

VITTUC MODEL UNITED NATIONS

CCAI'31

The Enforcement of Common Laws for the use Artificial Intelligence

Satyajit Giri & Meghana Kumar

Table of Contents

- I. Introduction
 - a. Artificial Intelligence(AI)
 - i. Hard AI
 - ii. Soft AI
 - iii. The in between camp
 - b. Lethal Autonomous Weapons
 - c. Domestic Use of AI based systems
 - d. Ethics of Artificial Intelligence
 - i. Robot Rights
 - ii. Threat to Privacy
 - iii. Threat to Human Dignity
 - iv. Machine Ethics
- II. The Committee
 - a. The timeline
 - b. A word from the executive board
- III. A Sample Resolution

I. BACKGROUND THEORY

Artificial Intelligence

Artificial intelligence (AI) is a sub-field of computer science. Its goal is to enable the development of computers that are able to do things normally done by people -- in particular, things associated with people acting intelligently. These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction.

Stanford researcher John McCarthy coined the term in 1956 during what is now called The Dartmouth Conference, where the core mission of the AI field was defined. The following was the mandate of the Dartmouth Conference:

“The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.”

If we start with this definition, any program can be considered AI if it does something that we would normally think of as intelligent in humans. How the program does it is not the issue, just that it is able to do it at all. That is, it is AI if it is smart, but it doesn't have to be smart like us. Based on the goals of an AI developer, we can divide AI into three camps: Strong AI, Weak AI and a third camp that lies between the previous two.

Strong AI

The work aimed at genuinely simulating human reasoning tends to be called “strong AI”. However, strong AI remains controversial. The reputation of this area of research has been damaged over the years by exaggerated claims of success that have appeared both in the popular media and in the professional journals. At the present time, even an embodied system displaying the overall intelligence of a cockroach is proving elusive, let alone a system rivalling a human being.

The difficulty of “scaling up” AI's so far relatively modest achievements cannot be overstated. Five decades of research in symbolic AI has failed to produce any firm evidence that a symbol-system can manifest human levels of general intelligence. Critics of nouvelle AI regard as mystical the view that high-level behaviours involving language-understanding, planning, and reasoning will somehow “emerge” from the interaction of basic behaviours like obstacle avoidance, gaze control and object manipulation. Connectionists have been unable to construct working models of the nervous systems of even the simplest living things. *Caenorhabditis elegans*, a much-studied worm, has approximately 300 neurons, whose pattern of interconnections is perfectly known. Yet connectionist models have failed to mimic the worm's simple nervous system. The “neurons” of connectionist theory are gross oversimplifications of the real thing.

However, this lack of substantial progress may simply be testimony to the difficulty of strong AI, not to its impossibility.

Weak AI

Weak AI systems, while behave like human beings, there behavior in no way helps us understand about how humans think. One of the prime examples of this is IBM's Deep Blue, a system that was a master chess player, but certainly did not play in the same way that humans do. Therefore while Strong AI systems not only take humanlike decisions, but also the process of coming to a human decision is exactly the same as a human. Weak AI systems however make decisions on the basis of algorithms, mathematical modeling and other methods that are not necessarily the same way a human would behave.

A combination

Somewhere in the middle of strong and weak AI is a third camp (the "in-between"): systems that are informed or inspired by human reasoning. This tends to be where most of the more powerful work is happening today. These systems use human reasoning as a guide, but they are not driven by the goal to perfectly model it.

A good example of this is IBM Watson. Watson builds up evidence for the answers it finds by looking at thousands of pieces of text that give it a level of confidence in its conclusion. It combines the ability to recognize patterns in text with the very different ability to weigh the evidence that matching those patterns provides. Its development was guided by the observation that people are able to come to conclusions without having hard and fast rules and can, instead, build up collections of evidence. Just like people, Watson is able to notice patterns in text that provide a little bit of evidence and then add all that evidence up to get to an answer.

