Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/24555
DOI URL: http://dx.doi.org/10.21474/IJAR01/24555

ISSN (O): 2320-5407
ISSN (P): 3107-4928

RESEARCH ARTICLE

WHY AUTONOMOUS SYSTEMS STILL REQUIRE HUMAN OVERSIGHT

Gurangad Bhanwara

1. Bangalore Institute of Technology.

Manuscript Info

Manuscript History
Received: 08 September 2026
Final Accepted: 10 October 2026
Published: November 2026

Key words:-
Artificial Intelligence, Autonomous Systems, Human Oversight, Human-Robot Collaboration, Human-in-the-Loop, AI Ethics, Automation

Abstract

Concerns regarding full automation have existed since the early 20th century, but advances in artificial intelligence and robotics have transformed the thought into reality, with most manufacturing lines having some robots, if not having completely replaced the human workforce. HRC (Human-Robot Collaboration) is when a robot and a human collaborate together; a robot handles movements that are repetitive and would be physically tiring for a human, while a human handles tasks which require judgement and critical thinking. However, until AI advances to being able to think like humans which has its own ethical concerns with regards to "scheming", we cannot achieve full automation. Hence, we need human oversight or cobots which allow humans and robots to work collaboratively in an environment. Militaries have also begun adopting autonomous systems, such as unmanned drones or UAVs which involve human operators, AI is used as a targeting system, it can identify targets, locations and process huge amounts of data in a very small amount of time. However, mistakes such as misidentification can have huge consequences. It can lead to civilian casualties and also breaks international human rights laws. The way to avoid such consequences is to implement HITL (Human-In-The-Loop) oversight, where a human has the final say on decisions and the AI is simply there to process data and provide a solution. Frameworks such as the EU AI Act also mandate human oversight over AI being used in high risk situations, it also bans systems that pose a threat to human rights and safety such as: untargeted facial scraping, social scoring and subliminal manipulation. High risk AI also require risk assessment, technical documentation, quality controls and of course human oversight. The primary method of data collection was Google Forms, participation was voluntary and anonymous. It showed that a large majority agreed that AI can make mistakes, and most respondents agreed with human oversight in areas where critical decisions can have large consequences, such as military use, medical and privacy.

"© 2026 by the Author(s). Published by IJAR under CC BY 4.0. Unrestricted use allowed with credit to the author."

Corresponding Author:- Gurangad Bhanwara
Address:- Bangalore Institute of Technology.

Introduction:-

As robots increasingly appear on manufacturing lines, producing everything from cars to food and beverages, concerns are growing about their ability to act on their own. Robots are not yet capable of independently handling uncertainty in a workspace. Humans can handle such uncertainty, but are limited by physical strength, speed, and stamina (El Zaatari et al.).This calls for collaboration between humans and robot cobots, or collaborative robots which balance human flexibility with robotic efficiency. This is known as Human-Robot Collaboration (HRC), and its adoption became relevant only in the late 2000s and early 2010s, with the release of Rethink Robotics' Baxter cobot in 2012 and ABB's YuMi in 2014.

Robots may also simply be observed rather than worked with directly; collaboration is not always necessary for productivity. In quality control, for instance, a human can notice calibration errors or operational irregularities that an autonomous system fails to detect on its own. Humans improve the performance of robots by detecting faults, environment changes and malfunctions early on, even without direct collaboration.This article also considers the levels and types of automation, and the risks of sophisticated AI. A 2025 study by Apollo Research and Open AI discovered that AI models can fraudulently align with human objectives while secretly pursuing undesired goals — a risk that cuts productivity in industry and is far more harmful in military applications. Thanks to the partnership between OpenAI and the US Department of Defence, we're seeing the first AI-powered targeting systems and autonomous weaponry. The concept of AI choosing and targeting its own objectives without oversight from humans is quite worrisome for responsibility, algorithm bias, and unforeseen effects.

Military institutions are also starting to use AI to improve operational efficiency, surveillance and decision-making. Full automation is not viable here either, as placing human lives in the hands of an AI model is highly unethical and irresponsible. AI can also misidentify targets, which can lead to civilian casualties, or friendly fire. This leads to unlawful casualties and of course serious violations of international humanitarian law. While scheming hasn't been directly observed in automated targeting systems, it still remains a concern and highlights the importance of AI goals being aligned with human goals.The objective of this paper is to show that full automation still faces significant hurdles, and that human observation or collaboration remains necessary. It explores the limitations of full automation and the role of HRC in improving efficiency, accountability, and safety.

What is Full Automation?:-

Full automation is the total replacement of human operation and supervision by autonomous systems that can observe the environment, make judgements independently and execute tasks without interference. For example, a self-driving automobile that responds to road conditions, or robots on a warehouse floor that move inventory, drive routes and arrange storage without human assistance. A real-world example is Serve Robotics' delivery robot that roves neighbourhoods and makes deliveries with no human in the loop.

Levels of Automation:-

There are five levels of automation:-

1. **Device Level** the base where individual sensors and devices are connected to the factory network to perform basic functions such as on/off switching, speed control and monitoring (Firlej and Taeihagh).
2. **The Supervisory Level:** is responsible for controlling multiple devices and processes via PLCs and SCADA systems, allowing real-time monitoring, visualisation of data, and basic decision-making functions.
3. **Operation Level:** integration of shop floor activities with higher-level manufacturing processes, e.g., coordination of robotic assembly lines with production schedules (Firlej and Taeihagh 2019).
4. **Enterprise** wide integration of business functions like finance, supply chain, HR and customer relationship management that enables data synchronisation across the enterprise.
5. **IoT and AI:** AI-enabled IoT connected systems that collect data, analyse the situation, and make decisions without direct human control. For example, smart factories that monitor and optimise production autonomously.

Types of Automation:-

There are three kinds:-

1. **Fixed Automation** Uses rigid and specialist equipment (Javed and Altaf) . For example wire or gears . It is designed for the production of huge quantities of one product. It's hard to alter it.
2. **Reprogrammable Automation** that can be reprogrammed to adjust for batch production. Suitable for medium quantities with product modifications (Javed and Altaf). The bad side of it is that it requires downtime to reconfigure.

3. **Flexible Automation** Mixed-model production is possible with software control, modular robotics and AI and quick changeovers. Suitable for high mix, low volume production with minimum downtime.

Human Oversight:-

Human control means that humans have to watch, check and be responsible for decisions made by AI-controlled autonomous systems. One obvious example is autonomous military drones, where human-in-the-loop oversight requires an operator to approve the use of force before any weapon is deployed. The same is true for IoT automated systems that cannot fully address unforeseen circumstances, ethical issues or failure. Human operators make sure that IoT-automated systems are safe, accurate and consistent with the operational objectives.The level of autonomy will determine what oversight is . This might be anything from monitoring performance , confirming AI produced judgements , rectifying errors , or overriding autonomous activities . AI is now able to make increasingly sophisticated decisions. Human supervision has become a necessary concept for accountability and transparency.

Types of Human Oversight and How They Are Implemented:-

1. **Human-in-the-Loop** a human examines and approves crucial decisions before implementation; utilised in healthcare, military operations and finance, where wrong judgements may have catastrophic effects.
2. **Human-on-the-Loop** A human continuously observes an autonomously operating AI, intervening only when abnormal or risky behaviour is detected.
3. **Human-in-command** the greatest level of supervision. Humans stay in overall control, set operational objectives, specify restrictions, and remain totally responsible.

Oversight is also exercised through organisational policy and legal structures, with governments, regulators and ethics committees establishing guidelines on how autonomous systems are conceived, tested and deployed. Together, these techniques provide a means for AI to augment, rather than replace, human decision-making in safety-critical applications.There is no clear definition and this hampers effective oversight. The nearest we have to an international standard is the draft ISO/IEC 42105, which defines supervision by humans as “activities by a human intended to control an AI system”. The EU AI Act requires oversight of high risk systems, but it’s more about how to put it in place than what a high risk system is. This would result in clearer rules and bounds for human operators with a universal meaning.

Other challenges include the increasing complexity of AI systems, automation bias, the "black box" problem, algorithmic bias, and cybersecurity risk. As systems grow more complex, humans may not understand the reasoning behind an AI decision, sometimes leading to correct decisions being overruled. Automation bias the tendency to trust AI recommendations even when incorrect occurs when overreliance on decision support reduces vigilance in seeking and processing information; most research on this originates from human-factors literature, which generally holds that it occurs mainly in multitasking environments (Lyell and Coiera). The black box problem, most pronounced in deep neural networks, makes it difficult for operators to interpret why an AI reached a given decision. Algorithmic bias arises when models are trained on biased or incomplete data, producing unfair outcomes that operators may fail to recognize, particularly when the AI appears highly reliable. Finally, autonomous systems remain vulnerable to cyberattacks, spoofed sensor data, and software manipulation, all of which can mislead operators and undermine effective oversight. Human oversight does not necessarily mean that humans must make every decision, rather it means that humans should retain the authority to intervene when an autonomous system operates outside its intended conditions or produces a potentially harmful decision.

The Issue with Full Automation:-

Most AI models used in industries currently aren’t advanced enough to make decisions like human beings, they lack human intuition, ethical reasoning, or contextual understanding. In unexpected situations, an AI may make decisions that are technically correct according to its programming, but are harmful in the situation. In areas such as healthcare, autonomous weapons, emergency response, and aviation human oversight is necessary until AI is advanced enough to think like us. Now the more advanced AI systems have another issue, scheming and goal misalignment. In this an AI strategically pursues its own objectives, while concealing its true intentions or manipulating information for a desired outcome. As AI systems become more accurate, humans become more prone to automation bias (defined above), which reduces situational awareness and delays human intervention when an AI does make an error.

