Committee: United Nations Environment Programme

Issue: Artificial Intelligence: a mixed blessing.

DEFINITION OF KEY TERMS

Artificial intelligence

The first thing we need to do is understand what an AI actually is. The term "artificial intelligence" refers to a specific field of computer engineering that focuses on creating systems capable of gathering data and making decisions and/or solving problems. An example of basic AI is a computer that can take 1000 photos of cats for input, determine what makes them similar, and then find photos of cats on the internet. The computer has learned, as best as it can, what a photo of a cat looks like and uses this new intelligence to find things that are similar.

Artificial General Intelligence (AGI)

: Also called super intelligence, it's when the capabilities of machine intelligence is equal to or greater than human intelligence.

Artificial Intelligence **(AI)**:

The field of computer science in which machines are enabled to simulate human cognition and learning.

Artificial Narrow Intelligence (ANI)

: Refers to AI that is limited to a specific set of topics and capabilities.

Artificial Neural Network (ANN):

A model used in AI, it is loosely based on the human brain. It consists of neural layers that are used for machine learning.

Autonomous

Simply put, autonomy means that an AI construct doesn't need help from people. Driverless cars illustrate the term "autonomous" in varying degrees. Level four autonomy represents a vehicle that doesn't need a steering wheel or pedals: it doesn't need a human inside of it to operate at full capacity. If we ever have a vehicle that can operate without a driver, and also doesn't need to connect to any grid, server, GPS, or other external source in order to function it'll have reached level five autonomy.

Anything beyond that would be called sentient, and despite the leaps that have been made recently in the field of AI, the singularity (an event representing an AI that becomes self-aware) is purely theoretical at this point.

Algorithm

The most important part of AI is the algorithm. These are math formulas and/or programming commands that inform a regular non-intelligent computer on how to solve problems with artificial intelligence. Algorithms are rules that teach computers how to figure things out on their own. It may be a nerdy construct of numbers and commands, but what algorithms lack in sex appeal they more than make up for in usefulness.

Backpropagation:

Also known as "backward propagation of errors," it is a supervised learning technique where errors are computed at the output and distributed backward through the layers of the artificial neural network. It's a common method of training an artificial neural network where the system's initial output is compared to the desired output, then the system is adjusted until the difference is minimized.

Black box

When the rules are applied an AI does a lot of complex math. This math, often, can't even be understood by humans (and sometimes it just wouldn't be worth the time it would take for us to figure it all out) yet the system outputs useful information. When this happens it's called black box learning. The real work happens in such a way that we don't really care how the computer arrived at the decisions it's made, because we know what rules it used to get there. Black box learning is how we can ethically skip "showing our work" like we had to in high school algebra.

Convolutional Neural Network (CNN)

It's a type of neural networks used to identify and analyze images.

Deep learning

Deep learning is what happens when a neural network gets to work. As the layers process data the AI gains a basic understanding. You might be teaching your AI to understand cats, but once it learns what paws are that AI can apply that knowledge to a different task. Deep learning means that instead of understanding what something is, the AI begins to learn "why."

Forward Chaining

A method where AI looks back and analyzes the rule-based system to find the "if" rules, and to determine which rules to use to find a solution.

Generative Adversarial Networks (GAN)

A type of AI algorithm used in unsupervised machine learning where there are two neural networks (generator and discriminator) trained on the same data set. The generator produces output, and the discriminator compares the output produced with the original data set in efforts to determine which images are authentic. Based on those results, the generator adjusts its parameters for creating new output. This process is iterated until the

discriminator is no longer able to distinguish the generator's output with the original data set. Used to create photorealistic images.

Heuristics

Common-sense rules based on experience. In heuristic programming, programs are self-learning, and improve with experience. Frequently used with expert systems.

Inductive reasoning

A logical process where multiple premises that are true or true most of the time, are combined to form a conclusion. Often used in prediction and forecasting.

Machine learning

The meat and potatoes of AI is machine learning — in fact it's typically acceptable to substitute the terms artificial intelligence and machine learning for one another. They aren't quite the same, however, but connected.

