Many of the presuppositions I will use in this analysis are derived from Nick Bostrom and his article “The Ethics of Artificial Intelligence”. The purpose of this analysis is to determine whether the experiments below have demonstrated if we are capable of generating an appropriate ethical framework for Artificial Intelligence. The types of Artificial Intelligence which will be analyzed in this poster are autonomous vehicles and autonomous weapons. Prior to evaluating the experiments themselves it is necessary for us to have a concrete understanding of Bostrom’s ethical requirements for Artificial Intelligence. Bostrom cites four necessary characteristics for an AI “algorithm intended to replace human judgement” (Bostrom, 3). These characteristics are: transparency, predictability, incorruptibility and responsibility.

Prior to making any judgements about whether these characteristics are realized in the following experiments, it is vital to understand how Bostrom’s framework functions in the real world. While it is easy to comprehend how each attribute is important to the foundation of an ethical theory, it is more complex to interpret the moral validity of these characteristics in practice.

The fundamental question which arises in regards to the practice of these ethical characteristics, then, is how engineers and programmers implement these attributes into the AI. A transparent AI will typically be based on decision trees or Bayesian networks as opposed to an intricate neural network or genetic algorithm. In order to make an AI predictable, programmers have to create an algorithm which teaches the AI to follow past precedent. Bostrom would ideally like to see the predictability of an AI mirror that of the legal system. This would “provide a predictable environment within which citizens can optimize their own lives” (Bostrom, 2). For incorruptibility to be achieved an AI’s built-in algorithm must be immune to coercion. We have already experienced the implementation of these types of machines in airports and stadiums. These machines know “for example, a shape that, placed next to a pistol in one’s luggage, would neutralize recognition of it.” (Bostrom, 2)

The next attribute, responsibility, is quite possibly the most ethically intertwined algorithmic trait an AI would be composed of. It is also the attribute most prevalent in our discussion of the following experiments. For an AI to be wholly responsible for its actions the trail leading back to its original designer would have to be affectively removed. This, of course, would require an AI to be created and programmed without the direct or indirect influence of another individual. If another being is even remotely associated with the actions of said AI, the burden of a mistake from the machine would rest in that beings hands. A deeper analysis of responsibility and AI will be discussed in the preceding sections.

The Moral Machine Experiment, conducted by researchers at MIT in 2017, is the perfect study to learn if we as a community have appropriate ethical standards for Artificial Intelligence. The researchers put the responsibility of the autonomous vehicle’s decisions solely in the participant’s hands. By doing this, not only were the researchers gathering the most sincere responses from the participants but it also allows us to analyze the moral decisions directly from the respondents themselves. The experiment also does a good job in making the reader aware of the general background of machine ethics and where we stand today. The authors stress the realization that “the ethical principles that will guide autonomous vehicles cannot be left solely to either the engineers or the ethicists” (Awad, et al). Or in other words, citizens will be held responsible for the actions of autonomous vehicles too. The underlying question in this experiment, then, is how people differ in regards to the action taken by the autonomous vehicle across different regions of the world. The authors make note of this underlying question as well when they claim that “we need a fine-grained understanding of how different individuals and countries may differ in their ethical preferences” (Awad, et al). The researches focused on nine different choices, five choices focused on certain characteristics like human vs animal, male vs female, young vs old, physical appearance (how fit you are), and social status. Four choices were geared towards more mechanical options like staying on course vs swerving, sparing passengers vs pedestrians, sparing pedestrians crossing legally vs illegally and sparing more lives vs fewer lives. In total, the researchers were confident that through a four-stage analysis of the responses to these choices that a significant amount of ethical data would be gathered. In the second half of this analysis we will be able to interpret the actual findings of the experiment and in turn begin to see where the ethical standard is at a global scale.

The results of The Moral Machine Experiment showed positive signs for policy makes and ethicists to build upon when considering the choices autonomous vehicles will (hopefully not) have to make. The data, gathered from 40 million different decisions in 233 different countries, showed “three strong preferences” (Awad, et al) shared by the respondents. The universal preference for sparing human lives, sparing more lives and sparing young lives “should be used by industry and government as a foundation for understanding how the public would react to the ethics of different design and policy decisions” (Hao, 2018). While these findings are extremely significant, the ethical discussions we are most in need of today should be focused on what to do about the choices which aren’t universally shared. One of the problems the study identified was that in different regions of the world, a choice is universally shared but only within that specific region is it shared. The researchers titled these groups: Christian Cultural Group, Confucianist Cultural Group and The Southern Cluster Group. This finding “suggests that geographical and cultural proximity may allow groups of territories to converge on shared preferences for machine ethics” (Awad, et al). I would argue that the cultural influence of this finding has more significance than the geographical one. Considering this, it will be quite complicated for policy makers to have universal machine ethics that keep everyone happy even in regions where cultural values are shared. A scarier problem the experiment uncovered was that a document titled German Ethics Commission on Automated and Connected Driving, “the first and only attempt so far to provide official guidelines for the ethical choices of autonomous vehicles” (Awad, et al), disagrees with the experiment’s universal principle of sparing more lives. The researchers stress that more work should be done to gain a better understanding of universal ethical beliefs. In order for this to happen, experiments which align machine ethics with human values should continue to be conducted.

