THE UNIVERSITY OF HONG KONG

SCHOOL OF COMPUTING AND DATA SCIENCE

School Coursework Requirements for Research Postgraduate Students[1] [2]

For MPhil/PhD students registered from February 2025 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 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.

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

[1] The syllabus is currently pending approval from the university and may be subject to change.
[2] As a transitional arrangement and on an exceptional basis, RPg students who registered in the academic year 2025-2026 may apply to adopt the course requirements valid in the year 2024-2025, subject to approval by the School Higher Degrees Committee. The requirement to attend at least 10 seminars per academic year, however, applies without exceptions to all four-year PhD students.

Statistics	STAT6010 Advanced probability
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
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: TBC

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 Advanced probability (6 credits)

This course provides a mathematical introduction to measure theory and probability theory to equip students with the necessary mathematical concepts important for research in probability, statistics and actuarial science. Contents include sigma-algebras, measurable spaces, measures and probability measures, independence and dependence, random variables, integration theory, modes of convergence, characteristic functions, conditional expectations and more.

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%)

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