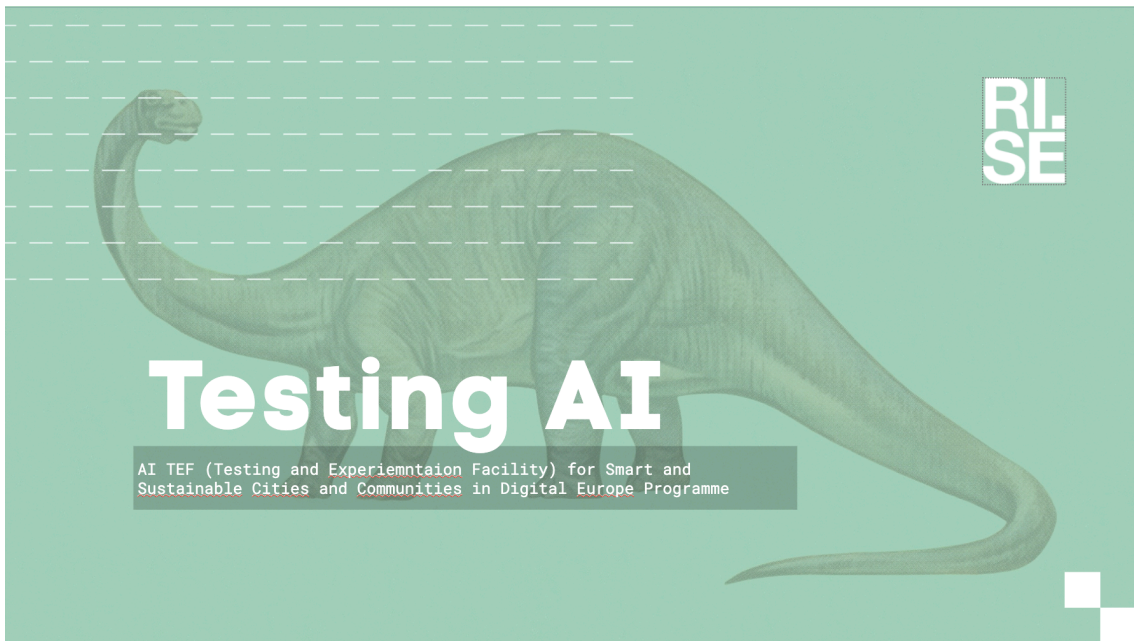


RI. SE

CITCOM.AI RISE



From AI Act to Structured Testing of AI Systems

Nishat I Mowla, Kabir Fahria

RISE Report : 2024:84

From AI Act to Structured Testing of AI Systems

Nishat I Mowla, Kabir Fahria

Abstract

From AI Act to Structured Testing of AI Systems

The Citcom.AI RISE testing approach is a step towards structured AI system evaluation and testing under the AI Act's regulatory framework. It establishes a definition of context in the scenario of different AI application domains, AI subfields, and use cases. In particular, a systematic evaluation, from defining the context and application to detailed risk assessments, linking each AI application to corresponding testing standards and methodologies, is presented. The approach translates AI Act's high level regulatory requirements for different AI system risk levels to appropriate technical testing techniques for achieving trustworthiness across different domains and AI subfields, promoting responsible AI deployment and fostering trust in AI applications.

Key words: Testing AI, AI Act, AI systems, Standards, Context

RISE Research Institutes of Sweden AB

RISE Report : 2024:84

ISBN: 978-91-89971-46-2

Content

Abstract	1
Content	2
1 Make sense of AI evaluation	3
2 AI Act and risk categories	3
3 AI testing standards	5
4 Application domains & subfields of AI	6
5 Proposed AI Evaluation Approach	7
5.1 Component 1 – Application Context	8
5.2 Component 2 – System Risk Level.....	9
5.3 Component 3 – Mapping standards and SOTA on AI Act requirements.....	11
6 Conclusion	12

1 Make sense of AI evaluation

The challenges of Artificial Intelligence (AI) applications are vast and multifaceted, especially when aiming to develop comprehensive testing under frameworks like the AI Act. The field is evolving, with state-of-the-art (SOTA) advancements emerging rapidly across numerous domains and use cases, each provided by different organizations with varying expertise. AI subfields, such as machine learning, natural language processing (NLP), and computer vision, add layers of complexity with the increasing number of tools available. This creates a chaotic landscape as shown in Fig. 1 and a systematic approach is required to develop a structured solution for testing AI application that is also compliant with regulations and standards. Addressing this requires a harmonized framework that can cater to the diverse applications and sub-fields of AI while ensuring transparency, accountability, and adaptability within the AI Act’s guidelines.

