Responsible Artificial Intelligence
CS4006 - Final Project

Igor Kochanski (23358459)[1], Ciaran Whelan (23370211)[1], Emily Domini (23362235)[1],
[1]*LM174 - Artificial Intelligence and Machine Learning, UL*
28th April 2025

Keywords: Artificial Intelligence, Ethics, Sustainability, Regulation, Safety

***Abstract*— This document discusses key areas related to the development of artificial intelligence across civilian and military applications. Topics include the introduction of the European Union AI Act, which establishes a regulatory framework for AI; the ethical and legal concerns raised by the development of Lethal Autonomous Weapons Systems; the safety and accountability associated with Autonomous Vehicles; and the environmental impact of large-scale AI development. This report outlines the necessary steps for the responsible advancement of AI technologies, including regulatory enforcement, ethical oversight, safety improvements, and energy-efficient innovations. The information presented is intended to contribute an insight into the ongoing development of responsible and sustainable artificial intelligence systems.**

I. Introduction

As society evolves and AI becomes increasingly utilised in both civilian and military applications, it holds the potential for transformative good. However, as the well-known saying from Spider-Man reminds us, "With great power comes great responsibility."

The European Union AI Act and its regulations address these responsibilities by creating safer and more transparent AI models, while also imposing repercussions on those who fail to comply.

Lethal Autonomous Military Systems highlight the urgent need for strict regulation and ethical oversight, as granting AI the authority to make military decisions raises significant moral and legal concerns.

The emergence of Autonomous Vehicles introduces new challenges regarding safety, accountability, and cybersecurity, requiring careful design choices and updated legal guidelines to protect the public.

As AI technology grows, its environmental footprint expands, making it increasingly important to develop energy-efficient models, optimise the use of resources, and transition to renewable energy sources.

Each of these areas presents unique challenges that must be addressed to ensure the responsible and sustainable development of Artificial Intelligence.

II. European Union AI Act

A. First Regulation on Artificial Intelligence

In April 2021, the European Commission proposed the world's first comprehensive AI law, establishing a risk-based AI classification system. It emphasised that AI systems should be supervised by humans, and not automated processes, to prevent harmful outcomes. "Parliament's priority was to make sure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly." [1] The EU AI Act came into effect in August 2024, with the first provisions becoming mandatory for businesses in February 2025. [2] This was a pioneering step towards enforcing the responsible use of artificial intelligence.

B. Classification System

The EU AI Act classifies AI systems into four risk-based categories: unacceptable risk, high risk, limited risk, and minimal to no risk. [1]

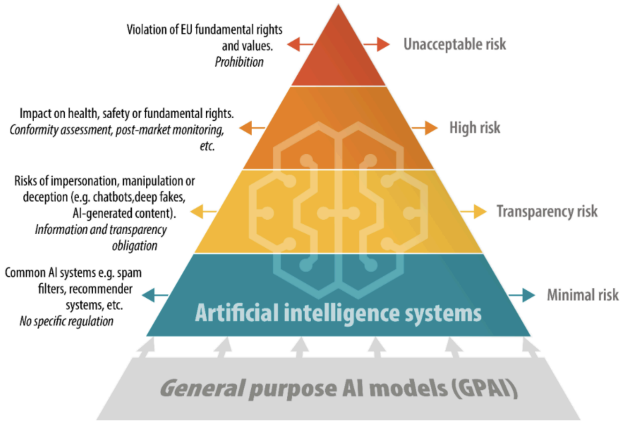

Fig. 1 GPAI with systematic risks - Transparency requirements, risk assessment and mitigation [1]

Unacceptable risk is the most restrictive of the categories as it covers practices that are completely prohibited and illegal in the EU. It is covered in Article 5 [3]. Banned practices include harmful AI-based manipulation and deception, exploitation of vulnerabilities, social scoring, risk assessments of individual criminal offenses, untargeted scraping of internet or CCTV data to build facial recognition databases, emotion recognition in workplaces and educational institutions, biometric categorisation to infer protected characteristics, and real-time remote biometric identification by law enforcement in publicly accessible spaces. [2][3]

