16-450 A: Robotics Systems Engineering

Robotics Systems Engineering Ethics Mini-Module Lesson Plan

Developed by

Shaquetta A. Johnson & Mollie Wild

of behalf of

David Wettergreen

May 4, 2021

Table of Contents

Introduction

This course consists of three smaller ethics “mini-modules” to be integrated into the curriculum for Robotics and Systems Engineering (16-450) at Carnegie Mellon University. The mini-modules will be spread throughout the course, although they will be more present towards the end of the course due to time constraints. Mini-modules 1 & 3 will integrate with lectures 1,6,9 and lecture 20, respectively. Mini-module 2 is flexible and may be place wherever the professor sees fit. Students will apply ethics concepts learned in this module to their final projects for the course.

Mini-Module #1: *The Value of Diversity*

The overall goal of this mini-module is to demonstrate to students why a diverse workplace is important by showing concrete rewards that come from a well-rounded group as well as the pitfalls and poor design that result from a homogeneous workplace. Diversity is a key component in the creation of ethical technology, so it is important that all students feel personally invested in promoting diversity in their future work.

Learning Objectives

By the end of this mini-module, students should…

- be able to articulate the importance of diversity in both experience and background for engineers involved in systems engineering
- be able to provide concrete examples of how that diversity benefits system development

Slide Guide (45 minutes)

Lecture 1, Slide 25: YOU! (10 minutes)

In this slide, students introduce themselves to the class. The definition of experience is expanded from the base material to include “Experience using things that were not made with you in mind” in order to encourage students to share what they bring to the class beyond their technical expertise.

Lecture 6, Slide 30: Define Scenario (15 minutes)

This activity, creating scenarios for a given use case, is extended to include both a nominal scenario and an exceptional scenario. The goal is to get students to think outside the box and see how their different perspectives help them to find more use cases and create a more robust product.

Lecture 9, Slide 3: Review (20 minutes)

This activity is intended to encourage students to speak up about their ideas in a group setting and to value the perspectives of others. Students can expand on the memos they

wrote for homework and will defend their ideas or explain questions. This activity should also highlight the importance of listening to other students' feedback.

Post-Lecture Assignment

"Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?" (Holstein et. al, 2019) should be appended to the end of assignment 1. Students should write a 1-3 paragraph answer to the following discussion question: "The reading includes a section on the challenge of 'blind spots.' What blind spots do you recognize in yourself that may be relevant to your project? What blind spots do you expect your project group to have, if you are working with multiple people? Finally, what steps can you take to account for these blind spots?" Their write-ups will be worth 3 points of extra credit.

The final project requirements specifications should be moved back a week, to be due at 5pm the class period before Lecture 8. Meanwhile, the memo to another group about the requirements for their project should be moved back a week, to be due at 5pm the day before Lecture 9 so that it can be discussed in class that day.

Mini-Module #2: *Idealism vs Realism (80 minutes)*

The overall goal of this mini-module is to help students identify and adopt innovative, inclusive, and realistic acceptance criteria. The students will be given a walk-through of the philosophy that drives the mindset of setting expectations. They will then apply these concepts through two class activities exploring the balance of requirements and expectations for both a hypothetical scenario and their final project. Engineers play a critical role in the development of infrastructure that affects society, the environment, and the economy so it is imperative for students to understand the implications and consequences of deploying underdeveloped or marginalizing systems.

Learning Objectives

By the end of this mini-module, students should…

- be able to recognize the realist and idealist limitations of ethical intervention
- be able to develop priorities of functional and nonfunctional requirements

Pre-Lecture Reading

Ethics of Artificial Intelligence. (2015). Nature, 521(7553), 415-418. Link; https://search-proquest-com.cmu.idm.oclc.org/scholarly-journals/ethics-artificial-intelligence/docview/1684645373/se-2?accountid=9902 (available through CMU VPN)

Slide Guide (20 minutes)

Slide 2: Managing Performance Expectations

Students should understand the importance of knowing the differences between realistic and idealistic expectations. The goal of this slide is to get students to recognize who and/or what we plan for and when. These may be rhetorical or students may be asked to answer these directly.

Slide 3: Managing Performance Expectations

This slide is intended to have the class interact and share perspectives on what is most important in system design. It is also the first use of the term “marginalizing features.” This is intentional, as it will allow the professor to explain how certain features of system design may exclude certain users (e.g. accessibility, physical attributes, etc.) This will allow the students to reflect on the importance of diversity in engineering emphasized in Mini-Module 1.

