

2025 Activity Report

RESEARCH CENTRE: Inria Centre at the University of Lille

IN PARTNERSHIP WITH: CNRS, Université de Lille

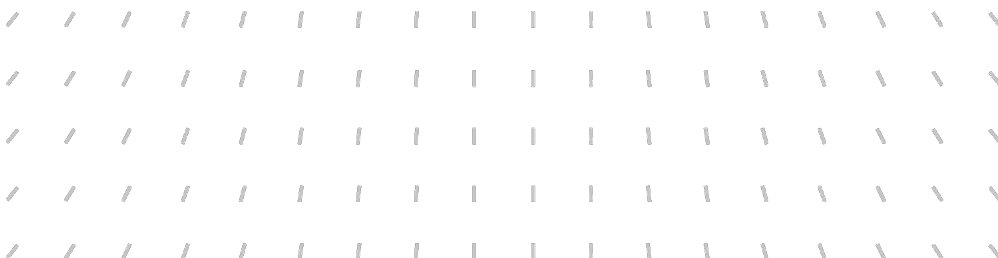

Project-Team

SCOOOL

Sequential decision making under uncertainty
problem



*In collaboration with Centre de Recherche en Informatique, Signal et Automatique
de Lille*



Project-Team SCOOL

Creation of the Project-Team: 2020 November 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

- A3. – Data and knowledge
 - A3.1. – Data
 - A3.1.1. – Modeling, representation
 - A3.1.1.4. – Uncertain data
 - A3.1.1.11. – Structured data
- A3.3. – Data and knowledge analysis
 - A3.3.1. – On-line analytical processing
 - A3.3.2. – Data mining
 - A3.3.3. – Big data analysis
- A3.4. – Machine learning and statistics
- A3.5.2. – Recommendation systems
- A5.1. – Human-Computer Interaction
- A8.6. – Information theory
- A8.11. – Game Theory
- A9. – Artificial intelligence
 - A9.2. – Machine learning
 - A9.2.1. – Supervised learning
 - A9.2.2. – Unsupervised learning
 - A9.2.3. – Reinforcement learning
 - A9.2.4. – Optimization and learning
 - A9.2.5. – Bayesian methods
 - A9.2.6. – Neural networks
 - A9.2.8. – Deep learning
 - A9.3. – Signal processing
 - A9.4. – Natural language processing
 - A9.7. – AI algorithmics

Other research topics and application domains

- B2. – Digital health
 - B3.1. – Sustainable development
 - B3.5. – Agronomy
 - B9.5. – Sciences
 - B9.5.6. – Data science

Contents

| | |
|---|-----------|
| Project-Team SCOOL | 1 |
| 1 Team members, visitors, external collaborators | 5 |
| 2 Overall objectives | 6 |
| 3 Research program | 7 |
| 4 Application domains | 7 |
| 5 Social and environmental responsibility | 8 |
| 6 Highlights of the year | 8 |
| 7 New results | 8 |
| 7.1 Bandits and RL theory | 8 |
| 7.2 Bandits and RL under Real-life constraints | 10 |
| 7.3 Bandits and RL for real-life: Deep RL and Applications | 11 |
| 7.4 Others | 13 |
| 7.4.1 Responsible AI and Algorithmic Auditing | 15 |
| 7.4.2 Formalization of mathematics | 17 |
| 8 Bilateral contracts and grants with industry | 17 |
| 8.1 Bilateral contracts with industry | 17 |
| 9 Partnerships and cooperations | 17 |
| 9.1 International initiatives | 18 |
| 9.1.1 Inria associate team not involved in an ILL or an international program | 18 |
| 9.2 International research visitors | 18 |
| 9.2.1 Visits of international scientists | 18 |
| 9.2.2 Visits to international teams | 19 |
| 9.3 National initiatives | 20 |
| 9.3.1 ANR projects | 20 |
| 9.3.2 PEPR projects | 20 |
| 9.3.3 Other projects in France | 21 |
| 9.3.4 Inria Exploratory Actions | 21 |
| 10 Dissemination | 22 |
| 10.1 Promoting scientific activities | 22 |
| 10.1.1 Scientific events: organisation | 22 |
| 10.1.2 Scientific events: selection | 22 |
| 10.1.3 Journal | 23 |
| 10.1.4 Invited talks | 23 |
| 10.1.5 Scientific expertise | 24 |
| 10.1.6 Research administration | 24 |
| 10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach | 24 |
| 10.2.1 Supervision | 25 |
| 10.2.2 Juries | 26 |
| 10.2.3 Educational and pedagogical outreach | 26 |
| 10.3 Popularization | 26 |
| 10.3.1 Productions (articles, videos, podcasts, serious games, ...) | 26 |
| 10.3.2 Participation in Live events | 27 |
| 10.3.3 Others science outreach relevant activities | 27 |

| | |
|---|-----------|
| 11 Scientific production | 27 |
| 11.1 Major publications | 27 |
| 11.2 Publications of the year | 28 |
| 11.3 Cited publications | 31 |

1 Team members, visitors, external collaborators

Research Scientists

- Riadh Akrouf [INRIA, ISFP, HDR]
- Debabrota Basu [INRIA, ISFP]
- Remy Degenne [INRIA, ISFP]
- Emilie Kaufmann [CNRS, Researcher, HDR]
- Odalric-Ambrym Maillard [INRIA, Researcher, HDR]
- Timothee Mathieu [INRIA, Researcher]

Faculty Members

- Philippe Preux [Team leader, UNIV LILLE, Professor, HDR]
- Juliette Achddou [UNIV LILLE, Associate Professor]

Post-Doctoral Fellows

- Sabrine Chebbi [INRIA, Post-Doctoral Fellow]
- Lorenzo Hermez [INRIA, Post-Doctoral Fellow, from Nov 2025]
- Tanguy Lefort [INRIA, Post-Doctoral Fellow, until Feb 2025]

PhD Students

- Ayoub Ajarra [INRIA]
- Mickael Basson [LILLY FRANCE, CIFRE]
- Yann Berthelot [SAINT GOBAIN RESEARCH, CIFRE]
- Hadrien Crassous [INRIA, from Nov 2025]
- Udvas Das [INRIA]
- Brahim Driss [INRIA]
- Anthony Kobanda [UBISOFT, CIFRE]
- Hector Kohler [UNIV LILLE, until Sep 2025]
- Penanklihi Cyrille Kone [INRIA]
- Matheus Medeiros Centa [UNIV LILLE, until Mar 2025]
- Nicolas Michalak [INRIA]
- Thomas Michel [INRIA]
- Adrien Prevost [INRIA]
- Waris Radji [INRIA]
- Adrienne Tuynman [INRIA]
- Sumit Vashishtha [UNIV LILLE]
- Redouane Yagouti [INRIA, from Nov 2025]

Technical Staff

- Alex Davey [INRIA, Engineer, until Sep 2025]
- Guillaume Pourcel [INRIA, Engineer, until Apr 2025]
- Julien Teigny [INRIA, Engineer]

Interns and Apprentices

- Francesca Hemetsberger [INRIA, Intern, from Sep 2025]
- Aurele Mingam [INRIA, Intern, from Apr 2025 until Aug 2025]
- Francois Muller [INRIA, Intern, from Apr 2025 until Sep 2025]
- Balthazar Tack [CENTRALE LILLE, Intern, from Apr 2025 until Aug 2025]
- Redouane Yagouti [INRIA, Intern, from May 2025 until Oct 2025]

Administrative Assistant

- Amélie Supervielle [INRIA]

2 Overall objectives

Scool is a machine learning (ML) research group. Scool's research focuses on the study of the sequential decision making under uncertainty problem (SDMUP). In particular, we consider bandit problems [57] and the reinforcement learning (RL) problem [56]. In a simplified way, RL considers the problem of learning an optimal policy in a Markov Decision Problem (MDP) [54]; when the set of states collapses to a single state, this is known as the bandit problem which focuses on the exploration/exploitation problem.

Bandit and RL problems are interesting to study on their own; both types of problems share a number of fundamental issues (convergence analysis, sample complexity, representation, safety, *etc.*); both problems have real life applications, different though closely related; the fact that while solving an RL problem, one faces an exploration/exploitation problem and has to solve a bandit problem in each state connects the two types of problems very intimately.

In our work, we also consider settings going beyond the Markovian assumption, in particular non-stationary settings, which represent a challenge common to bandits and RL. A distinctive aspect of the SDMUP with regards to the rest of the field of ML is that the learning problem takes place within a closed-loop interaction between a learning agent and its environment. This feedback loop makes our field of research very different from the two other sub-fields of ML, supervised and unsupervised learning, even when they are defined in an incremental setting. Hence, SDMUP combines ML with control: the learner is not passive, the learner acts on its environment, and learns from the consequences of these interactions; hence, the learner can act in order to obtain information from the environment. Naturally, the optimal control community is getting more and more interested by RL (see e.g. [55]).

We wish to go on, studying applied questions and developing theory to come up with sound approaches to the practical resolution of SDMUP tasks, and guide their resolution. Non-stationary environments are a particularly interesting setting; we are studying this setting and developing new tools to approach it in a sound way, in order to have algorithms to detect environment changes as fast as possible, and as reliably as possible, adapt to them, and prove their behavior, in terms of their performance, measured with the regret for instance. We mostly consider non parametric statistical models, that is models in which the number of parameters is not fixed (a parameter may be of any type: a scalar, a vector, a function, *etc.*), so that the model can adapt along learning, and to its changing environment; this also lets the algorithm learn a representation that fits its environment.

3 Research program

Our research is mostly dealing with bandit problems, and reinforcement learning problems. We investigate each thread separately and also in combination, since the management of the exploration/exploitation trade-off is a major issue in reinforcement learning.

On bandit problems, we focus on:

- structured bandits
- bandits for planning (in particular for Monte Carlo Tree Search (MCTS))
- non stationary bandits

Regarding reinforcement learning, we focus on:

- modeling issues, and dealing with the discrepancy between the model and the task to solve
- learning and using the structure of a Markov decision problem, and of the learned policy
- generalization in reinforcement learning
- reinforcement learning in non stationary environments

Beyond these objectives, we put a particular emphasis on the study of non-stationary environments. Another area of great concern is the combination of symbolic methods with numerical methods, be it to provide knowledge to the learning algorithm to improve its learning curve, or to better understand what the algorithm has learned and explain its behavior, or to rely on causality rather than on mere correlation.

We also put a particular emphasis on real applications and how to deal with their constraints: lack of a simulator, difficulty to have a realistic model of the problem, small amount of data, dealing with risks, availability of expert knowledge on the task.

4 Application domains

Scool has 2 main topics of application:

- health
- sustainable development

In each of these two domains, we put forward the investigation and the application of the idea of sequential decision making under uncertainty. Though supervised and non supervised learning have already been studied and applied extensively, sequential decision making remains far less studied; bandits have already been used in many applications of e-commerce (e.g. for computational advertising and recommendation systems). However, in applications where human beings may be severely impacted, bandits and reinforcement learning have not been studied much; moreover, these applications come along with a scarcity of data, and the non availability of a simulator, which prevents heavy computational simulations to come up with safe automatic decision making.

In 2022, in health, we investigated patient follow-up with Prof. F. Pattou's research group (CHU Lille, Inserm, Université de Lille) in project B4H. This effort came along with investigating how we may use medical data available locally at CHU Lille, and also the national social security data. We also investigated drug repurposing with Prof. A. Delahaye-Duriez (Inserm, Université de Paris) in project Repos. We also studied catheter control by way of reinforcement learning with Inria Lille group Defrost, and company Robocath (Rouen).