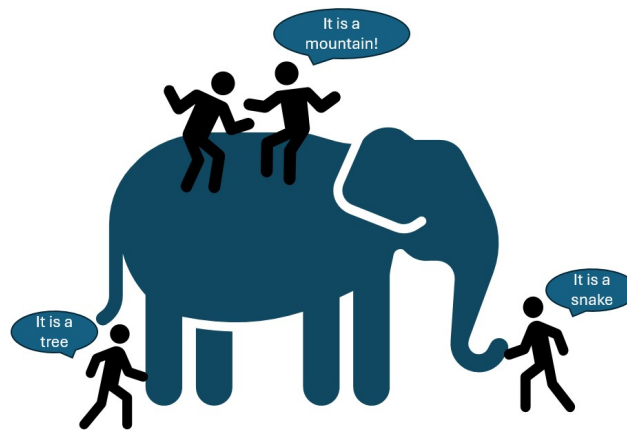


Fig. 1: Evaluation of AI compared to a big elephant showing a necessity for a holistic picture.

2 AI Act and risk categories

The Artificial Intelligence Act (AI Act)¹ established by the European Union (EU) is a comprehensive legislative framework designed to regulate the development, deployment, and use of AI across EU member states. According to the AI Act,

“AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.”

¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of the 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), <https://eurlex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>.

Written in a high-level language, the AI Act aims to address the risks and challenges posed by AI technologies while promoting innovation and the adoption of AI. It categorizes AI systems based on the level of risk they pose in terms of health, safety and fundamental rights—ranging from unacceptable to low or no risk—and sets out corresponding regulatory requirements as shown in Fig. 2. High-risk AI applications, for example, are subject to stringent compliance and transparency obligations to ensure they are safe and respect fundamental rights. The Act is part of the EU's broader strategy to become a global leader in *trustworthy AI*, ensuring that all AI systems that has societal impact are ethical, secure, and respect user privacy.

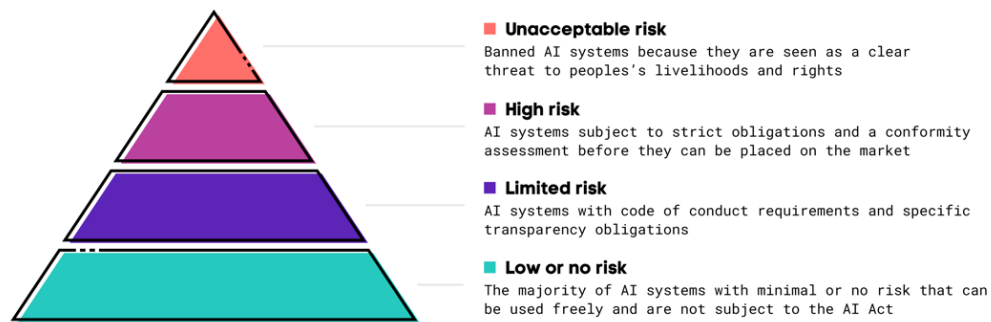


Fig. 2: AI Act risk levels.

The AI Act considers four key risk levels:

- **Unacceptable Risk:** AI systems classified in this category are banned from use because they pose a significant threat to people's livelihoods and rights. These types of AI are considered too hazardous to be allowed in practice.
- **High Risk:** AI systems that fall under this category are subject to strict regulatory obligations and must undergo a comprehensive conformity assessment before they can be put on the market. These systems require careful scrutiny to ensure they meet all regulatory standards due to their potential impact on society in terms of health, safety, and fundamental rights by materially influencing the outcome of decision making.
- **Limited Risk:** AI systems in this category need to adhere to specific codes of conduct and are subject to transparency obligations. While these systems can be used, they must meet certain conditions that ensure a lesser, but still significant, level of regulatory compliance.
- **Low or No Risk:** This category includes AI systems that are perceived to have minimal or no risk, allowing them to be used freely without stringent regulatory oversight under the AI Act. These systems are generally considered safe for widespread use with no significant implications for fundamental rights or societal values.

