Futuristic Committee for Control of Artificial Intelligence

Study Guide

AGENDA

Ethics in Artificial Intelligence

Executive Board

Pranav Satish
Chairperson

Jaivignesh Jayakumar
Vice Chairperson

Letter from the Executive Board

Greetings, delegates!

We are proud to welcome you to the Futuristic Committee for Control of Artificial Intelligence (FCCAI) simulation at VITTUC 2017. The quality of the council depends on the quality of your contribution to it, and if that is to be ensured, then reading this guide is a must. We hope that you've done your part in preparing for two days of intense debate that lies ahead. Ahead of this kind of a committee, we expect you to have significant knowledge of your country and the policies your country has adopted over the years. Knowing the friends, enemies and geopolitical dynamics of your country goes a long way in being a key player in the committee. Seeing as this is a committee concerning artificial intelligence, we expect you to have fundamental knowledge about **computer science, robotics and use of artificial intelligence** in day to day life. This agenda has been strategically chosen keeping in mind the impact it has on the international community. **Even though this is a futuristic committee, it is a realistic one and we expect the delegates to be pragmatic in their debate.** However, this guide is nothing more than a starting point in your research and will in no way hold as viable source in committee. If you choose to back up your research with substantial sources, please note that we will only be accepting reports from Reuters and UN offices as credible sources of information. Over the two days of the conference, we expect you to display adequate diplomatic capabilities and come up with innovative and feasible solutions to the agenda under discussion. We're pretty sure that by the end of this simulation, you'll have tapped the diplomatic potential and the speaking prowess in you. We hope that you will enjoy this experience and demonstrate a greater interest in international politics.

Godspeed, y'all!

About the Committee

Futuristic committees, in many aspects, represents the upper echelon of the Model United Nations circuit. While other councils may be equally difficult, the difficulty of a futuristic committee resides in the more research-oriented setting rather than the low probability of being called on in a large body. The most exciting part of such committees is that these committees move forward in time and can be affected by events that occur in the outside world. Wars may break out, natural disasters can occur, and scandals or corruption can be revealed. You will have many chances to speak; therefore, your speech must maintain substantive quality as well as strength. This requires you to respond intelligently with leadership coupled with logical debate. The ever-changing dynamics of the committee also means that lost momentum can easily be regained as crises develop. During the simulation, you should try your best to represent a citizen from your country and his or her interests by reacting to situations appropriately. However, its nature as a hypothetical council **does not make foreign policy any less important**—whatever approach you choose to take must be in line with your country's interests, and if that means a rejection of whatever the council comes up with, then that is what you will have to do. This opens the door to a **multitude of approaches**, both realistic and idealistic.

This committee deals with the ethical use of artificial intelligence, global challenges and threats to peace that stem out of the same and seeks out solutions to the challenges faced by the international community. It considers all legal and international security matters related to artificial intelligence that lie within the scope of the **Charter of the United Nations**, as well as principles governing the peaceful use of artificial intelligence and regulation of safeguards. It also encourages promotion of cooperative arrangements and measures aimed at strengthening stability through all levels of administration.

Introduction

Artificial intelligence (AI) is a sub-field of computer science. The goal of AI is to enable the development of computers to do things that are normally done by intelligent human beings. This includes the **procurement of information, rules for assimilating this information, reasoning to achieve 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."

A critical problem that occurs here is how we define AI or even intelligence as we cannot characterize what kinds of activities are considered intelligent. Mainstream thinking in psychology regards human intelligence not as a single ability or cognitive process but rather as an array of separate components. This is important because any law that governs AI will be rendered irrelevant if member states argue that these systems do not fulfil their **country's definition for AI**. Therefore, a relevant definition can positively shape the discourse and direction in which further AI technology is developed. Currently, the five main aspects of AI include **learning, reasoning, problem-solving, perception and linguistic-processing**.

In fact, AI has no real definition of intelligence to offer, not even in the sub-human case. Worms are intelligent, but what exactly must a research team achieve in order for it to be the case that the team has created an artefact as intelligent as a worm?

In the absence of a reasonably precise criterion for when an artificial system counts as intelligent, there is no way of telling whether a research program that aims at producing intelligent artefacts has succeeded or failed. One result of AI's failure to produce a satisfactory criterion of when a system counts as intelligent is that whenever AI achieves one of its goals, **critics state that it is not intelligent**. How can this problem be resolved?

This opens up the classification into three major groups: Strong AI, Weak AI and The In-between.