Regarding sustainable development, we have a set of projects and collaborations regarding agriculture and gardening. With Cirad and CGIAR, we investigate how one may recommend agricultural practices to farmers in developing countries. Through an associate team with Bihar Agriculture University (India), we investigate data collection. Inria exploratory action SR4SG concerns recommender systems at the level of individual gardens.

There are two important aspects that are amply shared by these two application fields. First, we consider that data collection is an active task: we do not passively observe and record data, we design methods and algorithms to search for useful data. This idea is exploited in most of these works oriented towards applications. Second, many of these projects include a careful management of risks for human beings. We have to take decisions taking care of their consequences on human beings, on eco-systems and life more generally.

5 Social and environmental responsibility

Sustainable development is a major field of research and application of Scool. We investigate what machine learning can bring to sustainable development, identifying challenges and obstacles, and studying how to overcome them.

Let us mention here:

- sustainable agriculture in developing countries;
- sustainable gardening.

More details can be found in Section 4.

6 Highlights of the year

The Scool team was selected to organize the 19th edition of the European Workshop on Reinforcement Learning in October 2026.

Debabrota Basu got the "top reviewer" recognition from NeurIPS and was awarded a complementary registration.

7 New results

We organize our research results in a set of categories. The main categories are: bandits and RL theory, bandits and RL under real life constraints, and applications.

Participants: Adrienne Tuynman, Alex Davey, Anthony Kobanda, Ayoub Ajarra, Brahim Driss, Debabrota Basu, Emilie Kaufmann, Guillaume Pourcel, Hector Kohler, Mickael Basson, Odalric-Ambrym Maillard, Penank-lihi Cyrille Kone, Philippe Preux, Remy Degenne, Riadh Akrou, Sumit Vashishtha, Thomas Michel, Udvas Das, Waris Radji.

7.1 Bandits and RL theory

Best-Arm Identification in Unimodal Bandits, [39]

We study the fixed-confidence best-arm identification problem in unimodal bandits, in which the means of the arms increase with the index of the arm up to their maximum, then decrease. We derive two lower bounds on the stopping time of any algorithm. The instance-dependent lower bound suggests that due to the unimodal structure, only three arms contribute to the leading confidence-dependent cost. However, a worst-case lower bound shows that a linear dependence on the number of arms is unavoidable in the confidence-independent cost. We propose modifications of Track-and-Stop and a Top Two algorithm that leverage the unimodal structure. Both versions of Track-and-Stop are asymptotically optimal for one-parameter exponential families. The Top Two algorithm is asymptotically near-optimal for Gaussian distributions and we prove a non-asymptotic guarantee matching the worse-case lower bound. The algorithms can be implemented efficiently and we demonstrate their competitive empirical performance.

The Batch Complexity of Bandit Pure Exploration, [40]

In a fixed-confidence pure exploration problem in stochastic multi-armed bandits, an algorithm iteratively samples arms and should stop as early as possible and return the correct answer to a query about the arms distributions. We are interested in batched methods, which change their sampling behaviour only a few times, between batches of observations. We give an instance-dependent lower bound on the number of batches used by any sample efficient algorithm for any pure exploration task. We then give a general batched algorithm and prove upper bounds on its expected sample complexity and batch complexity. We illustrate both lower and upper bounds on best-arm identification and thresholding bandits.

Pareto Set Identification With Posterior Sampling, [33]

The problem of identifying the best answer among a collection of items having real-valued distribution is well-understood. Despite its practical relevance for many applications, fewer works have studied its extension when multiple and potentially conflicting metrics are available to assess an item's quality. Pareto set identification (PSI) aims to identify the set of answers whose means are not uniformly worse than another. This paper studies PSI in the transductive linear setting with potentially correlated objectives. Building on posterior sampling in both the stopping and the sampling rules, we propose the PSIPS algorithm that deals simultaneously with structure and correlation without paying the computational cost of existing oraclebased algorithms. Both from a frequentist and Bayesian perspective, PSIPS is asymptotically optimal. We demonstrate its good empirical performance in real-world and synthetic instances.

FraPPE: Fast and Efficient Preference-based Pure Exploration, [30]

Preference-based Pure Exploration (PrePEX) aims to identify with a given confidence level the set of Pareto optimal arms in a vector-valued (aka multi-objective) bandit, where the reward vectors are ordered via a (given) preference cone C . Though PrePEX and its variants are well-studied, there does not exist a computationally efficient algorithm that can optimally track the existing lower bound for arbitrary preference cones. We successfully fill this gap by efficiently solving the minimisation and maximisation problems in the lower bound. First, we derive three structural properties of the lower bound that yield a computationally tractable reduction of the minimisation problem. Then, we deploy a Frank-Wolfe optimiser to accelerate the maximisation problem in the lower bound. Together, these techniques solve the maxmin optimisation problem in $\mathcal{O}(KL^2)$ time for a bandit instance with K arms and L dimensional reward, which is a significant acceleration over the literature. We further prove that our proposed PrePEX algorithm, FraPPE, asymptotically achieves the optimal sample complexity. Finally, we perform numerical experiments across synthetic and real datasets demonstrating that FraPPE achieves the lowest sample complexities to identify the exact Pareto set among the existing algorithms.

Leveraging Priors on Distribution Functions for Multi-Arm Bandits, [21]

We introduce Dirichlet Process Posterior Sampling (DPPS), a Bayesian non-parametric algorithm for multi-arm bandits based on Dirichlet Process (DP) priors. Like Thompson-sampling, DPPS is a probability-matching algorithm, i.e., it plays an arm based on its posterior probability of being optimal. Instead of assuming a parametric class for the reward generating distribution of each arm, and then putting a prior on the parameters, in DPPS the reward generating distribution is directly modeled using DP priors. DPPS provides a principled approach to incorporate prior belief about the bandit environment, and in the noninformative limit of the DP priors (i.e. Bayesian Bootstrap), we recover Non Parametric Thompson Sampling (NPTS), a popular non-parametric bandit algorithm, as a special case of DPPS. We employ stick-breaking representation of the DP priors, and show excellent empirical performance of DPPS in challenging synthetic and real world bandit environments. Finally, using an information-theoretic analysis, we show non-asymptotic optimality of DPPS in the Bayesian regret setup.

Towards Blackwell Optimality: Bellman Optimality Is All You Can Get, [48]

Although average gain optimality is a commonly adopted performance measure in Markov Decision Processes (MDPs), it is often too asymptotic. Further incorporating measures of immediate losses leads to the hierarchy of bias optimalities, all the way up to Blackwell optimality. In this paper, we investigate the problem of identifying policies of such optimality orders. To that end, for each order, we construct a learning algorithm with vanishing probability of error. Furthermore, we characterize the class of MDPs for which identification algorithms can stop in finite time. That class corresponds to the MDPs with a unique Bellman optimal policy, and does not depend on the optimality order considered. Lastly, we provide a tractable stopping rule that when coupled to our learning algorithm triggers in finite time whenever it is possible to do so.

7.2 Bandits and RL under Real-life constraints

Constrained Pareto Set Identification with Bandit Feedback, [35]

In this paper, we address the problem of identifying the Pareto Set under feasibility constraints in a multivariate bandit setting. Specifically, given a K -armed bandit with unknown means in \mathbb{R}^d , the goal is to identify the set of arms whose mean is not uniformly worse than that of another arm (i.e., not smaller for all objectives), while satisfying some known set of linear constraints, expressing, for example, some minimal performance on each objective. Our focus lies in fixed-confidence identification, for which we introduce an algorithm that significantly outperforms racing-like algorithms and the intuitive two-stage approach that first identifies feasible arms and then their Pareto Set. We further prove an information-theoretic lower bound on the sample complexity of any algorithm for constrained Pareto Set identification, showing that the sample complexity of our approach is near-optimal. Our theoretical results are supported by an extensive empirical evaluation on a series of benchmarks.

Bandit Pareto Set Identification in a Multi-Output Linear Model, [34]

We study the Pareto Set Identification (PSI) problem in a structured multi-output linear bandit model. In this setting, each arm is associated a feature vector belonging to \mathbb{R}^h , and its mean vector in \mathbb{R}^d linearly depends on this feature vector through a common unknown matrix $\Theta \in \mathbb{R}^{h \times d}$. The goal is to identify the set of non-dominated arms by adaptively collecting samples from the arms. We introduce and analyze the first optimal design-based algorithms for PSI, providing nearly optimal guarantees in both the fixed-budget and the fixed-confidence settings. Notably, we show that the difficulty of these tasks mainly depends on the sub-optimality gaps of h arms only. Our theoretical results are supported by an extensive benchmark on synthetic and real-world datasets.

Kriging and Gaussian Process Interpolation for Georeferenced Data Augmentation, [51]

Data augmentation is a crucial step in the development of robust supervised learning models, especially when dealing with limited datasets. This study explores interpolation techniques for the augmentation of geo-referenced data, with the aim of predicting the presence of *Commelina benghalensis* L. in sugarcane plots in La Réunion. Given the spatial nature of the data and the high cost of data collection, we evaluated two interpolation approaches: Gaussian processes (GPs) with different kernels and kriging with various variograms. The objectives of this work are threefold: (i) to identify which interpolation methods offer the best predictive performance for various regression algorithms, (ii) to analyze the evolution of performance as a function of the number of observations added, and (iii) to assess the spatial consistency of augmented datasets. The results show that GP-based methods, in particular with combined kernels (GP-COMB), significantly improve the performance of regression algorithms while requiring less additional data. Although kriging shows slightly lower performance, it is distinguished by a more homogeneous spatial coverage, a potential advantage in certain contexts.

Optimal Regret of Bandits under Differential Privacy, [25]

As sequential learning algorithms are increasingly applied to real life, ensuring data privacy while maintaining their utilities emerges as a timely question. In this context, regret minimisation in stochastic bandits under ϵ -global Differential Privacy (DP) has been widely studied. Unlike bandits without DP, there is a significant gap between the best-known regret lower and upper bound in this setting, though they "match" in order. Thus, we revisit the regret lower and upper bounds of ϵ -global DP algorithms for Bernoulli bandits and improve both. First, we prove a tighter regret lower bound involving a novel information-theoretic quantity characterising the hardness of ϵ -global DP in stochastic bandits. Our lower bound strictly improves on the existing ones across all ϵ values. Then, we choose two asymptotically optimal bandit algorithms, i.e. DP-KLUCB and DP-IMED, and propose their DP versions using a unified blueprint, i.e., (a) running in arm-dependent phases, and (b) adding Laplace noise to achieve privacy. For Bernoulli bandits, we analyse the regrets of these algorithms and show that their regrets asymptotically match our lower bound up to a constant arbitrary close to 1. This refutes the conjecture that forgetting past rewards is necessary to design optimal bandit algorithms under global DP. At the core of our algorithms lies a new concentration inequality for sums of Bernoulli variables under Laplace mechanism, which is a new DP version of the Chernoff bound. This result is universally useful as the DP literature commonly treats the concentrations of Laplace noise and random variables separately, while we couple them to yield a tighter bound.

