Associate Team proposal 2021-2023
Submission form

Title: Real-life bandits

Associate Team acronym: RELIANT

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

Principal investigator (Main team): Junya Honda, Kyoto University.

Other participants: *if the project involves other partners on either side name them here (Inria project-team, university, research center...)*

Kyushu University

New York University

The University of Tokyo

CNRS

INSERM

Key Words: Add key words with regard to: *Please refer to the online Scientific Cartography Portal:* `https://cartographie.inria.fr/SIKeywords/accueilKW`

A- Research themes on digital science: *(at most 5 keywords)*

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: *(at most 5 keywords)*

B2. Santé

B3.5. Agronomie

B9.5.6. Science des données

1 Partnership

1.1 Detailed list of participants

- 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 head of the AND JCJC project BADASS (BAnDits Against non-Stationarity and Structure) that is directly relevant to the project. More recently, he worked on society-oriented applications via several projects, including health-care and agroecology (see `http://odalricambrymmaillard.neowordpress.fr/research-projets`). Webpage: `http://odalricambrymmaillard.neowordpress.fr`.

- Junya Honda (Kyoto University, associate professor, since 2020) is the principal investigator of the partner, and has been working on theories of bandit algorithms and its application to experimental design. `http://stat.sys.i.kyoto-u.ac.jp/honda/index_e.html`

- Kohei Hatano (Kyushu University, associate professor, since 2017) has been worked on algorithms for bandit problems and online learning with their application to communication systems. `https://sites.google.com/view/koheihatano/`

- Junpei Komiyama (NYU, assistant professor, since 2020) has been collaborated with Junya Honda on the topics related to the proposal, especially on bandit problems and fairness in machine learning. `https://sites.google.com/view/junpeikomiyama/`

- Emilie Kaufmann (CNRS junior researcher, since 2015, member of Inria Scool) has collaborated with Odalric-Ambrym Maillard on the topics of the proposal and has recently been interested in applying bandit algorithms for drug development, especially clinical trials [40, 3] `http://chercheurs.lille.inria.fr/ekaufman/`

- 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 does also active research on the topics of privacy and fairness in machine learning.
`https://debabrota-basu.github.io/`

- Yuko Kuroki (The University of Tokyo, research associate, since 2021) had been supervised by Junya Honda, and is working on the computational aspects of bandit algorithms. `https://sites.google.com/g.ecc.u-tokyo.ac.jp/yukokuroki/`

- Charles Riou (The university of Tokyo, PhD student, until September 2023) is supervised by Junya Honda and working on nonparametric algorithms in bandit algorithms and their application to risk-sensitive environments.

- Taira Tsuchiya (Kyoto University, PhD student, until September 2023) is supervised by Junya Honda and working on stochastic bandit algorithms and methods jointly involving bandit algorithms and reinforcement learning. `https://tsuchhiii.github.io/`

- Dorian Baudry (PhD student, until November 2022) is co-advised by O-A. Maillard and E. Kaufmann, and has been working on non-parameteric methods for bandits.

  `https://dbaudry.github.io/`

- Clémence Réda (PhD student, until August 2022), is co-advised by E. Kaufmann and Andrée Delahaye-Duriez (INSERM) and is working on applications of sequential methods to drug repurposing `https://clreda.github.io/`

- Hassan Saber (PhD student, until November 2022), is advised by O-A. Maillard under a grant from Inria, and has been working on optimal generic strategies for structured multi-armed bandits. `https://hassansaber.com`
- Fabien Pesquerel (PhD student, until October 2023), is advised by O-A. Maillard, under a grant from ENS Paris, and is working on efficiently handling structures in reinforcement learning theory. `https://fabienpesquerel.github.io`

1.2 Nature and history of the collaboration

Describe the nature and complementarity of the collaboration and the past/existing activities between the participants. Expected length: half a page.