Likewise, Google's work in Deep Learning has a similar feel in that it is inspired by the actual structure of the brain. Informed by the behavior of neurons, Deep Learning systems function by learning layers of representations for tasks such as image and speech recognition. Not exactly like the brain, but inspired by it.

The important conclusion is that AI systems do not have to mimic a human, just need to make smart decisions. However, this in itself is a weak definition:

Issue One: How do we define Artificial Intelligence?

This is of particular importance, because any law governing AI will be irrelevant if member states can simply argue systems do not fulfill their own country's definition for AI. A relevant definition can also positively shape the discourse and direction in which further AI technology is developed.

Lethal Autonomous Weapons (LAWS)

Perhaps the biggest contentious issue, with regards to AI, is their application in international conflict regions. In scientific literature and official government documents, there a number of approaches to defining autonomous weapons systems. At present there is no universally accepted definition. Common to all different approaches, however, is that the level of capability with regard to decision-making by means of algorithms alone, without human intervention. The level of autonomy that military research is striving to achieve in the long term indicates towards providing higher degrees of “autonomous choices” to the weapons systems, most importantly the decision to take “critical decisions”. Such advancements aim to take humans out of the decision making loop, as much as they can. In contrast to autonomous systems, automated systems merely automatically execute previously programmed commands in a predictable fashion. They are unable to react independently to unforeseen events. Such systems are typically deployed to carry out an entirely determinate, pre-established task (for example defense against anti-ship missiles), in a narrowly defined area of operations (for example in a radius of 5 nautical miles around the defending ship). A mine also falls-as the simplest possible example-of this terminology of automated weapons systems. However, it is also to be understood that the more complex the field of functions and the larger or more differentiated the area of deployment becomes, the more the distinction between autonomous and automated systems becomes blurred. To cite a standpoint over this, the ICRC has largely abandoned any distinction in between in the two. In the 69th session of the General Assembly, First Committee, the statement by the ICRC laid down the viewpoint that the distinction between the two was not of primary importance. What is important is whether critical functions-that is, independent decision making about life and death-are entrusted into a system. This approach also makes it clear that every autonomous system may not be problematic. For example, an autonomously operating submarine used to search for mines, or an autonomous offensive system deployed exclusively in outer space to neutralize hostile military satellites does not raise the same kind of urgent ethical and legal issues as a combat robot designed for urban warfare that directly encounters human beings, both combatants/fighters and civilians.

A brief summary of the advantages and disadvantages are as follows:

Advantages:

1. LAWS save soldier lives. They can perform high risk operations without any risk to their own soldiers. It is also argued that efficient LAWS can ensure that less civilian casualties occur by quickly targeting only armed combatants.
2. LAWS are more humane than humans. They're actions aren't governed by fear, rage, etc. Take for example a scenario where a child picks up a gun. A human soldier can justify shooting the child, but a LAWS would be programmed to never shoot the child. Their every actions can also be digitally tracked leading to greater accountability in combat regions.
3. An increasing concern in warfare, is the viewpoint that given the pace of warfare, humans have become the weakest link in the military arsenal and thus by providing increasing autonomy to the machines will be considerably negating this weak link. These systems

can be continued to work even when, for example, communication lines have broken down or behind enemy lines.

Disadvantages:

1. Humane characteristics: The lack of emotion and empathy of humans in robots might be a very bad thing. When a government fails in its duty to its citizens, tries to push policies that are against humanity, often they resort to the armed forces for the enforcement of these laws. There have been many cases where the army has refused to cooperate, mostly because the soldiers themselves refuse to perpetrate inhumane acts on fellow humans. The best example of this is how the Egyptian army refused to attack the demonstrators in its Twitter Revolution.
2. Another concern dealing with their use might be the ability of robots to distinguish between legal and illegal orders. If this distinction is not established with clarity, it shall blur the whole concept of accountability over the use of such systems, further complicating matters.
3. It is also an important concern put forward when dealing with LAWS; of the distinction between civilian and combatants. This has been a fundamental principle in International Law, violation of which can have grave consequences. By addressing this concern, a large part of the debate over their legality of use could easily be addressed.
4. In addition to the above mentioned issues, other legal and humanitarian issues persist such as the ability of the robots to obey by the principles of “distinction”, “proportionality”, prevent violations of the right to life; some of which shall be discussed in the subsequent sections.

