



Instating regulations on the development of Artificial Intelligence



SPC1

Daphne de Haas, Sarah Porcheron
President, Deputy President

Forum: Special Conference 1 (SPC1)

Issue: Instating regulations on the development of Artificial Intelligence

Name: Daphne de Haas, Sarah Porcheron

Position: President, Deputy-President

Introduction

In 2020, a Kargu 2 drone from the Turkish company STM attacked a human target in Libya. According to a UN report: "The lethal autonomous weapons systems were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true 'fire, forget and find' capability". Artificial Intelligence allowed a weapon to guide itself to a target, bringing the idea of a killer robot to reality for the first time in recorded history. Nowadays, drone warfare has become increasingly common. On the Russo-Ukrainian battlefield, they constitute a substantial part of the opposing armies with Ukraine using kamikaze technology and the Russian Federation deploying autonomous mines capable of discerning home troops from enemy troops to avoid fratricide.

However, outside of the defense and space sector, AI is playing a central role due to its ability to simulate human intelligence. For example, in healthcare, AI applications are revolutionizing diagnosis, treatment planning, and patient care. In the contrasting sector of education, AI is redesigning learning experiences, adapting to individual needs, and evolving teaching methods. The integration of AI is a fundamental shift in the way we tackle complex problems and deliver services in these critical areas.

However, as seen in the context of weaponry, the implementation of AI comes with ethical, societal, and technical considerations. As AI algorithms become more sophisticated, discussions surrounding transparency, bias and privacy: implications of autonomous decision-making continue to gain significance. This report will oversee past and current regulations on the development of AI, from both a general and military perspective, in order to establish effective policies around this rapidly evolving sector.

Definition of Key Terms

Artificial Intelligence

Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or animals.

Algorithm

A set of rules or instructions designed to perform a specific task or solve a particular problem. In the context of AI, algorithms are used to enable machines to learn from data and make decisions.

Autonomous Weapons Systems (AWS) or Lethal Autonomous Weapon Systems

A type of autonomous military system that can independently search for and engage targets based on programmed constraints and descriptions.

Neural networks

Computational models inspired by the structure and functioning of the human brain. Neural networks are a fundamental component of deep learning, allowing machines to process information in a way similar to human thought processes.

Ethical AI

The consideration and implementation of ethical principles and guidelines in the development and deployment of AI systems, addressing issues such as bias, transparency, accountability, and the impact on society.

Transfer learning

A machine learning technique where a model trained on one task is adapted for a related task, leveraging knowledge gained from the initial training. Transfer learning is valuable for scenarios with limited labeled data.

Big data

The massive volume of structured and unstructured data generated by various sources, including sensors, social media, and online activities. Big data is crucial for training and improving AI models, as it provides the necessary information for learning patterns and making predictions.

Deepfake

Deepfake refers to the process of creating or modifying digital content using artificial intelligence, usually with the intention of portraying humans in unrealistic or unrepresentative environments or scenarios. It involves altering images, films, or audio to give the impression that someone said or did something they didn't actually do.

Natural Language Processing (NLP)

NLP alludes to the part of artificial intelligence that has to do with enabling PCs to comprehend, understand, and respond to composed and verbally expressed words similarly to how people can. NLP mixes factual, AI, and profound learning models with computational etymology — rule-based demonstrating of human language.

Reactive Machines

Reactive-generated machines are likely to function without relying on previous information and memories and therefore only respond to different stimuli.

Limited Memory

Limited memory machines give answers based on memory and the earlier activities. It also improves their responses as the more they develop the more accurate answers it can give.

Theory of Mind

The theory of mind is created to understand the needs of other intelligent entities. It will be able to better understand the entities it is interacting with by discerning their needs, emotions, beliefs, and thought processes. Development in other areas of Theory of Mind is also needed. Because to truly understand human needs, AI machines must see humans as individuals, whose thoughts can be influenced by multiple factors, and essentially "understand" humans.

