Subject to Approval

THE UNIVERSITY OF HONG KONG

SCHOOL OF COMPUTING AND DATA SCIENCE

School Coursework Requirements for Research Postgraduate Students

For MPhil/PhD students registered from February 2026 onwards, the coursework requirements are as follows:

1. MPhil and three-year PhD students are required to take two core courses in different areas OR two core courses in Statistics area OR two core courses in the Foundations area, along with one elective course from any areas.

2. Four-year PhD students are required to complete:
   a. Three core courses, each from different areas; or
   b. Two core courses from Statistics area OR two core courses from the Foundations area, and one core course from another area

   along with three elective courses from any areas. In addition, four-year PhD students must attend at least 10 seminars, while MPhil and three-year PhD students must attend at least 5 seminars,per academic year throughout their studies. The seminars can be chosen from the SCDS Distinguished Lectures, or from the seminar series of the three departments.

Notes

- Students may opt to take additional courses beyond the required number of courses if they wish.
- MPhil students may be allowed to replace one postgraduate elective course with two courses from the School of Computing and Data Science's Master programme or one postgraduate course offered by other departments, subject to the approval by the School Higher Degrees Committee.
- PhD students may be allowed to replace up to two postgraduate elective courses with courses from the School of Computing and Data Science's Master programme and/or postgraduate courses offered by other departments, at the rate of two Master programme courses (from the School) for one postgraduate elective course. They may also be allowed to replace one postgraduate course offered by other departments for one postgraduate elective course, subject to the approval by the School Higher Degrees Committee.
- On an exceptional basis, students may be permitted to take fewer courses if they have previously completed similar courses, subject to the approval by the School Higher Degrees Committee. Any courses cannot serve as replacements for core courses.

Area	**Core Courses**
Systems	COMP9102 Data management and information retrieval
Applications	COMP9501 Advanced machine learning
Foundations	COMP9601 Theory of computation and algorithms design
Foundations	COMP9602 Optimization
Foundations	DATA8014 Principles of deep representation learning
Statistics	STAT6008 Advanced statistical inference
Statistics	STAT6009 Research methods in statistics
Statistics	STAT6010 Probability theory
Statistics	STAT6011 Computational statistics and Bayesian learning

Area	Elective Courses	
Systems	COMP8301	Advanced computing systems
Applications	COMP8317	Advanced computer vision
Applications	COMP8503	Advanced topics in visual analytics
Systems	COMP8505	Advanced topics in language models
Foundations	COMP8601	Advanced topics in theoretical computer science
Systems	COMP8803	Advanced Security and Privacy in Artificial Intelligence
Foundations	DATA8001	High-dimensional Data Analysis
Statistics	DATA8002	Statistical Inference and Machine Learning for Network Data
Foundations	DATA8003	Theoretical Foundation of Deep Learning
Foundations	DATA8004	Optimization for Statistical Learning
Applications	DATA8005	Advanced Natural Language Processing
Foundations	DATA8006	Advanced Data Mining
Foundations	DATA8007	Foundations of Sequential Decision-Making
Foundations	DATA8008	Scalable Optimization Methods in Data Science
Applications	DATA8009	Advanced Deep Learning for Computer Vision
Applications	DATA8010	Embodied AI: Perception, Representation and Action
Applications	DATA8012	AI in Medicine
Applications	DATA8013	Ethical Issues in Big Data
Foundations	DATA8015	Mathematical Foundation of Data Science
Systems	DATA8016	Quantitative Neuroscience and AI: Modeling, Inference, and Shared Principles
Applications	DATA8017	Fundamentals of Autonomous Intelligent Systems
Applications	DATA8018	Deep Generative Models
Applications	DATA8019	Computer Security
Statistics	STAT6005	Special studies in statistics (shell course)
Statistics	STAT6018	Research frontiers in data science
Statistics	STAT6025	Special studies in machine learning (shell course)
Or any courses from the core course list above		

COURSE DESCRIPTION

Core Courses

COMP9102 Data management and information retrieval

Database management and information retrieval (IR) are fundamental topics in computer science. Database systems store and manage data in structured format; all data are stored in the form of tables, where rows are entities or relationships between them and columns are attributes. The data are accessed with the help of a query language, which specifies the information to be retrieved based on the database schema. On the other hand, IR systems store data in a less structured format (e.g., text documents) and the user retrieves relevant documents to query keywords. This course will cover fundamental topics in database management, including the relational data model; query languages; storage management; indexing; query processing; query optimization; and transaction management. We will also delve into advanced topics including parallel and distributed data management; big data and key value stores; advanced indexing for complex data types. Regarding IR, the students will have the chance to learn retrieval models, web search and link analysis, representation and search of linked data.