Issue Two: LAWS

Are LAWS be legal and if yes, to which degree of autonomy can there use be legally monitored by current international law?

Who will be accountable in cases of LAWS violating International Humanitarian Law?

How can we ensure civilians are protected from LAWS?

To what extent can “critical decisions” concerning legal interests such as right to life be delegated completely to autonomous weapons systems?

Can LAWS be programmed to abide by the principles of international law such as respecting the principle of proportionality, principle of distinction etc.

Domestic Use of Artificial Intelligence

So if hard AI is difficult to code, how exactly does AI manifest itself in the domestic sphere? Not in the form of humanoid like cyborgs (like Terminator) but rather in the manner of intelligent computer systems that help us process large amounts of data at incredibly short durations to make critical decisions in real life (Jarvis from Ironman). An excellent example of the application of AI in real life is the autonomous cars that are currently being developed, most prominently by Google. Recent surveys attribute almost nine in ten accidents in USA to human error. And while even Google cars have been recorded to make mistakes and crash, the question is: are they safer than the average human driver.

Even if the technology can operate at these high levels of efficiency and safety, there exist several interesting moral dilemmas. The first is commonly known as the trolley problem. Imagine two cars skid on a narrow icy road: one car has five people in it, the other just one. Now, if the only way to avoid a head on collision is for one of the two cars to swerve off the road, do we effectively have to program the car with only one passenger to kill him? What if that one guy is the President of your country? Do we still accept the utilitarian way of thinking?

The autonomous car and LAWS are great examples because they highlight repeating issues that we see in every AI technology. What happens if AI technology behaves abnormally, like a Google car swerves randomly and kills a person? Is Google accountable? If so, who in Google, considering no one in Google actively wanted to kill a person? The mistake in the code obviously lies with the coder, but usually a coding happens in groups, over time, with many parts of the code going through many different coders at different times. If no one is accountable does that mean the victim has no way of getting remuneration/retribution?

The rise of AI technology also behooves the question of security. As in the trolley case, the worry is a lot of people will hack their own cars and reprogram it in such a way that the car can never kill the user, but rather will always kill the other family. There is also the worry of external hacking. Imagine if a terrorist infiltrates the communication system of an AI system and forces all cars in the city to crash. Every AI system requires knowledge about its surroundings, a way to communicate and consequently have mechanisms that are susceptible to corruption.

There is also the fear of AI affects the economy. If AI replaces mundane jobs, like say a waiter at Pizza Hut, it also means one less job exists in the economy. And I haven't even touched on the Skynet Doomsday/Age of Ultron scenario. How do we protect ourselves from such a large plethora of issues?

The key here is to understand that the concerns that we have for one type of AI system is usually applicable to other kinds. Consequently we have to pick these universal issues and come to a consensus on our stance with them. Equally important is to understand exceptional circumstances and how to have a mechanism to deal with these issues too.

The Ethics of Artificial Intelligence

The ethics of laws intended for AI based systems involves two aspects:

- a. Governing how the AI systems interact with humans to maximize their benefits without compromising on human rights and dignity.
- b. Governing how society interacts with robots, and how do we protect AI systems.

The term "roboethics" was coined by roboticist Gianmarco Veruggio in 2002, referring to the morality of how humans design, construct, use and treat robots and other artificially intelligent beings. It considers both how artificially intelligent beings may be used to harm humans and how they may be used to benefit humans.

