Direction des relations internationales

Associate Team proposal 2026-2028

Submission form

Title: Foundations of Structured Inference and Control with Non-asymptotic Sequential Learning

Associate Team acronym: FosSIL

Principal investigator (Inria): Odalric-Ambrym Maillard, Inria Scool.

Principal investigator (Main team): Aditya Gopalan, Indian Institute of Science, Bangalore.

Other participants: Shubhada Agrawal, Indian Institute of Science, Bangalore.

Debabrota Basu, Inria Scool.

Key Words: Add key words with regard to: A- Research themes on digital science: A3.1.4. Données incertaines

A3.4. Apprentissage et statistiques

A3.4.3. Apprentissage par renforcement

A3.4.4. Optimisation pour l'apprentissage

A6.2.4 Méthodes statistiques

B- Other research themes and application areas: B9.5.6. Science des données

1 Partnership

1.1 Detailed list of participants

Permanent members on each side:

- Odalric-Ambrym Maillard (Inria, since 2015) is the principal investigator of the French partner, and has been working on advancing bandit and reinforcement learning theory. From 2016 to 2021, he was PI of the ANR JCJC project BADASS (BAnDits Against non-Stationarity and Structure), and from 2021 to 2024 of the French-Japan Associate Team RELIANT (Real-life Bandits), both precursor to this project. Besides these foundation work, he pushes society-oriented applications around health-care and agroecology (see http://odalricambrymmaillard.neowordpress.fr/research-projets). http://odalricambrymmaillard.neowordpress.fr.

- Aditya Gopalan (Associate Professor, Dept. of ECE, IISc, since 2014) is the principal investigator on the Indian side. His research interests revolve around foundational aspects of machine learning, with a focus on online and reinforcement learning, optimization and control, and statistical inference. He has made contributions to the modeling, design and analysis of algorithms for bandits and reinforcement learning such as Thompson sampling, Gaussian process-based black-box optimization, and policy gradient methods. He has advised 4 students towards their Ph.D. dissertations in the broad area of sequential decision making. He has also contributed to public health efforts undertaken during the COVID-19 pandemic in India by way of designing testing and sampling strategies. https://ece.iisc.ac.in/~aditya

- Debabrota Basu (Inria starting faculty, since 2020) is a colleague of Odalric-Ambrym Maillard in the Scool team, and also collaborating on the topics of designing safe and efficient algorithms for bandits and reinforcement learning. Debabrota also carries out active research on the topics of privacy and fairness in machine learning. He obtained ANR Young Researcher (JCJC) for working on the impact of constraints and structures in bandits and reinforcement learning. He has also started the Inria-Indian Statistical Institute Kolkata's associate team (SeRAI) to concurrently study computational and statistical efficiencies of distribution testing. https://debabrota-basu.github.io

- Shubhada Agrawal (Assistant Professor, Dept. of ECE, IISc, since 2025) is a colleague of Aditya Gopalan in the ECE department at IISc, India. Her research lies at the intersection of sequential statistics and machine learning, developing rigorous methods for sequential hypothesis testing and multi-armed bandits, with an emphasis on optimality, robustness, and finite-sample guarantees. Shubhada has been actively collaborating with the researchers at the Scool group, INRIA, Lille, since her visit to the team in 2022 as a PhD student. https://sites.google.com/view/shubhada-agrawal

PhD students on each side (along with the work packages to which they are expected to contribute):

- Soumya Ghatak (PhD student, August 2023 - July 2028, IISc) is advised by Aditya Gopalan and supported by the Ministry of Education (Govt. of India) scholarship. He is currently engaged in developing reinforcement learning algorithms based on Wasserstein gradient flow approaches, and possesses a masters degree in statistics (work packages **WP1A, WP1B**). https://soumyaghatak.github.io

- Aditya Siroutiya (PhD student, August 2025 - July 2030, IISc) is advised by Shubhada Agrawal and is supported by the Ministry of Education (Govt. of India) scholarship. He is passionate about the theoretical foundations of sequential learning with a particular interest in multi-armed bandits and sequential hypothesis testing (work packages **WP2A, WP2C**).

- Alhad Sethi (Master's student, August 2025 - December 2027, IISc) is advised by Shubhada Agrawal and is supported by the Ministry of Education (Govt. of India) scholarship. He is currently engaged in designing efficient and theoretically grounded sequential hypothesis tests in structured environments (work packages **WP1C, WP2B**).
- Adrien Prevost (PhD student, October 2024 - September 2027, PEPR Foundry) is co-advised by T. Mathieu and O-A. Maillard under PEPR Foundry grant. He is working on extending robust bandits to the setup of temporally structured multi-armed bandits (work packages **WP1C,WP2A,WP2C**). https://prevostad.github.io/Site
- Udvas Das (PhD student, October 2024 - September 2027, ANR Republic), is a first year PhD student at Scool in Inria Centre at the University of Lille, and co-advised by D. Basu and O-A. Maillard. His work focuses on impacts of constraints in sequential decision making, specifically bandits and reinforcement learning (work packages **WP1A,WP1B,WP1D**).
- Thomas Michel (PhD student, Octobre 2024- Septembre 2027, PEPR Foundry) is a second year PhD student at Scool in Inria Centre at the University of Lille. He is supervised by D. Basu and E. Kaufmann. His work focuses on impacts of robustness and privacy in bandits and sequential hypothesis testing problems (work packages **WP2A,WP2B**).
- Waris Radji (PhD student, October 2024 - September 2027, PEPR Foundry), is advised by O-A. Maillard under PEPER Foundry grant. He is working on adaptation of bandit techniques to continuous MDPs and deep reinforcement learning (work packages **WP1B,WP1D**). https://warisradji.com