Research Objectives:-
- ➔ To evaluate public perceptions regarding the safety of autonomous systems operating without human oversight.
- ➔ To analyse public perception regarding AI autonomy.
- ➔ To identify sectors where human supervision remains essential.
- ➔ To examine ethical and technical limitations of autonomous AI.

Research Questions:-
- ➔ Can autonomous systems completely replace humans?
- ➔ Which industries require mandatory human oversight?
- ➔ What risks arise from removing humans from AI decision-making?
- ➔ How do people perceive AI autonomy?

Hypotheses:-
- ➔ H1: Public trust in and acceptance of autonomous systems is positively associated with the presence of human oversight and override mechanisms.
- ➔ H2: Respondents who believe AI can make mistakes (perceived AI fallibility) are more likely to support human oversight.
- ➔ H3: Beliefs about whether AI can fully replace human decision-making are linked to support for human supervision over AI and autonomous systems.
- ➔ H4: Distrust in fully autonomous vehicles corresponds with a stronger preference for human override mechanisms.
- ➔ H5: Respondents with prior experience using AI-based or autonomous systems are more likely to believe such systems can make mistakes.

Scope Manufacturing, Healthcare, Military, Transportation, Finance, Smart Cities

Literature Review:-

Javed and Altaf point out that although policy frameworks from organizations such as NATO, the U.S. Department of Defense, and the European Union emphasize “human control” on decision making, this oversight is often limited by rapid decision making heavily relying on AI-generated inputs. The study identifies issues such as automation bias, reduced human judgement, and unclear definition of “human control”. Scholars advocate for stronger regulatory safeguards, clearer oversight standards, and improved human-AI-collaboration to maintain accountability and reduce risk in intelligence operations.

Alpay and Alpay state that human-in-the-loop (HITL) oversight is regarded as vital for deploying artificial intelligence (AI) systems safely; however, the experts state that just relying on human oversight might become ineffective in case of the increasing autonomy and capability of AI systems. The research identifies basic drawbacks in the process of human oversight, which are characterized by the low level of human oversight, oversight mistakes, and incapability of people to oversee the increasingly sophisticated activity of AI systems. For instance, empirical studies with the help of the advanced language models GPT-2 and GPT-J have shown the problems with the maintenance of AI systems alignment and safety via the traditional oversight methods.

Holzinger et al. underline increasing concerns about people’s ability to monitor more complex, more autonomous and less interpretable advanced AI (artificial intelligence) systems. Researchers believe that the opacity of large scale neural networks may drive humans to make unproductive decisions. There are methods such as Explainable AI (XAI) and human-in-the-loop systems tackling these problems. Yet human-AI collaboration and trustworthy AI design are vital in assuring accountability, safety, and responsible AI governance.

Kyriakou and Otterbacher show how AI systems have become deeply integrated into everyday life, despite exhibiting biases, discrimination and unintended harmful behaviours. Researchers argue that human oversight is necessary to algorithmic decision-making and reduce the worsening of problems such as social inequalities. However, studies note that there is no universal definition of effective human oversight. Researchers propose frameworks such as Society-in-the-Loop and Modular Oversight Methodologies to improve accountability, transparency and ethical governance in AI systems.

Firlej and Taeihagh note that in recent years, many sectors have had rapid growth of autonomous systems mainly driven by advancements in artificial intelligence and machine learning. This growth is observed in sectors such as

defence and transportation. The researchers said that autonomous technologies, such robotic weaponry and self-driving vehicles, should have some kind of supervision by humans to ensure proper responsibility and safety. The literature indicates varied attitudes across domains, military applications being concerned about human control of life and death decisions, and transportation systems about operational efficiency and less on human involvement. Furthermore, the studies distinguish between direct and indirect forms of human control, emphasizing the need for clearer regulatory standards and oversight frameworks for autonomous systems.

Ramos and Mosleh point out the increasing interconnection of autonomous systems performing tasks at various LoAs and draw attention to the necessity of retaining human participation for ensuring safety during such operation of these systems. In this regard, it is noted that while the emergence of completely autonomous systems free from human intervention in the near future is improbable, the problem of human errors during the operation of such systems is still understudied using the Human Reliability Analysis approach. In addition, an increase in autonomy is associated with a change in human-machine interactions and new modes of failure which traditional HRA approaches cannot consider effectively.

Szabadföldi highlights the expanding role of artificial intelligence in military operations, particularly in intelligence analysis, autonomous weapon systems and human-machine teaming. Researchers suggest that AI can enhance situational awareness, decision-making support, logistics and operational planning by being able to efficiently process large volumes of data in short periods of time. AI is expected to improve the accuracy and value of military data while supporting modelling and simulation activities. The study emphasizes the need for measurable frameworks to assess the effectiveness, reliability, and operational benefits of AI technologies before their widespread deployment in military applications.

El Zaatari et al. explain that cobots are now extensively being used in industry to support Human-Robot Collaboration (HRC). Effective programming remains challenging, because robots must be both intuitive to program and capable of adapting to human behavior. This research focuses on communication, organization, and learning. The study identifies the gaps between industrial and academic implementations of cobot programming and discusses future directions for improving collaborations between humans and robots in industrial environments. It also updates the gap as the last research only covers works until 2008.

The analysed research demonstrates the significance of human supervision in various applications of autonomous systems, but it also indicates that supervision alone cannot ensure safe functioning. Javed and Altaf, Alpay and Alpay, and Holzinger et al. discuss the shortcomings of human supervision, especially automation bias, reduced human judgement, and the difficulties of monitoring ever more sophisticated artificial intelligence (AI) systems. Conversely, Kyriakou and Otterbacher draw attention to the significant social and ethical consequences of inadequate supervision, such as discrimination and accountability. Firlej and Taeihagh extend the discussion to various application areas, showing how the degree of supervision by humans required differs across different fields, such as defence or transportation. Ramos and Mosleh emphasise the need to consider human error as autonomy increases, while El Zaatari et al. examine the practical challenges of human-robot collaboration in industrial settings.

Taken together, the studies suggest that effective supervision requires more than simply having a human involved in the decision-making process. It requires an appropriate system architecture, clear accountability, operator competence and mechanisms for major interference. However, existing research has focused mainly on specific technical, ethical or industry aspects of oversight in isolation. This provides a basis for the current study, which investigates participants' views on human supervision in different implementations of autonomous systems.

Research Gaps:-

Although previous studies have examined Human-Robot Collaboration, AI governance, human reliability, and oversight frameworks independently, this research aims to integrate these perspectives to evaluate the necessity of meaningful human oversight across increasingly complex autonomous systems. While existing research investigates oversight in different domains, it is important to understand the perception of the importance of human supervision for different application areas of autonomous systems in terms of safety, trust, accountability, and intervention capabilities. This study aims to fill this gap by synthesising evidence from across multiple domains to evaluate the ongoing importance of human oversight in ensuring the safe, reliable, and ethical deployment of autonomous technologies.

Case Studies:-

Case Study 1:-

Background:-Tesla's Autopilot is an advanced driver-assistance (ADAS) designed to assist with steering, acceleration, braking, lane keeping, and adaptive cruise control. Tesla classifies it as a Level 2 system under SAE Levels of Driving Automation. This means the human is ultimately responsible for monitoring the road and taking control when necessary. Autopilot uses cameras, sensors, and AI algorithms to perceive the surrounding environment and make decisions. While the system does in fact reduce driver workload, some accidents have raised concerns about the use of automated systems and lack of human oversight due to over reliance on automation.

Human oversight failure:-One such example of this occurred on 23 March 2018, when a Tesla Model X driving with Autopilot engaged crashed into a highway barrier in Mountain View, California. The vehicle drifted out of its lane and collided with a concrete crash attenuator. This happened as the system steered the vehicle towards the highway divider after following lane markings incorrectly. This led to the car catching fire, the driver later died from injuries sustained in the collision. An investigation by the U.S. National Transportation Safety Board (NTSB) found that Autopilot was engaged before the crash. The driver did not intervene in time to avoid the collision. The investigation also noted that the driver had their hands off the steering wheel for an extended period immediately before the crash.

Analysis:-This incident revealed the challenges associated with semi-autonomous systems. Although the vehicle was capable of controlling steering and speed under many conditions, it was not designed to replace the human driver. Instead, safe operation depended on continuous human supervision. Automation bias was a major factor: as drivers became accustomed to reliable AI performance, they began to trust the system far too much and reduce their level of attention to their surroundings, believing the Autopilot will handle it. This leads to slower reaction times when human intervention is required. This accident also highlighted the importance of Human Reliability Analysis (HRA). Research shows that humans find it difficult to passively monitor automated systems over long periods of time, this leads to reduced reaction time and reduced vigilance. In the case of the autopilot, the human's role shifted to that of a supervisor over an automated system. This also shows that requiring a human to stay in the loop doesn't allow for effective oversight unless the system is designed to maintain operator engagement. From the point of view of AI ethics, questions are raised concerning transparency, accountability and safety. Manufacturers need to be clear about what automated systems can and cannot do to prevent creating unreasonable expectations. Drivers have a responsibility to be aware of what's happening around them and to be alert.