Although this project is the first formal collaboration between our two teams, they have known each other for some years, especially regarding the work on multi-armed bandit theory. Indeed, we both publish on related topics in similar venues. The idea of this collaboration started after O.A. Maillard visited Japan as an invited speaker by J. Honda at IBIS conference, and visited him while in Tokyo (at that time), and meeting at a few other times. Our two teams are leading expert in multi-armed bandit theory, while using different, complementary approaches. In particular, they have a history on working simultaneously on closely related topics, like recently on non-parametric bandit algorithms [39, 5] or on non stationary bandits [8, 26]. One important motivation for creating this associate team is the identification of numerous challenges, acknowledged by both teams, to make bandit strategies applicable - for real - in applications such as health care, personalized medicine or agroecology. While our two teams are involved in medicine and agriculture-related projects with experts from the corresponding domains, we feel that connecting on the mathematical level with our complementary expertise on multi-armed bandit (MAB), sequential hypothesis testing (SHT) and Markov decision processes (MDP) may significantly advance the design of the next generation of sequential decision making algorithms for real-life applications.

2 Scientific program

2.1 Context

Outline the general scientific context of this collaboration: area, general problems, motivations... It is recommended to document this context by references Expected length: half a page.

The RELIANT project is about studying sequential decision making from a reinforcement learning (RL) and multi-armed bandit (MAB) theory standpoint (see [33] for a recent survey). In short, a bandit algorithm adaptively collects samples from a pool of actions (arms) which yield random rewards. In RL, often modelled by a Markov Decision Process (MDP), there is an underlying state that is also impacted by the chosen actions. Different objective can be considered, often related to maximizing rewards (equivalently, minimizing regret), or learning a good policy. Building on over a decade of leading expertise in advancing the field of MAB and RL theory, our two teams have also developed interactions with practitioners (e.g. in healthcare) in recent projects, in the quest to bring modern bandit theory to societal applications, for real. This quest for real-world reinforcement learning, rather than working in simulated and toyish environments is today's main grand-challenge of the field that hinders applications to the society and industry (see e.g. RWRL workshop https://sites.google.com/view/neurips2020rwrl). MAB are acknowledged to be the most applicable building block of RL, mostly due to several successes for online content optimization [34, 13]. However, as experts interacting with practitioners from different fields we have identify a number of key bottlenecks on which joining our efforts is expected to significantly impact the applicability of MAB to the real-world. Those as related to the typically small samples size that arise in medical applications [3], the complicated type of rewards distributions that arise, e.g., in applications to agriculture [5], the numerous constraints (such as fairness [10]) that should be taken into account to speed up learning, and the possible non-stationary aspects [8]. We elaborate on them in the next section.

2.2 Objectives (for the three years)

State the main scientific goals of this collaboration: research directions, anticipated challenges, intended approaches / methodologies / techniques... Expected length: half a page.

In order to tackle the previous challenges, we organize the research directions into four complementary work-packages. We expect the two teams to contribute to each.

[WP1] Optimality under practical constraints: fairness, ethics, and structures. Structures are omnipresent in real-life data, and so are the constraints on applicable policies. In the presence of structures among the arms (e.g. Lipschitz, linear or unimodal bandits), the optimal policies and also the the achievable *regret lower bounds* are significantly modified. Additionally, the raising concerns over algorithmic bias and fairness measures posit additional constraints on applicable policies. Despite key advances in efficiently exploiting some classical structures in bandits [42, 41, 43, 36, 14, 27, 44, 31], and mitigating specific fairness constraints [22, 35, 10], a unified study of fairness constraints as *structures*, and designing optimal and computationally efficient algorithms for them is still an open question.

[WP2] Dealing with small sample sizes. Pure exploration in bandits allows to formalize many practical problems such as the identification of the best treatment in a phase II clinical trial. The fixed-confidence formulation is the most popular setting in the literature but it provides sample complexity lower bound showing that the number of required samples is very large to guarantee meaningful error probabilities [16]. Thus considering different formulation of the problem becomes necessary to obtain a reasonable guarantee for a small number of samples (e.g., patients in a clinical trial). We plan to contribute to the under-studied *fixed-budget* setting [2] and also to propose different formulations such as the Bayesian pure exploration and identification of not the best but reasonably good arms [24, 23].

