6-Month MSc 2 Internship Proposal

# Minimal Reinforcement Learning Building Blocks: from Reactive to Learning Agents

**Supervisor:** Alexandre Muzy, Director of Research at CNRS, Email : alexandre.muzy@cnrs.fr.

**Level:** MSc 2. **Duration:** 5 to 6 months, first semester of 2027.

**Location:** International Laboratory on Learning Systems (ILLS), McGill – MILA – ETS – Université Paris-Saclay – CentraleSupélec – CNRS, Montreal, Canada.

## Financial Assistance

- Reimbursement of airfare and lodging expenses may be available, subject to the availability of funds.
- **DATAIA International Mobility Grant** (Institut DATAIA, Université Paris-Saclay): €1,000 per month plus a €500 travel lump sum, combinable with other funding up to €1,500 per month, for a stay of 6 weeks to 6 months. The 2026–2027 call is expected to open at the end of September 2026, with the Master 2 wave closing at the end of November 2026 (dates to be confirmed when the call opens). dataia.eu/appels-a-projets
- **IDEX International Internship Grant** (Université Paris-Saclay): excellence-based grant, open to students of Paris-Saclay and its member institutions, including ENS Paris-Saclay; minimum duration of 6 weeks (42 days) at master level. universite-paris-saclay.fr/bourses-de-stage-linternational-idex
- **Other Université Paris-Saclay mobility support** (Bonus International, ministry mobility aid): a single application form is filed on the outgoing students portal, from which the International Relations Office allocates the most suitable grant. Overview: financements-pour-partir-letranger — Portal: psud.moveonfr.com
- **ENS Paris-Saclay students** may additionally apply to the AMIS and AMI schemes. ens-paris-saclay.fr/financer-votre-projet-letranger
- Opportunity to apply to calls for thesis funding.

## Required Skills

- Programming skills: foundations of object-oriented programming (Python).
- Foundations in reinforcement learning.
- Interest in bio-inspired Artificial Intelligence (AI) and minimal cognition.
- Ability to design and analyze computational experiments.

## Introduction

Understanding intelligence in its simplest form is a central question in AI. The challenge of this internship can be stated as follows: what are the minimal computational building blocks an agent needs to behave reactively, adaptively and to learn?

Valentino Braitenberg's Vehicles (Braitenberg, 1984) offer a natural starting point. Built from nothing more than sensors directly connected to motors, these conceptual machines already display rich, purposeful-looking behaviors. They demonstrate that intelligent behavior does not always require complex architectures. Sometimes it can emerge from the right combination of extremely simple components. Between these minimal machines and the capabilities usually attributed to learning agents lies an intermediate class: the kineses, taxes and compass reactions catalogued by Fraenkel and Gunn (1961). That catalogue is descriptive but the strategies it names, from undirected changes in turning rate to steering by comparison of sensor readings and fixed-angle headings, form a concrete repertoire of behaviors to specify as minimal agents and then measure in simulation.

**The progression itself is the methodology.** Starting directly from a model-based reinforcement-learning agent, and asking whether it reproduces an observed behavior, leaves the necessity of the model unexamined. Simpler algorithms may account for the same behavior. Establishing what reactive, memory-based and adaptive agents can and cannot do is therefore a prerequisite to any claim about the necessity of learning or planning.

A central idea guides this exploration: computational activity reveals necessity. An agent's internal activity consists of different measures of changes over time (its state changes, events, computations). The choice of activity measure depends on the problem but it is consistent at each level of the agent's description (neuronal activity, movement activity, etc.). Activity offers direct evidence of which components of the agent matter during successful and unsuccessful behaviors: components that are consistently active (inversely, inactive) during a successful or unsuccessful behavior are likely necessary (inv. unnecessary) to achieve this behavior. Activity measures can be extended to the connections (e.g., synapses) between the components (e.g., neurons). By measuring internal activity, we can prune, simplify, and ultimately uncover minimal building blocks, or circuits of components, required to achieve a behavior.

To unify these scales, we rely on computational modeling and activity measures. This provides a coherent framework for comparing and simplifying models at every level, from simple neurons to agents and groups of agents.

## Objectives

- Design and implement agents built from minimal building blocks: sensors, internal states, adaptive mechanisms, learning rules.
- Start from Braitenberg-style vehicles, then progressively introduce orientation strategies, memory, adaptation, and reinforcement learning.
- Define and compute internal activity to determine which components contribute meaningfully.
- Use activity to identify minimal building blocks through pruning and simplification.
- Provide a systematic analysis linking behavior complexity, model complexity, environment complexity and activity.

## Methodology: A Graded Progression

The progression below spans five levels. **Levels 1 to 3 form the core of the internship and are expected to be completed. Level 4 is the target. Level 5 is an optional extension, to be attempted only if the earlier levels are secured.** This scoping is deliberate: the value of the work lies in the comparison across levels, not in reaching the last one.

**Level 1: Reactive agents (Braitenberg-like).** Minimal sensor–motor systems. Goal: characterize basic behavior, basic environment and activity measures. This goal remains valid at each subsequent level.

**Level 2 : Orientation strategies and memory-based agents.** Kineses and taxes (Fraenkel and Gunn, 1961), then agents equipped with elementary internal states. Goal: determine when and why memory becomes necessary.

**Level 3 : Adaptive agents.** Agents whose parameters evolve through experience. Goal: identify the minimal forms of adaptation.

**Level 4 : Reinforcement-learning agents.** Model-free RL and simple planning. Goal: justify the necessity of learning, relative to what Levels 1 to 3 already achieve.

**Level 5 (optional): Multi-agent and cooperative agents.** Agents interacting through activity-based coordination. Goal: determine the minimal structures enabling cooperation.

## Expected Outcomes

**Core:**

- A modular simulation platform (proof of concept) implementing Levels 1 to 3, extensible to the later levels. An existing visualization software of agents moving and interacting with their environment will be chosen rather than developed.
- Activity-based comparisons between agent architectures.
- Identification of the minimal structures required for each class of tasks addressed.
- A clear and rigorous internship report; clean, well-structured and documented code.

**If time allows:** Level 4, then Level 5, and a computational model of agents at each level covered.

## Bibliography


Braitenberg, V. (1984). Vehicles: Experiments in Synthetic Psychology. MIT Press.

Fraenkel, G. S., & Gunn, D. L. (1961). The Orientation of Animals: Kineses, Taxes and Compass Reactions. Oxford Clarendon Press.

Mascart, C., Muzy, A., et al. (2022). Time-Asynchronous Simulation of Large-Scale Spiking Neural Networks. Neural Computation.

Muzy, A., & Zeigler, B. P. (2017). Activity-Based Credit Assignment Heuristic for Simulation-Based Stochastic Search. IEEE Systems Journal.

Sabri, O., Muzy, A., et al. (2023). Gradient-Ascent Activity-Based Credit Assignment for Cooperative Multi-Agent Learning. Scientific Reports.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.

Zeigler, B. P., Muzy, A., & Kofman, E. (2018). Theory of Modeling and Simulation. Academic Press.