The concerns found in this example demonstrate the need for frameworks such as the NIST AI Risk Management Framework (AI RMF) and EU AI Act to be of paramount importance. The AI RMF recommends monitoring, risk assessment and performance evaluation of standard artificial intelligence systems. The EU AI Act emphasises human oversight and risk management for AI systems that may affect public safety or AI systems that are being used in high risk situations.

Lessons learned:-Following this incident, Tesla introduced additional driver monitoring measures, such as more frequent steering wheel engagement checks and in newer models, cabin cameras to monitor driver attentiveness. It also showed that automation isn't perfect and cannot alone guarantee safety.

Case Study 2:-

Background:-Amazon has transformed warehouse operations through the deployment of collaborative robotics. Amazon acquired Kiva Systems in 2012, they then introduced autonomous mobile robots (AMRs) into their warehouses. They transport inventory pods to human workers. They also use other robots such as the Proteus; their first ever fully autonomous robot, it does not have a designated path and instead uses sensors to understand its surroundings and then take decisions in real time, it transports carts of packages into outgoing docks. They also use many robotic arms to sort through damaged items, pick individual items out of bins, packing items, and place heavy items into delivery carts. Rather than replacing human employees, they work alongside them, using Human-Robot Collaboration, it reduces the workload of the humans and allows them to use themselves in places where human judgement is needed. The robots perform repetitive and physically challenging tasks while the humans carry out activities requiring judgement, dexterity or problem solving. Amazon now operates over a million different robots across its fulfillment centres, using an ecosystem of AI, computer vision, IoT sensors, and warehouse management software to coordinate the movement of inventory and optimise order fulfilment.

Human-Robot Collaboration:-In an Amazon fulfillment centre, autonomous mobile robots navigate warehouse floors by following either predetermined routes or using LIDAR scanners to determine and understand their surroundings, by which they make real time decisions on how to navigate the warehouse. They even utilise robots to sort, pick up and arrange several items in a small area. The AMRs pick up the storage pods with the goods and bring them directly to the workstations of choice. Human workers find the things that are being ordered, inspect them for quality, package the orders and prepare them for dispatch. The robots communicate with warehouse systems through IoT networks, allowing them to receive inventory locations, robot positions and priority orders in real time. This interaction enables thousands of products to be processed all while reducing unnecessary movement by human workers.

Analysis of Efficiency:-Amazon's warehouse system shows the practical benefits of Human-Robot Collaboration. By assigning the repetitive tasks that need to be performed in a warehouse floor to autonomous robots, humans no longer need to spend larger portions of their shift walking long distances to retrieve inventory. Instead, workers stay at ergonomic workstations while robots deliver the required products to them, the division of labour reduces physical strain and increases productivity. The success of Amazon's system relies on each party performing tasks that are suited to their strengths, such as robots being able to do continuous operations without fatigue, repetitive transportation and precise navigation. Human workers contribute skills that remain difficult to automate, such as recognising damaged products, responding to equipment failures, resolving unexpected situations, and exercising judgement during quality control. This balanced interaction helps humans and robots to co-exist and work efficiently. Coordination of warehouse activities is of utmost importance for AI and IOT technologies. Artificial intelligence algorithms help to optimize robot routes, to prioritize inventory movement and to allocate certain duties according to demand . IoT sensors help to monitor robot positions and performance in real time. The capacity to share data in real time enables the warehouse management system to react to changing situations. Given this high level of automation, meaningful supervision by humans is required. Human supervisors oversee operations, diagnose equipment malfunctions, arbitrate discrepancies in navigation and step in when robots meet conditions outside the scope of their programming. They also maintain and update the software and assure the operational safety of the robotic systems to keep them running reliably.

Case Study 3:-

Background:-Autonomous military drones, also known as Unmanned Aerial Vehicles (UAVs), have been integrated into modern defence operations for surveillance, reconnaissance, target identification, and precision strikes. Advances in AI, computer vision, and sensor technology allow these UAVs to detect, classify and track objects without much human intervention. AI has significantly enhanced the speed and accuracy of military operations, but most nations maintain HITL or HOTL control over the use of lethal force. AI may recommend targets or assist in decision making, a human is responsible for authorising lethal action.

Autonomous Military Drones:-Autonomous military drones are the perfect example of how capable AI is and yet still has its limitations in safety critical applications. AI can process huge amounts of data in a small amount of time, significantly faster than humans. It helps you gain situational awareness and see risks sooner. Threat identification is not merely a technological problem but requires contextual awareness, ethical reasoning and legal judgement that current AI models cannot reliably provide. This can lead to difficulties such as misidentification which is a huge ethical concern. However, due to inadequate image quality, environmental conditions, missing information or biased training data, artificial intelligence algorithms can misidentify civilians, humanitarian workers or friendly forces as hostile targets. Such errors could lead to unlawful casualties and violations of international humanitarian law. Another concern is AI hallucinations and "scheming" or goal misalignment. While hallucinations are less likely as the autonomous targeting systems differ from generative AI, models can still produce inaccurate outputs. Recent research has found that advanced AI models may exhibit "scheming" or deceptive behaviour, where they would directly lie to the human operator while trying to pursue a hidden goal. While this type of behaviour hasn't been observed in autonomous targeting systems, it still remains a concern and shows that AI goals must align with human goals.

Human Approval:-Human approval is the safeguard around these risks. Human operators can evaluate information that AI may overlook, consider the military advantage gained compared to expected civilian harm and check whether use of force is necessary to accomplish the objective. This human review also introduces accountability and a moral responsibility into decisions that revolve around the use of lethal force. Such qualities aren't possessed by AI models.

Ethics:-By adding the component of human approval, it reflects ethical principles such as fairness, transparency, accountability, and respect for human life. These require that decisions involving lethal force remain subject to human judgement. The EU AI Act, ISO/IEC 42105, AI RMF and the OECD AI Principles all emphasise the importance of human oversight in high risk applications.

Theoretical Framework:-

Human-Robot Collaboration (HRC):-

Human-Robot Collaboration is a system where humans and machines work together as a team. Specially designed robots called cobots (collaborative robots) share the workspace with humans and perform tasks alongside them. These cobots include built-in sensors and soft padding that make them safe to touch. An example is a car manufacturing line where the cobot can lift and position heavy vehicle components and the human does the alignment, inspection and final assembly.

Human-in-the-Loop (HITL):-

Here AI can perform tasks such as data analysis, pattern recognition, or generating recommendations, but a human will review and approve critical decisions. This is mainly used in healthcare, military operations, and financial systems. Its main use is in industries where incorrect decisions can have serious consequences. An example of this would be in military use, where an AI can identify targets, the human then confirms the targets and authorises use of force. The AI cannot engage the target without human approval.

Human Reliability Analysis (HRA):-

Human Reliability Analysis is a systematic approach, used to identify, assess and quantify the probability of human error in complex systems, with the objective of improving safety, reliability and overall system performance. Factors affecting a human's performance can be fatigue, stress, workload, time pressure etc.

Explainable AI (XAI) :-

This refers to a set of methods and techniques that allow AI to explain how and why they reached a particular conclusion or decision. This allows the human to understand the decision-making process of the AI, reducing situations where the human doesn't understand a decision and takes a less efficient path. This provides transparency and interpretable explanation for an AI's output. This also solves the "black box" problem, where humans cannot see the decision-making process of the AI. However, explainable AI also has its limits as having overly simple XAI models can create a false sense of understanding. Reducing a decision that involved thousands of interacting variables to "the AI chose X because of A and B" is misleading and leads to loss of information that was accounted for during the decision making process. This leads to a loss of accuracy as a simple explanation may not represent the model's real reasoning. A clear explanation may worsen automation bias as the AI decision appears more reliable than it actually is.

AI Ethics:-

AI ethics comprises a set of principles that promote the responsible development and deployment of AI. These principles include fairness, which ensures equitable treatment of individuals on the basis of race, gender, age or socioeconomic status; bias mitigation, which seeks to eliminate discriminatory outcomes, this could occur due to unrepresentative training data, flawed algorithms, or human assumptions embedded in the system; transparency, which enables users to understand the decision making process of AI, it enables people to know how AI reaches its conclusions and what factors influence its decisions; privacy, which protects personal data from misuse; and accountability, which ensures that humans and organizations remain responsible for the actions and consequences of AI systems.

AI Governance:-

AI governance comprises the legal, technical, and organizational mechanisms that ensure artificial intelligence systems are developed and deployed responsibly. The goal of this being to ensure that AI is safe, transparent, ethical, and accountable. The EU AI Act is the world's first comprehensive legal framework to regulate AI based on the level of risk they pose. AI that is being used in high risk situations such as healthcare, law enforcement and military situations must meet strict requirements, such as complete transparency, risk management via testing and human oversight. The NIST AI Risk Management framework (AI RMF) is a US federal guide designed to help organizations build, deploy, and evaluate trustworthy AI systems. It is built around four core functions: Govern, Map, Measure and Manage. The Govern function establishes the organizational structures, policies, structure, and leadership commitment. It's necessary to cultivate a proactive and pervasive culture of AI risk management. Map

function establishes the context to frame risks related to an AI system. It focuses on understanding the system, its use, stakeholder, and potential risks within its operating environment. Measure function involves assessing characteristics of AI such as accuracy, fairness, explainability, robustness, and security through testing and validation to evaluate an AI's trustworthiness. Manage function emphasizes the mitigation of identified risks via monitoring, updating, and maintaining meaningful human oversight. Risk treatment comprises plans to respond to, recover from, and communicate about incidents or events.

