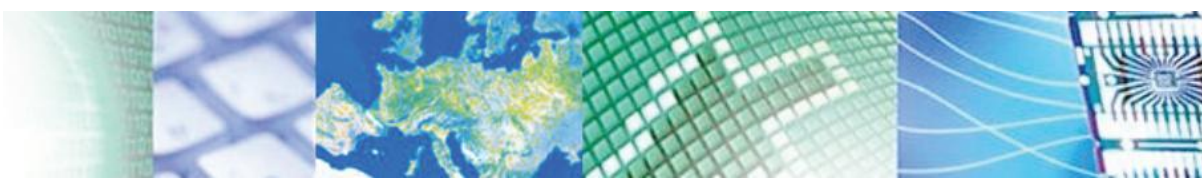




chist-era



CHIST-ERA Projects Seminar 2021

Topic XAI: CausalXRL Causal eXplainable Reinforcement Learning

***Aditya Gilra
Eleni Vasilaki***



The
University
Of
Sheffield.

Moritz Grosse-Wentrup ***Philippe Preux***



universität
wien

April 13, 2021

Inria

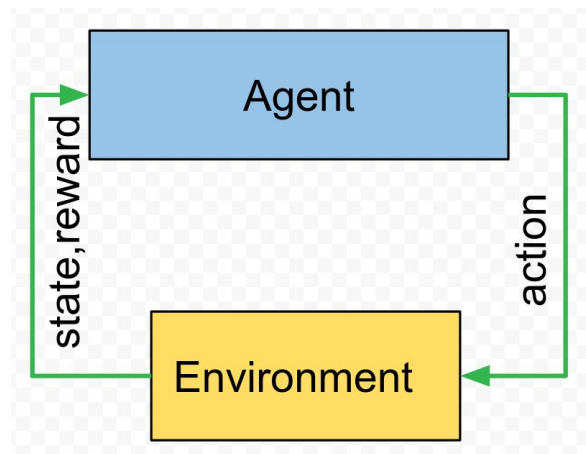


Programme co-funded by the
EUROPEAN UNION



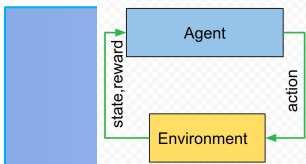
Introduction: Causal eXplainable Reinforcement Learning

- ◆ ML typically uses input-output correlations
- ◆ Reinforcement Learning (RL) uses random exploration
- ◆ To improve, similar to humans, we will:
 - Infer a causal model of the environment
 - Use that model to plan & suggest explainable actions on environment in RL loop

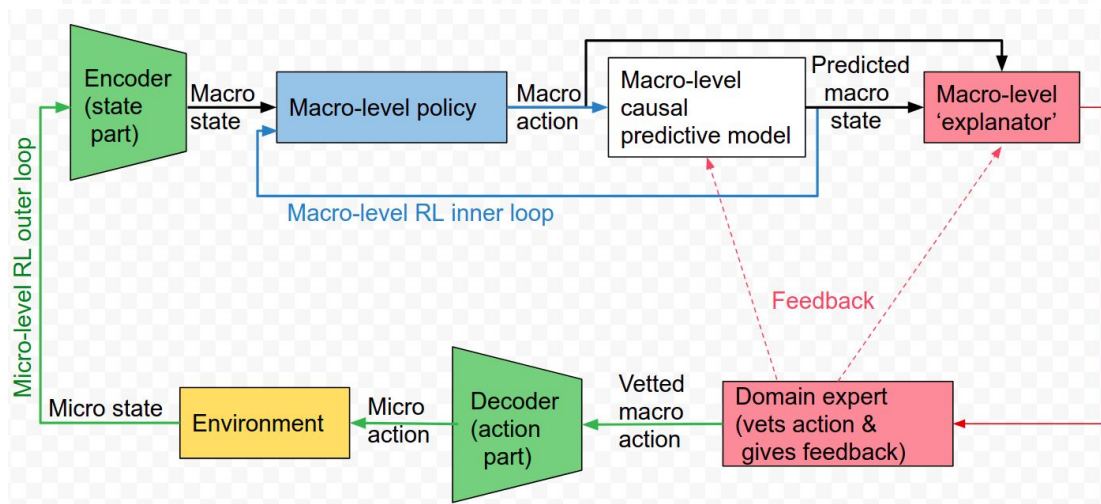
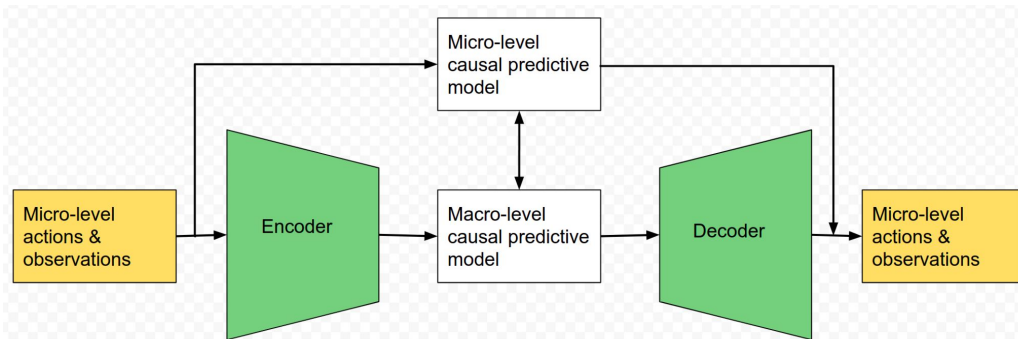




Methods and challenges



- ◆ Learn a micro/macro-level causal model
- ◆ Interpretable macro variables and actions
- ◆ Embed in an RL loop:
 - Macro-control loop
 - Macro-planning loop
 - Explanations & Expert in loop





Partners and Applications

Partners:

- ❖ University of Sheffield, UK:
 - Aditya Gilra & Eleni Vasilaki
neural RL & model learning
- ❖ University of Vienna, Austria:
 - Moritz Grosse-Wentrup
causal inference & neuro-rehab
- ❖ INRIA, Lille, France:
 - Philippe Preux
RL and applications
- ❖ Intensive care
(as baseline, not deployed)
- ❖ Post-stroke neuro-rehabilitation via
non-invasive brain stimulation
- ❖ Continuous-domain bio-plausible
implementation
- ❖ E-education
- ❖ Farming