

COURSE DESCRIPTIONS 科目簡介

COURSES FOR TAUGHT POSTGRADUATE PROGRAMMES

AIF501 Contemporary Issues in AI (3 credits)

This course synthesises and elucidates material covered in the other core courses of the MA. It also introduces the areas covered in the electives. It proceeds by considering concrete contemporary scenarios involving AI. It focuses on how machine learning systems have been successfully deployed in many contexts, from gaming to triage, and on the issues – epistemic, practical, social, and ethical – that these raise in the present and near future. In the written assignments, students will be free to address and improve their knowledge concerning their personal AI-related interests. The students will also gain some ‘hands on’ experience using AI systems, such as ChatGPT and OpenArt.

AIF502 Introduction to Automated Reasoning (3 credits)

Artificial intelligence is possible because we can automate reasoning. While the use of mechanical devices to solve problems goes back to the abacus, the 20th and 21st centuries have seen the development of extremely powerful methods for automating reasoning.

How can something like reasoning be done by a computer? And why do programmers use particular ways to automate it? This course provides an introduction to the automation of reasoning. It also gives students the tools from logic that they will need to understand other courses in the MAAIF.

AIF503 Ethics of AI (3 credits)

This course provides an introduction to the ethics of AI for students of the MAAIF. After introducing the main ethics theories, as well as ethically relevant features of AI systems, this course discusses a series of practical ethical issues that are or will be caused by the widespread use of artificial intelligence in human society. The topics discussed range from issues that are already relevant today to other that may be exacerbated by future technologies like superintelligent AI systems. The course also discusses issues of AI governance and the legal regulation of AI.

AIF504 Minds & Machines (3 credits)

We often think of artificial intelligence by analogy with human intelligence. Moreover, some of the most impressive machine learning systems, such as deep neural networks (DNNs), are modelled on the brain. But when, and to what extent, should we take such analogies seriously? Consider a chess-playing DNN, such as AlphaZero, for example. Does this generate representations of positions on chess boards, akin to mental representations?

And does it apply concepts such as piece mobility and space advantage, as human chess masters do, in deciding which moves to make? This course draws on the philosophy of mind and psychology to tackle such questions.

AIF601 Computability and Logic (3 credits)

(Prerequisite(s): AIF502 Introduction to Automated Reasoning)

What sorts of problems can be solved with a computer? This course provides students with a clear understanding of the possibilities and limitations of computable problems through an examination of fundamental concepts and results in proof theory, computability theory, and model theory. Specific topics of coverage may include: basic set theory including diagonalisation and uncountable sets; meta-logic of first-order logic, including soundness and completeness; effective computability, undecidability, and incompleteness; the Löwenheim-Skolem and compactness theorems; philosophical issues concerning the nature of physical computers; machine learning and formal learning theory.

AIF602 AI and Risk (3 credits)

As artificial intelligence becomes increasingly powerful and less restricted in scope, researchers face an increasing number of increasingly weightier risks in developing it. These include risks involving the loss of control of AI systems, the intentional misuse of AI systems, conflicts between the goals of AI systems and the goals of researchers, and relying on AI systems whose inner functioning is opaque. After an introduction to risk and related concepts, this course explores the key risks arising from AI development as well as strategies for preventing and mitigating these risks. More specifically, the course will cover issues like the following. What is a risk? How can you mitigate a risk? How does risk arise with respect to developing AI systems? Could AI systems become too advanced or powerful for their programmers to control? How can researchers ensure that the goals of AI systems “align” with human values? Can we trust the outputs of AI systems whose internal functioning is opaque? Are there risks of an AI “arms race?” Should regulation or policy be introduced to help mitigate these risks? Are the risks of catastrophe great enough to warrant giving up on AI development altogether?

AIF603 Sci-fi, AI, and Reality (3 credits)

This course aims to provide students with a comprehensive exploration of contemporary science-fiction stories and films, coupled with a historical and theoretical understanding of these works. Our primary objective is to introduce students to major works in SF literature and visual arts, with a specific focus on themes related to AI and reality. Throughout the course, we will examine how the effects of AI can be interpreted in both an optimistic (utopian) and a pessimistic (dystopian) way. For instance, many works of science fiction have addressed the idea that AI may eventually perform most of the work that is currently done by human beings. Some science fiction authors have written stories in which this is a positive development because it improves humans’ quality of life, while others have written stories in which this is a negative development because humans become too reliant on AI. In general, authors have explored various effects that AI might have on humans and human society through both positive and negative perspectives.

AIF604 Policy, Business, and Regulatory Issues in AI (3 credits)

(Prerequisite(s): AIF501 Contemporary issues in AI)

In the coming decades, we will increasingly come to rely on AI in making public policy and business decisions. The same advantages that enable AI systems to beat humans in chess

enable it to better predict the implications of governmental policies or beat human traders in the stock market. But while these AI technologies promise to improve our decision-making in policy and business, they also pose new and serious threats. In this course, we examine the potential risks and rewards of AI for policy and business. This will include studying current efforts to regulate AI systems, e.g. the recent EU proposed Artificial Intelligence Act, China's 2023 Interim Measures for the Management of Generative Artificial Intelligence Services, and court cases covering the use of copyrighted material in AI training data sets.

AIF605 Applications of AI in Science and Medicine (3 credits)

(Prerequisite(s): AIF501 Contemporary issues in AI)

From locating black holes in the night sky, to predicting the structure of proteins, to identifying potential causes of cancer, AI techniques are now widely applied in science and medicine. Organised around a set of detailed case studies, this course examines some of the most pressing questions raised by these applications, such as: How do AI techniques complement – or supplant – more traditional scientific or therapeutic methods? To what extent should we embrace automated scientific research? And what are the potential ethical implications of these applications for clinicians?

AIF606 Anticipating AI Futures (3 credits)

This course provides an introduction to topics in the field of speculative AI, that is, future developments of AI that are probable but not yet within the realm of the technologically possible. The course will first focus on the question of how we can try to estimate future technological and social development and then discuss a series of particular AI applications that will likely become relevant to humanity in the future; for example: superintelligence and the singularity, alternative futures, long-term AI risks, cyborgs, cyborg identity and human enhancement, post- and transhumanism, smart cities, AI and the environment, effects of AI on poverty, human freedom, dignity and social structures, and the future of work.

AIF701 Capstone Project (6 credits)

(Prerequisite(s): AIF501 Contemporary issues in AI, AIF502 Introduction to Automated Reasoning, AIF503 Ethics of AI, and AIF504 Minds & Machines)

This capstone project enables students to deploy conceptual frameworks and research methodologies learned throughout the MAAIF, to explore in-depth relevant topics of personal interest, and to develop skills of critical research, analysis, presentation, and academic writing. The capstone research is conducted through individual or group (max 2 persons) work and direct supervision, with consultation meetings between students and advisors arranged roughly once a month. Students will work with a supervisor whose expertise includes topics relevant to their project and who will provide guidance and academic support with developing, researching, and writing the MAAIF capstone paper. Students choosing to work in pairs will be required to write a final paper of 7000-8000 words. Those working individually will be required to write a final paper of 5000-6000 words. Footnotes and quotations from primary and secondary sources in the main text are included in the word count, but appendices, tables, image captions, and the bibliography are excluded from the final word tally. The capstone project may focus primarily on themes and

concepts from within the fields of AI and its impact.

For full-time students, capstone project planning and supervision normally occurs in Term 2. Part-time students may plan and consult with supervisors in the second half of Year 1 and proceed to research and writing across Year 2.

Last Updated: 26 Sep 2024