FLIPHAT: Joint Differential Privacy for High Dimensional Sparse Linear Bandits, [28]

High dimensional sparse linear bandits serve as an efficient model for sequential decision-making problems (e.g. personalized medicine), where high dimensional features (e.g. genomic data) on the users are available, but only a small subset of them are relevant. Motivated by data privacy concerns in these applications, we study the joint differentially private high dimensional sparse linear bandits, where both rewards and contexts are considered as private data. First, to quantify the cost of privacy, we derive a lower bound on the regret achievable in this setting. To further address the problem, we design a computationally efficient bandit algorithm, **Forgetful Iterative Private HARD Thresholding (FLIPHAT)**. Along with doubling of episodes and episodic forgetting, FLIPHAT deploys a variant of Noisy Iterative Hard Thresholding (N-IHT) algorithm as a sparse linear regression oracle to ensure both privacy and regret-optimality. We show that FLIPHAT achieves optimal regret up to logarithmic factors. We analyze the regret by providing a novel refined analysis of the estimation error of N-IHT, which is of parallel interest.

Stochastic Online Instrumental Variable Regression: Regrets for Endogeneity and Bandit Feedback, [31]

The independence of noise and covariates is a standard assumption in online linear regression with unbounded noise and linear bandit literature. This assumption and the following analysis are invalid in the case of endogeneity, i.e., when the noise and covariates are correlated. In this paper, we study the online setting of Instrumental Variable (IV) regression, which is widely used in economics to identify the underlying model from an endogenous dataset. Specifically, we upper bound the identification and oracle regrets of the popular Two-Stage Least Squares (2SLS) approach to IV regression but in the online setting. Our analysis shows that Online 2SLS (O2SLS) achieves $O(d^2 \log^2 T)$ identification and $O(\gamma \sqrt{dT \log T})$ oracle regret after T interactions, where d is the dimension of covariates and γ is the bias due to endogeneity. Then, we leverage O2SLS as an oracle to design OFUL-IV, a linear bandit algorithm. OFUL-IV can tackle endogeneity and achieves $O(d\sqrt{T} \log T)$ regret. For different datasets with endogeneity, we experimentally show efficiencies of O2SLS and OFUL-IV.

Unifying (Federated) (Private) High-Dimensional Bandits via ADMM, [45]

We study all possible variants of the high dimensional stochastic linear contextual bandit problem in federated and private settings. We propose a unifying algorithm design and analysis framework built on ADMM. Our method achieves existing state-of-the art guarantees in either setting for the central model. For the federated model, our results are entirely new and near-optimal in either setting. We also establish a novel lower bound on privacy-utility trade-off for the federated model in the private setting and demonstrate on suitable numerical experiments for all problem variants.

The Confusing Instance Principle for Online Linear Quadratic Control, [19]

We revisit the problem of controlling linear systems with quadratic cost under unknown dynamics with model-based reinforcement learning. Traditional methods like Optimism in the Face of Uncertainty and Thompson Sampling, rooted in multi-armed bandits (MABs), face practical limitations. In contrast, we propose an alternative based on the Confusing Instance (CI) principle, which underpins regret lower bounds in MABs and discrete Markov Decision Processes (MDPs) and is central to the Minimum Empirical Divergence (MED) family of algorithms, known for their asymptotic optimality in various settings. By leveraging the structure of LQR policies along with sensitivity and stability analysis, we develop MED-LQ. This novel control strategy extends the principles of CI and MED beyond small-scale settings. Our benchmarks on a comprehensive control suite demonstrate that MED-LQ achieves competitive performance in various scenarios while highlighting its potential for broader applications in large-scale MDPs.

7.3 Bandits and RL for real-life: Deep RL and Applications

Breiman meets Bellman: Non-Greedy Decision Trees with MDPs, [32]

In supervised learning, decision trees are valued for their interpretability and performance. While greedy decision tree algorithms like CART remain widely used due to their computational efficiency, they often produce sub-optimal solutions with respect to a regularized training loss. Conversely, optimal decision tree methods can find better solutions but are computationally intensive and typically limited to shallow trees or binary features. We present Dynamic Programming Decision Trees (DPDT), a framework that bridges the gap between greedy and optimal approaches. DPDT relies on a Markov Decision Process formulation combined with heuristic split generation to construct near-optimal decision trees with significantly reduced computational complexity. Our approach dynamically limits the set of admissible splits at each node while

directly optimizing the tree regularized training loss. Theoretical analysis demonstrates that DPDT can minimize regularized training losses at least as well as CART. Our empirical study shows on multiple datasets that DPDT achieves near-optimal loss with orders of magnitude fewer operations than existing optimal solvers. More importantly, extensive benchmarking suggests statistically significant improvements of DPDT over both CART and optimal decision trees in terms of generalization to unseen data. We demonstrate DPDT practicality through applications to boosting, where it consistently outperforms baselines. Our framework provides a promising direction for developing efficient, near-optimal decision tree algorithms that scale to practical applications.

How Hard is it to Confuse a World Model?, [53]

In reinforcement learning (RL) theory, the concept of most confusing instances is central to establishing regret lower bounds, that is, the minimal exploration needed to solve a problem. Given a reference model and its optimal policy, a most confusing instance is the statistically closest alternative model that makes a suboptimal policy optimal. While this concept is well-studied in multi-armed bandits and ergodic tabular Markov decision processes, constructing such instances remains an open question in the general case. In this paper, we formalize this problem for neural network world models as a constrained optimization: finding a modified model that is statistically close to the reference one, while producing divergent performance between optimal and suboptimal policies. We propose an adversarial training procedure to solve this problem and conduct an empirical study across world models of varying quality. Our results suggest that the degree of achievable confusion correlates with uncertainty in the approximate model, which may inform theoretically-grounded exploration strategies for deep model-based RL.

Hierarchical Subspaces of Policies for Continual Offline Reinforcement Learning, [43]

We consider a Continual Reinforcement Learning setup, where a learning agent must continuously adapt to new tasks while retaining previously acquired skill sets, with a focus on the challenge of avoiding forgetting past gathered knowledge and ensuring scalability with the growing number of tasks. Such issues prevail in autonomous robotics and video game simulations, notably for navigation tasks prone to topological or kinematic changes. To address these issues, we introduce HiSPO, a novel hierarchical framework designed specifically for continual learning in navigation settings from offline data. Our method leverages distinct policy subspaces of neural networks to enable flexible and efficient adaptation to new tasks while preserving existing knowledge. We demonstrate, through a careful experimental study, the effectiveness of our method in both classical MuJoCo maze environments and complex video game-like navigation simulations, showcasing competitive performances and satisfying adaptability with respect to classical continual learning metrics, in particular regarding the memory usage and efficiency.

A Continual Offline Reinforcement Learning Benchmark for Navigation Tasks, [42]

Autonomous agents operating in domains such as robotics or video game simulations must adapt to changing tasks without forgetting about the previous ones. This process called Continual Reinforcement Learning poses non-trivial difficulties, from preventing catastrophic forgetting to ensuring the scalability of the approaches considered. Building on recent advances, we introduce a benchmark providing a suite of video-game navigation scenarios, thus filling a gap in the literature and capturing key challenges : catastrophic forgetting, task adaptation, and memory efficiency. We define a set of various tasks and datasets, evaluation protocols, and metrics to assess the performance of algorithms, including state-of-the-art baselines. Our benchmark is designed not only to foster reproducible research and to accelerate progress in continual reinforcement learning for gaming, but also to provide a reproducible framework for production pipelines – helping practitioners to identify and to apply effective approaches.

StaQ it! Growing neural networks for Policy Mirror Descent, [44]

In Reinforcement Learning (RL), regularization has emerged as a popular tool both in theory and practice, typically based either on an entropy bonus or a Kullback-Leibler divergence that constrains successive policies. In practice, these approaches have been shown to improve exploration, robustness and stability, giving rise to popular Deep RL algorithms such as SAC and TRPO. Policy Mirror Descent (PMD) is a theoretical framework that solves this general regularized policy optimization problem, however the closed-form solution involves the sum of all past Q-functions, which is intractable in practice. We propose and analyze PMD-like algorithms that only keep the last M Q-functions in memory, and show that for finite and large enough M , a convergent algorithm can be derived, introducing no error in the policy update, unlike prior deep RL PMD implementations. StaQ, the resulting algorithm, enjoys strong theoretical guarantees and is competitive with deep RL baselines, while exhibiting less performance oscillation, paving the way for fully stable deep RL

algorithms and providing a testbed for experimentation with Policy Mirror Descent.

PB²: Preference Space Exploration via Population-Based Methods in Preference-Based Reinforcement Learning, [41]

Preference-based reinforcement learning (PbRL) has emerged as a promising approach for learning behaviors from human feedback without predefined reward functions. However, current PbRL methods face a critical challenge in effectively exploring the preference space, often converging prematurely to suboptimal policies that satisfy only a narrow subset of human preferences. In this work, we identify and address this preference exploration problem through population-based methods. We demonstrate that maintaining a diverse population of agents enables more comprehensive exploration of the preference landscape compared to single-agent approaches. Crucially, this diversity improves reward model learning by generating preference queries with clearly distinguishable behaviors, a key factor in real-world scenarios where humans must easily differentiate between options to provide meaningful feedback. Our experiments reveal that current methods may fail by getting stuck in local optima, requiring excessive feedback, or degrading significantly when human evaluators make errors on similar trajectories, a realistic scenario often overlooked by methods relying on perfect oracle teachers. Our population-based approach demonstrates robust performance when teachers mislabel similar trajectory segments and shows significantly enhanced preference exploration capabilities, particularly in environments with complex reward landscapes

Lagrangian-based Equilibrium Propagation: generalisation to arbitrary boundary conditions & equivalence with Hamiltonian Echo Learning, [52]

Equilibrium Propagation (EP) is a learning algorithm for training Energy-based Models (EBMs) on static inputs which leverages the variational description of their fixed points. Extending EP to time-varying inputs is a challenging problem, as the variational description must apply to the entire system trajectory rather than just fixed points, and careful consideration of boundary conditions becomes essential. In this work, we present Generalized Lagrangian Equilibrium Propagation (GLEP), which extends the variational formulation of EP to time-varying inputs. We demonstrate that GLEP yields different learning algorithms depending on the boundary conditions of the system, many of which are impractical for implementation. We then show that Hamiltonian Echo Learning (HEL) – which includes the recently proposed Recurrent HEL (RHEL) and the earlier known Hamiltonian Echo Backpropagation (HEB) algorithms – can be derived as a special case of GLEP. Notably, HEL is the only instance of GLEP we found that inherits the properties that make EP a desirable alternative to backpropagation for hardware implementations: it operates in a “forward-only” manner (i.e. using the same system for both inference and learning), it scales efficiently (requiring only two or more passes through the system regardless of model size), and enables local learning.

Studying Exploration in RL: An Optimal Transport Analysis of Occupancy Measure Trajectories, [18]

The rising successes of RL are propelled by combining smart algorithmic strategies and deep architectures to optimize the distribution of returns and visitations over the state-action space. A quantitative framework to compare the learning processes of these eclectic RL algorithms is currently absent but desired in practice. We address this gap by representing the learning process of an RL algorithm as a sequence of policies generated during training, and then studying the policy trajectory induced in the manifold of state-action occupancy measures. Using an optimal transport-based metric, we measure the length of the paths induced by the policy sequence yielded by an RL algorithm between an initial policy and a final optimal policy. Hence, we first define the Effort of Sequential Learning (ESL). ESL quantifies the relative distance that an RL algorithm travels compared to the shortest path from the initial to the optimal policy. Furthermore, we connect the dynamics of policies in the occupancy measure space and regret (another metric to understand the suboptimality of an RL algorithm), by defining the Optimal Movement Ratio (OMR). OMR assesses the fraction of movements in the occupancy measure space that effectively reduce an analogue of regret. Finally, we derive approximation guarantees to estimate ESL and OMR with a finite number of samples and without access to an optimal policy. Through empirical analyses across various environments and algorithms, we demonstrate that ESL and OMR provide insights into the exploration processes of RL algorithms and the hardness of different tasks in discrete and continuous MDPs.