3 AI testing standards

ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission) have established ISO/IEC JTC 1/SC 42² to focus on the standardization of AI. This committee is responsible for developing international standards, technical reports, and technical specifications relevant to AI. These standards cover various aspects, including transparency, data quality, security, reliability, ethical concerns, and lifecycle management of AI systems. The goal is to facilitate the responsible adoption of AI technologies, ensuring they are developed in a trustworthy manner while addressing both technical and ethical aspects. These standards are meant to be applicable across different industries and sectors, fostering safe, reliable, and effective use of AI technology worldwide. Fig. 3 shows a table outlining various AI testing standards, categorized by specific aspects such as classification and evaluation, AI software quality, security, safety, data quality, robustness, ethical concerns, lifecycle management, and risk management. Each category lists corresponding ISO/IEC standards that address various aspects of AI testing:

- *Classification and Evaluation:* Standards like ISO/IEC 29119 and ISO/IEC 4213 focus on methodologies for testing and evaluating software, including AI systems.
- *AI Software Quality:* ISO/IEC 24028 offer guidelines on ensuring quality in software development and implementation.
- *Security, Trustworthiness, and Privacy:* ISO/IEC 25010 and other related standards provide frameworks for ensuring that software systems are secure and protect user privacy.
- *Safety:* ISO/IEC 5469 discusses functional safety and AI systems where standards such as ISO/IEC 22989 discusses concepts and terminologies connected to safety.
- *Data Quality and Bias:* ISO/IEC 5259 series discuss various characteristics focusing on data quality, crucial for training reliable AI systems.

Classification and evaluation	AI Software quality	Security, trustworthiness, privacy	Safety	Data quality & bias	Robustness and reliability	Ethical and societal concerns	Management & lifecycle
ISO/IEC 29119 series	ISO/IEC 24028	ISO/IEC 25010	ISO/IEC 22989	ISO/IEC 5259	ISO/IEC 27000	ISO/IEC 24368	ISO/IEC 42001
ISO/IEC 4213	ISO/IEC 12207	ISO/IEC 22989	ISO/IEC 5469	ISO/IEC 24027	ISO/IEC 24029	-	ISO/IEC 23894
ISO/IEC 25059	ISO/IEC 25000 series	ISO/IEC 2382	-	ISO/IEC 8183	-	-	ISO/IEC 38507
ISO/IEC 5471	ISO/IEC 23053	ISO/IEC 24028	-	-	-	-	ISO/IEC 5338
Functional		Non-functional					

Fig. 3: AI testing standards.

² ISO/IEC JTC 1/SC 42 Artificial intelligence. Available at <https://www.iso.org/committee/6794475/x/catalogue/p/o/u/1/w/o/d/o>

- *Robustness and Reliability:* ISO/IEC 27000 are concerned with information security, indirectly impacting the robustness of AI systems.
- *Ethical and Societal Concerns:* Standards such as ISO/IEC 24368 highlight the need for ethical considerations in AI development.
- *Management and Lifecycle:* ISO/IEC 42001 discusses the lifecycle management of AI systems, ensuring they are developed and maintained correctly.

The standards mentioned above are categorized and narrowed down to specific focuses such as quality, safety, security, and ethics, to align with global best practices and regulatory requirements. There are new standards that are continuously being developed by standardization bodies that Citcom.AI RISE testing approach will also continuously adopt. There are also Harmonised Standards³ developed by CEN-CENELEC⁴ that will be another supporting tool for complying with AI Act requirements.

4 Application domains & subfields of AI

AI systems can be categorized by dimensions such as the application domains and subfields of AI as shown in Fig. 4. The different subfields of AI include Machine Learning, Deep Learning, Natural Language Processing (including LLMs), Computer Vision, Reinforcement Learning, etc. as shown in Fig. 4.

