CUHK Business School

Course Syllabus

DSME6635: Artificial Intelligence for Business Research

Spring 2024

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: February 13 (Chinese New Year), March 5 (Final Project Preview)

Class Will Meet on: All other Tuesdays from January 9 to April 16

Location: Cheng Yu-Tung (CYT) **LT5**

Course Website: https://github.com/rphilipzhang/AI-PhD-S24

Lecture Format: In person

Zoom Room: 996 4239 3764, Pass code: 386119

Lecture Videos: The lectures will be recorded and available on GitHub.

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 over the past 10 years. This Ph.D. course is designed to prepare students from a broad range of busi-

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

ness 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.

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);
- 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 its application 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-S24

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 people. Please submit the group to me by email before 11:59pm, Jan 10, 2023 (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** (we will ask the students to scribe the lecture notes)

- **Paper Presentation** (we will have about one paper presentation each week)
- **Homework Problem Sets** (we will have 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 Presentation, Problem Sets, and Final Project.

Grading: We follow the same grading practices as 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 Presentation** (10%)
- **Problem Sets** (40%)
- **Project** (40%)

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 for Session 1, *Introduction to Machine Learning*. 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 10, 2023 (Wednesday).

Paper Presentation

You will need to present the assigned paper with your group mate. We will have one or two presentations at the end of each class related to the topic of my lecture. Please sign up for the presentation schedule at the designated Google Sheet by 11:59pm, Jan 10, 2023 (Wednesday). All students are expected to have read the paper before the presentation and actively contribute to the discussions on it. The presenters are expected to discuss both the paper itself and, briefly, the relevant literature where it is positioned. Your presentations will be graded based on how well you

understand the paper (and the related literature), as well as how clear you introduce the paper to your classmates.

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, 8% 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:

- 8 means you have made sincere efforts to solve this problem set;
- 4 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 timeline of your final project:

- Session 1: Start thinking about how to generate the idea and data for your final project.
- Week of March 5: 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.
- 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.

Detailed Course Schedule

The course schedule below is tentative and subject to minor changes.

Session 1: Artificial Intelligence and Machine Learning in a Nutshell (Jan/09/2024)

Course

- Introduction to Each Other
- Course Introduction
- Introduction to Machine Learning and Artificial Intelligence

Recommended Reading

- *The Elements of Statistical Learning* (2nd Edition), 2009, by Trevor Hastie, Robert Tibshirani, Jerome Friedman, https://hastie.su.domains/ElemStatLearn/.
- *Probabilistic Machine Learning: An Introduction*, 2022, by Kevin Murphy, https://probml.github.io/pml-book/book1.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.
- Hofman, Jake M., et al. 2021. Integrating explanation and prediction in computational social science. *Nature* 595.7866: 181-188.
- Bastani, Hamsa, Dennis Zhang, and Heng Zhang. 2022. Applied machine learning in operations management. *Innovative Technology at the Interface of Finance and Operations*. Springer: 189-222.
- Kelly, Brian, and Dacheng Xiu. 2023. Financial machine learning, *SSRN*, https://ssrn.com/abstract=4501707.

Homework

- Problem Set 1: Python Basics; Bias-Variance Trade-off; Cross Validation.

Session 2: Introduction to Deep Learning (Jan/16/2024)

Course

- Homework Recap
- Introduction to Deep Learning
- Presentation: Gu, Shihao, Brian Kelly, and Dacheng Xiu. 2020. Empirical asset pricing with machine learning. *Review of Financial Studies* 33: 2223-2273.

Recommended Reading

- *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/.
- *Probabilistic Machine Learning: Advanced Topics*, 2023, by Kevin Murphy, https://probml.github.io/pml-book/book2.html.
- *Deep Learning with PyTorch*, 2020, by Eli Stevens, Luca Antiga, and Thomas Viehmann.

Homework

- Problem Set 2: Implementing a Simple Neural Network.

Session 3: Prediction in Business and Traditional Natural Language Processing (Jan/23/2024)

Course

- Homework Recap
- Prediction Problems in Business Research
- Pre-processiong and Word Representation in Natural Language Processing
- Presentation: Mullainathan, Sendhil, and Jann Spiess. 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31(2): 87-106.