1.2 Nature and history of the collaboration

The Indian and French teams have a rich history of scientific engagement and collaboration dating back to over two decades ago, in the broad areas of multi-armed bandits, Markov decision processes and statistical methods.

A. Gopalan and OA. Maillard were colleagues during their postdoctoral studies at the research group of Shie Mannor, Dept. of Electrical Engineering, Technion-Israel Institute of Technology. After returning to their home countries, they continued to collaborate on problems of mutual interest in sequential decision making, and were the recipients of an Indo-French IFCAM grant for collaborative research during 2016-2017. This was followed by visits of A. Gopalan to the Inria Scool group and OA. Maillard to IISc on other sponsored projects. Their collaboration has resulted in 3 joint articles along with their graduate students – [18], [12] (AISTATS), and more recently [13] (COLT).

S. Agrawal visited the Inria Scool group for an extended period of 3 months in 2022, initiating fruitful research collaborations with OA. Maillard and D. Basu. This has resulted in a joint article [2] (ALT). S. Agrawal also collaborates actively with other members at Inria Scool such as Rèmy Degenne and Timothée Mathieu.

Other joint articles of the team members include [25] and [28]. Both teams are leading experts in RL theory, while using different, complementary approaches. In particular, they have a history on working simultaneously on closely related topics, such as corrupted bandits, boundary crossing probabilities for exponential families and more generally concentration of measure for sequential processes, and structured (e.g linear, kernel) models of Markov decision processes.

One important motivation for creating this associate team is the common identification of similar research challenges arising not only at the theoretical level but also at the computational level, which hinder the practical availability of efficient allocation strategies. We feel that connecting on the mathematical level with our complementary expertise on Sequential Statistics, including multi-armed bandit (MAB), sequential hypothesis testing (SHT) and Markov decision processes (MDP) may significantly advance the design of the next generation of cost-efficient sequential decision making algorithms.

2 Scientific program

2.1 Context

Sequential statistical inference and decision-making encompass a broad family of problems, including sequential hypothesis testing, identification, sample complexity analysis, and regret minimization. Unlike classical (batched) statistical procedures that require the full dataset to be collected prior to decision-making, sequential approaches evaluate data as it arrives, enabling adaptive and more efficient decisions. Over time, a unifying information-theoretic perspective has profoundly advanced our understanding of adaptivity in sequential decision-making. In particular, the development of instance-dependent lower bounds [19], which characterize the fundamental limits of achievable performance, has emerged as a central organizing principle. For each problem formulation, these bounds exhibit an optimization problem that captures the best achievable allocation of information or sampling effort. Solving this problem yields both the asymptotic form of an optimal allocation strategy and its associated performance cost, providing a common foundation that connects adaptive hypothesis testing [3], multi-armed bandits [1, 31, 2, 24, 5, 10], and regret minimization in Markov decision processes (MDPs) [29]. This generic viewpoint has led to tight characterizations of optimal strategies in many structured bandit and reinforcement learning models [24, 29, 8, 7, 32, 10], and recent extensions in continuous settings [30]. However, several fundamental challenges remain.

First, the optimization problems underlying these information-theoretic bounds are often computationally demanding, and sometimes NP-hard. Consider, for instance, the standard (by-now) change-of-measure or data-processing based information-theoretic variational problem underlying the Track-and-Stop exploration algorithm for multi-armed bandits [16]. While it provides principled guidance on how to choose the exploration distribution as a function of the underlying problem instance, the variational problem as such is computationally challenging even for finite action spaces. Solving for optimal exploration distributions is exacerbated further with large or continuous action spaces, and more generally in MDPs [36, 7]. This motivates the search for *computationally tractable surrogates* with controlled sub-optimality possibly considering alternative performance criteria [20, 26] and modern computational methods for optimization over probability spaces such as optimal transport [33]. Further, even when tractable, multiple asymptotically optimal strategies (e.g., KLUCB [9], Thompson Sampling [10]) may exhibit markedly different finite-sample behaviors, such as variability in stopping times or concentration properties. This requires developing the emerging *non-asymptotic* theory of these lower bounds [17]. A related and somewhat overlooked question is how structural properties of environments, such as local similarity, geometrical invariance and compositionality [21, 22, 4], can be exploited to revisit lower bounds and efficient strategies.

Second, real-world applications increasingly require robust and privacy-preserving extensions of classical sequential models [6]. Recent advances in revisiting Wald's classical stopping rules [35, 14, 34] and adaptive experimentation in non-parametric [3], structured [15], and differentially private settings [27] highlight the growing demand for reliable decision-making under uncertainty, privacy constraints, and non-parametric or dependent data. Addressing these challenges calls for a new generation of models and algorithms that combine structured inference, computational tractability, and robust performance beyond the asymptotic regime.