Slide 4: Outlooks & Perspectives

This slide is not meant to be explored in depth until slide 8. However, it is shown to the students beforehand to guide their understanding of the context of the four terms to be

explained in the next 3 slides (realism, idealism, optimism, pessimism). A short explanation of the diagram as a spectrum of four quadrants will suffice at this point in the lecture.

Slide 5: Idealism vs Realism in Engineering

This slide is intended to give the philosophical background to why we think the way that we do. Plato's ideology of idealism is the embodiment of "life is what you make it." The key concept with this ideology is that invention is not a matter of if, but a matter of how. Aristotle developed realism based on the notion that "if everyone cannot be satisfied" or if "the non-functional requirements are unquantifiable" it is best to prioritize the quantifiable and tangible functional requirements.

Slide 6: Optimism

It is to be assumed that the students have in some way heard and understood the meaning of optimism. The goal is to showcase the benefits and consequences of optimism. Emphasize that optimism is not blind, and tit does not imply that all extraordinary ideas are to be taken at face value (e.g. timeline for projects, perfect functionality, etc.)

Slide 7: Pessimism

It is to be assumed that the students have in some way heard and understood the meaning of pessimism as well. The goal is to showcase the benefits and consequences of pessimism. The key concept here is that pessimism without critical thinking is detrimental to the success of a project. The students should understand that being skeptical is encouraged, as it allows uncertainty to be met with thorough questioning (e.g. challenging assumptions, providing cases where the uncertainty would become a problem).

Slide 8: Outlooks & Perspectives

These perspectives will be the core of the activities to follow later on. Now that the students are more familiar with the terminology, the professor should go into detail with the perspectives represented in each quadrant. The optimistic realist pursues ambitious goals with the intention of success, but acknowledgement of room for failure. The pessimistic realist is all or nothing; either 100% success/failure. The optimistic idealist adopts radical change and pursues high risk/high reward ventures. The pessimistic idealist allows for incremental change, but is not one to take risks.

Slide 9: Aspects to Consider

This slide presents the four broad categories of stakeholders/interests that are affected by engineering systems. The main takeaway in this slide is to understand that while these aspects are important, they may find themselves at a crossroads when they require system specifications that conflict.

Slide 10: Discussion (10 minutes)

The discussion is designed to allow students to think both objectively and subjectively about the four outlooks and perspectives. This will also give students insight on the composition of their final project team members. Encourage participation and emphasize that there is no wrong or right answer to this question. The questions will challenge students to assess their list of priorities with others'.

Slide 11: Outlooks & Perspectives Activity (20 minutes)

This activity is designed to challenge students to strategize on the deployment of an engineering system while taking into account both its potential and feasibility. This is done by using a common application of engineering technology to identify, assess, and set expectations through the lens of another perspective. Once the instructions for the assignment are understood, re-share slide 8 with the class so that the students may reference the characteristics of the four perspectives discussed earlier in the lecture. As the professor monitors the progress for each group, feel free to interact and question the students' assumptions to ensure alignment with their designated perspective.

Slide 12: Outlooks & Perspectives Final Project Activity

Now that the students have had a chance to acquire hands-on experience developing system requirements, they will apply this same analysis to their final projects. They may use the framework on the following slide to reflect on the process they used during the class activity. This activity is designed to ensure that the students comprehend the limits of potential while being encouraged to responsibly maximize their resources.

Slide 13: Ethical Analysis

This slide serves as a simple framework for conducting an ethical analysis on system requirements and expectations. Exclusionary metrics are features/requirements of design that can cause foreseeable and/or divisive ethical problems (e.g. unilingual language processing technologies, cameras being unable to recognize different skin tones, etc.). The students already understand functional and nonfunctional requirements, but this challenges them to categorize each into what is realistic, what is ideal, and if there is a solution that could bridge the gap.

Post-Lecture Discussion (15 minutes)

Students will be asked to evaluate the system requirements adopted for their final projects using the four steps given in the lecture. Considering the limitations, scope, timeline of the project, not all specifications are expected to be perfectly-inclusive. The goal here is to have the students identify the gaps in their designs and be able to foresee potential marginalizing features that would lead to ethical issues upon deployment.

