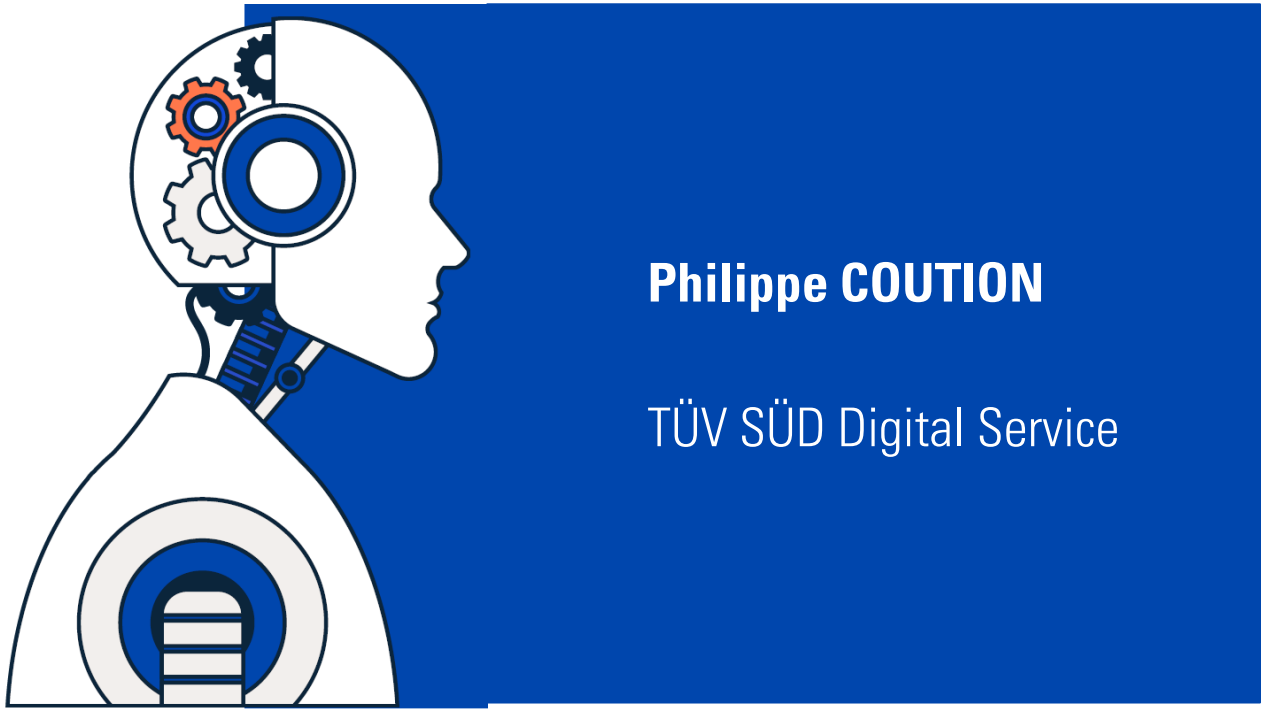
Security Conference

Key Steps towards Trustworthy AI with AI Quality Assurance

Presented by: Philippe COUTION

16.10.2023

With you today



Philippe COUTION

TÜV SÜD Digital Service

AI Quality Senior Consultant

- Engineer, 20 Years experience in Industry and Digital Transformation
- Start-up in Computer Vision

TÜV SÜD at a glance

By choosing TÜV SÜD, a dedicated team of global experts is committed to help you manage risks and access global markets through a comprehensive portfolio of technical solutions.

- Our logo is universally respected as an independent and impartial symbol of safety, security and sustainability.
- TÜV SÜD certification marks and certificates represent third-party endorsement by a globally renowned organisation, while our personnel certificates provide our customers with greater market opportunities.

