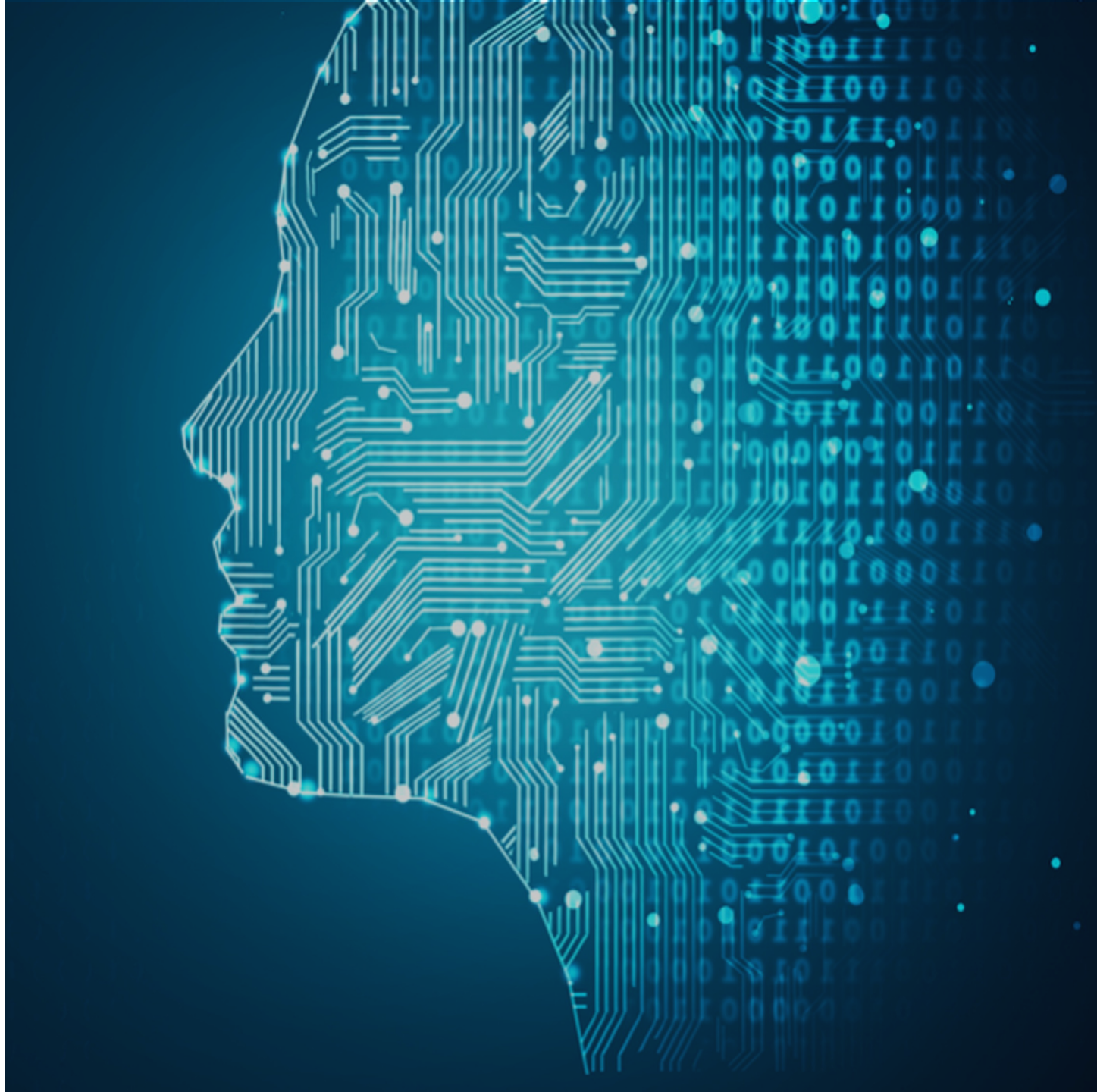
Horizontal vs. Sectoral Standards (Profile Concept)

PricewaterhouseCoopers Germany
October 2023



Risks of Artificial Intelligence can be observed by many sectors where AI systems are being used in