Mini-Module #3: *Context & Reinforcing Intent (30 minutes)*

The goal of this mini-module is to reinforce that their technologies don't exist in a vacuum; they have consequences for the context in which they are deployed. Students need to understand this, and consequence web diagrams provide a way for them to better predict what these consequences might be. With this fuller understanding, students will have a better idea of whether or not they have satisfied their own ethical priorities.

Learning Objectives

By the end of this mini-module, students should…

- Understand the consequences of unethical behavior
- Contextualize their final project's place in society
- Enumerate ethical priorities and verify that they have been followed

Pre-Lecture Activity

Engineering Ethics: Crash Course Engineering #27 (10 minutes)

This YouTube video provides a comprehensive overview of the ethical values, duties, and consequences that are bestowed upon the engineering profession. This activity will prime students to think about ethics at the forefront of design and provides real-life examples of engineering disasters. Link: https://youtu.be/5KZx81crb48

Slide Guide (10 minutes)

Lecture 20, Slide 16: Aspects to Consider (Review)

This slide is a review of the four categories of [potentially] affected stakeholders/interests discussed in mini-module 2. Reiterate that while these aspects are important, they may find themselves at a crossroads when they require system specifications that conflict.

Lecture 20, Slide 17: Intent vs Execution: Drone Technology

The purpose of this slide is to explore the ramifications of transformational technologies beyond the scope of intent. Students are to observe that there may be uses of their technologies that may either leverage or exploit its capabilities. The students should think about how to set system requirements in place that will align use with the their intentions and expectations along with the "unintended good."

Lecture 20, Slide 18: Define Scenario

This slide should introduce students to the idea of a "consequence web diagram," a diagram that shows the way a deployed technology affects and is affected by the outside world.

We are using a 3D printer as an example because that's a well-known technology with some obvious flaws and benefits (plastic, but open source design). In case the class has trouble getting started, see Fig. 1 for an example. Another example can be found on aiandhumanity.org under the gallery tab. This diagram isn't supposed to be too complex; it's just meant to be a demonstration that students can then apply to their own project. The extra questions should get students thinking about how their technology relates to the rest of the world.

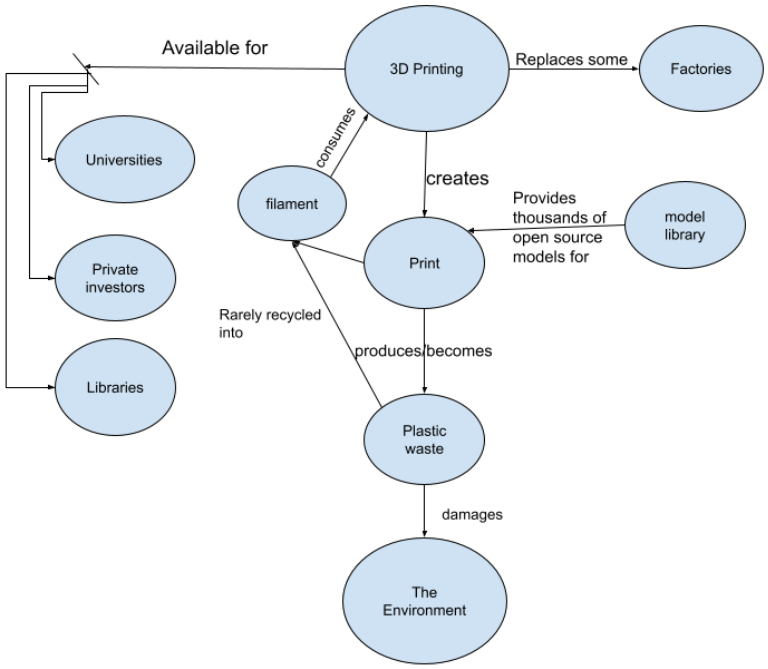

Figure 2: Possible Web Diagram Concepts

Post-Lecture Assignment

Now that students know what a consequence web diagram is, they should break into groups to create one for their own project. The goal is for students to better understand the possible ripples their project would create, should it go into production. The finished diagram can possibly be included in their final project documentation, depending how useful they find this activity.

