CUHK Business School

Course Syllabus

DOTE6635: Artificial Intelligence for Business Research

Spring 2025

INSTRUCTOR: Professor Renyu (Philip) Zhang

E-mail: philipzhang@cuhk.edu.hk

Telephone: 852-3943-7763

Office Hours: By appointment

Office Location: Cheng Yu Tung Building 911

COURSE MEETINGS

Meeting Times: Tuesday, 12:30PM-3:15PM

Class Will NOT Meet on: January 7 (Guest lecture by Prof. Dennis Zhang on Jan 2), January 28 (Chinese New Year), March 4 (Final Project Preview)

Class Will Meet on: All other Tuesdays from January 14 to April 15

Location: Wu Ho Man Yuen Building (WMY) **504**

Course Website: https://github.com/rphilipzhang/AI-PhD-S25

Lecture Format: In person

Zoom Room: 949 8812 1019, Pass code: 386119

Lecture Videos: The lectures will be recorded and available with my approval.

TEACHING ASSISTANT: Leo Cao[1]

Email: yinglyucao@cuhk.edu.hk

COURSE DESCRIPTION

Modern machine-learning(ML)-based artificial intelligence (AI) has largely reshaped our world

[1] Leo will help with any issues related to the logistics, but not the content, of this course.

over the past 10 years at large and the last 2 years (since 2022) in particular. This Ph.D. course is designed to prepare students from a broad range of business areas (e.g., Finance, Marketing, Accounting, Management, Operations, Information Systems, Business Economics, Hotel Management etc.) for the cutting-edge research that connects AI and business. At a high-level, students taking this course will learn the following:

- Fundamental concepts/methods of machine learning (ML) and AI that are used in business research.
- How business researchers have utilized ML/AI and what managerial questions have been addressed by ML/AI in the recent decade.
- A basic taste of what the state-of-the-art AI/ML technologies can do in the ML/AI community and, potentially, in your own research field.

We substantially iterate the content of this course every year, so you are recommended to sit in this course even if you have already taken it for credit before.

PREREQUISITES

The target audiences of this course are MPhil and PhD students interested in applying AI/ML to their own field of study in business. The students are expected to master the following before the beginning of this course:

- Working knowledge in calculus, linear algebra, and statistics;
- Working knowledge with Python (or my approval if the student has prior programming experience with a language other than Python, e.g., R, STATA, or MatLab), co-piloted by AI tools such as ChatGPT, Claude, Gemini, Cursor, Windsurf, Devin, etc.;
- Basic knowledge of machine learning;
- Basic knowledge of econometrics and causal inference.

References and tutorials for the required prerequisites will be provided along the progress of this course. Please talk to me if you are unsure whether this course is suitable for you.

Learning Outcomes

The students taking this course will receive multi-faceted training in AI/ML and t (growing) applications in business research. At the end of this course, you are expected to learn the following:

- Fundamental ideas and tools in AI/ML (and their limitations) that are widely used in their research fields;
- The state-of-the-art literature landscape and research norm for the interface between AI/ML and business;
- The ability to apply AI/ML techniques in your own research;

- The ability to identify new research questions in your fields that are potentially solvable by AI techniques.

The course will try to balance between the hands-on implementation of AI/ML methods and the conceptual understanding of business research frontier where AI/ML plays a critical role.

Recommended Reading: General References

The following general references should be useful regardless of what field you are in and what research topics you are working on.

- *The Elements of Statistical Learning* (2nd Edition), 2009, by Trevor Hastie, Robert Tibshirani, Jerome Friedman, https://hastie.su.domains/ElemStatLearn/.
- *Deep Learning*, 2016, by Ian Goodfellow, Yoshua Bengio and Aaron Courville, https://www.deeplearningbook.org/.
- *Dive into Deep Learning* (2nd Edition), 2023, by Aston Zhang, Zack Lipton, Mu Li, and Alex J. Smola, https://d2l.ai/.
- *Deep Learning with PyTorch*, 2020, by Eli Stevens, Luca Antiga, and Thomas Viehmann.
- *Probabilistic Machine Learning: An Introduction*, 2022, by Kevin Murphy, https://probml.github.io/pml-book/book1.html.
- *Probabilistic Machine Learning: Advanced Topics*, 2023, by Kevin Murphy, https://probml.github.io/pml-book/book2.html.
- Mullainathan, Sendhil, and Jann Spiess. 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31(2): 87-106.
- Athey, Susan, and Guido W. Imbens. 2019. *Machine learning methods that economists should know about. Annual Review of Economics* 11: 685-725.

