

课程详述

COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	计算伦理学 Computational Ethics
2.	授课院系 Originating Department	计算机科学与工程系 Department of Computer Science and Engineering
3.	课程编号 Course Code	CS340
4.	课程学分 Credit Value	3
5.	课程类别 Course Type	专业选修课 Major Elective Courses
6.	授课学期 Semester	春季 Spring
7.	授课语言 Teaching Language	中英双语 English & Chinese
8.	授课教师、所属学系、联系方式 (如属团队授课, 请列明其他授课教师) Instructor(s), Affiliation & Contact (For team teaching, please list all instructors)	危学涛, 副教授, 计算机科学与工程系, weixt@sustech.edu.cn Xuetao Wei, Associate Professor, Department of Computer Science and Engineering, weixt@sustech.edu.cn
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	待公布 To be announced
10.	选课人数限额(可不填) Maximum Enrolment (Optional)	

11. 授课方式 Delivery Method	讲授 Lectures	习题/辅导/讨论 Tutorials	实验/实习 Lab/Practical	其它(请具体注明) Other (Please specify)	总学时 Total
	32		32		64
12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	CS303 Artificial Intelligence				
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite					
14. 其它要求修读本课程的学系 Cross-listing Dept.					

教学大纲及教学日历 SYLLABUS

15. 教学目标 Course Objectives

This course covers the ethical issues students will face both as members of a technological society and as professionals in computer-related fields. The lecture topics include both classic philosophical issues, such as fairness and bias, and emerging state-of-the-art technologies, such as value alignment in large models. Through taking this course, students will understand the ethical issues that arise with the usage and development of computing technologies. More importantly, this course will guide them to learn the responsibilities of computer professionals and how to make appropriate decisions when faced with ethical issues in computing.

本课程涵盖了学生作为技术社会成员和计算机相关领域的专业人员将面临的科技伦理问题。课程内容既包括经典的哲学问题，如公平和偏见，也包括新兴的最先进技术，如大模型的价值对齐。通过学习本课程，学生将了解计算技术的使用和发展过程中出现的伦理道德问题。更重要的是，本课程将指导他们学习计算机专业人士的责任，以及在面临计算方面的伦理道德问题时如何做出适当的决定。

16. 预达学习成果 Learning Outcomes

Upon the completion of this course, the student should be able to:

Identify ethical issues that arise in emerging technologies and determine how to address them technically.

Understand the philosophical underpinnings of ethical metrics/algorithms and apply them in various computing tasks.

Recognize the ethical responsibility of ensuring computing technologies and assess ethical trade-offs in technical decisions.

完成本课程后，学生应能够：

1. 识别新兴技术中出现的伦理道德问题，并确定如何从技术上解决这些问题。

2. 理解伦理道德度量/算法的哲学基础，并将其应用于各种计算任务。

3. 认识到确保计算技术的伦理道德责任，并评估技术决策中的伦理道德权衡。

17. 课程内容及教学日历（如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）

Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

- Lecture 1: Introduction
- Lab: Setting up your environment.
- Lecture 2-3: Ethical Challenges
- Lab: Ethics tests.
- Lecture 4-5: Philosophical Foundations
- Lab: Understanding philosophical concepts.
- Lecture 6-8: Fairness and social bias
- Lab: Fairness measurement and debiasing.
- Lecture 9-10: Hate speech
- Lab: Hate speech identification.
- Lecture 11-12: Intellectual Property
- Lab: Intellectual property infringement detection.
- Lecture 13-15: Value Alignment
- Lab: Value alignment algorithms.
- Lecture 16: Other ethics principles (e.g., sustainability)
- Lab: No lab.
- 第 1 讲: 简介
- 实验: 设置实验环境。
- 第 2-3 讲: 伦理挑战
- 实验: 伦理测试。
- 第 4-5 讲: 哲学基础
- 实验: 理解哲学概念。
- 第 6-8 讲: 公平与社会偏见
- 实验: 公平性测量和去偏。
- 第 9-10 讲: 仇恨言论
- 实验: 仇恨言论识别。
- 第 11-12 讲: 知识产权
- 实验: 知识产权侵权检测。
- 第 13-15 讲: 价值对齐
- 实验: 价值对齐算法。
- 第 16 讲: 其他伦理原则 (如可持续性)
- 实验: 无实验。
-

18. 教材及其它参考资料 Textbook and Supplementary Readings

- 《人工智能伦理导引》，陈小平，中国科学技术大学出版社
- S. M. Liao, Ed., Ethics of Artificial Intelligence. New York, NY, USA: Oxford Univ. Press, 2020

课程评估 ASSESSMENT

19. 评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
出勤 Attendance		10%		
课堂表现 Class Performance				
小测验 Quiz				
课程项目 Projects		30%		

平时作业 Assignments		60%		3 assignments
期中考试 Mid-Term Test				
期末考试 Final Exam				
期末报告 Final Presentation				
其它（可根据需要 改写以上评估方 式） Others (The above may be modified as necessary)				

20. 记分方式 GRADING SYSTEM

- A. 十三级等级制 Letter Grading
 B. 二级记分制（通过/不通过） Pass/Fail Grading

课程审批 REVIEW AND APPROVAL

21. 本课程设置已经过以下责任人/委员会审议通过
This Course has been approved by the following person or committee of authority

