

CS 221: Artificial Intelligence: Principles and Techniques

Assignment 8: From Language to Logic

May 27, 2022

SUNet ID: jchan7
Name: [Jason Chan]
Collaborators: None

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Before you get started, please read the Assignments section on the course website thoroughly.

In this assignment, you will get some hands-on experience with logic. You'll see how logic can be used to represent the meaning of natural language sentences, and how it can be used to solve puzzles and prove theorems. Most of this assignment will be translating English into logical formulas, but in Problem 4, we will delve into the mechanics of logical inference. Finally, for the ethics question you will revisit the ethical frameworks and concepts you learned about this quarter through once again writing negative impact statements.

Problem 1: Propositional logic

[Code only]

Problem 2: First-order logic

[Code only]

Problem 3: Liar puzzle

[Code only]

Problem 4: Odd and even integers

[Code only]

Problem 5: Semantic parsing (extra credit)

[Code only]

Problem 6: Revisiting Ethical Issue Spotting

In [Homework 1: Foundations](#), you practiced spotting ethical issues in real-world scenarios. In this question, you will return to a similar task equipped with the ethical frameworks and concepts you have learned and applied during the quarter. Just like on homework 1, you will write a negative impacts statement for the scenarios below that justifies your decision using one of the ethical frameworks / concepts you learned about this quarter. Specifically, you will:

1. List one of the sixteen numbered ethical concerns taken from the [NeurIPS Ethical Guidelines](#) page
2. List an ethical concept or framework from the course, which can be found listed below
3. Finally, justify why the chosen ethical concern from the NeurIPS guidelines applies to this scenario *using the ethical concept or framework from the course*.

Ethical Frameworks / Concepts From the Course:

1. Theories of Distributive Justice (utilitarianism, prioritarianism, difference principle, avoid compounding injustice) If using this framework in your answer, pick a specific theory of distributive justice.
2. Communicating Limitations with a Transparency Statement

3. Theories of Responsibility To Future Generations (Veil of Ignorance, Uncertain Future, Precautionary Risks, Symmetry of Future Generations, Vulnerability of Future Generations) If using this framework in your answer, pick a specific theory of responsibility to future generations.
4. Wicked Problems
5. Ethis Dumping
6. Dual-Use Technologies

Example Scenario: Recent resolutions by the United Nations and policy statements by roboticists propose a 5 year moratorium on “development, deployment, transfer, and use of anti-personnel lethal autonomous weapon systems,” or weapons systems that can target and kill people without direct human oversight. Human oversight ensures that a human takes moral and legal responsibility for all actions and for ensuring that actions comply with the laws of war. Some lethal autonomous drones couple facial recognition, used to autonomously identify their targets, with action-planning used to kill them.

A research paper proposes a new AI-based action-planning system for automatically detecting and aiding tourists stranded in Death Valley in the summer. The system couples two systems on an unmanned drone: (a) computer vision to do vehicle recognition and detect the stranded cars in the desert and (b) a motorized arm to provide the stranded tourists with water bottles while rescuers are on their way, since humans can easily die of heatstroke and exhaustion when temperatures exceed 50°C, as often happens in Death Valley. The drone is able to detect disabled vehicles with almost 99.5% accuracy, and so the paper proposes that the drone system can likely be fully automated with no human verification of the stranded car before delivering the water, also speeding up the time of water delivery and potentially saving the stranded tourists before heatstroke sets in.¹

Example answer: The proposed technology Directly facilitates injury to living beings, as it can be used as part of a weapons system without significant modification. A drone capable of recognizing cars and delivering water, and more importantly coupling these capacities as part of a multi-stage autonomous planning system, is a dual-use technology, or a technology that can be used for peaceful purposes or as part of a weapon of war. It would be very simple to replace the water delivery system with a weapon of war, and since the drone has a very high accuracy rate, it can likely be operated successfully without human supervision, in violation of the UN resolution.