The two objectives are the central focus of the FOSSIL project. We elaborate on them in next section.

2.2 Objectives (for the three years)

Building on the recent advances outlined above, the FOSSIL project seeks to advance the theoretical and algorithmic foundations of sequential decision making by addressing two complementary objectives (O) organized into the following work packages (WP); We expect the two teams to contribute equally.

[Objective O1] Principled, computationally tractable surrogates for large reinforcement learning problems. Real-world reinforcement learning problems exhibit large or infinite state/action spaces, which does not permit theoretically grounded approaches based on tabular or finite MDPs to scale well, in both the computational and statistical sense.

Continuous Bandits [WP1A] & MDPs [WP1B]: To design computationally tractable surrogates that preserve the information-theoretic structure of optimal decision-making, we want to leverage modern tools from the theory of optimal transport and Wasserstein gradient flow over distribution spaces, and their associated mean-field particle discretizations. We will develop this approach for non-contextual and contextual bandits (WP1A) before moving to MDPs (WP1B). These formulations will serve as a foundation for developing particle-based and gradient-flow approximations that translate idealized, continuous dynamics into scalable algorithms. **Non-asymptotic analysis and relaxation [WP1C]:** To bridge the gap between asymptotic theory and realistic, data-limited settings, we aim to study non-asymptotic variants of the underlying optimization problems revisiting the very notion of optimality (following emerging notions of lenient or satisficing criteria) and relaxing strict optimality to avoid NP-hard formulations while retaining theoretical interpretability. **Geometric, compositional structure [WP1D]:** Complementarily, we propose to investigate how structural properties of environments, such as local similarity and compositionality, can be exploited to achieve improved sample efficiency and principled generalization across related tasks.

[Objective O2] Performance and robustness beyond classical models. The second objective focuses on performance guarantees and robustness in sequential inference and learning. Classical formulations (e.g., Wald's sequential tests or asymptotic regret bounds) capture expected performance but overlook variability, fairness, privacy, and distributional robustness, which are critical in modern data-driven applications such as large-scale A/B testing or adaptive experimentation.

Test Selectivity [WP2A]: We plan to study the beyond-expectation behavior of sequential procedures, including the distribution of stopping times, variance, and concentration properties, to complement understanding of their finite-sample performance and offer better selection properties for the practitioner. **Privacy-Efficiency tradeoff [WP2B]:** We propose to design privacy-preserving and robust sequential testing methods that extend recent advances in differentially private hypothesis testing to composite, non-parametric, and dependent-data settings. These investigations will ultimately contribute to a broader theory of robust sequential inference suitable for high-stakes or sensitive real-world environments. **Temporal structure [WP2C]:** Sequential learning is often used under temporal constraints, such as when decisions must come in a batch or group-sequential way, or when handling systems that exhibit autonomous dynamics. We plan to explore how to adapt the methodology for bandits and hypothesis testing to such settings.

2.3 Work-program (for the first year)

In year 1, we mainly focus on advancing WP1A and WP2A, which we detail below. We also expect early progress on WP1C with SA, AS and WP2C with OAM, AP.

Problem WP1A: Computationally Tractable surrogates for Continuous Bandits, more precisely **Exploration via particle Wasserstein gradient flow for contextual bandits.** We aim to tackle the following problem: *Given a bandit problem with a continuous action (or policy) space, design an action-sampling algorithm which, when run with an appropriate stopping rule, reliably identifies an ϵ-optimal arm (or policy). The sampling algorithm must be based on computationally tractable primitives and come with guarantees on its statistical efficiency (sample complexity) with regard to an oracle that employs an information-theoretically optimal sampling strategy (originating from, say, standard change-of-measure considerations [16]).*

- Methodology: We first obtain an appropriate information-theoretic lower bound, for contextual bandits, on the sample complexity of finding a (near) optimal arm (or more generally a policy), typically expressed as an optimization over the space of exploration distributions, see e.g., [16] for the case of non-contextual bandits. The next step is to view this measure optimization as taking place in the Wasserstein space of probability measures, and derive an idealized gradient flow, in continuous time, that converges to the optimal exploration distribution. We then aim to approximate this ideal continuous-time continuous-space evolution using ideas from mean-field approximations to Wasserstein gradient flow [11], to yield computationally tractable schemes that perturb each member of an empirical population with the Wasserstein gradient velocity vector at its location to gradually move the population distribution towards the optimal exploration distribution.
- Participants: Aditya Gopalan (AG), Odalric-Ambrym Maillard (OAM), Debabrota Basu (DB), Shubhada Agrawal (SA), Udvas Das (UD), Soumya Ghatak (SG)
- Planned exchange and visits: This activity is initiated and driven in Year 1. UD & OAM plan to travel from France to India in Year 1.