Strong AI

The overall aim of Strong AI is to build machines that have the same (or more) amount of intellectual abilities as compared to a human being. In essence, its intelligence would be **indistinguishable from that of a human**. Joseph Weizenbaum, of the MIT AI Laboratory, has described the ultimate goal of strong AI as being *"nothing less than to build a machine on the model of man, a robot that is to have its childhood, to learn language as a child does, to gain its knowledge of the world by sensing the world through its own organs, and ultimately to contemplate the whole domain of human thought"*.

There is a huge difficulty in developing AI systems that are capable of displaying the aforementioned abilities. After years of exaggerated claims of success, there has been a huge amount of damage to research in this field. In spite of six decades of research in AI, there is **very little evidence** to prove that these systems can manifest intelligence that is displayed by human beings. Currently, it is impossible to even develop a system that displays the overall intelligence of the smaller living beings, let alone humans. However, the **lack of progress** may simply be a testimony to the difficulty of strong AI, not to its impossibility.

Weak AI

Weak artificial intelligence is a form of AI specifically designed to be focused on a narrow task and to seem very intelligent at it. Weak AI is never taken as a general intelligence but rather a construct designed to be intelligent in the narrow task that it is assigned to. A very good example of a weak AI is Apple's Siri, which has the Internet behind it serving as a powerful database. Siri seems very intelligent, as it is able to hold a conversation with actual people, even giving snide remarks and a few jokes, but actually operates in a very narrow, predefined manner. However, the **"narrowness"** of its function can be evidenced by its inaccurate results when it is engaged in conversations that it is not programmed to respond to. Robots used in the manufacturing process can also seem very intelligent because of the accuracy and the fact that they are doing very complicated actions that could seem incomprehensible to a normal human mind. But that is the **extent of their intelligence**; they know what to do in the situations that they are programmed for, and outside of that they have no way of determining what to do. Even AI equipped for machine learning can only learn and apply what it learns to the scope it is programmed for.

The In-between

These are 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**.

The closest available commercial subset of this technology is Applied AI which is also known as advanced information-processing. This aims to produce commercially viable "smart" systems--such as, for example, a security system that is able to recognise the faces of people who are permitted to enter a particular building. Applied AI has already enjoyed a considerable rate of success.

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 behaviour 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.

Turing Test

The Turing Test was coined by Alan Turing in the year, 1950. It was designed to be a rudimentary way of testing whether or not a system counts as intelligent. This is similar to some sort of an **imitation game** where the system's job is to chat and convince (or trick) a human judge into thinking that it is a real human. There is a series of 30 minute conversations with a panel of judges. A system passes the test if it **convinces 33% of the judges** of its humanity. Some experts think that the Turing test as a test of system intelligence where a computer is intelligent if and only if the test fails to distinguish it from a human being. However, Turing himself in fact pointed out that his test cannot provide a definition of intelligence. It is possible, he said, that a computer which ought to be described as intelligent might nevertheless fail the test because it is not capable of successfully imitating a human being. For example, why should an intelligent robot designed to oversee mining on the moon necessarily be able to pass itself off in conversation as a human being? If an intelligent entity can fail the test, then the test cannot function as a definition of intelligence. What follows is a transcript of a conversation of a robot named Eugene which passed the Turing test in the year, 2012.

[16:11:25] Judge: what music do you like?
[16:11:32] Eugene: I like Eminem – because he doesn't have a thing to do with "polite-correctness".
[16:11:46] Judge: OMG i love eminem, he's my favourite artist!
[16:12:00] Eugene: I wonder what Eminem's nick mean. I associate it only with a word "enema". But he is a cool guy, I like his "The Real Slim Shady".
[16:12:55] Judge: i know what his nick name means, his real name is Marshall Mathers. From the initials, M&M phynetically sounding like Eminem.
[16:13:06] Eugene: Eminem is a famous crapper. No, raper... ripper... Yeah, I mean - rapper. I like his "Stan" and "The Real Slim Shady".
[16:13:32] Judge: I like "Hailies song"
[16:13:46] Eugene: Interesting tastes. I like science fiction books, by the way. Where do you came from, by the way? Could you tell me about the place where you live?
[16:14:22] Judge: Epsom, home of the derby. yourself?
[16:14:27] Eugene: My guinea pig says that name Derby sounds very nice.
[16:14:41] Judge: i should talk with him sometime
[16:14:48] Eugene: I think you should take some typing lessons. Or just type more accurately.
[16:15:04] Judge: that's very rude?
[16:15:09] Eugene: Ummm... Frankly, I didn't get your question:-(
[16:15:49] Judge: "Ummm" isn't a word. i think you need typing lessons
[16:15:58] Eugene: My thoughts are same. By the way, I still don't know your specialty - or, possibly, I've missed it?