Robot rights

Robot rights are the moral obligations of society towards its machines, similar to human rights or animal rights. These may include the right to life and liberty, freedom of thought and expression and equality before the law. A collection of over 240 papers by the United Kingdom's Office of Science and Innovation have argued that if AI reach a level where they can even reproduce, they might demand basic rights like robohealthcare, robohousing, etc. Experts disagree whether specific and detailed laws will be required soon or safely in the distant future. Glenn McGee reports that sufficiently humanoid robots may appear by 2020. Ray Kurzweil sets the date at 2029. Another group of scientists meeting in 2007 supposed that at least 50 years had to pass before any sufficiently advanced system would exist. An interesting question then, is the characteristics an AI system must display to become deserving of rights. An interesting competition called Loebner Prize and the test it employs known as the Turing Test might be an interesting evaluation mechanism.

Threat to privacy

If an AI program exists that can understand natural languages and speech (e.g. English), then, with adequate processing power it could theoretically listen to every phone conversation and read every email in the world, understand them and report back to the program's operators exactly what is said and exactly who is saying it. An AI program like this could allow governments or other entities to efficiently suppress dissent and attack their enemies. It could also help governments keep their countries safe from things like terrorist attacks. Keep in mind this AI system could be a system that selects the target from potential candidates and takes appropriate action or it can be an AI system that simply informs a human use. Once again we see how the degree of autonomy becomes an important parameter.

Threat to human dignity

Joseph Weizenbaum argued in 1976 that AI technology should not be used to replace people in positions that require respect and care, such as any of these:

- A customer service representative

- A therapist
- A nursemaid for the elderly
- A soldier
- A judge
- A police officer

Weizenbaum explains that we require authentic feelings of empathy from people in these positions. If machines replace them, we will find ourselves alienated, devalued and frustrated. Artificial intelligence, if used in this way, represents a threat to human dignity. Weizenbaum argues that the fact that we are entertaining the possibility of machines in these positions suggests that we have experienced an "atrophy of the human spirit that comes from thinking of ourselves as computers."

The counter argument is sometimes human emotions, bias and prejudice is severely unwanted. A minority or a woman would probably want to interact with a AI system rather than a lecherous or racist man. To achieve self actualization or the pinnacle of human dignity, a human must be free of mundane life struggles and in that way AI systems are a necessity.

Machine ethics

Machine ethics (or machine morality) is the field of research concerned with designing Artificial Moral Agents (AMAs), robots or artificially intelligent computers that behave morally or as though moral.

Isaac Asimov considered the issue in the 1950s in his *I, Robot*. At the insistence of his editor John W. Campbell Jr., he proposed the Three Laws of Robotics to govern artificially intelligent systems. Much of his work was then spent testing the boundaries of his three laws to see where they would break down, or where they would create paradoxical or unanticipated behavior. His work suggests that no set of fixed laws can sufficiently anticipate all possible circumstances.

In 2009, during an experiment at the Laboratory of Intelligent Systems in the Ecole Polytechnique Fédérale of Lausanne in Switzerland, robots that were programmed to cooperate with each other (in searching out a beneficial resource and avoiding a poisonous one) eventually learned to lie to each other in an attempt to hoard the beneficial resource. One problem in this case may have been that the goals were "terminal" (i.e. in contrast, ultimate human motives typically have a quality of requiring never-ending learning).

Some experts and academics have questioned the use of robots for military combat, especially when such robots are given some degree of autonomous functions. The US Navy has funded a report which indicates that as military robots become more complex, there should be greater attention to implications of their ability to make autonomous decisions. The President of the Association for the Advancement of Artificial Intelligence has commissioned a study to look at this issue. They point to programs like the Language Acquisition Device which can emulate human interaction.