The course includes 3 hours of lectures (by the instructor) per week. Homework includes both written exercises and programming exercises. Depending on the instructor or the need, the course can be offered with midterm and final exams or with a course project (including midterm proposal and final presentation and report). The weighting of coursework and examination is subject to approval.

Pre-requisites: Prior knowledge of data structures and algorithms at the introductory level or a computer science curriculum is necessary. Programming skills are essential for the delivery of programming exercises.

Assessment: coursework (100%)

COMP9501 Advanced machine learning

This course provides a journey through foundational concepts and cutting-edge methodologies in modern machine learning. Beginning with essential mathematical foundations in probability, statistics, and graphical models, the course progresses to core supervised learning techniques including linear/logistic regression, regularization strategies, and kernel-based methods. We will explore the theoretical underpinnings of support vector machines, neural networks, and nonparametric approaches like Gaussian processes while gaining practical insights into ensembling and mixture architectures. We will bridge classical and contemporary paradigms, covering traditional sequential models (HMMs, CRFs) alongside neural sequential architectures (LSTMs, Transformers), culminating in advanced generative frameworks such as diffusion models and flow matching. A belief introduction on probabilistic reasoning will also be covered, with modules on variational inference, graphical model analysis, and energy-based modeling.

The course includes 3 hours of lectures per week. Homework includes written assignments.

Pre-requisites: Prior knowledge of undergraduate linear algebra, multivariable calculus, and probability (MATH1853 or MATH2014), undergraduate machine learning (COMP3314), or equivalent courses.

Assessment: coursework (100%)

COMP9601 Theory of computation and algorithms design

This all-in-one introductory course presents the primary principles of theoretical computer science, focusing on algorithmic design and computational complexity. Topics include:

- Finite automata and regular expressions
- Turing machines
- Classical complexity classes (Decidability, P, NP, NP-complete problems, BPP for randomized algorithms)
- Modern complexity classes (PLS for gradient descent, PPAD for computation of equilibria, BQP for quantum)
- Polynomial-time algorithms for problems in P
- Polynomial-time approximation algorithms for NP-complete problems
- An advanced topic chosen by the instructor (online algorithms, randomized algorithms, etc.)

Pre-requisites: Pass in COMP3251 Algorithm Design or equivalent

Assessment: coursework (50%) and examination (50%)

COMP9602 Optimization

This core research course will provide you the theory, algorithms, and applications of convex optimization. Main topics to be discussed include:

On the theory of convex optimization: convex set and functions, linear programming, quadratic programming, semidefinite programming, geometric programming, integer programming, vector optimization, duality theory (dual, Lagrange multiplier, KKT conditions), etc.

On algorithms to solve convex optimization problems: gradient descent algorithm, Newton's method, interior point method, ellipsoid method, subgradient algorithm, decomposition methods, etc.

Pre-requisites: Knowledge in Linear algebra

Assessment: coursework (60%) and examination (40%)

DATA8014 Principles of deep representation learning

This course aims to provide a rigorous and systematic introduction to the mathematical and computational principles of deep learning. We achieve this by centering the course around a common and fundamental problem behind almost all modern practices of artificial intelligence and machine learning such as image recognition and generation. The problem is how to effectively and efficiently learn a low-dimensional distribution of data in a high-dimensional space and then transform the distribution to a compact and structure representation. Such a representation can be generally referred to as a memory learned from the sensed data.

Pre-requisites: Some background in undergraduate linear algebra, statistics, and probability is required. Background in signal processing, information theory, optimization, feedback control may allow you to appreciate better certain aspects of the course material, but not necessary all at once.

Assessment: Class participation (10%), Homework (50%) and Course project (40%)

STAT6008 Advanced statistical inference (6 credits)

This course covers the advanced theory of point estimation, interval estimation and hypothesis testing. Using a mathematically-oriented approach, the course provides a formal treatment of inferential problems, statistical methodologies and their underlying theory. It is suitable in particular for students intending to further their studies or to develop a career in statistical research. Contents include: (1)

Decision problem – frequentist approach: loss function; risk; decision rule; admissibility; minimaxity; unbiasedness; Bayes' rule; (2) Decision problem – Bayesian approach: prior and posterior distributions, Bayesian inference; (3) Estimation theory: exponential families; likelihood; sufficiency; minimal sufficiency; completeness; UMVU estimators; information inequality; large-sample theory of maximum likelihood estimation; (4) Hypothesis testing: uniformly most powerful (UMP) test; monotone likelihood ratio; UMP unbiased test; conditional test; large-sample theory of likelihood ratio; confidence set; (5) Nonparametric inference; bootstrap methods.