For one thing, winning a competition by pretending to be a child with gaping holes in their knowledge does not exactly reinforce the idea that machines are something to be scared of. Does this open the door to **new possibilities of testing systems intelligence?**

Lethal Autonomous Weapon Systems (LAWS)

Fully autonomous weapon systems are highly sophisticated weapon systems with 'artificial intelligence' that are programmed to independently determine their own actions, make complex decisions and adapt to their environment. 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.

There is a distinction between **automatic systems** and **autonomous systems** where the former operates with pre-programmed instructions to carry out a specific task, whereas the latter act dynamically to decide if, when, and how to carry out a task. Automatic systems therefore act based on deterministic (rule-based) instructions whereas autonomous systems act on stochastic (probability-based) reasoning, which introduces uncertainty. However, the future military systems would most likely be hybrids of automatic and autonomous systems.

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. Hence. there are four main drivers for military interest in increased overall autonomy for weapons platforms, which are linked to the **advantages of unmanned weapon systems** in general.

1. The potential for reduced operating costs and personnel requirements;
2. The potential for saving soldier lives in high risk operations;
3. The potential for increased safety in operating these platforms (compared to manned systems);
4. The potential for increased military capability by using one weapons platform to perform all functions – from identifying through to attacking a target.

Other drivers of autonomy in weapon systems include the potential for:

1. Force multiplication (i.e. greater military capability with fewer personnel);
2. Removal of risks to one's own forces;
3. Decreased reliance on communications links.

However, many of these advantages may still be possible while retaining remote control of the critical functions of selecting and attacking targets. There are also some functions, such as 'autopilot' in military and civilian aircraft, which have been autonomous for many years. For other functions, such as target selection and attack, direct human control is maintained for the vast majority of weapon systems today. What is important is whether critical functions such as 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. However, there are several limitations in the current technology of autonomous systems that are particularly relevant for military applications such as weapon systems.

1. The current autonomous systems are 'brittle' (not adaptable and easily break down), which makes them unreliable;
2. The existing autonomous systems still rely heavily on human input for many functions in order to correct mistakes;
3. They lack humane characteristics such as emotion and empathy;
4. They might not be able to distinguish between legal and illegal orders which thereby blurs the line of accountability;
5. They do not have the capacity to distinguish between civilians and combatants which is a fundamental principle in International Law;
6. They might not be able to obey by the principles of distinction and proportionality thereby violating the right to life;
7. They could be easily reprogrammed if fallen into the wrong hands and this could cause an asymmetric aggravation in the ongoing conflict;
8. There is no metric of accountability as robots cannot be jailed or punished for their actions;
9. There is a lack of standard methodologies to test and validate autonomous systems;
10. Finally, and perhaps the greatest barrier to development of autonomous weapon systems in particular, is the limited ability of autonomous robotic systems to perceive the environment in which they operate.

Even though there are no internationally agreed definitions of autonomous weapon systems, we can divide autonomous weapons into three types according to the level of autonomy and the level of human control:

1. *Autonomous weapon system (human 'out-of-the-the-loop'):* A weapon system that, once activated, can select and engage targets without further intervention by a human operator. Examples include some 'loitering' munitions that, once launched, search for and attack their intended targets over a specified area and without any further human intervention.

2. *Supervised autonomous weapon system (human 'on-the-loop'):* An autonomous weapon system that is designed to provide human operators with the ability to

intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur. Examples include defensive weapon systems used to attack incoming missile or rocket attacks with a human retaining supervision of the weapon operation.

3. *Semi-autonomous weapon system (human 'in-the-loop'):* A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator. Examples include 'homing' munitions that, once launched to a particular target location, search for and attack preprogramed categories of targets within the area.

There are three main considerations for assessing the implications of autonomy in a given weapon system:

1. The task the weapon system is carrying out;
2. The level of complexity of the weapon system;
3. The level of human control or supervision of the weapon system.

The critical functions of some weapons systems have been automated for many years and that a weapon system does not necessarily need to be highly complex for it to be autonomous. The autonomous weapon systems in use today conforming to the definitions provided are constrained in several respects:

1. They are limited in the tasks they are used for (e.g. defensive roles against rocket attacks, or offensive roles against specific military installations such as radar);
2. They are limited in the types of targets they attack (e.g. primarily vehicles or objects rather than personnel);
3. They are used in limited contexts (e.g. relatively simple and predictable environments such as at sea or on land outside populated areas).

