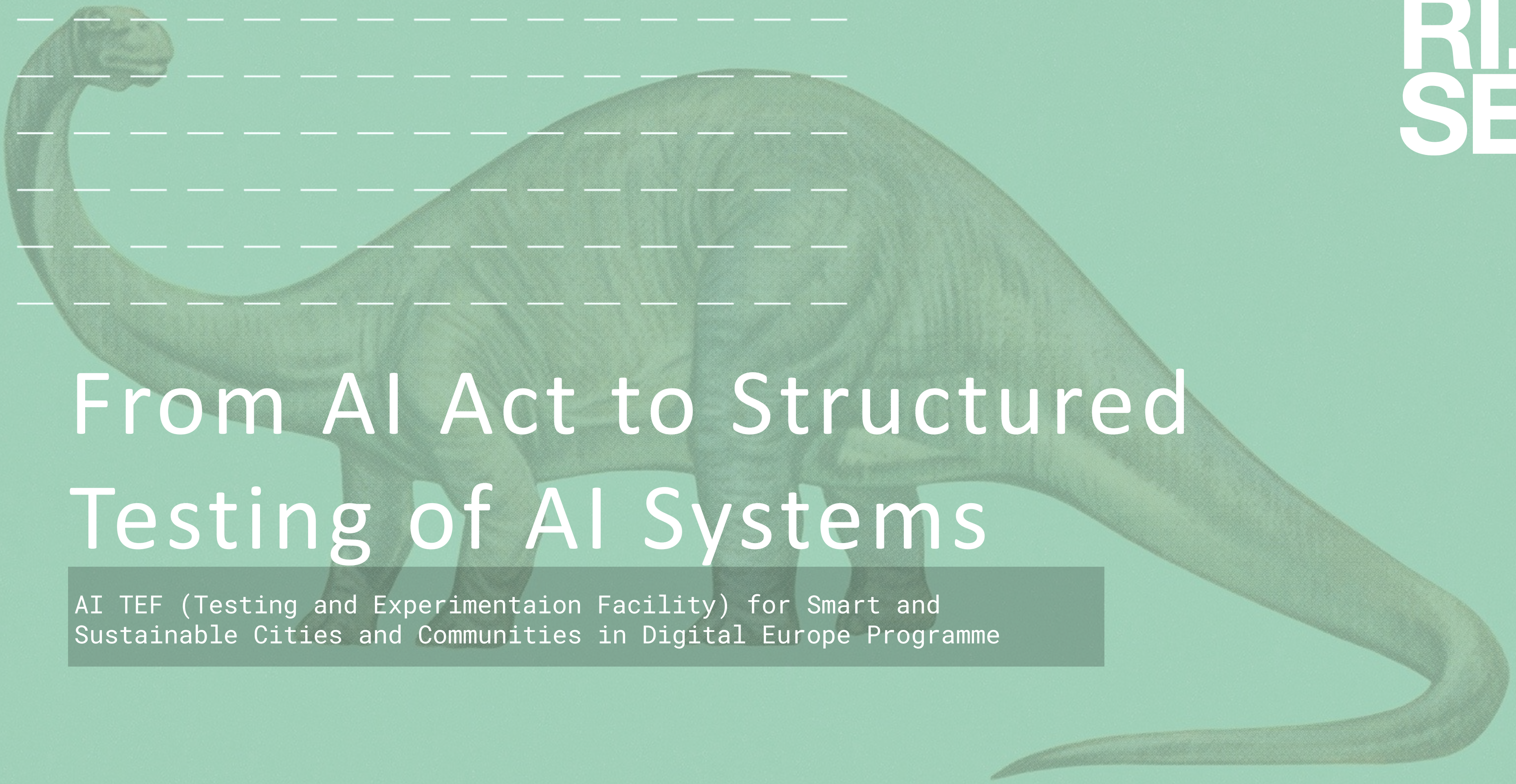


From AI Act to Structured Testing of AI Systems

AI TEF (Testing and Experimentaion Facility) for Smart and Sustainable Cities and Communities in Digital Europe Programme

Katya Mishchenko
Senior Scientist, RISE
kateryna.mishchenko@ri.se

AI Act translation to technical testing and Market Analysis
Optimization, AI Testing
Applied AI

Nishat Mowla
Senior Researcher, RISE
nishat.mowla@ri.se

Trustworthy AI
AI Act translation to technical testing
Machine Learning Scientist
Applied AI applications

Kabir Fahria
Research Engineer, RISE
kabir.fahria@ri.se

Software development
Testing AI
Machine Learning Operations
Large Language Models (LLMs)

Agenda

Advancing and outreaching AI testing by Citcom.AI/RISE

Testing AI approach in Citcom.AI

TEF Collaborations

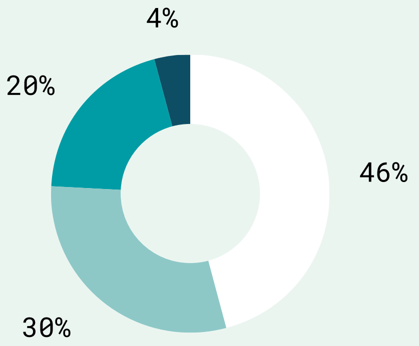
Expanding AI Testing Horizons

3 993

Net sales, MSEK

Operating result: 22 MSEK

Operating margin: 0,6%



Distribution of turnover

□	Industry	1 831 MSEK
▨	Public financiers	1 179 MSEK
▦	State base funding	812 MSEK
■	EU funds	171 MSEK

Alla siffror avser 2022.

Close to

3 300

employees

40%

women

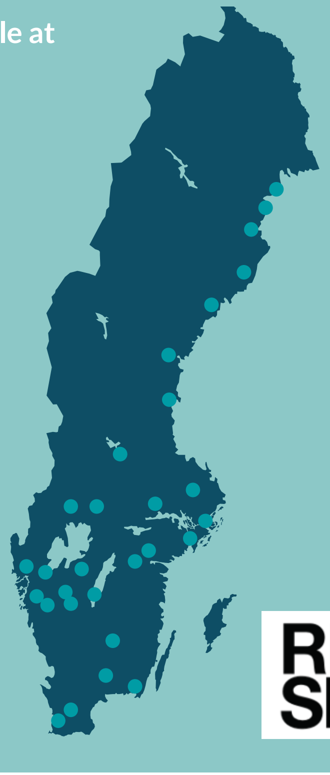
130+

Test and demonstration environment

We are available at

35

Places around Sweden



RI. SE

78

Satisfied Customer Index (2023)

TEF

Testing and Experimental Facilities

(RISE)

Health

Manufacturing

(RISE)

Smart and sustainable cities and societies

(RISE)

AgriFood

(+5)

Advancing and outreaching AI testing by Citcom.AI at RISE

Co-funded by
the European Union

This project has recieved co-funding from the European Union's Digital Europe Program under Grant Agreement No 101100728

About testing AI

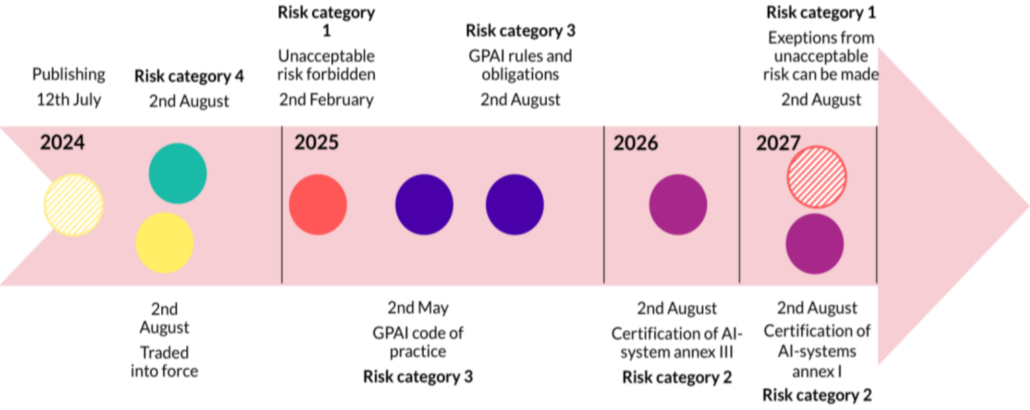
What: Testing AI systems is s a vital part of the development and deployment of AI systems since it ensures their accuracy, reliability, safety & security, efficiency and effectiveness

Why: AI testing builds trust and confidence in real-world applications and helps in identifying and rectifying potential issues early, thereby improving the quality of software releases.

