DDA4210/AIR6002 Advanced Machine Learning
Course Project Topics

Basic Scheme (please read carefully)

- **DDA 4210 Grading policy:** The project will be evaluated based on the final report&code **(25%)** and in-class group presentation **(75%=15%peer+25%TA+35%Instructor)**. The project has 35 points in total. The report should contain at least the following sections: *significance and novelty of the study, data collection or preprocessing, methodology, numerical or experimental results, and conclusion.* Determine the orders of these sections by yourself. The length of the report should be **3-4 A4 pages (10pt, single column, normal spacing, excluding references and appendices)**. Every presentation should be finished in 8 minutes. Exceeding the time limit will trigger a score deduction (-0.5 point per 0.5 minute). There will be a 2-minute QA session after each presentation.

- **AIR 6002 Grading policy:** The project will be evaluated based on the mid-term proposal **(10%)**, final report&code **(25%)** and in-class group presentation **(65%=10%peer+25%TA+30%Instructor)**. The project has 60 points in total. The report should contain at least the following sections: *significance and novelty of the study, data collection or preprocessing, methodology, numerical or experimental results, and conclusion.* Determine the orders of these sections by yourself. The length of the mid-term proposal should be a **2-pager (A4, single column, normal spacing, 10pt)** describing the project problem and the proposed methodology. The length of the report should be **8-10 A4 pages (10pt, single column, normal spacing, excluding references and appendices)**. Every presentation should be finished in 8 minutes. Exceeding the time limit will trigger a score deduction (-0.5 point per 0.5 minute). There will be a 2-minute QA session after each presentation.

- **Collaboration policy:** Students can form a group up to 4 people for this course project. The grading scheme will be the same regardless of the size of the group, as long as it is less than or equal to 4. Note that due to different grading policies, students from DDA4210 and AIR6002 cannot cross team up.

- **Conflict of Project Topics:** As part of the project guidelines, *students are required identify whether they have used or plan to use similar project contents in other courses*, such as CSC3160. *To ensure academic integrity and fairness, it is important for students to report any potential conflicts of interest or overlap in project contents with other courses.* **Failure to report such conflicts may result in penalties or disciplinary action.** While students are allowed to choose the same project for different courses, they must demonstrate additional effort and progress beyond what is required in each course. This approach is intended to ensure that students are fully engaged in the project and are able to develop their skills and knowledge to the fullest extent possible. Additionally, it allows students to integrate their learning across different courses and to apply their knowledge to real-world problems.

- **Submission Format:** Report and mid-term proposal templates are provided.

  - **Report Templates:** `https://www.overleaf.com/latex/templates/neurips-2023/vstgtvjwgdng`
  - (For AIR6002) **Proposal Templates:** `https://www.overleaf.com/latex/templates/my-geo2010-p2-template/pysgszqwjqys`
  - (For AIR6002) The file name of your proposal should be in the form of '**Proposal_Group_Num**'.
  - The file name of your presentation slides/PPT/PDF should be in the form of '**Pre_Group_Num**'.
  - The file name of your report PDF/zip should be in the form of '**Report_Group_Num**'.
  - Please show the names of all members of your group on the first page of your proposal/slides/PDF.
  - Submit your source code together with your report if there is any. Otherwise, you will lose 5% of the project score.

- **Evaluation criteria (for both report and presentation) and guidelines:**
  - **significance (25%):** the problem you studied should be interesting and hasn't been well-solved.
  - **novelty (25%):** at least one aspect of your problem, data, and method is novel. Do not apply an existing method to an existing dataset to solve an existing problem. You may collect or create a new dataset, propose a new problem, devise a new method, or modify an existing method.
  - **technical soundness (25%):** the techniques of data collection, processing, modeling, analyzing, etc., you used, should be reasonable, correct, and explainable.
  - **completeness (25%):** your report and presentation should contain the necessary and important information, description, explanation, or/and discussion, etc., about your task, data, method, and results.

- **Outcome Evaluation:**
  - (For DDA4210) Score = [Report (25%) + Presentation (75%)] * r_{ind}
  - (For AIR6002) Score = [Proposal (10%) + Report (25%) + Presentation (65%)] * r_{ind}
  - $r_{\text{ind}} \in \{0.5, 0.8, 1\}$: it is rated by your teammates on your contribution. 0.5 (or 0.8) means your contribution is less than 20% (or 50%) of the expected workload (namely, average workload, overall workload of the project divided by the number of members in the team). If your contribution is equal to or larger than 0.8 of the expected workload, you will get $r_{\text{ind}} = 1$.

- **Deadlines:** Detailed presentation schedules and report submission deadlines will be planned and announced. For AIR6002 students, the mid-term proposal deadline is **March 25th 2024**.