Domestic Use of Artificial Intelligence

Artificial intelligence is different from psychology because it emphasis on computation and is different from computer science because of its emphasis on perception, reasoning and action. It makes machines smarter and more useful. It works with the help of artificial neurons (artificial neural network) and scientific theorems (if then statements and logics). AI technologies have matured to the point in offering real practical benefits in many of their applications. Major Artificial Intelligence areas are Expert Systems, Natural Language Processing, Speech Understanding, Robotics and Sensory Systems, Computer Vision and Scene Recognition, Intelligent Computer Aided Instruction, Neural Computing. Artificial intelligence has the advantages over the natural intelligence **not in the form of humanoid like cyborgs** (like Terminator) but rather in the manner of intelligent computer systems that are more permanent, consistent, less expensive, has the ease of duplication and dissemination, can be documented and can perform certain tasks **much faster and better than the human** (like J.A.R.V.I.S). 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. While even Google cars have been recorded to make mistakes and crash, the question is if 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 resolution** 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 if 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 head of state? 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. There is a **gap in accountability** if the AI technology behaves abnormally. The rise of AI technology also introduces the **question of security**. As in the trolley case, the worry is a lot of people will hack their own cars and reprogram it. There is also the worry of **external hacking**. Every AI system requires knowledge about its surroundings, a way to communicate and consequently have mechanisms that are susceptible to corruption. The key here is to understand that the concerns that we have for one type of AI system is usually applicable to other kinds.

Ethics in Artificial Intelligence

The possibility of creating thinking machines raises a **host of ethical issues**. These questions relate both to ensuring that such machines do not harm humans and other morally relevant beings, and to the moral status of the machines themselves. There are various things that need to be addressed when discussing this topic.

1. The issues that may arise in the near future of AI;
2. The challenges for ensuring that AI operates safely as it approaches humans in its intelligence;
3. The assessment whether, and in what circumstances, AIs themselves have moral status;
4. The difference of AIs from humans in certain basic respects relevant to our ethical assessment of them;
5. The issues of creating AIs more intelligent than human, and ensuring that they use their advanced intelligence for good rather than ill;

A different set of ethical issues arises when we contemplate the possibility that some future AI systems might be candidates for having moral status. Our dealings with beings possessed of moral status are not exclusively a matter of **instrumental rationality**: we also have moral reasons to treat them in certain ways, and to refrain from treating them in certain other ways. Therefore, before we begin discussion on the aforementioned topics, there are a few concepts that everybody should be clear of:

1. X has **moral status** because X counts morally in its own right, it is permissible/impermissible to do things to it for its own sake;
2. **Sentience** is the capacity for phenomenal experience or qualia, such as the capacity to feel pain and suffer;
3. **Sapience** is a set of capacities associated with higher intelligence, such as self-awareness and being a reason-responsive agent;
4. **Principle of Substrate Non-Discrimination** states that if two beings have the same functionality and the same conscious experience, and differ only in the substrate of their implementation, then they have the same moral status.
5. **Principle of Ontogeny Non-Discrimination** states that if two beings have the same functionality and the same consciousness experience, and differ only in how they came into existence, then they have the same moral status.
6. **Principle of Subjective Rate of Time** states that in cases where the duration of an experience is of basic normative significance, it is the experience's subjective duration that counts.

Thus, the discipline of AI ethics, especially as applied to AGI, is likely to differ fundamentally from the ethical discipline of non-cognitive technologies, in that:

1. The local, specific behaviour of the AI may not be predictable apart from its safety, even if the programmers do everything right;
2. Verifying the safety of the system becomes a greater challenge because we must verify what the system is trying to do, rather than being able to verify the system's safe behaviour in all operating contexts;
3. Ethical cognition itself must be taken as a subject matter of engineering.

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. Some have argued that if AI reach a level where they can even reproduce, they might demand **basic rights** like robot healthcare, robot housing, etc. Experts disagree whether **specific and detailed laws** will be required soon or safely in the distant future. An interesting question then, is the characteristics an AI system must display to become **deserving of rights**. The most important thing that needs to be addressed is the **validity of their consciousness** in order to be eligible for rights. As humans, we **associate consciousness with pain** and seeing as a robot can't feel pain, are they truly conscious?

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

Sometimes, 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. 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.

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm;
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law;
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

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 behaviour**. His work suggests that no set of fixed laws can sufficiently anticipate all possible circumstances.

In 2009, during an experiment, robots that were **programmed to cooperate** with each other in searching out a beneficial resource and avoiding a poisonous one. They **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** where it in contrast, ultimate human motives typically have a quality of requiring never-ending learning). 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**.