How: One instrument to emphasize the importance and ensure the safety and reliability of AI systems is the **AI Act** .

It lays down harmonized rules on AI, aiming to balance the socio-economic benefits and potential risks of AI technologies placed on the European market.

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206



Publishing
12th July

Risk category 4
2nd August

Risk category 1
Unacceptable risk forbidden
2nd February

Risk category 3
GPAI rules and obligations
2nd August

Risk category 1
Exeptions from unacceptable risk can be made
2nd August

2024

2025

2026

2027

2nd August
Traded into force

2nd May
GPAI code of practice
Risk category 3

2nd August
Certification of AI-system annex III
Risk category 2

2nd August
Certification of AI-systems annex I
Risk category 2

Understanding challenges in Testing AI

Tasks/Challenges according to AI Act:

- Safety and Risk Compliance
- Data Quality
- Transparency
- Accountability
- Robustness and Security
- Ongoing monitoring
- Documentation

How to test AI in practice?

- Defining the risk category: who, how and by which means?
- Which legal regulations and standards are relevant?
- Reducing the risk category level: is it possible, how?
- Testing: which methods to apply, how it depends on the role of the actor, who?
- Which tools/platforms to use?
- …….

Testing AI at Citcom.AI by RISE. Framework

Goal

Provide services for testing of AI systems in order ensure safe and secure AI in society and industry

Needs

Establish mature methodologies, procedures and platform for testing of AI systems incl. models, documented experience, customer requirements, mappings to AI act, standards and technologies

Approach

With a use case and context driven approach we explore potential services in testing of AI within the TEF which in future aims to lead to offering certification of AI systems

Benefits

The UCs are envisioned to connect, relate, and define how an AI model is tested, secured, and monitored by ensuring relevant activities and identifying potential AI testing service offerings based on real needs

Testing AI at Citcom.AI by RISE. Activities and Competence

Market Analysis:
- Code of practice
- Actors active in testing AI
- Existing tools/platforms

Understanding the legal part:

- AI Act
- Relevant AI Standards and Guidelines
- Participation in SIS/ISO activities

Technical work:

- Technical state-of-the-art reviews
- Development of testing approach incl. methods and tools

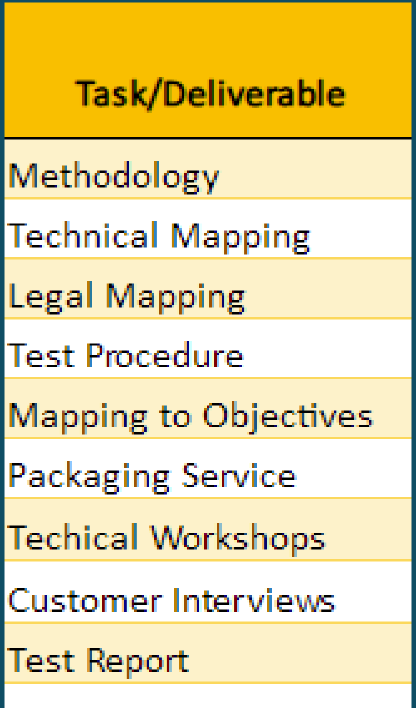
Use Cases design:

- Finding & Preparing cases and background for further testing
- Creation of pilot AI testing offerings

Ongoing UCs:

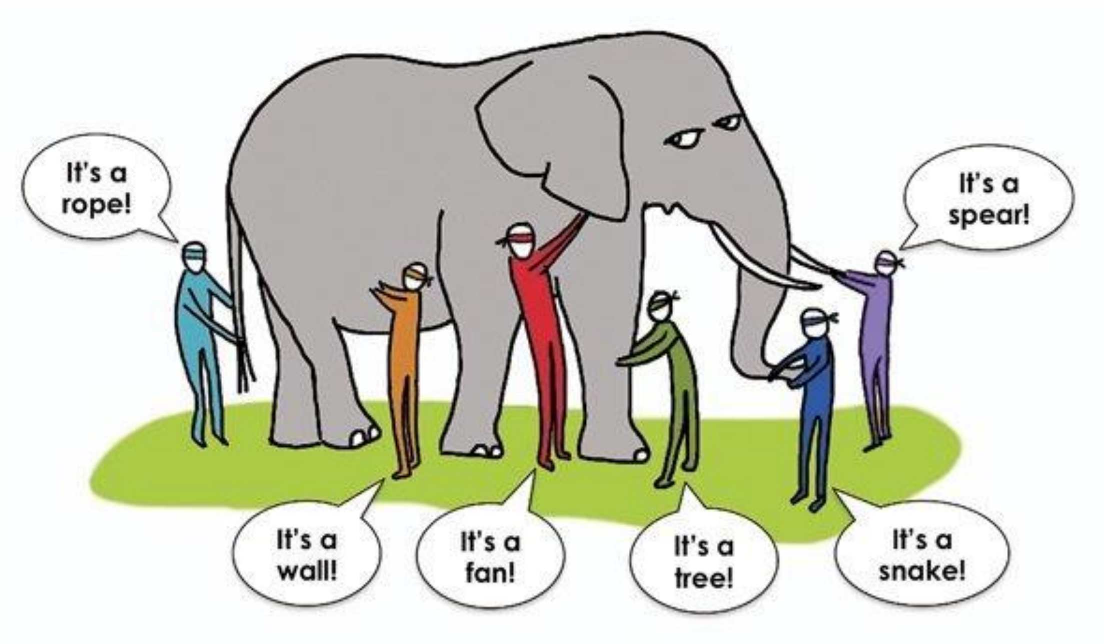
- • District Heating (electric power and district heating provider)
- Testing RAG LLM (Swedish National Financial Management Authority)
- Intrusion Detection System (Scania)
- ...

Promoting and Advertising testing AI
- Presentations and WSs
- Communication with potential partners

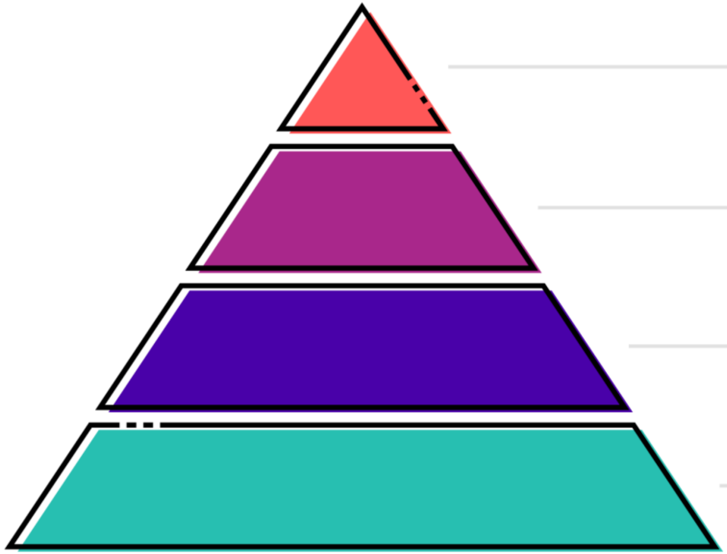


Task/Deliverable
Methodology
Technical Mapping
Legal Mapping
Test Procedure
Mapping to Objectives
Packaging Service
Techical Workshops
Customer Interviews
Test Report

Make sense of testing AI systems



AI Regulation - risk categories



🟥 Unacceptable risk

Banned AI systems because they are seen as a clear threat to peoples’s livelihoods and rights

🟪 High risk

AI systems subject to strict obligations and a conformity assessment before they can be placed on the market

🟪 Limited risk

AI systems with code of conduct requirements and specific transparency obligations