Self-aware

Self-aware AI is an AI that has evolved to be so similar to the human brain that it has developed self-awareness. This type of AI will not only understand and call up emotions in those it interacts with, but also have emotions, needs, beliefs, and potentially desires of its own. While the emergence of self-awareness has the capacity to propel our society forward at breakneck speed, it also carries the risk of catastrophic consequences. This is due to the fact that an AI that becomes self-aware could have concepts like self-preservation, which could either directly or indirectly mean the end of humanity. An entity with such intelligence could quickly outwit a human's intellect and devise complex plans to subjugate humanity.

General Overview

To create and ensure regulations on the development of Artificial Intelligence, it is important to be aware of the roots of its establishments and the dangers it can show. Especially since AI is having a massive influence on a global scale.

Artificial Intelligence (AI) is the intelligence of machines or software, as opposed to the intelligence of humans or animals. This replication is possible, because of the large data sets that it combines with intuitive processing algorithms. It can control algorithms by learning behavior patterns within data sets of sensors, social media and online activities.

In 1955 the first AI-generated program was released called 'The Logic Theorist' by Alan Newell and Herbert Simon. It was made to solve humans' problems and theorems by answering in a symbolic logic. This was probably the first *working* program to give an accurate simulation of the aspects of people's ability to solve complex conflicts.

The term 'Artificial Intelligence' was officially coined when John McCarthy spoke at the Dartmouth conference in 1956. He defined it as "the science and engineering of making intelligent machines". And ever since AI has developed massively, seeing there are more than 300 AI tools on the

world wide web. These tools can be different types of AI as there are seven forms in AI classification, of which four key machines are explained below:

The most influencing AI machine is currently the chatbot ChatGPT. It is a computer program that simulates conversations between humans to enhance experiences for users. It has been used for predefined conversation flows, but ChatGPT is also for users to answer questions in NLP and give automated responses in real time.

But the estimated dangers of AI go from biased programming and hacking into personal accounts to creating nuclear weapons. Therefore, it is important that regulations are taken amongst AI users and AI creators. Additionally, it is crucial to understand that different regulations are needed for short-term consequences and long-term consequences, as they require a contrasting observation.

Timeline of Key Events

Date	Event
12th of July 1955	The Logic Theorist is released
Summer 1956	Definition of Artificial Intelligence by McCarthy
11th of June 2020	UN's High-level Panel on Digital Cooperation
13th of December 2022	UNESCO's First Global Forum on AI Ethics
November 2022	UN's Convention on Certain Conventional Weapon
23rd of November 2021	UNESCO's Recommendation on the Ethics of AI
Spring 2023	UN Group of Governmental Experts on AI
6th of February 2024	UNESCO's Second Global Forum on AI Ethics (planned)

Major Parties Involved

OpenAI

OpenAI is the organization behind the development of ChatGPT. OpenAI's mission is to ensure that artificial general intelligence (AGI) benefits all of humanity. ChatGPT is one of OpenAI's language models that is part of the GPT (Generative Pre-trained Transformer) series. It's critical to take note of that while OpenAI creates and keeps up with the models, the genuine connections with ChatGPT happen through different points of interaction, for example, the OpenAI site or outsider applications that incorporate the OpenAI Programming interface. OpenAI looks to offset transparency and availability with dependable use, and it keeps on refining its models and arrangements in view of progressing exploration and client criticism.

Microsoft

Microsoft has been effectively engaged with advancing mindful and moral artificial intelligence advancement. Microsoft has framed its way of dealing with simulated intelligence in its man-made intelligence standards and has focused on keeping a bunch of rules to guarantee the mindful utilization of man-made intelligence innovations. Their efforts to incorporate these principles into their products and services reflect their commitment to responsible AI. It is essential to keep in mind that Microsoft may alter its policies and guidelines over time.