[WP3] Relaxing typical assumption on the rewards. Strong optimality properties for bandit algorithms are mostly obtained under restrictive assumptions on the rewards distributions (e.g., Bernoulli, Gaussian with known variance), and optimal algorithms need to know which distributions they are facing [11, 25]. We want to develop universal, practical bandit algorithms with near-optimal performance for a wide range of possible distributions, continuing the promising line of work around sub-sampling and reweighting strategies [4, 12, 6, 32, 39].

[WP4] Facing non-stationary, exploiting recurrence. Practitioners often face similar bandits/MDP instances over time. Whether the change happens at a known time or not (change-point detection [1, 37, 7]), this creates recurring patterns [43]. Exploiting previously solved MDPs/bandits could lead to significant regret efficiency over existing work (e.g. [9]). A natural question we want to solve is: How fast can we identify the MDP on which we are playing? Also, we want to address the *identification*-exploration-exploitation trade-off in an efficient way, revisiting lower bounds and combining RL with change-point and hypothesis-testing theory.

2.3 Work-program (for the first year)

Problem 1 [WP3]: Two powerful approaches for non-parametric settings have been recently studied by both teams, with complementary theoretical guarantees. [39] have proposed the Non-Parametric Thompson Sampling (NPTS) algorithm, that has been proved to be optimal for distributions with known bounded support while [6] have proved the optimality of sub-sampling algorithms for several families of parametric distributions, possibly unbounded. We want to understand how far one can go in designing a single bandit algorithm that can work for (almost) any distribution.

- Methodology: A first step is to understand under which conditions sub-sampling algorithms can have logarithmic regret. Another direction is to see whether the duelling architecture of sub-sampling algorithms can be combined with the random reweighting performed by NPTS.
- Technique: This project heavily rely on the expertise of Dorian Baudry on sub-sampling methods and on Non-Parametric Thompson Sampling [5], and that of Charles Riou and Junya Honda who proposed the NPTS algorithm.
- Participants: Dorian Baudry, Charles Riou, O-A. Maillard, E. Kaufmann, J. Honda.
- Planned exchange: Dorian Baudry (PhD student in Scool) will spend four months in Kyoto (March-June 2022) to work in collaboration with Charles Riou and Junya Honda.

Problem 2: [WP1] Mitigating algorithmic bias and ensuring fairness have emerged as new concerns as bandits are applied to real-life applications and decision making. Though there are multitudes of fairness definitions proposed in literature (e.g. fairness of exposure [45], smooth fairness [35], demographic parity [22] etc.), most of them can be formulated as constraints over the computed policy. How these constrains modify the best instance-dependent achievable performance and suggest modifications of state-of-the-art bandit strategies is an open challenge.

Recently, Junya Honda's team proposed an computationally efficient algorithm for non-convex optimization under fairness constraints [28]. Scool members also have developed the state-of-the-art fairness verifiers to measure the violation of fairness constraints by an ML algorithm [18, 17], and investigated consequences of using policy optimization algorithms for fair decision making [10]. We aim to merge these research directions to develop optimal bandit algorithms that satisfy fairness constraints with high probability. In addition, the modification of algorithms considering fairness constraints often involves computationally expensive procedures [29]. Therefore the computational aspects of algorithms must also be tackled.

- Methodology: The first step will be to compute the lower bounds of regret under fairness constraints. Then, we aim to leverage the lower bounds to design efficient and fair bandit algorithms. Extending this theory to applications will require ensuring computational efficiency, mitigating violation of fairness constraints, and incorporating structures.
- Technique: The project relies on expertise in structured bandits, optimization under constraints, and change-of-measure to establish lower-bounds and design efficient strategies.
- Participants: D. Basu, J. Komiyama, J. Honda, Hassan Saber, Yuko Kuroki
- Planned exchange: Yuko Kuroki (research associate in The University of Tokyo) will spend two weeks (September 2022) in Scool.

We also plan to have two visits in fall 2022, in order to start collaboration on WP2 and WP4. The precise research agenda will be established during next year.

3 Data Management Plan

Within the RELIANT project, we expect that sharing (medical) data will not be possible due to regulations. Hence, we plan to work on designing sound principled methods and collaborate at the fundamental level. Each team may apply the methods to their own data. 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.