🟩 Low or no risk

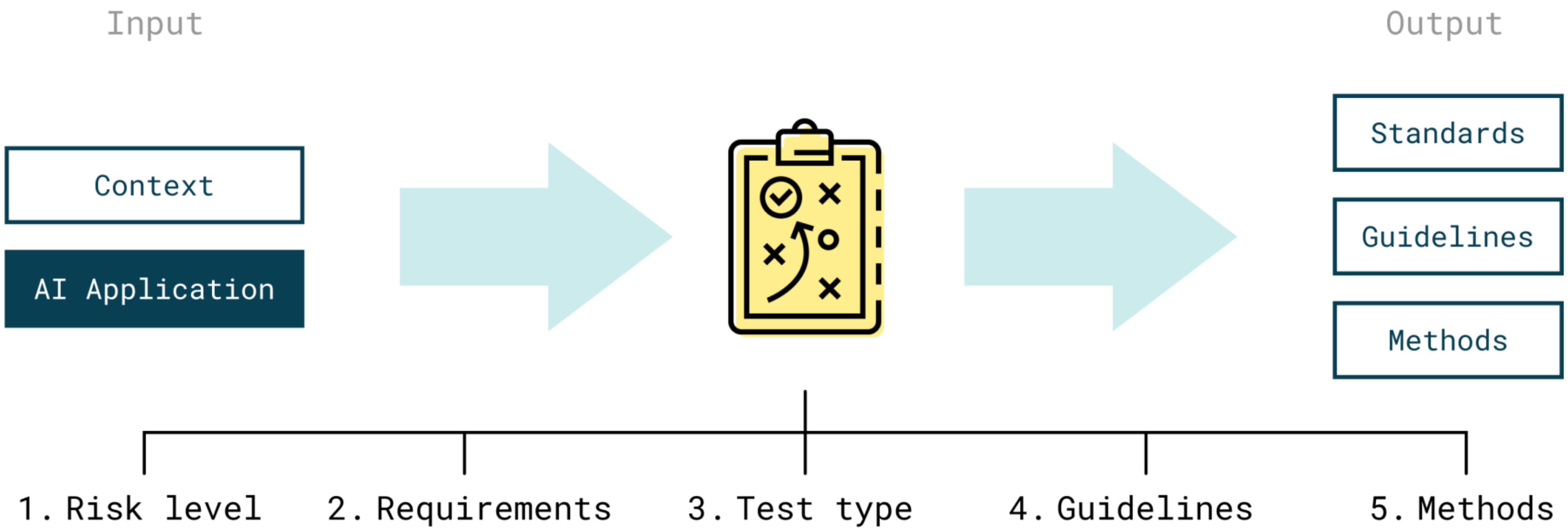
The majority of AI systems with minimal or no risk that can be used freely and are not subject to the AI Act

AI Regulation - risk categories

Requirements per risk category

■ Low or no risk	■ Limited risk	■ High risk		■ Unacceptable risk
✓ No certification required	Transparency ***Article 13***	Functionality ***Article 15***	Safety ***Article 15***	✕ Drop
	Documentation ***Article 11 & 12***	Performance ***Article 15***	Data quality ***Article 10***	
	Functional / non-functional	Security ***Article 15***	Risk management ***Article 9***	
		Transparency ***Article 13***	Human oversight ***Article 14***	
		Robustness ***Article 15***	Documentation ***Article 11 & 12***	