Problem WP2A: Beyond expected performance: moderate and large deviations for optimal sequential tests. Sequential hypothesis testing (HT) is the most fundamental setting of sequential learning. These questions were addressed in the sequential HT setup for certain parametric problems in the classical works such as [35, 14], and more recently, by [3] for composite and non-parametric problems. To uncover the robustness of optimized sequential tests, we aim to address the question: *Given a collection of statistically optimal sequential tests, which one must ultimately be chosen?*

- Methodology: While many tests may have the same expected sample complexity, their behavior can vary significantly along individual sample paths. We aim to analyze properties of the random stopping times of optimized sequential tests, such as their variance, concentration around the mean, and central limit behavior. Some progress has been made on this front in classical literature on simple and parametric HT framework (cf. [23, 34]). Building upon these as well as recent work [3], we will design sequential tests that exhibit not only good average performance but also desirable path-by-path guarantees, making them more reliable in practice.
- Participants: Shubhada Agrawal (SA), Odalric-Ambrym Maillard (OAM), Aditya Gopalan (AG), Aditya Siroutiya (AdS), Adrien Prevost (AP)
- Planned exchange and visits: This project will be initiated and driven in Year 1. SA will travel from India to France in Year 1. OAM will travel from France to India in Year 1.

3 Data Management Plan

Within the FOSSIL project, we plan to work on designing sound principled methods and collaborate at the fundamental level. The code of the algorithms linked to the published research articles issued from the project will be made publicly available on an open repository (e.g. github), as per the communication standards of the community. No critical data is expected to be considered.

4 Budget

4.1 Budget (for the first year)

Researcher Name	Status	Funding	Mission	Estimated cost	Objective
S. Agrawal	Researcher IISc	CEFIPRA	14 days, Spring 2026 India to France	4,009k	**WP1A,WP2A**
A. Sethi	Ph.D. IISc	CEFIPRA	90 days, May-July France to India	5.667k	**WP2B**
OA Maillard	Researcher Inria	Inria	15 days, Fall 2026 France to India	4.3k	**WP1A,WP2A**
U. Das	Ph.D. Inria	Inria	90days, July-Sept France to India	5.7k	**WP1A**

Total budget request for year 1 (Inria funding): 10k

Total budget request for year 1 (DST-CEFIPRA funding): 9.7k This budget is consistent with the strict guidelines from document https://cefipra.org/Docs/DSTI-GLine-International-Travel.pdf.

Total amount of co-funding (please specified if the co-funding has been secured or just applied to): D. Basu's travel till 2027 will be partially funded by ANR Republic. U. Das's PhD is funded by the same. T. Michel, A.Prevost and W. Radjis' PhDs are funded by PEPR Foundry. PhD is O.-A. Maillard has applied for an ANR and D. Basu is further applying for ERC grants which would fund a Phd student and partially fund the last years' of travel, if successful (see 4.3).

Master's and PhD students from India are supported by a fellowship from the Ministry of Education (Govt. of India). Shubhada Agrawal's research is being supported in part by the start-up (seed funding) grant by IISc, INSPIRE Faculty Fellowship by the Department of Science and Technology (Govt. of India), and the Pratiksha trust young investigator award, IISc.

4.2 Tentative Budget for second and third year

After the first year of the project, we expect to spend more time working in Year 2 on WP1B, WP1C, as well WP2A, and WP2C. In Year 3, we will focus more on WP1D, WP2B and WP2C. We expect to involve new PhD students joining the two teams. We aim to balance visits from France to India and from India to France.

Year 2 estimated budget for Inria: 10k

We expect at least 2 weeks visit from France to India and 3 weeks from India to France. Estimated cost: 5.4k (CEFIPRA) and 4.3k (Inria)

We also expect one longer visit (3 months) by one PhD student from IISc, and one PhD from Inria to visit IISc. Estimated cost: 5.7k (CEFIPRA) and 5.7k (Inria)

Year 3 estimated budget for Inria: 10k

We expect at least 2 weeks visit from France to India and 2 weeks plus 3weeks from India to France. Estimated cost: 9.6k (CEFIPRA) and 4.3k (Inria).

We also expect one longer visit (3 months) by one PhD student from IISc, and one PhD from Inria to visit IISc. Estimated cost: 5.7k (CEFIPRA) and 5.7k (Inria)

4.3 Strategy to get additional funding

The work on instance-dependent lower bounds and robustness will be partly supported by the PEPR Foundry. PEPR Foundry funds two PhD students that participate to the project. On the French side, OA Maillard and colleagues are preparing an ANR grant proposal that should fund a PhD student on a relevant parts of the project. In order to complement the funding of the associate team, we intend to resort to the Raman-Charpak Fellowship of Campus France implemented by the Indo-French Centre for Promotion of Advanced Research. This is open for IISc PhD students to visit French Institutes and hosting the French Masters & PhD Students at IISc, hence especially complementary to the associate team. Last, and crucially, by strengthening our collaborations during Year 1 of the project, we expect to apply for a CEFIPRA grant (submission next August). This will directly enable to hire and co-supervise PhD students and Postdoctoral fellows, hence considerably expand the impact and ambitions of the FOSSIL project.

A. Gopalan and colleagues, on the Indian side, separately plan to seek out funding for supplementing the thrusts in the FOSSIL project via grant calls anticipated from the India AI mission on foundational ML and AI. This can be strengthened also via grant proposals planning to be submitted to the recently initiated Anusandhan National Research Foundation's (ANRF) Advanced Research Grant (ARG) scheme from 2026 onwards.

5 Added value

On both sides, the project will directly benefit the PhD students and early-career researchers involved, by exposing them to complementary research cultures and by fostering collaboration on shared theoretical questions from different methodological angles. The associate team will provide a fertile environment for exchanging ideas, co-supervising young researchers, and consolidating a joint scientific community around the emerging theory of structured and non-asymptotic sequential learning.

Inria – Scool team: The Scool team has long-standing expertise in bandit algorithms, reinforcement learning, and sequential decision-making, with strong contributions to the information-theoretic understanding of exploration–exploitation trade-offs and the design of adaptive strategies. Over the past years, Scool has also increasingly engaged with applications of adaptive experimentation in science and engineering. The collaboration will reinforce these directions by creating a dynamic space for confronting viewpoints, sharing technical advances, and co-developing unified frameworks that bridge statistical inference and control. Beyond the scientific outcomes, the partnership naturally extends long-standing informal interactions between both groups and will accelerate ongoing efforts toward a more principled theory of non-asymptotic and structured learning.

Indian partner team: The Indian group has complementary strengths in theoretical statistics, information-efficient bandit algorithms, and sequential inference under structural or robustness constraints. Their work combines statistical precision with algorithmic insight, often focusing on fine-grained finite-sample analysis and information-theoretic optimality. Collaborating closely with Scool will provide a stimulating environment to confront and refine shared ideas, to explore unified viewpoints on open problems in reinforcement learning and structured bandits, and to build a coherent transnational research axis at the interface of statistics and control.

Overall, the associate team brings together two groups that have been evolving along parallel paths for years. Joining forces will both accelerate theoretical progress and strengthen the visibility of this line of research within the international community.

6 Previous Associate Teams

The French investigator O.-A. Maillard has been PI of the French-Japan Associate Team RELIANT (2021-2024) between Inria Lille and University of Kyoto, Link, co-headed with J. Honda (U. Kyoto).

The French co-investigator D. Basu is PI of the Indo-French Associate Team SeRAI (2025-2027) between Inria Lille and Indian Statistical Institute, Kolkata, Link, co-headed with A. Ghosh and S. Chakraborty (ISI Kolkata)

7 Impact

The proposed associate team FOSSIL addresses fundamental challenges that are also of high practical relevance across multiple application domains where decisions must be made sequentially under uncertainty. The theoretical advances expected from this project, particularly in designing computationally tractable, data-efficient, and privacy-preserving algorithms, have strong potential for transfer to industry and applied research settings:

A first area of impact concerns **adaptive experimentation**, which is central to modern data-driven innovation in the life sciences and medicine. Sequential decision methods can dramatically accelerate clinical trials, drug testing, and agroecological field experiments by adaptively reallocating resources toward promising treatments or conditions while maintaining rigorous statistical guarantees. In agroecology, such approaches could guide the design of adaptive trials for sustainable crop systems or precision-farming interventions, improving both efficiency and reproducibility. In healthcare and biomedical research, robust and privacy-aware sequential testing methods developed under Objective O2 could directly enhance the reliability of personalized medicine protocols or digital health monitoring, where data arrive continuously and confidentiality is essential.

Beyond the life sciences, the project's outcomes are expected to impact **industrial optimization** and **digital experimentation** at large, ranging from A/B testing and online recommendation systems to autonomous systems control and robotics, where exploration–exploitation trade-offs and non-asymptotic robustness are key. The collaboration's emphasis on computational surrogates (Objective O1) and non-asymptotic performance guarantees (Objective O2) naturally aligns with the needs of these data-intensive sectors.

The team will promote the dissemination and reuse of its findings primarily through open and reproducible research practices, including the publication of theoretical results, open-source prototypes, and benchmark datasets. While the project remains fundamentally academic, its methodological advances, especially in adaptive experimental design, non-asymptotic optimization, and privacy-aware decision protocols, bear natural relevance for data-driven applications in agroecology, biomedical research, and public health, with whom each PI has led interdisciplinary connections. These domains increasingly rely on sequential, uncertainty-aware decision-making to optimize experimental resources or interventions. In this context, the project's developments could eventually inform collaborations with public or private actors engaged in adaptive experimentation in life sciences and beyond, without departing from the team's primarily scientific and exploratory mission.

By bridging theory and practice in sequential inference and reinforcement learning, this collaboration is poised to advance both the foundational understanding of adaptive learning and its practical deployment in socially and economically critical domains.

8 Intellectual Property Right Management

8.1 Background

There is no intellectual property but the one represented by the published articles available publicly as per the standards of the community.

8.2 Protective measures

None is required. We expect to publish articles jointly and release them in proceedings of international conferences or journals of the field.