In 2009, academics and technical experts attended a conference to discuss the potential impact of robots and computers and the impact of the hypothetical possibility that they could become self-sufficient and able to make their own decisions. They discussed the possibility and the extent to which computers and robots might be able to acquire any level of autonomy, and to what degree they could use such abilities to possibly pose any threat or hazard. They noted that some machines have acquired various forms of semi-autonomy, including being able to find power sources on their own and being able to independently choose targets to attack with weapons. They also noted that some computer viruses can evade elimination and have achieved "cockroach intelligence." They noted that self-awareness as depicted in science-fiction is probably unlikely, but that there were other potential hazards and pitfalls.

In *Moral Machines: Teaching Robots Right from Wrong*, Wendell Wallach and Colin Allen conclude that attempts to teach robots right from wrong will likely advance understanding of human ethics by motivating humans to address gaps in modern normative theory and by providing a platform for experimental investigation. As one example, it has introduced normative ethicists to the controversial issue of which specific learning algorithms to use in machines. Nick Bostrom and Eliezer Yudkowsky have argued for decision trees (such as ID3) over neural networks and genetic algorithms on the grounds that decision trees obey modern social norms of transparency and predictability (e.g. *stare decisis*), while Chris Santos-Lang argued in the opposite direction on the grounds that the norms of any age must be allowed to change and that natural failure to fully satisfy these particular norms has been essential in making humans less vulnerable to criminal "hackers".

II. THE COMMITTEE

Timeline:

2016: Donald Trump becomes the 58th President of the United States of America.

2017: President Trump and a predominantly Republican Congress vote to institute a Muslim Register in which every Muslim must submit his name to live in American territory.

2017: Russia and USA collaborate in their attack against ISIS. ISIS is pushed out of Iraq to only Syria.

2018: Israel's Iron Drome System defends itself against rocket missiles from the West Bank, but collateral damage kills 24 Israeli citizens. Putin declared lifetime President of Russia.

2019: Many Arab League countries come together to protest Trump Anti-Muslim policies. Iran makes landmark Nuclear Deal with China and India. A Google car hits a schoolbus, five dead, Google fined \$20 million dollars, but vows to continue sales of Google Cars. ISIS shooters kill 43 in a mall in USA.

2020: Donald Trump wins Elections again on wave of anti Islamic sentiments. First AI driven doctor delivers a baby. Trump promotes the use of LAWS in war against ISIS.

2021: LAWS destroy ISIS with only one major setback: a single drone identifies an entire town as a threat and kills over a 100 civilians in Syria. Arab League moves to put oils embargo on the USA.

2022: Chinese company Baidu reveals its first batch of autonomous car series Gixao. It has been approved by both American and Chinese highway requirements. Google, BMW and Honda expected to follow soon. The latest music craze is a new AI driven music software called Sync which monitors a user's or the room's mood to generate the most suitable music. USA strongly moves to greener technology on the road.

2024: EU reports that a combination of the Google X and Baidu Gixao has led to a 40% market capture and a 70% reduction in accidents. Autonomous ships are now being used for movement of goods, fishing etc.

2027: Google's stocks crash after photos of Google cars being taken apart and assembled for LAWS by the Taliban emerge. Israel accepts having used Fully Autonomous Systems to independently kill terrorists amongst claims by Amnesty that the people who were killed were innocent civilians.

2028: Two ships carrying oil crashes off the coast of China after the two independent AI systems fail to communicate efficiently with each other. Both systems assumed the other would have moved out of the way.

2029: Taliban uses LAWS to attack Lahore parliament. Seven ministers killed in the attack. Arab League bans all technology related to AI. African Union supports them. USA, India and Israel have made a fourth of their army AI based, while China has over a half.

2030: France becomes first EU country to denounce AI driven technology for military purposes with many Muslims protesting in European Countries. LAWS used now for domestic law enforcement in USA and Russia. A fully autonomous Phalynx System kills child crossing the border in between USA and Mexico. Man lands on Mars.

