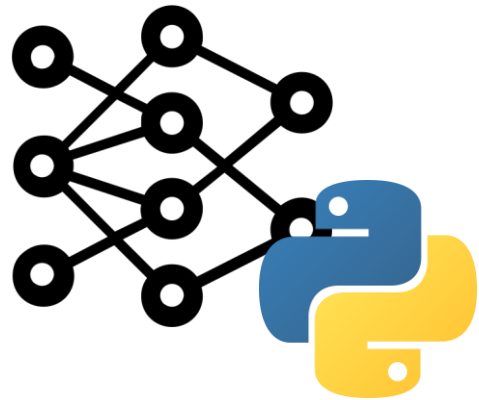
Adaptive DOE Sampling for Machine Learning suitable dataset generation

- Supporting companies: **GammaTech Engineering**



- Start date: September/October 2026
- Project duration: 6÷8 months
- Site:
  - GTE's offices in Turin
- Compensations: Meals, travel expenses
- Motivations and Project Scope:
  - Machine Learning surrogates of physical models generalize only if their input features are **as uniformly distributed as possible** over the operating domain: clustered or poorly covered regions give biased predictions
  - In datasets generated from physics-based models, the DOE variables swept in the physical model (boundary conditions, control settings, operating points) are often **not the neural network input features**, which are non-linear coupled outputs of the simulation: a space-filling DOE can then yield strongly non-uniform features
- Thesis proposal:
  - Formalize a methodology to solve the feature-space filling issue by designing a **DOE automatic sampling strategy**, structured in two interconnected steps:
    - **Step 1:** Formalize a DOE-to-feature mapping function and identify the most appropriate **feature-space coverage metrics** to quantify uniformity and detect sparsely covered operating conditions
    - **Step 2:** Develop an **adaptive sampling loop**: initial space-filling DOE, physical model runs, surrogate estimation of the DOE-to-feature map, selection of the next DOE points filling the sparsest feature regions, iterating until the target uniformity metric is satisfied
  - The physical model/use case adopted as case study can be agreed with the candidate