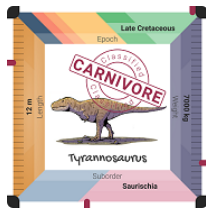# AI TOOLS | Impact of glass-boxing the Machine Learning pipeline on learning

MOBOTS Lab - Prof. Francesco Mondada

Project Type: Semester project for master students

Section: IC or STI

Assistants: Valentina Ferraioli   (Valentina.ferraioli@epfl.ch)

Context:

Machine learning (ML) becomes increasingly prevalent in everyday technology, understanding its foundational principles is essential for future generations. However, ML systems—often described as "black boxes"—can be challenging to comprehend, presenting a unique challenge when introducing these concepts to novices or students.

ML can be deconstructed into several pipeline stages: Data Collection, Data Cleaning, Feature Selection, Model Training, and Evaluation. While recent research indicates that exposing students to data preparation helps their understanding of AI (Hitron et al., 2019), few tools have systematically examined whether giving access to the model logic improves understanding. Currently, tools often approach model learning merely by executing a pre-defined model without encouraging further interaction (Marques et al., 2020).

The rationale for this project is to test the concept of "Pipeline Transparency." by evaluating the impact of Data preparation and Model Logic on their understanding. Does uncovering data preparation steps bring more value than uncovering the model logic? To investigate this we use decision trees (DT) as they are intuitive for students.

Project Description:

The goal of this project is to develop a custom web-based platform, and use it to conduct a study comparing three levels of transparency. The project builds on an existing unplugged version (paper based) of this experiment that needs to be adapted to the digital environment and on previous semester projects that have implemented a website for basic interaction with a DT.The student will implement a single web application (e.g., using React/Vue and Python/Node) that challenges users to predict a dinosaur's diet ("Herbivore" vs. "Carnivore") based on field notes. The core task is implementing three distinct experimental modes to isolate the learning variables:

- ***C1 (Glass Box):***
  - ***Data***: *Users must manually Collect and Clean Data from PDF field notes (handling errors/ambiguity).*
  - ***Model***: *Users must Construct Logic manually using a drag-and-drop Decision Tree UI with impurity feedback (Gini/Entropy).*
- ***C2 (Gray Box):***
  - ***Data***: *Users must manually Collect and Clean Data (same as C1).*
  - ***Model***: *The logic is Automated. Users perform feature selection, but a "Train AI" button handles the tree building.*
- ***C3 (Black Box):***
  - ***Data***: *Users receive Pre-loaded Data (read-only). They interact with the data purely as consumers.*
  - ***Model***: *The logic is Automated also here they are just consumers. Users simply click "Train."*