| AAPG2026 | APosTeriori | | PRC |
|---|---|---|---|
| Coordinated by | Emilie Morvant | 48 months | 536 711€ |
| CES 23 - Axe E.02 : Intelligence artificielle et science des données | | | |

# APosTeriori: Actionable PAC-Bayesian Theoretical Certifications for Multi-Armed Bandits

*Certifications théoriques PAC-Bayesiennes actionnables pour les bandits manchots*

## I Context, positioning and objectives of the project

This project is a **fundamental research project** that is a continuation of the ANR project APRIORI (ANR-18-CE23-0015). Concretely, APRIORI laid the foundations for developing machine learning algorithms that directly optimize a theoretical certification result, allowing the construction of prediction models that come with built-in theoretical guarantees. In the **APosTeriori** project, we aim to reach two new objectives to develop not only practically usable theoretical results for a framework where this kind of guarantees are lacking but also to improve model quality and trustworthiness in line with major current challenges and concerns in artificial intelligence.

### a Project objectives and state of the art

**General context.** With the development of **Machine Learning (ML)** methods and their widespread use, **the theoretical analysis of their quality and confidence** when used in production has become **crucial** to ensuring robust, reliable, and trustworthy applications. It is thus essential not only to measure how well models perform on existing data but also to ensure they can generalize to new situations and data (that we can encounter while the model is in production). This is where theoretical guarantees come into play, helping us understand and trust our models when they encounter data they have not seen before. Classical results from **statistical learning theory**, like PAC generalization bounds [29], estimate the so-called generalization gap, which measures the deviation between the performance of the model on observed data and its performance on new unseen data. Not only do these results **justify that a model will be effective**, but they also provide **a better understanding of its behavior**. However, classic approaches typically prove guarantees after the algorithm is designed, without integrating certification into the learning process. During the previous APRIORI project, we derived novel theoretical bounds on the generalization gap for classification tasks (*e.g.*, for neural networks [25, 26] or majority votes [23, 31]) that are directly optimizable. The resulting algorithms are known as **self-bounding algorithms** [7], meaning that a model is learned by optimizing a generalization bound to get a model that comes directly with guarantees. This approach makes the bounds "**usable in practice**" and enhances the trustworthiness of the model. The key results of APRIORI are based on the **PAC-Bayesian theory** [14, 21]. This theory, which has seen renewed interest in recent years, has provided a new path to proving surprisingly rich theoretical results in various settings (*e.g.*, supervised learning [4], domain adaptation [22]). To the best of our knowledge, the PAC-Bayesian theory offers a unique feature among learning theory tools: It has the ability to provide simultaneously generalization guarantees and learning algorithms.

**APosTeriori's objectives.** We aim to tackle the main issue of **designing self-bounding algorithms for another setting** related to sequential learning in ML: the **multi-armed bandit setting**. Instead of optimizing the generalization gap, the objective is to focus on the **notion of regret**, which raises different challenges. Additionally, motivated by making our expected fundamental results "usable in practice", we will pay great attention to making our theoretical outcomes **applicable to major challenges in machine learning**: fairness in decision [1, 10], robustness against malicious attacks [3, 13, 16, 24] and transfer learning to new situations [11, 15, 17, 22].

**Position of APosTeriori in relation to the state of the art.** In the **multi-armed bandit setting** [18, 28], which has numerous applications such as recommender systems, feature selection, resource allocation, clinical trials [see, *e.g.*, 8, 9, 12, 19, 27, 30], the model, akin to a gambler following a policy (*i.e.*, a strategy), sequentially interacts with an environment consisting of multiple slot machines (or arms), each associated with an unknown reward distribution. At each time step, the model must select one arm, with the goal to maximize the cumulative reward over time. As mentioned above, a key performance measure is the **regret**, which quantifies the difference between the cumulative rewards obtained by the optimal policy and those achieved by the gambler's chosen policy. Classical *optimistic* strategies learn a policy (minimizing the regret) by constructing and refining confidence intervals around the rewards of each arm [*e.g.*, 2]. Since the true regret is unknown, most studies focus on establishing bounds on the *expected* regret, which serve as theoretical guarantees. To the best of our knowledge, **no PAC-Bayesian self-bounding algorithm exists**

**that minimizes the regret**, meaning no algorithm directly learns a policy by optimizing a regret upper bound, whose guarantee is obtained by design, rather than through a separate post-hoc analysis.

