# Toward the Generalization of Reinforcement Learning


## Abstract

Conventional Reinforcement Learning (RL) involves training a unimodal agent on a single, well-defined task, guided by a gradient-optimized reward signal. This framework does not allow us to envisage a learning agent adapted to real-world problems involving diverse modality streams, multiple tasks, often poorly defined, sometimes not defined at all. Hence, we advocate for transitioning towards a more general framework, aiming to create RL algorithms that more inherently versatile.

To advance in this direction, we identify two primary areas of focus. The first aspect involves improving exploration, enabling the agent to learn from the environment with reduced dependence on the reward signal. We present Latent Go-Explore (LGE), an extension of the Go-Explore algorithm. While Go-Explore achieved impressive results, it was constrained by domain-specific knowledge. LGE overcomes these limitations, offering wider applicability within a general framework. In various tested environments, LGE consistently outperforms the baselines, showcasing its enhanced effectiveness and versatility. The second focus is to design a general-purpose agent that can operate in a variety of environments, thus involving a multimodal structure and even transcending the conventional sequential framework of RL. We introduce Jack of All Trades (JAT), a multimodal Transformer-based architecture uniquely tailored to sequential decision tasks. Using a single set of weights, JAT demonstrates robustness and versatility, competing its unique baseline on several RL benchmarks and even showing promising performance on vision and textual tasks. We believe that these two contributions are a valuable step towards a more general approach to RL. In addition, we present other methodological and technical advances that are closely related to our core research question. The first is the introduction of a set of sparsely rewarded simulated robotic environments designed to provide the community with the necessary tools for learning under conditions of low supervision. Notably, three years after its introduction, this contribution has been widely adopted by the community and continues to receive active maintenance and support. On the other hand, we present Open RL Benchmark, our pioneering initiative to provide a comprehensive and fully tracked set of RL experiments, going beyond typical data to include all algorithm-specific and system metrics. This benchmark aims to improve research efficiency by providing out-of-the-box RL data and facilitating accurate reproducibility of experiments. With its community-driven approach, it has quickly become an important resource, documenting over 25,000 runs.

These technical and methodological advances, along with the scientific contributions described above, are intended to promote a more general approach to Reinforcement Learning and, we hope, represent a meaningful step toward the eventual development of a more operative RL agent.



## Keywords

reinforcement learning; multimodal agents; exploration; general-purpose agent; sparsely rewarded environments

# Vers la Généralisation de l’Apprentissage par Renforcement


## Résumé

L’apprentissage par renforcement conventionnel implique l’entraînement d’un agent unimodal sur une tâche unique et bien définie, guidé par un signal de récompense optimisé pour le gradient. Ce cadre ne nous permet pas d’envisager un agent d’apprentissage adapté aux problèmes du monde réel impliquant des flux de diverses modalités, des tâches multiples, souvent mal définies, voire pas définies du tout. C’est pourquoi nous préconisons une transition vers un cadre plus général, visant à créer des algorithmes d’apprentissage par renforcement plus adaptables et intrinsèquement polyvalents.

Pour progresser dans cette direction, nous identifions deux domaines d’intérêt principaux. Le premier est l’amélioration de l’exploration, qui permet à l’agent d’apprendre de l’environnement en dépendant le moins possible du signal de récompense. Nous présentons *Latent Go-Explore* (LGE), une généralisation de l’algorithme Go-Explore qui, malgré ses résultats impressionnants, était limité par une forte contrainte de connaissance du domaine. LGE atténue ces limitations et permet une application plus large dans un cadre plus général. LGE démontre son efficacité et sa polyvalence accrues en surpassant de manière significative les lignes de base dans tous les environnements testés. Le deuxième domaine d’intérêt est celui de la conception d’un agent polyvalent qui peut fonctionner dans une variété d’environnements, impliquant ainsi une structure multimodale et transcendant même le cadre séquentiel conventionnel de l’apprentissage par renforcement. Nous présentons *Jack of All Trades* (JAT), une architecture multimodale basée Transformers, spécialement conçue pour les tâches de décision séquentielle. En utilisant un seul ensemble de poids, JAT démontre sa robustesse et sa polyvalence, rivalisant avec son unique référence sur plusieurs benchmarks d’apprentissage par renforcement et montrant même des performances prometteuses sur des tâches de vision et textuelles. Nous pensons que ces deux contributions constituent une étape importante vers une approche plus générale de l’apprentissage par renforcement. En outre, nous présentons d’autres avancées méthodologiques et techniques qui sont étroitement liées à notre question de recherche initiale. La première est l’introduction d’un ensemble d’environnements robotiques simulés à récompense éparse, conçus pour fournir à la communauté les outils nécessaires à l’apprentissage dans des conditions de faible supervision. Trois ans après son introduction, cette contribution a été largement adoptée par la communauté et continue de faire l’objet d’une maintenance et d’un support actifs. D’autre part, nous présentons Open RL Benchmark, notre initiative pionnière visant à fournir un ensemble complet et entièrement enregistré d’expériences d’apprentissage par renforcement, allant au-delà des données typiques pour inclure toutes les métriques spécifiques à l’algorithme et au système. Ce benchmark vise à améliorer l’efficacité de la recherche en fournissant des données prêtes à l’emploi et en v vi facilitant la reproductibilité précise des expériences. Grâce à son approche communautaire, il est rapidement devenu une ressource importante, documentant plus de 25 000 exécutions.

Ces avancées techniques et méthodologiques, associées aux contributions scientifiques décrites ci-dessus, visent à promouvoir une approche plus générale de l’apprentissage par renforcement et, nous l’espérons, représentent une étape significative vers le développement à terme d’un agent plus opérationnel.



## Mots clés

apprentissage par renforcement ; agents multimodaux; exploration ; agent à usage général ; environnements à récompenses rares