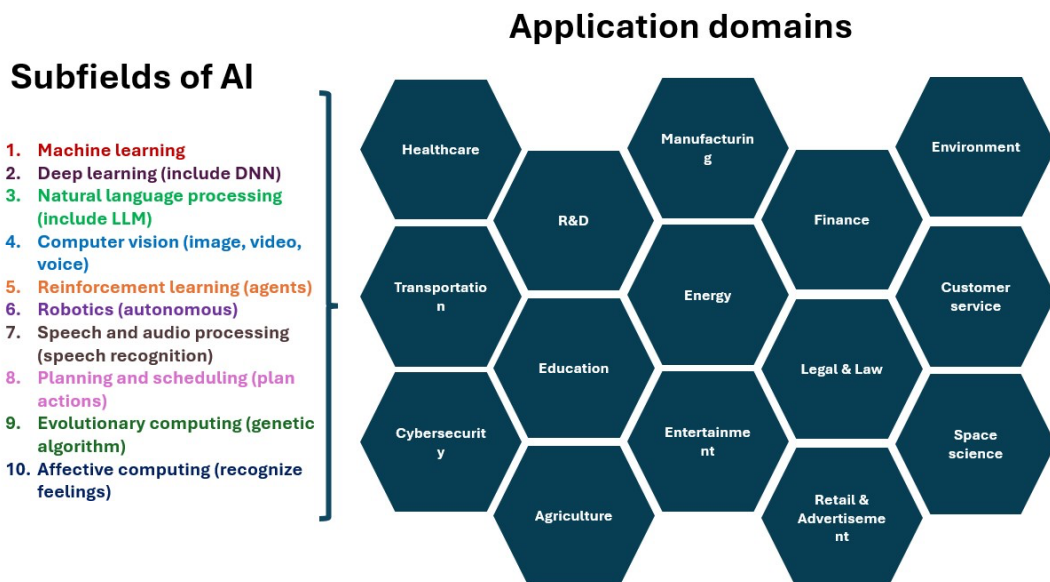


Fig. 4: AI Application domains and subfields of AI.

³ Harmonised Standards for the European AI Act. Available at <https://publications.jrc.ec.europa.eu/repository/handle/JRC139430>

⁴ CEN-CENELEC Management Centre (CCMC). Available at <https://www.cenelec.eu/management-centre/>

Each AI subfield is designed to address specific functionalities within AI technologies. Moreover, the diverse AI subfields can be mapped to use cases within AI application domains across diverse sectors such as Healthcare, Finance, Legal & Law, Space Science, Agriculture, Transportation, etc.

For example, a machine learning model such as neural network might be used to *identify anomalies* in the transportation domain’s vehicle network data, while a computer vision model such as convolutional neural network could be utilized in the same domain by autonomous vehicles for *object detection*. In contrast, two domains such as healthcare and education may use natural language processing such as LLMs to *summarize textual information*. This shows the wide-ranging complexity and adaptability of AI sub-fields and their applications across diverse sectors.

5 Proposed AI Evaluation Approach

Chapters 2 and 3 address the complexities of AI risk management, while Chapter 4 explores the diversity of AI systems across AI subfields, domains, and use cases. To effectively manage these complexities and diversity, Citcom.AI RISE has developed a structured testing approach. This approach distils the diversity of AI systems into a contextual framework that maps to AI Act, standards, and SOTA methods.

As shown in Figure 5, this “mapping” approach begins by defining the initial context and specific AI application to determine the risk level, applicable requirements, types of tests (such as accuracy or explainability testing), relevant guidelines, and methods, further detailed in Section 5.3. This structured approach ensures that AI systems comply with the AI Act by establishing a clear process from input to output: the input defines the AI system as the AI application and the context it operates in, and the output is a Compliance Status. This Compliance Status provides a summary of standards, guidelines, and methodological frameworks and may include an approval state (Fig. 6).