1-STOP SOLUTIONS PROVIDER

150 + YEARS OF SAFETY, SECURITY & SUSTAINABILITY

1,000 + LOCATIONS WORLDWIDE

25,000+ EMPLOYEES

€2.7 BILLION IN ANNUAL REVENUE

100% INDEPENDENT & IMPARTIAL

55,000 SYSTEM CERTIFICATES

500,000 PRODUCT CERTIFICATES

50,000 PERSONNEL CERTIFICATES

Adding value across the business cycle

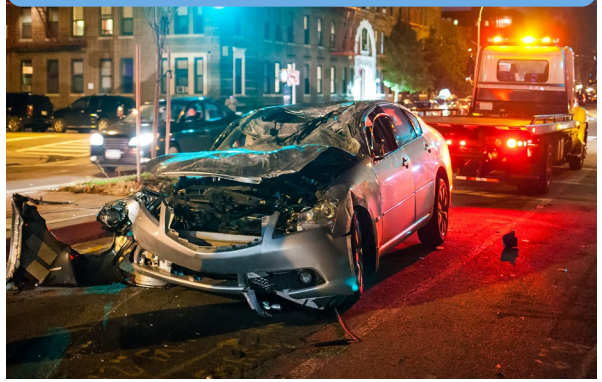
- TESTING AND PRODUCT CERTIFICATION
- AUDITING AND SYSTEM CERTIFICATION
- TRAINING
- INSPECTION
- ADVISORY

Navigating AI Regulations for Business Success

1	**Trustworthy AI: AI risks and pitfalls**
2	**Quality for trustworthy AI**
3	**Industrial Use Case**