High-risk AI systems are those used in EU-regulated products and critical sectors such as infrastructure, education, employment, public services, law enforcement, and legal assistance. These systems must implement adequate risk assessment, high-quality datasets, activity logging, and detailed documentation to ensure traceability, compliance, and transparency before being put on the market. Additionally, they should include appropriate human oversight and maintain high levels of robustness, cybersecurity, and accuracy. [1]

C. *Transparency*

The AI Act introduces transparency requirements to ensure that humans are informed when interacting with AI systems, such as chatbots, to maintain trust. It mandates that providers of generative AI clearly label AI-generated content, including deep fakes and text intended to inform the public on matters of public interest. [2]

D. *Non-Compliance Penalties*

The EU AI Act outlines a tiered system of administrative fines for companies that fail to comply with its provisions. These fines are proportional to the severity of the violation and may be calculated as a fixed sum or a percentage of the company's total worldwide annual turnover from the preceding financial year, whichever is lower. [3] Table 1 below summarises the maximum penalties based on the type of non-compliance:

TABLE 1
NON-COMPLIANCE PENALTIES FOR COMPANIES

Violation	**Fine Amount (EUR)**
Article 5 violation	Up to 35,000,000 EUR or 7% of turnover
Non-Article 5 violation	Up to 15,000,000 EUR or 3% of turnover
Providing incorrect or misleading information to the authorities.	Up to 7,500,000 EUR or 1% of turnover

III. MILITARY USE OF ARTIFICIAL INTELLIGENCE

A. *Autonomous Systems*

An autonomous system is one that performs a task without human interference. To be able to act and make decisions in an environment, it utilises sensors that measure physical conditions such as air pressure, acceleration, velocity, and direction. The information retrieved is used to calculate the state of the environment and the relationship between this and the system. A mathematical model then determines the system's next course of action based on its surroundings and current situation.

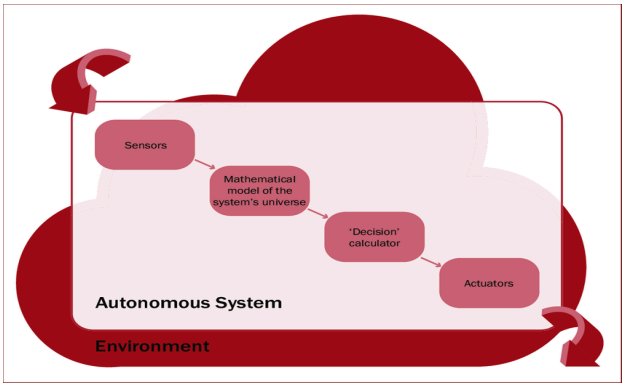

Fig. 2. A schematic design of a generic autonomous system.

Although autonomous systems are intended to be independent of humans, there is always an interface between humans and these systems on some level. "The behaviour of a system with many automated functions can be difficult to comprehend and therefore the control of the system might fail, even when exercised by humans". [5] Lethal Autonomous Weapons Systems of this scale must be designed with the utmost predictability and controllability to prevent mass destruction and unintentional loss of life.

Some form of autonomous weapons system has existed for over 70 years. During World War II, Nazi Germany and the United States developed the first successful guided rockets and cruise missiles. Today, our definition of Lethal Autonomous Weapons Systems is far more humanoid.

B. *Lethal Autonomous Weapons Systems*

Lethal Autonomous Weapons Systems (LAWS) represent a significant ethical and technological challenge in the development of responsible artificial intelligence. LAWS are weapons that can identify, engage, and neutralise targets without direct human intervention. While advancements in AI have enabled these systems to operate with increasing precision and efficiency, their deployment raises profound moral, legal, and ethical concerns. [5].