Assessment: coursework (40%) and examination (60%)

STAT6009 Research methods in statistics (6 credits)

This course introduces key statistical concepts and methods, with a focus on state-of-the-art statistical techniques and their underlying theory. Contents may be selected from: (1) Basic asymptotic methods: modes of convergence; stochastic orders; laws of large numbers; central limit theorems; delta method; (2) Parametric and nonparametric likelihood methods: high-order approximations; profile likelihood and its variants; signed likelihood ratio statistics; empirical likelihood; (3) Nonparametric statistical inference: sign and rank tests; Kolmogorov-Smirnov test; nonparametric regression; density estimation; kernel methods; (4) Computationally-intensive methods: cross-validation; bootstrap; permutation methods; (5) Robust methods: measures of robustness; M-estimator; L-estimator; R-estimator; estimating functions; (6) Other topics as determined by the instructor.

Assessment: coursework (40%) and examination (60%)

STAT6010 Probability theory (6 credits)

This course provides a mathematical introduction to probability theory and measure theory. Emphasis is put on mathematical arguments and their proofs. Contents include sigma-algebras, measurable spaces, measures and probability measures, independence and dependence, random variables, integration theory, modes of convergence, characteristic functions and conditional expectation.

Assessment: coursework (40%) and examination (60%)

STAT6011 Computational statistics and Bayesian learning (6 credits)

This course aims to give students an introduction on modern computationally intensive methods in statistics, with a strong focus on Bayesian methods. The role of computation as a fundamental tool in data analysis and statistical inference will be emphasized. The course will introduce topics including the generation of random variables, optimization techniques, and numerical integration using quadrature and Monte Carlo methods. This course will then cover the fundamental Bayesian framework, including prior elicitation, posterior inference and model selection. For posterior computation, Monte Carlo methods such as importance sampling and Markov chain Monte Carlo will be introduced. Methods for approximate inference such as variational Bayes will also be covered. Advanced Bayesian modeling with nonparametric Bayes will then be explored, with applications in machine learning. This course is particularly suitable for students who intend to pursue further studies or a career in research.

Assessment: coursework (50%) and examination (50%)

Elective Courses

COMP8301 Advanced computing systems

This course will discuss and study research topics of cutting-edge computing systems (especially, parallel and distributed systems), including three major fields: AI, security, and data science.

The central theme of this course is "the rapid growth and global distribution of data have pushed computing systems in almost all fields to become increasing parallel and distributed, in order to process the massive data and to harness the great power of computing hardware; this course will introduce cutting-edge computing systems in the perspective of parallel and distributed computing, in three major fields: AI systems, security (including Fintech and blockchain) systems, and database and data science oriented systems".

Pre-requisites: Students are expected to have completed a relevant operating system course during their undergraduate studies.

Assessment: coursework (40%) and examination (60%)

COMP8317 Advanced computer vision

This course is an advanced course in computer vision, and it covers three fundamental tasks that modern computer vision strives to achieve: image understanding, image generation, and 3D reconstruction. Hence, it consists of three parts and each part consists of a coherent set of topics and methods.

For image understanding, we will cover image classification, image segmentation and object detection. For image generation, we cover the typical image generation methods, including Variational Auto-Encoder, Generative Adversarial Networks, Diffusion Models, as well as their typical applications. For 3D reconstruction and generation part, we may cover multiple-view geometry that enables reconstructing 3D geometry from feature points, lines, and planes, as well as from regular textural patterns. We will also introduce how to represent a 3D scene, both implicit and explicit representations.

Other than introducing basic theory and methods,, this course emphasizes hands-on implementation and problem solving skills of the students. The target students are year 4 undergraduate students or year 1 graduate students.

Pre-requisites: Pass in COMP3340 Introduction to Deep Learning or COMP3317 Introduction to Computer Vision

Assessment: coursework (50%) and examination (50%)

COMP8503 Advanced topics in visual analytics

This course covers concepts, methods, and recent progress in visual analytics. It includes topics in both data analytics and information visualization, both of which are critical components of visual analytics. Topics in data analytics include density estimation, advanced clustering analysis such as density-based clustering, visualization friendly dimension reduction and manifold learning algorithms such as multidimensional scaling and t-SNE; topics in information visualization include basic charts and graphs, techniques for visualizing hierarchical structures and trees, methods for graph drawing and network visualization, methods for text and document visualization, interactive visualization techniques, etc. Students also have the chance to learn the latest progress in visual analytics by reading and presenting recent papers in information visualization and visual analytics conferences and journals.

The course includes 3 hours of lectures per week.

Pre-requisites: Prior knowledge of undergraduate linear algebra, probability, and discrete mathematics.