4 Budget

4.1 Budget (for the first year)

Researcher Name	Status	Institution	Mission	Estimated cost	Objective
Dorian Baudry	Ph.D.	Inria	01/03/22-01/07/22 France to Japan	6k (2k secured)	WP3, Pb1
Yuko Kuroki	Research Associate	Tokyo Univ.	01/09/22-14/09/22 Japan to France	3k (2k secured)	WP1, Pb2
Taira Tsuchiya	Ph.D.	Kyoto Univ.	01/09/22-30/09/22 Japan to France	4k (3k secured)	WP2
Fabien Pesquerel	Ph.D.	Inria	2 weeks, fall 2022 France to Japan	3k	WP4

Total budget request for year 1 (Inria funding): 9k

Total amount of co-funding (please specified if the co-funding has been secured or just applied to): 2k from CNRS (travel grant associated with the PhD thesis of Dorian Baudry, will cover at least plane tickets), 2k from University of Tokyo and 3k from Kyoto University (see 4.3).

4.2 Tentative Budget for second and third year

After the first year of the project, we expect to spend more time working on WP2 and WP4 (on top of WP1 and WP3), possibly involving new PhD students joining the two teams. We will balance visits from France to Japan and from Japan to France.

Year 2 estimated budget for Inria: 8k

We expect at least 2 weeks visit from France to Japan and 2 weeks from Japan to France. Estimated cost: 6k (3k for Inria).

We also expect one longer visit (4 months) by one PhD student. Estimated cost: 6k (5k for Inria).

Year 3 estimated budget for Inria: 8k

We expect at least 2 weeks visit from France to Japan and 2 weeks from Japan to France. Estimated cost: 6k (3k for Inria) .

We also expect one longer visit (2-4 months) by one or two PhD students. Estimated cost: 6k (5k for Inria).

4.3 Strategy to get additional funding

The work on fairness in online learning and bandits will be partly supported by Regalia, a pilot project of Inria. Regalia will fund a PhD student in this topic whom Debabrota Basu is going to co-supervise. Debabrota Basu is also applying for an ANR-JCJC grant, which is going to support the research works in bandits and reinforcement learning under constraints and structures. In Japan side, Junya Honda is advising each PhD student to apply for ACT-X grant that supports young researchers, which Taira Tsuchiya and Yuko Kuroki have already got. The associate team is also planning to apply for Sakura program, which promotes bilateral joint research supported by each country.

5 Added value

On both side, the RELIANT project will directly benefit students involved in the project, by providing them with exposure to other research environments and an international experience.

Inria Scool team: The RELIANT associate team will benefit the research of Inria team Scool in the following way. The SequeL team (ancestor of Scool) has been advancing theory of multi-armed bandit and reinforcement learning over the past +10 years, contributing to many state-of-the-art strategies. The Scool project has a more applied flavor, collaborating on society-oriented applications, which creates a number of unprecedented challenges not typically addressed in the MAB/MDP community. The Japan team has faced similar questions when interacting with practitioners, and brings a complementary standpoint on such questions. They also have a leading expertise in statistics and hypothesis testing that complements that of Scool. The associate team will further bring visibility to such challenges at an international level, paving a different avenue of research in RL than considered by the mainstream industry.

J. Honda team: The Japan team benefits from the collaboration in this associate team from the following way. To widen the applicable fields of the decision-making algorithms, it is vitally important to consider non-i.i.d. settings that are typically represented by MDPs. On the other hand, the researchers in Japan team are not strong at this region and have been mostly considering the problems of classic bandit problems and around there. Therefore the research scope would be drastically enhanced by collaboration with Inria where many works on MDPs and reinforcement learning have been done.

6 Previous Associate Teams

None.

7 Impact

The traditional bandit algorithms try to improve the accuracy in terms of simple regret or regret while aiming to detect the best arm or to mitigate the total loss in this process [33]. This often lead to recommender and web advertisement systems, and clinical trial policies that focus on empowering only the stronger arm from the pool of arms, i.e. songs/movies, products, medicines etc. [30, 38]. But this often narrow down the diversity among arms [10], and also the exposure of the arms to the audience (imagine a recommender or advertisement system) [45, 20]. This invokes the fairness concerns and our work in this project will enable us to mitigate that.