Mapping testing methods to AI Act



Input

Context

AI Application

Output

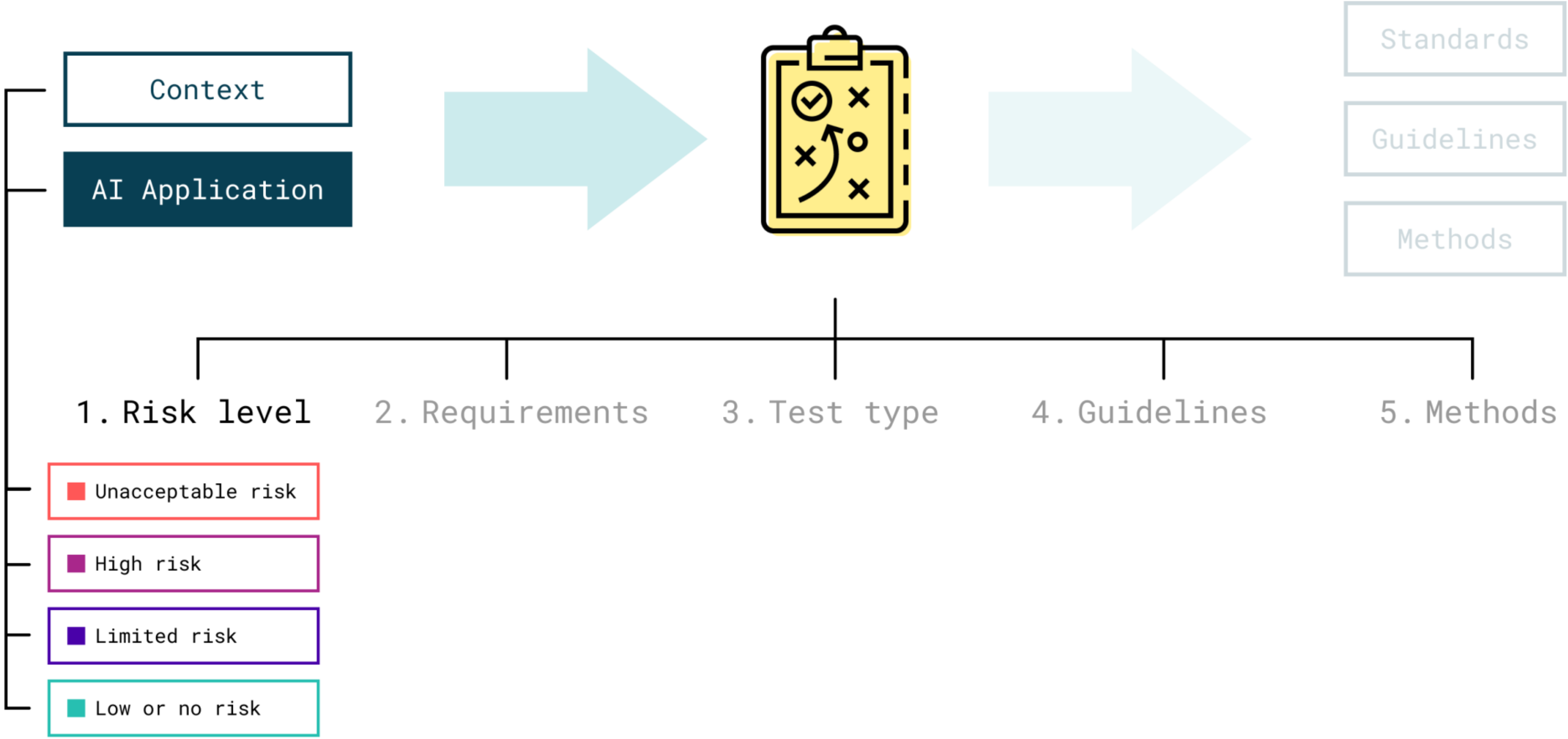
Standards

Guidelines

Methods

1. Risk level
2. Requirements
3. Test type
4. Guidelines
5. Methods

Mapping testing methods to AI Act



Context

AI Application

Standards

Guidelines

Methods

1. Risk level

2. Requirements

3. Test type

4. Guidelines

5. Methods

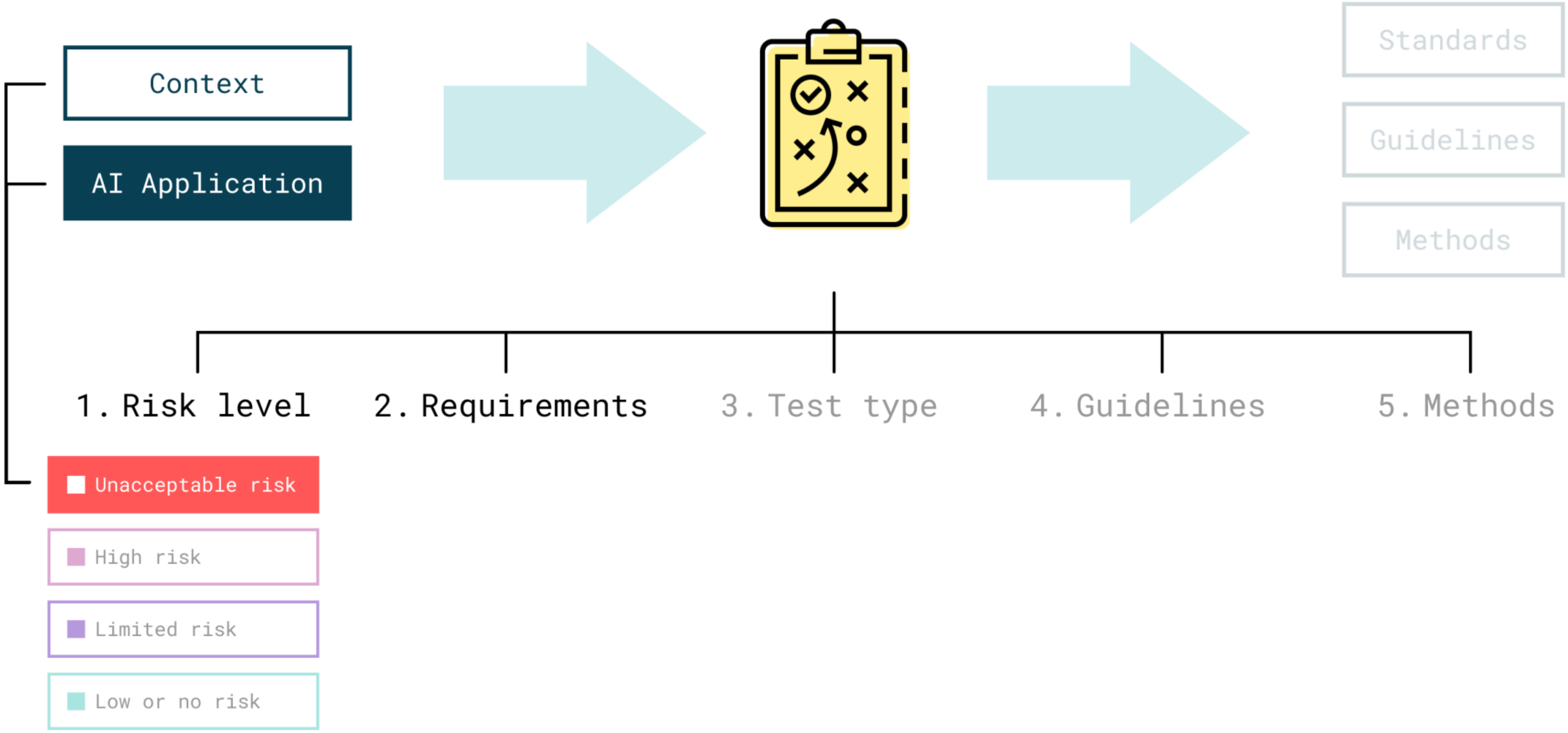
🟥 Unacceptable risk

🟪 High risk

🟪 Limited risk

🟩 Low or no risk

Mapping testing methods to AI Act



Context

AI Application

Standards

Guidelines

Methods

1. Risk level

2. Requirements

3. Test type

4. Guidelines

5. Methods

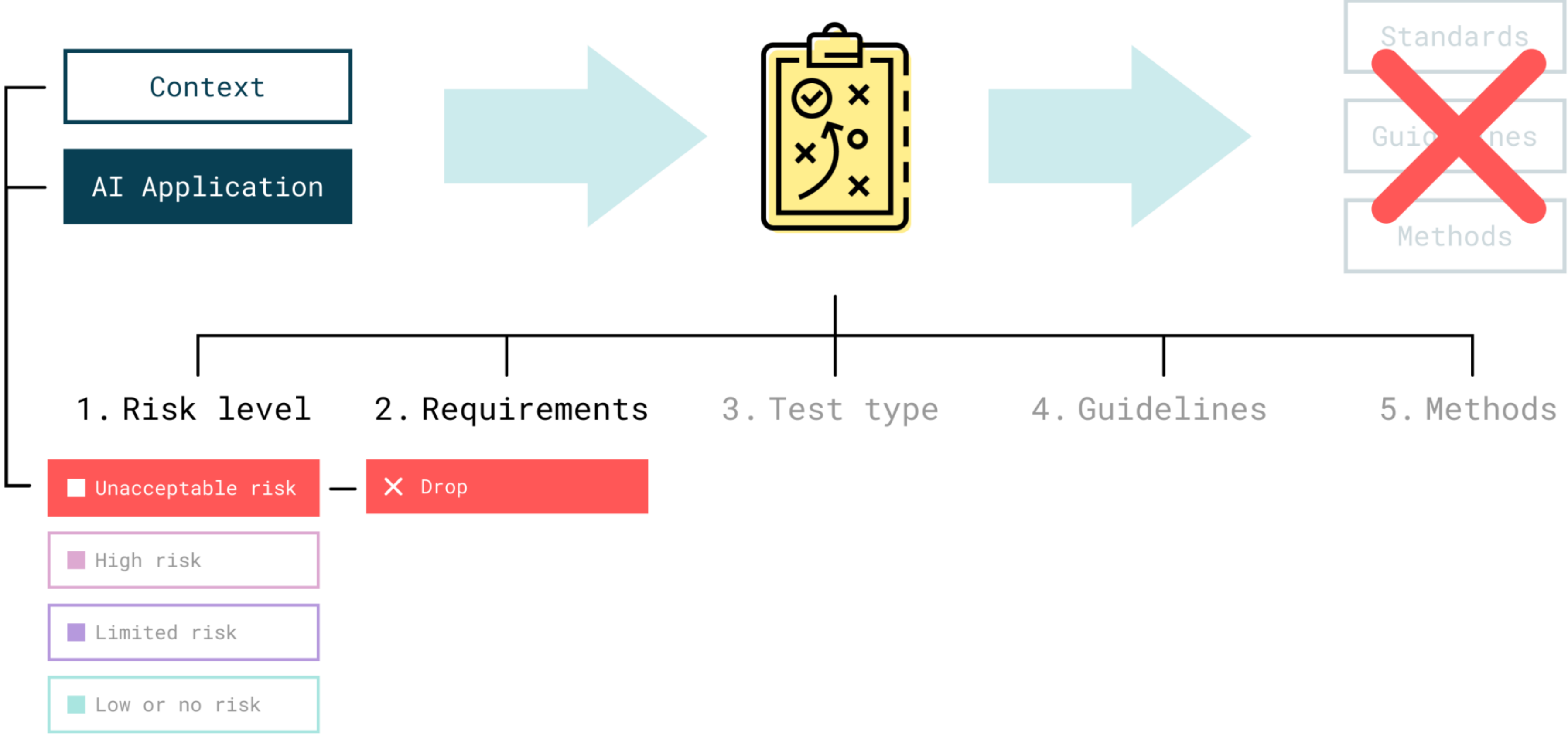
■ Unacceptable risk

■ High risk

■ Limited risk

■ Low or no risk

Mapping testing methods to AI Act



Context

AI Application

Standards

Guidelines

Methods

1. Risk level

2. Requirements

3. Test type

4. Guidelines

5. Methods

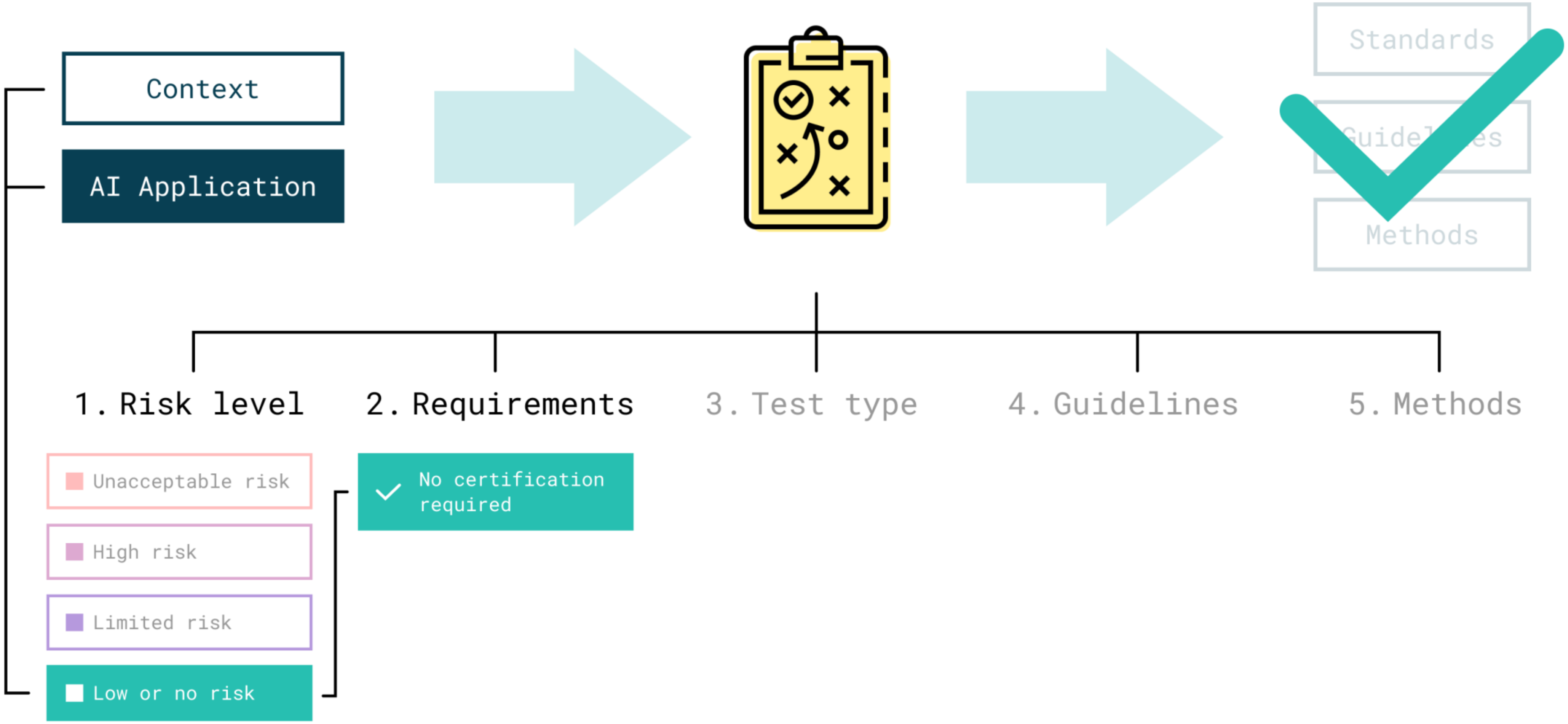
□ Unacceptable risk — ✕ Drop

■ High risk