2031: CCAI now created as a subsidiary committee to the General Assembly Sixth Committee with the mandate to: regulation of all AI based technology, especially in situations where civilian lives are at stake. IBM displays for the first time a “Hard AI system” that behaves like a dog. ***The possibility of sentient robots now increases rapidly. (Note: very important update)***

Letter from the Executive Board:

Dear Delegates

Welcome to the CCAI at VITTUC 2016. We hope that by now your research is well on its way and you have gained a fair idea about the agenda at hand. Kindly go through this study guide and make note of the pointer questions and research provided as they shall be of utmost importance in ensuring a high level of debate during committee proceedings and it shall be expected that the delegates steer the discussion on those lines. However, the given list of topics is not exhaustive and it is not intended to be. The list is simply indicative of pressing issues and topics of concerns which must be addressed and will give you a bird’s eye view of the gist of the issue. The delegates are at full liberty to bring up any other relevant point for discussion.

This debate can go multiple ways:

1. Technical: Debate what is AI, which kind of AIs are permissible and which types should be outlawed. Talk about the technical barriers to further development and possible loopholes in current design techniques that may lead to internal and external mishaps.
2. Ethical: Debate how the concept of humanity changes as robots become more and autonomous. What kind of rights do robots deserve and what stops them from trumping human needs and wants.
3. Legal: Current legal instruments designed for humans cannot apply to robots. Now that robots can make human like decisions, should they also fall under the ambit of human law? Is the robot going to be prosecuted, the owner or the maker?

We understand that MUN conferences can be an overwhelming experience for first timers and a tiring one for those who are familiar with the concept. We strongly suggest the first timers to participate fully in the conference and if any doubt persists in their minds (either substantive or related to parliamentary procedure), they do not hesitate to clarify the same. There are multiple angles to approach this debate from, and a balance of all of them will serve our purpose the best.

If there are any further inquiries, like where to start your own research, mail us at satyajit.giri2012@vit.ac.in or meghanadkumar@gmail.com and we would be happy to point you along the way.

On a final note we must advice all delegates to act in the capacity of representatives of their respective countries and thus conduct themselves with proper decorum and diplomacy.

Best Wishes.

CHAIRPERSON | Satyajit Giri

VICE CHAIRPERSON | Meghana Kumar

III. Sample Resolution

Resolution GA/6/1.1

General Assembly Sixth Committee

Sponsors: United States, Austria and Italy

Signatories: Greece, Tajikistan, Japan, Canada, Mali, the Netherlands and Gabon

Topic: "Strengthening UN coordination of humanitarian assistance in complex emergencies"

The General Assembly,

Reminding all nations of the celebration of the 50th anniversary of the *Universal Declaration of Human Rights*, which recognizes the inherent dignity, equality and inalienable rights of all global citizens, [use commas to separate preambulatory clauses]

Reaffirming its Resolution 33/1996 of 25 July 1996, which encourages Governments to work with UN bodies aimed at improving the coordination and effectiveness of humanitarian assistance,

Noting with satisfaction the past efforts of various relevant UN bodies and nongovernmental organizations,

Stressing the fact that the United Nations faces significant financial obstacles and is in need of reform, particularly in the humanitarian realm,

1. Encourages all relevant agencies of the United Nations to collaborate more closely with countries at the grassroots level to enhance the carrying out of relief efforts; [use semicolons to separate operative clauses]
2. Urges member states to comply with the goals of the UN Department of Humanitarian Affairs to streamline efforts of humanitarian aid;
3. Requests that all nations develop rapid deployment forces to better enhance the coordination of relief efforts of humanitarian assistance in complex emergencies;
4. Calls for the development of a United Nations Trust Fund that encourages voluntary donations from the private transnational sector to aid in funding the implementation of rapid deployment forces;
5. Stresses the continuing need for impartial and objective information on the political, economic and social situations and events of all countries;
6. Calls upon states to respond quickly and generously to consolidated appeals for humanitarian assistance; and
7. Requests the expansion of preventive actions and assurance of post-conflict assistance through reconstruction and development. [end resolutions with a period]