As mentioned previously, APosTeriori seeks to develop guarantees and PAC-Bayesian self-bounding algorithms "usable in practice" that offer **two key advantages** over traditional guarantees. First, by deriving **computable upper bounds on the regret**, we will be able to **design algorithms directly from these bounds**. This is in contrast with traditional methods, where algorithms are usually developed first, followed by tailored and static theoretical guarantees to analyze the regret after learning[1]. This approach limits the direct application of theoretical results to algorithm design, implying a significant gap between theory and practice. Second, the "usable in practice" nature of our bounds enables **real-time performance monitoring during learning**. Since the bounds are computable and interpretable at runtime, they can be updated and visualized dynamically as the algorithm runs, something that is not possible with expectation-based guarantees, which are static and algorithm-specific.

The partners of APosTeriori agree on the fact that our novel approach we propose will not only deepen the understanding of multi-armed bandits, but will also enable the visualization of theoretical guarantees for existing algorithms, including those originally paired with fixed, non-visualizable guarantees.

## b Methodology

We will address the **development of self-bounding algorithms for multi-armed bandits** through 2 fundamental Work Packages intended to cover the study and formalization of the relationships between PAC-Bayes and multi-armed bandits, along with practical algorithms: (**WP1**) for classical multi-armed bandits, and (**WP2**) for contextual/structured bandits. To maximize the practical impact of our work, a third work package (**WP3**) will focus on applying **WP1-2**'s results on major issues in ML, at the heart of current concerns, and for which the consortium has expertise. A last work package, (**WP4**) will have the objective to provide a PAC-Bayesian toolbox for ML practitioners.

**WP1: Self-bounding algorithms for classical multi-armed bandit.** This WP represents the theoretical foundations of APosTeriori. The goal is to establish a rigorous PAC-Bayesian framework for multi-armed bandits and to design the first self-bounding algorithms. The recent survey by Flynn et al. [6] on PAC-Bayes and multi-armed bandits highlights that there are several unresolved challenges that could exploit the power of PAC-Bayesian certification. Notably, there is no self-bounding algorithm for multi-armed bandits, *i.e.*, that the regret bound is not simultaneously certified and optimized by the algorithm itself. We will revisit standard PAC-Bayesian generalization bounds through the lens of regret minimization, with the objective to derive PAC-Bayesian regret bounds that are both tight and computable (online). In other words, we want to explore how PAC-Bayes can be extended for sequential decision-making under uncertainty, and how our new approach will compare to existing confidence-based of Bayesian methods such as UCB, Thompson sampling, or KL-UCB. The results of **WP1** will directly feed into **WP2** for contextual/structured extensions.

**WP2: Self-bounding algorithms for contextual multi-armed bandit.** Building upon **WP1**, this **WP** will focus on deriving self-bounding algorithms for more sophisticated and realistic learning settings. In particular, we will study contextual multi-armed bandits, where decisions (policies) depend on context vectors, enabling the algorithm to adapt its choices based on additional information for more flexible and adaptive decisions. A first objective will draw inspiration from the offline contextual bandit setting studied by Sakhi et al. [20] (where past interactions are logged from a fixed policy) in order to extend PAC-Bayesian principles to the online contextual bandit framework. Specifically, we expect our PAC-Bayesian bounds to naturally lead to self-bounding algorithms for online contextual bandits that are both computationally efficient and effective in regret minimization, while maintaining tight theoretical guarantees. Beyond this, we will explore multi-armed bandit models with specific structures, such as (generalized) linear bandits, neural bandits, or kernelized bandits.

