JCSDS

第四届统计与数据科学联合会议

THE 4TH JOINT CONFERENCE ON STATISTICS AND DATA SCIENCE

主办单位：

第四届统计与数据科学联合会议会务组

贵州财经大学数学与统计学院、大数据统计学院

JCSDS

贵州财经大学 GUIZHOU UNIVERSITY OF FINANCE AND ECONOMICS

第四届统计与数据科学联合会议

THE 4th JOINT CONFERENCE ON STATISTICS AND DATA SCIENCE

暨IMS China双年会

IMS CHINA BIENNIAL MEETING

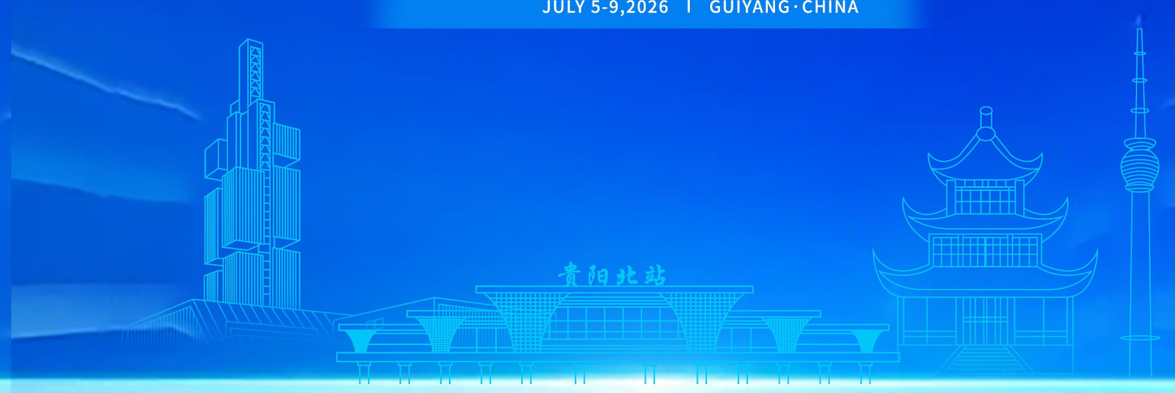
短期课程

SHORT COURSE

会议手册

2026年7月5日-9日 | 中国·贵阳

JULY 5-9,2026 | GUIYANG·CHINA



目　录

一、课程指引

（一）授课时间地点

- **授课时间**：2026年7月5日~9日
- **授课地点**：贵州财经大学
- **授课地址**：贵州省贵阳市花溪区贵州财经大学修德楼2楼218室

（二）报到安排

- **报到时间：**2026年7月4日14:00~20:00
- **报到地点：**昆朋酒店(贵阳花溪大学城店）一楼大厅
- **报到流程**

① 已完成线上注册的参会人员，请核对个人信息并签字确认；未完成缴费者请现场办理缴费手续。

② 物料领取：领取参会证件与会议手册，请妥善保管，会场入场需凭参会证通行。

③ 酒店入住：前往酒店办理入住登记。

注：2026年7月4日以后请至贵州财经大学修德楼2楼206室报到

（三）餐食安排

- 所有学员在报到**当天**，可以选择**一次性**购买短课期间午餐券

（用餐凭证，**请妥善保管**），届时凭借餐券，至贵州财经大学竹香园（第二食堂）三楼用餐。

（四）交通信息

- **贵阳北站→贵州财经大学：**全程约30公里，打车约40分钟。
- **贵阳东站→贵州财经大学：**全程约40公里，打车约50分钟。
- **贵阳龙洞堡国际机场→贵州财经大学：**全程约35公里，打车约45分钟。
- **地铁到达：**乘坐贵阳地铁**3号线**至花溪公园站（**B口**），换乘公交402路或404路花溪大学城方向至花燕路，步行100米至贵州财经大学。