Below we exemplify some potential project topics. **Students are free to choose any topics related to machine learning, not limited to the following examples provided that cover topics in all the lectures of DDA4210/AIR6002.** The goal of this course project is to encourage creativity and flexibility in exploring personal interests and strengths in machine learning.

1. **Generative AI**

   Watermarking. The proliferation of Large Language Models (LLMs) like ChatGPT has ushered in a new era of AI-driven applications, offering unprecedented capabilities in text generation, comprehension, and interaction. However, this rise also brings challenges in intellectual property protection, model authenticity, and misuse prevention. The concept of watermarking for LLMs proposes embedding distinctive markers within the model's outputs to verify its origin, ownership, or authenticity. Inspired by the work of Kirchenbauer et al., a potential project is to analyze the fundamental limit; explore, implement, and evaluate effective watermarking techniques for LLMs, addressing both the technical intricacies and ethical considerations of such implementations.

   References:

   [1] Kirchenbauer, John, Jonas Geiping, Yuxin Wen, Jonathan Katz, Ian Miers, and Tom Goldstein. "A watermark for large language models." arXiv preprint arXiv:2301.10226 (2023).

   LLMs as decision models. Recent advancements in Large Language Models (LLMs) have demonstrated their potential beyond text generation, extending into the realm of complex decision-making and policy implementation. Inspired by groundbreaking research, including the works of Brooks et al. and Yang et al., you may delve into the capabilities of LLMs in executing policy iteration and facilitating collective decision-making processes. For example, you may investigate how LLMs can be leveraged to simulate decision-making scenarios, influence policy-making, or enhance collaborative decision-making among human groups.

   References:

   [1] Brooks, E., Walls, L., Lewis, R. L., & Singh, S. (2024). Large language models can implement policy iteration. Advances in Neural Information Processing Systems, 36.

   [2] Yang, J. C., Korecki, M., Dailisan, D., Hausladen, C. I., & Helbing, D. (2024). LLM Voting: Human Choices and AI Collective Decision Making. arXiv preprint arXiv:2402.01766.

   Other topics include prompting LLM agents for your customized tasks.

2. **Trustworthy ML with ChatGPT**

   ChatGPT is a large language model that has the potential to generate human-like responses to a wide range of queries, but the model's decision-making process is often seen as a black box. Moreover, how to obtain a comprehensive understanding of the ability of ChatGPT is still an open problem. XAI tools can help to shed light on the internal workings of the model and provide explanations for the responses it generates.

   One potential task is on developing a ChatGPT-based virtual assistant that is capable of explaining its reasoning to users. The goal of the project would be to develop a system that not only provides accurate responses to user queries but also provides an explanation of how it arrived at those responses.

   Another possible approach would be to use trustworthy machine learning techniques to identify and analyze biases in the responses generated by ChatGPT. ChatGPT is trained on a large corpus of text, which may contain biases that are reflected in the model's responses. Trustworthy machine learning techniques could be used to identify and quantify these biases, enabling researchers to develop strategies for mitigating them and making the outputs more reliable.

Besides, developing useful prompts for interesting practical scenarios is also a relevant topic.

References:

[1] https://github.com/f/awesome-chatgpt-prompts

[2] Regulating ChatGPT and other Large Generative AI Models[1]

[3] Goal Driven Discovery of Distributional Differences via Language Descriptions[2]

[4] Causal-Discovery Performance of ChatGPT in the context of Neuropathic Pain Diagnosis[3]

3. **Detect Contexts Generated by AI**

   The proposed course project involves building a system to identify whether a given text was generated by AI tools such as ChatGPT, or by a human. Students would create a dataset of text samples, extract features from the samples, train a machine learning model to classify the text, evaluate the performance of the model, and discuss the results and potential applications of the model. The project would provide students with practical experience in building machine learning models for text classification and exploring the challenges and ethical implications of identifying machine-generated text.

4. **Detect Images Generated by AI**

   With the development of deep generative models such as GANs and diffusion models, it is very difficult to identify whether an image is made by human beings or AI. In this project, you need to construct a model or simply a classifier, to determine whether a given image is true or fake. For simplicity, you may focus on face images or animal images.

5. **Online Learning in Control and Decision-Making.**

   Online learning is an important area of research that enables algorithms to adapt to changing data distributions in real-time. There have been lots of recent advances in online learning in the interdisciplinary fields such as control and decision-making. This project will be focusing on designing new models related to online learning and optimization and developing online learning algorithms with theoretical guarantees.

   References:

   [1] Book: *Introduction to Online Convex Optimization* by Elad Hazan[4]

   [1] Ghai, Udaya, et al. "Robust Online Control with Model Misspecification." Learning for Dynamics and Control Conference. PMLR, 2022.