**WP3: On major ML issues.** Drawing on the algorithms and certifications obtained in the early stages of the project, as well as the expertise of the consortium, we will specifically address the following issues: fairness, robustness, and transfer learning. Indeed, while **WP1-2** are mainly motivated by the theoretical aspects in quite general settings, we are convinced that self-bounding algorithms can be used to tackle these important challenges in ML, while maintaining rigorous theoretical guarantees. For instance, *(transfer*

[1] Note that some modern bandit methods are inspired by regret analysis, but require separate and static post-hoc guarantees.

*learning)* the intrinsic nature of contextual bandits allows adapting a model to new situations, *(fairness)*, we can handle subgroup fairness, where each subgroup's regret could be explicitly controlled and certified with a self-bounding algorithm, or *(robustness)* since multi-armed bandits focus on intervals rather than exact predictions, models can become more resilient to adversarial perturbations, and attackers may struggle to influence the entire interval, making it harder for them to deceive the model.

**WP4: Development of a PAC-Bayes toolbox.** The goal of this **WP** is to develop a general library to learn models or policies through PAC-Bayesian methods. Indeed, there exists no general library dedicated to PAC-Bayes methods. We believe that the current state of research makes this the right moment to initiate such a development. This toolbox will capitalize on recent advances in PAC-Bayes, with the ambition of providing the community with practical tools for designing models with certifications that are both visualizable and, potentially, interpretable. More precisely, our goal is to develop Python code, compatible with PyTorch and scikit-learn, to implement both existing and newly proposed self-bounding algorithms and their corresponding PAC-Bayesian generalization bounds.

**The APosTeriori's organization is summarized in Figure 1.** Additionaly, to tackle each aspect of the project, we would like to recruit three Ph.D. students (**PhD1**, **PhD2**, **PhD3**) who will be co-supervised by two partners, depending on their areas of expertise.

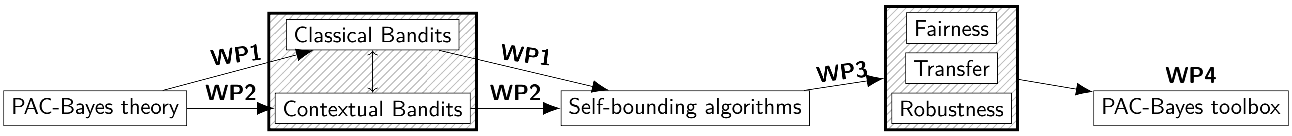


**Figure 1.** Summary of the organization of the work packages of APosTeriori.

## c Expected results and impacts

**The expected scientific results** cover new theoretical results for multi-armed bandits, along with self-bounding algorithms supported by empirical studies on datasets; applications in fairness, robustness, and transfer learning are expected. The consortium will publish its results in major ML venues (*e.g.*, NeurIPS, ICML, ICLR) and journals (*e.g.*, JMLR, MLJ), and we will particularly take care to ease the **diffusion and the transfer of our results**: The source code of our algorithms will be **open-source and publicly available** on repositories (*e.g.*, GitHub). Especially, another important outcome will be a PAC-Bayesian toolbox compatible with ML libraries (*e.g.*, scikit-learn or PyTorch), following the model of open projects such as *POT: Python Optimal Transport* [5]).

**The project is expected to have a positive socio-economical impact** since one goal of the project is to develop learning algorithms, making ML models safer and trustworthy, capable of reasoning, with certification, about uncertainty in their predictions through self-bounding algorithms while using theoretical guarantees usable in practice. This socio-economic impact may directly impact some specific ML applications since we aim to apply our fundamental and theoretical results to fairness, robustness, and transfer learning.

## d Ability of the project to address the research issues of "Axe E.02" of CES 23

APosTeriori fits naturally within the theoretical **machine learning** theme, particularly addressing issues related to **(statistical) learning theory**. Indeed, we are motivated by getting novel theoretical results on method certifications, from which we will derive algorithms by optimizing these certifications. This will lead to a better understanding of the learning process, while improving model **trust** and **efficiency**. In parallel, we will extend these results to tackle ML challenges, particularly those related to **fairness**, **safety** (with robustness), and **transfer learning**.

