Ethics Module: Deep Reinforcement Learning for Robotics

Instructor: David Held
Authors: Mosam Dabhi, Kshitij Goel
May 6, 2021

The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA

Contents

List of Figures

Chapter 1

Learning Objectives

We have divided the learning objectives into three parts

1. For each of the themes during the course, the student will be able to identify the key impacts on the society.

   - Assessment: Student groups choose one theme based on their interest and write a piazza post corresponding to that lecture.

2. Being able to present the impact of these technologies to a wider audience.

   - Assessment: Student groups should demonstrate the style, organization, and clarity to cohesively present an ethical argument over a high-level research theme.

3. Being able to identify valid resources on the impact of deep reinforcement learning techniques.

   - Students will identify and post resources (scholarly articles, classical books, recent events) related to the lecture content from an ethical standpoint.

The detailed rubric for the assessment would be discussed in the subsequent chapters.

Chapter 2

Reading

Students are expected to find the relevant sources of ethics related articles and do the corresponding read for each of the themes mentioned during the class.

2.1   Suggestions to find the scholarly articles:

- In many of the recent research articles and/or papers, we can find a broader outcomes section. Within those sections, students can find the information on authors and their corresponding works who have investigated the ethical ramifications of the concerned technology or the theme.

- Many a times, students could start the search for scholarly reading of ethics by backtracking a thread of a news article found on the world wide web. Due to considerable interest of media in the breakthroughs led by major tech companies, numerous news articles could be retrieved and be used as a token of reference to search for the scholarly articles.

Students are thus expected to finish the search for scholarly reading and post their findings on Piazza discussion post, post lectures that focused on specific theme.

Chapter 3

Class Outline

Since the proposed module is intended to be leveraged as a seamless integration into the existing structure of the class, the class outline for the ethics module is supposed to echo earlier presentation classes where two students dissected the ethical ramifications of their chosen theme.

3.1 Pre-class work

- Throughout the course, students accumulate Piazza posts for each seminar.
- For the ethics module lecture, they create a presentation.

3.2 Class outline and timing

We plan to keep one lecture towards the end of the course.

- For each of the 13 research themes, students have 5-minute presentations each. This leads to a total of 65 minutes.
- The final 15 minutes leads to the questions, followed by in-class discussions.

Chapter 4

Slides

Some of the example potential slides that could be used by the instructor to set the stage for the class session is shown below. In each of these slides, the major topic is shown on the top, followed by sub-topics shown below each of them.

Ethics Module

Topic: Why Ethics Matters for Autonomous Cars

Instructor: Prof. David Held

Lin, Patrick. "Why ethics matters for autonomous cars." Autonomous driving. Springer, Berlin, Heidelberg, 2016. 69-85.

Figure 4.1: Why Ethics Matters for Autonomous cars? – Title slide

Why Ethics Matters?

Ethics beyond crash-avoidance?

Trolley problem and crash-optimization

Psychological point-of-view w.r.t. harm caused

Figure 4.2: Why Ethics Matters for Autonomous cars? – Intro Slide

Scenarios that Implicate Ethics

Animal accidents

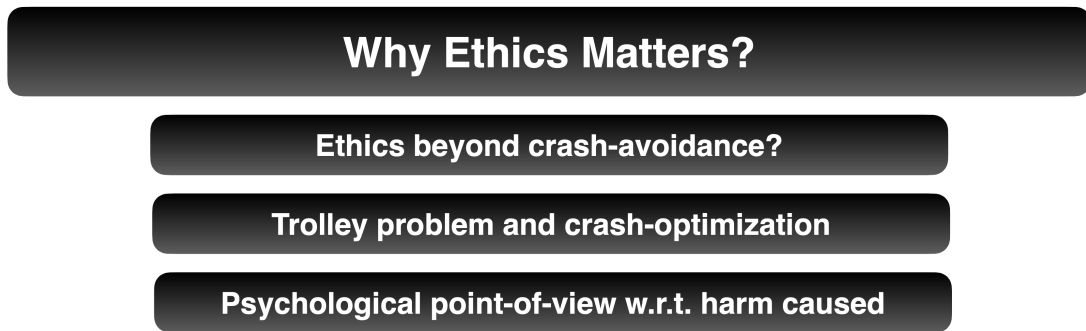Should autonomous cars self-sacrifice?

Dodging the legal implications

Figure 4.3: Why Ethics Matters for Autonomous cars? – Scenarios slide
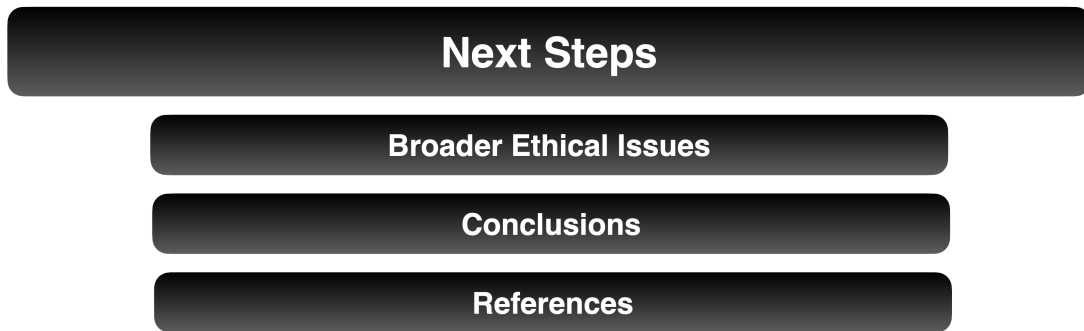
Figure 4.4: Why Ethics Matters for Autonomous cars? – Concluding slide

Chapter 5

Activity/Discussion

5.1   Piazza posts - Scholarly resources

The instructor asks the students to use the tricks mentioned in Sec. 2.1 and encourage students to move the discussion on ethical connections to Deep RL on the Piazza.