9 Ethical Issues

Working on making bandits and RL strategies more applicable naturally aligns with taking into account ethical issues. Establishing mathematically sound strategies with performance guarantees in realistic scenarios under small-data is indeed a crucial way to tackle the ethical issue of recommending actions or treatments in real-world applications, leading to certified algorithms in the context of Trustable AI. On a different point, fairness is directly considered in one work package. Further, we conduct this research at a fundamental level first, while interacting in our respective lab with expert practitioners in health-care and agriculture to minimize ethical issues raised by directly applying RL in practice. Last, we expect communications with Inria Coerle on specific points.

10 Others

11 References

11.1 Joint publications of the partners

References

[1] Agrawal, S., Mathieu, T., Basu, D. and Maillard, O.A., "CRIMED: Lower and Upper Bounds on Regret for Bandits with Unbounded Stochastic Corruption", in *International Conference on Algorithmic Learning Theory*. PMLR, 2024, March, pp. 74–124.

[2] Chowdhury, S.R., Saux, P., Maillard, O. and Gopalan, A., "Bregman deviations of generic exponential families", in *The Thirty Sixth Annual Conference on Learning Theory*, PMLR, 2023, July, pp. 394–449.

[3] Chowdhury, S.R., Gopalan, A. and Maillard, O.A., "Reinforcement learning in parametric mdps with exponential families," in *International Conference on Artificial Intelligence and Statistics* PMLR, 2021, March, pp. 1855–1863.

[4] Gopalan, A., Maillard, O.A. and Zaki, M. "Low-rank bandits with latent mixtures', arXiv preprint arXiv:1609.01508, 2016.

11.2 Main publications of the participants relevant to the project

References

[1] **Agrawal, S.**, Prashanth, L.A., and Maguluri, S.T.. "Policy evaluation for variance in average reward reinforcement learning", *41st International Conference on Machine Learning (ICML), 2024*.

[2] **Agrawal, S.**, Koolen, W.M., and Juneja, S., "Optimal best-arm identification methods for tail-risk measures", *Advances in neural information processing systems (NeurIPS), 2021*.

[3] Bandyopadhyay, A., Juneja, S., and **Agrawal, S.**, "Optimal top-two method for best arm identification and fluid analysis", *Advances in Neural Information Processing Systems (NeurIPS), 2024*.

[4] Takemori, S., Umeda, Y., and **Gopalan, A.**, "Instance-Optimal Pure Exploration for Linear Bandits on Continuous Arms", *42nd International Conference on Machine Learning (ICML)*, 2025.

[5] Chowdhury, S.R., Saux, P., **Maillard, O.-A.**, and **Gopalan, A.**, "Bregman Deviations of Generic Exponential Families", *36th Annual Conference on Learning Theory (COLT)*, 2023.

[6] Pesquerel, F. and **Maillard, O.A.**. IMED-RL: Regret optimal learning of ergodic Markov decision processes. In advances in Neural Information Processing Systems, 35:26363-26374, 2022

[7] Agrawal, S., Mathieu, T., **Basu, D.** and **Maillard, O.A.**, "CRIMED: Lower and Upper Bounds on Regret for Bandits with Unbounded Stochastic Corruption", in *International Conference on Algorithmic Learning Theory*. PMLR, 2024, March, pp. 74–124.

[8] Azize, A., Jourdan, M., Al Marjani, A. and **Basu, D.**, "On the complexity of differentially private best-arm identification with fixed confidence", in *Advances in Neural Information Processing Systems*, 2023, December, pp.71150-71194.

[9] Ouhamma, R., **Basu, D.**, and **Maillard, O.**, *Bilinear exponential family of MDPs: frequentist regret bound with tractable exploration & planning*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, June, Vol. 37, No. 8, pp. 9336-9344.

[10] Carlsson, E., **Basu, D.**, Johansson, F., and Dubhashi, D., "Pure exploration in bandits with linear constraints", in *International Conference on Artificial Intelligence and Statistics*, 2024, April, pp. 334-342, PMLR.

11.3 Other references

References

[1] Shubhada Agrawal. *Bandits with Heavy Tails: Algorithms Analysis and Optimality*. PhD thesis, Tata Institute of Fundamental Research, 2023.

[2] Shubhada Agrawal, Timothée Mathieu, Debabrota Basu, and Odalric-Ambrym Maillard. Crimed: Lower and upper bounds on regret for bandits with unbounded stochastic corruption. In *International Conference on Algorithmic Learning Theory*, pages 74–124. PMLR, 2024.

[3] Shubhada Agrawal and Aaditya Ramdas. On stopping times of power-one sequential tests: Tight lower and upper bounds. *arXiv preprint arXiv:2504.19952*, 2025.

[4] Mahsa Asadi, Mohammad Sadegh Talebi, Hippolyte Bourel, and Odalric-Ambrym Maillard. Model-based reinforcement learning exploiting state-action equivalence. In *Asian Conference on Machine Learning*, pages 204–219. PMLR, 2019.

[5] Achraf Azize, Marc Jourdan, Aymen Al Marjani, and Debabrota Basu. On the complexity of differentially private best-arm identification with fixed confidence. *Advances in Neural Information Processing Systems*, 36:71150–71194, 2023.

[6] Moitree Basu. *Privacy-preserving AI using declarative constraints*. PhD thesis, Université de Lille, 2024.