Additionally, in clinical trials and other applications where getting data is cost-intensive, the good guarantees obtainable for traditional algorithms are not realisable [16]. The works on small sample analysis of bandits will impact these application fields. Specially, we aim to impact the clinical trials where the bandit algorithms are still little applied in practice due to their limited realisability in small sample domains [3].

In most applications scenarios, the typical reward does not have a natural parametric form (e.g. yield in agriculture, health score of a patient). Building decision making strategies adaptive to non-parametric reward distributions is expected to directly impact the applicability of RL to realistic application scenarios. Here, we target an agronomy application with CIRAD via the DSSAT crop model [21, 19].

Properly handling recurring patterns in sequential decision making can have a direct impact in the industry, for instance in daily calibration tasks that are time consuming and often involve a small number of recurring problems, e.g. [15]. Transfer learning of RL solutions from tasks to tasks will reduce the sampling requirement.

8 Intellectual Property Right Management

This section can be filled out with the help of a CPPI from your research centre. The goal of this section is to describe the measures to protect the background knowledge and joint results obtained in the framework of the collaboration.

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 involves taking into account ethical constraints. Fairness is directly considered as one work package. Further, establishing mathematically sound strategies with performance guarantees in realistic scenarios under small-data is another way to tackle the ethical issue of recommending actions or treatments in real-world applications. Further, we conduct this research at a fundamental level first, and interact with expert practitioners in health-care and agriculture to minimize ethical issues raised by directly applying RL in practice. Last, we expect communications with Coerle on specific points.

10 Others

Any other element you would like to add. Expected length: half a page.

11 References

11.1 Joint publications of the partners

None.

11.2 Main publications of the participants relevant to the project

List the main publications of the participants that are relevant for the project. List at most 5 publications for each partner.

References

[1] C. Riou and J. Honda, "Bandit Algorithms Based on Thompson Sampling for Bounded Reward Distributions," in *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, vol. 117, 2020, pp. 777–826.

[2] Y. Kuroki, L. Xu, A. Miyauchi, J. Honda, and M. Sugiyama, "Polynomial-Time Algorithms for Multiple-Arm Identification with Full-Bandit Feedback," *Neural Computation*, vol. 32, no. 9, pp. 1733–1773, 2020.

[3] T. Tsuchiya, J. Honda, and M. Sugiyama, "Analysis and Design of Thompson Sampling for Stochastic Partial Monitoring," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 8861–8871.

[4] J. Komiyama, A. Takeda, J. Honda, and H. Shimao, "Nonconvex optimization for regression with fairness constraints," in *Proceedings of the 35th International Conference on Machine Learning*, 2018, pp. 2737–2746.

[5] J. Honda and A. Nakamura, *Theory and Algorithms for Bandit Problems (in Japanese)*. Machine Learning Professional Series, Kodansha Scientific, 2016.

[6] Dorian Baudry, Romain Gautron, Emilie Kaufmann, and Odalric Maillard. Optimal Thompson Sampling strategies for support-aware cvar bandits. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

[7] Hassan Saber, Léo Saci, Odalric-Ambrym Maillard, and Audrey Durand. Routine bandits: Minimizing regret on recurring problems. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 3–18. Springer, 2021.

[8] Hippolyte Bourel, Odalric-Ambrym Maillard, and Mohammad Sadegh Talebi. Tightening exploration in upper confidence reinforcement learning. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1056–1066. PMLR, 13–18 Jul 2020.

[9] Maryam Aziz, Emilie Kaufmann and Marie-Karelle Riviere On Multi-Armed Bandit Designs for Dose-Finding Clinical Trials. *Journal of Machine Learning Research*, Volume 22, pages 1-38, 2021.

[10] Bishwamittra Ghosh, Debabrota Basu. and Kuldeep S. Meel Justicia: A Stochastic SAT Approach to Formally Verify Fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35-9, pp. 7554-7563, 2021.

11.3 Other references

References

[1] Reda Alami, Odalric Maillard, and Raphael Feraud. Restarted Bayesian online change-point detector achieves optimal detection delay. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 211–221. PMLR, 13–18 Jul 2020.