AI in Medicine

Impact on human body

Perpetuation health biases

Use of sensitive biometric data

AI in Finance

Individual financial consequences

General financial instability

Perpetuation existing inequalities

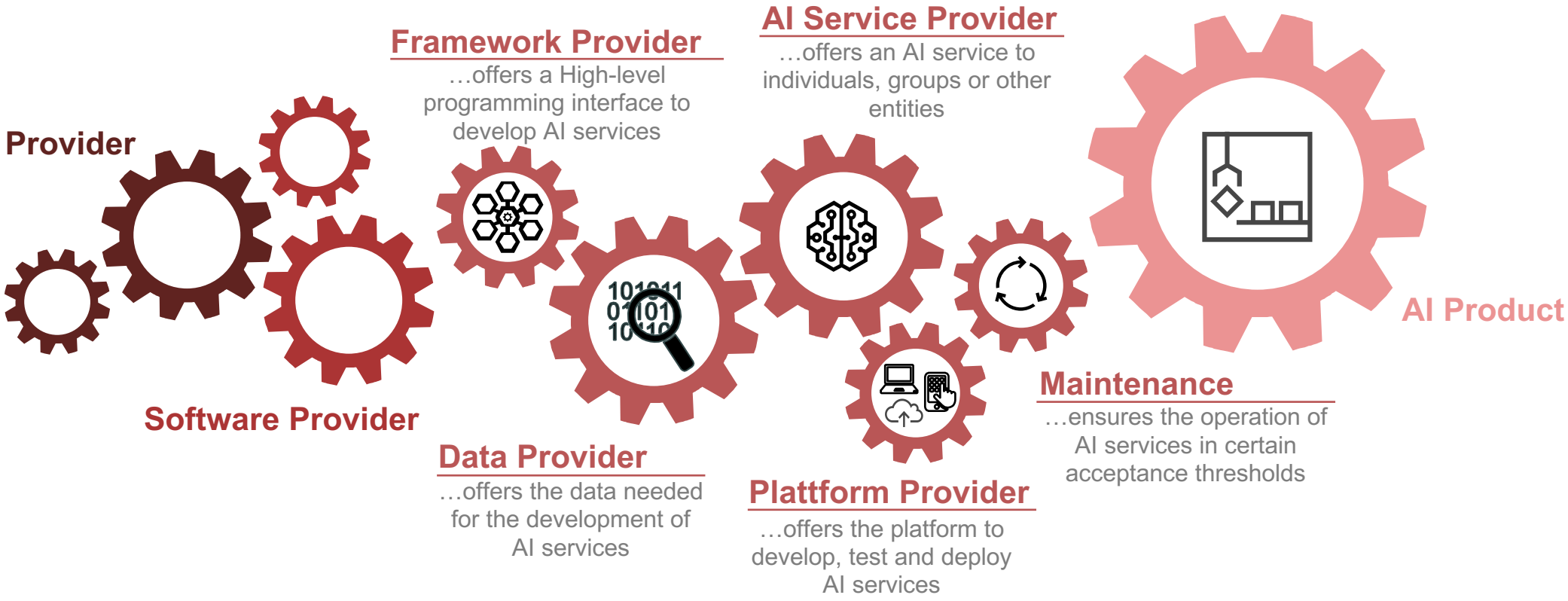
AI in Automotive

Uncontrolled environment

User without AI knowledge

Ethical real-time decisions

The modularity of AI systems lead to an increased complexity in proving compliance for trustworthy AI



Sectoral standards should be created to function as a direct answer to sector specific risks to artificial intelligence

Sectoral examples

Automotive	Medical
ISO/AWI 8800 Risk factors influencing AI underperformance	**ISO 14971:2019** **Medical Devices** Standard on risk management process
ISO/AWI TS 5083 Verification and validation of the safety of AI based on principles and goals	**ISO 13485:2016** **Medical Devices:** Standard on quality management systems
ISO/IEC DTR 5469 Specific requirements for functional safety of AI	**IEC 62304:2006** **Medical device software:** Standard for life cycle requirements

Currently no AI specific sectoral standards exist

Sectoral AI standards would help to:

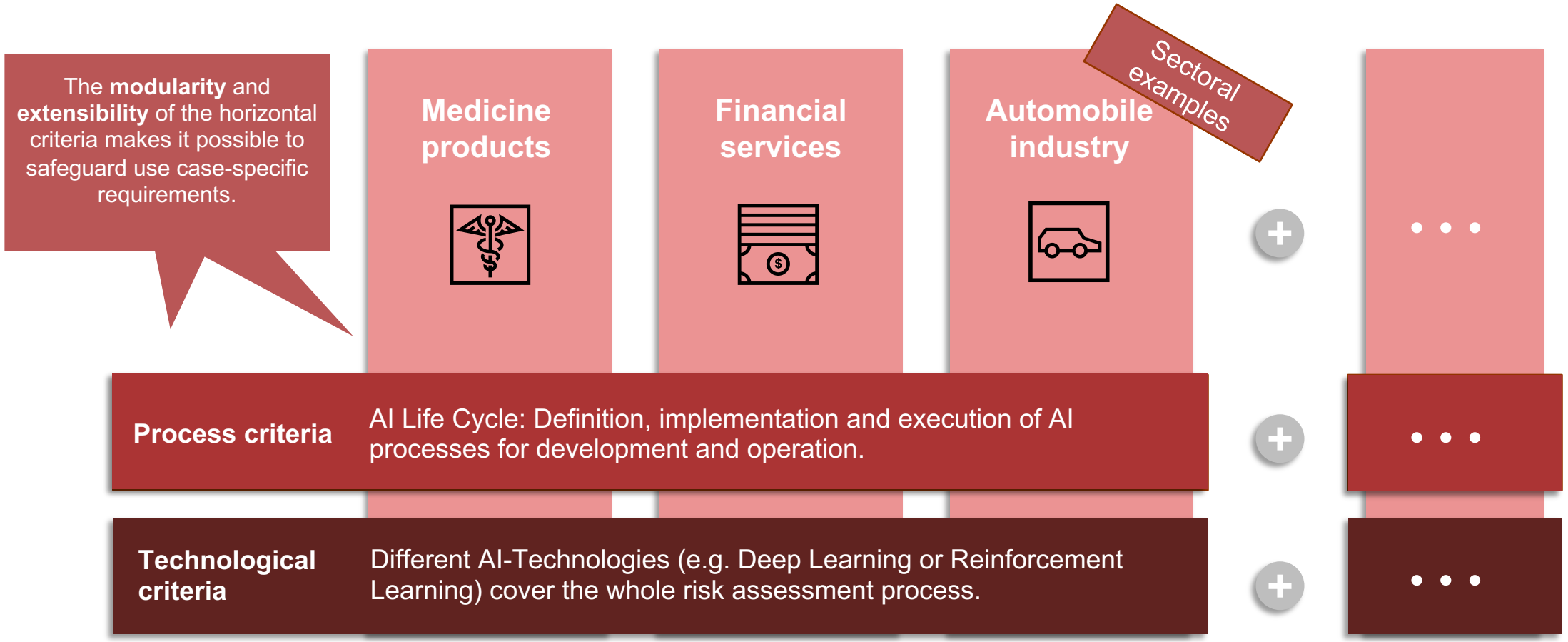
Address **the specific environment of AI applications**

Enable industries to understand how **AI best practices** look like.

Gear towards and optimize for **specific needs of the sector**

Address and **mitigate sector specific AI risks**

Artificial Intelligence should be regarded as a cross-sectoral technology needing both horizontal and sectoral criteria



The EU AI Act has laid out the foundation for a horizontal standard for trustworthy AI

EU AI Act aims to develop a uniform framework

Trustworthiness has to be ensured over the **entire AI value** chain.

It proposes **horizontal AI standards** with criteria **agnostic to all sectors.**

