PHD SPOTLIGHT 2024

Elia Piccoli
elia.piccoli@phd.unipi.it

Formalize and Leverage Prior Knowledge in Reinforcement Learning

Università di Pisa

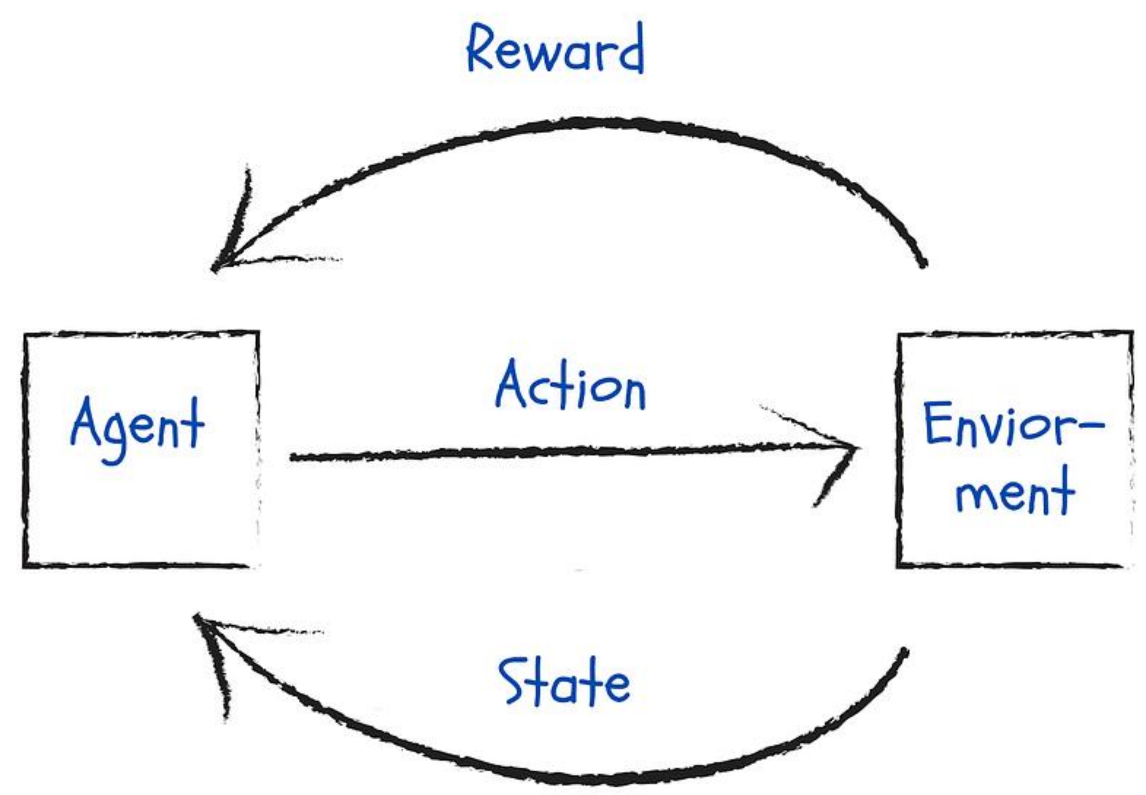
PHD SPOTLIGHT FROM STUDY TO RESEARCH

Reinforcement Learning

In Reinforcement Learning (RL) a problem can be framed as a **Markov Decision Process** (MDP), which is defined by:

- $\mathcal{S}$ - State space
- $\mathcal{A}$ - Action space
- $\mathcal{R}$ - Reward function
- $\mathcal{T}$ - Transition function

In RL, an **agent** tries to solve a task in a certain environment. At each timestep *t*, the agent performs an **action**. Based on the action, it receives a **reward** and the new **state** of the environment.



The agent's objective is to maximize the **expected discounted return**, where the return G_t is a function that maps the reward that the agent will accumulate from time step *t*.