[2] J-Y. Audibert, S. Bubeck, and R. Munos. Best Arm Identification in Multi-armed Bandits. In *Proceedings of the 23rd Conference on Learning Theory*, 2010.

[3] Maryam Aziz, Emilie Kaufmann, and Marie-Karelle Riviere. On multi-armed bandit designs for dose-finding clinical trials. *Journal of Machine Learning Research*, 22:14:1–14:38, 2021.

[4] Akram Baransi, Odalric-Ambrym Maillard, and Shie Mannor. Sub-sampling for multi-armed bandits. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 115–131. Springer, 2014.

[5] Dorian Baudry, Romain Gautron, Emilie Kaufmann, and Odalric Maillard. Optimal thompson sampling strategies for support-aware cvar bandits. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

[6] Dorian Baudry, Emilie Kaufmann, and Odalric-Ambrym Maillard. Sub-sampling for efficient non-parametric bandit exploration. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[7] Lilian Besson, Emilie Kaufmann, Odalric-Ambrym Maillard, and Julien Seznec. Efficient change-point detection for tackling piecewise-stationary bandits, 2020.

[8] Lilian Besson, Emilie Kaufmann, Odalric-Ambrym Maillard, and Julien Seznec. Efficient change-point detection for tackling piecewise-stationary bandits. *arXiv:1902.01575*, 2021.

[9] Hippolyte Bourel, Odalric-Ambrym Maillard, and Mohammad Sadegh Talebi. Tightening exploration in upper confidence reinforcement learning. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1056–1066. PMLR, 13–18 Jul 2020.

[10] Thomas Kleine Buening, Meirav Segal, Debabrota Basu, and Christos Dimitrakakis. Fair set selection: Meritocracy and social welfare. *arXiv preprint arXiv:2102.11932*, 2021.

[11] 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.

[12] Hock Peng Chan. The multi-armed bandit problem: An efficient nonparametric solution. *The Annals of Statistics*, 48(1):346–373, 2020.

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

[14] Thibaut Cuvelier, Richard Combes, and Eric Gourdin. Statistically efficient, polynomial-time algorithms for combinatorial semi-bandits. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(1):1–31, 2021.

[15] Audrey Durand, Theresa Wiesner, Marc-André Gardner, Louis-Émile Robitaille, Anthony Bilodeau, Christian Gagné, Paul De Koninck, and Flavie Lavoie-Cardinal. A machine learning approach for online automated optimization of super-resolution optical microscopy. *Nature communications*, 9(1):1–16, 2018.

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

[17] Bishwamittra Ghosh, Debabrota Basu, and Kuldeep S Meel. Algorithmic fairness verification with graphical models. *arXiv preprint arXiv:2109.09447*, 2021.

[18] Bishwamittra Ghosh, Debabrota Basu, and Kuldeep S Meel. Justicia: A stochastic sat approach to formally verify fairness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7554–7563, 2021.

[19] Gerrit Hoogenboom, CH Porter, KJ Boote, V Shelia, PW Wilkens, U Singh, JW White, S Asseng, JI Lizaso, LP Moreno, et al. The dssat crop modeling ecosystem. *Advances in crop modeling for a sustainable agriculture*, pages 173–216, 2019.

[20] Olivier Jeunen and Bart Goethals. Top-k contextual bandits with equity of exposure. In *Fifteenth ACM Conference on Recommender Systems*, pages 310–320, 2021.

[21] James W Jones, Gerrit Hoogenboom, Cheryl H Porter, Ken J Boote, William D Batchelor, LA Hunt, Paul W Wilkens, Upendra Singh, Arjan J Gijsman, and Joe T Ritchie. The dssat cropping system model. *European journal of agronomy*, 18(3-4):235–265, 2003.

[22] Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in learning: classic and contextual bandits. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 325–333, 2016.

[23] Hideaki Kano, Junya Honda, Kentaro Sakamaki, Kentaro Matsuura, Atsuyoshi Nakamura, and Masashi Sugiyama. Good arm identification via bandit feedback. *Machine Learning*, 108(5):721–745, 2019.