European Union (EU)

The EU has been actively working on guidelines and regulations to govern the development and use of artificial intelligence (AI). One significant initiative is the "Artificial Intelligence Act" proposed by the European Commission. The parliament's need is to ensure that man-made intelligence frameworks utilized in the EU are protected, straightforward, recognizable, non-oppressive and harmless to the ecosystem. Simulated intelligence frameworks ought to be administered by individuals, instead of via computerization, to forestall destructive results.

Possible Solutions

Policy Frameworks

When regulating AI, establishing a dedicated body is crucial, if not essential. With the rapidly evolving landscape, regulations must adhere to regular evaluation and updates to keep pace with innovations. Flexibility in regulation would allow adjustments based on emerging ethical and

technological concerns. As developers, researchers and technology experts emerge as the most experienced, possessing knowledge of its characteristics and possible danger, engaging these actors in the regulatory process could help attain effective and practical policies while avoiding unnecessary restrictions on technical freedom. They are capable of ensuring the consideration of limitations and technical possibilities whether in general AI as well as for AWS, with the goal to guide data privacy, algorithmic transparency, accountability, safety standards, and ethical AI usage.

Means of Information

Both within a company and in relation to the users, transparency and explainability campaigns can aid in building trust. Encouraging the installation of industry-wide standards can guide researchers, companies and policymakers to set benchmarks for safety, security and ethical AI design. Making algorithms understandable could improve communication and clarity between company directors and their shareholders.

Users are also involved in the extensive data available to AI, so safeguarding it against breaches and misuse is imperative in accordance with UNESCO's Recommendation on the Ethics of Artificial Intelligence. The general public would furthermore benefit from the possibility of making informed decisions and participating in discussions shaping AI policies.

Enforcing an Ethical Guidelines

Guidelines for the development and deployment of AI are important in defining the ethical principles that AI systems must adhere to fairness, transparency, accountability, and inclusivity. Conducting impact assessments and ensuring cybersecurity measures can identify potential risks and prevent misuse of AI.

Especially in the perilous context of AI in the military, its creators and users must have a definite responsibility in its conduct to compensate for misapplications and misfires of AWS. Humans must retain ultimate responsibility for decisions involving the use of force therefore an international definition must be devised to mark a clear distinction. This includes principles emphasizing human control, proportionality, accountability, and compliance with international humanitarian law. Implementing rigorous risk assessment frameworks to evaluate the potential risks and impacts of AI applications in this setting is therefore quite literally of vital importance. Establishing predefined guidelines to follow in the

defense and space industry would certify a standard of safety but may hinder the policies of the free market.

Bibliography

(2023, October 30). , - YouTube. Retrieved December 28, 2023, from

<https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/?sh=5f1ca78a233e>

Bishop, C. (2023, May 23). *Chatbots vs. conversational AI: What's the difference?* Zendesk.

Retrieved December 28, 2023, from <https://www.zendesk.nl/blog/chatbot-vs-conversational-ai/#georedirect>

EU AI Act: first regulation on artificial intelligence | News. (2023, June 8). European Parliament.

Retrieved December 28, 2023, from

<https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

What are the risks of artificial intelligence (AI)? (n.d.). Tableau. Retrieved December 28,

2023, from <https://www.tableau.com/data-insights/ai/risk>

Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, (July 2022)

[CCW/MSP/2022/7](https://www.un.org/disarmament/convention-weapons/2022/07/)

[A Military Drone With A Mind Of Its Own Was Used In Combat, UN Says](https://www.forbes.com/sites/davidhambling/2023/10/17/ukraines-ai-drones-attack-russian-forces-without-human-oversight/?sh=f309b3366da6)

<https://www.forbes.com/sites/davidhambling/2023/10/17/ukraines-ai-drones-attack-russian-forces-without-human-oversight/?sh=f309b3366da6>

<https://moderndiplomacy.eu/2022/08/10/a-war-where-the-machine-decides-who-to-kill-laws-wars/>