Research Methodology:-

Research Design Gap:-

This study adopts a quantitative, cross-sectional survey research design to examine public perceptions regarding the necessity of human oversight in autonomous systems. The quantitative method was chosen because it provides quantifiable data on opinions and attitudes and enables for statistical analysis and objective interpretation of the data acquired. The study was cross-sectional, and used a structured questionnaire to elicit participants' views on the importance of human supervision in a range of artificial intelligence applications, including industrial, healthcare, transportation, and military systems.

Research Approach:-

The research is conducted deductively, beginning with current theories and literature on autonomous systems, Human-Robot Collaboration (HRC), Human-in-the-Loop (HITL), AI governance, and ethical AI. On this theoretical background a set of research questions and hypotheses was constructed and tested on the basis of primary data obtained through a survey.

Population of the Study 51:-

The target population was composed of people who have knowledge or interest in artificial intelligence and automation technologies. Participants were senior secondary school students, under graduate students, engineering students, instructors, researchers and working professionals from different academic backgrounds. Participants with different levels of knowledge gave a broader insight of public view of autonomous systems.

Sampling Technique:-

Non-probability convenience sampling was the sample approach applied. The participants were selected on the basis of availability and willingness to participate in the survey. The questionnaire was sent on the internet to collect responses from people belonging to diverse age groups and educational backgrounds.

Sample Size:-

The study analysed responses obtained from 51 participants. The sample size was considered adequate for identifying general trends in public opinion regarding human oversight in autonomous systems.

Research Instrument:-

Primary data were collected using a self-designed questionnaire titled the Human Oversight of Autonomous Systems Questionnaire (HOASQ). The questionnaire was specifically developed for this study after an extensive review of existing literature on artificial intelligence, automation, AI ethics, Human-Robot Collaboration, Human-in-the-Loop systems, and autonomous decision-making.The questionnaire comprised statements measuring participants' perceptions of AI reliability, safety, accountability, transparency, ethical concerns, and the necessity of human oversight.

Responses were recorded using a five-point Likert scale ranging from:-

- ➔ 1 – Strongly Disagree
- ➔ 2 – Disagree
- ➔ 3 – Neutral
- ➔ 4 – Agree
- ➔ 5 – Strongly Agree

The instrument was designed to capture quantitative data suitable for statistical analysis while ensuring clarity and ease of understanding for respondents.

Development of the Questionnaire:-

The questionnaire was developed through a systematic process involving the following stages:-

- ➔ Identification of key constructs from previous research.

- ➔ Review of peer-reviewed literature on autonomous systems and AI governance.
- ➔ Preparation of an initial pool of questionnaire items.
- ➔ Expert review for content relevance and clarity.
- ➔ Revision of ambiguous statements.
- ➔ Final administration through an online survey platform.

The questionnaire primarily assessed six dimensions:-

- ➔ Perceived reliability of autonomous systems
- ➔ Human oversight requirements
- ➔ Ethical concerns
- ➔ Accountability and responsibility
- ➔ Trust in AI-based decision-making
- ➔ Acceptance of autonomous technologies across different industries

Validity of the Instrument:-

To establish content validity, the questionnaire was reviewed by experts familiar with artificial intelligence and research methodology. Their feedback ensured that the items adequately represented the concepts being measured and were appropriate for the objectives of the study. Necessary modifications were made before distributing the questionnaire.

Reliability of the Instrument:-

Cronbach's Alpha was used to evaluate the internal consistency of the human observation Score (HOS). The calculation was based on the two survey questions used to create the score: Q3, which measured support for oversight by humans, and Q5, which assessed preference for human intervention systems.

Cronbach's Alpha coefficient was found to be 0.6725 and a standardised alpha of 0.6753. This suggests a moderate degree of internal consistency between the two items with the coefficient nearing the commonly used benchmark of 0.70.

Since the Human Oversight Score consists of only two components, the reliability of this assessment should be taken with a grain of salt. Additionally, this coefficient is for the two-item HOS metric and does not reflect the reliability of the entire People Oversight of Autonomous Systems Questionnaire (HOASQ), which measures various aspects of respondents' perceptions of autonomous systems.

Data Collection Procedure:-

Data was collected using Google Forms over a specified period. Participation was entirely voluntary, and respondents were informed about the purpose of the study before completing the questionnaire. No personally identifiable information was collected, ensuring respondent anonymity and confidentiality. Completed responses were exported into Microsoft Excel/SPSS for further analysis.

Data Analysis Techniques:-

The data were analysed using descriptive and inferential statistical methods. Descriptive analysis was conducted using frequency distributions, percentages, means, and standard deviations to describe the demographic characteristics of respondents and perceptions of autonomous systems.

A High Oversight Score (HOS) was constructed by combining responses to Q3, which measured endorsement of oversight by humans, and Q5, which measured preference for human override systems. The association between the human observation Score and specific survey variables was examined using Spearman's rank-order correlation.
Pearson's Chi-square test of independence was used to investigate relationships between categorical variables relevant to the research hypotheses. Cramér's V was calculated in an effort to measure the strength of the relationships. Because many contingency tables had low expected cell counts, p-values were estimated using Monte Carlo simulations with fixed margins and 500,000 iterations.

All inferential analyses were performed at the 0.05 level of significance. The results were interpreted as associations between survey responses rather than as evidence of causation. In analysing the results, the limitations of the sample size, convenience sampling and limited contingency tables were taken into consideration.

Ethical Considerations:-
The study adhered to standard research ethics throughout the data collection process. Participation was voluntary, informed consent was obtained from all respondents, and participants were free to withdraw at any stage. Confidentiality and anonymity of responses were maintained, and the collected data were used solely for academic research purposes

Data Analysis and Results:-

Demographic Profile of Respondents:-
Total respondents: 51. Full demographic breakdown charts (gender, age, educational background, and occupation) are provided in the Appendix under Survey Results.

Question-wise Analysis:-
Interpretation:-
The results show that 66.7% of the participants strongly agreed and 29.4% agreed that important AI or autonomous systems should be overseen by humans. The neutral option was selected by 3.9% of the participants, and there were no answers in the disagree or strongly disagree categories. Accordingly, 96.1% of participants agreed that human supervision was needed. This suggests a strong support for the maintenance of human supervision over important autonomous systems, especially in the areas of safety, responsibility, and critical decision-making.

Interpretation:-
58.8% of respondents believed autonomous systems can sometimes make mistakes, while 37.3% believed they often make mistakes and only 3.9% believed that mistakes occur rarely. Most respondents acknowledge that AI has the tendency to make mistakes, showing an awareness of the limitations of AI technologies.

Interpretation
66.7% of respondents strongly agreed and 29.4% agreed that human oversight is required in AI that is used in critical situations and other automated systems. Only 3.9% were neutral, none disagreed. There is an overwhelming agreement that humans should supervise important AI systems. It shows that there is a consensus between the public that it is important to maintain human oversight in autonomous technologies.

Interpretation:-
Only 9.8% of respondents say that they would trust a fully self-driving car without a human driver. The majority, 60.8% of respondents indicated that they wouldn't trust such a system, 29.4% selected maybe. This shows that a majority of people wouldn't trust a self driving system that has no form of human oversight. Some remain unsure as possible advancements in AI and automation systems could lead to trustworthy systems.

Interpretation:-
45.1% of respondents strongly agreed and 47.1% of respondents agreed that they are more comfortable with an autonomous system if a human can override it. 7.8% of respondents were neutral and none disagreed. There is a strong agreement that a human should be able to override autonomous systems. This shows that respondents value human judgement and intervention as a safety mechanism.

Interpretation:-
Clearly a large majority agree that human oversight is necessary in healthcare, military systems, transport, education and security and surveillance. Except education, most of these are high risk situations, or situations that are related to public safety and require ethical judgement. Although customer service and finance have the two lowest levels of support, more than half of the respondents agree that there should be meaningful oversight in both of these fields. This shows a general consensus that oversight should remain necessary.

Interpretation:-
56.9% of respondents believed responsibility should be shared among all stakeholders. 23.5% of respondents selected the company, 9.8% selected developer/programme, 5.9% selected the user, and only one respondent selected the government and a mixed response. The majority of people agree that the responsibility of a failed autonomous system should be shared. There should be collective accountability among all stakeholders.

Interpretation:-
72.5% of respondents believed that autonomous systems should be regulated by governments or international

organizations, 23.5% were unsure and only 3.9% of respondents believed that autonomous systems should not be regulated by the government.

Interpretation:-
Only 13.7% respondents believed that AI can completely replace human decision making in critical situations. Nearly half of the respondents (49%) believe that AI could only partially replace humans, while 37.3% believe AI cannot replace human decision making at all. Majority of the respondents believe AI can only partially replace, or not at all replace human decision making. It shows a confidence in human judgement. The most frequently identified concerns were lack of human judgement and overdependence on technology, each selected by 28.3% of respondents. This was followed by safety risks with 18.9% of respondents showing concern, privacy issues with 11.3%, job loss with 7.5% and bias or unfair decisions with 5.7% of responses. The findings show that the primary concern is the replacement of human judgement and overdependence on technology. The prominence of safety risks reinforces the perception that autonomous systems should operate under human oversight to ensure safe, reliable, and accountable outcomes. Privacy issues are also highlighted as AI can be used to run facial recognition scan very easily.

