# CS 4973-05: Responsible Machine Learning — Fall 2023 — Avijit Ghosh

Day 2 — Preparation Questions For Class
Due: Tuesday 9/12/2023 at 9:00 am on Canvas

Name: [Put your name here]

You may consult any and all resources in answering the questions. Your goal is to have answers that are ready to be shared with the class (or on a hypothetical job interview) as written. Your answers should be as concise as possible. When asked to explain a figure, your response should have the following structure: provide context (state what experiment was being run / state what problem is being solved), state what has been plotted, remark on what we observe from the plots, and interpret the results.

We recommend you use Overleaf for easy editing of this TeX document.

**Directions:** Read 'Machine learning: Trends, perspectives, and prospects'

- Read the whole paper

**Question 1.** *What do the authors mean by learning problem? What are the three paradigms of learning problems?*

**Response:**

**Question 2.** *Explain Figure 2. In your explanation, explain how it connects to the claim that 'Supervised learning systems generally form their predictions via a learned mapping f(x), which produces an output y for each input x'. See the preamble above for comments on how to explain a figure.*

**Response:**
Context:
What is plotted:
What we observe:
Interpretation:

**Question 3.** *What is Big Data? What new challenges has Big Data created as compared to earlier Machine Learning techniques?*

**Response:**

**Question 4.** *Explain Figure 3.*

**Response:**

**Question 5.** *If you created a Machine Learning algorithm that teaches a robot to walk by starting with random movements, but learning from failures and falls and eventually beginning to walk normally, what paradigm does this type of learning fall under? Explain why.*

**Response:**

**Question 6.** *What is differential privacy? Give two different examples of machine learning problems where differential privacy would be a useful technique.*

**Response:**

**Question 7.** *Although this paper was written in 2015, the authors presciently suggest that environmental/online data of people could be collected to prevent the spread of a communicable disease during a global pandemic. Based on what you know from COVID-19, do you think their suggested techniques would have been helpful? What are some potential pitfalls? (Hint: Think about privacy and equity concerns in such a data driven approach)*

**Response:**