■ Limited risk

■ Low or no risk

Mapping testing methods to AI Act



Context

AI Application

Standards

Guidelines

Methods

1. Risk level

2. Requirements

3. Test type

4. Guidelines

5. Methods

Unacceptable risk

High risk

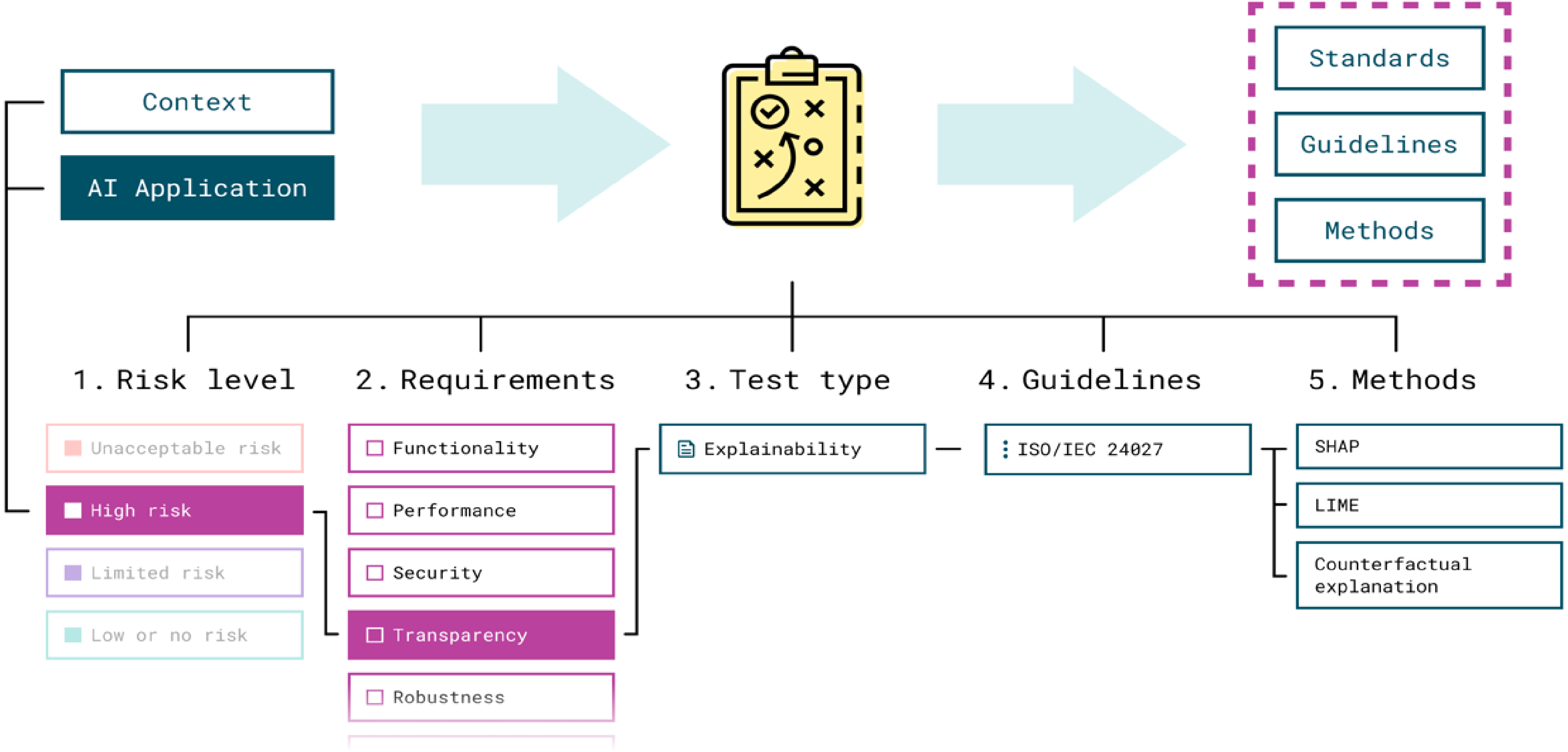
Limited risk

Low or no risk

✓ No certification required

RI. SE

Mapping testing methods to AI Act



Context

AI Application

Standards

Guidelines

Methods

1. Risk level

- Unacceptable risk
- High risk
- Limited risk
- Low or no risk

2. Requirements

- □ Functionality
- □ Performance
- □ Security
- □ Transparency
- □ Robustness

3. Test type

- Explainability

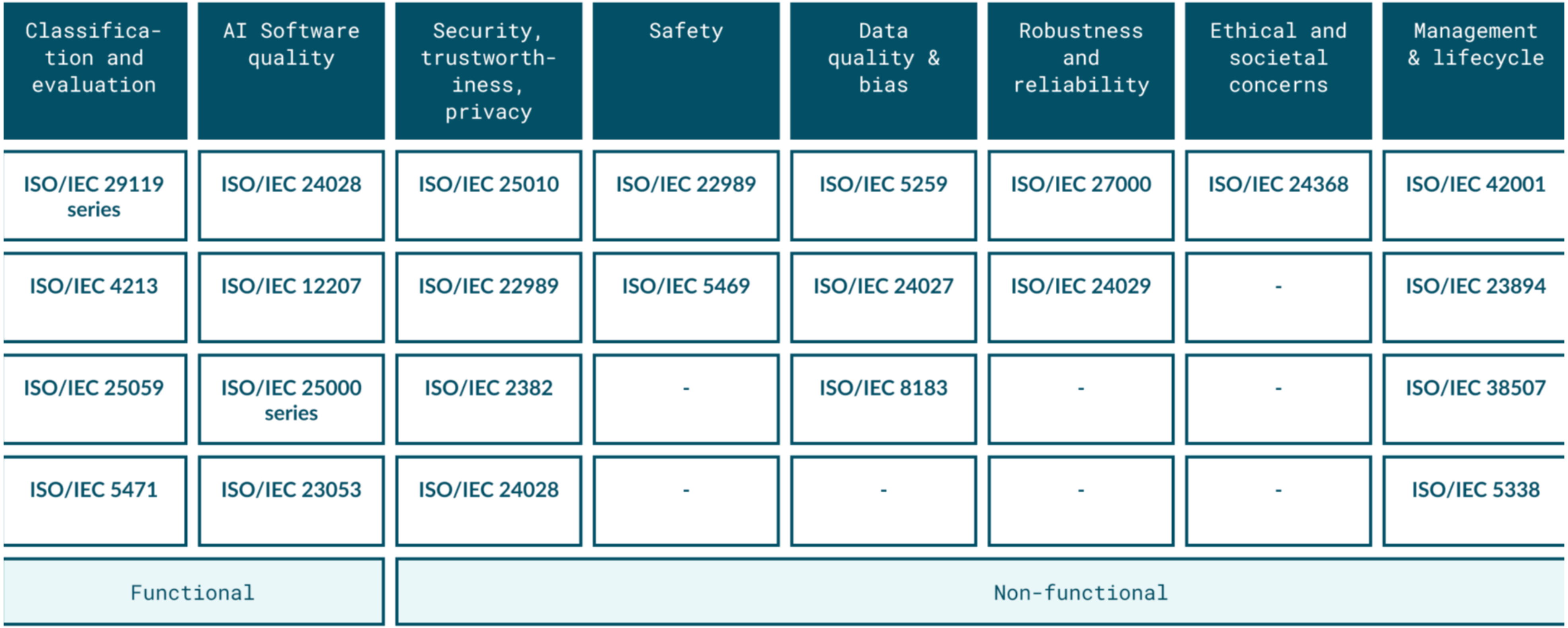
4. Guidelines

- ISO/IEC 24027

5. Methods

- SHAP
- LIME
- Counterfactual explanation

Mapping Standards

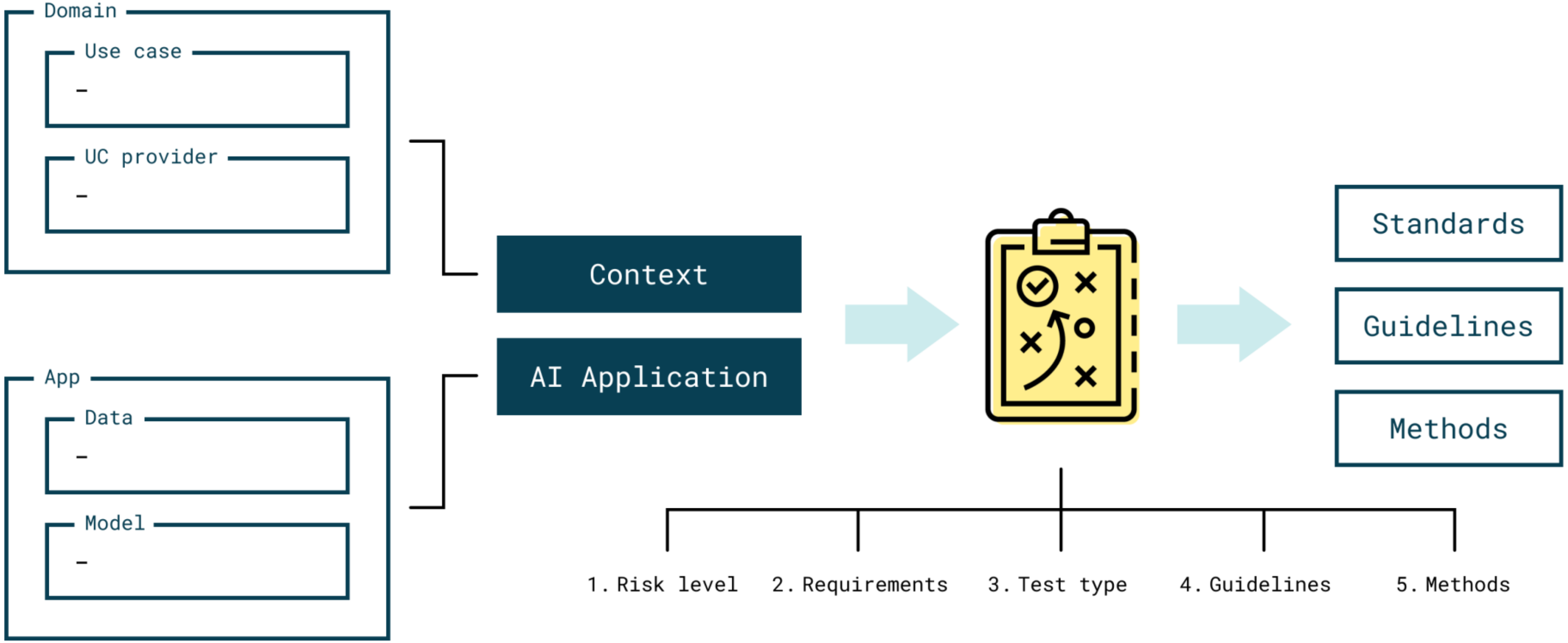


Classification and evaluation	AI Software quality	Security, trustworthiness, privacy	Safety	Data quality & bias	Robustness and reliability	Ethical and societal concerns	Management & lifecycle