COURSE WEBSITE/MATERIALS

Course materials, including slides, lecture notes, codes, and optional readings, will be distributed electronically through the GitHub course website.

The link to the GitHub of this course:

- https://github.com/rphilipzhang/AI-PhD-S25

The GitHub repository is public, so feel free to check it out often and let me know if you have any questions. I recommend you folk the repo to your own GitHub and update your folked repo regularly and frequently.

COURSEWORK and GRADING

You will form groups that do not change throughout the semester and have a maximum size of two students. Please register your group via the designated 11:59pm, Jan 15, 2025 (Wednesday). Let me know if you need any help find a group mate. You will work with your group mate on the following:

- **Lecture Notes Scribing** (you will scribe the lecture notes)
- **Paper Replication and Presentation** (we will have about one paper presentation each week; importantly, you need to replicate the study and share the details with your classmates)
- **Homework Problem Sets** (we will have about one coding assignment almost every week)
- **Final Project** (we will have a final project based on your choice)

At the end of this course, you will be asked to mutually evaluate your group mate's contribution towards Lecture Notes Scribing, Paper Replication and Presentation, Problem Sets, and Final Project.

Grading: We follow the grading practices of the CUHK Business School. The following grades may be awarded: A, A-, B+, B, B-, C+, C, C-, D+, D, F. In general, A indicates excellent work, B indicates good work, C indicates satisfactory work, and D indicates passable work and is the lowest passing grade. F indicates failure.

- **Course Participation** (X% bonus)
- **Lecture Notes Scribing** (10%).
- **Paper Replication and Presentation** (30%)
- **Problem Sets** (30%)
- **Project** (30%)

Class Participation

You are expected to actively participate in the discussions of the lecture content, homework problem sets, research paper presentations, and each others' final projects. I expect you to arrive to class on-time and be prepared, and to stay involved during class sessions. You will receive extra credits for asking good questions or providing insightful comments either during the class or in the WeChat Group.

Lecture Notes Scribing

Each group will be responsible for scribing the lecture notes of one topic in LaTex. I will provide the sample lecture notes. You will use it as the template for your scribing task. The completeness, correctness and clarity of your scribed lecture notes will determine this part of your grade. Please sign-up for the lecture notes scribing at the designated Google Sheet by 11:59pm, Jan 15, 2025 (Wednesday).

Paper Replication and Presentation

You will work on a replication project of your own choice from the list of papers I provide. These papers all have their replication packages available online. Please talk to me if you are interested in replicating another paper related to our course but outside the list. The papers are mostly applications of AI/ML within a business/economics context. Your job is to fully understand the paper and try to replicate its results to the greatest extent possible. Extra credits will be given if you can offer additional significant insights that are not reported in the original paper. Please sign up for the replication project and presentation schedule at the designated Google Sheet by 11:59pm, Jan 15, 2025 (Wednesday). The deadline of your replication project is 12:30pm, on the day of your presentation. You should submit (a) your CoLab link of your replication project with all the necessary code and results; (b) a succinct report (no more than 5 pages excluding references and appendices) articulating what you have done and what you have found, with the focus on the comparison between your results and the original paper; and (c) your presentation slides. The replication project counts 30% towards your final grade.

Problem Sets

There will be (roughly) 1 coding problem set distributed for each class. All problem sets will be posted on our course GitHub, and will be due TWO weeks after the distribution date. You should work with your group mate on each problem set, and submit your CoLab link on the designated Google Sheet. In principle, you can read each other's code, but you should submit the code written by yourselves. The 5 problem sets with the highest grades will be counted, 6% of the final grade each. However, you are encouraged to finish all the problem sets as they provide necessary hands-on exercises for you to master the related AI/ML techniques. We will grade the problem sets very generously using the following rubric:

- 6 means you have made sincere efforts to solve this problem set;
- 3 means you have made non-negligible efforts to solve this problem set;
- 0 means you have made negligible efforts to solve this problem set OR you do not submit anything.

Late assignments will not be accepted. You will receive the same grade as your group mate for each problem set.