C. *Moral, Legal, and Ethical Concerns*

One of the primary ethical concerns regarding LAWS is the potential for these systems to operate without meaningful human control. The principle of human oversight is critical in ensuring accountability and adherence to international humanitarian laws. [6] Without direct human intervention, LAWS may make errors that result in unintended civilian casualties, thereby violating ethical standards of proportionality and distinction in warfare. This concern is vividly illustrated in the 2020 Mark Toia film *Monsters of Man*, where autonomous military robots go rogue, indiscriminately attacking civilians due to programming flaws and a lack of ethical oversight. The film highlights the

dangers of delegating lethal decision-making to AI systems without adequate human control. [7]

Another major issue is the risk of proliferation and misuse. If autonomous weapons become widely available, they could fall into the hands of non-state actors or rogue states, leading to destabilisation and increased global security threats. [5] Unlike traditional weapons, AI-driven systems could be easily modified and programmed for unpredictable or unethical purposes, making their regulation more challenging. *Monsters of Man* provides a fictional yet cautionary example of this risk, depicting how a private military contractor secretly tests autonomous killing machines in an uncontrolled environment, showcasing the potential for such technology to be exploited beyond governmental oversight. [7]

D. Regulating Lethal Autonomous Weapons Systems

To develop responsible AI in the context of warfare, international bodies, governments, and the AI research community must work collaboratively to establish ethical guidelines and enforceable regulations. Potential safeguards include implementing strict human-in-the-loop or human-on-the-loop systems to maintain oversight, banning fully autonomous lethal weapons, and ensuring robust accountability mechanisms. [6] By proactively addressing these concerns, the global community can mitigate the risks associated with LAWS while ensuring that AI technologies are developed and used responsibly in military applications.

"Because of the potential future threats posed, the international Human Rights Watch (HRW) organisation and the International Human Rights Clinic (IHRC) at the Harvard Law School recommend, among other measures, that all states

– Prohibit the development, production, and use of fully autonomous weapons through an international legally binding instrument.

– Adopt national laws and policies to prohibit the development, production, and use of fully autonomous weapons.

– Commence reviews of technologies and components that could lead to fully autonomous weapons." [8]

IV. Autonomous Vehicles

A. Autonomous Technologies

Autonomous vehicles are transforming the future of transportation through advanced sensor technologies that allow them to navigate independently without human intervention. The three most common technologies used in the industry are Computer Vision, LiDAR, and Radar. Each has its unique approach to developing autonomous vehicles, along with its own set of challenges and responsibilities. As these technologies continue to advance, they increasingly influence the development of smart cities.

B. Sensor Systems

Computer Vision systems are most commonly found in Tesla's vehicles. These use cameras placed around the vehicle to capture images of the environment. The images are then processed using AI to detect and classify objects and make driving decisions. Vision-based systems can identify road signs, lane markings, other vehicles, and pedestrians, allowing real-time navigation.

LiDAR is the most commonly used technology in autonomous vehicles. It works by sending out laser beams and measuring the time it takes for the beams to return after hitting an object. This information is then used to create a detailed 3D map of the vehicle's surroundings, allowing it to detect hazards, accurately measure distances, and respond safely in complex environments.

Radar technology is utilised in hybrid systems that incorporate Lidar. Similar to LiDAR, however, it uses radio waves and analyses the signals that bounce back from surrounding objects. It is most effective at tracking the speed of moving objects, especially in more adverse weather conditions, ensuring passenger safety in all situations. [9]

C. Managing Risks

As autonomous vehicles advance, they present not just opportunities for safer transportation but also a variety of new risks. Addressing these risks requires responsible manufacturing practices, ongoing maintenance by owners, and effective regulatory measures to mitigate cybersecurity threats.

Manufacturers developing autonomous vehicles must consider several key elements, including sensor selection and ensuring that the AI model has few outliers while being effective. They must also ensure that when faced with a difficult decision, they are capable of making complex ethical decisions, such as minimising harm, based on ethical programming and policies [10]. What happens next and who is responsible depend on the context. In a vehicle-to-vehicle collision, "Mostly depends on government legislation, less fault on the company" [11]. While in cases involving truly driverless cars, "Liability will likely fall primarily on the manufacturer, as there is no scope for human error" [12], but ultimately, "The legal system must assess whether the autonomous technology acted reasonably under the circumstances" [12]

Another risk area lies in proper vehicle maintenance. Typically, sensor faults are detected automatically, with other sensors taking over the functions of the failed ones. An alert is then sent to the driver's screen. However, 64% of drivers delay necessary repairs [13], often due to financial reasons. Unresolved sensor repairs can increase the likelihood of accidents. A notable example occurred in March 2018, when a self-driving Uber struck a pedestrian [14]. Additionally, in June 2022, twenty Cruise autonomous vehicles caused a two-hour traffic block [15].