  乘坐贵阳地铁**3号线**至桐木岭（省委党校）站，后打车至贵州财经大学约10分钟。

•接驳车安排（7月 5日~9日）：每日 08:30 昆朋酒店 → 贵州财经大学，每日 17:30 贵州财经大学 → 昆朋酒店

注：为方便进校报备事宜，请您从贵州财经大学西门进出校园。

二、短课日程安排

日期	时间	课程名称	主讲嘉宾
7月5日	09:00-12:00	Intelligent Agent: Statistical Framework, Theory and Applications	严晓东 西安交通大学
	14:30-17:30	CP-Factorization for High Dimensional Tensor Time Series and Double Projection Iterations	常晋源 西南财经大学
7月6日	09:00-12:00	Generative Models for Statistical Inference and Decision-Making	Yao Xie Georgia Institute of Technology
	14:30-17:30	Foundations of Diffusion Generative Models	Gen Li Chinese University of Hong Kong
7月7日	09:00-12:00	Spectral Methods for Data Science: A Statistical Perspective	Cong Ma University of Chicago
	14:30-17:30	Reinforcement Learning and Its Application in LLM	周帆 上海财经大学
7月8日	09:00-12:00	Model Averaging and Related Studies	张新雨 中国科学院数学与系统科学研究院
	14:30-17:30	Causal Inference	Peng Ding University of California, Berkeley
7月9日	09:00-12:00	Real-world Causal Problems and Solutions: Bias Reduction and Data Integration	Yong Chen University of Pennsylvania Yumou Qiu Peking University
	14:30-17:30	The Factor Modelling Approach to High-dimensional Time Series Analysis	涂云东 北京大学

三、短课授课专家及课程简介

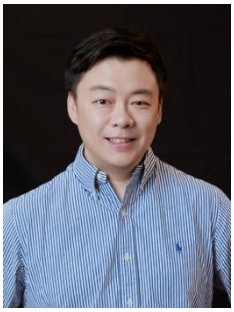
（按姓名字母排序）

常晋源，西南财经大学光华首席教授，国家杰青，享受国务院政府特殊津贴专家，四川省基础研究领衔科学家，教育部中央高校优秀青年团队带头人，主要从事大规模复杂数据分析相关的研究。先后担任统计学，计量经济学和运筹管理国际顶级学术期刊Journal of the Royal Statistical Society Series B, Journal of Business&Economic Statistics,Journal of the American Statistical Association和Operations Research的副主编，获霍英东教育基金会高等院校青年教师奖一等奖和青年科学奖一等奖，教育部高等学校科学研究优秀成果奖，四川省青年科技奖等多项奖励。

课程名称：CP-Factorization for High Dimensional Tensor Time Series and Double Projection Iterations

课程简介：We adopt the canonical polyadic (CP) decomposition to model high-dimensional tensor time series. Our primary goal is to identify and estimate the factor loadings in the CP decomposition. We propose a one-pass estimation procedure through standard eigen-analysis for a matrix constructed based on the serial dependence structure of the data. The asymptotic properties of the proposed estimator are established under a general setting as long as the factor loading vectors are linearly independent, allowing the factors to be correlated and the factor loading vectors to be not nearly orthogonal. The procedure adapts to the sparsity of the factor loading vectors, accommodates weak factors, and demonstrates strong performance across a wide range of scenarios. To further reduce estimation errors, we also introduce an iterative algorithm based on a novel double projection approach. We theoretically justify the improved convergence rate of the iterative estimator, and derive the associated limiting distribution. A consistent estimator of the asymptotic variance is also provided, which plays a key role in the related inference problems. All results are validated through extensive simulations and two real data applications.