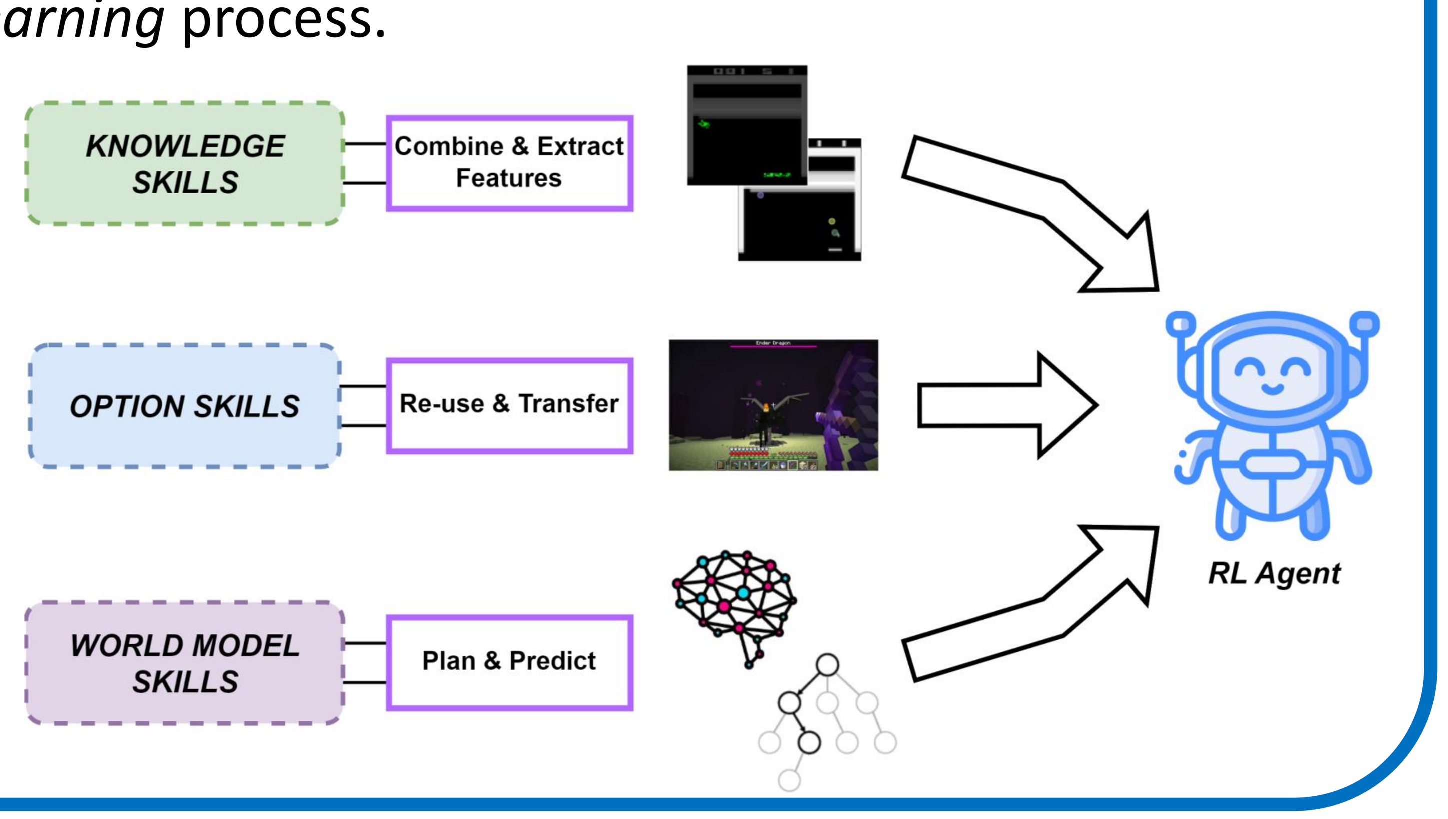
$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

Problems & Challenges

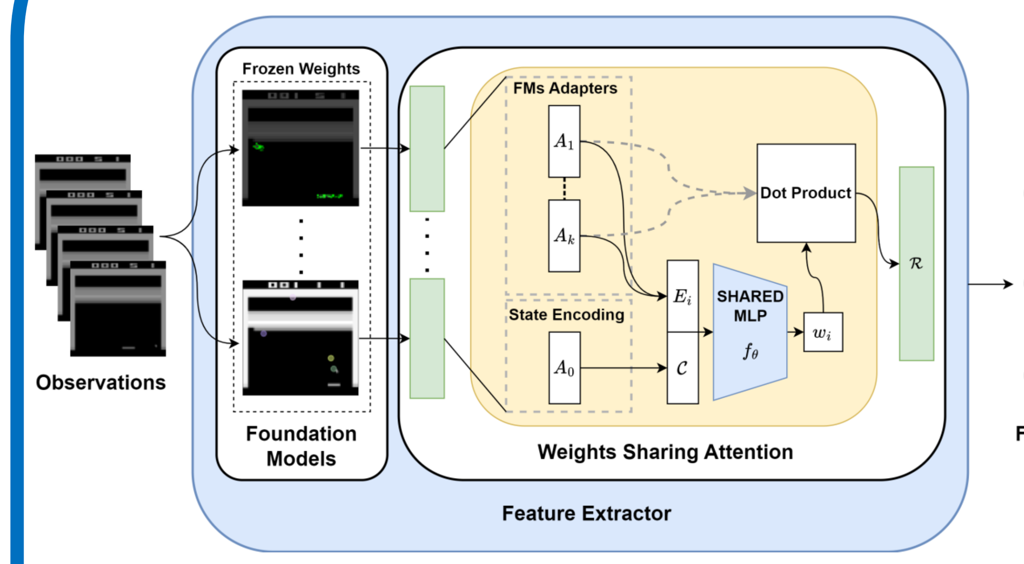
- The learning process is based on **trial-and-error**, challenging the **exploration-exploitation** dilemma.
- Agents do not have prior knowledge about the environment, **representation learning** is a key to learn optimal and robust policies.
- RL is *data hungry*. Agents need a lot of interaction with the environment thus **fast**, **reliable** and **simple simulators** are key for research.
- Generalization to **multiple task** with different *objectives* or different *observations* is hard to achieve.
- **Knowledge transfer** between tasks requires additional information and how to exploit it.

