LETTER OF TRANSMITTAL

Haider Malik
haider05@my.yorku.ca
Lassonde School of Engineering, York University

Mar 20, 2025

To: Professor Kevin Gingerich
ENG 2003 - Effective Engineering Communication
York University

Subject: Submission of Term Project 01 - Phase 4 Final Technical Report

For Term Project 01, I am happy to present the completed technical report, which is titled "Enhancing Safety Certification of AI-Driven Autonomous Systems through Explainable AI, Regression Testing, and Standardized Regulations." This thorough research discusses the significant obstacles to guaranteeing the security of AI-driven autonomous systems and offers strong alternatives to guarantee their dependability and public confidence.

This final proposal has several important improvements, including:

- A thorough examination of the difficulties in obtaining AI safety certification, emphasizing the intricacies brought about by AI's probabilistic nature and the absence of established certification standards at the moment.
- An emphasis on Explainable AI (XAI), Regression Testing, and Simulations for AI Safety Testing, together with case studies and real-world applications, the suggested solutions are thoroughly examined.
- Examining current and developing frameworks, including the EU AI Act and NIST AI Standards, and their implications for AI certification, this discussion focuses on regulatory activities.
  Incorporation of visual elements, such as figures and diagrams, to improve comprehension of difficult ideas.
- A thorough revision summary appendix that describes the comments received and the changes made to the report as a result.
- I value the time you took to read this final proposal, and I eagerly await any additional comments you might have.

Sincerely

Haider Malik

Enhancing Safety Certification of AI-Driven Autonomous Systems through Explainable AI, Regression Testing, and Standardized Regulations

Submitted By Haider Malik

Lassonde School of Engineering

ENG 2003 - Effective Engineering Communication

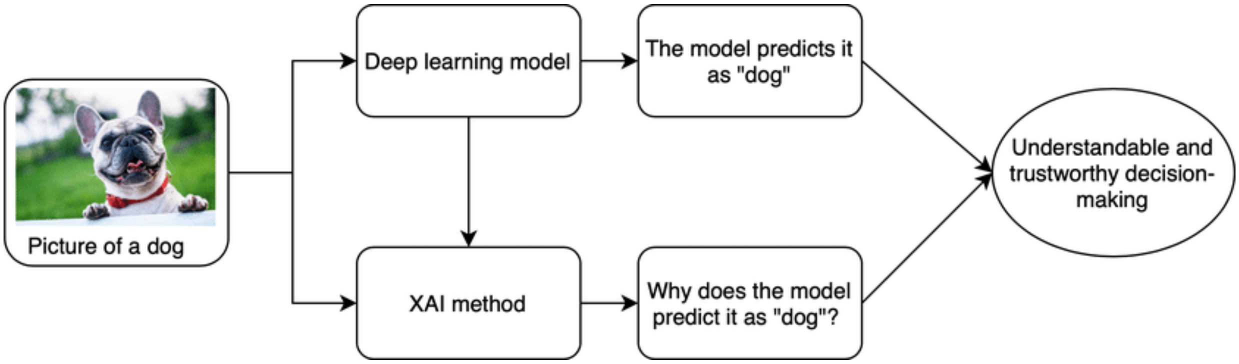
Date: Mar 20, 2025

EXECUTIVE SUMMARY

Artificial intelligence (AI) has transformed operational capabilities and efficiency through its quick integration into autonomous systems, which span industries like aircraft, healthcare, and transportation. Nevertheless, this development poses serious difficulties for guaranteeing the security and dependability of these AI-powered systems. For AI applications with probabilistic behaviours and learning capabilities, traditional certification procedures—which were created for deterministic systems—are frequently insufficient.

