Ph.D. Thesis Proposal

Verified Causal Reinforcement Learning for Robust Transfer under Distribution Shift

Type: 3-year doctoral thesis
Start date: last quarter of 2026
Level: Master's (M2) degree or engineering diploma (or equivalent) in Applied Mathematics, Data Science, Computer Science, and related fields.
Funding: Full doctoral contract according to current French regulations (gross salary as per university scale)
Location: LITIS, Université de Rouen Normandie, INSA Rouen Normandie, Rouen

1 Scientific Context

Deep neural networks can learn high-dimensional representations and policies from large datasets or repeated interaction, yet their success does not imply that the learned representations capture the mechanisms that generate the environment [1, 4]. In this scenario, learning agents are therefore vulnerable to distribution changes that alter nuisance variables, correlations, observation processes, or parts of the environment dynamics [3, 11]. This gap between training and deployment is particularly problematic when failures are rare, expensive, or difficult to reproduce. For instance, in autonomous driving, aerospace, industrial control, and healthcare, the available data may systematically under-represent precisely the states in which reliable reasoning is most important.

Nevertheless, many systems contain knowledge that is more stable than the observed data distribution. Physical laws, component relationships, conservation principles, anatomical constraints, and other relational properties may remain unchanged while weather, sensor characteristics, user populations, or operational regimes vary [9]. Instead of transferring an undifferentiated neural representation, a model should identify and reuse causal mechanisms that remain invariant across environments. This perspective connects transfer learning to causal reasoning. In this case, causal models are intended to describe how variables respond to specific actions (i.e., interventions) and, under suitable assumptions, can support predictions beyond the observational regime from which data were collected [7, 8].

Reinforcement learning (RL) provides a natural setting for this, as actions already have an intervention-like interpretation. In a standard Markov decision process (MDP), however, the transition law $P(s' \mid s, a)$ is typically treated as a conditional distribution without an explicit representation of the mechanisms producing the next state. A causal formulation instead models the factors of the state and reward through structural causal mechanisms, making it possible to distinguish conditioning from intervention and to reason about mechanism changes, confounding, transport, and counterfactual alternatives. Causal MDP formulations have shown that prior causal knowledge can reduce the complexity of sequential decision making in structured settings [5]; invariant causal prediction has also been connected to state abstraction and generalization across environments [12].

A second limitation of contemporary learning systems is that empirical performance alone provides weak guarantees in high-stakes settings. A policy may attain a high average reward while violating a safety requirement in rare states. Formal verification offers a complementary approach: rather than relying solely on testing, one states assumptions and desired properties mathematically and checks proofs with a proof assistant such as Lean [6], which provides an expressive dependent type theory and a growing mathematical library. Recent works have begun to formalize core RL convergence results in Lean [13], and demonstrated substantial machine-checked infrastructure for causal inference [10].

Thus, this thesis aims to answer the research question: *can causal invariance provide a transferable representation for reinforcement learning under distribution shift, and can the semantics and selected guarantees of such a causal RL system be formally verified?*

2 Problem Statement

Transfer under distribution shift. Conventional deep RL can exploit predictive correlations that are specific to the training distribution. Domain randomization, representation learning, multi-task learning, and continual learning methods can improve robustness, but they do not, by themselves, identify which dependencies correspond to stable causal mechanisms. Consequently, transfer may fail when the target environment changes the relationship between variables that were spuriously correlated during training. The thesis focuses on mechanism-aware transfer. An environmental change is represented not merely as a different joint distribution but as a change to selected causal mechanisms. This distinction allows one to ask which parts of a learned model should be reused and which should be adapted.

Causal semantics in sequential decision making. In observational data, $P(Y \mid A = a)$ need not equal $P(Y \mid \mathrm{do}(A = a))$ when action selection is confounded. The distinction is especially important in offline and partially observed settings, where policies are learned from historical behavior rather than randomized interaction. Recent work has characterized settings in which confounding prevents consistent offline policy evaluation and has developed alternatives under explicit assumptions [2]. This thesis will use the distinction between observation and intervention to make the RL model's assumptions explicit and to formalize the conditions under which the induced causal transition semantics are valid.

Verification gap. Formal verification of reinforcement-learning theory and formalization of causal inference are emerging areas, but they have largely developed independently. Formal RL work has begun to establish machine-checked convergence theorems [13], while formal causal-inference infrastructure is now appearing [10]. The research gap targeted here is the formal interface between the two: a reusable semantics for causal sequential decision models together with proofs about intervention-induced transitions, invariant transfer, and policy safety. The goal is not to show that causality automatically solves catastrophic forgetting, nor that a theorem prover can certify an arbitrary neural network end-to-end. Instead, continual learning will be treated as one possible evaluation regime for transfer, while formal verification will target a mathematically defined causal abstraction and the properties enforced at its interface with a learned policy.

3 Research Objectives

This thesis aims to develop and evaluate a framework for verified causal reinforcement learning that leverages invariant causal mechanisms and relational background knowledge to enable robust transfer under distribution shift, and uses Lean to machine-check selected semantic, transfer, and safety properties.