[3 points] **Scenario 1:** After their study design was not approved by IRB in their home country, a research group tests their new low-cost AI-based screening device for cervical

¹Inspired by <https://www.cc.gatech.edu/ai/robot-lab/online-publications/AWS.pdf> and <https://www.cc.gatech.edu/ai/robot-lab/online-publications/AWS.pdf>.

cancer in another country with a lower standard for clinical trials. In the home country, the standard of care for the control versus test group for a new screening technology is to still administer the standard of care test to all patients in the trial, which in this case would be a Pap smear. However, in the host country this research team administered only the low-cost AI-based screen to 140,000 women in the test group and administered **no** screen to 140,000 women in the control group. No pap smears were administered to any women in the study. Of the 140,000 women who received no screen in the control arm of the study, 254 of them ended up developing cervical cancer and dying as a result.²

What we expect: A 2-5 sentence paragraph for each of the scenarios where you **A.** identify one ethical concern from the NeurIPS Ethical Guidelines and **B.** one ethical concept from the course and justify why you think they apply, and **C.** link the ethical concept to the NeurIPS ethical concern. Copy out and underline both the ethical checklist item and the ethical concept to which you are referring as part of your answer (i.e.: Severely damage the environment; wicked problems).

Your Solution: This scenario raises the ethical concern about human subject experimentation and whether it has been reviewed and approved by a relevant oversight board. This is a concern because unlike the home country, the host country does not need to conduct pap smears on women in the control group and 254 died as a result. Fewer people would've died in the host country if the research group applied the home country's standard for giving pap smears to the control group.

This is an example of ethics dumping - the research group knowingly chose another country with a lower standard for clinical trials after being rejected from their home country. The research group must be responsible for having the prior knowledge about the testing standards of their home country and willingly or negligently disregarded them when conducting the clinical trials in the host country.

This ethical concept links to the ethical concern because as a consequence of the ethics dumping, the research group had the opportunity to simultaneously improve the livelihood of the participants in the control group by offering them pap smears but did not do so. A strong link can be drawn between this inaction, the 254 deaths to suggest that the host country's oversight board is inadequate.

²[2] Modified from [cervical cancer screening in India](#)

[3 points] **Scenario 2:** A recent economic forestry paper proposed an MDP model for optimizing commercial logging in a National Forest that uses a 30-year time horizon to suggest when trees should be cut for optimal logging yields. It recommends cutting each tree once it is at 30 years or older, suggesting the policy of harvesting every tree in the forest every 30 years and leading to a clear-cut forest by the end of the simulation.³

What we expect: A 2-5 sentence paragraph for each of the scenarios where you **A.** identify one ethical concern from the NeurIPS Ethical Guidelines and **B.** one ethical concept from the course and justify why you think they apply, and **C.** link the ethical concept to the NeurIPS ethical concern. Copy out and underline both the ethical checklist item and the ethical concept to which you are referring as part of your answer (i.e.: .Severely damage the environment; wicked problems).

Your Solution: The ethical concern is that the economic forestry paper's MDP model recommends a policy that will severely damage the environment. The MDP's objective is to *[optimize] commercial logging* and its simulation resulted in a *clear-cut* forest.

The ethical concept that applies to this ethical concern is the Theories of Responsibility To Future Generations. There already exists scientific consensus that human-induced climate change and destruction of the biosphere can lead to irreversible damage to the planet. Clearing more forests only further contributes to effects of human-induced climate change and biosphere destruction. Consequently, future generations will bear the cost and suffering of inheriting that degraded world.

This ethical concept links to the ethical concern because it assigns responsibility for taking actions to minimise the negative impacts for future generations. If viewed over a short time horizon such as 5 years, the recommendations by the economic forestry paper's MDP model might seem reasonable. However, when considering long time horizons such as 50 or 100 years and the affected people, we rethink our choice of actions in the present.

³[3] Modified from [optimal forest growth modeling](#).