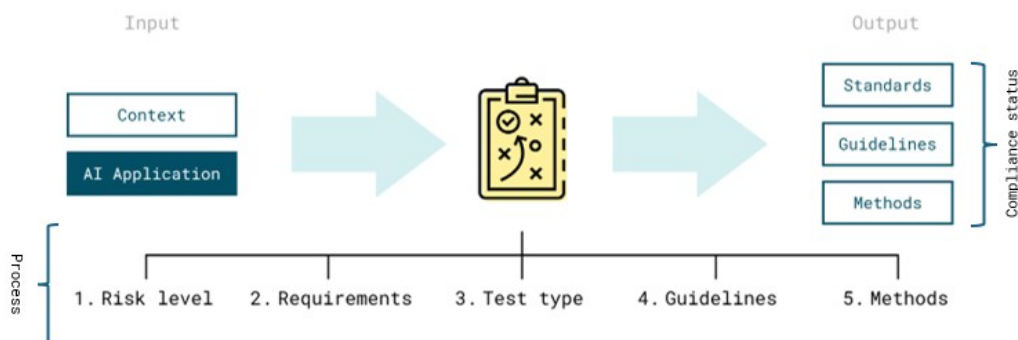


Fig. 5: System Overview: Testing AI from input to output perspective.

The testing AI approach is laid out through three components as shown in Fig. 6:

Component 1 - application context: We define context in terms of domain, use case, and use case provider, and we define AI application in terms of data & AI models.

Component 2 – system risk level: The output from component 1 (application context) is mapped on to the risk categorization to define a risk level for a specific application. Component 1 allows a clear distinction between different AI subfields, use cases, and their providers while considering the specific data and models used within each application. This decoupling enables a systematic evaluation of AI solutions across various industries—ranging from healthcare to agriculture—facilitating the identification of risk levels (i.e., unacceptable, high, medium, and low) for component 3.

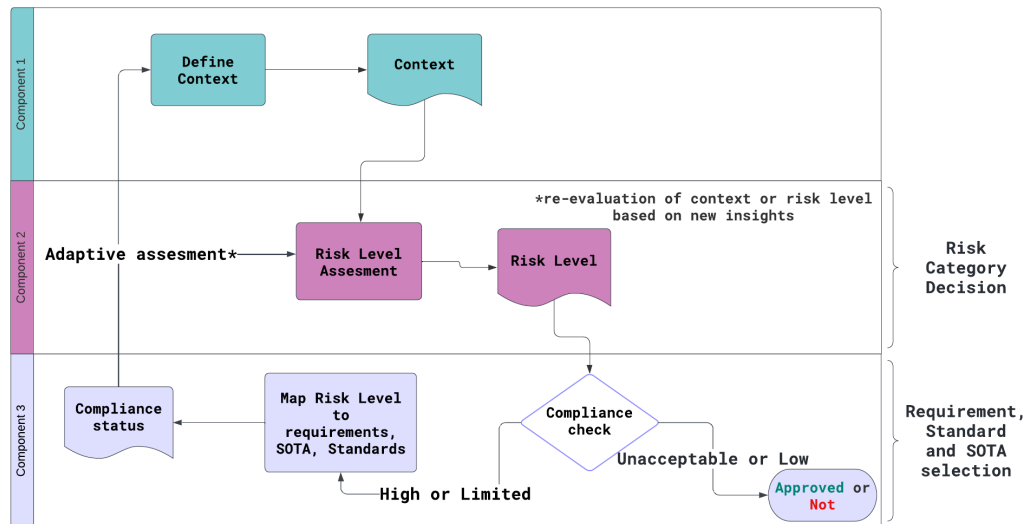


Fig. 6: Component flow-chart.

Component 3 - mapping standards and SOTA on AI Act requirements: The final component of Citcom.AI RISE’s testing approach focuses on aligning AI systems with the regulatory requirements of the AI Act, incorporating relevant standards and SOTA techniques. This component ensures that AI applications are accurately assessed and classified based on identified contexts and regulatory guidelines. By mapping each application to a defined Compliance Status, which may include an Approval State, this step guarantees that applications meet applicable standards and best practices.

The structured evaluation supports responsible AI deployment by fostering informed decision-making and encouraging organizations to confidently address the complexities of AI deployment. Detailed description of component 1, component 2, and component 3 are discussed below.

5.1 Component 1 – Application Context

The context in which an AI application is deployed plays a critical role in determining its regulatory obligations. It is not simply the tool itself that is categorized as high or low risk; rather, the level of risk is assessed once the tool is placed within its specific context. This means an AI application *does not inherently possess a risk level per se*; rather, it is the *specific use case and domain in which it operates that defines its acquired risk level*. The application’s risk level – whether high or low – can be accurately identified only by understanding the intended domain, use case, and operational environment, ensuring that regulatory standards are met in a manner suited to its actual impact and usage.