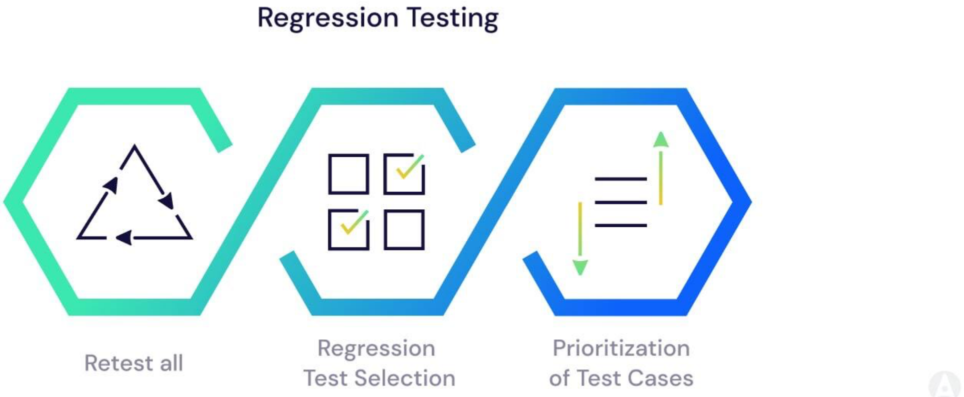
This paper explores the challenges of certifying autonomous systems powered by artificial intelligence and suggests a comprehensive strategy to improve safety certification. Among the primary solutions mentioned are:

- Figure 1 Below shows a simple flowchart of how XAI could work


Picture of a dog
Deep learning model
XAI method
The model predicts it as "dog"
Why does the model predict it as "dog"?
Understandable and trustworthy decision-making

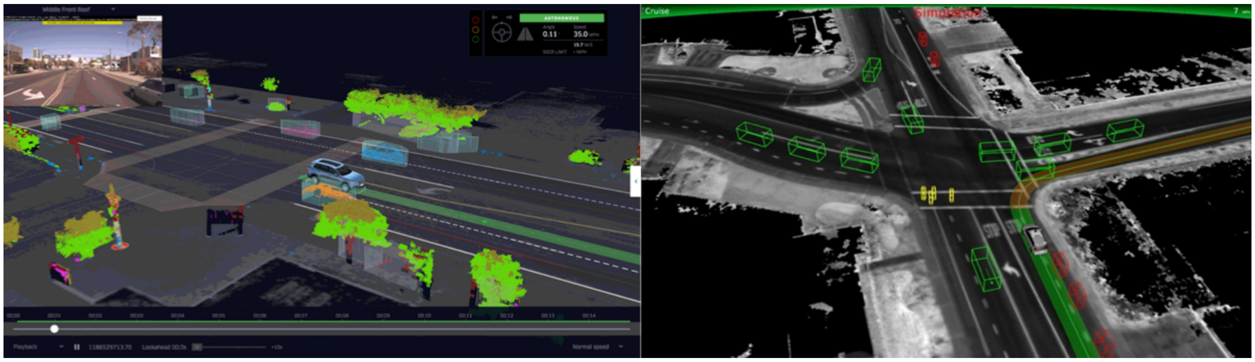
1. Explainable AI (XAI): Improving the interpretability and transparency of AI judgments to promote comprehension and confidence among interested parties.

- Figure 2 Below shows the repeated cycle of regression testing



2. Regression Testing: Applying methodical testing techniques to make sure that AI system updates or adjustments don't jeopardize current safety regulations.

- Figure 3 shows Waymo and Tesla’s autonomous driving simulator



3. Simulations for AI Safety Testing: By employing sophisticated simulation settings to thoroughly assess AI performance in a range of harsh circumstances, risks are reduced prior to real-world implementation.