Criteria	Points	Rating		
		Excellent	Good	Poor
Node count	2	At least 8 relationships and 6 nodes	Fewer than 6 nodes but at least 4	Fewer than 4 nodes
Nuance	3	Web contains detailed relationships between many different areas of society and provides a good global picture	Web contains a few areas, but a significant area relevant to the project is missing (ex. 3D printing graph does not include anything about environmental impact)	Web only investigates a narrow slice of society

Figure 3: Consequence Web Diagram Rubric

References

Cascio, Jamais. “Open the Future: What’s Your Future?” Http://Www.openthefuture.com, The Long View, 22 May 2006, www.openthefuture.com/2006/05/whats_your_future.html.

Cohen, LeoNora M. “Four General or World Philosophies.” Philosophical Perspectives in Education, Oregon State University School of Education, 1999, oregonstate.edu/instruct/ed 416/PP2.html.

Raman, Usha. “Idealism Inspires, Realism Works.” The Hindu, The Hindu, 18 Oct. 2016, www.thehindu.com/features/education/Idealism-inspires-realism-works/article14416673. ece.

Stanier, James. “Can Optimism Ever Be Bad?” The Engineering Manager, 1 June 2018, www.theengineeringmanager.com/growth/pessimism-optimism-idealism-and-realism/.

Want more fun? Challenge your students to the Robotics Systems Engineering Ethics Refresher Kahoot game!

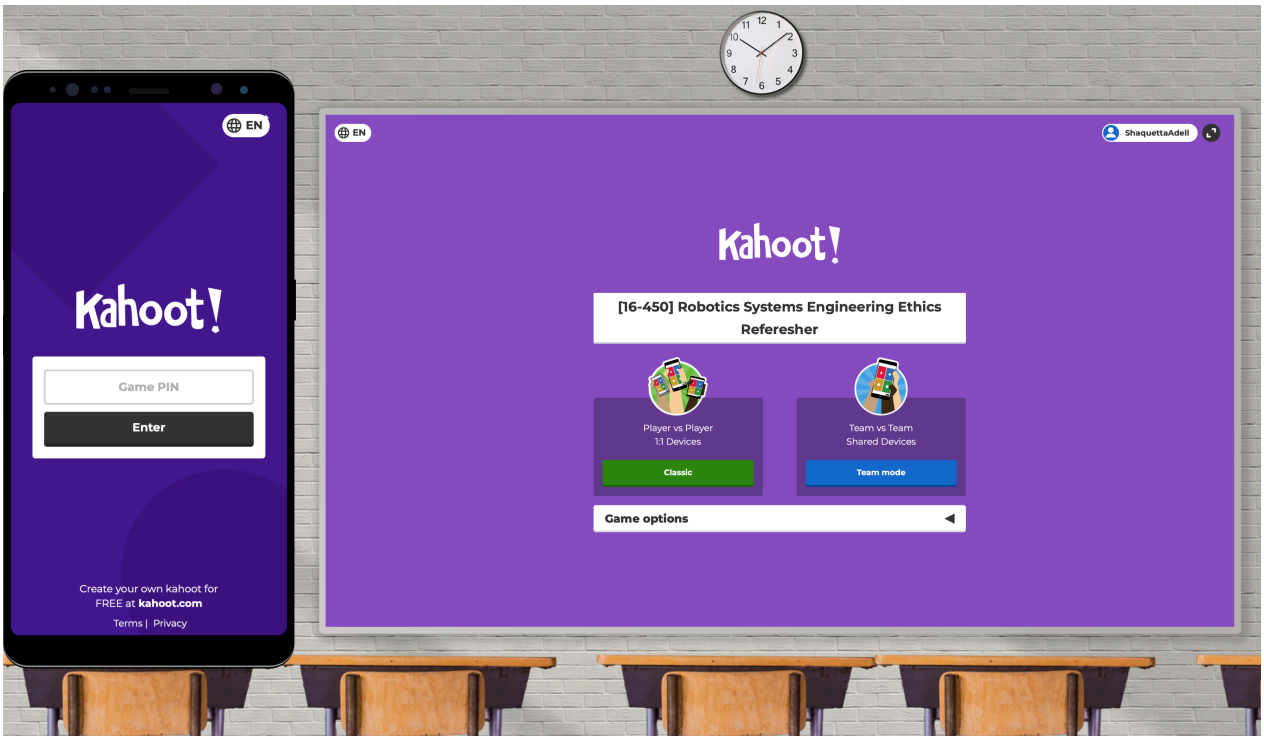

Figure 4: Robotics Systems Engineering Ethics Refresher Kahoot Game