ISO/IEC 29119 series	ISO/IEC 24028	ISO/IEC 25010	ISO/IEC 22989	ISO/IEC 5259	ISO/IEC 27000	ISO/IEC 24368	ISO/IEC 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					

AI Testing guidelines

	Risk management	Data and data governance	Technical documentation	Record keeping	Transparency	Human oversight	Accuracy, robustness & cybersecurity
AI Act article	Article 9	Article 10	Article 11	Article 12	Article 13	Article 14	Article 15
EU ALTAI	Requirement 7,	Requirement 3,	Requirement 7,	Requirement 7,	Requirement 4,	Requirement 1,	Requirement 2,
Requirement	Risk management testing	Data quality testing	Documentation testing	Record keeping testing	Transparency testing	Human oversight testing	Security testing, safety testing

Our AI testing approach



Domain

Use case

–

UC provider

–

App

Data

–

Model

–

Context

AI Application

Standards

Guidelines

Methods

1. Risk level

2. Requirements

3. Test type

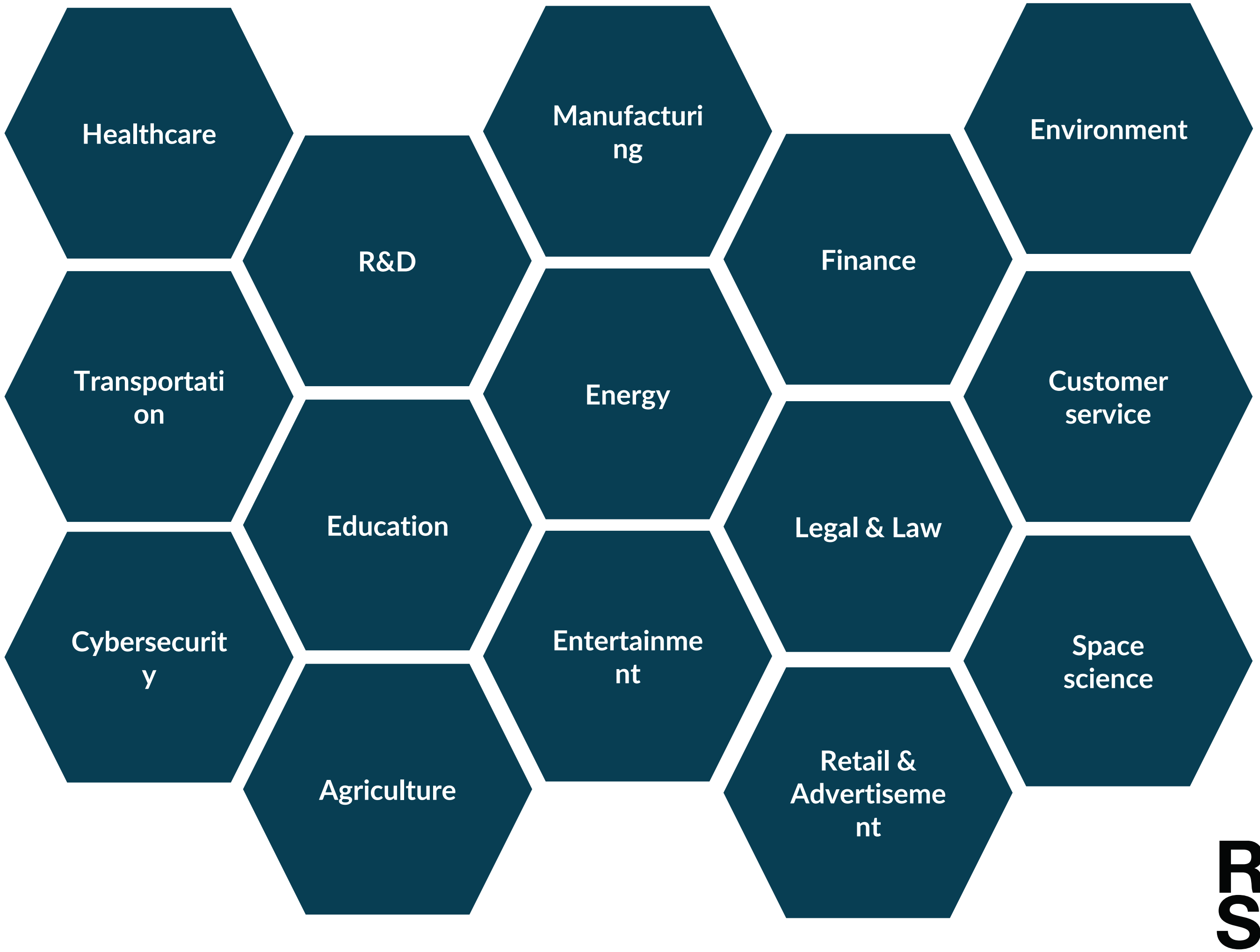
4. Guidelines

5. Methods

Application domains
Subfields of AI

Subfields of AI:

1. Machine learning
2. Deep learning (include DNN)
3. Natural language processing (include LLM)
4. Computer vision (image, video, voice)
5. Reinforcement learning (agents)
6. Robotics (autonomous)
7. Speech and audio processing (speech recognition)
8. Planning and scheduling (plan actions)
9. Evolutionary computing (genetic algorithm)
10. Affective computing (recognize feelings)



Healthcare

Transportation

Cybersecurity

R&D

Education

Agriculture

Manufacturing

Energy

Entertainment

Finance

Legal & Law

Retail & Advertisement

Environment

Customer service

Space science

Application domains
Subfields of AI

Subfields of AI:

1. Machine learning
2. Deep learning (include DNN)
3. Natural language processing (include LLM)
4. Computer vision (image, video, voice)
5. Reinforcement learning (agents)
6. Robotics (autonomous)
7. Speech and audio processing (speech recognition)
8. Planning and scheduling (plan actions)
9. Evolutionary computing (genetic algorithm)
10. Affective computing (recognize feelings)

Healthcare

Transportation

Cybersecurity

R&D

Education

Agriculture

Manufacturing

Energy

Entertainment

Finance

Legal & Law

Retail & Advertisement

Environment

Customer service

Space science

Explainability in Automotive Intrusion Detection System

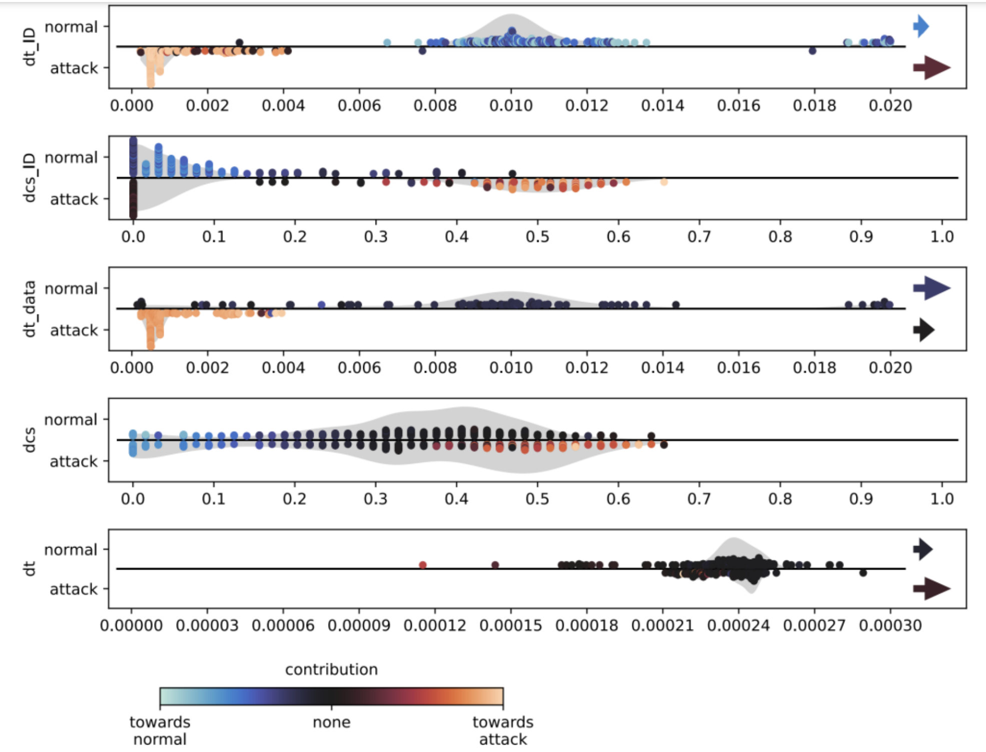


FIGURE 2. **VisExp | A pseudo-global visualization-based explanation, using SHAP values. It shows the features in the dataset in swarm plot-like strips for normal and attack classifications. Each point is an instance from the train data. The x-axes are the feature values, and the color represents the SHAP values. The color of the arrows represent the mean of the SHAP values outside of the diagram, and their relative size represents how many data points there are.**

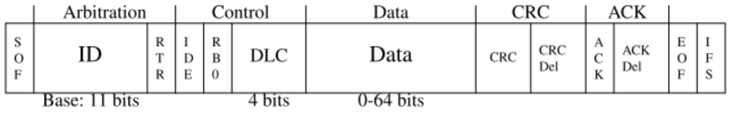
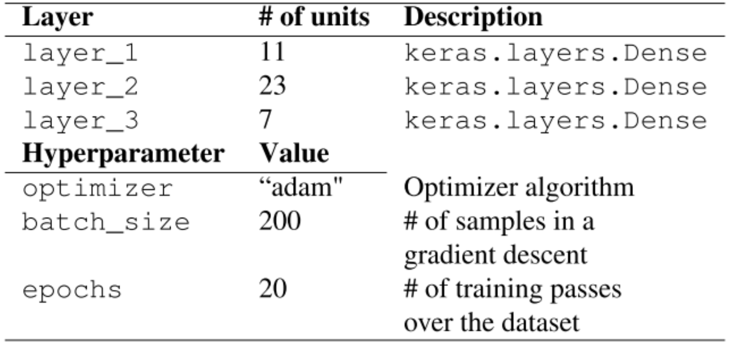


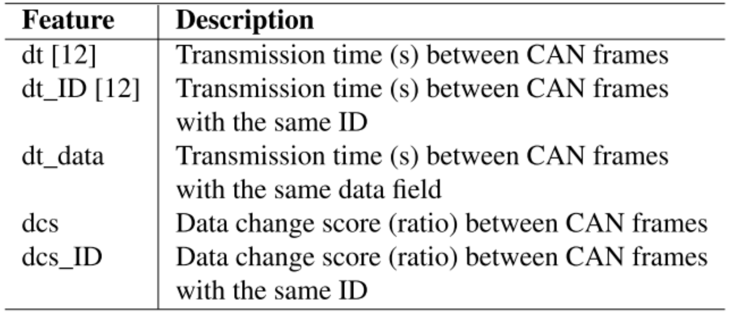
FIGURE 1. **CAN frame | The Survival dataset has features of the ID, DLC and data field, along with the timestamp of when a CAN frame is transmitted.**

TABLE 1. **DNN hyperparameters | Parameters and their values as specified when building the DNN in keras.**



Layer	# of units	Description
layer_1	11	keras.layers.Dense
layer_2	23	keras.layers.Dense
layer_3	7	keras.layers.Dense
Hyperparameter	**Value**	
optimizer	“adam"	Optimizer algorithm
batch_size	200	# of samples in a gradient descent
epochs	20	# of training passes over the dataset

