

# Spring 2022

## Introduction to Artificial Intelligence

### Project Proposal

Mar. 1, 2022

#### Introduction

In the final project, you will work in groups of **three to four students** to apply methods learned in the class to a real-world problem that you are interested in. In this final project, the topic scope is **AI for Social Impact**, which has been a crucial issue in recent years. Since this is a class project instead of a top conference, we note that anything that happens in your life can also be framed as a real-world social problem.

For instance, you go to Dining Hall 2 to buy dinner every day and discover some patterns that you need to line up for minutes. Thus, you want to design machine learning methods that can predict the waiting time based on some conditions so that you can make a better arrangement yourself. This is not a huge issue in society, but this solution can benefit most students for scheduling and avoiding the crowd, which is also a social impact!

Here we provide some keywords in AAAI AI for Social Impact Track, but not limited to these domains: Agriculture, Education, Financial, Mobility/Transportation, Ethical Issues, Security and Privacy, Social Welfare, Justice, Fairness and Equality, and Urban Planning. We encourage you to come up with novel and interesting ideas!

#### Proposal Requirement

The proposal should include the following items. Note that you should include all of these items in your report.

- **Problem Statement and Task Definition.** What does your system do (what is its input and output)? What real-world problem does this system try to solve? Make sure that the scope of the project is not too narrow or broad. For example, building a system to predict the exact stock price is unrealistic, whereas predicting the relative changes about stocks is more reasonable.
- **Description of the challenges.** What are the challenges of this task? Which topics (e.g., search, MDPs, etc.) might be able to address those challenges (at a high-level, since we haven't covered any techniques in detail at this point)?

- **Input/Output Behavior with Concrete Examples.** Concrete examples of both the inputs and outputs (explain what the input and output variables should look like and how they interact with the system). You should collect some preliminary data that you can use in your description of the input and output behavior. Specifically, reference what data you are using. We recommend using the following open source data (of course you can choose from other sources). Note that you should not use data with ethical issues.
  - Open data platform of Taiwan government
  - Kaggle
  - Papers with code
- **Related works.** Search the Internet for similar projects and mention the related research and projects.
- **Methodology.** How will you approach the problem? Why does this method fit with your problem? Identify the challenges of building the system and the phenomena in the data that you're trying to capture. How should you model the task (e.g., using search, machine learning, logic, etc.)? There will be many ways to do this, but you should pick one or two and determine how the methods address the challenges as well as any pros and cons. What algorithms are appropriate for handling the models that you came up with, and what are the tradeoffs between accuracy and efficiency? Are there any implementation choices specific to your problem?
- **Evaluation Metrics.** How will you measure the success of your system? Why does this metric work best for this problem? For this, you need to obtain a reasonably sized dataset of example input-output pairs, either from existing sources, or collecting one from scratch. A natural evaluation metric is accuracy, but it could be memory or running time.
- **Baselines.** Before developing your primary approach, you should implement baselines and oracles. These are really important as they give you intuition for how easy or hard the problem you're solving is. Intuitively, baselines give lower bounds on the performance you will obtain. If this gap is too small, then you probably don't have a good task. Importantly, baselines should be relatively easy to implement and can be done before you invest a lot of time in implementing a fancier approach. In this part of the proposal, you should describe what your baselines are (we expect at least 2-3 baselines in your final project presentation). While you do not need to implement it for the proposal, it is recommended that you have started implementing them by this point. Note that baselines can be simple algorithms like a rule-based method or a simple classifier.

## An Example

We provide an imaginary example of the first three bullets to show how to approach the final project (you can find there are some impractical things).

- **Problem Statement and Task Definition.** As mentioned above, the motivation for the task is we found that we would line up for minutes when buying dinner in NYCU. Hence, given the crowd data within a month, the problem we want to solve is to predict the time we need to wait for line up.
- **Description of the challenges.** There are three challenges in our task. First, there isn't public data describing the number of people in Dining Hall 2 and the waiting time of each vendor. Thus, we plan to collect by ourselves and limit the targets only at the first floor. Besides, we plan to collect only from 17:00 to 19:00 on weekdays to avoid the imbalance issue. Second, this problem might be a time-series forecasting problem. That is, given the information between 17:00 to 18:00, the goal is to predict the time between 18:00 to 19:00. Third, it is hard to be solved by applying machine learning models since machine learning models fail to consider the relationship between different time slots.
- **Input/Output Behavior with Concrete Examples.** The input sample includes [time, vendor\_a, the number of crowd, menu of vendor\_a], and the predicted output is the time that we can buy dinner from now on. Here we skip the example to avoid some potential problems, but we suggest you use an illustration for easy understanding.

## Work Plan

Since this is a team project, we need to make sure everyone in the team contributes to the project. However, we understand that it is hard and meaningless for you to pre-assign work to team members. Therefore, instead of a work division table, we require you to provide these information in the proposal:

- Time schedule
  - No need for an exact date, just list all the tasks and sub-goals of your project, then tell us which part will be done first.
- Discussion
  - We want to see how your team works together throughout this project, so please open a discussion board/channel (e.g., Hackmd note, Trello board, Notion) and record your discussion throughout the project.
  - The record should be the summary of your regular discussion, which might be what's the progress of your work (something has completed, something hasn't completed, next targets), the trade-off between different implementation methods and your final choice, and something the prototype needs to be

modified. Notice that this is a **tool** for you to develop your application, please do not only see it as a class work.

- In the proposal, just open the discussion board/channel and provide the link in the slide.
- Repo
  - Please open a github/gitlab repo for your project, you need to upload the source code of the project in the repo and a README to describe your functionality and the steps of reproducibility at the end of the project.
  - (Optional) We encourage you to do version control and show the contribution of each member using git, you will get extra points if you do this.
  - In this proposal, just open the repo and provide the link in the slide.

## Discussion

TAs had opened a channel **Final Project 討論區** on Microsoft Teams of the course, you can ask questions about the final project in the channel. TAs will answer questions in the channel as soon as possible.

## Submission

1. **The deadline for this homework is 4/18 (Mon.) 23:55:00.**
2. **Please fill the [registration form](#) before 3/8 12:00:00** or we will arrange for grouping unsatisfied students after this first deadline.
3. The proposal should be **at most 4 pages** excluding references. Both Chinese and English are acceptable.
4. Submit one proposal with the filename of **Proposal\_Team{ID}.pdf** (e.g., Proposal\_Team05.pdf). Please write 05 if you are Team 5 for example.
5. Late submission leads to a score of  $(\text{original score}) \times 0.85^{\text{days}}$ , for example, if you submit your homework right after the deadline, you will get  $(\text{original score}) \times 0.85$  points.
6. **We only accept one pdf file**, wrong format or naming format cause -10 points to your score (after considering late submission penalty).
7. Notice that the proposal will count as part of your final score, please design your project as detailed as possible.
8. **Only 1 team member needs to submit the proposal.**

9. We might select some outstanding proposals and invite you to present during the class and would have a bonus if you present.
10. If there is anything you are not sure about submission, ask in the discussion forum.
11. You can refer [here](#) for some motivations.