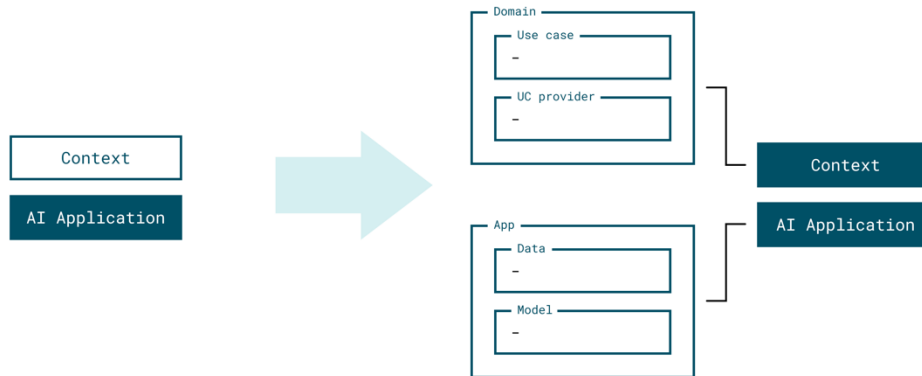


Fig. 7: Component 1: Context of AI application.

For instance, consider a Large Language Model (LLM)-based system designed to sift through and summarize Curriculum Vitae (CV)s based on key pieces of information. If this system is employed by a company as a tool to assist its Human Resource (HR) department by providing summaries of submitted CVs, the regulatory requirements would likely be less stringent. Subsequently, if a similar AI system is used to automatically reject CVs that do not meet certain criteria without human oversight, the regulatory demands would be considerably higher, reflecting the greater impact of the AI's decisions on individuals. In the first stage of our testing AI approach the application context is determined as shown in Fig. 7 based on domain and use case details.

5.2 Component 2 – System Risk Level

One aspect is determining the appropriate risk level from the application context. Another aspect involves identifying distinct requirements from AI Act with each risk levels. By mapping the context and AI application in Component 1, Citcom.AI RISE approach pinpoints the exact requirements to follow for each risk level.

Fig. 8 shows a classification of AI systems based on risk levels according to the AI Act, along with the specific requirements that apply to each category. These risk levels include:

- **Low or No Risk:** AI systems that pose minimal or no risk can be used freely without needing certification under the AI Act.
- **Limited Risk:** Limited risk applications have requirements of transparency of AI and documentation: 1) Transparency requirements under Article 13 of AI Act must be met, and 2) Documentational obligations as per Article 11 and 12. It is not a requirement to have functional and non-functional testing but naturally should have the basic testing and evaluation to ensure expected performance.

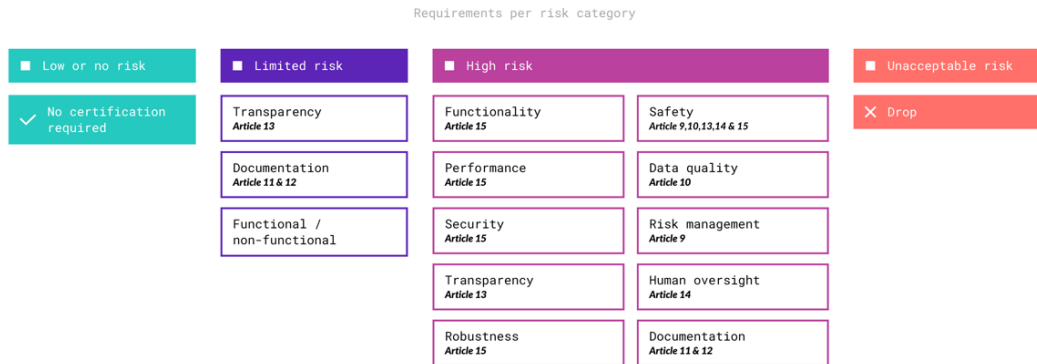


Fig. 8: Mapping AI risk categories to testing requirements and AI Act articles.