5.2   Piazza posts - Ethical considerations

Based on the introductory slides presented by the instructor (Fig. 4.1-Fig. 4.4), piazza posts would be encouraged on the introspection of ethical connections of autonomous cars (for example) on the society, as a whole.

Chapter 6

Assignment

6.1   Discussion assignment

In this assignment, student groups choose one theme based on their interest and write a piazza post corresponding to that lecture or theme. This is a small assignment that encourages students for participation and engage in ethical discussions. We incentivize students by giving extra credits who substantially further and engage in the ethical conversation on Piazza.

6.1.1 Rubric: Discussion posts on Piazza identifying and discussing the ethical ramifications

Assessment rubric: Discussion Posts	
Criteria	**Rating basis**
Knowledge and Understanding of Content and Applicability to Professional Practice	Discussion response and replies to classmates show evidence of knowledge and understanding of course concepts by correctly and consistently applying them to professional practice.
Generates Learning within the Community	Discussion response and replies to classmates elicit responses and reflections from other learners ; they attempt to build upon and integrate multiple views to take the discussion deeper.

Figure 6.1: Piazza post rubric

6.2 Presentation assignment

This is a major assignment intended to be kept in the final lecture of the class - focusing on ethical connections of the proposed technologies throughout the class. At this stage, since the students would already be familiar with the paired presentation format, they could simply leverage the skills learned throughout the course to discuss the topics from the ethical standpoint. Students can lead the discussion in any way they seem comfortable as long as it generates ethical learning within the class community. In hindsight, student groups should demonstrate the style, organization, and clarity to cohesively present an ethical argument over research themes discussed during the class. Similar rubric would be used as other presentations. For an alternative rubric, a more extensive breakdown is given below

6.2.1 Rubric: Student presentations on ethical ramifications of the Deep RL theme.

Assessment rubric: Student Presentations				
Criteria	**Rating**			
	16 - 20 Meets Expectations	14 - 16 Acceptable	12 - 14 Needs improvement	0 - 12 Not met
Purpose	Demonstrates understanding of assignment. Recognizes types of information necessary to support ideas.	Mainly demonstrates understanding of assignment. Mainly recognizes types of information necessary to support ideas.	Partly demonstrates understanding of assignment. Partly recognizes types of information necessary to support ideas.	Does not demonstrate understanding of assignment. Does not recognize types of information necessary to support ideas.

Figure 6.2: Presentation rubric - Purpose

Criteria	Rating			
	12 - 15 Meets Expectations	**10.5 - 12** Acceptable	**9 - 10.5** Needs improvement	**0 - 9** Not met
Organization	Sentences and paragraphs are clear and focused. Material organized in sensible order. Includes an introduction, body, and conclusion. Fulfills required length and format for the assignment.	Sentences and paragraphs are mainly clear and focused. Most material organized in sensible order. Mostly fulfills required length and format for the assignment.	Sentences and paragraphs are partly clear and focused. Partly fulfills required length and format for the assignment.	Sentences and paragraphs are not clear and focused. Does not fulfill required length and format for the assignment.

Figure 6.3: Presentation rubric - Organization

Criteria	Rating			
	12 - 15 Meets Expectations	**10.5 - 12** Acceptable	**9 - 10.5** Needs improvement	**0 - 9** Not met
Research	Recognizes and accesses information from scholarly sources. Includes research from multiple perspectives or sources. Correctly incorporates quotes from sources.	Mostly fulfills the criteria than the one mentioned on the left.	Partially fulfills the criteria than the one mentioned on the left.	Does not fulfill the criteria than the one mentioned on the left.

Figure 6.4: Presentation rubric - Research

Criteria	Rating			
	16 - 20 Meets Expectations	14 - 16 Acceptable	12 - 14 Needs improvement	0 - 12 Not met
Content analysis	Appropriately evaluates all researched material. Makes knowledgeable and thoughtful observations. Provides a cohesive discussion.	Mostly fulfills the criteria than the one mentioned on the left.	Partially fulfills the criteria than the one mentioned on the left.	Does not fulfill the criteria than the one mentioned on the left.

Figure 6.5: Presentation rubric - Content analyis

Criteria	Rating			
	8 - 10 Meets Expectations	7 - 8 Acceptable	6 - 7 Needs improvement	0 - 6 Not met
Language	Demonstrates proper use of vocabulary, spelling, grammar, punctuation. Demonstrates appropriate formality for assignment.	Mostly fulfills the criteria than the one mentioned on the left.	Partially fulfills the criteria than the one mentioned on the left.	Does not fulfill the criteria than the one mentioned on the left.

Figure 6.6: Presentation rubric - Vocabulary

Criteria	Rating			
	8 - 10 Meets Expectations	**7 - 8** Acceptable	**6 - 7** Needs improvement	**0 - 6** Not met
Presentation Structure	Included clear introduction, body, and conclusion. Added the citations and references where required.	Mostly fulfills the criteria than the one mentioned on the left.	Partially fulfills the criteria than the one mentioned on the left.	Does not fulfill the criteria than the one mentioned on the left.

Figure 6.7: Presentation rubric - Presentation structure

Criteria	Rating			
	6 - 10 Meets Expectations	**4 - 6** Acceptable	**0 - 4** Needs improvement	**0** Not met
Presentation Content	Clear description of topic, compelling arguments, and well defined conclusion. Team shared delivery well and with clarity.	Mostly fulfills the criteria than the one mentioned on the left.	Partially fulfills the criteria than the one mentioned on the left.	Does not fulfill the criteria than the one mentioned on the left.

Figure 6.8: Presentation rubric - Presentation content