PhD Thesis Topic

Meta-Learning and Artificial General Intelligence

for a computational theory of assistance to human learning

Supervision: Aomar Osmani and Jean-Philippe Kotowicz
Host laboratory: LITIS – INSA Rouen Normandy
Field: Computer Science / Artificial Intelligence / Applied Mathematics
Contact: aomar.osmani@insa-rouen.fr
Keywords: Meta-learning • Artificial General Intelligence • Learner modeling • Continuous personalization • Cognitive education • Explainable AI (XAI)

Abstract

Recent advances in artificial intelligence have profoundly renewed the capabilities of learning systems, in particular through meta-learning, reinforcement learning, adaptive architectures, and models able to integrate heterogeneous knowledge. These developments open the way toward Artificial General Intelligence (AGI), capable of learning to learn, transferring knowledge across contexts, and reasoning about its own learning mechanisms.

In parallel, contemporary educational sciences and cognitive sciences show that human learning is an intrinsically dynamic process, structured by internal properties such as motivation, intention, learning schemas, desires, reward mechanisms, regulation of cognitive effort, and transfer ability. These properties play a central role in learning speed, the quality of acquired knowledge, and its reuse in new contexts.

In an educational context marked by a persistent paradox—*each learner is unique, yet learning systems are often standardized*—the challenge is to design assistants capable of fine-grained, *real-time* personalization, not only based on explicit knowledge, but also from cognitive and motivational signals (attention, engagement, forgetting, effort, confidence).

This PhD proposes to develop a theoretical and computational meta-learning framework, oriented toward AGI, explicitly integrating these properties as objects of modeling and reasoning. The goal is to establish a theory of assistance to human learning in which an intelligent assistant dynamically adapts its support, assessment, and transfer strategies according to a learner model and the evolution of context.

Scientific context and positioning

Artificial learning systems rely on a sequence of structuring choices related to knowledge representation, data decomposition, evaluation functions, and exploration strategies. These choices induce representational and reasoning biases that strongly condition performance, generalization, and explainability.

In human learning, these biases are not artifacts to remove but functional mechanisms: motivation drives engagement, intention structures goals, learning schemas organize understanding, rewards—internal or external—guide adaptation, and transfer is a fundamental criterion of successful learning. A direct consequence is that learning assistance cannot be reduced to a static knowledge model: it must incorporate the dynamics of attention, effort, forgetting, and context changes.

Meta-learning provides a unifying framework to make these biases explicit, represent them formally, and learn how to adapt them. In this perspective, motivational, intentional, and affective properties are not treated as purely descriptive psychological concepts, but as internal functions for evaluation, selection, and regulation of learning trajectories, which can be learned, transferred, and explained.

This PhD lies at the intersection of AGI, meta-learning, computational cognitive science, and the learning sciences, with the ambition to move beyond classical educational models toward a computational, dynamic, and individualized approach to learning. In particular, it aims to design a *meta-model* capable of optimizing pedagogical strategies from diverse experiences and adapting to a new learner with limited initial data, while integrating advances within the GAIAHL framework. The GAIAHL project addresses this fundamental challenge by proposing the design of an AGI architecture specialized in optimizing human learning. The core objective is to overcome the limitations of static models by developing a meta-model that learns to optimize pedagogical strategies in real time, adjusting not only to knowledge, but also to the learner's cognitive processes (attention, engagement, forgetting).

Problem statement and research questions

The overall objective is to formalize a computational theory of assistance to human learning, in which the assistant (i) builds a dynamic learner model, (ii) learns to choose and adapt pedagogical interventions, and (iii) guarantees the transferability and explainability of its decisions.

Key questions include:

- How can we unify knowledge state with metacognitive variables (confidence, effort, attentional biases) and motivational variables within a dynamic learner model?
- How can meta-learning be used to learn assistance strategies (activity sequences, explanations, feedback) that generalize across contexts and diverse profiles?
- How can we ensure effective transfer of *pedagogical know-how* across skill domains and rapid adaptation to a new learner (*few-shot*)?
- What forms of explainability are relevant to connect the assistant's decisions to internal mechanisms (objectives, rewards, regulation) and to observable indicators?

Scientific objectives

The main objectives of the PhD are:

1. Define a general meta-learning framework explicitly integrating representational, reasoning, and evaluation biases.

2. Formalize core properties of human learning (motivation, intention, schemas, rewards, transfer) as computational objects.
3. Build a dynamic model of human learning integrating progress, stagnation, forgetting, and recovery, and linking these dynamics to metacognitive variables (e.g., effort, confidence, engagement).
4. Study generalization and transfer mechanisms from an AGI perspective, including rapid adaptation to new learners and new domains.
5. Establish a theory of assistance to human learning based on metacognitive adaptation, real-time personalization, and explainability.

Theoretical aspects

The PhD will address in particular:

- formal modeling of motivation, intention, and desire as internal functions that guide learning;
- multi-level reward mechanisms and their role in exploration, engagement, and transfer;
- learning schemas as recurrent structures for representation and reasoning;
- the temporal dynamics of human learning (attention, engagement, forgetting, recovery) and their implications for assistance;
- meta-reasoning and adaptation of learning strategies, including generation/recommendation of activity sequences and feedback;
- links between personalization and explainability (XAI) for education: justification of interventions, traceability of choices, open learner models.

Methodology

The work will combine theoretical modeling, algorithmic development (meta-learning, reinforcement learning, adaptive learning), the design of multi-level meta-modeling architectures, and experimental validation on human learning scenarios.

Particular attention will be given to:

- *learner modeling* (including metacognitive/motivational variables);
- meta-learning of assistance policies (model-based, metric-based, optimization-based) able to recommend or generate interventions at time step t;
- empirical evaluation: proof-of-concept on targeted use cases (e.g., technical skills, continuing education), then controlled studies measuring learning gain, engagement, transfer, and robustness.

Indicative profile

We are looking for a candidate with an MSc degree or engineering diploma in computer science, data science, AI, computational cognitive science, or mathematics, with a strong interest in research.

Desired skills:

- solid foundations in ML/DL;
- interest in cognitive science, learning sciences/education, or optimization;
- strong interest in mathematical modeling, as well as modeling and programming;
- knowledge of meta-learning, RL, and sequential models (RNN/Transformers) is a significant plus.

Environment:

- multidisciplinary project (AI, cognitive science, instructional engineering) with strong societal impact;
- computing resources and data for large-scale experiments;
- expected dissemination in top-tier international venues (NeurIPS, ICLR, AIED, etc.).

Selected references

References

[1] B. Lake et al. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 2017.

[2] C. Finn et al. Model-agnostic meta-learning for fast adaptation. *ICML*, 2017.

[3] T. Hospedales et al. Meta-learning in neural networks: A survey. *IEEE TPAMI*, 2021.

[4] Y. Bengio. From System 1 to System 2 deep learning. *NeurIPS*, 2019.

[5] S. Dehaene. *How We Learn*. Viking, 2020.

[6] P. Winne. Self-regulated learning and open learner models. *IJAIED*, 2021.

[7] H. Khosravi et al. Explainable artificial intelligence in education. *Computers and Education: AI*, 2022.

[8] B. Lake et al. Human-like learning in machines. *Annual Review of Psychology*, 2023.