Research Questions

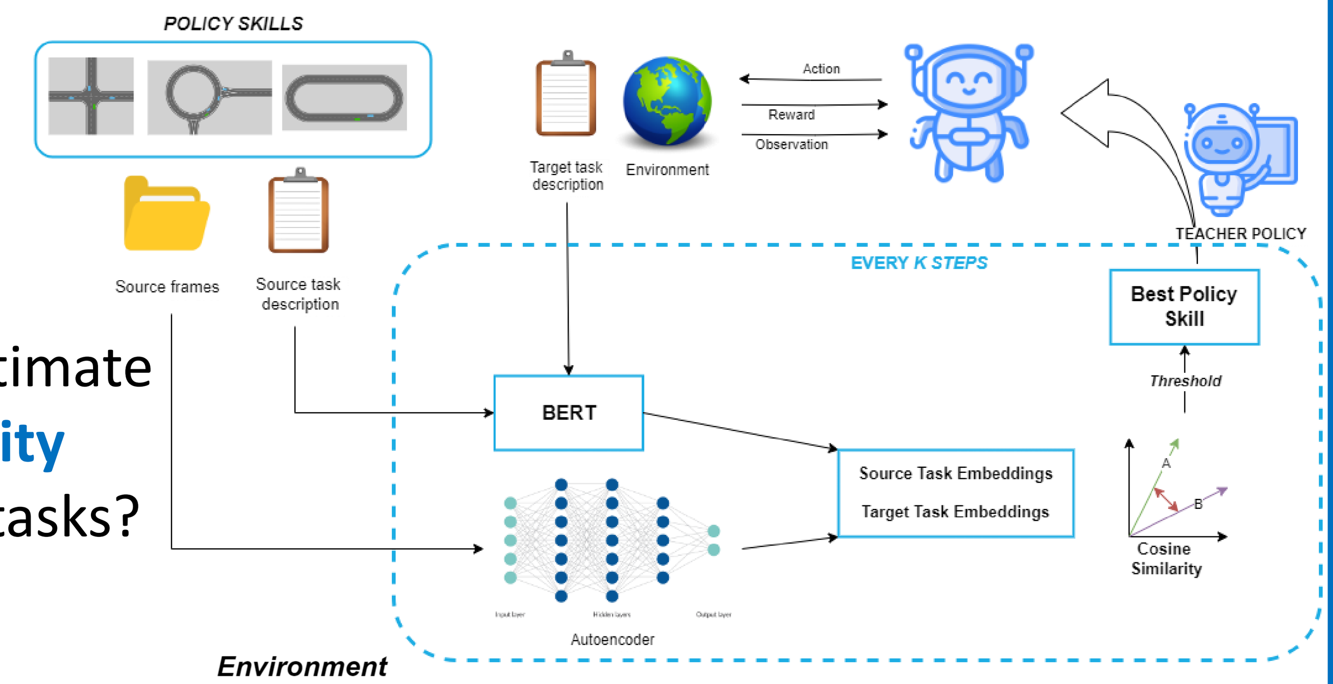
- **Formalize** prior knowledge as a family of different *Skills.*
- How to **leverage** and **combine** different skills to improve the *learning* process.