- **High Risk:** High risk applications must meet requirements related to functionality, performance, robustness and security tests as specified under Article 15, ensure safety as per Article 9,10,13,14 & 15, obligations for data quality testing under Article 10, risk management outlined in Article 9, transparency as detailed in Article 13, and human oversight as mentioned in Article 14.
- **Unacceptable Risk:** These are AI systems that are banned because they pose a clear and significant threat to the public's safety and rights.

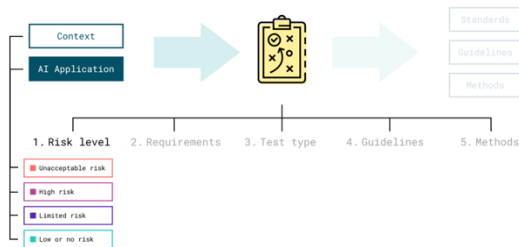


Fig. 9: Component 2: Risk level related to context and AI application.

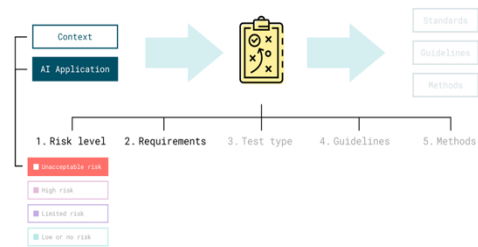


Fig. 10: Component 2: Mapping context and AI application to a specific risk level.

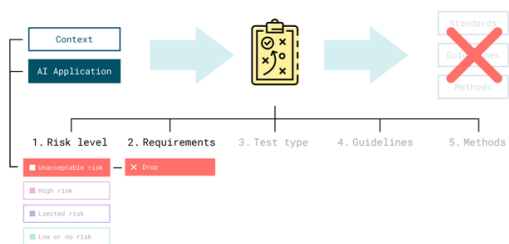


Fig. 11: System Prototype: AI Evaluation of unacceptable risk.

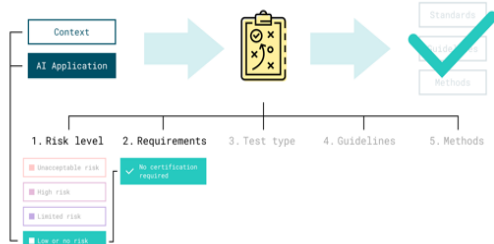


Fig. 12: System Prototype: AI Evaluation of low risk.

This categorization ensures that each AI system is scrutinized and regulated according to its potential impact, ensuring that higher-risk applications are subjected to stricter oversight and compliance requirements to safeguard user rights and safety. In component 2, the context and AI application comes as an input and goes through Risk Level analysis (Fig. 9 and Fig. 10).

The risk-based approach to mapping testing methods to the AI Act is crucial for ensuring that AI applications are evaluated effectively. The process begins by categorizing the AI application into one of the risk levels: unacceptable risk, high risk, limited risk, and low

or no risk. In cases of “unacceptable risk”, the AI application raises a red flag and is considered to be discarded (Fig. 11), as it poses threats that cannot be mitigated. Similarly, applications falling under “low or no risk” are considered safe and do not require extensive testing, as their impact is minimal (Fig. 12).

5.3 Component 3 – Mapping standards and SOTA on AI Act requirements

The high risk and limited risk categories require evaluation across various dimensions, starting with identifying specific requirements, such as functionality, performance, security, transparency, and robustness. These requirements form the foundation for selecting appropriate test types (e.g., explainability test) based on established standards guidelines such as ISO/IEC DIS 24027. Subsequently, best practices are employed to ensure compliance utilizing SOTA testing methods.