3.1 Research questions

RQ1: Causal representation. How can relational and physical background knowledge be incorporated into a causal representation of sequential decision problems so that task-relevant state factors and mechanisms are distinguished from environment-specific correlations?

RQ2: Causal transfer. Under what classes of distribution shift does transferring invariant causal mechanisms improve sample efficiency and out-of-distribution policy performance compared with non-causal deep RL, relational RL, and standard transfer baselines?

RQ3: Formal semantics. Can a finite causal Markov decision process, including interventions, induced transition kernels, policies, value functions, and environment invariance, be formalized in Lean with machine-checked proofs of its core semantic properties?

RQ4: Verified decision making. Can a formally verified causal safety layer constrain or filter the actions of a learned policy such that specified safety properties are preserved under explicit model assumptions and selected forms of distribution shift?

4 Expected Contributions

1. **Methodological contribution:** a causal-relational RL architecture in which prior domain knowledge constrains a factored causal model and transfer operates at the level of invariant mechanisms.
2. **Theoretical contribution:** explicit conditions connecting invariant causal submodels to policy/value preservation or safe reuse across environments.
3. **Formal-methods contribution:** a Lean library for a finite causal RL core, including interventions, induced transition semantics, policies, values, environment invariance, and selected machine-checked transfer and safety theorems.

4.1 Research Laboratory

The LITIS (Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes) is a public research unit (EA 4108) in computer science, jointly affiliated with the University of Rouen Normandy, the INSA Rouen Normandy, and the University of Le Havre Normandy. A member of the CNRS research federation NormaSTIC (FR 3638), LITIS conducts advanced research in machine learning, pattern recognition, information processing, and intelligent systems. Its activities span a broad spectrum, from fundamental algorithmic research to applied system design, across domains such as health, smart mobility, and smart territories. The laboratory has a strong tradition of technological transfer and industrial collaboration, working with partners such as Airbus, BioMérieux, Orange Labs, Siemens, and Valeo. In the context of this thesis, the LITIS will provide academic supervision and expertise in causal inference, adaptive machine learning, and hybrid modeling, supporting the methodological development of causal digital twins for building energy management.

4.2 Doctoral school

The Doctoral School MIIS (Mathematics, Computer Science, and Systems Engineering) brings together doctoral research and training activities in digital sciences and engineering across the Normandy region. It federates several higher education institutions, including INSA Rouen Normandy, the University of Rouen Normandy, and the University of Le Havre Normandie, in close collaboration with recognized research laboratories such as LITIS and LMRS.

The MIIS Doctoral School aims to train high-level researchers and engineers capable of addressing complex scientific and technological challenges at the intersection of mathematical modeling, artificial intelligence, computer systems, and engineering. It actively promotes interdisciplinary collaboration and industrial partnerships, particularly through CIFRE programs, which bridge academic research and applied innovation. By providing high-quality scientific supervision, support for research valorization, and strong international openness, MIIS offers an excellent environment for developing the scientific, technical, and professional skills of its doctoral candidates.

5 Profile

We are looking for candidates interested in causal reinforcement learning and formal verification who hold (or are about to obtain) a Master's degree or engineering diploma in applied mathematics or computer science, with a solid background in machine learning. Prior exposure to reinforcement learning, causal inference, or Lean is appreciated but not required. A good level of English and an interest in both theoretical analysis and rigorous experimentation are strongly recommended.

6 Application

Applications should include a detailed CV, a cover letter, and academic transcripts from the Master's or engineering years, and contact details of one or two references.

7 Supervisors

Alessandro Leite (a{lastname}[at]insa-rouen[dot]fr)

References

[1] Ahmed M Abdel-Zaher and Ayman M Eldeib. “Breast cancer classification using deep belief networks”. In: *Expert Systems with Applications* (2016).

[2] Chinmaya Kausik et al. “Offline policy evaluation and optimization under confounding”. In: *International Conference on Artificial Intelligence and Statistics*. 2024.

[3] Barbara Landau, Linda B Smith, and Susan S Jones. “The importance of shape in early lexical learning”. In: *Cognitive development* (1988).

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* (2015).

[5] Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. “Efficient reinforcement learning with prior causal knowledge”. In: *Conference on Causal Learning and Reasoning*. 2022.

[6] Leonardo de Moura and Sebastian Ullrich. “The Lean 4 theorem prover and programming language”. In: *International Conference on Automated Deduction*. 2021.

[7] Judea Pearl. *Causality: Models, Reasoning, and Inference*. 2nd. Cambridge University Press, 2009.

[8] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017.

[9] Adam Santoro et al. “Relational recurrent neural networks”. In: *Advances in Neural Information Processing Systems*. 2018.

[10] Jiyuan Tan and Vasilis Syrgkanis. “CausalForge: A Formally Grounded, Self-Improving Agentic Framework for Automated Research in Causal Inference”. In: *arXiv:2607.22511* (2026).

[11] Sebastian Thrun. “Lifelong learning algorithms”. In: *Learning to learn*. 1998.

[12] Amy Zhang et al. “Invariant causal prediction for block mdps”. In: *International Conference on Machine Learning*. 2020.
[13] Shangtong Zhang. “Towards formalizing reinforcement learning theory”. In: *arXiv:2511.03618* (2025).