

Embedded EthiCS @Harvard: S20 Repository Entries

Lyndal Grant

1. **Course.** CS189: Autonomous Robot Systems
2. **Course Level.** Upper-level undergraduate
3. **Course Description.** “Building autonomous robotic systems requires making **Robots that Observe, Reason, and Act**. How does a robot make sense of the world from raw and noisy sensor inputs? How does it control its actions reliably and recover from failures? When does it need to reason about the world and when can it just react? How does it balance short-term problems versus long-term goals? How does it operate in a world where others (human and robots) exist? And how do we program a robot to achieve all these things? The goal of creating a robot is the goal of creating **Embodied Artificial Intelligence**. In this class we will study methodologies for achieving embodied AI through a hands-on and ground up approach of programming your own.”¹ (Course description)
4. **Module Topic.** The Ethics of Lethal Autonomous Weapons
5. **Module Author.** Lyndal Grant
6. **Semesters Taught.** Spring 2019-2020
7. **Tags.** robotics (CS), automation (CS), war (phil), risk (phil)
8. **Module Overview.** This module addresses the question *Is it morally acceptable to deploy lethal autonomous weapons systems (LAWS) in combat?* We consider the ethical case for and against the use of LAWS, focusing on both whether the use of LAWS would be likely to lead to better outcomes (e.g. less loss of life) and whether it would violate ethical norms concerning how life and death decisions are made in combat. Students brainstorm ways in which introducing LAWS to the battlefield might lead to morally better and worse outcomes, drawing on course material about how autonomous robot systems operate. After identifying and evaluating the most promising outcome-based arguments for and against the use of LAWS, we consider a distinct ethical concern: whether entrusting life-and-death decisions to machines is consistent with appropriate respect for human dignity.
9. **Connection to Course Material.** During the course of the semester, students gain experience programming robots, starting with low-level control (e.g. motion, vision, feedback) and moving on to more advanced, higher-level reasoning (e.g. navigation/mapping). One potential application of these skills is to program LAWS. This module encourages students to think critically about whether LAWS should be built and deployed, and (if so) how much human control ought to be maintained. We also consider whether technologists should pledge not to contribute to the

¹ <https://canvas.harvard.edu/courses/37276>

development of LAWS (as some prominent critics of LAWS have argued) – a question that could be directly relevant to students’ future career options.

10. Module Goals.

1. Identify important moral arguments for and against the development and deployment of LAWS.
2. Distinguish consequentialist and non-consequentialist arguments for/against LAWS.
3. Distinguish different things it might mean for a weapons system to be “autonomous,” and discuss implications for the debate about the moral permissibility of deploying LAWS.
4. Understand the difference between legal and moral permissibility, and consider the ethical implications of LAWS’ unclear status under international law.

11. Key Philosophical Questions.

1. The legal permissibility of deploying LAWS in combat is murky (see “The Martens Clause”). What, if anything, follows about whether it is *morally* permissible to deploy LAWS?
2. Advocates of LAWS often claim deploying them in combat will lead to better outcomes, such as fewer casualties. What reasons do we have to accept this claim? Are there good reasons to be skeptical of it?
3. Critics of LAWS sometimes argue that deploying LAWS in combat is inconsistent with proper respect for human dignity. What do they mean by this, and are they right?

12. Key Philosophical Concepts.

- Legal vs. moral permissibility
- Consequentialist moral arguments
- Non-consequentialist moral arguments
- Human dignity

13. Assigned Readings.

https://www.youtube.com/watch?v=eSIW2ahGBE8&list=PLUFVjdQqDvCccJrSZqwyYwLATM38hH_wd&index=12&t=0s

Marginal note: In this video, Stuart Russell (a computer scientist and well-known critic of LAWS) and Paul Scharre (a defense analyst) discuss whether concerns about “killer robots” being deployed in the near future are well-placed. Russell argues that the technology to create LAWS is already available, and that we should be doing whatever we can to prevent their deployment. Sharre demurs, arguing that the scenarios Russell identifies remain unrealistic and that the concerns he raises are little more than fearmongering. We assigned the video because it provides an engaging way to familiarize students with some of the most important arguments for and against the use of LAWS (without requiring extensive background reading).

14. Class Agenda.

1. Introduction: autonomy and intelligence in robot systems.
2. Legal vs. moral permissibility; the Martens Clause.
3. The outcome-based case for LAWS.
4. The outcome-based case against LAWS.
5. The case against LAWS based on human dignity.
6. The ethical responsibilities of technologists.

15. Sample Class Activity. In the first part of the class session, students are introduced to some of the basic capabilities of currently existing LAWS, as well as ways technologists foresee LAWS being developed and deployed in the near future. The class is then broken up into small groups, and students brainstorm potential risks and benefits of using LAWS in combat with their peers before reporting back to the full class. One student reports back from each group.

Marginal note: Students were highly engaged in this activity, in large part because it gave them an opportunity to work together with their peers to apply knowledge they had gained from the class (as well as other computer science coursework) to the topic under discussion. The discussion focused on an empirical question, rather than an ethical one – what kinds of results should we expect if LAWS were deployed on the battlefield, given the current state of the relevant technologies? This worked well for two reasons: it helped students more familiar with technical material jump right in, and it laid the groundwork for a rich discussion of the ethical significance of different kinds of answers to that empirical question.

16. Lessons Learned.

- The topic of this module really resonated with students, who were immediately enthusiastic about participating in the discussion and had many insightful things to contribute. We suspect this is partly due to the fact that the topic is intrinsically interesting and partly due to the fact that the discussion connected strongly with prior course material.
- This was one of our first modules conducted over Zoom, and we were pleased to find that the format worked just as well as in-person module delivery. We used the “breakout rooms” function to conduct small group discussions, which allows the instructor to “drop in” on breakout rooms during the activity. This is a good substitute for walking around the classroom during small group activities, and helps the instructor get a sense of what students are thinking about in preparation for the full-class debrief.