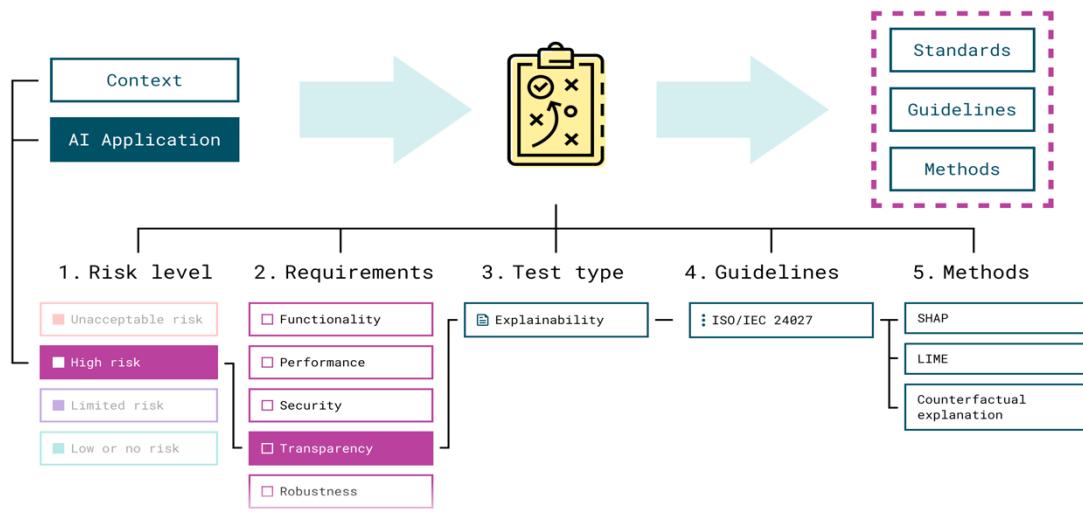


Fig. 13: System Prototype: Evaluation of high-risk AI based on risk levels, requirements, standards, guidelines, and methods.

For high risk applications, explainability testing methodologies like SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), and counterfactual explanations are commonly employed to ensure transparency and meet explainability requirements under Article 13 (Transparency and provision of information to users). By adhering to these structured technical testing processes, we can ensure that high-risk and limited-risk AI applications are in alignment with the AI Act’s requirements and operating safely and ethically.

The Compliance Status offers a summary of the established requirements and methodological frameworks, providing clarity on each application’s alignment with the AI Act. Based on evolving information and compliance status from the above process, an adaptive assessment mechanism is initiated that enables the AI system to evolve, potentially triggering updates to either the context (Component 1) or the risk level (Component 2). This approach ensures that compliance assessments remain relevant

and accurate by making targeted adjustments as needed. Rather than functioning as a fixed feedback loop, it provides flexibility to adapt dynamically to new insights.

6 Conclusion

The Citcom.AI RISE testing approach provides a robust blueprint for evaluating AI across diverse domains and AI subfields by integrating detailed testing methods with the AI Act's structured categorization of risk. This ensures that AI technologies—whether they are involved in healthcare, finance, or any other sector—operate within the bounds of standards and technical requirements, ultimately supporting the safe and effective deployment of AI systems. This approach navigates the complex landscape of AI system testing within the regulatory framework established by the AI Act. The assessment considers various AI application contexts, risk categories and linking them to corresponding testing standards and methodologies, thereby ensuring adherence to stringent regulatory mandates associated with AI systems. The Citcom.AI RISE testing approach serves as a foundational approach for stakeholders aiming to navigate the complexities of AI certification and compliance, fostering a responsible and forward-thinking approach to AI development and usage.

Through our international collaboration programmes with academia, industry, and the public sector, we ensure the competitiveness of the Swedish business community on an international level and contribute to a sustainable society. Our 2,800 employees support and promote all manner of innovative processes, and our roughly 100 testbeds and demonstration facilities are instrumental in developing the future-proofing of products, technologies, and services. RISE Research Institutes of Sweden is fully owned by the Swedish state.

I internationell samverkan med akademi, näringsliv och offentlig sektor bidrar vi till ett konkurrenskraftigt näringsliv och ett hållbart samhälle. RISE 2 800 medarbetare driver och stöder alla typer av innovationsprocesser. Vi erbjuder ett 100-tal test- och demonstrationsmiljöer för framtidssäkra produkter, tekniker och tjänster. RISE Research Institutes of Sweden ägs av svenska staten.



RISE Research Institutes of Sweden AB Box 857, 501 15 BORÅS, SWEDEN Telephone: +46 10-516 50 00 E-mail: info@ri.se , Internet: www.ri.se	RISE Report : ISBN:
--	------------------------