Dr. Yong Chen is Professor of Biostatistics and Biomedical Informatics at the University of Pennsylvania Perelman School of Medicine and Founding Director of the PennCIL Lab (Penn Computing, Inference and Learning) and a type I center CHASE (Center for Health AI and Synthesis of Evidence) with 30 trainees and staff. His work focuses on generating regulatory-grade real-world evidence using large-scale EHR and oncology datasets, including Flatiron, TriNetX, and Epic Cosmos. This work on real-world data has been featured by The New York Times and TIME, and one study was selected as a Top 10 Clinical Research Achievement nationally by the Clinical Research Forum—one of the most prestigious honors in U.S. clinical research. He develops AI and statistical frameworks to address data quality, bias calibration, and scalable evidence generation for drug repurposing, label expansion, and trial optimization across distributed research networks. Dr. Chen has published over 300 papers, including more than 100 in leading clinical journals, and serves as Statistical Editor for Annals of Internal Medicine, Statistical Consultant for NEJM AI, and Associate Editor for JASA and The Annals of Applied Statistics. He is an elected Fellow of the ASA and American College of Medical Informatics.



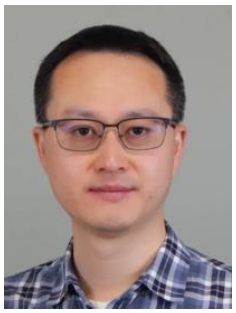
Yumou Qiu is a tenured Associate Professor in the School of Mathematical Sciences and the Center for Statistical Science at Peking University. He received his Ph.D. in Statistics from Iowa State University. Prior to joining Peking University, he held faculty positions at Iowa State University and the University of Nebraska–Lincoln. His research in statistical methodology and theory focuses on high-dimensional data analysis, causal inference, missing data, and survey sampling. Motivated by fundamental scientific questions and real-world applications, his work also spans interdisciplinary areas, including plant science, epidemiology, forensic science, and ocean science. He has taught over ten different courses and has experience with tutorials and short courses.

课程名称：Real-world Causal Problems and Solutions: Bias Reduction and Data Integration

课程简介：Modern biomedical research increasingly relies on multi-institutional data networks built on electronic health records (EHRs), claims databases, and registries, such as OHDSI, PCORnet, and Sentinel. These infrastructures enable large-scale studies across diverse populations but introduce two fundamental challenges: (1) data-sharing restrictions that prevent exchange of individual-level data across institutions, and (2) residual systematic bias arising from unmeasured confounding, measurement error, and heterogeneous data-generating processes across sites. Addressing these challenges requires statistical frameworks that jointly account for privacy constraints and inferential validity.

This short course introduces a unified framework for federated evidence generation that integrates federated learning with bias calibration using negative control outcomes (NCOs). Federated approaches enable collaborative analyses using only summary-level information, while negative controls provide a principled mechanism for diagnosing and calibrating residual bias in observational studies.

The course presents methodological foundations and practical workflows for combining these ideas in distributed research networks. Topics include privacy-preserving distributed algorithms, summary-level causal inference, negative control diagnostics, and bias calibration for large-scale real-world evidence studies



Peng Ding is an Associate Professor in the Department of Statistics at UC Berkeley. He obtained my Ph.D. from the Department of Statistics, Harvard University in May 2015 and worked as a postdoctoral researcher in the Department of Epidemiology, Harvard T. H. Chan School of Public Health until December 2015. Previously, he received my B.S. in Mathematics, B.A. in Economics, and M.S. in Statistics from Peking University.

课程名称：Causal Inference

课程简介：This course will give a quick overview of causal inference, based on Peng Ding's book "A first course in causal inference".

Introduction to the content of the book: The past decade has witnessed an explosive growth in interest in causal inference in research and education, thanks to its widespread applications in biomedical research, social sciences, artificial intelligence, and other fields. This book is based on the causal inference course taught by the author at the University of California, Berkeley over the past seven years, and only requires basic knowledge of probability theory, statistical inference, linear and logistic regression. It covers causal inference from a statistical perspective and includes examples and applications from biostatistics and econometrics.

Gen Li is currently an assistant professor in the Department of Statistics and Data Science at the Chinese University of Hong Kong. He received the Ph.D. in the Department of Electronic Engineering at Tsinghua University in 2021, and received the bachelor's degree from the Department of Electronic Engineering and Department of Mathematics at Tsinghua University in 2016. His research interests include diffusion based generative model, reinforcement learning, high-dimensional statistics, machine learning.