7.4 Others

Improving Diffusion Models for the Traveling Salesman Problem (TSP) by Leveraging the Structure of the Solution Space, [26]

In this paper we show how leveraging the structure of the solution space of the Traveling Salesman Problem (TSP) can lead to a dramatic improvement of the performance of state of the art diffusion based neural solvers. Building on recent approaches of DIFUSCO and T2TCO which pipeline a diffusion-based solution generation with a local search procedure, we propose IDEQ (constrained Inverse Diffusion and EQuivalence class-based training of diffusion models for combinatorial optimization). IDEQ improves the quality of the solutions by leveraging the constrained structure of the TSP state space. Indeed, the solution space consists of locally optimal Hamiltonian tours which is a much smaller space than the space of adjacency matrices used in previous works. Also, the local search procedure defines an equivalence class of Hamiltonian tours: all elements of this equivalence class reach the same local optimum after the application of the local search. This should be aligned with the supervised training objective of the diffusion. IDEQ addresses these two points. Our experiments show that IDEQ achieves 0.3% to 0.4% optimality gap on TSP instances made of 500 cities, and 0.5% to 0.6% optimality gap on TSP instances with 1000 cities. This sets a new SOTA for neural based methods solving the TSP. IDEQ also performs well on the instances of the TSPLib, a reference benchmark in the TSP community, outside of the training distribution, with optimality gaps ranging from 0.9 to 1.1 %.

Yara: An Ocean Virtual Environment for Research and Development of Autonomous Sailing Robots and Other Unmanned Surface Vessels, [20]

Overall, a big challenge in building a sailboat USV relies on the development of an autonomous system for guidance, navigation, and control (GNC) because both sail and rudder angle must be cooperatively adjusted to correct the navigation direction -traditional propelled boats can be more easily controlled with a straightforward control task to set the rudder angle. Moreover, sailing upwind requires special maneuvers to reach a given target in that unfeasible direction. Reinforcement learning emerges as a promising technique for building autonomous GNCs for sailing robots, but training the neural network with a real sailboat is impractical due to long periods of training and safety reasons. Even traditional control-based approaches are mainly tested in simulated environments due to the difficulties in building and operating a real sailboat. The issue that arises is the fidelity of these simulated environments. In this context, we propose Yara, an oceanic virtual environment with a reliable physics simulation for developing, training, and evaluating autonomous agents to operate digital twins of sailing robots in reinforcement learning and other paradigms. An autonomous sailing robot digital twin is available within the virtual environment, with the foil dynamics constructed based on a real sailing robot. We coupled these foil dynamics in Gazebo's physics engine to compute the lift and drag forces acting on the sail, rudder, and keel. The simulated world feeds sensors such as cameras, wind sensors, and GPS. The Robot Operating System communicates these sensors' data through topics, facilitating users' implementation and testing of new GNC solutions. Yara provides a reliable solution for foil dynamic simulated physics that achieves a simulation speedup of 300 times on an i7 laptop with 8 GB of RAM, powered by a Nvidia RTX 3060 and running Ubuntu 20.04. With this speedup, it is possible to complete a million time steps of deep reinforcement learning training in approximately eight hours. Evaluation scenarios were presented to highlight specific features of the simulator, like the maneuverability of the sailing robot digital twin and applications to train, evaluate, and compare reinforcement learning agents and other control solutions.

Efficient Active Imitation Learning with Random Network Distillation, [27]

Developing agents for complex and underspecified tasks, where no clear objective exists, remains challenging but offers many opportunities. This is especially true in video games, where simulated players (bots) need to play realistically, and there is no clear reward to evaluate them. While imitation learning has shown promise in such domains, these methods often fail when agents encounter out-of-distribution scenarios during deployment. Expanding the training dataset is a common solution, but it becomes impractical or costly when relying on human demonstrations. This article addresses active imitation learning, aiming to trigger expert intervention only when necessary, reducing the need for constant expert input along training. We introduce Random Network Distillation DAgger (RND-DAgger), a new active imitation learning method that limits expert querying by using a learned state-based out-of-distribution measure to trigger interventions. This approach avoids frequent expert-agent action comparisons, thus making the expert intervene only when it is useful. We evaluate RND-DAgger against traditional imitation learning and other active approaches in 3D video games (racing and third-person navigation) and in a robotic locomotion task and show that RND-DAgger surpasses previous methods by reducing expert queries. [Link](#)

Exploring Flow-Lenia Universes with a Curiosity-driven AI Scientist: Discovering Diverse Ecosystem

Dynamics, [37]

We present a method for the automated discovery of system-level dynamics in Flow-Lenia—a continuous cellular automaton (CA) with mass conservation and parameter localization—using a curiosity-driven AI scientist. This method aims to uncover processes leading to self-organization of evolutionary and ecosystemic dynamics in CAs. We build on previous work which uses diversity search algorithms in Lenia to find self-organized individual patterns, and extend it to large environments that support distinct interacting patterns. We adapt Intrinsically Motivated Goal Exploration Processes (IMGEPs) to drive exploration of diverse Flow-Lenia environments using simulation-wide metrics, such as evolutionary activity, compression-based complexity, and multi-scale entropy. We test our method in two experiments, showcasing its ability to illuminate significantly more diverse dynamics compared to random search. We show qualitative results illustrating how ecosystemic simulations enable self-organization of complex collective behaviors not captured by previous individual pattern search and analysis. We complement automated discovery with an interactive exploration tool, creating an effective human-AI collaborative workflow for scientific investigation. Though demonstrated specifically with Flow-Lenia, this methodology provides a framework potentially applicable to other parameterizable complex systems where understanding emergent collective properties is of interest.

7.4.1 Responsible AI and Algorithmic Auditing**Active Fourier Auditor for Estimating Distributional Properties of ML Models, [22]**

With the pervasive deployment of Machine Learning (ML) models in real-world applications, verifying and auditing properties of ML models have become a central concern. In this work, we focus on three properties: robustness, individual fairness, and group fairness. We discuss two approaches for auditing ML model properties: estimation with and without reconstruction of the target model under audit. Though the first approach is studied in the literature, the second approach remains unexplored. For this purpose, we develop a new framework that quantifies different properties in terms of the Fourier coefficients of the ML model under audit but does not parametrically reconstruct it. We propose the Active Fourier Auditor (AFA), which queries sample points according to the Fourier coefficients of the ML model, and further estimates the properties. We derive high probability error bounds on AFA’s estimates, along with the worst-case lower bounds on the sample complexity to audit them. Numerically we demonstrate on multiple datasets and models that AFA is more accurate and sample-efficient to estimate the properties of interest than the baselines.

When Witnesses Defend: A Witness Graph Topological Layer for Adversarial Graph Learning, [23]

Capitalizing on the intuitive premise that shape characteristics are more robust to perturbations, we bridge adversarial graph learning with the emerging tools from computational topology, namely, persistent homology representations of graphs. We introduce the concept of witness complex to adversarial analysis on graphs, which allows us to focus only on the salient shape characteristics of graphs, yielded by the subset of the most essential nodes (i.e., landmarks), with minimal loss of topological information on the whole graph. The remaining nodes are then used as witnesses, governing which higher-order graph substructures are incorporated into the learning process. Armed with the witness mechanism, we design Witness Graph Topological Layer (WGTL), which systematically integrates both local and global topological graph feature representations, the impact of which is, in turn, automatically controlled by the robust regularized topological loss. Given the attacker’s budget, we derive the important stability guarantees of both local and global topology encodings and the associated robust topological loss. We illustrate the versatility and efficiency of WGTL by its integration with five GNNs and three existing non-topological defense mechanisms. Our extensive experiments across six datasets demonstrate that WGTL boosts the robustness of GNNs across a range of perturbations and against a range of adversarial attacks, leading to relative gains of up to 18%.

Sublinear Algorithms for Wasserstein and Total Variation Distances: Applications to Fairness and Privacy Auditing, [16]

Resource-efficiently computing representations of probability distributions and the distances between them while only having access to the samples is a fundamental and useful problem across mathematical sciences. In this paper, we propose a generic algorithmic framework to estimate the PDF and CDF of any sub-Gaussian distribution while the samples from them arrive in a stream. We compute mergeable summaries of distributions from the stream of samples that require sublinear space w.r.t. the number of observed samples. This allows us to estimate Wasserstein and Total Variation (TV) distances between any two sub-Gaussian

distributions while samples arrive in streams and from multiple sources (e.g. federated learning). Our algorithms significantly improves on the existing methods for distance estimation incurring super-linear time and linear space complexities. In addition, we use the proposed estimators of Wasserstein and TV distances to audit the fairness and privacy of the ML algorithms. We empirically demonstrate the efficiency of the algorithms for estimating these distances and auditing using both synthetic and real-world datasets.

The Fair Game: Auditing & debiasing AI algorithms over time, [17]

Abstract An emerging field of AI, namely Fair Machine Learning (ML), aims to quantify different types of bias (also known as unfairness) exhibited in the predictions of ML algorithms, and to design new algorithms to mitigate them. Often, the definitions of bias used in the literature are observational, i.e. they use the input and output of a pre-trained algorithm to quantify a bias under concern. In reality, these definitions are often conflicting in nature and can only be deployed if either the ground truth is known or only in retrospect after deploying the algorithm. Thus, there is a gap between what we want Fair ML to achieve and what it does in a dynamic social environment. Hence, we propose an alternative dynamic mechanism, "Fair Game", to assure fairness in the predictions of an ML algorithm and to adapt its predictions as the society interacts with the algorithm over time. "Fair Game" puts together an Auditor and a Debiasing algorithm in a loop around an ML algorithm. The "Fair Game" puts these two components in a loop by leveraging Reinforcement Learning (RL). RL algorithms interact with an environment to take decisions, which yields new observations (also known as data/feedback) from the environment and in turn, adapts future decisions. RL is already used in algorithms with pre-fixed long-term fairness goals. "Fair Game" provides a unique framework where the fairness goals can be adapted over time by only modifying the auditor and the different biases it quantifies. Thus, "Fair Game" aims to simulate the evolution of ethical and legal frameworks in the society by creating an auditor which sends feedback to a debiasing algorithm deployed around an ML system. This allows us to develop a flexible and adaptive-over-time framework to build Fair ML systems pre- and post-deployment.

DP-SPRT: Differentially Private Sequential Probability Ratio Tests, [36]

We revisit Wald's celebrated Sequential Probability Ratio Test for sequential tests of two simple hypotheses, under privacy constraints. We propose DP-SPRT, a wrapper that can be calibrated to achieve desired error probabilities and privacy constraints, addressing a significant gap in previous work. DP-SPRT relies on a private mechanism that processes a sequence of queries and stops after privately determining when the query results fall outside a predefined interval. This OutsideInterval mechanism improves upon naive composition of existing techniques like AboveThreshold, potentially benefiting other sequential algorithms. We prove generic upper bounds on the error and sample complexity of DP-SPRT that can accommodate various noise distributions based on the practitioner's privacy needs. We exemplify them in two settings: Laplace noise (pure Differential Privacy) and Gaussian noise (Rényi differential privacy). In the former setting, by providing a lower bound on the sample complexity of any ϵ -DP test with prescribed type I and type II errors, we show that DP-SPRT is near optimal when both errors are small and the two hypotheses are close. Moreover, we conduct an experimental study revealing its good practical performance.