TABLE 2. **The engineered features.**



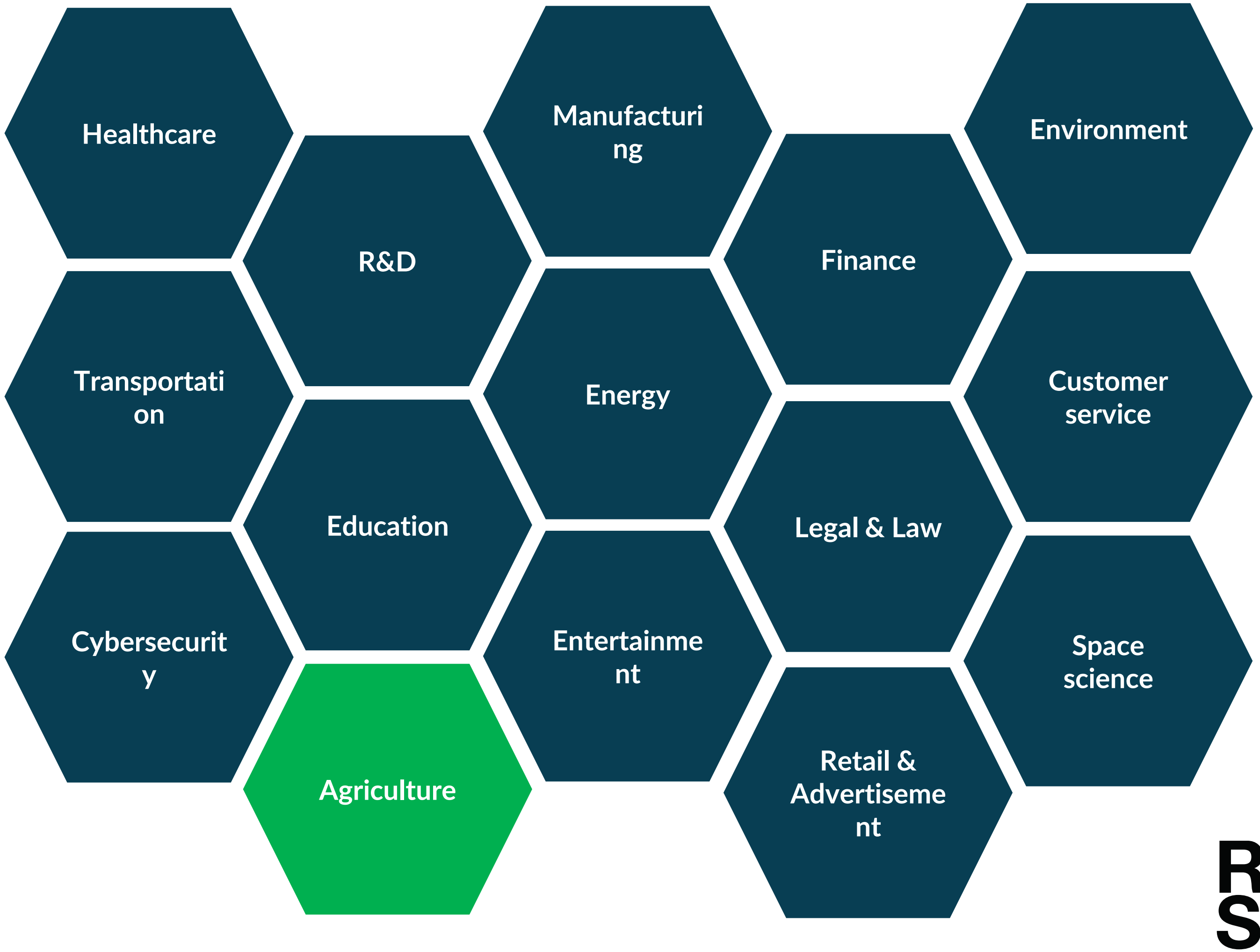
Feature	Description
dt [12]	Transmission time (s) between CAN frames
dt_ID [12]	Transmission time (s) between CAN frames with the same ID
dt_data	Transmission time (s) between CAN frames with the same data field
dcs	Data change score (ratio) between CAN frames
dcs_ID	Data change score (ratio) between CAN frames with the same ID

Hampus Lundberg, Nishat I Mowla, Sarder Fakhrul Abedin, Kyi Thar, Aamir Mahmood, Mikael Gidlund, Shahid Raza, “Experimental Analysis of Trustworthy In-Vehicle Intrusion Detection System Using eXplainable Artificial Intelligence (XAI),” IEEE Access, vol. 10, September 2022. (Link)

Application domains
Subfields of AI

Subfields of AI:

1. Machine learning
2. Deep learning (include DNN)
3. Natural language processing (include LLM)
4. Computer vision (image, video, voice)
5. Reinforcement learning (agents)
6. Robotics (autonomous)
7. Speech and audio processing (speech recognition)
8. Planning and scheduling (plan actions)
9. Evolutionary computing (genetic algorithm)
10. Affective computing (recognize feelings)



Healthcare

Transportati
on

Cybersecurit
y

R&D

Education

Agriculture

Manufacturi
ng

Energy

Entertainme
nt

Finance

Legal & Law

Retail &
Advertiseme
nt

Environment

Customer
service

Space
science

Machine learning to classify peppers

- AI in an industrial computer vision system
- Trained deep learning model
- Images from 3D camera to distinguish between good and bad peppers
- Parallel robotic with pneumatic end-effector performs a sorting task of the peppers on the conveyor belt

ILVO
Flanders Research Institute for
Agriculture, Fisheries and Food

Expanding AI Testing Horizons: Expertise, Compliance, and Partnership

Despite being at the beginning of our journey, we have extensive experience in testing AI systems and are continuously gaining new insights.

We have developed specialized testing approaches tailored to meet unique challenges of testing AI. Our team is well-versed not only in the technical aspects of testing but also, we are learning to map the legal and regulatory requirements, ensuring smooth implementation and compliance across projects.

We are seeking partners and co-developers to expand our use-case-based approach, designed to create pilot testing procedures across the AI lifecycle.

Collaboration with Citcom.AI

Enhance Your AI Projects with Expert Testing & Evaluation

- We invite all AI projects to consider integrating comprehensive testing and evaluation of AI systems within your project scopes

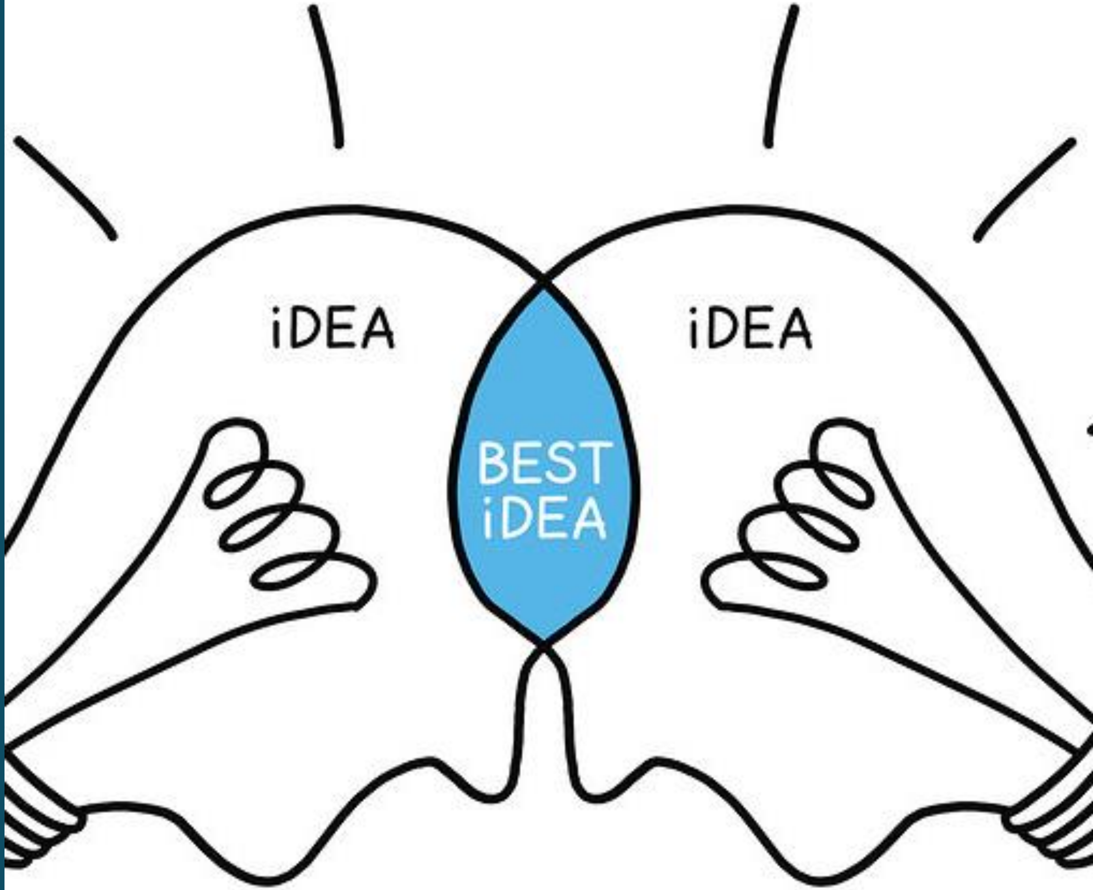
- Adding structured testing and evaluation not only aligns with AI Act compliance but also strengthens your project by demonstrating a commitment to quality and reliability

How We Can Help

- Expert Guidance: Assist you in crafting testing and evaluation of AI systems

- Flexible Collaboration: Choose to conduct the testing yourselves or partner with us for joint execution

Reach out!



Nishat Mowla
Senior Researcher, RISE
nishat.mowla@ri.se

Trustworthy AI
AI Act translation to technical testing
Machine Learning Scientist
Applied AI applications

Katya Mishchenko
Senior Scientist, RISE
kateryna.mishchenko@ri.se

AI Act translation to technical testing and Market Analysis
Optimization, AI Testing
Applied AI

Kabir Fahria
Research Engineer, RISE
kabir.fahria@ri.se

Software development
Testing AI
Machine Learning Operations
Large Language Models (LLMs)

Thanks!

RI. SE