AI Risks

Safety



Legal

Performance

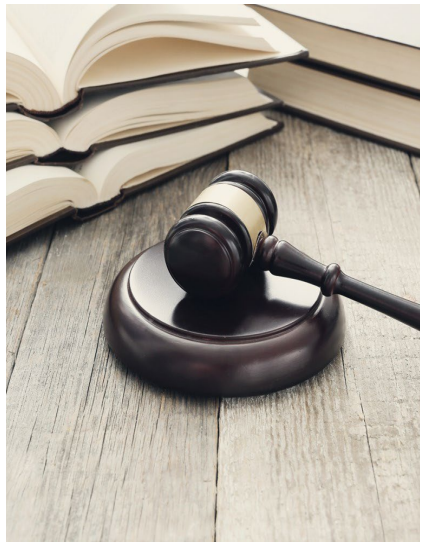
Security

Ethics

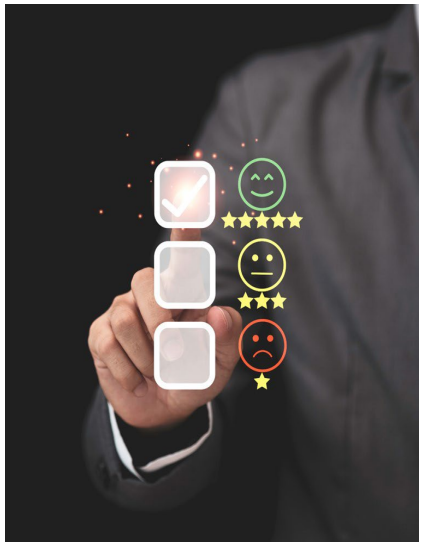
Sustainability

Consequences of AI incidents

Scale of potential consequences



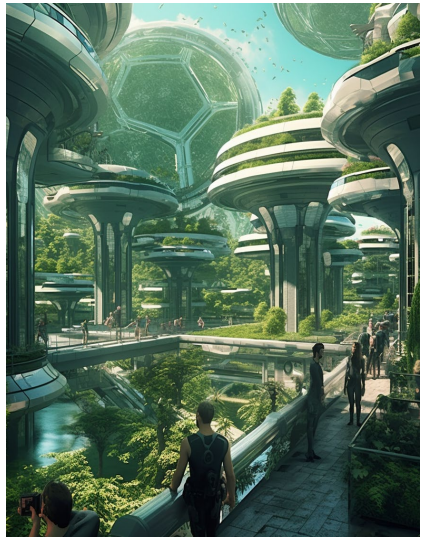
Legal



Reputational

Monetary

Society



Environment

Why AI Quality matters

Control and manage risks

- Assure regulatory compliance
- Avoid failures

Build trust with customers

- Provide clarity and assurance
- Meet the expectations of customers

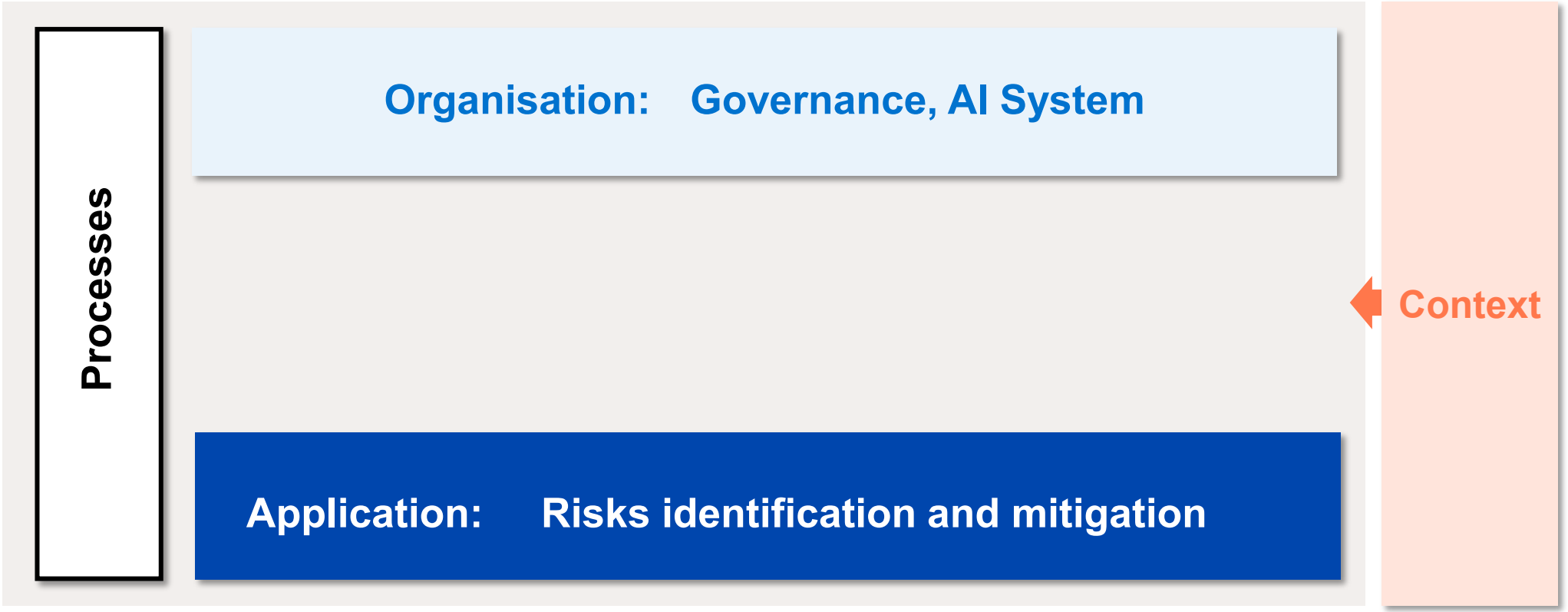
Scale the adoption of AI

- Fully utilize the potential of AI solutions

We need an **AI Quality Model**

AI Quality Model

A comprehensive AI Quality approach needs to address Organization, System, Process, and Application in its context



But how to precisely define those requirements?

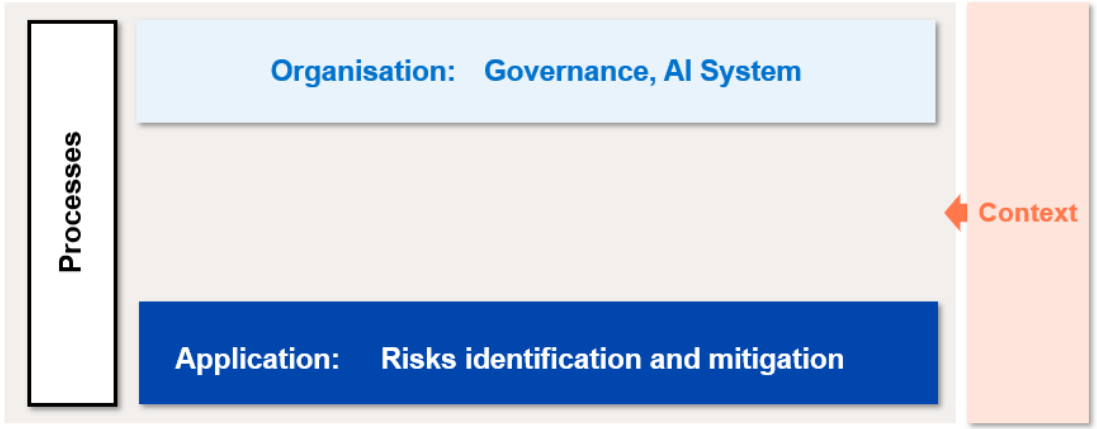
AI Quality is key for compliance and scaling

We compiled all relevant information to define the AI quality model

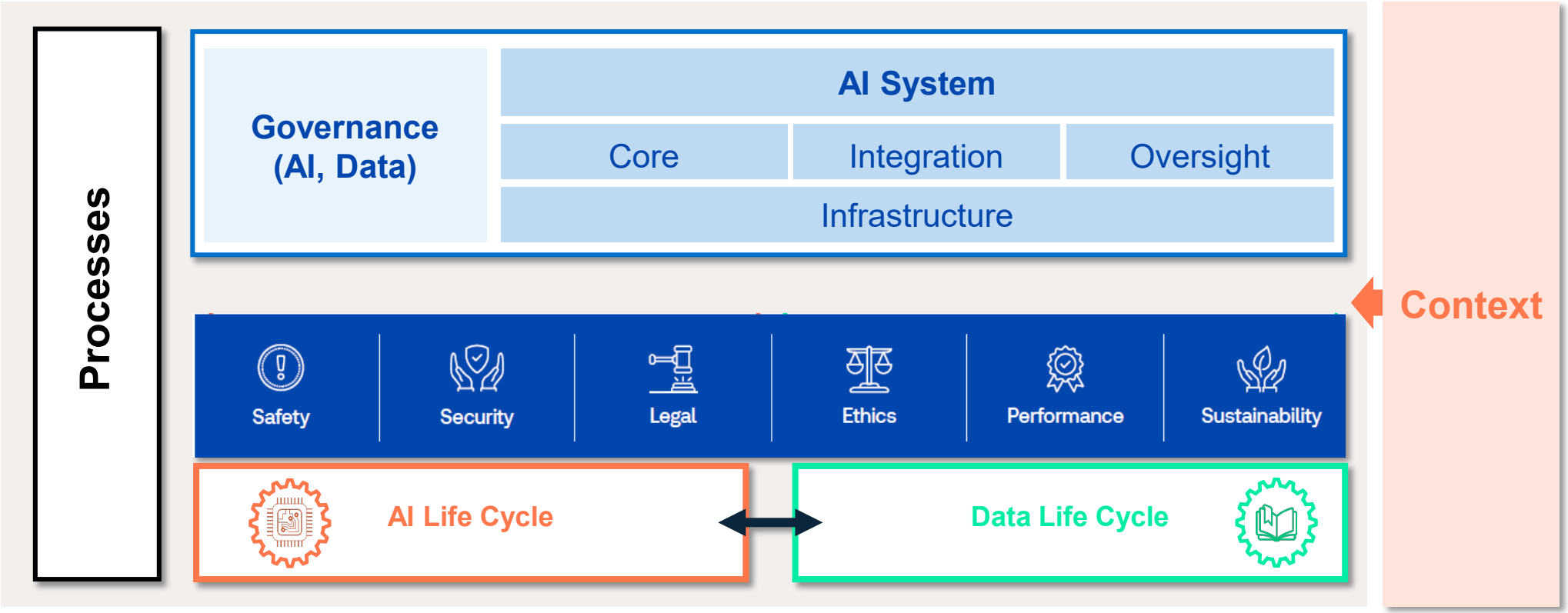
Standards

EU AI Act
National regulations

Best Practice
(QMS, Risk Management,…)



AI Quality model in details

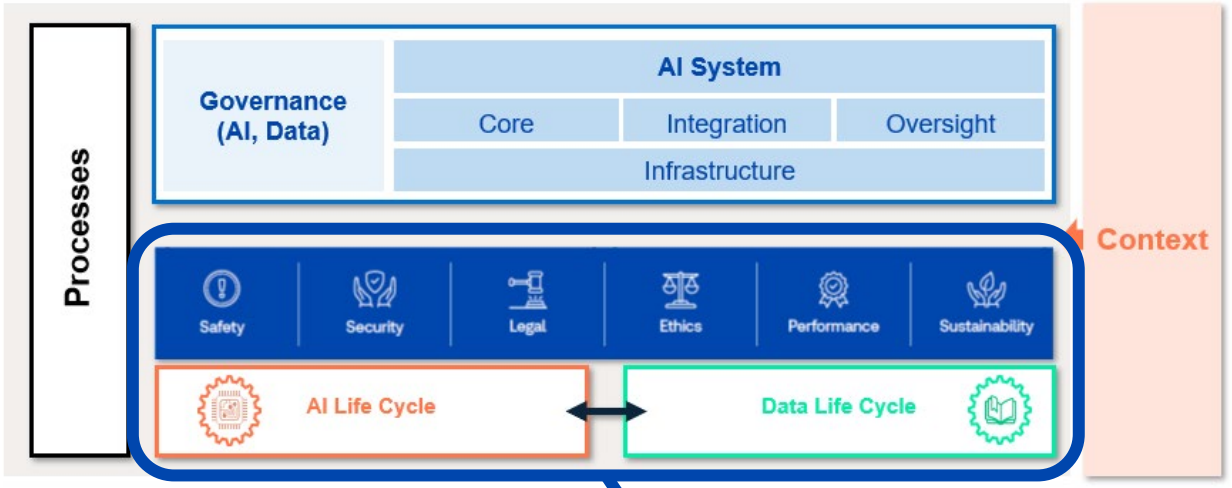


Navigating AI Regulations for Business Success

1	Trustworthy AI: AI risks and pitfalls
2	Quality for trustworthy AI
3	Industrial Use Case

Quality for Trustworthy AI

Let's focus on a specific part of the quality model: the risk profiling



Risk profiling of an AI application

AI Risks

Safety

Has the AI system the potential to directly or indirectly harm anyone/anything?

- Predictability
- Testability
- Traceability
- ...

Legal

Is the AI system and usage compliant with regulations?

- Obligations
- Governance
- Auditability
- ...

Performance

Is the system fit for use? Are unsubstantiated claims made?

- Suitability
- Efficiency
- Reliability
- ...

Security

Does the AI system increase cybersecurity risks?

- Confidentiality
- Authenticity
- Recoverability
- ...

Ethics

Does it violate accepted ethical principles for impacted stakeholders?

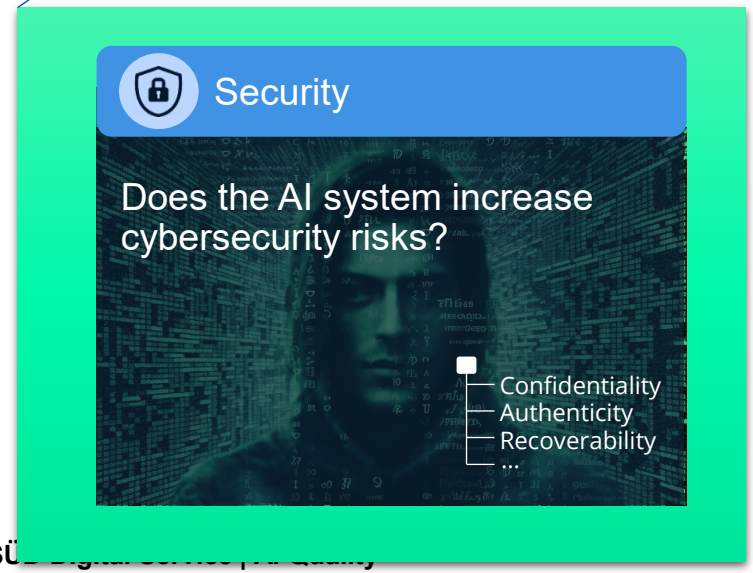
- Transparency
- Non-discrimination
- Accountability
- ...

Sustainability

Has the development or operation been conducted with environmental considerations?

- Resource footprint
- Proportionality
- Reusability
- ...

AI Risks – 9 quality characteristics for Security



Accountability

Property that an individual is answerable to proper authority for a given object, information, event, or action.

Authenticity

Property that an entity is what it claims to be.

Availability

Property of being accessible and usable on demand by an authorized entity.

Confidentiality

Property that information is not made available or disclosed to unauthorized individuals, entities, or processes.

Integrity

Property of accuracy and completeness.

Non-repudiation

Inability to deny a claim.

Recoverability

Ability to recreate a prior state.

Reliability

Property of consistent intended behavior and results.

Repeatability

Precision under repeatability conditions, i.e., Same method on identical items in the same context within short intervals of time.