Recommended Reading

- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. Prediction policy problems. *American Economic Review* **105**(5): 491-495.
- Gu, Shihao, Brian Kelly, and Dacheng Xiu. 2020. Empirical asset pricing with machine learning. *Review of Financial Studies* 33: 2223-2273.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. Human decisions and machine predictions. *Quarterly Journal of Economics* **133**(1): 237-293.
- Bajari, Patrick, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang. 2015. Machine learning methods for demand estimation. *American Economic Review*, **105**(5): 481-485.
- Farias, Vivek F., and Andrew A. Li. 2019. Learning preferences with side information. *Management Science* **65**(7): 3131-3149.
- Chapters 2, 12 & 13, *Introduction to Information Retrieval*, 2008, Cambridge University Press, by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, https://nlp.stanford.edu/IR-book/information-retrieval-book.html.

Homework

- Problem Set 3: NLP Pre-processing.

Session 4: Natural Language Processing (II): Word2Vec (Jan/30/2024)

Course

- Homework Recap
- N-gram and Naïve Bayes
- NLP Performance Evaluations
- Word-to-Vector
- Presentation: Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. Transparency and deliberation within the FOMC: A computational linguistics approach. *Quarterly Journal of Economics*, **133**(2): 801-870.

Recommended Reading

- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. Text as data. *Journal of Economic Literature*, **57**(3): 535-574.
- Gentzkow, Matthew, Shapiro, Jesse M., and Taddy, Matt. 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, **87**(4): 1307-1340.
- Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. Transparency and deliberation within the FOMC: A computational linguistics approach. *Quarterly Journal of Economics*, **133**(2): 801-870.
- Cui, Ruomeng, Santiago Gallino, Antonio Moreno, and Dennis J. Zhang. 2018. The operational value of social media information. *Production and Operations Management*, **27**(10): 1749-1769.
- Parts I - II, *Lecture Notes for CS224n: Natural Language Processing with Deep Learning*, by Christopher D. Manning, https://web.stanford.edu/class/cs224n/.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. 2013. Efficient estimation of word representations in vector space. *ArXiv Preprint* arXiv:1301.3781.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems (NeurIPS)* 26.
- Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- Li, Kai, Feng Mai, Rui Shen, Xinyan Yan. 2021. Measuring corporate culture using machine learning. *Review of Financial Studies*, **34**(7): 3265-3315.

Homework

- Problem Set 4: Dictionary-based Sentiment Classification; Naïve Bayes; Pre-trained Word2Vecs.

Session 5: Natural Language Processing (III): RNN, Seq2Seq, and Attention (Feb/6/2024)

Course

- Homework Recap
- Recurrent Neural Network and Sequence-to-Sequence Models
- Attention Mechanism and Self-Attention
- Presentation: Timoshenko, Artem, and John R. Hauser. 2019. Identifying customer needs from user-generated content. *Marketing Science*, **38**(1): 1-20.

Recommended Reading

- Part V, *Lecture Notes for CS224n: Natural Language Processing with Deep Learning*, by Christopher D. Manning, https://web.stanford.edu/class/cs224n/.
- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, **27**.
- Bahdanau, Dzmitry, KyungHyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *ArXiv preprint* arXiv:1409.0473, https://arxiv.org/abs/1409.0473.
- Timoshenko, Artem, and John R. Hauser. 2019. Identifying customer needs from user-generated content. *Marketing Science*, **38**(1): 1-20.
- Sarzynska-Wawer, Justyna, Aleksander Wawer, Aleksandra Pawlak, Julia Szymanowska, Izabela Stefaniak, Michal Jarkiewicz, and Lukasz Okruszek. 2021. Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research*, **304**, 114135.

Homework

- Problem Set 5: Word2Vec for Classification; NLP APIs.

Session 6: Natural Language Processing (IV): Transformers (Feb/20/2024)

Course

- Homework Recap
- Transformer, BERT, GPT
- Pre-training, Fine-tuning
- Presentation: Huang, Allen H., Hui Wang, and Yi Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, **40**(2): 806-841.

Recommended Reading

- Part 10, *Lecture Notes for CS224n: Natural Language Processing with Deep Learning*, by Christopher D. Manning, https://web.stanford.edu/class/cs224n/.
- Chapter 11, Chapter 15.8, *Dive into Deep Learning* (2nd Edition), 2023, by Aston Zhang, Zack Lipton, Mu Li, and Alex J. Smola, https://d2l.ai/.
- Vaswani, Ashish, et al. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, **30**.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv preprint* arXiv:1810.04805, https://arxiv.org/abs/1810.04805.
- Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- Brown, Tom, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, **33**, 1877-1901.
- Huang, Allen H., Hui Wang, and Yi Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, **40**(2): 806-841.

