课程编号 课程名称 深度强化学习

1. **课堂讲授学时 Lecture Hours 16 学时**
2. **课堂实验学时 Laboratory Hours 16 学时**
3. **课下研讨学时 Colloquia Hours 0**
4. **学生课下投入学时 Individual Study Hours 16 学时**
5. **学分 Credits 2 学分**
6. **开课学年学期**（如果有强制性的要求则必须填，否则可以不填）**Occurrence：1st year, 2nd year, 3rd year, 4th year; Autumn, Spring**
7. **先修课程 Prerequisite(s): 必须先修的课程直接写课程编号和课程名称，建议先修的课程在课程名称后用*号标注，并在下一行注明：*Recommended, not required as prerequisite 机器学习、深度学习、线性代数、概率论、微积分**
8. **课程概要 Course Description:** 100 字以内，学习内容以学术关键词出现。

本课程覆盖深度强化学习基础与前沿，含无模型 / 有模型、离线强化学习、模仿学习、元强化学习、分层强化学习、LLM 优化、机器人 Sim-to-Real，侧重 PyTorch 实践与高维决策。

9. **课程预期学习成果 Course Outcomes:** 用数字 1 到 9 列出每一项主要学习成果

(1) 掌握深度强化学习核心算法
(2) 理解马尔可夫决策过程（MDPs）
(3) 掌握策略梯度与演员 - 评论家方法
(4) 掌握离线强化学习与奖励学习
(5) 掌握面向大语言模型的 RL 偏好优化与推理
(6) 掌握基于模型的深度强化学习
(7) 掌握多任务、目标条件、元强化学习
(8) 掌握分层 RL 与模仿学习
(9) 能使用 PyTorch 实现深度强化学习模型并应用于机器人与语言任务

10. **教学内容与学时分配 Course Content, Laboratories and Laboratory Hours（有则填，没有则不填）, Colloquia Hours（有则填，没有则不填）：**各章节目录与学时，实验内容与学时，研讨内容与学时

(1) 课程导论、MDP、模仿学习、PyTorch 教程：2 学时 讲授 + 2 学时 实验
(2) 策略梯度、演员 - 评论家方法：2 学时 讲授 + 2 学时 实验
(3) 高级演员 - 评论家、离线策略、Q 学习：2 学时 讲授 + 2 学时 实验
(4) 离线强化学习、奖励学习：2 学时 讲授 + 2 学时 实验
(5) RL+LLM 偏好优化、推理：2 学时 讲授 + 2 学时 实验

(6) 基于模型 RL、多任务、目标条件 RL：2 学时 讲授 + 2 学时 实验

(7) 元强化学习、探索策略：2 学时 讲授 + 2 学时 实验

(8) 分层 RL、模仿学习、机器人自主学习、Sim-to-Real、领域前沿：2 学时 讲授 + 2 学时 实验

11. 考核与成绩评定 Grading:

Homework + Project: 100%

12. **教材，参考书 Text & Reference Book:** 作者，书名，版本，年份，国际标准书号 ISBN

13. 编写教师 Course Lecturer：

编写教师 Course Lecturer（签字）：Ayush Joshi, Sanyuan Zhao

附录：英文课程教学大纲模板

附录：

Deep Reinforcement Learning

1. 课堂讲授学时 Lecture Hours: 16
2. 课堂实验学时 Laboratory Hours: 16
3. 课下研讨学时 Colloquial Hours: 0
4. 学生课下投入学时 Individual Study Hours: 32
5. 学分 Credits: 2
6. 开课学年学期（如果有强制性的要求则 必须填， 否则可以不填） Occurrence: Summer Course
7. 先修课程 Prerequisite(s): Programming-related courses* (*Recommended, not required as prerequisite)

   Machine Learning, Deep Learning, Linear Algebra, Probability Theory, Calculus
8. 课程概要 Course Description: 100 字以内 ， 学习内容以学术 关键词出现。

   This course covers the fundamentals and frontiers of deep reinforcement learning, including model-free/model-based methods, offline reinforcement learning, imitation learning, meta reinforcement learning, hierarchical reinforcement learning, LLM optimization, and robotic Sim-to-Real, with an emphasis on PyTorch practice and high-dimensional decision-making.
9. 课程预期学习成果 Course Outcomes:

   Students are expected to master fundamental topics in deep reinforcement learning and apply them to complex domains.

   Specific measurable achievements and course deliverables include:

(1) Mastering Deep RL Algorithms: Students will learn practical methods for learning behavior from experience using deep neural networks, specifically focusing on high-dimensional observations.

(2) Core and Advanced Techniques: The curriculum requires students to understand both model-based and model-free deep RL methods, how to learn from demonstrations and offline datasets, and advanced multi-task techniques such as goal-conditioned RL, meta-RL, and unsupervised skill discovery.

(3) Application to Complex Domains: Students will learn how to instantiate these methods in domains with high-dimensional state and action spaces, with a strong emphasis on use-cases in robotics (e.g., visual navigation and control) and language modeling.

(4) Practical Implementation and Research: Students are expected to gain hands-on experience by training neural networks in PyTorch

10. 教学内容与学时分配 Course Content, Laboratories and Laboratory Hours（有则填，没有则不填）, Colloquial Hours（有则填，没有则不填）;

(1) Module 1[2 hour Lecture, 2 hour labs]: Course introduction and Markov Decision Processes (MDPs), followed by Imitation Learning and an additional PyTorch tutorial TA session.

(2) Module 2[2 hour Lecture, 2 hour labs]: Introduction to Policy Gradients and Actor-Critic Methods.

(3) Module 3[2 hour Lecture, 2 hour labs]: Advanced Actor-Critic methods with Off-Policy Actor Critic and Q-learning, supplemented by an extra section on Q-learning.

(4) Module 4[2 hour Lecture, 2 hour labs]: Offline Reinforcement Learning and Reward

Learning.

(5) Module 5[2 hour Lecture, 2 hour labs]: We Focus on Modern Innovations on RL for LLMs: Preference Optimization and RL for LLMs: Reasoning.

(6) Module 6[2 hour Lecture, 2 hour labs]: Model-based Reinforcement Learning, followed by Multi-Task and Goal-Conditioned RL.

(7) Module 7[2 hour Lecture, 2 hour labs]: Meta-Reinforcement Learning and exploration strategies.

(8) Module 8[2 hour Lecture, 2 hour labs]: Hierarchical RL and Imitation Learning (IL), followed by Autonomous Learning for Robots. Lecture on Sim-to-Real Transfer for robotics and concluding with a lecture on the Frontiers of the field.

11. 考核与成绩评定：

Homework + Project: 100%

12. 教材，参考书 Text & Reference Book: Yang, C. (2017). Introduction to GIS Programming and Fundamentals with Python and ArcGIS. (1st edition). CRC Press

13. 编写教师Course Lecturer: Ayush Joshi, Sanyuan Zhao

编写教师 Course Lecturer（签字）：