Risk Profiling

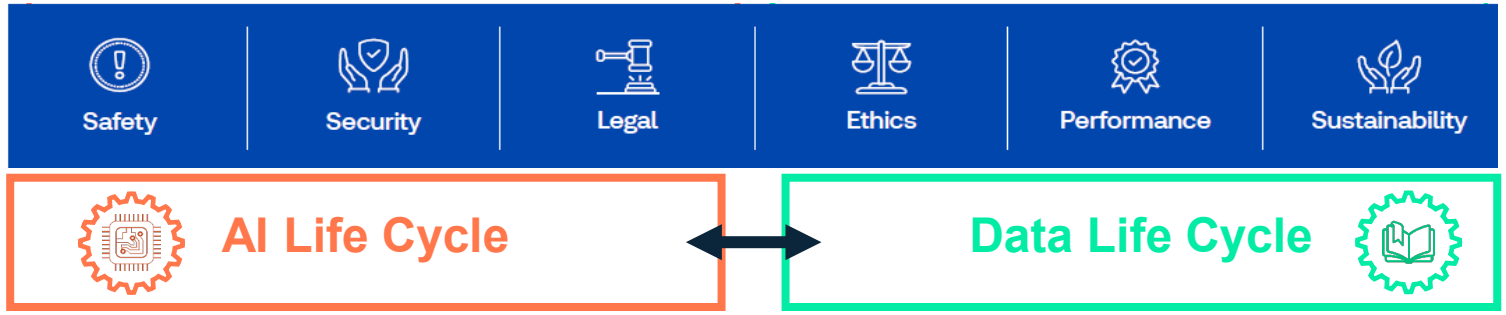
- Safety
  - ...
- Security
  - Confidentiality
  - Authenticity
  - Recoverability
  - ...
- Legal
  - ...
- Ethics
  - ...
- Performance
  - ...
- Sustainability
  - ...

Likelihood * **Severity** * **Scope**

Likelihood	Severity
Often	Critical
Likely	Major/Severe
Moderate	Serious
Unlikely	Moderate/Small
Rare	None/Minor

System	User	Business
Infrastructure / Environment	Indirect User	External
System Components	Secondary User	Company
Component	Primary User	Project Team

Risk profiling reveals risk characteristics of AI system



Identified residual risks for security (example)

Navigating AI Regulations for Business Success

1	Trustworthy AI: AI risks and pitfalls
2	Quality for trustworthy AI
3	Industrial Use Case

Use Cases show benefits of AI Quality Assessments

AI components for AVs

Start-up in the Automotive and Logistics sector developed AI-based software product to drive automated heavy-duty vehicles

Challenge: Evaluate the preparedness of the organization to ensure the quality of their AI system

Outcome: Demonstrated dedication to quality, enhancing credibility and ethical standards via compliant, trustworthy AI development.

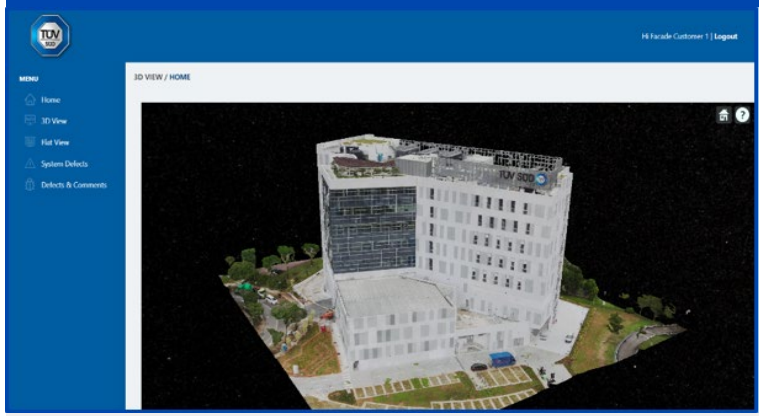
AI for brain health diagnosis

Healthcare start-up developed an AI-based product to diagnose mental health

Challenge: Limited understanding and capabilities of controlling AI and meeting ethical requirements

Outcome: Company demonstrates quality commitment, enhances credibility, and upholds ethics via trustworthy, compliant AI development.