[7] Victor Boone. Asymptotically optimal regret in communicating markov decision processes. *arXiv preprint arXiv:2505.18064*, 2025.

[8] Victor Boone and Odalric-Ambrym Maillard. The regret lower bound for communicating markov decision processes. *arXiv preprint arXiv:2501.13013*, 2025.

[9] O. Cappé, A. Garivier, O-A. Maillard, R. Munos, and G. Stoltz. Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Annals of Statistics*, 41(3):1516–1541, 2013.

[10] O. Chapelle and L. Li. An empirical evaluation of Thompson Sampling. In *Advances in Neural Information Processing Systems*, 2011.

[11] Sinho Chewi, Jonathan Niles-Weed, and Philippe Rigollet. Optimal transport. In *Statistical Optimal Transport: École d'Été de Probabilités de Saint-Flour XLIX–2019*, pages 1–36. Springer, 2025.

[12] Sayak Ray Chowdhury, Aditya Gopalan, and Odalric-Ambrym Maillard. Reinforcement learning in parametric mdps with exponential families. In *International Conference on Artificial Intelligence and Statistics*, pages 1855–1863. PMLR, 2021.

[13] Sayak Ray Chowdhury, Patrick Saux, Odalric Maillard, and Aditya Gopalan. Bregman deviations of generic exponential families. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 394–449. PMLR, 2023.

[14] Roger H Farrell. Asymptotic behavior of expected sample size in certain one sided tests. *The Annals of Mathematical Statistics*, pages 36–72, 1964.

[15] Greg Fields, Tara Javidi, and Shubhanshu Shekhar. Sequential one-sided hypothesis testing of markov chains. *arXiv preprint arXiv:2501.13187*, 2025.

[16] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference On Learning Theory*, 2016.

[17] Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. Explore first, exploit next: The true shape of regret in bandit problems. *Mathematics of Operations Research*, 44(2):377–399, 2019.

[18] Aditya Gopalan, Odalric-Ambrym Maillard, and Mohammadi Zaki. Low-rank bandits with latent mixtures. *arXiv preprint arXiv:1609.01508*, 2016.

[19] Todd L Graves and Tze Leung Lai. Asymptotically efficient adaptive choice of control laws incontrolled markov chains. *SIAM journal on control and optimization*, 35(3):715–743, 1997.

[20] Hossein Hajiabolhassan and Ronald Ortner. Online regret bounds for satisficing in markov decision processes. *Mathematics of Operations Research*, 2025.

[21] Pavan Karjol, Rohan Kashyap, Aditya Gopalan, and AP Prathosh. A unified framework for discovering discrete symmetries. In *International Conference on Artificial Intelligence and Statistics*, pages 793–801. PMLR, 2024.

[22] David Kuric, Guillermo Infante, Vicenç Gómez, Anders Jonsson, and Herke van Hoof. Planning with a learned policy basis to optimally solve complex tasks. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 34, pages 333–341, 2024.

[23] Tze L Lai and David Siegmund. A nonlinear renewal theory with applications to sequential analysis ii. *The Annals of Statistics*, pages 60–76, 1979.

[24] Stefan Magureanu. *Efficient Online Learning under Bandit Feedback*. PhD thesis, KTH Royal Institute of Technology, 2018.

[25] Timothée Mathieu, Debabrota Basu, and Odalric-Ambrym Maillard. Bandits with stochastic corruption: Lower bounds on regret and robust optimistic algorithms. *Transactions on Machine Learning Research Journal*, 2024.

[26] Nadav Merlis and Shie Mannor. Lenient regret for multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8950–8957, 2021.

[27] Thomas Michel, Debabrota Basu, and Emilie Kaufmann. DP-SPRT: Differentially private sequential probability ratio tests. *arXiv preprint arXiv:2508.06377*, 2025.

[28] Reda Ouhamma, Debabrota Basu, and Odalric Maillard. Bilinear exponential family of mdps: frequentist regret bound with tractable exploration & planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9336–9344, 2023.

[29] Fabien Pesquerel and Odalric-Ambrym Maillard. IMED-RL: Regret optimal learning of ergodic markov decision processes. *Advances in Neural Information Processing Systems*, 35:26363–26374, 2022.

[30] Waris Radji and Odalric-Ambrym Maillard. The confusing instance principle for online linear quadratic control. In *Reinforcement Learning Conference*.

[31] Hassan Saber. *Structure adaptation in bandit theory*. PhD thesis, Université de Lille, 2022.

[32] Hassan Saber, Fabien Pesquerel, Odalric-Ambrym Maillard, and Mohammad Sadegh Talebi. Logarithmic regret in communicating MDPs: Leveraging known dynamics with bandits. In *Asian Conference on Machine Learning*, pages 1167–1182. PMLR, 2024.

[33] Adil Salim, Anna Korba, and Giulia Luise. The Wasserstein proximal gradient algorithm. *Advances in Neural Information Processing Systems*, 33:12356–12366, 2020.

[34] David Siegmund. On the asymptotic normality of one-sided stopping rules. *The Annals of Mathematical Statistics*, 39(5):1493–1497, 1968.

