



JOHNS HOPKINS

WHITING SCHOOL  
*of* ENGINEERING

# Autonomous weapons and course recap

12/5

# Course Recap

---

1. What are 1-3 things from this class that you will try to apply in the future? Can you give a specific example of how you might apply them?
2. What do you think should be top priorities for AI Ethics researchers? What about AI practitioners, educators, or policy makers?

# Final Project Report

---

- Due Friday Dec 15 (email to instructors by midnight)
- ACM template
  - Double column
  - 6-8 pages
  - [We didn't take off points for format for the proposals but we will for the final report]
- Most reports will take the format of a standard research paper (introduction, methods, results, discussion, related work)
- It's ok to copy content from your previous submissions
- Don't forget authorship statement explaining how each individual contributed to the project

# Current Autonomous Weapons



**Predator**  
remote-controlled UAV used by the US and allies to strike targets worldwide



**Phalanx CWIZ**  
Ship-mounted Defense system



**Land mine**  
Deadly, dumb autonomous weapon, most countries have limited their use



**Korean SGR-A1**  
Sentry bot guarding The DMZ

The Samsung SGR-A1 presumes any person entering the [DMZ](#) is an enemy and, upon detection, will attempt to identify the target through voice recognition... the system can choose between sounding an alarm, firing rubber bullets or engaging the target with other weapons.

# Two Perspectives: Peter Asaro

- *(Definition) autonomous weapon systems (are) any automated system that can initiate lethal force without the specific, conscious, and deliberate decision of a human operator, controller, or supervisor.*
- *The **empirical question** is whether a computer, machine, or automated process could make each of these decisions of life and death and achieve some performance that is deemed acceptable.*
  - Vulnerability to hacking
  - Lowering threshold of going to war
- *But the **moral question** is whether a computer, machine or automated process ought to make these decisions of life and death at all.*
  - Threats to human rights and dignity

# Two Perspectives: Ron Arkin

---

- *Robots not only can be better than soldiers in conducting warfare in certain circumstances, but they also can be more humane in the battlefield than humans.*
- *It is my contention that robots can be built that do not exhibit fear, anger, frustration, or revenge, and that ultimately (and the key word here is ultimately) behave in a more humane manner than even human beings in these harsh circumstances and severe duress.*
- *People have not evolved to function in these conditions, but robots can be engineered to function well in them.*

# Discussion Questions

---

- To what extent are computer scientists motivated to see the best in their field?
- The argument that *AI can be more just than humans* echoes similar arguments around reducing bias (e.g. in criminal justice system)
  - What have practical examples suggested about these types of arguments?
  - Is the problem imperfect models or is it impossible to “correct” an unjust system with AI?
- Are there some tasks that fundamentally should not be automated regardless of model performance and impact?

# Course Recap

---

1. What are 1-3 things from this class that you will try to apply in the future? Can you give a specific example of how you might apply them?
2. What do you think should be top priorities for AI Ethics researchers? What about AI practitioners, educators, or policy makers?