Assessment: coursework (100%)

COMP8505 Advanced topics in language models

This advanced course in language models provides a comprehensive exploration of the latest techniques and approaches in natural language processing (NLP) and large language models (LLMs). Structured as a seminar with class discussions, the course fosters deep learning through group dialogue and intellectual exchange. Students will delve into cutting-edge neural architectures, state-of-the-art training and inference algorithms, and the practical applications of these models. The curriculum includes hands-on experience with leading pre-trained models such as GPT-4, Claude, Gemini, LLaMA-3, and Mistral, alongside analysis of the most recent research breakthroughs and innovations in the field. Ethical considerations, bias mitigation, alignment techniques, and the integration of multimodal models like CLIP, Stable Diffusion, and Sora are also key components. By critically evaluating different language modeling approaches through collaborative discussions and developing independent research projects, students will gain the skills necessary to contribute to advanced NLP and LLM research and applications in this rapidly evolving field.

Assessment: coursework (100%)

COMP8601 Advanced topics in theoretical computer science

This course explores advanced probabilistic and randomized techniques in theoretical computer science, where randomness serves as a powerful tool for solving complex problems efficiently. For many fundamental tasks, randomized algorithms offer simpler and faster solutions than their deterministic counterparts. Students will learn key techniques for using randomness in computation, along with rigorous mathematical tools to analyze the correctness and efficiency of randomized algorithms. The course first reviews probability theory fundamentals, then introduces advanced techniques and algorithms, including Monte Carlo methods, randomized data structures, randomized graph algorithms, random walks, probabilistic methods, and approximation algorithms.

Pre-requisites: Students should be familiar with basic concepts in probability theory, linear algebra, and algorithms, including probability spaces, random variables, expectation, variance, matrices, linear systems, graph theory, etc. While a strong background in these areas will help students engage more deeply with the material, comprehensive mastery is not required upfront. The course will review key foundational concepts as needed.

Assessment: coursework (100%)

COMP8803 Advanced Security and Privacy in Artificial Intelligence

This advanced course offers a deep dive into the latest research and practical challenges in securing AI systems and preserving the privacy of data and models involved in their training and deployment. It explores how machine learning (ML) models can be compromised and how such threats can be detected, mitigated, or prevented. We begin with model robustness, highlighting the fragility of ML models and discussing both formal and empirical approaches to enhance their resilience. We then examine privacy threats, including membership inference, model inversion, and model extraction, followed by privacy-enhancing technologies such as differential privacy, secure computation, and federated learning. We also address emerging issues in copyright and ownership, including model watermarking, data misuse prevention, and machine unlearning. Finally, the course delves into the growing risks of generative AI systems, such as jailbreak attacks, safety alignment, and broader governance challenges.

Through five thematic modules, students will engage with foundational and emerging research, assignments, and a research project, developing the technical depth and critical perspective needed to contribute to cutting-edge AI security and privacy research.

Pre-requisites:	Students need to have completed COMP3523, or provide proof of equivalent knowledge.
Assessment:	To be determined

DATA8001 High-dimensional Data Analysis

This graduate course introduces basic geometric and statistical concepts and principles of low-dimensional models for high-dimensional signal and data analysis, spanning basic theory, efficient algorithms, and diverse applications. We will discuss recovery theory, based on high-dimensional geometry and non-asymptotic statistics, for sparse, low-rank, and low-dimensional models – including compressed sensing theory, matrix completion, robust principal component analysis, and dictionary learning etc. We will introduce principled methods for developing efficient optimization algorithms for recovering low-dimensional structures, with an emphasis on scalable and efficient first-order methods, for solving the associated convex and nonconvex problems. We will illustrate the theory and algorithms with numerous application examples, drawn from computer vision, im- age processing, audio processing, communications, scientific imaging, bioinformatics, information retrieval etc. The course will provide ample mathematical and programming exercises with sup- porting algorithms, codes, and data. A final course project will give students additional hands-on experience with an application area of their choosing. Throughout the course, we will discuss strong conceptual, algorithmic, and theoretical connections between low-dimensional models with other popular data-driven methods such as deep neural networks (DNNs), providing new perspectives to understand deep learning.

The course includes 3 hours of lectures (by the instructor) and 1 hour discussion session (by a GSI) per week. Homework includes both written exercises and programming exercises. A final course project includes a midterm proposal and final presentation and report.

Pre-requisites:	We require students to have prior knowledge in undergraduate linear algebra, statistics, and probability. Background in signal processing, optimization, machine learning, and computer vision may allow you to appreciate better certain aspects of the course material, but not necessary all at once.
Assessment:	class participation (10%), homework (50%) and a course project (40%)

DATA8002 Statistical Inference and Machine Learning for Network Data

This course provides a concise and practical introduction to statistical inference and machine learning methods for analyzing network data, centered around three essential topics: community detection, network reconstruction, and graph representation learning. Within each of these topics, we will introduce the basic concepts then gradually move towards state-of-the-art techniques, along the way introducing key methods from statistical inference and machine learning that can be applied in more general contexts. We will illustrate the theory and algorithms with application examples from across the social and natural sciences. Students will gain hands-on experience through homework consisting of mathematical and/or programming exercises, as well as a final course project in the form of a research paper and presentation.