Homework

- Problem Set 6: Sentiment Analysis with Transformers.

Session 7: Natural Language Processing (V): Large Language Models and Generative Artificial Intelligence (Feb/27/2024)

Course

- Homework Recap
- Prompting, Instruction Fine-tuning, Reinforcement Learning with Human Feedback, In-Context Learning, Chain-of-Thought, Emergent Abilities, LLM Training, and Beyond
- LLM and Generative AI in Business Research
- Presentation: Noy, Shakked and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, **381**: 187-192.

Recommended Reading

- Wei, Jason, et al. 2021. Finetuned language models are zero-shot learners. *ArXiv preprint* arXiv:2109.01652, https://arxiv.org/abs/2109.01652.
- Wei, Jason, et al. 2022. Emergent abilities of large language models. *ArXiv preprint* arXiv:2206.07682, https://arxiv.org/abs/2206.07682.

- Ouyang, Long, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, **35**, 27730-27744.
- Wei, Jason, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, **35**, 24824-24837.
- Kaplan, Jared. 2020. Scaling laws for neural language models. *ArXiv preprint* arXiv:2001.08361, https://arxiv.org/abs/2001.08361.
- Hoffmann, Jordan, et al. 2022. Training compute-optimal large language models. *ArXiv preprint* arXiv:2203.15556, https://arxiv.org/abs/2203.15556.
- Shinn, Noah, et al. 2023. Reflexion: Language agents with verbal reinforcement learning. *ArXiv preprint* arXiv:2303.11366, https://arxiv.org/abs/2303.11366.
- Reisenbichler, Martin, Thomas Reutterer, David A. Schweidel, and Daniel Dan. 2022. Frontiers: Supporting content marketing with natural language generation. *Marketing Science*, **41**(3): 441-452.
- Noy, Shakked and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, **381**: 187-192.
- Romera-Paredes, B., Barekatain, M., Novikov, A. et al. 2023. Mathematical discoveries from program search with large language models. *Nature*, https://doi.org/10.1038/s41586-023-06924-6.
- Boiko, D.A., MacKnight, R., Kline, B. et al. 2023. Autonomous chemical research with large language models. *Nature*, **624**: 570–578. https://doi.org/10.1038/s41586-023-06792-0.

Homework

- Problem Set 7 (Bonus): Testing the Persuasiveness of AI with LLM API.

Session 7.5: Final Project Proposal Discussion (Week of Mar/5/2024, No Class)

Session 8: Image Processing and Computer Vision (I) (Mar/12/2024)

Course

- Homework Recap
- Image Classification, Traditional Classifiers, Convolutional Neural Networks (LeNet, AlexNet, ResNet, and Beyond)
- Presentation: Zhang, Mengxia and Lan Luo. 2023. Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from Yelp. *Management Science* **69**(1): 25-50.

Recommended Reading

- *Course Notes for CS231n: Deep Learning for Computer Vision*, by Fei-Fei Li, https://cs231n.github.io/.

- Chapters 7 and 8, *Dive into Deep Learning* (2nd Edition), 2023, by Aston Zhang, Zack Lipton, Mu Li, and Alex J. Smola, https://d2l.ai/.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, **25**.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun. 2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
- Jean, Neal, Marshall Burke, Michael Xie, Matthew W. Davis, David B. Lobell, and Stefand Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science*, **353**(6301), 790-794.
- Zhang, Mengxia and Lan Luo. 2023. Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from Yelp. *Management Science* **69**(1): 25-50.

Homework

- Problem Set 8: Implementing AlexNet.

Session 9: Computer Vision (II) (Mar/19/2024)

Course

- Homework Recap
- Data Augmentation, Transformer for Image Recognition, Video Understanding, Objective Detection, Image Segmentation
- Presentation: Liu, Liu, Dzyabura, Daria, Mizik, Natalie. 2020. Visual listening in: Extracting brand image portrayed on social media. *Marketing Science* **39**(4): 669-686.

Recommended Reading

- *Course Notes for CS231n: Deep Learning for Computer Vision*, by Fei-Fei Li, https://cs231n.github.io/.
- Chapter 14, *Dive into Deep Learning* (2nd Edition), 2023, by Aston Zhang, Zack Lipton, Mu Li, and Alex J. Smola, https://d2l.ai/.
- Dosovitskiy, Alexey, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv preprint* arXiv:2010.11929: https://arxiv.org/abs/2010.11929.
- Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
- Liu, Liu, Dzyabura, Daria, Mizik, Natalie. 2020. Visual listening in: Extracting brand image portrayed on social media. *Marketing Science* **39**(4): 669-686.