This section will function more as a discussion as opposed to an analyzation of a specific experiment like above. The discussion surrounding autonomous weapons has been happening since the start of the Bush administration but recently has hit a massive upscale due to the frequent use of drones by the Obama administration. In this discussion, we will focus on three articles which cover the majority of the ethical issues related to autonomous weapons. The first article, written by an adherent for the use of drones in a military setting, weighs the pros and cons of Washington’s use of drones in the Middle East. There are two sides to the discussion of drones in the Middle East. Those in favor claim that drones have killed many key leaders of al Qaeda as well as many (equally dangerous) lower-level terrorists. In his article, “Why Drones Work: The Case for Washington’s Weapon of Choice”, Daniel Byman cites that drones have already killed around 3,300 terrorists. While this number is massive, it remains to be seen exactly how many civilians were killed in these same attacks. The article presents numbers ranging from 0 to 500 killed civilians. Not only is the possibility of 500 civilians being killed by drones a scary thought but the vagueness surrounding the exact number is even more troubling. I would argue that prior to any more drone interference in the Middle East a concrete method for identifying every individual killed in a drone attack is necessary. For such a wide-scale, destructive force the ethical standards surrounding the entire process from hours before the attack to after the attack must be as thorough as the technology itself. Byman agrees, “Washington needs to set forth a clear policy now on extrajudicial and extraterritorial killings of terrorists — and stick to it” (Byman, 2013). While it is important for the U.S to take the first step in cleaning the ethical issues surrounding drones oversees, it is imperative for other countries to do the same. It will not be long until countries driven by generating revue begin to mass produce these types of weapons for unqualified civilians. The U.S has put the burden on themselves for beginning the development of ethical laws with respect to autonomous weapons.

Regardless of if you are pro drone or anti drone, we all must agree that improvements and adjustments to our current ethical standard is a necessity. I would be interested to see what data an experiment would gather if it was conducted like The Moral Machine Experiment but with respondents choosing the choices for autonomous weapons. I would not be surprised if there were different universal preferences even though the ending is the same in both scenarios. Weapons are a much less common tool for the average person than autonomous vehicles are. Given the unfamiliarity so many people have with weapons it will be interesting if public input will be considered in as much significance as it will be regarding the choices of autonomous vehicles. Relating back to what Bostrom taught us about necessary characteristics AI machines must have built in to them, it is extremely concerning to see none of the authors in both The Moral Machine Experiment and the articles about autonomous weapons speak about autonomous machines possessing any of these built-in attributes. Rather, they are more focused on how efficient these technologies are at doing their sole function. It seems as though the implications for the U.S’s use of drones as not totally gone into effect yet. I would not be surprised if, as more information regarding civilian deaths in drone attacks is revealed, the government receives stronger backlash from it. My suggestion, speaking on behalf of the U.S, would be to quit while we’re ahead. We should appreciate the amount of bad guys the autonomous weapons have killed for us already but I do not see a long-term future for this type of technology without any major world crisis. While I think the technology is extremely interesting, I believe drones should be more concerned with helming civilians in natural disasters and relief efforts. I believe in this sense, drones have a long-term future for providing care in large-scale disasters. Hopefully with more time and knowledge on the ethics and technology of AI we will find the best solution for what to do with drones. In the immediate future I would like to see a decrease in the use of autonomous weapons and a stronger focus on the good AI can produce.

The Moral Machine Experiment as well as the articles we discussed surrounding autonomous weapons reveal how our current ethical standards for AI are not suitable for a full rollover of autonomous vehicles and weapons. Most of the authors agree that what we have now is a relatively divided conception of how these technologies should function. While the experiment reveals how there are rooted universal preferences worldwide, the articles show how there are many more contrasting beliefs than common ones. A greater understanding of human values and universal machine ethics will be required for this technology to be implemented worldwide.

1. Should policy makers consider having autonomous vehicles make different decisions on a choice based on what region of the world the autonomous vehicles are in?
2. How can studies like The Moral Machine Experiment become more random while also reach out to more willing respondents? The importance of this question stems from the growing concern policy makers will have about responding to the critical backlash of their choices if a fatal accident were to occur.
3. Should Washington we weary of drone technology falling into the wrong hands? What necessary steps are they taking to ensure this will not happen?
4. If you were an ethicist: would you consider the military killing innocent civilians in pursuit of a terrorist group as moral or immoral? How many civilians would need to die before you considered it immoral?
5. Are drones a short-term or long-term solution for combating terrorists groups in the Middle East? If short-term, what will come next?

Bostrom, Nick, and Eliezer Yudkowsky. “The Ethics of Artificial Intelligence.” *The Cambridge Handbook of Artificial Intelligence*, pp. 316–334

Awad, Edmond, et al. “The Moral Machine Experiment.” *Nature*, vol. 563, no. 7729, 2018, pp. 59–64

Byman, Daniel L. “Why Drones Work: The Case for Washington’s Weapon of Choice.” *Brookings.edu*, The Brookings Institution, 28 July 2016

Hao, Karen. “Should a Self-Driving Car Kill the Baby or the Grandma? Depends on Where You're From.” *Medium.com*, Medium, 29 Oct. 2018