Some Targets Are Harder to Identify than Others: Quantifying the Target-dependent Membership Leakage, [24]

In a Membership Inference (MI) game, an attacker tries to infer whether a target point was included or not in the input of an algorithm. Existing works show that some target points are easier to identify, while others are harder. This paper explains the target-dependent hardness of membership attacks by studying the powers of the optimal attacks in a fixed-target MI game. We characterise the optimal advantage and trade-off functions of attacks against the empirical mean in terms of the Mahalanobis distance between the target point and the data-generating distribution. We further derive the impacts of two privacy defences, i.e. adding Gaussian noise and sub-sampling, and that of target misspecification on optimal attacks. As by-products of our novel analysis of the Likelihood Ratio (LR) test, we provide a new covariance attack which generalises and improves the scalar product attack. Also, we propose a new optimal canary-choosing strategy for auditing privacy in the white-box federated learning setting. Our experiments validate that the Mahalanobis score explains the hardness of fixed-target MI games.

Dimension Agnostic Testing of Survey Data Credibility through the Lens of Regression, [46]

Assessing whether a sample survey credibly represents the population is a critical question for ensuring the validity of downstream research. Generally, this problem reduces to estimating the distance between two high-dimensional distributions, which typically requires a number of samples that grows exponentially with the dimension. However, depending on the model used for data analysis, the conclusions drawn from the data

may remain consistent across different underlying distributions. In this context, we propose a task-based approach to assess the credibility of sampled surveys. Specifically, we introduce a model-specific distance metric to quantify this notion of credibility. We also design an algorithm to verify the credibility of survey data in the context of regression models. Notably, the sample complexity of our algorithm is independent of the data dimension. This efficiency stems from the fact that the algorithm focuses on verifying the credibility of the survey data rather than reconstructing the underlying regression model. Furthermore, we show that if one attempts to verify credibility by reconstructing the regression model, the sample complexity scales linearly with the dimensionality of the data. We prove the theoretical correctness of our algorithm and numerically demonstrate our algorithm’s performance.

7.4.2 Formalization of mathematics

Formalization of Brownian motion in Lean, [50]

Brownian motion is a building block in modern probability theory. In this paper, we describe a formalization of Brownian motion using the Lean theorem prover. We build on the existing measure-theoretic foundations in Lean’s mathematical library, Mathlib, and we develop several key components needed for the construction of Brownian motion, including the Carathéodory and Kolmogorov extension theorems, Gaussian measures in Banach spaces, and the Kolmogorov-Chentsov theorem for path continuity.

Markov kernels in Mathlib’s probability library, [49]

The probability folder of Mathlib, Lean’s mathematical library, makes a heavy use of Markov kernels. We present their definition and properties and describe the formalization of the disintegration theorem for Markov kernels. That theorem is used to define conditional probability distributions of random variables as well as posterior distributions. We then explain how Markov kernels are used in a more unusual way to get a common definition of independence and conditional independence and, following the same principles, to define sub-Gaussian random variables. Finally, we also discuss the role of kernels in our formalization of entropy and Kullback-Leibler divergence.

8 Bilateral contracts and grants with industry

Participants: Odalric-Ambrym Maillard, Philippe Preux, Mickael Basson, Yann Berthelot, Anthony Kobanda.

8.1 Bilateral contracts with industry

- contract with Ubisoft, 2023–2026: PI: Odalric-Ambrym Maillard.
This contract is related to Anthony Kobandas Ph.D. “Continual Reinforcement Learning with changing environments: Application to Video Games”
- contract with Lilly Group, 2023–2026, PI: Philippe Preux.
This contract is related to Mickael Basson’s Ph.D. “Reinforcement learning to solve combinatorial optimization problems”.
- contract with Saint-Gobain Research, 2023–2026, PI: Philippe Preux.
This contract is related to Yann Berthelot’s Ph.D. “Reinforcement learning for advanced control of industrial processus”.

9 Partnerships and cooperations

Participants: Philippe Preux, Odalric-Ambrym Maillard, Emilie Kaufmann, Remy De-
genne, Debabrota Basu, Riadh Akrou, Timothee Mathieu, Hector Kohler,
Penanklihi Cyrille Kone.

9.1 International initiatives

9.1.1 Inria associate team not involved in an IIL or an international program

SeRAI

Title: Sequential Testing and Learning Algorithms for Verifiably Robust and Responsible AI

Duration: 2025 -> 2027

Coordinator: Arijit Ghosh (arijitiitkgpster@gmail.com)

Partners:

- Indian Statistical Institute , Calcutta (Inde)

Inria contact: Debabrota Basu

Summary: Artificial Intelligence (AI) and Machine Learning (ML) have emerged as the technologies of our times, and presently, they are getting widely deployed in real-life applications with socioeconomic consequences. The reckoning of AI/ML has motivated development of robust and responsible AI algorithms to ensure social alignment of them, and also auditing algorithms to verify different properties of the AI/ML systems before and after deployment. Instead of the plethora of existing robust and responsible AI/ML algorithms, recent research has exposed a gap between what existing algorithms achieve in terms of a socioeconomic indicator of interest and what we socially want them to achieve. Internationally, it has led to a surge in verifying and auditing different properties of the ML and AI algorithms, where EU AI regulations push the frontiers. Distribution and property testing play a pivotal role in ensuring ethical and transparent AI operations, aligning with established guidelines and societal norms. These tests facilitate the scrutiny of data distributions used in AI model training, critical for upholding ethical and fairness standards. They further allow verification of biases and other unintended behaviours that may lead to adverse consequences. Presently, there are two complementary approaches to these testing problems: (a) statistical learning based methods, like using online learning, sequential tests, active learning etc., and (b) complexity theory-based and formal methods, like query and memory complexities of testing Boolean functions and properties of graphs, and distributions corresponding to their input-output domains. While Scool team is known for their expertise in the statistics-based approach, the Indian Statistical Institute team is known for their expertise in the other one. In SeRAI, our scientific aim is to marry these two complementary approaches (statistics and formal methods) to AI and harness their strengths to create verifiable and responsible but also efficient AI/ML algorithms. Specifically, we aim to explore three research directions. 1. Sample-efficient auditing bias of real-life survey data from the lens of different learning models with and without privacy constraints. 2. Statistically and computationally efficient auditing of different statistical assumptions on structured distributions required to design theoretically provable ML and RL algorithms. 3. Statistically and computationally efficient auditing of properties of large-scale ML models with high-dimensional data under structures, like symmetry, sparsity, and permutation-invariance. Thanks to the complementary expertise of the partners, the SeRAI associate team aims to contribute to the aforementioned three problems and design statistically and computationally efficient distribution and property testing algorithms under sequential interactions, which are frugal and practical in the pursuit of a verifiable and responsible AI.

9.2 International research visitors

9.2.1 Visits of international scientists

Other international visits to the team

Ahana Deb**Status** PhD**Institution of origin:** University of Pompeu Fabra**Country:** Spain**Dates:** April to July 2025

Context of the visit: Many real world AI applications involve the use of Reinforcement Learning (RL), such as medical trials, drug design, recommendation systems, automated robots, self-driving cars, etc. However most of the RL problems are modelled on a Markovian assumption which does not account for any dependency on the history of the agent-environment interaction. But such dependencies are often found in many real life scenarios, e.g., patient histories (including symptoms, test results, previous treatments, etc) always play a crucial role in diagnosis and consequent treatment. During her visit, we studied the challenges and solutions of designing a computationally-statistically efficient algorithm for these problems and deriving the corresponding convergence analysis.

Mobility program/type of mobility: Research visit funded by ELIAS PhD mobility grant.

Pratik Karmakar**Status** PhD**Institution of origin:** National University of Singapore**Country:** Singapore**Dates:** April 2025

Context of the visit: Pratik presented his work with Pierre Senellart (ENS, Paris) on ProvSQL: Provenance and Probabilistic Querying in Uncertain Databases.

Mobility program/type of mobility: Research visit.

9.2.2 Visits to international teams**Research stays abroad****Hector Kohler****Visited institution:** University of Alberta**Country:** Canada**Dates:** Jul-Aug.

Context of the visit: Hector visited the RLAI Lab, a highly recognized lab in reinforcement learning.

Mobility program/type of mobility: research stay.

Cyrille Koné**Visited institution:** University of Washington**Country:** United States**Dates:** June-August

Context of the visit: Cyrille visited the group of Kevin Jamieson to work on best policy identification in reinforcement learning.

Mobility program/type of mobility: research stay founded by UW.

9.3 National initiatives

9.3.1 ANR projects

Scool is involved in 5 ANR projects:

- ANR JCJC **FATE**, PI: Remy Degenne, 2023–2027
- ANR JCJC **REPUBLIC**, PI: Debabrota Basu, 2023–2026
- ANR **BIP-UP**, partnership: Scool/Inserm (CHU de Lille), PI: Adrien Prevost, 2023–2026.
- ANR JCJC NeuRL, PI: Riadh Akrou, 2024–2028
- ANR JCJC STRESS, PI: Timothee Mathieu, 2025–2029

9.3.2 PEPR projects

Scool is involved in 2 PEPR:

- PEPR AI: project **FOUNDRY**, local head: Emilie Kaufmann (description below);
- PEPR « **Agroécologie et numérique** », **PI@ntAgroEco**, local head: Odalric-Ambrym Maillard

Title: FOUNDRY

Duration: July 2024 → June 2028

Coordinator: Panayotis Mertikopoulos, Polaris, Univ. Grenoble Alpes

Partners:

- POLARIS: a joint research team between the CNRS, Inria, and Univ. Grenoble Alpes.
- ENS Lyon: faculty from the pure and applied mathematics department of ENS Lyon.
- Inria FAIRPLAY: a joint team between Criteo, IP Paris (ENSAE and Ecole Polytechnique), and Inria.
- LTCI: the informations and communications laboratory of Télécom Paris.
- MILES: the machine intelligence and learning systems of the LAMSADE lab at Paris Dauphine.
- Inria Scool

Inria contact: Emilie Kaufmann

Summary: From automated hospital admission systems powered by machine learning (ML), to flexible chatbots capable of fluent conversations and self-driving cars, the wildfire spread of artificial intelligence (AI) has brought to the forefront a crucial question with far-reaching ramifications for the society at large: Can ML systems and models be relied upon to provide trustworthy output in high-stakes, mission-critical environments? These questions invariably revolve around the notion of *robustness*, an operational desideratum that has eluded the field since its nascent stages. One of the main reasons for this is the fact that ML models and systems are typically data-hungry and highly sensitive to their training input, so they tend to be brittle, narrow-scoped, and unable to adapt to situations that go beyond their training envelope. On that account, the core vision of the proposed research is that robustness cannot be achieved by blindly throwing more data and computing power to larger and larger models with exponentially growing energy requirements (and a commensurate carbon footprint to boot). Instead, our proposal intends to rethink and develop the core theoretical and methodological FOUNDations of Robustness and reliabilitY (FOUNDRY) that are needed to build and instill trust in ML-powered technologies and systems from the ground up.

Title: PI@ntAgroEco

Duration: July 2024 → June 2028

Coordinator: Alexis Joly, Inria Zenith, and Pierre Bonnet CIRAD, AMAP.

Partners:

- INRAE
- INRIA
- IRD
- CIRAD
- Tela Botanica
- Université de Montpellier
- Université Paris-Saclay

Inria contact: Odalric-Ambrym Maillard

Summary: Agroecology necessarily involves crop diversification, but also the early detection of diseases, deficiencies and stresses (hydric, etc.), as well as better management of biodiversity. The main stumbling block is that this paradigm shift in agricultural practices requires expert skills in botany, plant pathology and ecology that are not generally available to those working in the field, such as farmers or agri-food technicians. Digital technologies, and artificial intelligence in particular, can play a crucial role in overcoming this barrier to access to knowledge.

The aim of the PI@ntAgroEco project will be to design, experiment with and develop new high-impact agro-ecology services within the PI@ntNet platform. This includes :

- research in AI and plant sciences ;
- agile development of new components within the platform;
- organization of participatory science programs and animation of the PI@ntNet user community.

Ce programme de travail a pour but de produire une amélioration de la détection et reconnaissance des maladies végétales, de l'identification des niveaux infraspécifiques. Il permettra le développement d'outils d'estimation de la sévérité des symptômes, carences, stades de déclin et stress hydrique ou de caractérisation des associations d'espèces à partir d'images multi-spécimens. Il améliorera la connaissance des espèces.

Le projet PI@ntAgroEco rassemble des forces complémentaires en matière de recherche, de développement et d'animation. S'ajouteront à l'équipe pluridisciplinaire chargée de la plateforme PI@ntNet de nouvelles forces de recherche ayant une expertise reconnue dans les sciences participatives. Le consortium rassemblera 10 partenaires incluant des organismes de recherche, des universités, des acteurs de la société civile et des partenaires internationaux.

9.3.3 Other projects in France

Scool is involved in the Regalia pilot-project.

Other collaborations:

- L. Richert, R. Thiébaud, Inria SISTM, Bordeaux, bandits for vaccine clinical trials.
- W. M. Koolen, CWI Amsterdam & University of Twente, concentration of information divergences.

9.3.4 Inria Exploratory Actions

Emilie Kaufmann obtained an Action Exploratoire grant BETA-3K (Bandit Exploration for Treatment Allocation in Phase III Trials with $K > 2$ arms) in 2025 to start a collaboration with the University of Cambridge (MRC Biostatistics Unit, team of Sofia Villar) on adaptive clinical trials.

Effective start of the Action Exploratoire AuDaCiTi (Autonomous Data Collection and Labelling Through Interaction) with the hiring of Hadrien Crassous as a PhD student in November 2025, under the supervision of Riadh Akrouf. AEx in collaboration with the Robot Learning group of Joni Pajarinen at Aalto University.

Remy Degenne obtained an Action Exploratoire grant FORMAL (formal proofs for machine learning).

10 Dissemination

Participants: Philippe Preux, Odalric-Ambrym Maillard, Emilie Kaufmann, Remy Degenne, Debabrota Basu, Riadh Akrou, Timothee Mathieu, Juliette Achddou, Julien Teigny, Adrienne Tuynman.

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

- Debabrota Basu co-organized with A. Gilra of CWI Amsterdam a research semester program on “*Control Theory and Reinforcement Learning: Connections and Challenges*”. It included
 1. Spring School on Control Theory and Reinforcement Learning, 17-21 March, 2025.
 2. Workshop on Themes across Control and Reinforcement Learning, 24-25 March, 2025.
 3. Workshop on Modern Applications of Control Theory and Reinforcement Learning, 20-21 May, 2025.
 4. Workshop on Theory of Control and Reinforcement Learning, 19-20 June, 2025.
- Odalric-Ambrym Maillard: Organization of “Séminaire Itinérant” of the AI transversal Axis of the CRIStaL laboratory, together with M. Keller (Inria Magnet)

10.1.2 Scientific events: selection

Member of the conference program committees

- Debabrota Basu: member of the PC at AAAI, PETS, CCS.
- Remy Degenne: member of the PC at ALT.
- Emilie Kaufmann: member of the PC at COLT.
- Philippe Preux: member of the PC at IJCAI and ECML.

Reviewer

- Juliette Achddou: reviewer at NeurIPS
- Riadh Akrou: reviewer at ICML, NeurIPS, ICLR
- Debabrota Basu: reviewer at ICML, NeurIPS, AISTATS, AAAI, EWRL, AAMAS.
- Remy Degenne: reviewer at COLT, ICML, ALT
- Emilie Kaufmann: reviewer at AISTATS, NeurIPS.
- Odalric-Ambrym Maillard: reviewer at RLC, ACML, EWRL.
- Timothee Mathieu: reviewer at COLT, AISTATS and EWRL

10.1.3 Journal

Reviewer - reviewing activities

- Debabrota Basu: reviewer for JMLR, TMLR, IEEE TPAMI, IEEE Access, ACM Journal on Responsible Computing, IEEE TAI, IEEE TKDE, IEEE Information Theory, Communications in Statistics – Theory and Methods.
- Remy Degenne: reviewer for JMLR, Annals of Formalized Mathematics.
- Emilie Kaufmann: reviewer for JMLR.
- Odalric-Ambrym Maillard: reviewer for JMLR, Mathematics of Operation Research.
- Timothee Mathieu: reviewer for Biometrika, Annals of Statistics, ALEA, JMLR, JRSSB.

10.1.4 Invited talks

- Debabrota Basu: Exploration–Exploitation Dilemma in RL: Bridging Theory-to-Practice Gaps, Centre for AI, IIT Delhi, September 2025.
- Debabrota Basu: When Privacy meets Partial Information: Privacy-Utility Trade-offs in Bandits, CNI Seminars, IISc Bangalore, September 2025.
- Debabrota Basu: Actors & Critics: Function Approximation & Policy Gradients in RL, Spring School on Control Theory & RL, CWI Amsterdam, March 2025.
- Debabrota Basu: Exploration–Exploitation in RL: Calibrated Optimism in Face of Uncertainty, Spring School on Control Theory & RL, Amsterdam, March 2025.
- Remy Degenne: Lean for PDEs workshop, Simons Laufer Mathematical Institute, Berkeley, California, October 2025.
- Remy Degenne: Stochastic Analysis and Mathematical Finance seminar, Oxford, November 2025.
- Remy Degenne: ItaLean conference on Bridging Formal Mathematics and AI, Bologna, December 2025.
- Remy Degenne: Imperial Lean study group, London, December 2025.
- Emilie Kaufmann: Colloquium Polaris, Lille, February 2025.
- Emilie Kaufmann: ENSAE Statistics seminar, Saclay, May 2025.
- Emilie Kaufmann: Colloquium of the MAP 5, Paris, June 2025.
- Emilie Kaufmann: RL Theory Workshop, CWI, Amsterdam, June 2025.
- Emilie Kaufmann: Workshop on Regret, Optimization and Games, IHP, Paris, November 2025.
- Emilie Kaufmann: Symposium de l'Association d'Informatique Medicale (AIM), Lille, November 2025.
- Emilie Kaufmann: Algorithmic Statistics Workshop, Oxford, November 2025.
- Odalric-Ambrym Maillard: talk at Colloque L'Agriculture au prisme des data-sciences organized by Alliance Harvest, regarding Pl@ntAgroEco project and collaboration in Agroecology, Palaiseau, February 2025.
- Odalric-Ambrym Maillard: talk at Inria Breizh-Carnot festival, Rennes, November 2025.

10.1.5 Scientific expertise

- Odalric-Ambrym Maillard: evaluation of an ANR JCJC project.
- Philippe Preux:
 - evaluation of 2 ANR project proposals.
 - member of the scientific committee of the MathNum department at Inrae
 - member of the scientific committee of **PEPR agroécologie et numérique**
 - member of the scientific and ethical committee of INCLUDE (data warehouse of CHU Lille).
 - member of the **éthique en commun** joint committee of Inrae, IRD, Ifremer, and Cirad.
- Adrienne Tuynman: evaluation of bachelor programs of the University of Science and Technology of Hanoi, Vietnam, as part of a committee of the international section of HCERES

10.1.6 Research administration

- Odalric-Ambrym Maillard
 - Scientific coordinator of **AI transversal Axis** of CRIStaL laboratory, together with M. Keller (Inria Magnet)
- Philippe Preux:
 - scientific coordinator of **CPER Cornelia**
 - member of the DAS¹ health at Région Hauts-de-France
 - member of the **“Hub Santé” board of the “Initiative d’excellence”**
 - member of the BCEP at Inria Lille.

10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

- Juliette Achddou: “Introduction au Deep Learning”, M2 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Machine Learning”, M1 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Algorithmique numérique pour l’optimisation”, M1 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Classification supervisée”, L3 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Régression Linéaire”, L3 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Data Mining”, L3 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Algèbre Linéaire Numérique”, L3 Informatique et Statistiques, Polytech’ Lille.
- Juliette Achddou: “Introduction aux Environnements Virtuels”, M1 Informatique et Statistiques, Polytech’ Lille.
- Riadh Akrou: “Option Machine Learning”, L3 Informatique, Université de Lille.
- Riadh Akrou: “Sequential Decision Making”, M2 Data Science, Ecole Centrale de Lille.
- Debabrota Basu: “Sequential Decision Making”, M2 in Data Science, Centrale Lille and Université de Lille.
- Debabrota Basu: “Research Reading Group”, M2 in Data Science, Centrale Lille and Université de Lille.

¹strategic activity domain

- Debabrota Basu: “Advanced Machine Learning and Decision Making”, Centrale Lille
- Remy Degenne: “Sequential Learning”, Master MVA, ENS Paris-Saclay.
- Remy Degenne: “Sequential Learning”, Centrale Lille.
- Emilie Kaufmann: “Statistics 2”, M1 Data Science, Ecole Centrale de Lille.
- Odalric-Ambrym Maillard: “Reinforcement Learning Research Challenges”, Executive Master Ecole Polytechnique.
- Odalric-Ambrym Maillard: “Reinforcement Learning for the Industry”, Inria Academy.
- Philippe Preux: “Prise de décision séquentielle dans l’incertain”, M2 in Computer Science, Université de Lille.
- Philippe Preux: “Apprentissage par renforcement”, M2 in Computer Science, Université de Lille.
- Philippe Preux: “Science des données II”, L3 MIASHS, Université de Lille.
- Philippe Preux: “Science des données III”, L3 MIASHS, Université de Lille.
- Philippe Preux: “Réseaux de neurones”, L1 Maths-Informatique, Université de Lille.
- Philippe Preux: “IA et apprentissage automatique”, DU IA & Santé, Université de Lille.
- Adrienne Tuynman: “Cryptographie” (practical sessions), M1 in Applied Mathematics, University of Lille

10.2.1 Supervision

- Riadh Akrou:
 - Ph.D. students: Brahim Driss, Hadrien Crassous
 - M2 Research Internship: Francois Muller
- Debabrota Basu and Emilie Kaufmann: Ph.D. student: Thomas Michel
- Debabrota Basu and Odalric-Ambrym Maillard: Ph.D. student: Udvas Das
- Remy Degenne and Emilie Kaufmann : Ph.D. students: Redouane Yagouti, Adrienne Tuynman
- Emilie Kaufmann : Ph.D. student: Penanklihi Cyrille Kone
- Timothee Mathieu and Odalric-Ambrym Maillard: Ph.D. student: Adrien Prevost
- Odalric-Ambrym Maillard: Ph.D. students: Sumit Vashishtha, Anthony Kobanda, Waris Radji.
- Philippe Preux:
 - Ph.D. students: Matheus Medeiros Centa, Mickael Basson
- Philippe Preux and Riadh Akrou: Ph.D. students: Hector Kohler, Yann Berthelot
- Philippe Preux and Emilie Kaufmann: Ph.D. student: Thomas Michel
- Philippe Preux and Debabrota Basu: Ph.D. students: Ayoub Ajarra