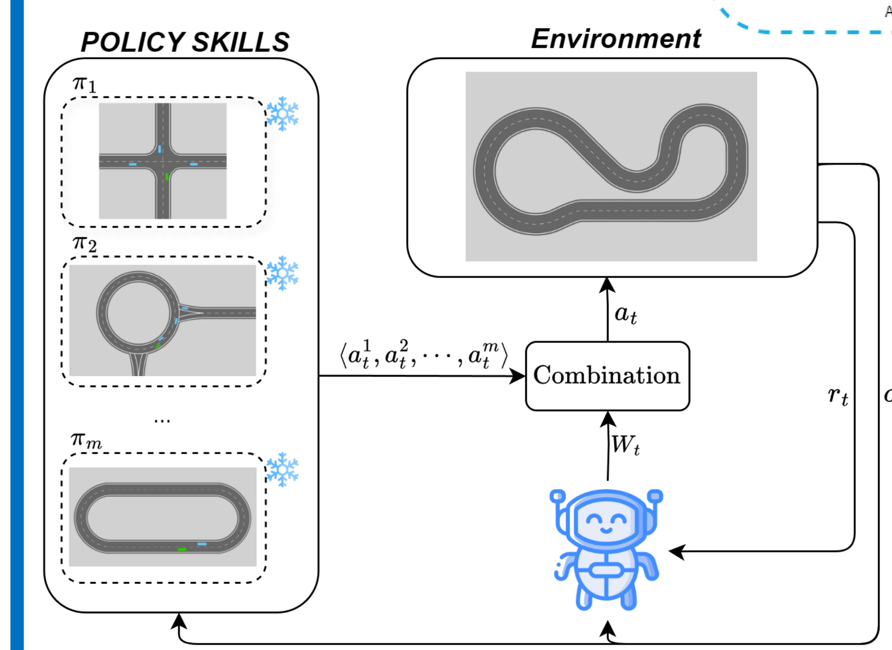
Research Projects



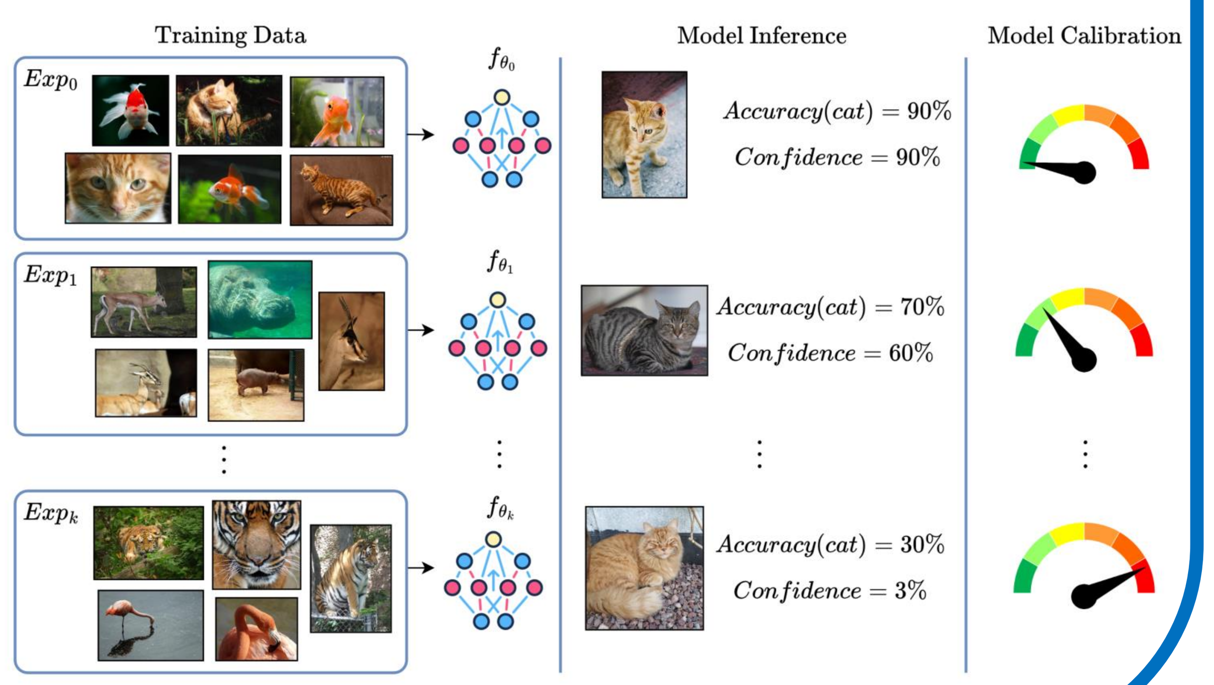
How can different *foundational models'* **representations** be combined?



How to estimate **similarity** between tasks?



Can agents learn **how to re-use** policies to solve new tasks?



Is the model able to **learn** and **update** the **confidence** in its predictions?