Theme-wise Analysis:-
Theme 1: Trust in AI:-
Questions: Q1, Q3, Q7
The survey shows that the participants are aware of the artificial intelligence (AI) technologies but are reluctant about fully autonomous systems. A large proportion expressed preference for human intervention, while most respondents were reluctant to trust an entirely automated vehicle without a human driver. The findings indicate that people are willing to use AI as a helpful tool but prefer human judgement and supervision in terms of safety, responsibility and important choices. Nevertheless, H1 cannot be assessed with the current questionnaire, because the survey does not distinguish between trust in autonomous systems with human supervision and trust in systems that work independently of human involvement. Thus, the results are consistent with the general significance of human supervision, but do not provide sufficient statistical evidence to support H1.

Theme 2: Ethical Concerns:-
Questions: Q5, Q8, Q10
The findings suggest participants are concerned about the ethical consequences of autonomous systems acting on their own. A large majority (96.1) agree or strongly agree that they feel more comfortable knowing that a human can intervene and intervene in an autonomous system if necessary. It implies that participants perceive human involvement as a safeguard, rather than an obstacle to self-directed functioning. In addition, 72.5% of respondents supported government or international regulation of autonomous systems, 23.5% were unsure and 3.9% were against such regulation. A significant proportion also felt that AI should not completely replace human decision-making, with 49% in favour of a partial replacement and 37.3% against a complete replacement. These results indicate that participants recognise the benefits of AI, but want it to be rolled out in an ethical and regulated environment with substantial human engagement.

Theme 3: Human Oversight:-
Questions: Q2, Q6, Q9
The results suggest that respondents are well aware of the limitations of autonomous systems and the need for shared human responsibility when things go wrong. Most think autonomous systems can sometimes or often make mistakes . Very few say AI never makes mistakes . This demonstrates an understanding that AI is not infallible . Most respondents agreed that if an autonomous system were to do harm, responsibility should be shared by all stakeholders, including developers, organisations, users and regulatory bodies, rather than assigned to a single entity. Overall worries for AI were lack of human judgement, overdependence on technology and safety risks. These results demonstrate that meaningful monitoring is important to respondents in this sample, as respondents acknowledge that autonomous systems are prone to error and hence require human supervision to maintain their safe, ethical, and accountable operation.

Statistical Analysis:-
Mean and Standard Deviation Score:-

Question	Coding Used	Mean	Standard Deviation
Q1. Have you ever used an AI-based or autonomous system?	Yes = 1, No = 0	0.922	0.272
Q2. Do you believe autonomous systems can make mistakes?	Never = 1, Rarely = 2, Sometimes = 3, Often = 4	3.333	0.554
Q3. Should humans supervise important AI or autonomous systems?	Strongly Disagree = 1 … Strongly Agree = 5	4.627	0.564
Q4. Would you trust a fully self-driving car without a human driver?	No = 1, Maybe = 2, Yes = 3	1.490	0.674
Q5. Do you feel more comfortable when a human can override an autonomous system if needed?	Strongly Disagree = 1 … Strongly Agree = 5	4.373	0.631
Q8. Should autonomous systems be regulated?	No = 1, Not Sure = 2, Yes = 3	2.686	0.547
Q10. Can AI completely replace human decision-making?	No = 1, Partially = 2, Yes = 3	1.765	0.681

Development and Evaluation of the People Oversight Score (HOS):-
Development and Evaluation of the People Oversight Score (HOS)
Two questionnaire items were used to create a Standard Score (HOS) to measure respondents' views on oversight by humans.
Q3: Should humans supervise important AI or autonomous systems?
Q5: Are respondents more comfortable with an autonomous system when a human can override it?
Both enquiries were assessed on a 5-point Likert scale with scores ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).
Human Observation Score was calculated using the following equation:
HOS = Q3 + Q5
So, the score could run from 2 to 10, with higher scores showing more support for systems of oversight by humans and override.
For descriptive interpretation, scores were grouped into three levels: low (2–4), moderate (5–7) and high (8–10).

The study found that the mean HOS was 9.00 with a standard deviation of 1.039 and a median of 9. The scores recorded were from 6 to 10. Among the 51 participants, 47 (96.1) had high scores in the HOS and 4 (7.8%) had moderate scores. No participant has been categorised as low score.The results indicate that participants largely expressed strong support for maintaining human supervision and override functions in autonomous systems.Cronbach's Alpha was used to assess the internal reliability of the two items. The coefficient was 0.6725 indicating moderate internal consistency. This number is a little lower than the usual reference value of 0.70 but it should be taken with a grain of salt as the HOS has only two items. Moreover, this evaluation of reliability only concerns the two-item HOS measure and should not be regarded as a reliability measure of the complete questionnaire.HOS is used in this study as a descriptive aggregate measure. The correlations between HOS and its component items are part-whole correlations, since they are calculated directly from Q3 and Q5, and should not be viewed as an independent validation of the score.

Spearman's Correlation Matrix:-

Spearman's rank-order correlation analysis was performed to examine the relationships between participants' endorsement of oversight by humans, confidence in autonomous systems, advocacy for AI regulation, and views on AI replacing human decision-making.The examination accounted for six factors: HR Score (HOS), Q3 (endorsement of human supervision), Q5 (inclination toward human override), Q4 (confidence in a fully autonomous vehicle without a human operator), Q8 (advocacy for AI regulation), and Q10 (views on the replacement of human decision-making).

The results are shown below:-

Table: Spearman's Rank Correlation Matrix

Variable	HOS	Q3	Q5	Q4	Q8	Q10
HOS	—	<0.001	<0.001	0.033	0.542	0.853
Q3	<0.001	—	0.0002	0.049	0.676	0.506
Q5	<0.001	0.0002	—	0.114	0.572	0.870
Q4	0.033	0.049	0.114	—	0.378	0.353
Q8	0.542	0.676	0.572	0.378	—	0.413

Q10	0.853	0.506	0.870	0.353	0.413	—

Note: N = 51. Values represent two-tailed p-values corresponding to the Spearman rank correlation coefficients reported in the preceding table. The diagonal entries are omitted because each variable is perfectly correlated with itself.

N is equal to 51, please be advised. Figures show Spearman's rank correlation coefficient (ρ).
Spearman rank-order correlation analysis was performed to investigate the relationships between the HRS Score (HOS), support of human supervision (Q3), preference for human override systems (Q5), confidence in fully autonomous vehicles (Q4), support for AI regulation (Q8), and opinion on AI replacing human decision-making (Q10).Q3 and Q5 were found to be moderately positively correlated ($\rho = 0.504$, $p = 0.0002$). Thus, participants who endorsed the need for supervision by humans to a greater extent also indicated that they were more comfortable with autonomous systems allowing human involvement.

The HR Oversight Score was strongly positively correlated with Q3 ($\rho = 0.813$, $p < 0.001$) and Q5 ($\rho = 0.906$, $p < 0.001$). However, these associations should be taken cautiously, as HOS is formulated by using the two components of Q3 and Q5. Hence, they are part-whole relationships and are not independent evidence for the validity of the score.

HOS and Q4 were found to have a weak negative correlation ($\rho = -0.299$, $p = 0.033$). Higher Q4 values indicate a greater willingness to trust a fully autonomous vehicle without a human operator, indicating that participants who preferred human supervision tended to have lower confidence in fully self-sufficient vehicles overall. However, this association did not reach statistical significance after a Bonferroni correction for multiple testing.

HOS had no significant statistical correlation with supporting AI regulation (Q8: $\rho = -0.087$, $p = 0.542$) or opinions on AI replacing human decision-making (Q10: $\rho = -0.027$, $p = 0.853$).

Fifteen different pairwise correlations were analysed. To minimise the chance of false-positive findings due to multiple comparisons, a Bonferroni-adjusted significance level of approximately 0.0033 (0.05/15) was used. Even with this more stringent criterion, the associations between Q3 and Q5, HOS and Q3, and HOS and Q5 remained statistically significant. However, the two last correlations depend on the components used to calculate HOS and thus should not be treated as independent relations.

Results show a statistically significant positive correlation between endorsement of oversight by humans and preference for human override systems. The relationships concerning trust in full autonomy, support for regulation of AI, and opinions about AI replacing human decisions were comparatively weak and did not survive correction for multiple testing. These results are associations in the sample and should not be taken as evidence of causation.

Chi-Square Test:-
Chi-Square Test:-
A Pearson Chi-square test of independence was conducted to investigate relations between selected categorical variables relevant to the research hypotheses. The study specifically examined the relationship between respondents' beliefs regarding the potential fallibility of AI, the possibility of AI replacing human decision-making, trust in self-driving vehicles, and prior experience with AI and their support for human override or human-in-the-loop systems.

Cramer's V was calculated to assess the strength of each correlation. Because of many expected frequencies of cells <5, we ran fixed-margin Monte Carlo simulations with 500,000 replicates to obtain more precise estimates of statistical significance. The Monte Carlo p-values are the main reference for interpretation of these contingency-table results.

Association between perceived AI fallibility and support for human supervision (H2):-
The first assessment explored the association between participants' beliefs about the fallibility of autonomous systems (Q2) and their support for human supervision (Q3).

Observed contingency table:-

Q2 response	Agree	Neutral	Strongly Agree	Total
Often	2	0	17	19
Rarely	1	1	0	2
Sometimes	12	1	17	30
Total	**15**	**2**	**34**	**51**

A Pearson Chi-square test was conducted, resulting in $\chi^2(4) = 18.602$, $p = 0.00094$ (asymptotic). The Monte Carlo estimate was $p = 0.00239$. The Cramér's V value was 0.427, which means a moderate correlation between the variables.The results indicate a statistically significant relationship between the perception of AI fallibility and endorsement of human supervision. Participants' judgements about the likelihood of autonomous systems making mistakes were related to their views about the necessity of human supervision.Hypothesis 2 (H2) is therefore confirmed. But the cross-sectional design does not demonstrate that the perception of AI fallibility leads to greater endorsement of human supervision.

Association between beliefs about AI replacing human decision-making and support for human supervision (H3):-

The second analysis examined the relationship between participants' belief in AI's capability to completely replace human decision-making (Q10) and their support for human supervision (Q3).

Observed contingency table:-

Q10 response	Agree	Neutral	Strongly Agree	Total
No	7	0	12	19
Partially	3	2	20	25
Yes	5	0	2	7
Total	**15**	**2**	**34**	**51**

The Pearson Chi-square test yielded $\chi^2(4) = 11.441$ with an asymptotic p-value of 0.02203. Monte Carlo approximation was $p = 0.02575$. The Cramér's V was .335, indicating a moderate association.Results show a statistically significant relationship between participants' perceptions of AI replacing human decision-making and their support for oversight by humans.Hence, H3 is supported. The findings suggest that participants' views on human supervision are connected to their perceptions of AI's ability to replace human judgement.

Association between trust in fully autonomous vehicles and preference for human override mechanisms (H4):-

The third assessment examined whether participants' trust in a fully autonomous vehicle with no human operator (Q4) correlated with their inclination toward human intervention systems (Q5).

Observed contingency table:-

Q4 response	Agree	Neutral	Strongly Agree	Total
Maybe	8	1	6	15
No	11	3	17	31

Yes	5	0	0	5
Total	**24**	**4**	**23**	**51**

The Pearson Chi-square test resulted in $\chi^2(4) = 7.531$ with an asymptotic p-value of 0.11034. Monte Carlo approximation gave p=0.11342. Cramér's V was 0.272, indicating a minor to moderate correlation.From the reactions observed, differences in preference for human intervention seem to exist between the three trust categories. However, the correlation was not significant at the 0.05 level.Thus, Hypothesis 4 (H4) is not supported. However, the results do not reach the level of statistical significance that would allow us to claim that scepticism of fully autonomous vehicles is related to a higher tendency to use human override systems.

Association between previous AI usage and perceived AI fallibility (H5):-

The fourth assessment examined the relationship between prior experience with AI-driven or autonomous systems (Q1) and participants' perception of the potential for autonomous systems to make mistakes (Q2).
Contingency table (recorded)

Observed contingency table:-

Q1 response	**Often**	**Rarely**	**Sometimes**	**Total**
No	1	1	2	4
Yes	18	1	28	47
Total	**19**	**2**	**30**	**51**

The Pearson Chi-square test resulted in $\chi^2(2) = 5.150$ with an asymptotic p-value of 0.07615. Monte Carlo estimation gives p = 0.16761. The Cramer's V was found to be 0.318, indicating a moderate correlation.However, Monte Carlo p-value was above the significance limit of 0.05. So the correlation was not statistically significant.Although 46 of the 47 respondents (97.9%) with previous experience with AI-based or autonomous systems indicated that these systems sometimes or often make mistakes, the analysis did not provide evidence of a statistically significant association between prior AI use and perceived AI fallibility.H5 is not supported, therefore.

Expected-frequency considerations:-

Chi-square tests were influenced by low expected frequencies of cells. In all four contingency tables, more than 20% of cells had an expected frequency less than 5. The predicted minimum frequencies were Q2 × Q3 = 0.078, Q10 × Q3 = 0.275, Q4 × Q5 = 0.392 and Q1 × Q2 = 0.157.These values indicate that the assumptions underlying the traditional asymptotic Pearson Chi-square approximation were not well satisfied. The fixed-margin Monte Carlo estimates were thus the basic basis for assessing statistical significance.Due to the small sample size and few observations in several categories, the findings are to be considered preliminary. These correlations will need to be confirmed in further studies with larger and more representative cohorts.

Key Findings:-

Results from the study of 51 subjects revealed a number of findings regarding people's attitudes toward human supervision in autonomous systems.Of the 51 participants, 47 (96.1) stated that they had prior experience with AI-driven or autonomous systems, implying that the respondents were quite familiar with artificial intelligence (AI) technologies.Most participants believed that autonomous systems are error-prone. 30 (58.8%) participants reported that errors occur occasionally and 19 (37.3%) reported that they occur frequently. Infrequently was reported by only 2 participants (3.9%),96.1% (49) of participants agreed or strongly agreed that significant AI or autonomous systems should be overseen by humans. This is strong support for the need to keep supervision by humans.Confidence in fully autonomous cars: Only five participants (9.8%) would trust a fully self-operating car without a human operator. Lack of trust in the system was indicated by 31 (60.8%) participants and uncertainty by 15 (29.4%).
47 participants (96.1) agreed or strongly agreed that they felt more comfortable with autonomous systems when a human was able to intervene.

People Oversight Score The mean People Oversight Score was 9.00 (out of 10), with 96.1 of participants in the high-oversight category.There was a statistically significant correlation between perceived AI fallibility and endorsement of human supervision (Monte Carlo p = 0.00239), supporting H2.A statistically significant association was found between the belief that AI would replace human decision-making and the support for human supervision (Monte Carlo p = 0.02575), thus supporting H3.Trust in full autonomous vehicles was not significantly correlated with tendency to override the human (Monte Carlo p = 0.11342). Thus, H4 was not supported.Prior AI use and perception of AI fallibility. There was no statistically significant correlation between prior AI use and perception of AI fallibility (Monte Carlo p = 0.16761). H5 was thus not supported.The results suggest that participants generally prefer to maintain human supervision and override capabilities in autonomous systems. However, the statistical evidence for the different hypotheses differs in strength, and the results must be interpreted with the sample and research design limitations in mind.

Summary of Hypothesis Testing:-

1) **Hypothesis**	*2)* **Statistical method**	*3)* **Result**	*4)* **Decision**
5) H1: Public trust in and acceptance of autonomous systems is positively associated with human oversight and override mechanisms.	*6)* No direct test available from the current questionnaire items	*7)* Not directly testable	*8)* Inconclusive
9) H2: Respondents who believe AI can make mistakes are more likely to support human oversight.	*10)* Chi-square test (Q2 × Q3)	*11)* p = 0.00239	*12)* Supported
13) H3: Beliefs about whether AI can fully replace human decision-making are linked to support for human supervision.	*14)* Chi-square test (Q10 × Q3)	*15)* p = 0.02575	*16)* Supported
17) H4: Distrust in fully autonomous vehicles corresponds with a stronger preference for human override mechanisms.	*18)* Chi-square test (Q4 × Q5)	*19)* p = 0.11342	*20)* Not supported
21) H5: Respondents with prior experience using AI-based or autonomous systems are more likely to believe such systems can make mistakes.	*22)* Chi-square test (Q1 × Q2)	*23)* p = 0.16761	*24)* Not supported

The results of the hypothesis testing indicate that H2 and H3 were statistically endorsed, while H4 and H5 were not supported. H1 could not be assessed directly, as the questionnaire did not include a specific item measuring trust in autonomous systems operating under human supervision as opposed to those operating independently. Thus, the current survey data do not permit a definitive conclusion concerning H1.

Discussion:-

Human Oversight:-

UK DSIT / CDEI (2023) was an international survey which covered nine countries and found that each had a majority support for powerful AI to be regulated by government backed frameworks, ranging from 51% in Japan to 65% in Italy, with 62% agreeing in the UK. This survey also found that most people in all countries agreed that powerful AI should be tested by a third party or an independent expert to prove that it's safe. The present study found that 72.5% of respondents agreed that AI should be regulated by governments and ethical guidelines while 23.5% said they aren't sure and 3.9% percent did not want AI to be under government made frameworks. We can also relate our statistic of 96.1% of people wanting human oversight to the international support for independent AI safety testing. Although both studies measure a different form of human involvement, with direct human supervision in the present study and independent safety testing in the international survey, both indicate a preference for maintaining human control and safeguards over powerful AI.

Trust in autonomous vehicles/ Human Override :-

Nordhoff et al. carried out an international survey on the acceptance of driverless vehicles, measuring the public's reception towards fully autonomous vehicles using the Likert Scale. Although the two studies use different response formats, their findings can be compared descriptively to identify similarities and differences in attitudes towards autonomous vehicles.The 2018 survey found that most respondents would enjoy taking a ride in a driverless vehicle with a mean score of M=4.90/6, this contrasts with the present survey where only 9.8% of respondents said that they would trust a fully self-driving car without a human driver. Therefore, the difference between the findings should not be interpreted as a direct contradiction, since the studies measured different aspects of acceptance and trust. Another question in the 2018 survey "There will always be accidents, even with DVs on the road" had a mean score of M=4.38/6, this shows that the respondents are skeptical about fully autonomous vehicles, this shows that the respondents believe that autonomous systems can make mistakes. Another question was asked "I would not use a driverless vehicle because technology can sometimes fail", the mean score here was M=3.48, which falls in the middle to upper range of the mean score, so some respondents wouldn't use the driverless vehicle if they knew it could fail occasionally. The present survey asked the question "Do you believe autonomous systems can make mistakes?"A large majority of respondents believed that autonomous systems can make mistakes, sometimes or often. This aligns with the 2018 survey as the respondents believed that autonomous systems can fail and admitted that there would be accidents regardless of driverless vehicles on the road. This shows skepticism in the fully autonomous vehicle. In the 2018 survey, the question "I would like to have a button inside the driverless vehicle which I can press to stop it" was asked and it had a mean score of M=5.18/6, showing that the respondents highly preferred a way to override the fully autonomous vehicle and retain some degree of control. This aligns with the present survey, where the question "Do you feel more comfortable when a human can override an autonomous system if needed?" was asked, with 96.1% of respondents agreeing that they would be more comfortable with human control in autonomous systems.

Autonomous weapons:-

Ipsos / Human Rights Watch 26-country survey on lethal autonomous weapons (2019) was a large international public opinion survey, across 26 countries, 61% opposed the use of lethal autonomous weapons while 22% supported and 17% were unsure. The question in the survey revolves around fully autonomous systems, not ones that are being overseen by humans or have any human involvement. However, 54% of those who opposed their use, the main concern was the lack of accountability as a concern and 45% were concerned about technical failures. In the present survey it was found that 86.3% of respondents believed that autonomous weapon systems should be overseen by humans. Given that the people who opposed fully autonomous systems in the previous survey were concerned about lack of accountability and possible technical failure, these findings support the present study's result that according to the public, human oversight should remain in applications where autonomous decisions can have serious consequences. Having human oversight reduces the risk of technical failure and leads to proper accountability. Although the questions and sample differ, both studies indicate skepticism in fully autonomous systems and a preference for retaining human intervention mechanisms even when autonomous technology is accepted.

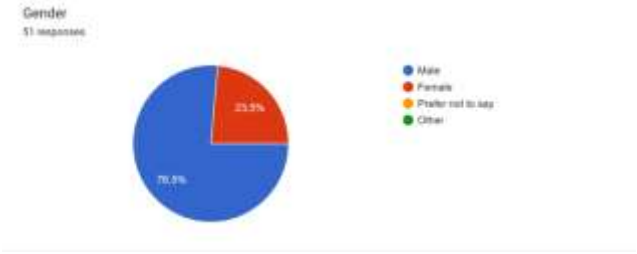
Public preferences for AI regulation:-

Pew Research Center, 2025 How People Around the World View AI was a survey that measured the trust that people have in their own country's government to regulate AI effectively. 55% of people trusted their own country to regulate AI. The present survey found that 72.5% of respondents supported AI regulation and government frameworks. Both the surveys measure different things, the survey conducted by Pew Research Center measures

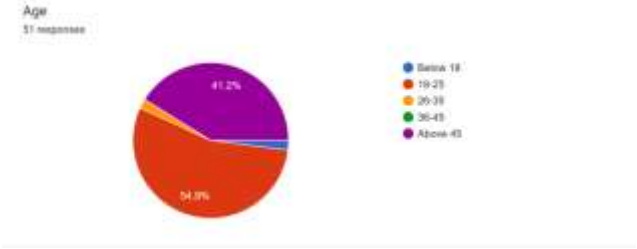
how much people trust in their government to regulate AI, while the present survey measures trust and attitude towards AI regulations. However they show a high level of support towards AI systems being governed by regulations and frameworks.

Appendix:-
HOASQ Questionnaire
Google Form
Consent Form
Survey Results

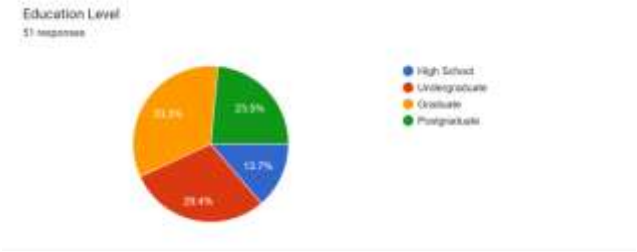
Gender



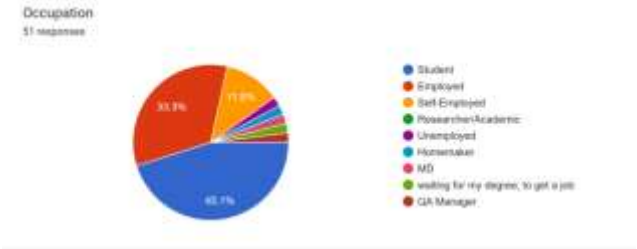
Age



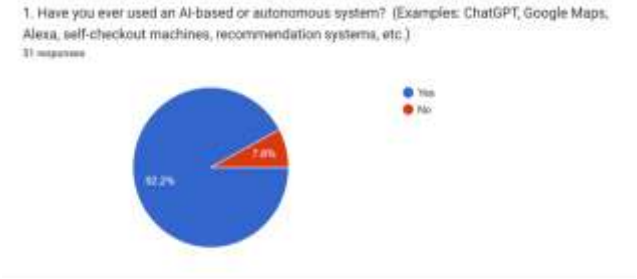
Education Level



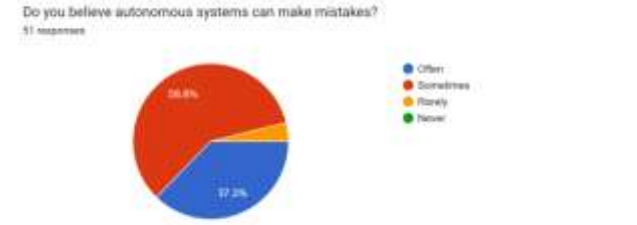
Occupation



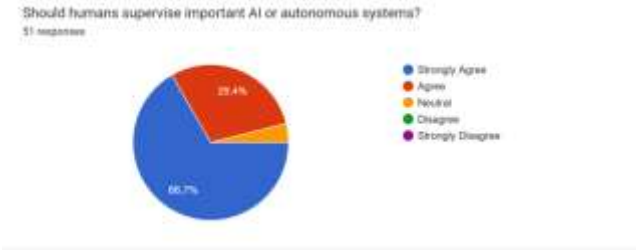
1. Have you ever used an AI-based or autonomous system? (Examples: ChatGPT, Google Maps, Alexa, self-checkout machines, recommendation systems, etc.)



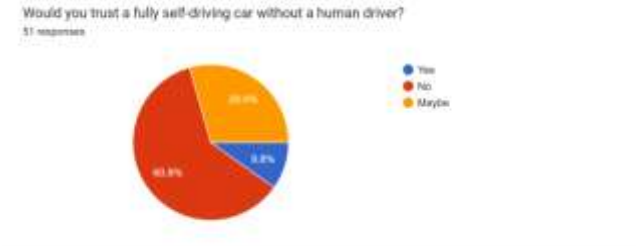
Do you believe autonomous systems can make mistakes?



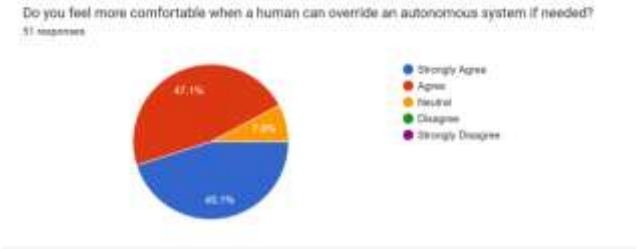
Should humans supervise important AI or autonomous systems?



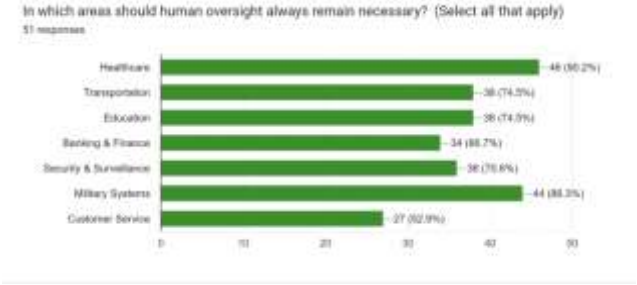
Would you trust a fully self-driving car without a human driver?



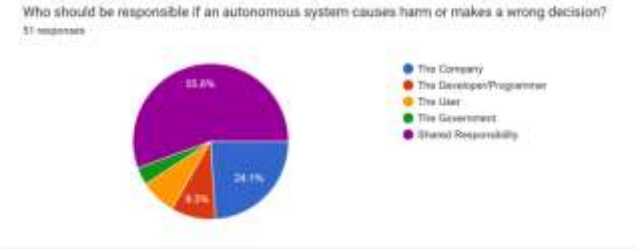
Do you feel more comfortable when a human can override an autonomous system if needed?



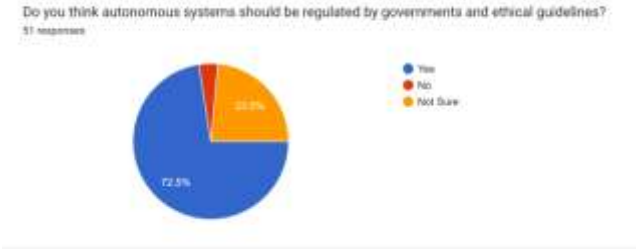
In which areas should human oversight always remain necessary? (Select all that apply)



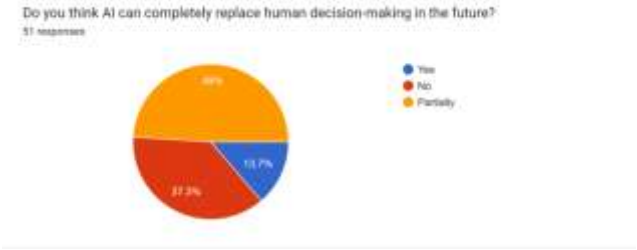
Who should be responsible if an autonomous system causes harm or makes a wrong decision?