   [2] Agarwal, Naman, et al. "Online control with adversarial disturbances." International Conference on Machine Learning. PMLR, 2019.

6. **Causal Feature Learning.**

[1] https://arxiv.org/pdf/2302.02337.pdf

[2] https://arxiv.org/pdf/2302.14233v1.pdf

[3] https://arxiv.org/pdf/2301.13819.pdf

[4] https://sites.google.com/view/intro-oco/

Causal feature learning is a special caual learning method, as a causal inference framework rooted in the language of causal graphical models in 2009. The goal of causal feature learning is to discover high-level causal relations from low-level data, and at reducing the experimental effort to understand confounding among the high-level variables. The focus of this project will be implementing existing causal feature learning algorithms and developing new causal models to adapt causal feature learning to more applications with confounders and high-dimensional data.

References:

[1] Chalupka, Krzysztof, Frederick Eberhardt, and Pietro Perona. ”Causal feature learning: an overview.” Behaviormetrika 44 (2017): 137-164.

[2] Chalupka, Krzysztof, Pietro Perona, and Frederick Eberhardt. ”Visual causal feature learning.” arXiv preprint arXiv:1412.2309 (2014).

[3] Kinney, David, and David Watson. ”Causal feature learning for utility-maximizing agents.” International conference on probabilistic graphical models. PMLR, 2020.

[4] Li, Xin, et al. ”Confounder identification-free causal visual feature learning.” arXiv preprint arXiv:2111.13420 (2021).

7. **Trustworthy AI in Physical Worlds**

   As we have introduced in our first lecture, applying ML methods to physical applications is a hard task. Because in real-world cyber-physical systems such as power grids, transportation systems, etc., there are often physical constraints. The project may contain several tasks, including researching challenges associated with applying machine learning to cyber-physical systems, identifying a specific real-world system, developing a machine learning model to predict its behavior, designing and implementing trustworthiness mechanisms, and evaluating the performance of both the model and the trustworthiness mechanisms. The project will be assessed based on criteria such as clarity, quality of research, effectiveness of the model, understanding of physical constraints, and quality of the report and presentation. The project will provide students with valuable hands-on experience in designing trustworthy AI systems for cyber-physical systems.

8. **Reinforcement Learning with Human Feedback**

   Developing an RL-based system that takes into account human factors, such as attention and decision-making, to optimize a task or process. The proposed course project involves developing an RL-based system that takes into account human factors, such as attention and decision-making, to optimize a task or process. Students would identify a task, model it using RL techniques, analyze the human factors that influence it, integrate these factors with the RL model, implement and evaluate the RLHF system, and discuss the results and implications of their work. The project provides students with hands-on experience in developing RLHF systems and exploring the challenges involved in designing such systems.

9. **Construct a recommendation system for students at CUHK-Shenzhen**

   We have learned a few methods of recommendation systems. Now it is the time to make practice using your knowledge and skill. In this project, you may construct a recommendation system for the students at CUHK-Shenzhen to select their favorite dining rooms, foods, courses, academic supervisors, etc. You may need to collect data by yourself.

Hint: You can collect data by making surveys in the Wechat groups of your courses. You may also collect some anonymous information about the users. Designing questionnaires is an art.

10. **Construct a generative model to generate images related to CUHK-Shenzhen**

    Consider generating images related to our campus or your university life.

11. **Graph neural networks for daily life**

    You may try to use GNN to solve one real problem in our daily life. Do not use the benchmark datasets for node or graph classifications. You need to collect data by yourself.

    Hint: You can classify APPs, websites, food, cars, animals, medicine, movies, actors/actresses, companies, cities, countries, etc, whenever there are available graphs or you can construct graphs using some side information rather than the original features.

12. **Learning theory for GCN-based semi-supervised learning**

    We have introduced the learning theory for supervised learning. You may study the learning theory for GCN-based semi-supervised learning. There will be two differences. The first one is that there is a graph presented in both the training and testing stages in GCN. The second one is that the number of test samples (unlabeled nodes) is given and fixed, while in supervised learning, the test samples are unlimited and we care about only the expectation.

13. **Causal inference or discovery in our daily life**

    In this project, you need to identify the causality of some data in our daily life. You need to collect the data by yourself.

    Hint: You can try to study the causality between: 1) the GDPs of China, US, and Europe; 2) the stock markets of China and US; 3) the prices of different stocks; 4) the prices of different food; 5) education and economy; 5) war and poverty; etc.

14. **Interpretability of diffusion models**

    Diffusion models have shown excellent performance in generating realistic samples. However, the interpretability of diffusion models is much lower than those of VAE and GANs.

15. **Privacy and fairness in ensemble learning**

    Privacy and fairness in ensemble learning are rarely studied in the literature, although they are very important in real problems. You may present fair or secure algorithms for ensemble learning.