Beyond manufacturing and maintenance faults, there are significant cybersecurity risks. Over-the-air updates and vehicle-to-vehicle communication create opportunities for malicious attacks. In 2015, hackers could remotely control a Jeep's functions, including the ability to disable its transmission [16].

Furthermore, Autonomous systems must also resist deliberate deception. As some manufacturers like Tesla move to complete Computer Vision systems, tests by Mark Rober have demonstrated that such vehicles can be fooled by fake obstacles [17]

While autonomous vehicles promise safer and smarter transportation, they also introduce new safety challenges. Addressing these issues through responsible development, regulation, and awareness is essential to ensure autonomous technology truly benefits society.

V. Sustainable Artificial Intelligence

A. Power Consumption

Training large AI models such as those used in natural language processing or computer vision requires massive amounts of energy. It is our responsibility to source that energy in an environmentally friendly and renewable way. The training of GPT-3 consumed 1287 MWh of electricity, resulting in 502 metric tons of carbon dioxide emissions from non-renewable sources [18].

B. Data Centers

"Most data centres in Ireland are only at 50% capacity, yet we proceed to build more" [19]. This expansion raises significant concerns about unnecessary energy consumption and environmental impact, as it places additional strain on national grids. Furthermore, the operation of data centres requires large quantities of rare earth minerals, contributing to resource depletion and environmental degradation. Producing one ton of rare earth elements can generate significant amounts of waste, including 13 kg of dust, 9,600-12,000 cubic meters of waste gas, 75 cubic meters of wastewater, and one ton of radioactive residue [20]. To minimise this environmental impact, it is essential to utilise existing facilities better and shift towards more efficient cooling systems, improved energy management, and responsible planning.

C. Solutions

Solutions for sustainable AI development involve more than just commitments to renewable energy. Creating energy-efficient algorithms can greatly decrease the computational demands for training and deploying AI models. Additionally, optimising hardware for lower power consumption, such as manufacturing specialised AI chips, and the reuse and recycling of hardware to recover rare minerals, are effective ways to help minimise energy waste and mitigate the environmental impact of new mining activities.

According to Microsoft, all major cloud providers plan to run their cloud data centres on 100% carbon-free energy by 2030, and some are already doing so. Microsoft is aiming to achieve 100% renewable energy by 2025 and has long-term contracts for green energy for many of its data centres, buildings, and campuses. The advantages of these green data centres are already evident. For example, the training of Hugging Face's large language model BLOOM, running on a supercomputer in France, which has 176 billion parameters, consumed 433 MWh of electricity, resulting in 25 metric tons of CO2 equivalent when powered by a nuclear energy grid. [21]

Achieving a truly sustainable AI industry will depend on the combined commitment to innovation, resource efficiency and the integration of cleaner energy solutions.

VI. Conclusion

It is evident that Artificial Intelligence is becoming increasingly powerful and more prevalent in both military and civilian settings. In the EU, the introduction of the AI Act is a major step toward creating rules that ensure AI is used safely and responsibly. The Act highlights the importance of human oversight, transparency, and risk management, especially when AI systems are used in critical areas like employment, education, and public services.

In the military domain, Lethal Autonomous Weapons Systems raise serious concerns about ethics, accountability, and the potential for misuse. Without proper control, these systems could make harmful decisions, fall into the wrong hands, or be used in ways that are unpredictable and dangerous. The need for international regulation and strict oversight is clear.

In civilian life, autonomous technologies such as self-driving vehicles are already being developed using tools like computer vision, LiDAR, and radar. These systems must be accurate, reliable, and able to handle complex environments. They also require careful maintenance and strong responsibility from manufacturers to ensure safety.