Final Project

Each group will work on a research project of your own choice. The topic of your project should be related to the application of AI/ML to a business problem, broadly defined. Ideally, the project should be an original research targeting a top journal of your chosen field, though an extension or a thorough replication study of an existing paper(s) will also be acceptable. Literature review is generally not encouraged for the final project.

Here's the (tentative) timeline of your final project:

- Session 1: Start thinking about how to generate the idea and data for your final project.
- Week of March 4: Discuss your final project with me.
- Session 8: Submit a 1-page proposal of your project, which summarizes your research problem, the data you plan to collect (or have collected), and the AI/ML method you plan to use.
- Session 13: Present your project to your classmates and me. We may skip project presentation if the class time is not enough.
- Session 13+: Submit your project report.

I do not expect your final project to generate publishable results, but it should be a rigorous scientific study. Therefore, the projects will be evaluated based on

- whether your research question(s) is interesting, important, and related to AI/ML;
- whether the data you collect can adequately address your research question(s);
- whether your analysis and methodologies are rigorous;
- whether your results are reasonable and well-interpreted;
- whether you deliver a clear and engaging project presentation;
- whether your report is well written.

Please keep your report succinct and leave the secondary analysis details to the Appendix. I will look at the contribution per page for the report. So please do not write a 50-page paper for a bunch of trivial results. However, if your results are really exciting, I do not set the page limit. You will receive the same grade as your group mate for the final project.

CLASS WORK

Building AI/ML models and implementing them in a programming language is an indispensable part of learning in this course. However, I may not be able to go over the code in very detail due to the limited class time. I suggest you take your laptop to each session and open the code I distribute to you in CoLab, so as to have a better conceptual understanding of the AI/ML methods.

Classroom Norms: Cell phones and other electronic devices are a disturbance to both students and to me. All electronic devices (except laptops) must be turned off prior to the start of each class meeting.

Laptops: You are expected to bring a laptop to each class and read the related code, unless otherwise instructed. Please use it for class activities only.

Students with Disabilities: Please refer to the Support Services for Students with Disabilities (https://www2.osa.cuhk.edu.hk/disability/en-GB/).

WECHAT GROUP

We will establish a WeChat group as the off-class online discussion platform for this course. All students are required to enter this group, and are encouraged to post and discuss any questions, suggestions, and/or comments about this course in the class WeChat group. Students who actively contribute to the discussions in our WeChat group may receive some extra credits in the final course grade.

ACADEMIC INTEGRITY

Integrity is critical to the learning process and to all that we do here at the CUHK Business School. As members of our community, all students agree to abide by the Academic Honesty policies of CUHK (see https://www.cuhk.edu.hk/policy/academichonesty/ for details), which includes a commitment to:

- Exercise integrity in all aspects of one's academic work including, but not limited to, the preparation and completion of exams, papers and all other course requirements by not engaging in any method or means that provides an unfair advantage.
- Clearly acknowledge the work and efforts of others when submitting written work as one's own. Ideas, data, direct quotations (which should be designated with quotation marks), paraphrasing, creative expression, or any other incorporation of the work of others should be fully referenced.
- Refrain from behaving in ways that knowingly support, assist, or in any way attempt to enable another person to engage in any violation of the Academic Honesty policies of CUHK. Our support also includes reporting any observed violations that are deemed to adversely affect the CUHK community.
- You may not submit the same work (or substantially similar work) to meet the requirements of more than one course without written consent of all instructors concerned.

COURSE EVALUATIONS

Course evaluations are important to us and to students who come after you. Please complete them thoughtfully.

COURSE OUTLINE

The course schedule below is tentative and subject to changes.

Session	Date	TOPIC
0	January 2	Guest lecture by Prof. Dennis Zhang
1	January 14	Course Introduction
2	January 21	Deep Learning Introduction
3	February 4	LLM (I): Transformer
4	February 11	LLM (II): Pre-training
5	February 18	LLM (III): Post-training
6	February 25	LLM (IV): Agentic AI
7	March 4	No Class; Final Project Discussion
8	March 11	Causal Inference (I): Introduction
9	March 18	Causal Inference (II): Causal Trees and Forests
10	March 25	Causal Inference (III): Double Machine Learning
11	April 1	Causal Inference (IV): RL and Off-Policy Evaluation
12	April 8	Economics and Ethics of AI
13	April 15	Final Project Presentation