课程名称：Foundations of Diffusion Generative Models

课程简介： Diffusion models have emerged as a powerful paradigm for generative modeling, enabling high-fidelity sampling by reversing a stochastic corruption process. Despite their empirical success, diffusion models face fundamental challenges in practice, especially in data-limited and compute-constrained regimes. This necessitates understanding and improving the sample complexity and computational efficiency of diffusion models. This lecture provides a unified mathematical framework for diffusion models, covering important algorithmic developments and theoretical foundations. A central goal is to characterize the fundamental principles governing diffusion-based generation and to highlight computationally efficient algorithms. The lecture aims to equip attendees with a principled understanding of diffusion models and a modern toolbox for analyzing and designing faster and more reliable diffusion-based methodologies, along with a synthesis of open problems and future research directions.

Cong Ma is an assistant professor in the Department of Statistics at the University of Chicago. Previously, he was a postdoctoral researcher at UC Berkeley, advised by Professor Martin Wainwright. he obtained his Ph.D. at Princeton University in 2020, advised by Professor Yuxin Chen and Professor Jianqing Fan. Prior to the graduate school, he received his bachelor's degree in Electrical Engineering from Tsinghua University in 2015. He is broadly interested in mathematics of data science with a current focus on reinforcement learning, transfer learning, multi-modal learning, high-dimensional statistics, and nonconvex optimization.

课程名称：Spectral Methods for Data Science: A Statistical Perspective

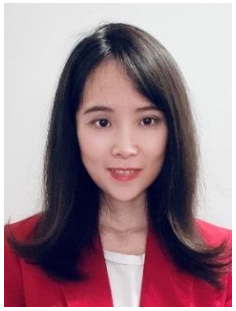
课程简介： Spectral methods have emerged as a simple yet surprisingly effective approach for extracting information from massive, noisy and incomplete data. In a nutshell, spectral methods refer to a collection of algorithms built upon the eigenvalues (resp. singular values) and eigenvectors (resp. singular vectors) of some properly designed matrices constructed from data. A diverse array of applications have been found in machine learning, data science, and signal processing. Due to their simplicity and effectiveness, spectral methods are not only used as a stand-alone estimator, but also frequently employed to initialize other more sophisticated algorithms to improve performance.

While the studies of spectral methods can be traced back to classical matrix perturbation theory and methods of moments, the past decade has witnessed tremendous theoretical advances in demystifying their efficacy through the lens of statistical modeling, with the aid of non-asymptotic random matrix theory. This monograph aims to present a systematic, comprehensive, yet accessible introduction to spectral methods from a modern statistical perspective, highlighting their algorithmic implications in diverse large-scale applications. In particular, our exposition gravitates around several central questions that span various applications: how to characterize the sample efficiency of spectral methods in reaching a target level of statistical accuracy, and how to assess their stability in the face of random noise, missing data, and adversarial corruptions? In addition to conventional $\ell 2$ perturbation analysis, we present a systematic $\ell\infty$ and $\ell 2,\infty$ perturbation theory for eigenspace and singular subspaces, which has only recently become available owing to a powerful "leave-one-out" analysis framework.

涂云东，北京大学博雅特聘教授，联合受聘于光华管理学院商务统计与经济计量系和北京大学统计科学中心。入选“日出东方”北大光华青年人才，北京大学优秀博士学位论文指导教师(2017,2021,2024)，北京大学优秀研究生导师(2024)，教育部“长江学者奖励计划”青年长江学者，国家杰出青年科学基金获得者，《计量经济学杂志》(Journal of Econometrics)和《计量经济学评论》（Econometric Reviews）当选会士(Fellow, 2025)。先后获武汉大学理学学士学位(2004)和经济学硕士学位(2006)、加州大学经济学博士学位(2012，河滨分校)。环亚太青年计量经济学者(YEAP)会议发起人和主要组织者。50余篇学术论文发表在多个国际国内知名专业杂志。著作教材《时间序列分析》由人民邮电出版社于2022年9月出版。研究领域涵盖时间序列分析、非参数计量方法、大数据分析、金融计量和预测等。