10.2.2 Juries

- Riadh Akrou: Ph.D. defenses: Hector Kohler (supervisor)
- Debabrota Basu:
 - Ph.D. defense: Y. Wang (Inria Lille and Orange)
 - CS: G. Richardeau (Inria Rennes, Univ. Rennes)
- Emilie Kaufmann:
 - Ph.D. defenses: A. Rio (Université de Grenoble, reviewer), D. Tiapkin (CMAP, Ecole Polytechnique, reviewer), C. Fiegel (ENSAE), R. Zhang (LSS, CentraleSupélec, reviewer), A. Gouverneur (KTH).
- Odalric-Ambrym Maillard:
 - HdR defense: R. Combes (Centrale-Supélec, Palaiseau, rapporteur),
 - Ph.D. defenses: F. Morri (Inria, U. Lille, examinateur), O. Rossini (IMAG, U. Montpellier, examinateur), S. Lindstahl (KTH, Stockholm, opposant), G.J. Molina (UPF, Barcelona, examinateur), F. Fabre (U. Reunion, co-superviseur).
 - CSI: Q.L Ta. (Intitut Polytechnique de Paris, rapporteur).
 - CR hiring committee of the Mathnum department at Inrae (23,24,25,26).
- Philippe Preux:
 - Participation to a hiring committee for an associate professor position in computer science applied to humanities at Paris-Sorbonne.
 - Ph.D. defenses: Th. Firmin (Université de Lille), P-A. Le Tolguenec (Toulouse, reviewer), M. Zouitine (Toulouse), Hector Kohler (supervisor), F. Fabre-Ferber (La Réunion, reviewer)
 - HdR defense: Riadh Akrou.

10.2.3 Educational and pedagogical outreach

- E. Kaufmann participated to a table ronde "Femmes et mathématiques" organized on Pi Day (03/14) at LILLIAD and targeting undergrad students at the University of Lille

10.3 Popularization

10.3.1 Productions (articles, videos, podcasts, serious games, ...)

- Timothee Mathieu: Video for the Summer of Maths Exposition on Sequential testing ([link to the video](#)).
- Emilie Kaufmann participated to the book "Tout comprendre (ou presque) sur l'intelligence artificielle" by Olivier Cappé and Claire Marc at CNRS Editions (chapter 12).
- Debabrota Basu drafted a book chapter in the "AI Bias in Education" book, which got highlighted in Forbes' 2025 reading list on AI.

10.3.2 Participation in Live events

- Emilie Kaufmann gave a presentation “Tout Comprendre (ou presque) sur l’Intelligence artificielle” and organized a “table ronde” on the same topic for high-school students and maths teacher at LILLIAD during the NSI (Numérique Science Informatique week) week ([link](#)). T. Michel (PhD student) participated to the table ronde.
- Philippe Preux gave a presentation on “AI, for the best, and the worst” at the “Fête de la science”
- Julien Teigny:
 - 1 hour presentation on the “alignment of AIs” to:
 - * a group of ENS Rennes students visiting Lille
 - * teachers of the “académie de Lille”
 - presentation of the [bariatric surgery website](#) to a group of ENS Rennes students,
 - presentation on AI using a game to “collégiens 3è”,
 - co-supervision of a serious game jam on “Equité en Intelligence Artificielle” for people of INSPE ([link to the video](#)),
- Odalric-Ambrym Maillard: Interview for France Culture "8h45", regarding AI and agroecology, February 26 2025.

10.3.3 Others science outreach relevant activities

- Philippe Preux was interviewed by senator A. Basquin on AI.

11 Scientific production

11.1 Major publications

- [1] B. Balle and O.-A. Maillard. ‘Spectral Learning from a Single Trajectory under Finite-State Policies’. In: *International conference on Machine Learning*. Proceedings of the International conference on Machine Learning. Sidney, France, July 2017. URL: <https://hal.archives-ouvertes.fr/hal-01590940>.
- [2] D. Baudry, R. Gautron, E. Kaufmann and O.-A. Maillard. ‘Optimal Thompson Sampling strategies for support-aware CVaR bandits’. In: 38th International Conference on Machine Learning. proceedings of machine learning research. Virtual, United States, 18th July 2021. URL: <https://hal.science/hal-03447244>.
- [3] L. Besson and E. Kaufmann. ‘Multi-Player Bandits Revisited’. In: *Algorithmic Learning Theory*. Mehryar Mohri and Karthik Sridharan. Lanzarote, Spain, Apr. 2018. URL: <https://hal.inria.fr/hal-01629733>.
- [4] G. Dulac-Arnold, L. Denoyer, P. Preux and P. Gallinari. ‘Sequential approaches for learning datum-wise sparse representations’. In: *Machine Learning* 89.1-2 (1st Oct. 2012), pp. 87–122. DOI: [10.1007/s10994-012-5306-7](https://doi.org/10.1007/s10994-012-5306-7). URL: <https://hal.inria.fr/hal-00747724>.
- [5] Y. Flet-Berliac and P. Preux. ‘Only Relevant Information Matters: Filtering Out Noisy Samples to Boost RL’. In: *IJCAI 2020 - International Joint Conference on Artificial Intelligence*. Yokohama, Japan, July 2020. DOI: [10.24963/ijcai.2020/376](https://doi.org/10.24963/ijcai.2020/376). URL: <https://hal.inria.fr/hal-02091547>.
- [6] A. Garivier and E. Kaufmann. ‘Optimal Best Arm Identification with Fixed Confidence’. In: *29th Annual Conference on Learning Theory (COLT)*. Vol. 49. JMLR Workshop and Conference Proceedings. New York, United States, June 2016. URL: <https://hal.archives-ouvertes.fr/hal-01273838>.
- [7] B. Ghosh, D. Basu and K. S. Meel. ‘Justicia: A Stochastic SAT Approach to Formally Verify Fairness’. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Conference on Artificial Intelligence. Vol. 35. Proceedings of the AAAI Conference on Artificial Intelligence 9. Virtual, Canada, Feb. 2021, pp. 7554–7563. URL: <https://hal.science/hal-03445831>.

- [8] M. Jourdan, R. Degenne, D. Baudry, R. de Heide and E. Kaufmann. ‘Top Two Algorithms Revisited’. In: *NeurIPS 2022 - 36th Conference on Neural Information Processing System*. Advances in Neural Information Processing Systems. New Orleans, United States, 28th Nov. 2022. URL: <https://hal.science/hal-03825103>.
- [9] H. Kadri, E. Duflos, P. Preux, S. Canu, A. Rakotomamonjy and J. Audiffren. ‘Operator-valued Kernels for Learning from Functional Response Data’. In: *Journal of Machine Learning Research* 17.20 (2016), pp. 1–54. URL: <https://hal.archives-ouvertes.fr/hal-01221329>.
- [10] E. Kaufmann and W. M. Koolen. ‘Monte-Carlo Tree Search by Best Arm Identification’. In: *NIPS 2017 - 31st Annual Conference on Neural Information Processing Systems*. Advances in Neural Information Processing Systems. Long Beach, United States, Dec. 2017, pp. 1–23. URL: <https://hal.archives-ouvertes.fr/hal-01535907>.
- [11] E. Kaufmann, P. Ménard, O. Darwiche Domingues, A. Jonsson, E. Leurent and M. Valko. ‘Adaptive reward-free exploration’. In: *Algorithmic Learning Theory*. Paris, France, 2021. URL: <https://hal.science/hal-02864574>.
- [12] O.-A. Maillard. ‘Boundary Crossing Probabilities for General Exponential Families’. In: *Mathematical Methods of Statistics* 27 (2018). URL: <https://hal.archives-ouvertes.fr/hal-01737150>.
- [13] O.-A. Maillard, H. Bourel and M. S. Talebi. ‘Tightening Exploration in Upper Confidence Reinforcement Learning’. In: *International Conference on Machine Learning*. Vienna, Austria, July 2020. URL: <https://hal.archives-ouvertes.fr/hal-03000664>.
- [14] T. Mathieu, R. Della Vecchia, A. Shilova, M. Medeiros Centa, H. Kohler, O.-A. Maillard and P. Preux. ‘AdaStop: adaptive statistical testing for sound comparisons of Deep RL agents’. In: *Transactions on Machine Learning Research Journal* (2024). URL: <https://inria.hal.science/hal-04132861>.
- [15] F. Pesquerel and O.-A. Maillard. ‘IMED-RL: Regret optimal learning of ergodic Markov decision processes’. In: *NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems*. Thirty-sixth Conference on Neural Information Processing Systems. New-Orleans, United States, 28th Nov. 2022. URL: <https://hal.science/hal-03825423>.

11.2 Publications of the year

International journals

- [16] D. Basu and D. Chanda. ‘Sublinear Algorithms for Estimating Wasserstein and Total Variation Distances: Applications to Fairness and Privacy Auditing’. In: *Transactions on Machine Learning Research Journal* (21st Jan. 2026). URL: <https://hal.science/hal-05027482> (cit. on p. 15).
- [17] D. Basu and U. Das. ‘The Fair Game: Auditing & debiasing AI algorithms over time’. In: *Cambridge Forum on AI: Law and Governance* 1 (4th June 2025), p. 27. DOI: [10.1017/cfl.2025.8](https://doi.org/10.1017/cfl.2025.8). URL: <https://hal.science/hal-05111387> (cit. on p. 16).
- [18] R. M. Nkhumise, D. Basu, T. J. Prescott and A. Gilra. ‘Studying Exploration in RL: An Optimal Transport Analysis of Occupancy Measure Trajectories’. In: *Transactions on Machine Learning Research Journal* (May 2025). URL: <https://hal.science/hal-04702986> (cit. on p. 13).
- [19] W. Radji and O.-A. Maillard. ‘The Confusing Instance Principle for Online Linear Quadratic Control’. In: *Reinforcement Learning Journal* 6 (5th Aug. 2025), pp. 811–828. URL: <https://hal.science/hal-05322160> (cit. on p. 11).
- [20] E. C. Vasconcellos, Á. P. F. Negreiros, A. P. D. de Araújo, R. Guerra, P. Preux, D. H. dos Santos, L. M. G. Gonçalves and E. W. G. Clua. ‘Yara: An Ocean Virtual Environment for Research and Development of Autonomous Sailing Robots and Other Unmanned Surface Vessels’. In: *Journal of Intelligent and Robotic Systems* 111.3 (5th July 2025), p. 78. DOI: [10.1007/s10846-024-02212-1](https://doi.org/10.1007/s10846-024-02212-1). URL: <https://inria.hal.science/hal-05148829> (cit. on p. 14).
- [21] S. Vashishtha and O.-A. Maillard. ‘Leveraging priors on distribution functions for multi-arm bandits’. In: *Reinforcement Learning Journal* 6 (2025), pp. 1600–1623. DOI: [10.48550/arXiv.2503.04518](https://doi.org/10.48550/arXiv.2503.04518). URL: <https://hal.science/hal-05310340>. In press (cit. on p. 9).