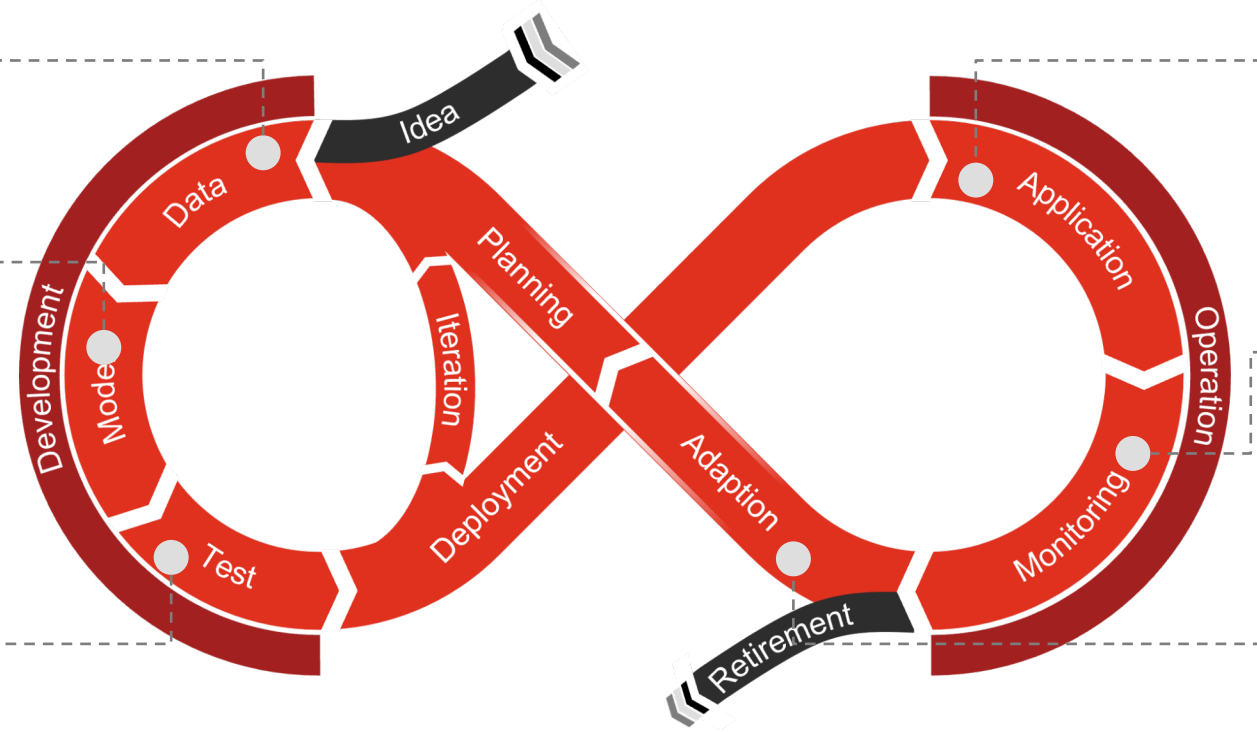
GRC Layer	Lifecycle Layer		Conformity Layer
Risk Management & Impact Assessment	Data Management & Data Quality	Transparency & User Information (including Explainability)	Technical Documentation & Quality management system
Resource Management, Roles & Rights	Design, Development & Testing Procedures	Logging, Record Keeping Traceability	Conformity Assessment, (CE) Marking
Responsibilities + Accountability	Accuracy, Robustness & Cybersecurity	Monitoring & (Human) Oversight	Database Registration & Reporting to Authorities

Compliance with the relevant trustworthy AI criteria should be ensured along the whole AI lifecycle

Data
- AI is only as good as the data used
- Data needs to be checked for bias
- Use of personal data by AI must be specifically assessed

Model
- Assessment of attacks on the AI service / model.
- Robustness of the model must be ensured
- Definition of software best practices

Test
- Select test metrics according to needs
- Identify and mitigate bias of the model



Application
- Model must be protected against data and model theft
- Clear and easy to understand manuals on the use of the AI

Monitoring
- Model metrics must be monitored and limits defined
- Model anomalies must be identified

Adaptation
- Identification of new requirements and vulnerabilities
- Updates for continued and secure operation must be provided

A uniform standardization approach should address all relevant areas to enable trustworthy AI

Evaluation types

- Conformity assessment
- Self-assessment
- Attestation
- Evaluation
- Certification

Criteria for Trustworthy AI

- Robustness
- Compliance
- Explainability
- Autonomy and control
- Bias and fairness
- Safety and Security

AI Ecosystem

- Embedded AI
- AI as a service
- AI supply chain
- Cloud-Edge-Systems
- Life Cycle Phases

Profile concepts: Aiming to give guidelines on how to approach assessments for sectoral und horizontal criteria.

Example: The idea of profile concepts combines both horizontal and sectoral aspects related to use cases

Profile concepts

Combination of **horizontal and sectoral principles** for technological and sectoral risks.