AI for automated defect detection



TUV SUD developed AI-based defect detection of building facades

Challenge: Limited understanding of degree of compliance with data privacy and robustness

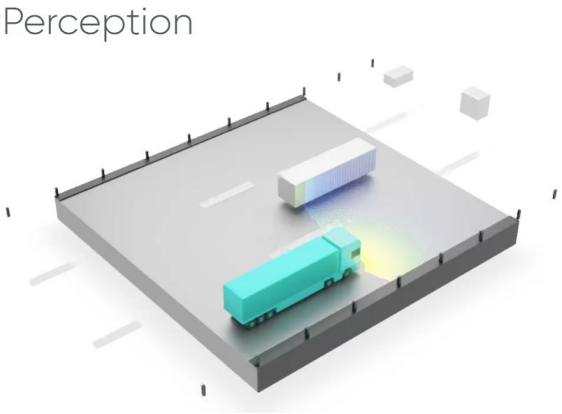
Outcome: Company demonstrates commitment to AI quality bolstering credibility and upholding privacy standards and high accuracy.

Use Case - Assessing AI System of AV Software

driveblocks

The Company

driveblocks is a software and AI technology provider for autonomous driving functions for heavy-duty vehicles – in container terminals, mining, agriculture, and hub-to-hub logistics



The Product

driveblocks offers a Mapless Autonomy Perception platform, overcoming the challenges associated with high-definition maps by replacing them with a sophisticated feature detection and sensor-fusion approach.

The technology combines recent advances in AI, such as transformer neural networks with a geometrically interpretable sensor-fusion, to achieve explainability and enable certification.

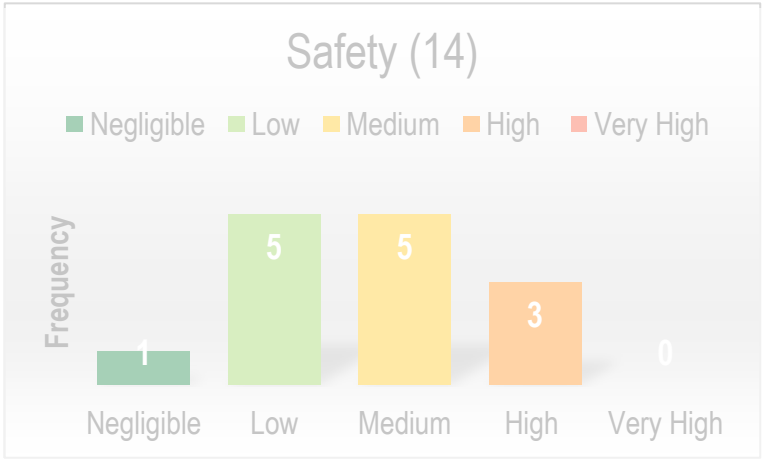
The software provides a consistent environment model, including drivable space detection, lane structure detection, and object detection.

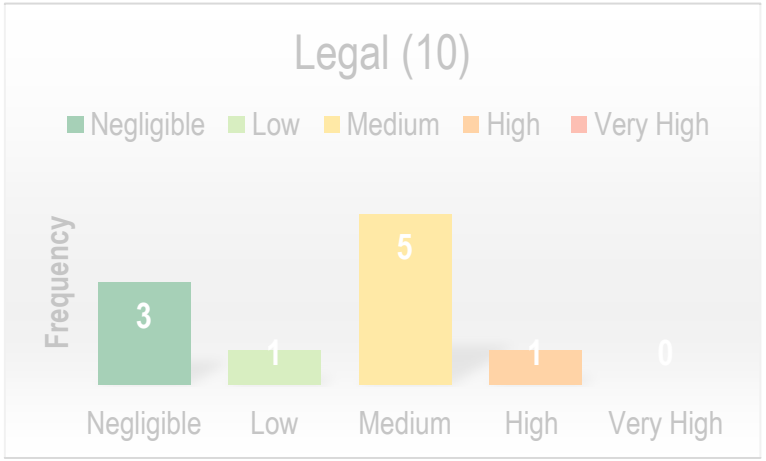
The Challenge

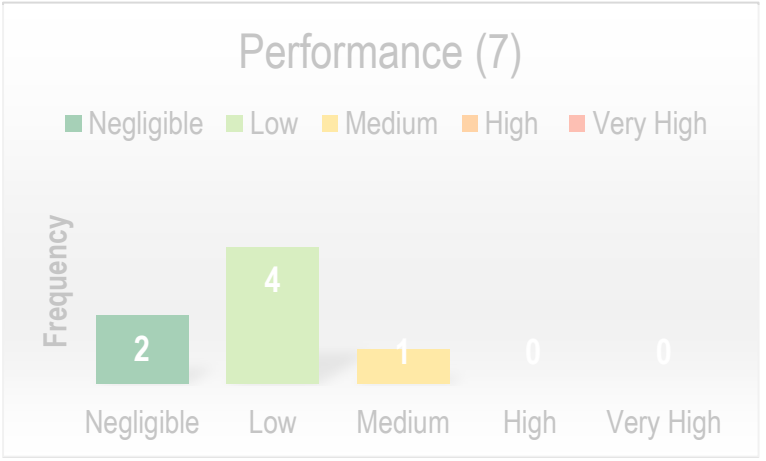
Identify the risks and their mitigations