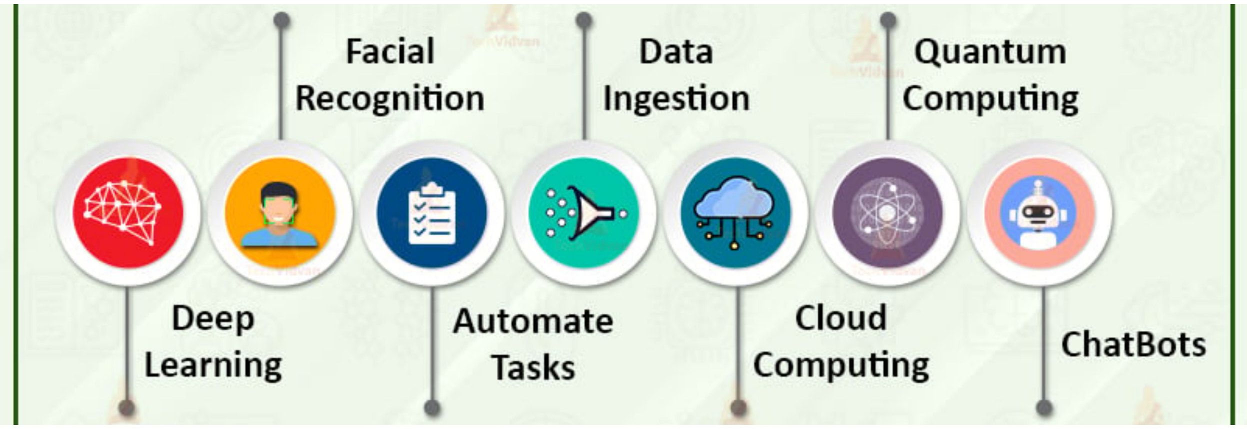
The research highlights the need for standardized frameworks to oversee AI safety certification and looks at current regulatory initiatives. The objective is to create a strong basis for the secure and dependable implementation of AI-powered autonomous systems in a range of sectors by combining these solutions.

TABLE OF CONTENTS

List Of Figure

INTRODUCTION AND BACKGROUND

- Figure 4 below shows some common tasks that AI can achieve



Artificial intelligence has emerged as a key component of contemporary technological development, spurring innovation in a variety of fields. AI-powered autonomous systems are now able to carry out activities like driving cars, diagnosing illnesses, and flying airplanes that were previously only possible by humans. Although there are many advantages to these advancements, there are also many drawbacks, especially when it comes to guaranteeing the security and dependability of AI-driven systems.

The predictability of deterministic systems, whose behaviours can be predicted and evaluated under specified conditions, is the foundation of conventional safety certification procedures. However, probabilistic models that learn and adapt from data are the foundation of AI systems, particularly those that use machine learning techniques. The implementation of traditional certification techniques is made more difficult by this inherent unpredictability, which calls for the creation of novel strategies catered to the particulars of artificial intelligence.

Recent changes in regulatory environments highlight how urgent it is to address AI safety. For example, proposals have been made in the UK to speed up AI safety laws so that tech firms are required to submit their AI models for stringent regulatory testing. The goal of this action is to shield the general public from the possible dangers of uncontrolled AI deployment. However, the difficulties in creating thorough AI safety regulations are highlighted by the delays in such laws, which are caused by alignments in world policy.

Given this, it is crucial to investigate and put into practice reliable techniques that may successfully certify the security of AI-driven autonomous systems. In order to create a thorough safety certification framework, this research explores the difficulties involved in this undertaking and suggests solutions that include Explainable AI, Regression Testing, and sophisticated simulation approaches.

THE CHALLENGE: CERTIFYING AI-DRIVEN AUTONOMOUS SYSTEMS

The following considerations are the main causes of the complex challenge of certifying the safety of AI-driven autonomous systems:

The absence of international AI safety standards**:** The creation of universal safety standards has lagged behind the quick growth of AI technologies. The creation of generally recognized safety standards is made more difficult by the lack of standardized certification procedures, which results in inconsistent safety evaluations across various industries and geographical areas.

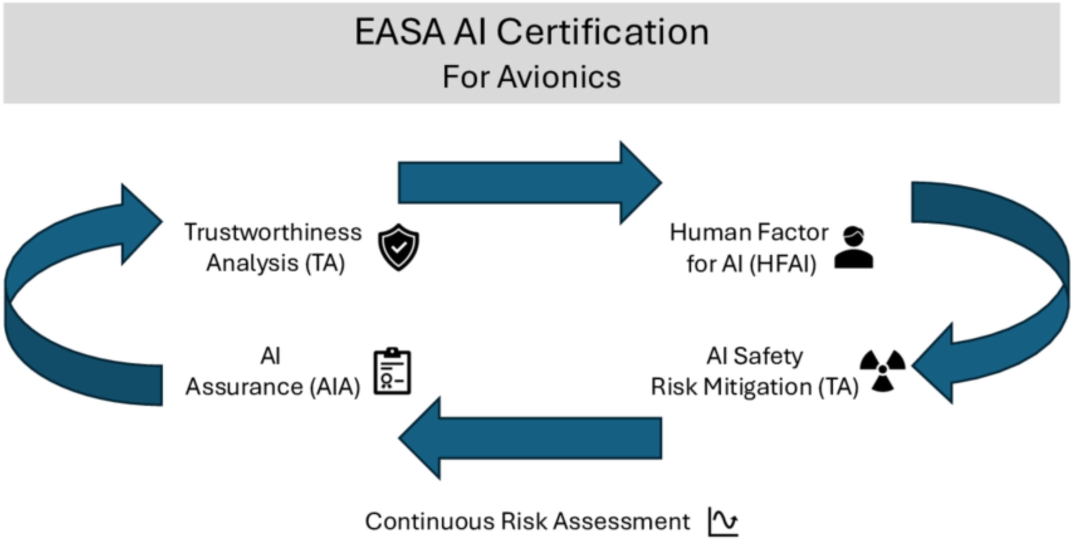
AI's Probabilistic Nature: AI models—especially those based on machine learning—function probabilistically, in contrast to traditional software systems that adhere to deterministic principles. They make decisions based on patterns and learn from data, which can result in unexpected behaviours in new circumstances. For conventional testing and certification techniques that depend on preset situations and results, this unpredictability presents serious difficulties.

Applications at High Risk: AI is being used more and more in safety-critical fields including autonomous driving, healthcare, and aviation. System failures in these situations can have disastrous effects, including fatalities. Thorough testing and validation procedures that may take into consideration a broad range of conceivable scenarios and edge cases are necessary to guarantee the security of AI systems in such high-stakes situations.

Changing Regulatory Activities: Initiatives to create legal frameworks for AI, such as the National Institute of Standards and Technology's (NIST) AI Standards and the European Union's AI Act, are still in the early stages of development and lack international harmonization. The certification process for AI systems meant for global deployment is made more difficult by the differences in regulatory approaches between various jurisdictions.

In order to overcome these obstacles, novel certification techniques that take into account the special qualities of AI systems must be developed. Potential ways to improve the safety certification procedure for AI-driven autonomous systems are examined in the report's following parts.

- Figure 5 below shows a example of an existing certification for avionics by the European Union


EASA AI Certification
For Avionics
Trustworthiness Analysis (TA)
Human Factor for AI (HFAI)
AI Safety Risk Mitigation (TA)
AI Assurance (AIA)
Continuous Risk Assessment

PROPOSED SOLUTIONS

A variety of approaches are suggested to successfully handle the difficulties involved with certifying AI-driven autonomous systems. These consist of the use of simulations for AI safety testing, Explainable AI (XAI), and Regression Testing. Every strategy makes a distinct contribution to improving the security and dependability of AI systems.

1. XAI, or Explainable AI

Improving Interpretability and Transparency

The term "explainable AI" (XAI) describes strategies and tactics that help people comprehend how AI systems make decisions. XAI improves transparency, fosters trust, and helps identify any biases or inaccuracies in the AI system by offering transparent insights into how AI models reach particular conclusions.

2. Regression testing

Maintaining Stability in the Face of Constant Development

Verifying that recent code changes have not negatively impacted current functionalities is the goal of the software testing technique known as regression testing. Regression testing makes assurance that changes or updates to AI models don't cause new mistakes or impair system performance in the context of AI-driven autonomous systems. This procedure is essential for preserving the security and dependability of AI programs, particularly as they develop over time.

3. Automated Regression testing

Manual regression testing can be time-consuming and prone to error because of the complexity and size of AI systems. In order to ensure thorough coverage and consistency, automated regression testing uses tools and frameworks to carry out test cases in a methodical and effective manner. Regression testing can be made even more successful by including AI. AI-driven technologies, for example, can forecast regions of the codebase that are more prone to errors, find trends in faults, and intelligently choose and prioritize test cases. In addition to speeding up testing, this clever automation enhances the identification of minute problems that conventional approaches might miss.

4. Visual Regression Testing

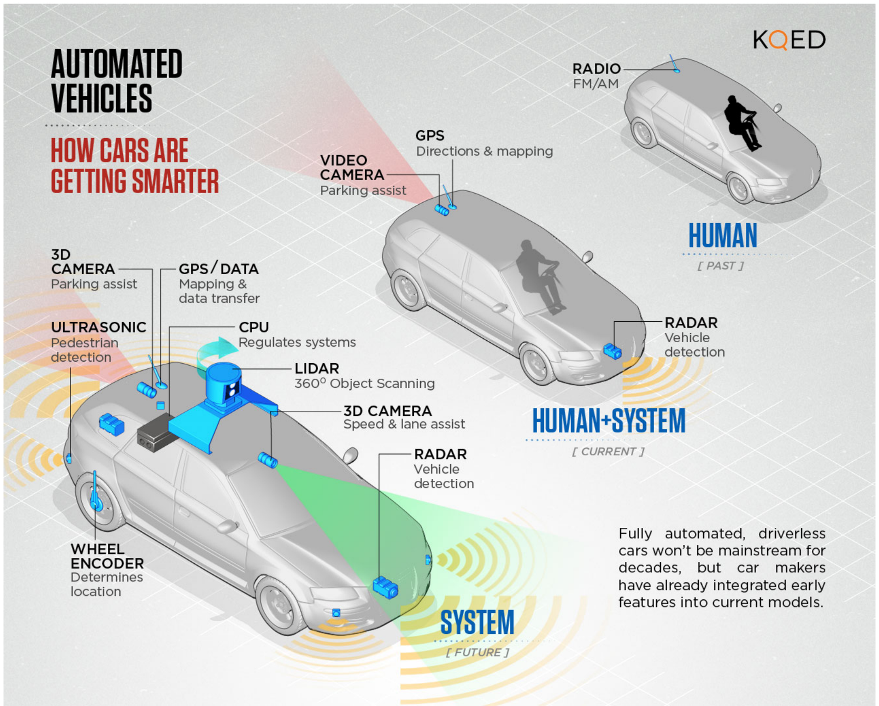
For AI-driven systems, especially those with user interfaces, visual integrity is just as important as functional correctness. The goal of visual regression testing is to identify inadvertent modifications to an application's visual elements. By comparing visual outputs before and after code modifications, artificial intelligence (AI) and computer vision technologies can be used to spot differences that might compromise user experience. This strategy guarantees that the visual display of the system stays constant and meets user expectations.

5. Simulations for AI safety testing

Establishing Regulated Settings for Strict Assessment

AI-driven autonomous systems can be tested in simulations under a variety of conditions, including uncommon and dangerous ones that are difficult to replicate in the real world. Developers can evaluate the system's behaviour, spot possible weak points, and make the required adjustments before to deployment by modelling a variety of scenarios.

- Figure 6 below shows the amount of technology that a automated vehicle possess compared to a human system


AUTOMATED VEHICLES
HOW CARS ARE GETTING SMARTER
KQED
RADIO FM/AM
GPS Directions & mapping
VIDEO CAMERA Parking assist
HUMAN [PAST]
3D CAMERA Parking assist
GPS/DATA Mapping & data transfer
ULTRASONIC Pedestrian detection
CPU Regulates systems
LIDAR 360° Object Scanning
RADAR Vehicle detection
3D CAMERA Speed & lane assist
HUMAN+SYSTEM [CURRENT]
RADAR Vehicle detection
WHEEL ENCODER Determines location
Fully automated, driverless cars won't be mainstream for decades, but car makers have already integrated early features into current models.
SYSTEM [FUTURE]

- Advantages of Simulation-Based Testing

Risk Reduction: By evaluating AI systems in simulated settings, possible harm can be avoided during the development stage without subjecting them to real-world hazards.

Scalability: Scaling simulations to test multiple situations at once offers comprehensive coverage and speeds up the testing process.

cost-effectiveness: the cost effectiveness of this Simulations can drastically reduce development expenses by eliminating the need for actual prototypes and real-world testing.

- Applications in Autonomous Vehicles

The creation of autonomous vehicles is a well-known use case for simulation-based testing. In order to evaluate how vehicles react to different traffic situations, weather conditions, and unforeseen impediments, businesses use advanced simulators to simulate virtual driving environments. Without the risks associated with on-road testing, this method makes it easier to validate safety features and improve decision-making algorithms.

Discussion: Real-World Applications and Limitations

- Implementing Solutions in Practice

To improve the safety and dependability of AI-driven autonomous systems, some sectors have embraced the integration of Explainable AI, regression testing, and simulation-based testing.

Healthcare

AI systems help with diagnosis and therapy recommendations in the medical field. By ensuring that medical practitioners can understand the reasoning behind AI-generated recommendations, XAI promotes trust and makes informed decision-making easier. As AI models are updated with fresh medical data, regression testing is used to ensure their accuracy and dependability. Before being used in the real world, simulation-based testing ensures the safety and efficacy of AI applications by enabling their virtual trial in healthcare settings.

Aerospace

AI is used in the aircraft sector for activities like autonomous flight and predictive maintenance. For AI judgments to comply with safety procedures and legal requirements, explainability is essential. Regression testing makes sure that current safety measures are not jeopardized by upgrades to AI systems. Simulations offer a way to validate performance without putting people in danger by testing AI behaviour under a variety of flying settings, including emergency situations.

- Limitations and challenges

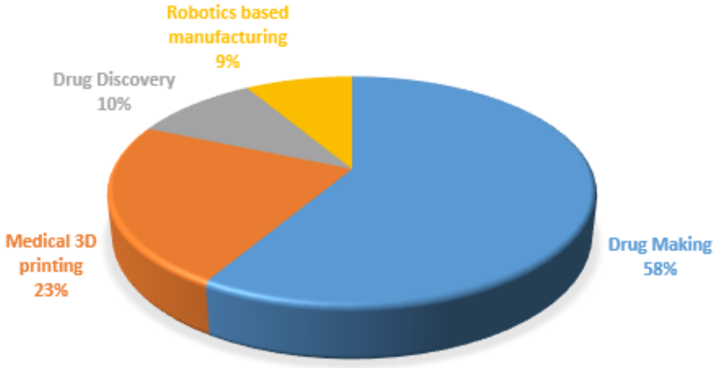
Despite their advantages, these solutions have certain drawbacks:

Complexity of explainability: It's still difficult to strike a balance between interpretability and model complexity. Although very complex models may perform better, explainability may suffer as a result.

Regression Testing's Resource Intensiveness: Especially for large-scale AI systems, thorough regression testing can be time-consuming and resource-intensive, requiring a significant amount of processing power.

Simulations' fidelity: The realism of the simulated environment determines how accurate simulation-based testing is. Simulations that are inaccurate or overly simplistic could produce deceptive findings and possibly miss important problems.

- figure 7 below shows a pie chart of the applications of medicine using AI



Conclusion and Future Directions

- The path to a strong AI safety certification

A comprehensive strategy combining Explainable AI, regression testing, and simulation-based testing is required to guarantee the security of AI-driven autonomous systems. Even though there has been a lot of progress, more research and development is necessary to overcome current constraints and adjust to new ones.

- Future Research Directions

Improving Explainability: Research is required to provide techniques that improve the interpretability of intricate AI models without sacrificing their functionality

Optimizing Regression Testing: Advances in automated regression testing, especially those that use artificial intelligence (AI), can increase efficacy and efficiency while also increasing scalability.

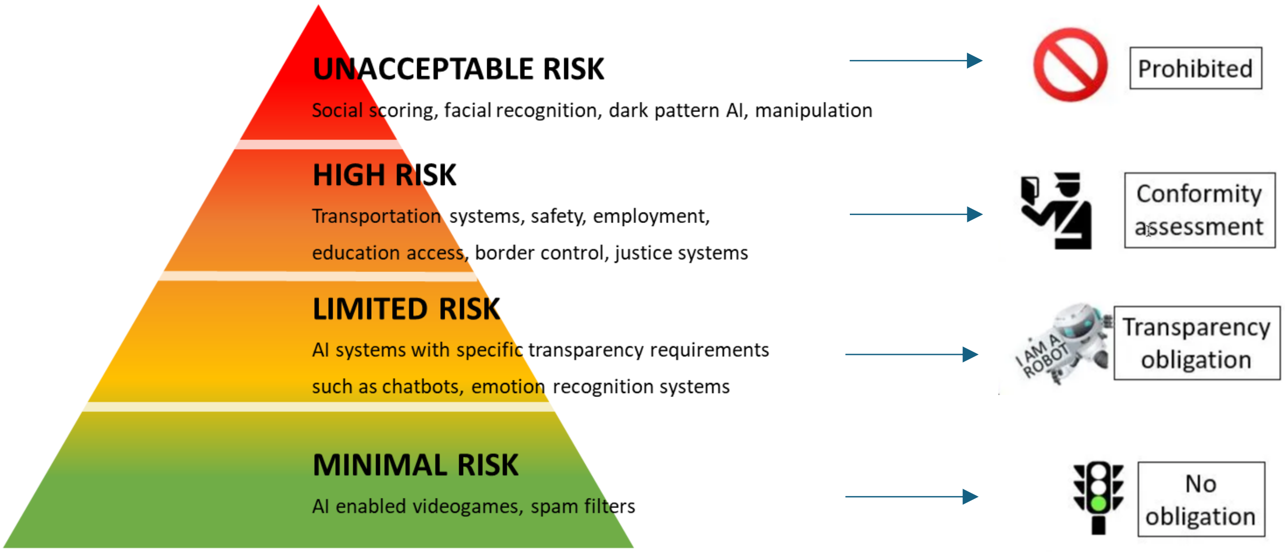
Enhancing Simulation Fidelity: The trustworthiness of simulation-based testing results will be increased by creating high-fidelity simulation environments that faithfully mimic real-world circumstances.

- Developments in Regulation

Standardized regulatory frameworks must be established if AI technologies are to be widely accepted and trusted. The goal of projects like the National Institute of Standards and Technology (NIST) and the European Union's AI Act is to provide thorough standards for AI safety certification. To create and execute efficient certification procedures, regulatory agencies, industry participants, and the scientific community must constantly collaborate.

To sum up, by combining these approaches and encouraging cooperation, we can clear the path for the secure and dependable implementation of AI-driven autonomous systems, guaranteeing that they function in a way that is open, trustworthy, and consistent with social norms.

- Figure 8 shows the existing EU's Artificial intelligence act for risk levels of AI


UNACCEPTABLE RISK
Social scoring, facial recognition, dark pattern AI, manipulation
Prohibited
HIGH RISK
Transportation systems, safety, employment, education access, border control, justice systems
Conformity assessment
LIMITED RISK
AI systems with specific transparency requirements such as chatbots, emotion recognition systems
I AM A ROBOT
Transparency obligation
MINIMAL RISK
AI enabled videogames, spam filters
No obligation

REFERENCES

1. A. H. Abdul Rahman, A. Z. Talib, and S. A. Noah, "Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions," *arXiv preprint arXiv:2112.11561*, 2021. Available: [https://arxiv.org/abs/2112.11561]

2. M. Kwiatkowska and X. Zhang, "When to Trust AI: Advances and Challenges for Certification of Neural Networks," *arXiv preprint arXiv:2309.11196*, 2023. Available: [https://arxiv.org/abs/2309.11196]

3. Holistic AI, "AI Regulations for Autonomous Vehicles [Updated 2025]," *Holistic AI Blog*, Feb. 2025. Available: [https://www.holisticai.com/blog/ai-regulations-for- autonomous-vehicles]

4. M. B. Islam, M. J. Hossain, and M. S. Kaiser, "Reliability and Safety of Autonomous Systems Based on Semantic Modelling for Self-Certification," *Robotics*, vol. 10, no. 1, pp. 1-21, 2021. Available: [https://www.mdpi.com/2218-6581/10/1/10]

5. M. S. Kaiser, M. B. Islam, and M. J. Hossain, "Autonomous Vehicles: Evolution of Artificial Intelligence and the Current Industry Landscape," *AI*, vol. 4, no. 4, pp. 1-20, 2023. Available: [https://www.mdpi.com/2504-2289/8/4/42]

6. Katalon, "Regression Testing: Embracing the Power of AI and Automation. Jan 25, 2025 "Available" [https://katalon.com/resources-center/blog/ai-in-regression-testing]

7. Anja Manuel, "It's Time for Limited, Mandatory Testing for AI." Financial Times Sep 25, 2024. Available: [https://www.ft.com/content/8b190ef1-32d8-4196-9643-3396a44d3bcd