- Glaeser, Edward L., Scott D. Kominers, Michael Luca, and Nikhil Naik. 2018. Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, **56**(1): 114-137.

- Yang, Jeremy, Juanjuan Zhang, and Yuhan Zhang. 2023. Engagement that sells: Influencer video advertising on TikTok. *Available at SSRN*: https://ssrn.com/abstract=3815124.

Homework

- Final Project.

Session 10: Unsupervised Learning (I) (Mar/26/2024)

Course

- Homework Recap

- Clustering, EM Algorithm, Dirichlet Latent Allocation, Topic Modeling

- Presentation: Netzer, Oded, Alain Lemaire, and Michal Herzenstein. 2019. When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, **56**(6): 960-980.

Recommended Reading

- Chapters 8.5 and 14, *The Elements of Statistical Learning* (2nd Edition), 2009, by Trevor Hastie, Robert Tibshirani, Jerome Friedman, https://hastie.su.domains/ElemStatLearn/.

- Blei, David M., Ng, Andrew Y., and Jordan, Michael I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, **3**(Jan): 993-1022.

- Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li. 2018. Modeling consumer footprints on search engines: An interplay with social media. *Management Science* **65**(3): 1363-1385.

- Netzer, Oded, Alain Lemaire, and Michal Herzenstein. 2019. When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, **56**(6): 960-980.

- Liu, Jia and Olivier Toubia. 2018. A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science* **37**(6): 930-952.

- Mejia, Jorge, Shawn Mankad, and Anandasivam Gopal. 2021. Service quality using text mining: Measurement and consequences. *Manufacturing & Service Operations Management* **23**(6): 1354-1372.

- Mueller, Hannes, and Christopher Rauh. 2018. Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, **112**(2): 358-375.

Homework

- Problem Set 10: EM Algorithm; Topic Modeling.

Session 11: Unsupervised Learning (II) (Apr/2/2024)

Course

- Homework Recap
- Image Generation Models, Variational Auto Encoder, Stable Diffusion
- Presentation: Yang, Jeremy, Juanjuan Zhang, and Yuhan Zhang. 2023. Engagement that sells: Influencer video advertising on TikTok. *Available at SSRN*: https://ssrn.com/abstract=3815124.

Recommended Reading

- *Course Notes for CS231n: Deep Learning for Computer Vision*, by Fei-Fei Li, https://cs231n.github.io/.
- *Denoising Diffusion-based Generative Modeling*, CVPR2022 Tutorial, https://cvpr2022-tutorial-diffusion-models.github.io/.
- Kingma, Diederik P. and Max Welling. 2019. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, https://arxiv.org/abs/1906.02691.
- Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* **33**: 6840-6851.
- Song, Yang et al. 2020. Score-based generative modeling through stochastic differential equations. *ArXiv preprint* arXiv:2011.13456, https://arxiv.org/abs/2011.13456.

Homework

- Final Project

Session 12: Algorithm and Fairness (Apr/9/2024)

Course

- Fairness and Discrimination from Computer Science and Economics Perspectives
- Testing Discrimination in Practice; Designing Algorithms with Discrimination Constraints
- Anti-Discrimination Law and Algorithm
- Presentation: Kallus, Nathan, Xiaojie Mao, and Angela Zhou. 2022. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*. **68**(3): 1959-1981.

Recommended Reading

- Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2017. Fairness in machine learning. *NeurIPS Tutorial*.

- Bertrand, Marianne, and Esther Duflo. 2017. Field experiments on discrimination. *Handbook of Economic Field Experiments (Volume 1, Chapter 8)*: 309-393.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. Human decisions and machine predictions. *Quarterly Journal of Economics* **133**(1): 237-293.
- Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. Racial bias in bail decisions. *Quarterly Journal of Economics*, **133**(4): 1885-1932.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunsteind. 2020. Algorithms as discrimination detectors. *Proceedings of the National Academy of Sciences*, **117**(48): 30096-30100.
- Kallus, Nathan, Xiaojie Mao, and Angela Zhou. 2022. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*. **68**(3): 1959-1981.
- Lambrecht, Anja and Catherine Tucker. 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science* **65**(7): 2966-2981.

Homework

- Final Project

Session 13: Final Project Presentation (Apr/16/2024)