## e Scientific obstacles / ambitious nature of the project

To the best of our knowledge, the scientific scope of this project, that is designing "usable bounds in practice" for the multi-armed bandit problem is new. The main scientific obstacles are due to these exploratory aspects, which are hard to predict, this is a reason why we chose to focus on different ML issues. We are confident that the way we organize the project ensures a structured approach, reducing the risks of no-result, since it follows a methodological protocol from theory to algorithms, and to applications to key issues in ML in which the consortium has a strong expertise.

# II Partnership (consortium)

**The scientific coordinator** of the project is Dr. Emilie Morvant (implication of 25pm). She is an Assistant Professor (MCF-HDR) in the ML group at Laboratoire Hubert Curien (Data Intelligence, LabHC, UMR CNRS 5516), where she has been since 2014. She is an active researcher in statistical machine learning with contributions in major ML venues, especially on PAC-Bayes theory driven by the development of algorithms[2]. She was the scientific coordinator of **APRIORI** project, which laid the foundations for **APosTeriori**.

**The consortium** is composed of 3 partners: **Laboratoire Hubert Curien** (**LabHC**), **Inria Rennes**, and **Inria Lille**. Intellectual properties considerations were discussed: open science and software packages are promoted. LabHC brings expertise in statistical machine learning, including PAC-Bayes & learning theory, transfer learning, optimal transport, robustness, fairness, and involves Dr. Farah Cherfaoui, Pr. Rémi Emonet, Pr. Amaury Habrard, Dr. Emilie Morvant. Inria Rennes brings expertise in statistical machine learning, especially in bandit theory, PAC-Bayes & learning theory, optimal transport, and involves Dr. Romaric Gaudel, Pr. Romain Tavenard, Dr. Paul Viallard. Inria Lille brings expertise in statistical machine learning theory, especially in bandit theory & regret bounds, reinforcement learning, robustness, fairness, and involves Dr. Debabrota Basu, and Dr. Odalric-Ambrym Maillard. The consortium is strengthened by the partnership with Pascal Germain and Audrey Durand from University Laval (Québec), both academic members of MILA, a leading institute in Montreal's AI ecosystem. P. Germain is an international expert in PAC-Bayes and was a member of APRIORI, and A. Durand is an international expert in bandits motivated by real-world applications.

# III References related to the project


[1] H. Atbir, **F. Cherfaoui**, G. Metzler, **E. Morvant**, and **P. Viallard**. PAC-Bayesian bounds on constrained $f$-entropic risk measures. In *Preprint arXiv:2510.11169*, 2025.

[2] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Mach. Learn.*, 2002.

[3] D. Baudry, P. Saux, and **O.-A. Maillard**. From optimality to robustness: Adaptive re-sampling strategies in stochastic bandits. In *NeurIPS*, 2021.

[4] O. Catoni. Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning. *IMS Lecture Notes Monograph Series*, 2007.

[5] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, **R. Tavenard**, A. Tong, and T. Vayer. POT: Python Optimal Transport. *JMLR*, 2021.

[6] H. Flynn, D. Reeb, M. Kandemir, and J. Peters. PAC-Bayes Bounds for Bandit Problems: A Survey and Experimental Comparison. *IEEE TPAMI*, 2023.

[7] Y. Freund. Self Bounding Learning Algorithms. In *COLT*, 1998.

[8] C. Gauthier, **R. Gaudel**, and É. Fromont. Bandit Algorithm for both Unknown Best Position and Best Item Display on Web Pages. In *IDA*, 2021.

[9] C. Gauthier, **R. Gaudel**, É. Fromont, and B. A. Lompo. Parametric Graph for Unimodal Ranking Bandit. In *ICML*, 2021.

[10] B. Ghosh, **D. Basu**, and K. S. Meel. Algorithmic fairness verification with graphical models. In *AAAI*, 2022.

