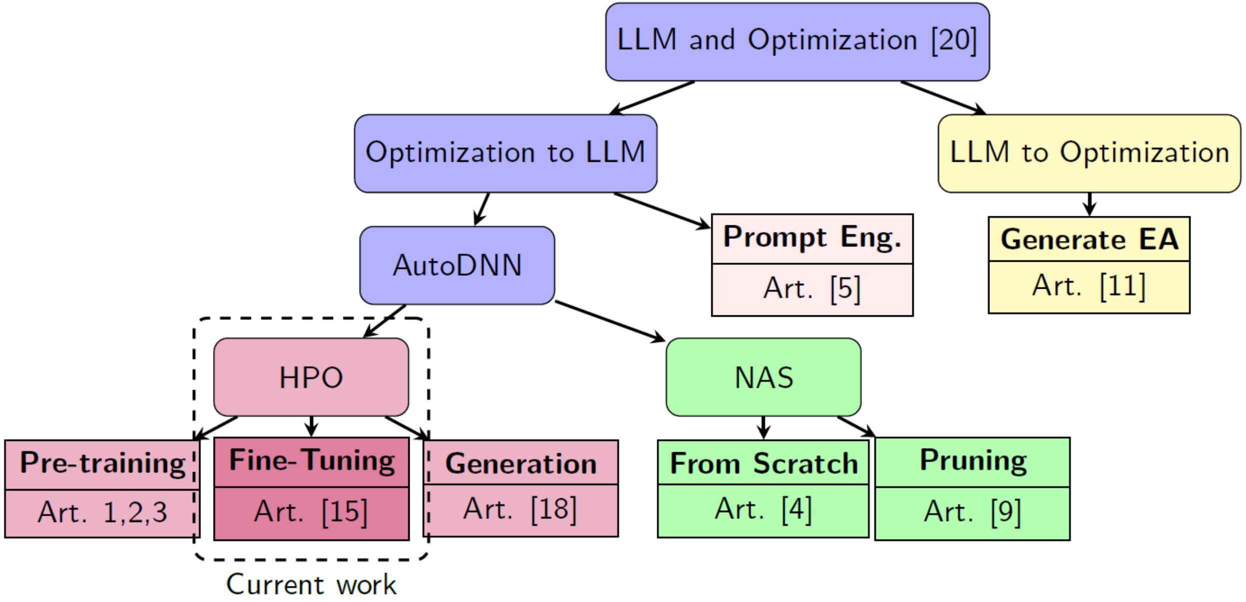
3. Optimization and Large Langage Models (LLMs)

Optimization Problems



Optimization Problems

•



N. Davouse, E-G. Talbi, LLM fine tuning using Bayesian optimization, OLA'2025 Optimization & Learning Conference

Parallel Optimization Algorithms

- Decomposition algorithms
  - Fractal decomposition: DIRECT, FRACTAL, SOO, …
  - Massively parallel → Towards Exascale
  - Costly
- Bayesian optimization
  - Efficient
  - Intrinsically sequential
- Complementarity between decomposition and Bayesian optimization