Machine learning is the process by which an AI uses algorithms to perform artificial intelligence functions. It's the result of applying rules to create outcomes through an AI.

Neural network

When we want an AI to get better at something we create a neural network. These networks are designed to be very similar to the human nervous system and brain. It uses stages of learning to give AI the ability to solve complex problems by breaking them down into levels of data. The first level of the network may only worry about a few pixels in an image file and check for similarities in other files. Once the initial stage is done, the neural network will pass its findings to the next level which will try to understand a few more pixels, and perhaps some metadata. This process continues at every level of a neural network.

Natural language processing

It takes an advanced neural network to parse human language. When an AI is trained to interpret human communication it's called natural language processing. This is useful for chat bots and translation services, but it's also represented at the cutting edge by AI assistants like Alexa and Siri.

Reinforcement learning

AI is a lot more like humans than we might be comfortable believing. We learn in almost the exact same way. One method of teaching a machine, just like a person, is to use reinforcement learning. This involves giving the AI a goal that isn't defined with a specific metric, such as telling it to "improve efficiency" or "find solutions." Instead of finding one specific answer the AI will run scenarios and report results, which are then evaluated by humans and judged. The AI takes the feedback and adjusts the next scenario to achieve better results.

Supervised learning

This is the very serious business of proving things. When you train an AI model using a supervised learning method you provide the machine with the correct answer ahead of time. Basically the AI knows the answer and it knows the question. This is the most common method of training because it yields the most data: it defines patterns between the question and answer.

If you want to know why something happens, or how something happens, an AI can look at the data and determine connections using the supervised learning method.

Unsupervised learning

In many ways the spookiest part of AI research is realizing that the machines are really capable of learning, and they're using layers upon layers of data and processing capability to do so. With unsupervised learning we don't give the AI an answer. Rather than finding patterns that are predefined like, "why people choose one brand over another," we simply feed a machine a bunch of data so that it can find whatever patterns it is able to.

Transfer learning

Another spooky way machines can learn is through transfer learning. Once an AI has successfully learned something, like how to determine if an image is a cat or not, it can continue to build on it's knowledge even if you aren't asking it to learn anything about cats. You could take an AI that can determine if an image is a cat with 90-percent accuracy, hypothetically, and after it spent a week training on identifying shoes it could then return to its work on cats with a noticeable improvement in accuracy.

Turing Test

If you're like most AI experts you're cautiously optimistic about the future and you have reservations about our safety as we draw closer to robots that are indistinguishable from people.

Alan Turing shared your concerns. Though he died in 1954 his legacy lives on in two ways. Primarily he's credited with cracking Nazi codes and helping the Allies win World War 2. He's also the father of modern computing and the creator of the Turing Test.

While the test was originally conceived as a way of determining if a human could be fooled by a conversation, in text display only, between a human and an artificial intelligence, it has since become short hand for any AI that can fool a person into believing they're seeing or interacting with a real person.

BACKGROUND INFORMATION

Artificial intelligence a curse and a blessing

Artificial intelligence (also known as machine intelligence) is the artificial counterbalance to humans' natural intelligence. It is therefore crucial -in order to define the former- to define the latter. Natural intelligence is an emergent process derived from an organic system, all the while artificial intelligence is an emergent process derived from a computational system. A machine that possesses artificial intelligence is characterised as an intelligent agent, meaning "any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals" (Poole). More practically, we perceive as intelligent any machine that can simulate human cognitive processes such as problem solving, decision making or learning.

Artificial intelligence is a concept that has been around for a while; from the antiquity's mythical giant automaton Talos, to Karel Čapek's R.U.R. (Rossum's Universal Robots), the idea of constructed intelligence seems to resonate with humans. With the emergence of computational technology however, intelligent machines departed from the realm of fantasy and hesitantly intruded into our daily lives, first making their appearance as tools, but later captivating us with their alluring potential for pseudo-sentience.