[35] Abraham Wald and Jacob Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, pages 326–339, 1948.

[36] Raymond Zhang and Richard Combes. Thompson sampling for combinatorial bandits: Polynomial regret and mismatched sampling paradox. *Advances in Neural Information Processing Systems*, 37:89437–89467, 2024.

12 Letter of Intent

See the attached letter.

JOINT PROPOSAL SUBMISSION AGREEMENT

(To be filled, signed and stamped by the Principal Collaborators who submit the Proposal)

We have completed the proposal as per the instructions and the completed proposal is submitted jointly by us for consideration. We also enclose the certificates from the Heads of the respective institutions in the format given. In the event of our proposal being accepted, we agree to abide by the terms and conditions to be stipulated for DST-Inria Targeted Programme.

Signature
Name Aditya Gopalan
(Principal Collaborator from India)

Signature
Name Odalric-Ambrym Maillard
(Principal Collaborator from France)

To:

The Director,
Indo–French Centre for the Promotion of Advanced Research
5B, Ground floor, India Habitat Centre
Lodhi Road, New Delhi–110 003 INDIA

ಭಾರತೀಯ ವಿಜ್ಞಾನ ಸಂಸ್ಥೆ
ಬೆಂಗಳೂರು 560012, (ಭಾರತ)
भारतीय विज्ञान संस्थान
बैंगलोर 560 012, (भारत)
Indian Institute of Science
Bangalore 560 012, (India)
Telefax : 080-23600757, Ph: 22932440/444
Email : registrar@iisc.ac.in

Regr(ECE)/IFCP Proposal/59/2025 10th October 2025

Certificates from the Heads of the Institutions of the Principal Collaborators from India/France

In order to result in effective, long lasting interactions between scientists of the two countries, it is desired that collaborative programmes supported by DST-INRIA should correspond to scientific interests of all collaborating laboratories, and not only to those of isolated investigators from these laboratories. For this reason, heads of collaborating laboratories are requested to commit themselves, by verifying that the proposed programme corresponds effectively to a major interest of the group they are responsible for, and by indicating the amount of resources (in terms of recurring expenses, equipment and manpower) allocated to the project from their own budget.

Project Title:

(1) Certified that this institution agrees to the participation of ADITYA GOPALAN, ASSOCIATE PROFESSOR in this Institution as Principal Collaborator from India for the above project which is being submitted for support to DST-INRIA programme.

(2) Certified that the infrastructural facilities related to the project activity available in this institution including equipment, manpower and other facilities, and all necessary administrative support will be extended for the project.

(3) This institution assumes to undertake the management of financial and other responsibilities of the part of the project work which will be conducted in this country.

Capt Sridhar Warrier (Retd) (Oct 13, 2025 15:19:21 GMT+5.5)

Place: Bengaluru

Name & Signature of the
Head of the Institution

Indian Institute of Science * Bangalore-560 012 *

ಕ್ಯಾಪ್ಟನ್ ಶ್ರೀಧರ್ ವಾರಿಯರ್ (ನಿವೃತ್ತ)/कप्तान श्रीधर वारियर (सेवानिवृत्त)
Capt. Sridhar Warrier (Retd.)
ಕುಲಸಚಿವರು/कुलसचिव/REGISTRAR
ಭಾರತೀಯ ವಿಜ್ಞಾನ ಸಂಸ್ಥೆ/भारतीय विज्ञान संस्थान/INDIAN INSTITUTE OF SCIENCE
ಬೆಂಗಳೂರು/बेंगलूरु/BENGALURU - 560 012

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inria

Certificates from the Heads of the Institutions of the Principal Collaborators from India/France

fn order to result in effective, long lasting interactions between scientists of the two countries, it is desired that collaborative programmes supported by DST-INRIA should correspond to scientific interests of all collaborating laboratories, and not only to those of isolated investigators from these laboratories. For this reason, heads of collaborating laboratories are requested to commit themselves, by verifying that the proposed programme corresponds effectively to a major interest of the group they are responsible for, and by indicating the amount of resources (in terms of recurring expenses, equipment and manpower) allocated to the project from their own budget.

Project Title: **Foundations of Structured Inference and Control with Non-asymptotic Sequential Learning (FOS-SIL)**

(1) Certified that this institution agrees to the participation of Odalric-Ambrym Maillard, Chargé de recherche Inria in this Institution as Principal Collaborator from India for the above project which is being submitted for support to DST-INRIA programme.

(2) Certified that the infrastructural facilities related to the project activity available in this institution including equipment, manpower and other facilities, and all necessary administrative support will be extended for the project.

(3) This institution assumes to undertake the management of financial and other responsibilities of the part of the project work which will be conducted in this country.

Date:10/10/2025

Place :Villeneuve d'Ascq

Stéphane Huot,
Director of Inria Center at the
University of Lille

Signé électroniquement par
Stephane HUOT
Le 09/10/2025 à 19:22

Inria centre at the University of Lille
Parc scientifique de la Haute-Borne
40 avenue Halley - Bât. A - Park Plaza
59650 Villeneuve d'Ascq - France
Phone: +33 (0) 3 59 57 78 00

www.inria.fr/en