[11] T. Kerdoncuff, **R. Emonet**, and M. Sebban. Metric Learning in Optimal Transport for Domain Adaptation. In *IJCAI*, 2021.

[12] T. Lattimore, K. Crammer, and C. Szepesvári. Optimal Resource Allocation with Semi-Bandit Feedback. In *UAI*, 2014.

[13] T. Mathieu, **D. Basu**, and **O.-A. Maillard**. Bandits with Stochastic Corruption: Lower Bounds on Regret and Robust Optimistic Algorithms. *TMLR*, 2024.

[14] D. McAllester. Some PAC-Bayesian Theorems. In *COLT*, 1998.

[15] F. Painblanc, L. Chapel, N. Courty, C. Friguet, C. Pelletier, and **R. Tavenard**. Match-And-Deform: Time Series Domain Adaptation through Optimal Transport and Temporal Alignment. In *ECML-PKDD*, 2023.

[16] J. Patracone, **P. Viallard**, **E. Morvant**, G. Gasso, **A. Habrard**, and S. Canu. A Theoretically Grounded Extension of Universal Attacks from the Attacker's Viewpoint. In *ECML-PKDD*, 2024.

[17] I. Redko, **E. Morvant**, **A. Habrard**, M. Sebban, and Y. Bennani. *Advances in Domain Adaptation Theory*. Elsevier, 2019.

[18] H. Robbins. Some aspects of the sequential design of experiments. *Bull. Am. Math. Soc.*, 1952.

[19] H. Saber, L. Saci, **O.-A. Maillard**, and **A. Durand**. Routine bandits: Minimizing regret on recurring problems. In *ECML-PKDD*, 2021.

[20] O. Sakhi, P. Alquier, and N. Chopin. PAC-Bayesian Offline Contextual Bandits With Guarantees. In *ICML*, 2023.

[21] J. Shawe-Taylor and R. C. Williamson. A PAC Analysis of a Bayesian Estimator. In *COLT*, 1997.

[22] **P. Germain**, **A. Habrard**, F. Laviolette, and **E. Morvant**. A New PAC-Bayesian Perspective on Domain Adaptation. In *ICML*, 2016.

[23] **P. Viallard**, **P. Germain**, **A. Habrard**, and **E. Morvant**. Self-bounding Majority Vote Learning Algorithms by the Direct Minimization of a Tight PAC-Bayesian C-Bound. In *ECML-PKDD*, 2021.

[24] **P. Viallard**, G. Vidot, **A. Habrard**, and **E. Morvant**. A PAC-Bayes Analysis of Adversarial Robustness. In *NeurIPS*, 2021.

[25] **P. Viallard**, **P. Germain**, **A. Habrard**, and **E. Morvant**. A general framework for the practical disintegration of PAC-Bayesian bounds. *Mach. Learn.*, 2024.

[26] **P. Viallard**, **R. Emonet**, **A. Habrard**, **E. Morvant**, and V. Zantedeschi. Leveraging PAC-Bayes Theory and Gibbs Distributions for Generalization Bounds with Complexity Measures. In *AISTATS*, 2024.

[27] **R. Gaudel** and M. Sebag. Feature Selection as a One-Player Game. In *ICML*, 2010.

[28] W. R. Thompson. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, 1933.

[29] L. Valiant. A Theory of the Learnable. *Communications of the ACM*, 1984.

[30] S. S. Villar, J. Bowden, and J. Wason. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical Science*, 2015.

[31] V. Zantedeschi, **P. Viallard**, **E. Morvant**, **R. Emonet**, **A. Habrard**, **P. Germain**, and B. Guedj. Learning Stochastic Majority Votes by Minimizing a PAC-Bayes Generalization Bound. In *NeurIPS*, 2021.


[2] See the tutorial "*How to Make Use of Learning Theory to Learn Efficient ML Models: From PAC-Bayesian Generalization Bounds to (Self-Bounding) Learning Algorithms*" of E. Morvant & P. Viallard given at the top-tier conference COLT 2025.