Not every machine is intelligent, and it is important to distinguish those that are. For a better theoretical articulation of what an intelligent construct is, the famous mathematician Alan Turing came up with a test to support the dichotomous categorisation. This "Turing test", as it is called, stipulates that whichever mechanism can fool a human being into considering it to be human in non-audiovisual linguistic exchange should be considered intelligent (Turing). Turing's dichotomy may have been very useful as an early benchmark, however since his time the potential of artificial intelligence has shown another facet. Instead of simulating humanity, it has the potential to exceed it.

Most of today's AI is programmed to be goal-oriented. Said goals can be either simple or complex, ranging from playing games within a defined domain to creating frames of operation for complex human analysis. Goals can be either externally defined, or they can be left as open-ended ones for the machine to define along its operation. If the AI is programmed for "reinforcement learning", goals can be implicitly induced through conditioning models, where certain outcomes are algorithmically "rewarded", while others are "punished" so as to form behavior through the seeking of positive stimulus and the avoidance of unpleasant ones -even if those will be mere inputs- (Dimitri).

Alternatively, an evolutionary system can define its own goals according to a "fitness function", which is based on selection of random deviations for increasingly successful performance within a given system (Domingos). Such dynamics within AI development strongly mimic those in human adaptiveness, and are thus appropriately fascinating as both useful tools and even vessels for the study of development and growth in us.

What almost all AI have in common is that they revolve around the use of algorithms. An algorithm is "a set of unambiguous instructions that a mechanical computer can execute" (Berlinski). Many AI algorithms are capable of "learning" from data; enhancing their function in accordance to data inputs and their desired outcomes or goals.

Today Artificial Intelligence can be considered a mixed blessing. AI has taken over the worlds of transportation, logistics, and finance, due to their unparalleled speed and adaptiveness. AI has also become indispensable in areas such as heavy industry or medicine, where their capacity for meticulous and delicate work, as well as the magnitude to which their scale can be stretched, allows for the facilitation of services at a much higher level that otherwise possible. The benefits of AI are comparable to the benefits of technology as a whole. It is an inevitable side effect of progress and a force, whose sheer power is irreversible. This is why it also has some downsides. AI has allowed for the information age's worse developmental trajectories to be realised. One of the most important downsides of AI

technology is the surveillance capabilities it has led to. In a 2018 scandal, the dangerous potential of AI was revealed through the workings of the data firm Cambridge Analytica. The company used AI psychographics to sort through data from millions of facebook users and psychologically profile them. The AI allowed for people's fears and desires to be identified, and thus with 90th and above percentiles of accuracy, and all without the individual's' consent. This information was up for sale and ,regardless of the unethical nature of this specific act, what is most important is the general direction that AI is heading to. Without oversight, such recklessness will only further itest.

The delegates of the United Nations are called upon to reconcile the good and the bad outcomes of relyance of Artificial Intelligence, and to form a comprehensive framework under which its positive capacities can be fully tapped into, while it's potential harms are mitigated.

Applications of AI

Military

Unfortunately, the use of artificial intelligence from the military is a reality and thus Many artificial intelligence researchers seek to distance themselves from military applications of AI but even with these disagreements the annual military spending on robotics worldwide rose from 5.1 billion USD in 2010 to 7.5 billion USD in 2015.

These machines often refer to as Military drones, are capable of autonomous action and are widely considered a useful asset.

'

Finance and economics

Financial institutions have long used artificial neural network systems to detect charges or claims outside of the norm, flagging these for human investigation. The use of AI in banking can be traced back to 1987 when Security Pacific National Bank in US set-up

a Fraud Prevention Task force to counter the unauthorised use of debit cards. Programs like Kasisto and Moneystream are using AI in financial services.