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**. 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.

Superintelligence is one of several "existential risks" as defined by Bostrom (2002): a risk "where an adverse outcome would either **annihilate Earth**-originating intelligent life or permanently and drastically curtail its potential". Conversely, a positive outcome for superintelligence could **preserve Earth**-originating intelligent life and help fulfil its potential. It is important to emphasize that smarter minds **pose great potential benefits as well as risks**.

Although current AI offers us few ethical issues that are not already present in the design of cars or power plants, the approach of AI algorithms toward more humanlike thought portends predictable complications. **Social roles** may be filled by AI algorithms, implying new design requirements like **transparency and predictability**. Sufficiently general AI algorithms may no longer execute in predictable contexts, requiring new kinds of **safety assurance** and the engineering of **artificial ethical considerations**. AIs with sufficiently **advanced mental states**, or the right kind of states, will have moral status, and **some may count as persons**—though perhaps persons very much unlike the sort that exist now, perhaps governed by different rules. And finally, the prospect of AIs with superhuman intelligence and **superhuman abilities** presents us with the extraordinary challenge of stating an algorithm that outputs **super ethical behaviour**.

Timeline of events

February 08, 2018: President of the United States of America, Donald J. Trump and a predominantly Republican Congress vote to institute the use of LAWS in the ongoing conflict in the Middle East.

March 01, 2019: LAWS destroy ISIS in Iraq with only one major setback where a single autonomous drone identifies an entire town as a threat and kills over 100 civilians. Many countries follow suit to institute LAWS in their armed forces.

April 04, 2020: Baidu reveals its first batch of autonomous car series, *Gixao*. It has been approved by both American and Chinese road safety requirements. Google, BMW and Honda are expected to follow soon.

May 15, 2021: The European Union reports that a combination of the Google *X* and Baidu *Gixao* has led to a 70% reduction in accidents. Owing to the success of these systems, Autonomous ships are now being tested for movement of goods, fishing etc.

June 21, 2022: There have been various reports that the autonomous cars are being taken apart and assembled for LAWS by multiple non-state actors. These systems are used to attack armed convoys carrying aid into the Middle East.

July 7, 2023: IBM displays for the first time a Strong AI system that behaves like a dog. The possibility of sentient robots now increases rapidly. Funding increases rapidly into research and development of these systems.

August 18, 2024: LAWS are being used for domestic law enforcement in multiple countries. There is a widespread feeling that there is a difference between a robot killing someone and a cop pulling the trigger. Several states have passed bans on putting weapons on drones.

April 22, 2025: *Futuristic Committee for Control of Artificial Intelligence* is now created as a subsidiary organ to the *United Nations General Assembly* with the mandate to regulate AI based technology and suggest measures to control and ensure the ethical use of AI.

Questions a resolution must answer (QARMA)

1. What is a legally binding definition of AI?
2. What types of AI are permissible and which types should be outlawed?
3. What are the measures to correct the loopholes in the current design techniques that might lead to internal as well as external mishaps?
4. What kind of rights do robots deserve with increase in autonomy?
5. What kind of a legal framework should be applied to robots and is there going to be a provision for prosecution?
6. Is the use of LAWS legal and if yes, to which degree of autonomy can their use be legally monitored by current international law?
7. Can LAWS be programmed to abide by principles of international law?
8. To what extent can “critical decisions” concerning legal interests such as right to life be delegated completely to autonomous weapons systems?
9. What kind of a failsafe could be designed to assure the protection of human interests when these systems don’t perform as planned?

Conclusion

For a simpler understanding of the ethical principles revolving around the agenda, watch the following video.

Do Robots Deserve Rights? What if Machines Become Conscious?

A reasonable knowledge of theoretical approaches to international relations will, thus, be useful in preparing for this council. Realism, liberal internationalism, constructivism, and so on, are a few ideas to brush up on, because delegates will have to create a resolution from scratch.

These approaches, when informed by the tools, technical feasibility and the context of the nation you are representing, should help you narrow down the kind of policy that you will want to pursue as a result of this council. Delegates should also brush up on historical AI conflicts and how they were resolved.

Ideological clashes between nations must also be understood in order to attain a clearer picture of why countries cannot often easily surmount their rivalries even when they stand to gain from them.

For further clarification, feel free to contact any of us!

Pranav Satish – pranavsatish06@gmail.com; +91 97104 13496
Jaivignesh Jayakumar - jaivignesh.jayakumar@gmail.com; +91 89393 19119