[24] E. Kaufmann, W.M. Koolen, and A. Garivier. Sequential test for the lowest mean: From Thompson to Murphy Sampling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[25] E. Kaufmann, N. Korda, and R. Munos. Thompson Sampling : an Asymptotically Optimal Finite-Time Analysis. In *Proceedings of the 23rd conference on Algorithmic Learning Theory (ALT)*, 2012.

[26] Junpei Komiyama, Edouard Fouché, and Junya Honda. Finite-time analysis of globally nonstationary multi-armed bandits. *arXiv:2107.11419*, 2021.

[27] Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Regret lower bound and optimal algorithm in finite stochastic partial monitoring. *Advances in Neural Information Processing Systems*, 28:1792–1800, 2015.

[28] Junpei Komiyama, Akiko Takeda, Junya Honda, and Hajime Shimao. Nonconvex optimization for regression with fairness constraints. In *International conference on machine learning*, pages 2737–2746. PMLR, 2018.

[29] Junpei Komiyama, Akiko Takeda, Junya Honda, and Hajime Shimao. Nonconvex optimization for regression with fairness constraints. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2737–2746, 2018.

[30] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems–a survey. *Knowledge-based systems*, 123:154–162, 2017.

[31] Yuko Kuroki, Atsushi Miyauchi, Junya Honda, and Masashi Sugiyama. Online dense subgraph discovery via blurred-graph feedback. In *International Conference on Machine Learning*, pages 5522–5532. PMLR, 2020.

[32] Branislav Kveton, Csaba Szepesvari, Sharan Vaswani, Zheng Wen, Tor Lattimore, and Mohammad Ghavamzadeh. Garbage in, reward out: Bootstrapping exploration in multi-armed bandits. In *International Conference on Machine Learning*, pages 3601–3610. PMLR, 2019.

[33] Tor Lattimore and Csaba Szepesvari. *Bandit Algorithms*. Cambridge University Press, 2019.

[34] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *WWW*, 2010.

[35] Yang Liu, Goran Radanovic, Christos Dimitrakakis, Debmalya Mandal, and David C Parkes. Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*, 2017.

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

[37] Odalric-Ambrym Maillard. Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds. In Aurélien Garivier and Satyen Kale, editors, *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, volume 98 of *Proceedings of Machine Learning Research*, pages 610–632. PMLR, 22–24 Mar 2019.

[38] Rishabh Mehrotra, Niannan Xue, and Mounia Lalmas. Bandit based optimization of multiple objectives on a music streaming platform. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3224–3233, 2020.

[39] Charles Riou and Junya Honda. Bandit algorithms based on thompson sampling for bounded reward distributions. In *Algorithmic Learning Theory (ALT)*, pages 777–826, 2020.

[40] Clémence Réda, Emilie Kaufmann, and Andrée Delahaye-Duriez. Machine learning applications in drug development. *Computational and Structural Biotechnology Journal*, 18:241–252, 2020.

[41] Hassan Saber, Pierre Ménard, and Odalric-Ambrym Maillard. Forced-exploration free strategies for unimodal bandits, 2020.

[42] Hassan Saber, Pierre Ménard, and Odalric-Ambrym Maillard. Optimal strategies for graph-structured bandits, 2020.

[43] Hassan Saber, Léo Saci, Odalric-Ambrym Maillard, and Audrey Durand. Routine bandits: Minimizing regret on recurring problems. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 3–18. Springer, 2021.

[44] Taira Tsuchiya, Junya Honda, and Masashi Sugiyama. Analysis and Design of Thompson Sampling for Stochastic Partial Monitoring. In *Advances in Neural Information Processing Systems*, volume 33, pages 8861–8871, 2020.

[45] Lequn Wang, Yiwei Bai, Wen Sun, and Thorsten Joachims. Fairness of exposure in stochastic bandits. *arXiv preprint arXiv:2103.02735*, 2021.

12 Letter of Intent

Please attach to this submission form the signed letter of intent by the partner institution. Template available here: https://partage.inria.fr/share/page/site/appel-2022-equipes-associes/document-details?nodeRef=workspace://SpacesStore/842d22df-5e9f-4f40-b25b-13899aa1a1d9

See the attached letter.