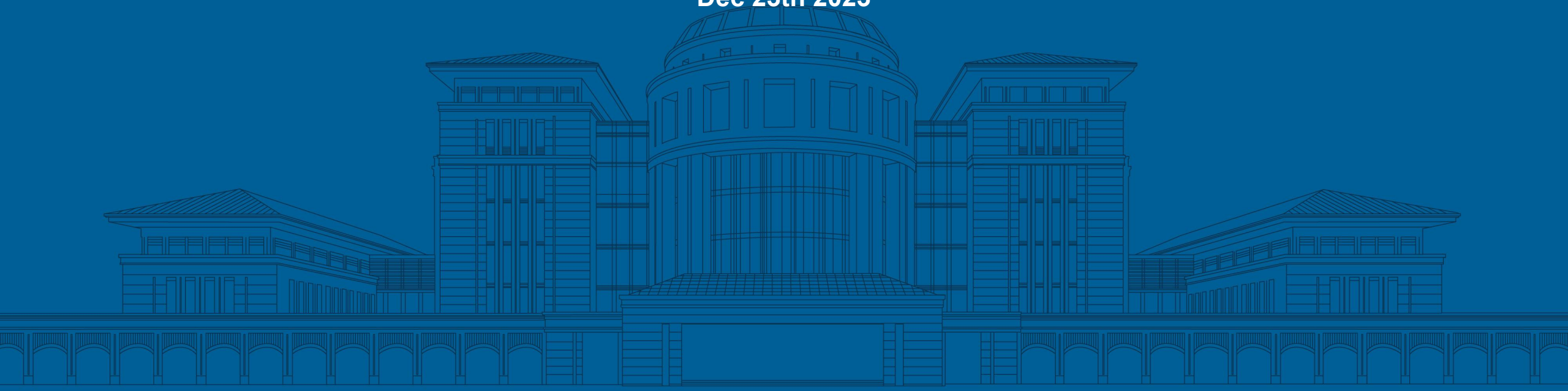
澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Reinforcement Learning for Generative AI:State of the Art, Opportunities and Open Research Challenges

speaker: Xiongyi Li

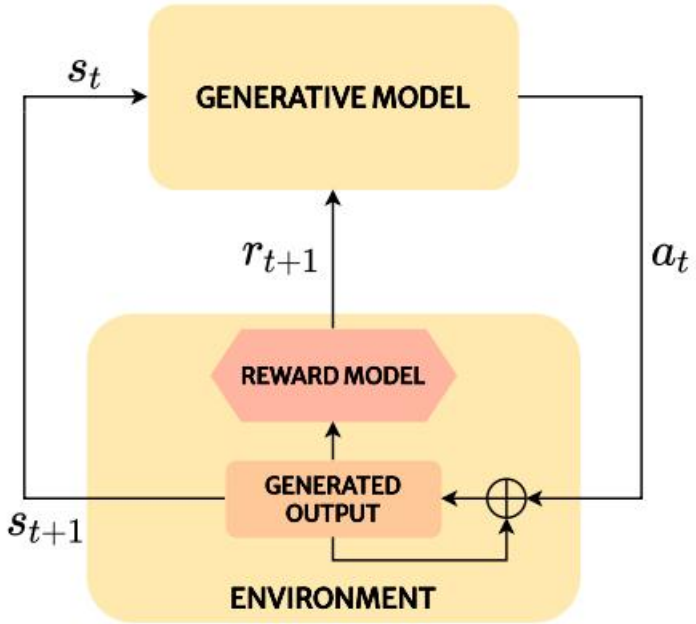
Dec 23th 2023



Main Part

- Reinforcement Learning’s application:
- as an alternative way for generation without specified objectives;
- as a way for generating outputs while concurrently maximizing an objective function;
- as a way of embedding desired characteristics, which cannot be easily captured by means of an objective function, into the generative process.

RL for mere generation



- the implementation of agent itself
- the definition of the dynamics of the system(transition)
- reward structure

} depend on the task

represent the classic supervised target

RL for mere generation

- **Reward**
- SeqGAN discriminative signal as the actual reward.
- LeakGAN hierarchical RL

RL for mere generation

- **Advantages**
- Derive generative models, even if target loss is non-differentiable
- Adapts GAN to sequential tasks
- Can implement RL strategies,like hierarchical RL, reduce the model dependence on training data

RL for mere generation

- **Limitations**
- Learning without supervision is hard
- Large action space, causing pre-training can prevent an appropriate exploration

RL for objective maximization

- The use of non-differentiable metrics as reward functions for generative learning capturing a variety of requirements and constraints.
- Adopted in text generation, molecular generation and image generation

RL for objective maximization

- **Advantages**
- Generators can be adapted for particular domains or for specific problems;
- Pre-trained models can be fine-tuned according to given requirements and specifications.;
- Any desired and quantifiable property can now be set as reward function;

RL for objective maximization

- **Limitations**
- Not every desirable property is quantifiable or easy to get
- how should we evaluate the model we derive

RL for improving not easily quantifiable characteristics

- **Reward**
- Reward modeling: learning the reward function from interaction with the user and then optimizing the agent through RL

RL for improving not easily quantifiable characteristics

- **Advantages**
- Satisfies nonquantifiable requirements

RL for improving not easily quantifiable characteristics

- **Limitations**
- Get user preferences is expensive
- Users might misbehave, disagree, or be biased
- Reward modeling is difficult (not correctly represent the population of end users or marginalized categories)

Thank You!

Avenida da Universidade, Taipa, Macau, China
Tel : (853) 8822 8833 Fax : (853) 8822 8822
Email : mc35289@um.edu.mo Website : www.um.edu.mo