Should be created for AI products, software and services.

Give **detailed guidelines through case archetypes** on how ensure AI conformity.

Are developed and provided based on **the target of evaluation.**

Case study: Object classification of traffic participants

Technology object classification:

Aims to localizing and classifying the object

Need for low latency in processing

Object detection

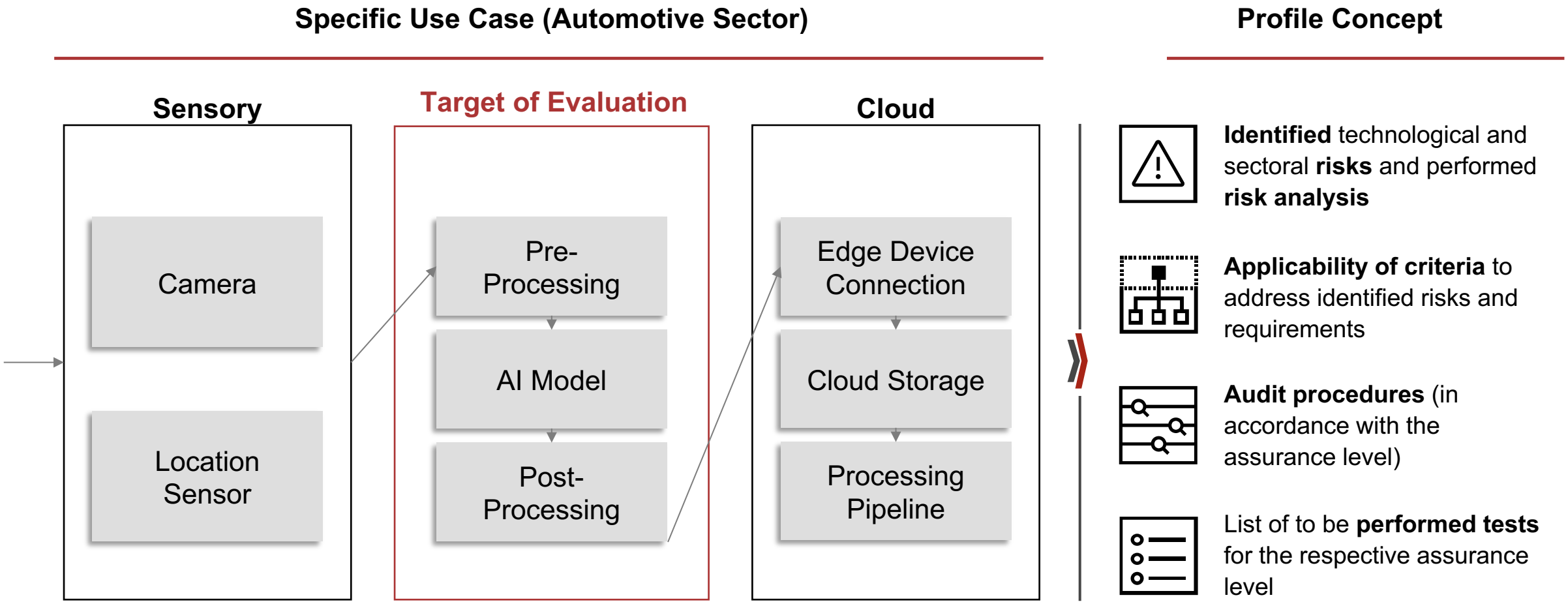
Automotive sector:

Car malfunction

Very high safety requirements

User without AI knowledge

Example: In a profile, the characteristics relevant for a specific use case are defined to ensure compliance.



In order to address all relevant aspects of AI, both horizontal and sectoral criteria are required

Mitigation of AI risks

General as well as sector specfic risks need to be mitigated

AI specific standards needed

IT general standards are not sufficient to adress AI specific challenges

Holistic approach required

Current approaches do not fully address the need for horizontal and sectoral standardizationon

Guidance needed

Guidelines on how to approach assessments for sectoral und horizontal criteria

Horizontal vs. Sectoral Standards (Profile Concept)

Q&A

Thank you.

Hendrik Reese
Partner

+49 151 704 23 201

hendrik.reese@pwc.com

pwc.de