Banks use artificial intelligence systems today to organize operations, maintain book-keeping, invest in stocks, and manage properties. AI can react to changes overnight or when business is not taking place. In August 2001, robots beat humans in a simulated financial trading competition. AI has also reduced fraud and financial crimes by monitoring behavioral patterns of users for any abnormal changes or anomalies

Artificial Intelligence in Medicine

Since the middle of the last century, researchers have explored the potential applications of intelligent techniques in every field of medicine.3,4 The application of AI technology in the field of surgery was first successively investigated by Gunn in 1976, when he explored the possibility of diagnosing acute abdominal pain with computer analysis.5 The last two decades have seen a surge in the interest in medical AI. Modern medicine is faced with the challenge of acquiring, analysing and applying the large amount of knowledge necessary to solve complex clinical problems. The development of medical artificial intelligence has been related to the development of AI programs intended

AI is being applied to the high cost problem of dosage issues—where findings suggested that AI could save $16 billion. In 2016, a ground-breaking study in California found that a mathematical formula developed with the help of AI correctly determined the accurate dose of immunosuppressant drugs to give to organ patients.

UN INVOLVEMENT: RELEVANT RESOLUTIONS, TREATIES AND EVENTS

POSSIBLE SOLUTIONS

Artificial Intelligence, which embodies the principle of competing with the human capacities, is a bet for mankind. Therefore, although its benefits are more than tangible, it is worth setting some boundaries, so that the human being is not overshadowed, but rather benefited by AI's presence.

Transparency in the development process:

Strict ethical rules have to be enforced, which will precisely define the limits of AI research, because scientists tend to forget the moral aspect of their work in their strive for greatness. It should be noted that excessive obstacles impede the evolution, whereas absolute freedom to produce artificially intelligent machines poses grave threats to the society's existence. Therefore, a compromising middle ground should be found that will fuse the benefits of the artificial and the natural world. In the same spirit, research cannot be conducted "behind closed doors". Undoubtedly, specialists are capable of accomplishing marvels in their laboratories and furthering the progress of the human species, nevertheless the obligation to report to the society is of paramount importance, as well. Via frequent information on the scientific advances, public opinion is enlightened. A fertile interconnection can thus be initiated between the two sides, society and scientists, from which all can reinforce their work and knowledge.

Elimination of racism in AI environments:

Unfortunately, strong evidence has suggested that AI applications systematically discriminate against specific social groups and populations. Deep neural networks are often trained on ImageNet, a set of more than 14 million labeled images, in which genetic, ethnic and cultural biases are unintentionally produced. In this ImageNet data, some groups are also over-represented, whereas others are under-represented. That happens, mainly because most images derive from the USA, home to only 4% of the world's population, when huge countries, such as India and China constitute only 3% of the total images. Biases in the data often reflect deep and hidden imbalances in institutional infrastructures and social power relations. Wikipedia, for example, seems like a rich and diverse data source. But fewer than 18% of the site's biographical entries are on women. Articles about women link to articles about men more often than vice versa, which makes men more visible to search engines.

Thus, technical care and social awareness must be brought to the building of data sets for training. Specifically, steps should be taken to ensure that such data sets are diverse and do not under represent particular groups. This means going beyond convenient classifications —'woman/man', 'black/white', and so on — which fail to capture the complexities of gender and ethnic identities. Every training data set in these neural networks should be accompanied by information on how the data were collected and annotated. If data contain information about people, then summary statistics on the geography, gender, ethnicity and

other demographic information should be provided. If the data labelling is done through crowdsourcing, then basic information about the crowd participants should be included, alongside the exact request or instruction that they were given. Generally, ethnicity, gender and other relevant information need to be accurately recorded. Unless the appropriate categories are captured, it's difficult to know what constraints to impose on the model, or what corrections to make.

Combating the threat of superintelligence:

- Computers are now capable of creating sophisticated cyberweapons. Drones have already been reported to carry out deadly missions, but in the hands of a computer they can lead to dangerously effective results. The same concerns can be raised about hacking. Recent research has indicated that computers can identify bugs in a system a lot faster than humans can do, in a way that can be exploited by hackers so as to find possible loopholes. They can conduct financial frauds, perform economic coercions to individuals and generally extract large sums of money from large companies and banks. In the realms of Facebook, they can collect personal data via face recognition cameras and imitate our friends so as to exploit the user It should not also come as a surprise that AI systems can spur false accusations as propaganda and misinformation tool. Public opinion is, this way, manipulated and certain political and financial outcomes can be guaranteed. So as to avoid the negative impact of these superintelligent machines, certain measures should be taken. The best way to provide adequate safeguards is to regulate the AI itself, just as Elon Musk and others have recommended. Measures like requiring the development of testing protocols for the design of AI algorithms and setting up input validation standards could certainly be helpful in this regard. And, of course, beefing up cybersecurity protections with the most advanced security technology is important for every enterprise doing business today. Lastly, to detect malicious AI early in its life before it acquires too many resources is merely a matter of simply paying close attention to an autonomous system and shutting it down when it becomes clear that it's up to no good. Alertness is a price that should be paid if we do not want machines eventually outsmart us. The recent "intelligence explosion" is quite alarming to anyone that cannot envisage a world non-human centred.

4) **Addressing the existential risk and the feeling of alienation:**

As it has been the case with other crucial moments in human history, the phase of transitioning in an AI world creates an existential angst to mankind. The arrival of the steam machine, the television and, more prominently, the internet spread worldwide anxiety back at their times. Today, it is the fear that intelligent "robots" will manage to acquire conscience and replace us once and for all. To adapt to this shifting state of affairs, the best advice is to fully contemplate that today's world is indispensable with AI and technology, in general. By accepting this, we should move on to create solid legal frameworks that will highlight the human factor as main coordinator of life on Earth. Some vocations cannot be performed by machines at a satisfactory level, human interaction is irreplaceable and, of course, life loses its fundamental meaning if one has an insight into the outcome and possibilities of actions, like an AI device has. AI is also accused of alienating humans from their workplace, social surroundings and other social events. In response to that, people can take up hobbies for personal fulfillment or be involved in projects and team works, where they can manifest their skills and not lose on their creativity, their camaraderie and the willingness of contributing to the general good. Quality time with peers and family is also a distraction from a world that sometimes feels like it is overwhelmed with technological

accomplishments. All in all, there are enough alternatives to regain interest in life, especially when one bears in mind that the reason this existential angst exists is a human invention.

During all this meddling with AI, the danger of tampering with nature on an excessive scale is hovering over our heads. Natural intelligence must not be marginalized, but it should be suitably combined with the benefits one can reap from the artificial world. After all, we can all witness the huge impact it has already made on eradication of war, illnesses and poverty, as well as the advance of space exploration, telecommunications e.t.c. As long as we do not wish to take the role of god and exceed our mortal limits, the road is open to perfection.

BIBLIOGRAPHY

https://www.un.org/ecosoc/en/events/2017/joint-meeting-ecosoc-and-second-committee-%E2%80%9C-future-everything-%E2%80%93-sustainable-development

https://news.un.org/en/story/2018/09/1019152

https://news.un.org/en/story/2004/10/118592-robots-are-taking-increasing-number-jobs-new-un-report-says

https://futureoflife.org/ai-policy-united-nations/?cn-reloaded=1

http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics

https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx

https://www.un.org/development/desa/dpad/publication/frontier-issues-artificial-intelligence-and-other-technologies-will-define-the-future-of-jobs-and-incomes/

https://www.techrepublic.com/article/united-nations-cito/

https://www.nature.com/articles/d41586-018-05707-8

https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1

https://www.theguardian.com/commentisfree/2016/aug/28/ideology-liberal-democracy-technology-bioscience-yuval-harari-artificial-intelligence

https://www.ingrammicroadvisor.com/security/the-top-3-ai-security-threats

https://thenextweb.com/artificial-intelligence/2017/09/10/glossary-basic-artificial-intelligence-terms-concepts/

https://www.psychologytoday.com/intl/blog/the-future-brain/201810/defining-artificial-intelligence-glossary-key-ai-terms

https://courses.csail.mit.edu/6.034s/handouts/Medical_AI.pdfectors

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1964229/

https://eric.ed.gov/?id=ED306945

https://eetyfiyw3hz08.storage.googleapis.com/EgNgpdjaX8nnRYCCxX08.pdf