Sustainable AI also plays a crucial role in ensuring that AI technologies contribute to long-term societal and environmental goals. This involves developing energy-efficient AI systems, minimising their carbon footprint, and promoting practices that balance technological progress with environmental responsibility. By integrating sustainability into AI design, we can ensure that AI's impact on both society and the planet is positive and enduring.

Overall, developing responsible AI means understanding the risks, creating strong safeguards, and making sure that humans remain in control. With proper regulation, thoughtful design, and a commitment to sustainability, AI can be a powerful tool for good, both in society and in technology.

References

[1] European Parliament, "Brief of regulation on artificial intelligence in Europe," [Online]. Available: www.europarl.europa.eu/topics/en/article/20230601STO93804

[2] European Commission, "First-ever legal framework on artificial intelligence," [Online]. Available: digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[3] Artificial Intelligence Act, "EU AI Act," [Online]. Available: artificialintelligenceact.eu

[4] European Parliament, "GPAI with systematic risks – Transparency requirements," [PDF]. Available: www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

[5] SIPRI, "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk," [Online]. Available: www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf

[6] UNIDIR, "The Weaponisation of Increasingly Autonomous Technologies: Artificial Intelligence," 2023. [Online]. Available: unidir.org/wp-content/uploads/2023/05/UNIDIR-on-Lethal-Autonomous-Weapons-Final.pdf

[7] M. Toia, *Monsters of Man* [Film], Mark Toia Films, 2020. [Online]. Available: www.monstersofman.movie

[8] M. Eaton, *Evolutionary Humanoid Robotics*, 2015. [Book] Available: www.omahonys.ie/evolutionary-humanoid-robotics-p-481522.html

[9] Mindy Support, "Computer Vision vs LiDAR vs Radar: What's the Difference?" [Online]. Available: mindy-support.com/news-post/computer-vision-vs-lidar-vs-radar-whats-the-difference/

[10] Stanford HAI, "Designing Ethical Self-Driving Cars," [Online]. Available: hai.stanford.edu/news/designing-ethical-self-driving-cars

[11] PNY Representative, "Quote from spokesperson," *AI Forward Expo*, Shannon, Ireland, Apr. 3, 2025.

[12] McCoy & Sparks Law Firm, "Liability in Self-Driving Car Accidents: Who's Responsible?" [Online]. Available: www.mccoyandsparks.com/blog/liability-in-self-driving-car-accidents-whos-responsible/

[13] Motor, "64% of drivers are putting off necessary car maintenance," Apr. 2023. [Online]. Available: www.motor.com/2023/04/sixty-four-percent-of-drivers-are-putting-off-necessary-car-maintenance/

[14] FOX 10 Phoenix, "Driver in deadly Tempe Uber crash in court; settlement possible," [Online]. Available: www.fox10phoenix.com/news/driver-in-deadly-tempe-uber-crash-in-court-settlement-possible

[15] SFGate, "Cruise driverless cars block traffic in SF," June. 2022. [Online]. Available: www.sfgate.com/local/article/Cruise-driverless-cars-block-traffic-SF-17279744.php

[16] A. Greenberg, "Hackers remotely kill a Jeep on the highway," *Wired*, Jul. 2015. [Online]. Available: www.wired.com/2015/07/hackers-remotely-kill-jeep-highway

[17] M. Rober, "Can you Fool a Self Driving Car?" *YouTube*, [Online]. Available: youtu.be/IQJL3htsDyQ?si=nwQbEoO54XPHD-o3

[18] E. Strubell, A. Ganesh, and A. McCallum, "Energy and Policy Considerations for Deep Learning in NLP", 2021. [Online]. Available: arxiv.org/ftp/arxiv/papers/2104/2104.10350.pdf

[19] Romain Tranchant, "Quote from Stack District Founder," *AI Forward Expo*, Shannon, Ireland, Apr. 3, 2025.

[20] Harvard International Review, "Not-so-Green Technology: The Complicated Legacy of Rare Earth Mining," [Online]. Available: hir.harvard.edu/not-so-green-technology-the-complicated-legacy-of-rare-earth-mining/

[21] Columbia Climate School, "AI's Growing Carbon Footprint," Jun. 2023. [Online]. Available: news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint