Exploring AI Testing: Introduction and Methodology

March 12th, 2024

Kateryna Mishchenko & Nishat I Mowla

RISE

Introduction to AI testing

About AI testing

What: Testing AI systems is s a vital part of the development and deployment of AI systems since it ensures their accuracy, reliability, safety, 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** (enters into force 2024-2026).

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.

AI Act: risk-based approach

UNACCEPTABLE RISK:
AI systems prohibited, because considered a clear threat to the safety, livelihoods and rights of people

HIGH-RISK:
AI system subject to strict obligations and a conformity assessment before they could be put on the market

LIMITED-RISK:
AI system with specific transparency obligations

LOWER-RISK:
They represent the vast majority of AI system with minimal risk or no risk, allowed to be used freely without any obligation from the discussed regulation

Source: https://www.iasonltd.com/doc/jit/2021/European_Commission_Regulation_on_AI.pdf

About the Standards related to AI Act

- **ISO/IEC TR 29119-11:2020 Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems**

  Provides an introduction to AI-based systems, new challenges and opportunities for testing them.

  This document explains those characteristics which are specific to AI-based systems and explains the corresponding difficulties of specifying the acceptance criteria for such systems.

- **ISO/IEC AWI TS 29119-11 Software and systems engineering — Software testing — Part 11: Testing of AI systems**

  Describes testing techniques applicable for AI systems in the context of the AI system life cycle model stages

  Shows how AI and ML assessment metrics can be used in the context of those testing techniques. It also maps testing processes to the verification and validation stages in the AI system life cycle.

- **ISO/IEC 25059 Software engineering. Systems and software Quality Requirements and Evaluation (SQua**RE)

  Outlines a quality model for AI systems and provide guidelines for measuring and evaluating the quality of AI systems, focusing on characteristics like accuracy, interpretability, robustness, fairness, privacy, and security.

Challenges in AI Testing

Some challenges related to AI testing

- Testing AI systems comes with unique challenges, such as the unpredictability of AI behaviour, the difficulty in defining the right metrics for success, and the complexity of creating diverse and representative test cases.
- ISO/IEC AWI TS 29119-11 "Software and systems engineering — Software testing — Part 11: Testing of AI systems" describes testing techniques and metrics for AI systems in the context of the AI system life cycle model stages. According to it, some of **challenges** are:

Data testing:

issues with data quality, diversity, privacy, labeling, temporal sequencing, data drift, and potential biases.

Explainability:

Arises from "black box", nature, making it difficult to understand why they make certain decisions.

Continuous Learning:

often learn and adapt over time, which means they need to be continuously tested and monitored

Transparency:

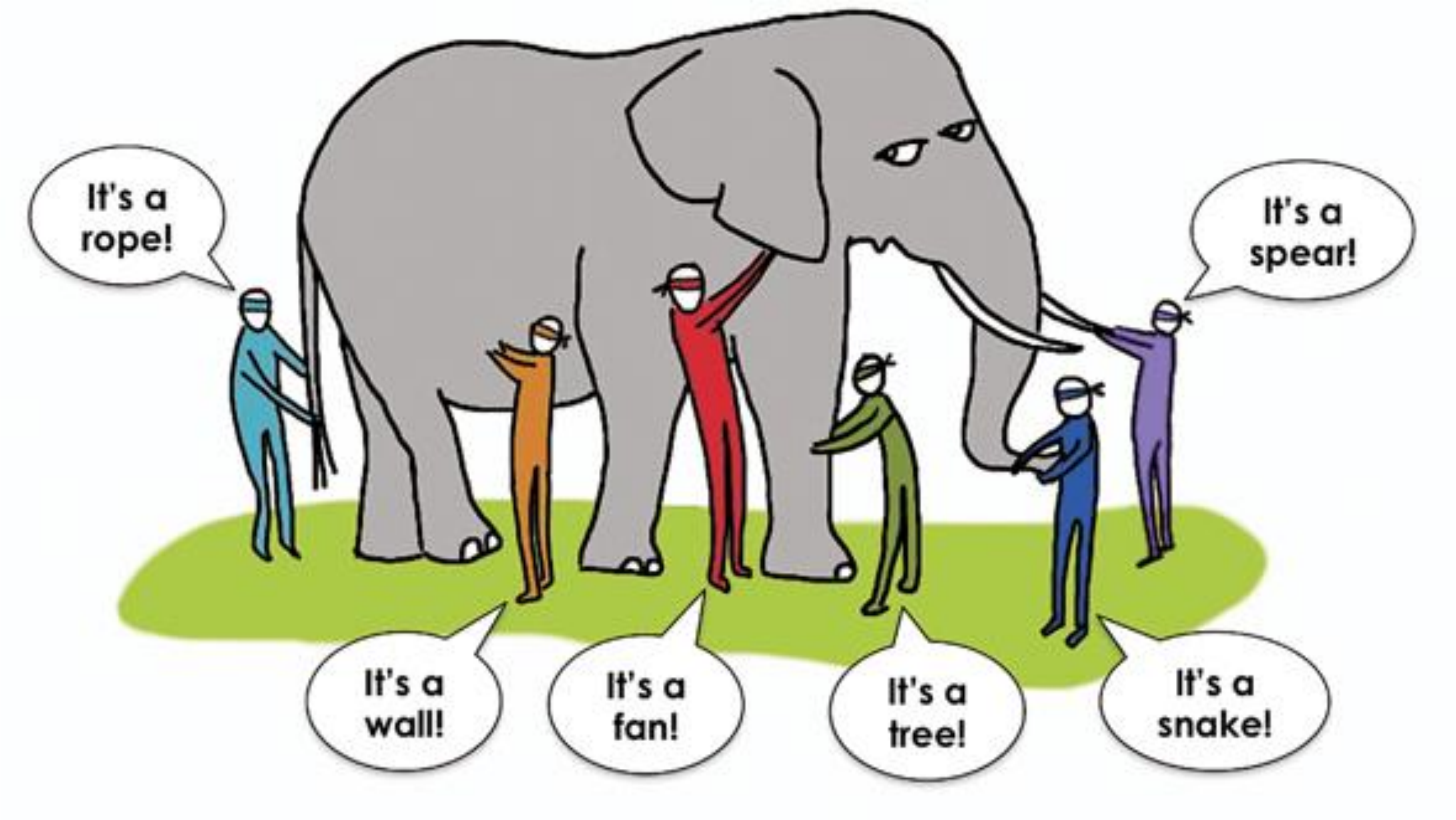
arises "black box" nature, sensitivity of training data, dynamic learning, potential for bias, and the trade-off between model accuracy and explainability.

Trustworthiness:

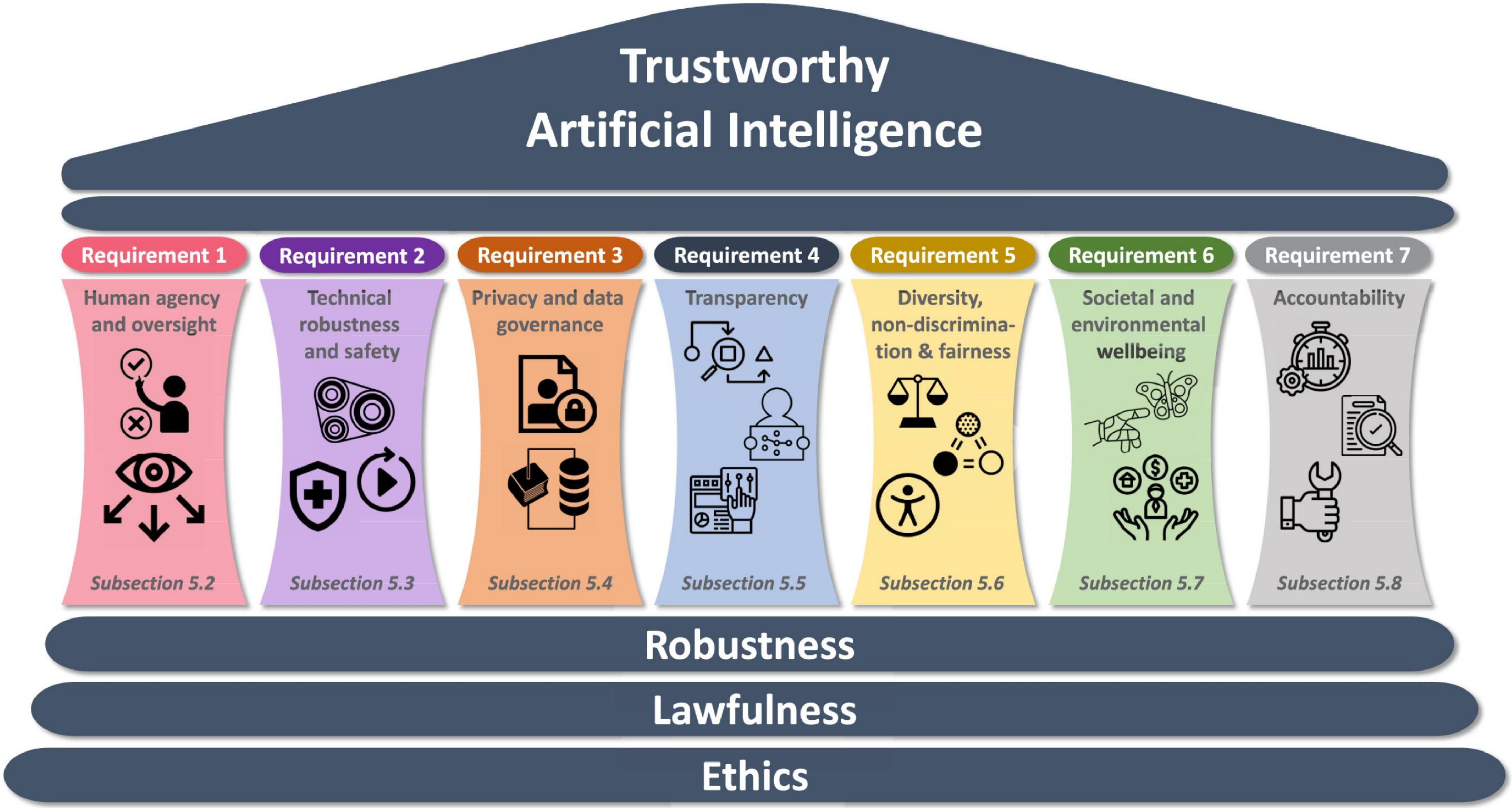
arises from the "black box" nature, the need for security against manipulation, the requirement for data privacy, the necessity for accountability, and the complexity of ensuring fairness and non-discrimination.

AI testing methodology

The AI Testing Elephant



Assessment list of Trustworthy AI (ALTAI)



Ethics guidelines for Trustworthy AI: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
Image source: https://www.sciencedirect.com/science/article/pii/S1566253523002129

AI Testing Standards

Classification and evaluation	AI Software quality	Security, trustworthiness, privacy	Safety	Data quality & bias	Robustness and reliability	Ethical and societal concerns	Management & Lifecycle	Risk management
ISO/IEC 29119 series	ISO/IEC 23053	ISO/IEC 22989	ISO/IEC 22989	ISO/IEC 5259	ISO/IEC 27001	ISO/IEC 24368	ISO/IEC 42001	ISO/IEC 23894
ISO/IEC 4213	ISO/IEC 24028	ISO/IEC 20547	ISO/IEC 5469	ISO/IEC 24027	ISO/IEC 24029		ISO/IEC 42006	ISO/IEC 31000
ISO/IEC 25059	ISO/IEC 25000	ISO/IEC 24028					ISO/IEC 38507	
ISO/IEC 42102							ISO/IEC 5338	
Functional		Non-functional						

Application domains and subfields of AI

Subfields of AI:

1. Machine learning
2. Deep learning (DNN)
3. Natural language processing (LLM)
4. Computer vision (image, video, voice)
5. Reinforcement learning (agents)
6. Multi-agent systems
7. Robotics (autonomous)
8. Expert systems (reasoning)
9. Speech processing (speech recognition)
10. Planning and scheduling (plan actions)
11. Knowledge representation and reasoning
12. Evolutionary computing (genetic algorithm)
13. Affective computing (recognize feelings)

Healthcare

Manufacturing

Environment

R&D

Finance

Transportation

Energy

Customer service

Education

Legal & Law

Cybersecurity

Entertainment

Robotics

Agriculture

Retail

Performing AI Testing

Performing AI Explainability Testing

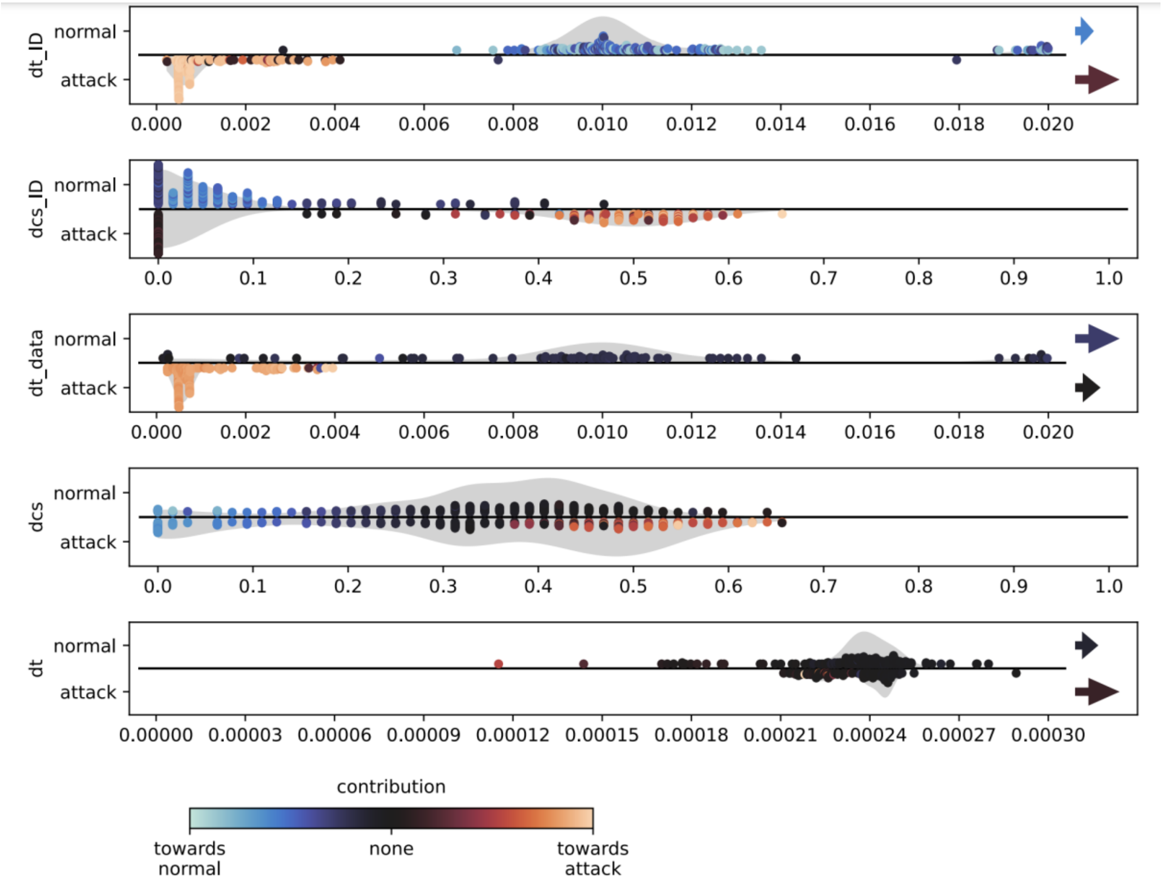


FIGURE 2. VisExp | A pseudo-global visualization-based explanation, using SHAP values. It shows the features in the dataset in swarm plot-like strips for normal and attack classifications. Each point is an instance from the train data. The x-axes are the feature values, and the color represents the SHAP 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.

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)

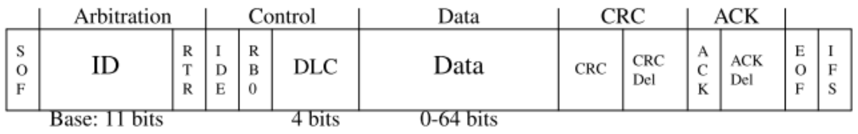


FIGURE 1. CAN frame | The Survival dataset has features of the ID, DLC and data field, along with the timestamp of when a CAN frame is transmitted.

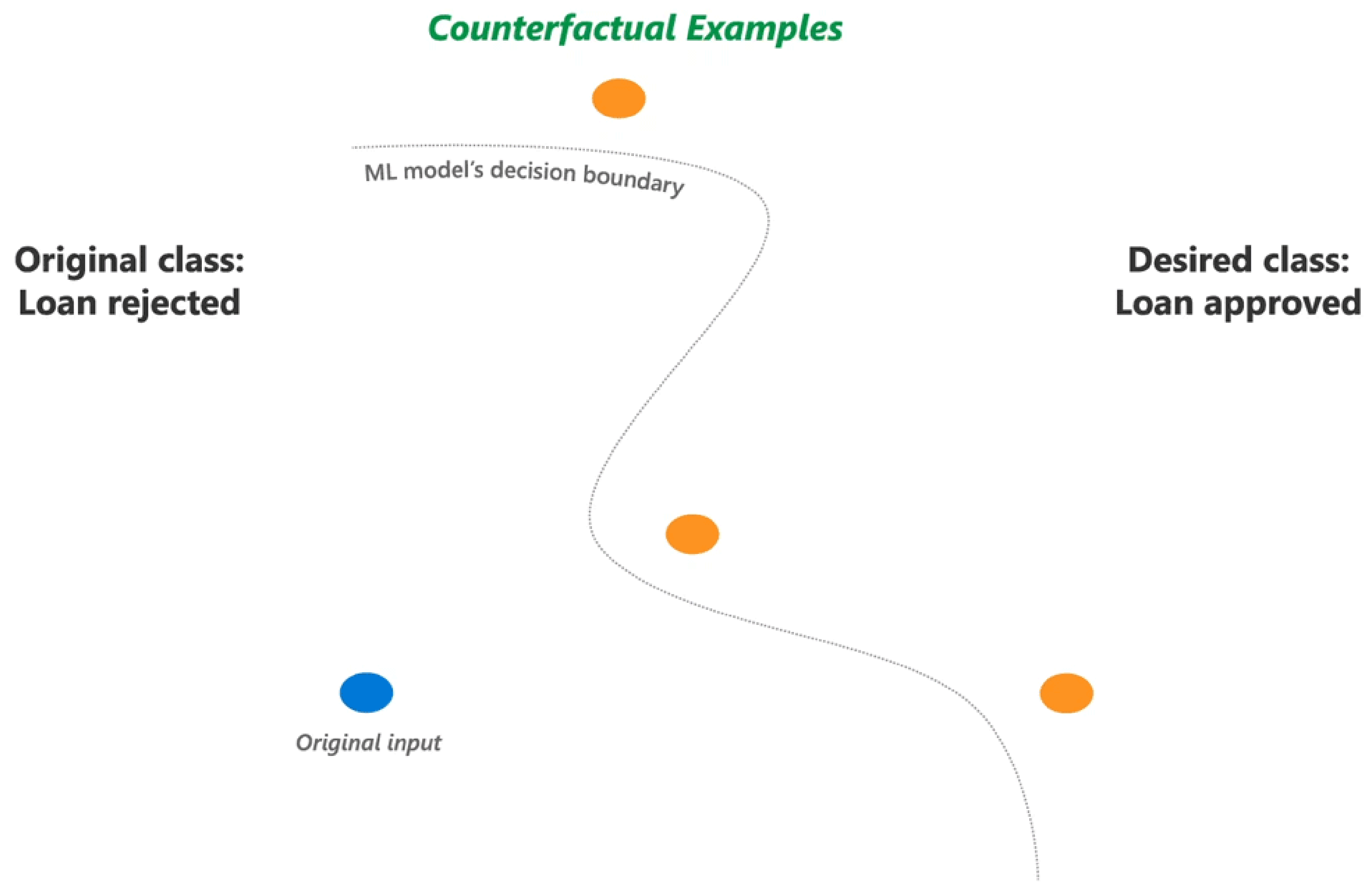
TABLE 1. DNN hyperparameters | Parameters and their values as specified when building the DNN in keras.

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

Performing AI Explainability Testing



Microsoft Research

Quality of AI

Quality AI requires quality data

But quality AI is more than data

- **Cybersecurity**
- **Transparency**
- **Robustness**
- **more**

Thanks!

kateryna.mishchenko@ri.se
nishat.mowla@ri.se