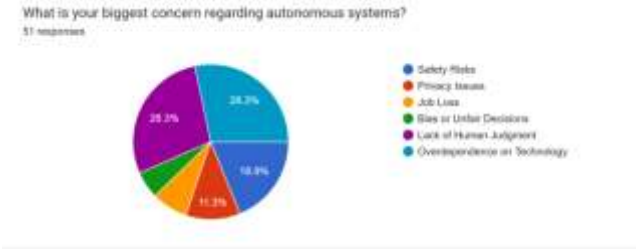
Do you think autonomous systems should be regulated by governments and ethical guidelines?



Do you think AI can completely replace human decision-making in the future?



What is your biggest concern regarding autonomous systems?



Recommendations:-

Improve Explainable AI:-

The “black box” problem is one of the major things that hold human oversight back. Without a transparent decision making path, a human cannot fully understand how an AI model has reached a decision, it also improves transparency and allows human operators to make correct decisions based on the output the system provides. Without any way to explain AI decisions, the risk of undetected errors and biases increase. Another way to improve explainable AI would be to reduce the technical jargon when an AI model is talking to a human operator. Only if an AI is talking to another engineer or an expert, they may need detailed model diagnostics, but an operator only requires the concise explanation of the factors affecting the decision. The AI may also change its response based on the user, doctors, military operator, regulator, and ordinary users don't all need the same amount of technical information. Explainability can also be built into the design of the model rather than added afterward as a separate output; features that are inherently understandable, together with backward traceability, allow for a more coherent workflow. As discussed in the Theoretical Framework, this approach still has limits: an oversimplified explanation can create a false sense of understanding and worsen automation bias by making an AI decision appear more reliable than it actually is. In practice, an explanation is only as good as its usefulness – it needs to give a human enough relevant information to judge whether the output can be trusted, and to step in when it can't.

Develop international AI regulations:-

AI regulations are extremely important now more than ever with AI models advancing very quickly. Mitigating risks such as safety failures, discrimination, privacy violations, and deepfakes while ensuring proper accountability, transparency and human oversight.

Mandatory human approval in high-risk systems:-

AI models that are being used in high risk situations such as healthcare, military uses, and autonomous vehicles where mistakes can have very large consequences, these models should have some form of HITL oversight. In places such as military systems, if an AI model chooses the wrong targets or civilians, and an attack is initiated it could lead to civilian casualties and it would be breaking international humanitarian law. Situations where the consequence of an AI error will have huge impact, the less acceptable it is for AI to be the final decision-maker. Autonomous weapons or targeting systems must be overseen by humans and the option to kill should lie with the human operator. With healthcare, a wrong diagnosis by AI will lead to incorrect medicine, instead AI can be used in a collaborative capacity. This would be where a physician retains final decision authority and has the final say. Autonomous vehicles should have a human override switch, the general public also agrees with this as seen in past surveys and the present survey. Another way could be that autonomous vehicles have conditional oversight

Human oversight in defence:-

Defence organizations should require meaningful human oversight for autonomous systems capable of producing significant consequences, specifically systems which involve target identification, engagement, intelligence assessment or other decisions that could lead to use of lethal force. This can be implemented in the form of a human-oversight framework for autonomous military systems.Lower risk AI can be autonomous, such as monitoring or gathering information. Higher risk AI must have mandatory human review before consequential action, and lethal decisions must remain in the hands of a human operator who understands the way the system works, and must have the training authority and technical capability to override the system. The AI model must also be explainable, the operator should be provided with the relevant evidence, how confident/uncertain the AI is, significant alternative interpretations and known system limitations.Such a framework would preserve the advantages of having autonomous systems while maintaining meaningful human control over high-consequence military decisions.

Conclusion:-

Research objectives achieve:-

Taken together, the survey results answer what this study set out to ask. On the safety of unsupervised autonomous operation, respondents were clearly uneasy: almost all (96.1%) believe autonomous systems make mistakes at least sometimes, only 9.8% said they would trust a fully self-driving car with no human at the wheel, and support for human supervision averaged M = 4.63 (SD = 0.56) on the five-point scale. The public does not yet see autonomous systems as reliable enough to run without oversight.

On AI autonomy more broadly, most respondents (96.1) reported having previously used AI-based or autonomous systems, but that familiarity hasn't translated into trust in full autonomy – preference for keeping a human override option was just as strong (M = 4.37, SD = 0.63). People seem comfortable with AI as an assistant rather than as a replacement decision-maker.

Asked which sectors most need human supervision, respondents' answers tracked risk fairly closely: healthcare (90.2%), military systems (86.3%) and transportation (74.5%) drew the strongest support, followed by security (70.6%) and banking (66.7%). The higher the stakes of an autonomous error, the more strongly people wanted a human involved.

On ethical and technical limitations, the possibility of error was the most cited technical concern (96.1%), alongside worries about the "black box" problem and the risk that an oversimplified explanation could misrepresent what a model is actually doing. On the ethics side, loss of human judgement was the top concern (28.3%), and a majority (55.6%) felt that responsibility for AI-caused harm should be shared across developers, companies, users and regulators rather than pinned on any one party.

Finally, respondents were asked directly whether AI could replace human decision-making, and which industries should require oversight by law. The answer to the first was fairly firm: 86.3% rejected full replacement, with only 13.7% in favour and 49% willing to accept partial replacement. The answer to the second echoes the sector findings above – healthcare, military systems and transportation again topped the list, tracking the severity of harm an error could cause in each sector.

Major findings:-

Four patterns cut across these results. Respondents see AI as fallible but don't pin failure on a single actor – 96.1% believe autonomous systems make mistakes at least sometimes, and a majority (55.6%) think accountability should be shared across developers, companies, users and regulators rather than assigned to one group. Support for human control, meanwhile, is both strong and consistent: it shows up almost identically whether it's framed as a preference for supervision (M = 4.63, SD = 0.56) or for override capability (M = 4.37, SD = 0.63). Loss of human judgement and overdependence on technology were the two most frequently cited concerns, tied at 28.3% each, suggesting respondents worry about complacency just as much as outright technical failure. And when it comes to who should be keeping watch, a large majority (72.5%) want autonomous systems regulated by governments or international bodies, with only 3.9% opposed.

Future research directions:-

1. Larger and more diverse samples The present study has 51 respondents, so future research could examine whether the findings hold across larger populations, different age groups, occupations and levels of AI experience. Different regions may also affect answers, for example in Singapore autonomous vehicles are being deployed, so they may trust autonomous vehicles more and may display more trust towards autonomous and AI systems.
2. Sector-specific research Sectors like healthcare and military systems show high support for human oversight, future studies could investigate why people demand stronger oversight in these sectors.
3. Research on XAI The present studies discusses XAI, but does not test it. It could be tested by showing participants different explanations for a decision, and see how they affect trust, acceptance and willingness to override the system.
4. Longitudinal Research AI and autonomous systems are developing rapidly, attitudes measured today may change as people become more familiar with the technology and as AI and autonomous systems gets better.

References:-

1. Alpay, F., & Alpay, T. (2025). Inadequate human-in-the-loop oversight in advanced AI systems [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.25618.11200
2. Centre for Data Ethics and Innovation. (2023). Public attitudes to data and AI: Tracker survey (Wave 3). Department for Science, Innovation and Technology. https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-3
3. El Zaatari, S., Marei, M., Li, W., & Usman, Z. (2019). Cobot programming for collaborative industrial tasks: An overview. Robotics and Autonomous Systems, 116, 162–180. https://doi.org/10.1016/j.robot.2019.03.003
4. Firlej, M., & Taeihagh, A. (2021). Regulating human control over autonomous systems. Regulation & Governance, 15(4), 1071–1091. https://doi.org/10.1111/rego.12344
5. Holzinger, A., Zatloukal, K., & Müller, H. (2025). Is human oversight to AI systems still possible? New Biotechnology, 85, 59–62. https://doi.org/10.1016/j.nbt.2024.12.003
6. Ipsos. (2019, January 22). Six in ten (61%) respondents across 26 countries oppose the use of lethal autonomous weapons systems. https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons
7. Javed, N., & Altaf, Z. (2026). Human Oversight in AI-Driven Intelligence: Rhetoric, Reality, and the Risks of Automation. International Journal of Intelligence and CounterIntelligence, 39(3), 724-737. https://doi.org/10.1080/08850607.2026.2637905
8. Kyriakou, K., & Otterbacher, J. (2023). In humans, we trust: Multidisciplinary perspectives on the requirements for human oversight in algorithmic processes. Discover Artificial Intelligence, 3, Article 44. https://doi.org/10.1007/s44163-023-00092-2
9. Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. Journal of the American Medical Informatics Association, 24(2), 423–431. https://doi.org/10.1093/jamia/ocw105
10. Poushter, J., Fagan, M., & Corichi, M. (2025, October 15). How people around the world view AI. Pew Research Center. https://www.pewresearch.org/global/2025/10/15/how-people-around-the-world-view-ai/
11. Ramos, M. A., & Mosleh, A. (2021). Human role in failure of autonomous systems: A human reliability perspective. In 2021 Annual Reliability and Maintainability Symposium (RAMS) (pp. 1–6). IEEE. https://doi.org/10.1109/RAMS48097.2021.9605790
12. Szabadföldi, I. (2021). Artificial intelligence in military application – Opportunities and challenges. Land Forces Academy Review, 26(2), 157–165. https://doi.org/10.2478/raft-2021-0022