课程名称：The Factor Modelling Approach to High-dimensional Time Series Analysis

课程简介：This lecture covers a series of recent theoretical and empirical studies on factor models with structural instabilities, group structures, nonlinear extensions, and novel modal structures. We first introduce generalized factor models accommodating multiple structural breaks, multiple threshold effects, and their unified framework that jointly captures breaks and regime-switching behaviors in the form of factor trees. Second, we develop group identification and homogeneity pursuit methods for high-dimensional approximate factor models, including multilevel factor structures with endogenous grouping. Third, we extend factor-augmented regressions to nonlinear forecasting and quantile prediction under structural instability and model uncertainty. Finally, we propose a new class of modal factor models centered on mode-based factor analysis. Collectively, these contributions advance the econometric methodology for high-dimensional data modeling, structural inference, grouping, and prediction in macroeconomics and finance.



Yao Xie is the Coca-Cola Foundation Chair and Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology, where she also serves as Associate Director of the Machine Learning Center (ML@GT). She received her Ph.D. in Electrical Engineering with a minor in Mathematics from Stanford University and was previously a Research Scientist at Duke University. Her research lies at the intersection of statistics, machine learning, and optimization, with a focus on statistically sound and computationally efficient methods for high-dimensional, sequential, and spatio-temporal data. She is a member of the 2026 cohort of the National Academies' New Voices in Sciences, Engineering, and Medicine program and the IEEE Information Theory Society Distinguished Lecturer for 2026 to 2027. Her honors include the NSF CAREER Award, the INFORMS Gaver Early Career Award for Excellence in Operations Research, and the C.W.S. Woodroofe Award. She serves as an Associate Editor for multiple journals, including Operations Research, IEEE Transactions on Information Theory, and the Journal of the American Statistical Association (Theory and Methods), and as an Area Chair for NeurIPS, ICML, and ICLR.

报告题目: Generative Models for Statistical Inference and Deceision Making

报告简介: Generative models, including diffusion models, normalizing flows, and flow-matching methods, have become central tools for representing high-dimensional distributions. While they are often introduced as methods for learning and sampling from data distributions, they also provide a useful framework for statistical inference and decision-making under uncertainty. This short course will present a unified view of generative modeling as flows in probability space, with an emphasis on the mathematical ideas that connect modern generative models with statistics, optimization, and operations research.

The course will begin with the basic principles of flow-based and diffusion-based generative models, including transport maps, continuous-time dynamics, and matching-based training. We will then discuss how these models can be used for statistical inference, including uncertainty quantification, posterior sampling, conformal prediction, and inference under structured distributional shift. The final part of the course will focus on decision-making applications, including robust optimization, worst-case generation, causal interventions, counterfactual prediction, and mean-field game formulations. Throughout the course, the emphasis will be on how generative models can be used not only for data imitation, but also for representing how distributions change under constraints, interventions, objectives, or strategic interactions.

严晓东，西安交通大学教授，博士生导师，入选国家级青年人才项目和校内青拔A类支持计划，研究方向为智能体统计学习。荣获“华为火花奖”，“滴滴盖亚学者”。学术成果发表在综合性著名期刊PNAS，统计或经济著名期刊JRSSB, AOS, JASA, JOE以及人工智能顶会 NeurIPS, ICML, AAAI 等。在“高等教育出版社”等出版了《机器学习》、《数据科学实践基础-基于R》、《数据科学导论-基于Python》和《数据科学时间序列分析》等5部教材。

课程名称：Agentic AI: Statistical Framework, Theory and Applications

课程简介：Multi-agent systems represent a critical technical pathway for advancing artificial intelligence from content generation to autonomous task execution. However, the absence of rigorous mathematical foundations severely limits the interpretability, optimality, and trustworthiness of such systems. This talk systematically addresses the theoretical foundations and practical applications of multi-agent learning. First, we establish a formal mathematical framework for multi-agent learning, providing axiomatic definitions of state spaces, action spaces, policy mappings, and value functions. Second, we develop the statistical theory of multi-agent learning under the nonlinear expectation framework, including the strategic law of large numbers and the strategic central limit theorem, which provide rigorous probabilistic characterizations of asymptotic policy optimality. Third, we derive the Bellman optimality equation in multi-agent settings and construct the core algorithmic framework of policy iteration and value iteration. Furthermore, we demonstrate the application of these theoretical results to mathematical optimization and statistical inference problems. Finally, we present practical deployments in domestic chip reliability detection and ride-hailing dispatch test. This talk aims to provide a complete theoretical chain from mathematical framework to engineering application for multi-agent artificial intelligence.

张新雨，中国科学院数学与系统科学研究院研究员。2010年在中科院系统所获博士学位，曾是TAMU博士后和PSU的Research Fellow。担任期刊《Journal of Systems Science and Complexity》领域主编、期刊《Statistical Analysis and Data Mining》Associate Editor、期刊《系统科学与数学》和《应用概率统计》编委，是中国统筹法优选法与经济数学研究会数据科学分会副理事长和国际统计学会当选会员。先后主持国家自然科学基金委优秀和杰出青年研究基金项目，曾获得中国管理学青年奖和中科院优秀博士学位论文等奖励。主要从事统计学和计量经济学的理论和应用研究工作,具体研究方向包括模型平均、机器学习、组合预测和卫生统计等。发表了50多篇学术论文,其中20余篇论文发表在Annals of Statistics、Biometrika、JASA、JRSSB、Journal of Econometrics和Econometric Theory。

课程名称：Model Averaging and Related Studies

课程简介： In this lecture, I will introduce the fundamental concepts and cutting-edge advances of model averaging. Especially, this lecture will cover interdisciplinary research integrating model averaging with transfer learning, prediction-assisted inference and neural networks.

周帆，上海财经大学统计与数据科学学院副教授，国家级青年人才，博士毕业于美国北卡罗来纳大学教堂山分校，现担任统计学顶刊JASA的副主编与人工智能顶会NeurIPS等的领域主席。研究兴趣包括深度学习，强化学习的算法与理论，大模型，因果推断，在包括JASA, JMLR, NeurIPS, ICML, ICLR等统计学，机器学习顶刊和顶会上发表一作通讯文章数十篇，曾获泛华统计协会国际会议新研究者奖,UNC James E. Grizzle Distinguished Alumnus Award和Barry H. Margolin Award.

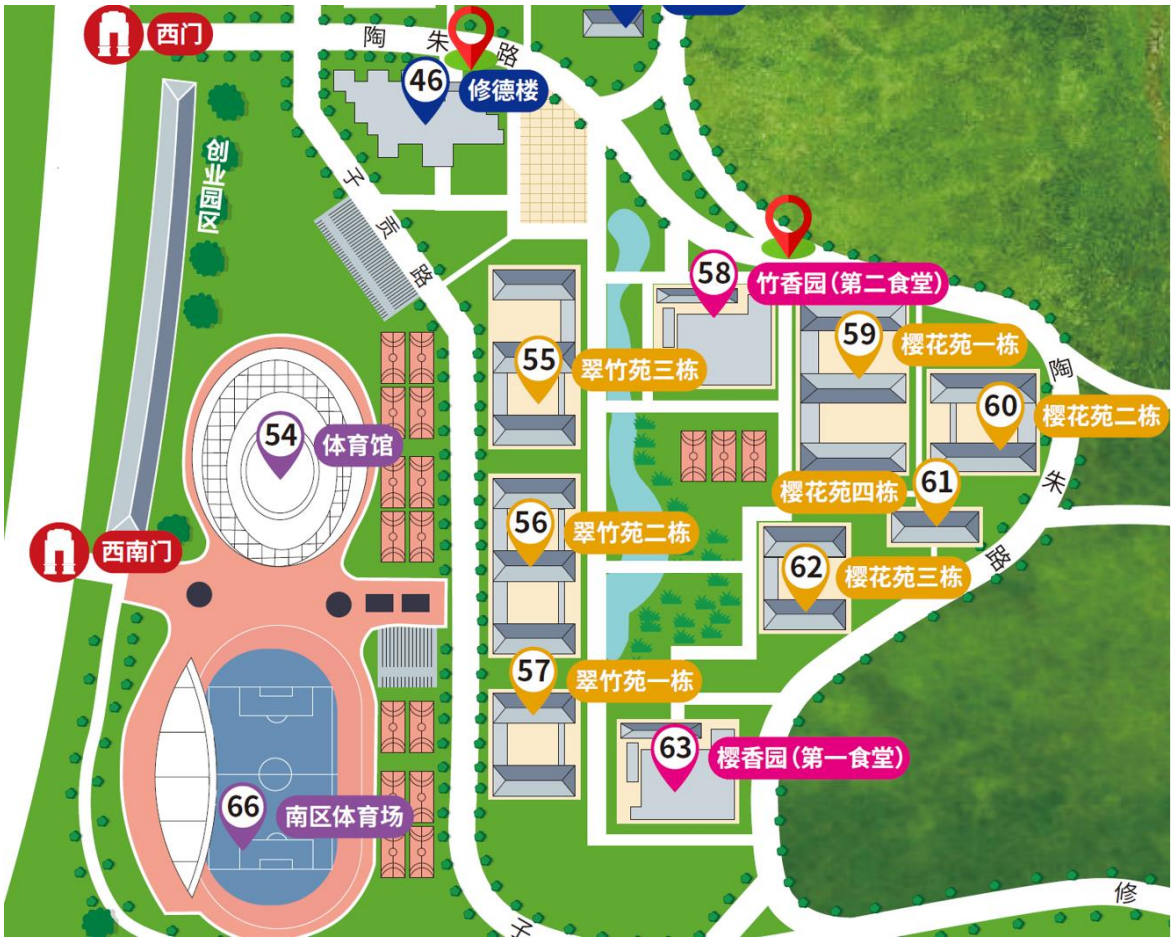
课程名称：Reinforcement Learning and Its Application in LLM

课程简介：本报告将系统介绍强化学习（Reinforcement Learning, RL）的基本理论及其在大语言模型（Large Language Models, LLMs）中的关键应用。报告首先从马尔可夫决策过程、价值函数和策略优化等强化学习基础知识出发，进一步介绍近年来快速发展的分布式强化学习（Distributional Reinforcement Learning）及其在不确定性刻画和决策优化中的优势。在此基础上，结合大语言模型的发展背景，重点讨论如何利用强化学习实现模型与人类偏好的对齐（Reinforcement Learning from Human Feedback, RLHF），并介绍PPO、DPO、GRPO等主流优化方法的核心思想与技术特点。最后，以ChatGPT系列模型和DeepSeek-R1等代表性工作为例，分析强化学习在提升大模型推理能力、决策能力和价值对齐能力中的重要作用，展望强化学习与大语言模型融合发展的前沿趋势与未来挑战。

四、授课地点平面图

授课周边地图概览

- ✧ 授课地点位于贵州财经大学修德楼218教室，位置位于地图西北角 处。
- ✧ 就餐食堂为贵州财经大学竹香园（第二食堂），位置如图东北方 处，步行约100米。
- ✧ 为方便进校报备事宜，请您从贵州财经大学西门进出校园。如您自行乘坐网约车或的士，可在到达西门后，于学校岗亭处出示听课证进入校园，步行50米即可到达修德楼。



会议笔记

会议笔记

会议笔记

会议笔记

会议笔记