Pre-requisites: A solid foundation in linear algebra, probability, and statistics will be important for understanding the course material. Such materials should at least be mastered at the level of MATH1853, but preferably at the levels of MATH2101/2102 for linear algebra and MATH3603 or STAT2601/2602 or STAT3902 for probability and statistics at HKU, or equivalent courses at other universities.

Some exposure to Bayesian inference, data structures and algorithms, machine learning, and optimization is also recommended but not strictly necessary.

If you're curious about whether you would benefit from this course or your academic background is appropriate, contact the instructor for details.

Assessment: homework (40%), final Project (50%), attendance/participation (10%)

DATA8003 Theoretical Foundation of Deep Learning

Deep learning has achieved great success in many real-world applications. However, the reason why deep learning is so powerful remains elusive. The goal of this course is to introduce theoretical tools and methods that are developed to understand and explain the success of deep learning. In particular, this course will cover multiple aspects of machine learning, including landscape analysis, optimization, generalization, and algorithm designs. We will start with the introduction of the basic setup of machine learning problems, including loss function, training algorithms, and generalization performance evaluation.

Then we will further introduce the conventional optimization theory and statistical learning theory, and discuss its limitation in studying over-parameterized deep neural network models. We will also introduce the neural tangent kernel (NTK) theory, a modern theoretical method can handle over-parameterization and nonconvex issues in deep learning. Finally, we will discuss representation learning and benign-overfitting of over-parameterized learning models, and their connections to the optimization and generalization in deep learning. The instructor will give lectures on the selected topics. Students will need to complete the homework (including programming and mathematical derivations) and a course project.

This course will be similar to STATS214 / CS229M: Machine Learning Theory in Stanford, and CS269: Foundations of Deep Learning in UCLA.

Pre-requisites: We require students to have prior knowledge in undergraduate linear algebra, statistics, probability, and calculus. Background in optimization and machine learning is not required but preferred.

Assessment:	attendance (10%), paper presentation (20%), homework (20%) and final project (50%)

DATA8004 Optimization for Statistical Learning

Optimization has long been playing an important, indispensable role in statistics and machine learning. This has become even more so in recent years, since the current era of big data presents many theoretical and computational challenges to the statistics and machine learning communities. This course aims at introducing a series of core topics lying at the intersection of modern optimization and statistics. We will particularly focus on the interplay between sample and computational complexity.

Pre-requisites:	The prerequisites to this course are undergraduate linear algebra, calculus, real analysis, probability theory and convex optimization.
Assessment:	homework (30%), exam (40%) and group project (30%)

DATA8005 Advanced Natural Language Processing

Natural language processing (NLP) is the study of human language from a computational perspective. This course is an introductory graduate-level course on natural language processing aimed at students who are interested in doing cutting-edge research in the field. In this class, we will cover recent developments on core techniques and modern advances in NLP, especially large language models and language grounding. We will also survey some recent NLP research topics including multimodality, interactivity, and interoperability for NLP.

Students will gain the necessary skills and experience to understand, design, implement, and test large language models through a final project. We will also introduce cutting-edge research topics and learn how to conduct NLP research through paper readings and discussions. We will potentially also host invited speakers for talks.

Pre-requisites:	We require students to have prior knowledge undergraduate linear algebra, probability and statistics, machine learning, or deep learning. Introduction to natural language processing is recommended.
Assessment:	homework (60%) and attendance (40%)

DATA8006 Advanced Data Mining

This advanced data mining course establishes a comprehensive foundation in Data-centric AGI (Artificial General Intelligence) while emphasizing practical applications. The course systematically covers core foundational principles and their real-world implementations across key domains. Foundation Knowledge: The curriculum builds essential expertise in: i) Foundation Knowledge of Large Language Models (architecture, training, and capabilities); ii) Retrieval-Augmented Generation (RAG) principles and mechanisms; iii) Autonomous Agents design and decision-making frameworks; iv) Graph Machine Learning fundamentals; v) Recommender Systems and Information Retrieval; vi) Multi-Modal Understanding across different modalities. By completing this course, students will gain both deep foundational knowledge in data-centric AGI and hands-on experience in building practical application systems.

The course consists of 3 hours of weekly lectures delivered by the instructor. Homework assignments include programming exercises and student presentations. Assessment methods are flexible and may include either midterm and final examinations or a comprehensive course project featuring a midterm proposal, final presentation, and written report. The specific weighting of coursework components and examinations is determined based on instructor preference and pedagogical needs, subject to departmental approval.

Pre-requisites:	Prior knowledge of programming and machine learning will facilitate your understanding of the data-centric AGI development process, which includes data collection, model training, system integration, and performance evaluation. This advanced course is intended for graduate students seeking to master both foundational knowledge in data-centric artificial intelligence and practical skills in building real-world application systems across various AI domains.
Assessment:	participation (10%), presentation (40%) and project (50%)

DATA8007 Foundations of Sequential Decision-Making

The digital world has a wealth of data, such as internet of things (IoT) data, business data, health data, mobile data, urban data, security data, and many more, in the era of "Big Data" and the Fourth Industrial Revolution (Industry 4.0 or 4IR). Extracting knowledge or useful insights from these data can be used for intelligent and automated decision-making in various applications domains. Advanced analytics methods including machine learning modeling can provide actionable insights or deeper knowledge about data, which makes the computing process automatic and smart. Data- driven discovery is revolutionizing the modeling, prediction, and control of dynamic complex systems.

This graduate course introduces the foundations of sequential decision-making models and algorithms (Markov decision processes, dynamic programming, Q-learning, TD learning, SARSA, actor-critic, policy gradient and bandits). We will illustrate the theory and algorithms via numerous application examples, drawn from the areas of finance, logistics, supply chain management, pricing and revenue management, and robotics etc. By the end of the course, the student will build solid understanding to conduct research on sequential decision-making problems. They also should be able to apply the theories and analysis skills in modelling dynamic engineering problems and designing algorithms to solving sequential decision-making problems in manufacturing and service applications.

The course includes 3 hours of lectures per week. Homework includes both written exercises and programming exercises. Depending on the instructor or the need, the course can be offered with a final quiz and a course project (including presentation and report).

Pre-requisites: Real Analysis, Linear Algebra, Operational Research, Statics and Probability, Optimization (Linear and Convex). In general, the course will be very much self-contained.

Assessment: homework (50%) and final project (50%)

DATA8008 Scalable Optimization Methods in Data Science

This course introduces students to fast and scalable optimization algorithms that play important roles in data science. Students will not only learn about implementation of the algorithms but also understand why they are correct and fast (theoretical guarantees). Both mathematical analysis skills and programming skills of the student will be trained.

Pre-requisites: Linear Algebra, calculus, Statistics and Probability. In general, the course will be very much self-contained.

Assessment: attendence (20%), homework/class tests (50%) and final project (30%)

DATA8009 Advanced Deep Learning for Computer Vision

This graduate course introduces basic and advanced topics in deep learning and its application in computer vision. We will discuss deep learning architectures including CNN, RNN, LSTM, and transformers, optimization and regularization of neural networks, and discriminative and generative models for solving different computer vision tasks, including visual perception for image, video, and 3D data, visual content generation, and recent advances in the field.

The course includes 3 hours of lectures per week. Assignments include both written exercises and programming exercises. A course project with a final presentation and a report is required for each student.

Pre-requisites: We require students to have prior knowledge in linear algebra, statistics, and programming.

Assessment: participation (10%), project (40%) and presentation (50%)

DATA8010 Embodied AI: Perception, Representation and Action

This course will explore various topics in Embodied Artificial Intelligence (AI), which is concerned with the perception and representation of the physical world by autonomous agents and their consequent physical interactions. Specifically, the course will cover how an agent can infer the physical and semantic states of the scene, e.g., via 3D reconstruction and semantic parsing; how these perceptions can be represented in a way that facilitates reasoning with both efficiency and explainability; and how actions can be made in order to achieve downstream interaction tasks. By the end of this course, students will have gained an understanding of the fundamentals of perception, representation, and control in AI systems, as well as learned how embodied agents can be used to interact with and manipulate the physical world.

This course is designed for students with electrical engineering or computer science backgrounds who aim to perform research in the intersection of computer vision, machine learning, and robotics.

Pre-requisites:	Basic knowledge of linear algebra, probability and random processes, optimization, signal processing, and machine learning will ease the digestion of the course materials.
Assessment:	attendance (15%), course presentation (30%), project proposal (15%), project presentation and report (40%)

DATA8012 AI in Medicine

In this course, students will learn a variety of AI and other computational approaches to address important problems in medicine, including disease diagnosis, risk assessment, patient stratification, infectious disease monitoring, mental health promotion, etc. In addition to the introduction to fundamental methods and classic case studies, students will also learn state-of-the-art by reading latest publications (mostly in the past 1-2 years). Students are required to form groups to work on a project with open source data.

Pre-requisites:	Basic knowledge in college-level mathematics and statistics will be required.
Assessment:	class participation (10%), homework (40%) and a course project (50%)

DATA8013 Ethical Issues in Big Data

This course will be an introduction to ethical dimensions arising in the practice of statistics and data science, including in particular moral puzzles, problems and dilemmas that arise in the application of machine learning and artificial intelligence to everyday decision making in politics, business, and ordinary life. Throughout the course we will gradually develop an interdisciplinary systematic framework for thinking about such problems which draws on multiple philosophical traditions, reasoning tools, and statistical methods. We will also draw on legal theory and case law.

Pre-requisites:	Some background in philosophy and statistics is recommended.
Assessment:	final project/paper (70%), attendance and class presentation (30%)

DATA8015 Mathematical Foundation of Data Science

This course aims to equip the students with fundamental mathematical tools frequently used in data science and prepare them for more advanced study in various directions of data science. It is designed for students who have basic knowledge of linear algebra, calculus, and probability (e.g., undergraduate courses in these topics) and would like later to pursue an in-depth investigation of data science.

For this goal, the course will cover fundamental mathematical tools for data science and focus particularly on their connections to data science. Topics include basic concepts in probability/statistics, linear algebra, and optimization, such as MLE/MAP, elementary information theory, concentration, spectral decomposition, convexity, gradient descent, etc. These are illustrated with the corresponding examples in data science, such as statistical learning, PCA, SGD on empirical loss, etc.

Pre-requisites: Basic knowledge of linear algebra, calculus, and probability (e.g., undergraduate courses in these topics) is recommended. The course is open to beginning graduates with the required math background.

Assessment: homework (50%) and final exam (50%)

DATA8016 Quantitative Neuroscience and AI: Modeling, Inference, and Shared Principles

This doctoral level class examines the interdisciplinary nexus of neuroscience and artificial intelligence (NeuroAI). The course provides doctoral students with core theoretical frameworks and methodological skills essential for conducting original research in this field. It is designed for students specializing in either machine learning or neuroscience, establishing a foundational basis for NeuroAI research. The curriculum critically evaluates two interconnected themes: the application of contemporary artificial intelligence architectures as models of neural function, and the potential utility of neuroscientific principles for advancing artificial intelligence.

The course integrates two primary components: (1) Critical Analysis: Examination of current methodologies and theoretical debates in computational neuroscience through lectures, assigned readings, and structured discussion. (2) Empirical Research: A substantial, project-based component involving computational assignments and an independent research project. This project is designed to enable students to formulate empirically grounded perspectives and has the potential to yield publishable results. Course content focuses on the comparative analysis of representational strategies in biological neural systems (across species) and artificial intelligence models.

Core domains include: sensory processing, motor control, language, and higher cognition. Students will learn and critically assess state-of-the-art techniques for employing AI models as hypotheses of brain function, including their strengths and limitations. The seminar will also address generative mechanisms in both biological and artificial systems and explore the potential for their integration.

Pre-requisites: Proficiency in intermediate statistics or machine learning is required. Students must possess working knowledge of deep learning fundamentals, linear algebra, and Python programming. Experience with array-based computation libraries (e.g., PyTorch, NumPy, TensorFlow) is essential.

Prior coursework in neuroscience is advantageous but not mandatory. Students without neuroscience background will receive introductory materials at the semester outset. Prior coursework in neuroscience is advantageous but not mandatory. Students without neuroscience background will receive introductory materials at the semester outset. Students with prior trainings or experiences should consult the instructor before registration.

The course includes structured support for scholars transitioning from non-neuroscience disciplines. Prospective students uncertain about prerequisite fulfillment should reach out to the instructors!

Assessment: reading and in-class quiz (15%), participation (15%), homework (40%) and final project (30%)

DATA8017 Fundamentals of Autonomous Intelligent Systems

This course aims at introducing the fundamentals in algorithms, data and systems of the autonomous intelligent systems, which often refers to the autonomous driving and robotics applications. As the fast advances in the field of AI, how to utilize the learning-based, data-driven approaches to improve the applications for the better human life, becomes very pivotal. We will address the key challenges in this domain, such as (i) how to formulate a system that is equipped with generalization, intelligence and reliability merits. (ii) How to balance the data distribution between simulation and real-world data. (iii) Is scaling law the only pathway towards high-level AGI.

We will introduce the concepts, principles and knowhow to build the autonomous intelligent systems. The basic fundamentals would be detailed in the lectures, with tutorials and hands-on training sessions. All the important topics will be covered, such as imitation learning, reinforcement learning, and so on, with a focus on the applications in autonomous driving and robotics. The highlights in this course would consist of several guest lectures from outside renowned speakers from both industry and academiato address the latest advances in this field. The hands-on session is akin to tutorials or hackathons where students learn the recipe of technologies from scratch quickly. These features would be complementary to the main lecture and facilitate the final group presentation.

The course includes 2 hours of lectures (by the instructor or guest lecture) and 1 hour discussion/tutorial/hands-on session (by a TA) per week. Homework includes both written exercises and programming.

Pre-requisites: The basic programming skill is needed, e.g. python and C++.

We require students to have prior knowledge in undergraduate linear algebra, statistics, and probability. Background in machine learning, and computer vision may allow you to better appreciate certain aspects of the course material, but not necessarily all at once.

The course is open to research postgraduates (i.e. MPhil, PhD). Senior undergraduates (Year 2 to 4) are welcome, with consent from the instructor. If you’re curious about whether you would benefit from this course, contact the instructor for details.

Assessment: paper reading presentation(30%), mid-term proposal(20%), final-project presentation & report (50%)

DATA8018 Deep Generative Models

Generative AI is now at the center of Artificial Intelligence in both academia and industry, with both significant potential and complex challenges. Such a transformative moment is brought by the remarkable advances in parameterizing generative models with deep neural networks. In this graduate course, students will systematically go through deep generative models in detail, equipping themselves with comprehensive knowledge and skills to address various tasks with deep generative models or understand in principle the issues of them. At the core of the curriculum is various representative deep generative models that are commonly used in research and practice, such as diffusion models, generative adversarial networks, variational autoencoders and autoregressive models.

In this course we will cover their theoretical principles, implementation details as well as relative pros and cons. In this way students are empowered with a clear understanding of these models and a practical guidance of using them to develop task-specific innovative solutions. On top of a principled journey into the secrets behind generative ai, this course will also demonstrate the frontier of practical application of deep generative models, with case studies in language modeling, art and embodied ai. This enhances their ability to catch up with current circumstances and opportunities ahead.

Finally, this course will also place a strong emphasis on the responsible use of deep generative models, by discussing how to evaluate these models, detect their output, and handle their ethical issues. To improve their learning efficiency, students will conduct example hands-on training of various deep generative models utilizing open-source datasets, and at the same time participate in literature survey and presentation of latest academic and industrial advances. In this way their problem-solving abilities will be strengthened, and their vision will be broadened. Upon completion of this course, students are expected to build a scientific and structured understanding of generative ai, and master necessary skills to make their own contributions for relevant academia and industry.

Pre-requisites: Basic knowledge in college-level coding and mathematic. The course is open to undergraduate students, subject to the course leader's approval.

Assessment: attendance (10%), course presentation (40%) and project (50%)

DATA8019 Computer Security

This course introduces advanced topics in modern computer security, such as software security, systems security, web security, mobile security, applied cryptography and cryptographic protocols, privacy enhancing technologies, usable security, machine learning security, blockchains, and cybercrimes. Research papers are the main instructional materials. The students read research papers before lectures and then discuss their merits and shortcoming during lectures. This course trains the students to do research in groups, to present their research results, and to write research papers. It prepares the students to pursue a career in security research or to apply security technologies in their professions.

Pre-requisites: College-level programming skills. Familiarity with operating systems, computer networking, and machine learning is recommended but not required. The course is open to undergraduate students subject to the instructor's approval.

Assessment: research paper reading, course project presentation, project report

---

STAT6005 Special studies in statistics

The aim of the course is to introduce students to the statistical topics which are of relevance to their research study but have not been taken previously. Students will be instructed to attend one course or a combination of courses from the department as prescribed by the supervisor(s) and approved by the Chairman of the Departmental Research Postgraduate Committee. Alternately this course may consist of supervised reading supplemented by written work and prescribed coursework.

Students are permitted to replace this course by another RPG course from the MPhil/PhD curricula offered by other Departments, subject to the approval of the Departmental Research Postgraduate Committee.

Note: Students should not be taking or have taken STAT6025 Special studies in machine learning

Assessment: To be determined

STAT6018 Research frontiers in data science

This course aims to equip postgraduate students with the latest knowledge and practical skills in data science and relevant domains, in order to enhance their research capabilities. The course comprises of different modules and guest lectures, with topics encompassing: 1) big data analytics; 2) machine learning; 3) image processing and computer vision; 4) high-dimensional data analysis; 5) statistical methods and their applications in medical research; 6) time series econometrics; 7) data analytics in actuarial science; and 8) other areas as determined by the instructor.

Assessment: 100% coursework

STAT6025 Special studies in machine learning

The aim of the course is to introduce students to the machine learning topics which are of relevance to their research study but have not been taken previously. Students will be instructed to attend one course or a combination of courses from the department as prescribed by the supervisor(s) and approved by the Chairman of the Departmental Research Postgraduate Committee. Alternately this course may consist of supervised reading supplemented by written work and prescribed coursework.

Students are permitted to replace this course by another RPG course from the MPhil/PhD curricula offered by other Departments, subject to the approval of the Departmental Research Postgraduate Committee.

Note: Students should not be taking or have taken STAT6005 Special studies in statistics

Assessment: To be determined