International peer-reviewed conferences

- [22] A. Ajarra, B. Ghosh and D. Basu. ‘Active Fourier Auditor for Estimating Distributional Properties of ML Models’. In: AAI Conference on Artificial Intelligence. Philadelphia, United States, Feb. 2025. URL: <https://hal.science/hal-04733059> (cit. on p. 15).
- [23] N. A. Arafat, D. Basu, Y. Gel and Y. Chen. ‘When Witnesses Defend: A Witness Graph Topological Layer for Adversarial Graph Learning’. In: AAI Conference on Artificial Intelligence. Philadelphia, United States, Feb. 2025. URL: <https://inria.hal.science/hal-04708183> (cit. on p. 15).
- [24] A. Azize and D. Basu. ‘Some Targets Are Harder to Identify than Others: Quantifying the Target-dependent Membership Leakage’. In: AISTATS 2025 – International Conference on Artificial Intelligence and Statistics. Phuket, Thailand, May 2025. URL: <https://hal.science/hal-04615701> (cit. on p. 16).
- [25] A. Azize, Y. Wu, J. Honda, F. Orabona, S. Ito and D. Basu. ‘Optimal Regret of Bandits under Differential Privacy’. In: NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems. San Diego (USA), United States, 2025. URL: <https://hal.science/hal-05111413> (cit. on p. 10).
- [26] M. Basson and P. Preux. ‘Improving Diffusion Models for the Traveling Salesman Problem (TSP) by Leveraging the Structure of the Solution Space’. In: *Lecture Notes in Computer Science*. 11th Annual Conference on machine Learning, Optimization and Data science (LOD 2025). LNCS. Riva del Sole, Toscane, Italy, 2025. URL: <https://hal.science/hal-05396558> (cit. on p. 13).
- [27] E. Biré, A. Kobanda, L. Denoyer and R. Portelas. ‘Efficient Active Imitation Learning with Random Network Distillation’. In: ICLR. Singapore, Singapore, 14th Apr. 2025. URL: <https://hal.science/hal-05126339> (cit. on p. 14).
- [28] S. Chakraborty, S. Roy and D. Basu. ‘FLIPHAT: Joint Differential Privacy for High Dimensional Sparse Linear Bandits’. In: AISTATS 2025 – International Conference on Artificial Intelligence and Statistics. Phuket, Thailand, May 2025. URL: <https://hal.science/hal-04615697> (cit. on p. 10).
- [29] U. Das and D. Basu. ‘Learning to Explore with Lagrangians for Bandits under Unknown Linear Constraints’. In: Twenty-Ninth Annual Conference on Artificial Intelligence and Statistics (AISTATS). Tangier, Morocco, 2nd May 2025. URL: <https://hal.science/hal-04784911>.
- [30] U. Das, A. Shukla and D. Basu. ‘FraPPE: Fast and Efficient Preference-based Pure Exploration’. In: NeurIPS 2025 - 39th Annual Conference on Neural Information Processing Systems. San Diego (USA), United States, Dec. 2025. URL: <https://hal.science/hal-05308357> (cit. on p. 9).
- [31] R. Della Vecchia and D. Basu. ‘Stochastic Online Instrumental Variable Regression: Regrets for Endogeneity and Bandit Feedback’. In: AAI Conference on Artificial Intelligence. Philadelphia, United States, Feb. 2025. URL: <https://hal.science/hal-03831210> (cit. on p. 11).
- [32] H. Kohler, R. Akrouf and P. Preux. ‘Breiman meets Bellman: Non-Greedy Decision Trees with MDPs’. In: *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. KDD 2025 - The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Vol. 2. Toronto, Canada: ACM, 2025, pp. 1207–1218. DOI: [10.1145/3711896.3736868](https://doi.org/10.1145/3711896.3736868). URL: <https://hal.science/hal-05097382> (cit. on p. 11).
- [33] C. Kone, M. Jourdan and E. Kaufmann. ‘Pareto Set Identification With Posterior Sampling’. In: *Proceedings of Machine Learning Research*. AISTATS 2025 - 28th International Conference on Artificial Intelligence and Statistics. Phuket, Thailand, 5th May 2025. URL: <https://hal.science/hal-05125257> (cit. on p. 9).
- [34] C. Kone, E. Kaufmann and L. Richert. ‘Bandit Pareto Set Identification in a Multi-Output Linear Model’. In: *Proceedings of Machine Learning Research*. AISTATS 2025 - 28th International Conference on Artificial Intelligence and Statistics. Phuket, Thailand, 5th May 2025. URL: <https://hal.science/hal-05125311> (cit. on p. 10).
- [35] C. Kone, E. Kaufmann and L. Richert. ‘Constrained Pareto Set Identification with Bandit Feedback’. In: *Proceedings of Machine Learning Research*. ICML 2025 - 42nd International Conference on Machine Learning. Vancouver, Canada, 17th July 2025. URL: <https://hal.science/hal-05125339> (cit. on p. 10).

- [36] T. Michel, D. Basu and E. Kaufmann. ‘DP-SPRT: Differentially Private Sequential Probability Ratio Tests’. In: Twenty-Ninth Annual Conference on Artificial Intelligence and Statistics (AISTATS). Tangier, Morocco, 2nd May 2026. doi: [10.48550/arXiv.2508.06377](https://doi.org/10.48550/arXiv.2508.06377). URL: <https://inria.hal.science/hal-05230052> (cit. on p. 16).
- [37] T. Michel, M. Cvjetko, G. Hamon, P.-Y. Oudeyer and C. Moulin-Frier. ‘Exploring Flow-Lenia Universes with a Curiosity-driven AI Scientist: Discovering Diverse Ecosystem Dynamics’. In: *Artificial Life Conference Proceedings 37*. ALIFE 2025 - Conference on Artificial Life. Vol. 2025. 1. Kyoto / Virtual, Japan, 2025, p. 68. doi: [10.1162/ISAL.a.896](https://doi.org/10.1162/ISAL.a.896). URL: <https://hal.science/hal-05333302> (cit. on p. 15).
- [38] A. Mukherjee, M. Bullo, D. Basu and D. Gündüz. ‘Test-time Verification via Optimal Transport: Coverage, ROC, & Sub-optimality’. In: The Fourteenth International Conference on Learning Representations (ICLR). Rio de Janeiro (BRAZIL), Brazil, 21st Oct. 2025. URL: <https://hal.science/hal-05520715>.
- [39] R. Poiani, M. Jourdan, E. Kaufmann and R. Degenne. ‘Best-Arm Identification in Unimodal Bandits’. In: *Proceedings of Machine Learning Research*. AISTATS 2025 - 28th International Conference on Artificial Intelligence and Statistics. Phuket, Thailand, 5th May 2025. URL: <https://hal.science/hal-05125233> (cit. on p. 8).
- [40] A. Tuynman and R. Degenne. ‘The Batch Complexity of Bandit Pure Exploration’. In: *PMLR*. ICML 2025 - 42nd International Conference on Machine Learning. Vol. 267. Vancouver, Canada, 2025, pp. 60442–60468. URL: <https://inria.hal.science/hal-05305415> (cit. on p. 8).

Conferences without proceedings

- [41] B. Driss, A. Davey and R. Akrou. ‘PB²: Preference Space Exploration via Population-Based Methods in Preference-Based Reinforcement Learning’. In: EWRL — Eighteenth European Workshop on Reinforcement Learning. Tübingen, Germany, 17th Sept. 2025. URL: <https://hal.science/hal-05116301> (cit. on p. 13).
- [42] A. Kobanda, O.-A. Maillard and R. Portelas. ‘A Continual Offline Reinforcement Learning Benchmark for Navigation Tasks’. In: IEEE Conference on Games. Lisboa, Portugal, 3rd June 2025. URL: <https://hal.science/hal-05126447> (cit. on p. 12).
- [43] A. Kobanda, R. Portelas, O.-A. Maillard and L. Denoyer. ‘Hierarchical Subspaces of Policies for Continual Offline Reinforcement Learning’. In: MDCD Workshop at ICLR 2025. Singapore, Singapore, 11th Apr. 2025. URL: <https://hal.science/hal-05126349> (cit. on p. 12).
- [44] A. Shilova, A. Davey, B. Driss and R. Akrou. ‘StaQ it! Growing neural networks for Policy Mirror Descent’. In: EWRL — Eighteenth European Workshop on Reinforcement Learning. Tübingen, Germany, 2025. doi: [10.48550/arXiv.2506.13862](https://doi.org/10.48550/arXiv.2506.13862). URL: <https://hal.science/hal-05118839> (cit. on p. 12).
- [45] A. Shukla and D. Basu. ‘Unifying (Federated) (Private) High-Dimensional Bandits via ADMM’. In: EWRL – Eighteenth European Workshop on Reinforcement Learning. Tuebingen, Germany, Germany, Sept. 2025. URL: <https://hal.science/hal-05319278> (cit. on p. 11).

Reports & preprints

- [46] D. Basu, S. Chakraborty, D. Chanda, B. D. Das, A. Ghosh and A. Ray. *Dimension Agnostic Testing of Survey Data Credibility through the Lens of Regression*. 10th Oct. 2025. URL: <https://hal.science/hal-05308416> (cit. on p. 16).
- [47] D. Basu, U. Das, B. Driss and U. Mukherjee. *Performative Policy Gradient: Optimality in Performative Reinforcement Learning*. 23rd Dec. 2025. URL: <https://hal.science/hal-05448545>.
- [48] V. Boone and A. Tuynman. *Towards Blackwell Optimality: Bellman Optimality Is All You Can Get*. 15th Oct. 2025. URL: <https://inria.hal.science/hal-05317876> (cit. on p. 9).
- [49] R. Degenne. *Markov kernels in Mathlib’s probability library*. 5th Oct. 2025. URL: <https://inria.hal.science/hal-05302581> (cit. on p. 17).

- [50] R. Degenne, D. Ledvinka, E. Marion and P. Pfaffelhuber. *Formalization of Brownian motion in Lean*. 25th Nov. 2025. URL: <https://hal.science/hal-05383328> (cit. on p. 17).
- [51] F. Fabre Ferber, D. Gay, J.-C. Soulié, J. Diatta and O.-A. Maillard. *Kriging and Gaussian Process Interpolation for Georeferenced Data Augmentation*. 10th Jan. 2025. URL: <https://hal.science/hal-04879232> (cit. on p. 10).
- [52] G. Pourcel, D. Basu, M. Ernoult and A. Gilra. *Lagrangian-based Equilibrium Propagation: generalisation to arbitrary boundary conditions & equivalence with Hamiltonian Echo Learning*. 6th June 2025. URL: <https://hal.science/hal-05111430> (cit. on p. 13).
- [53] W. Radji and O.-A. Maillard. *How Hard is it to Confuse a World Model?* 23rd Oct. 2025. URL: <https://hal.science/hal-05326972> (cit. on p. 12).

11.3 Cited publications

- [54] M. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994 (cit. on p. 6).
- [55] B. Recht. ‘A Tour of Reinforcement Learning: The View from Continuous Control’. arxiv preprint 1806.09460. 2018 (cit. on p. 6).
- [56] R. Sutton and A. Barto. *Reinforcement Learning: an Introduction*. 2nd ed. <http://incompleteideas.net/book/the-book-2nd.html>. MIT Press, 2018 (cit. on p. 6).
- [57] C. Szepesvári and T. Lattimore. *Bandit Algorithms*. Cambridge University press, 2019 (cit. on p. 6).