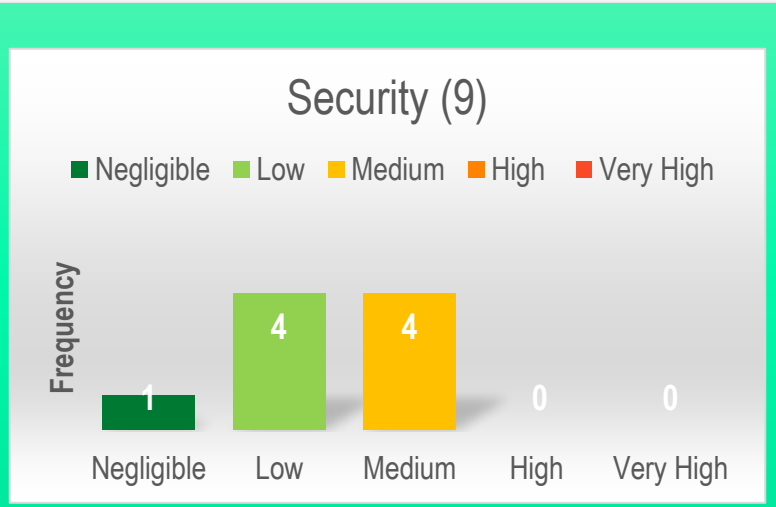
Risk profiling over all quality pillars

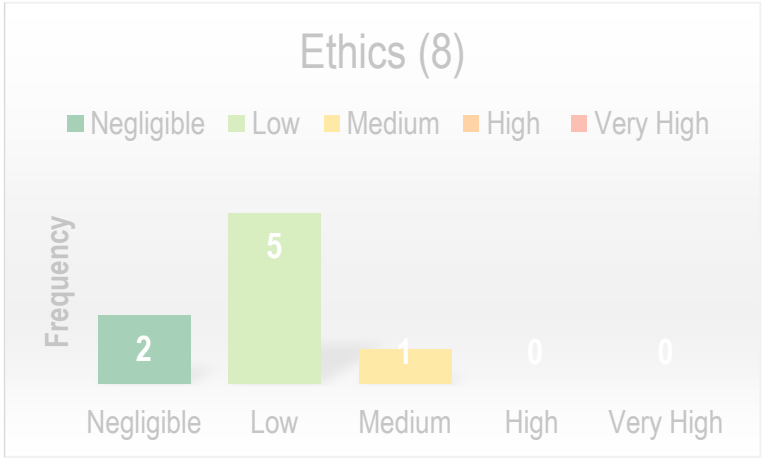
Assessment with guiding questions, ranking potential risk frequency, severity, and scope of impact.

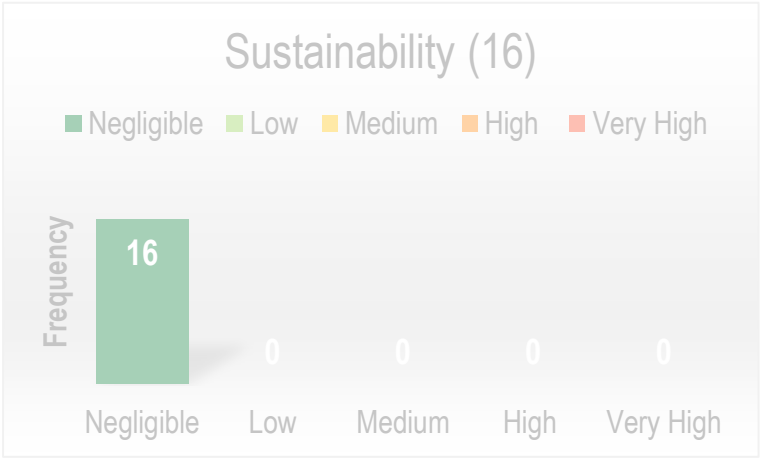




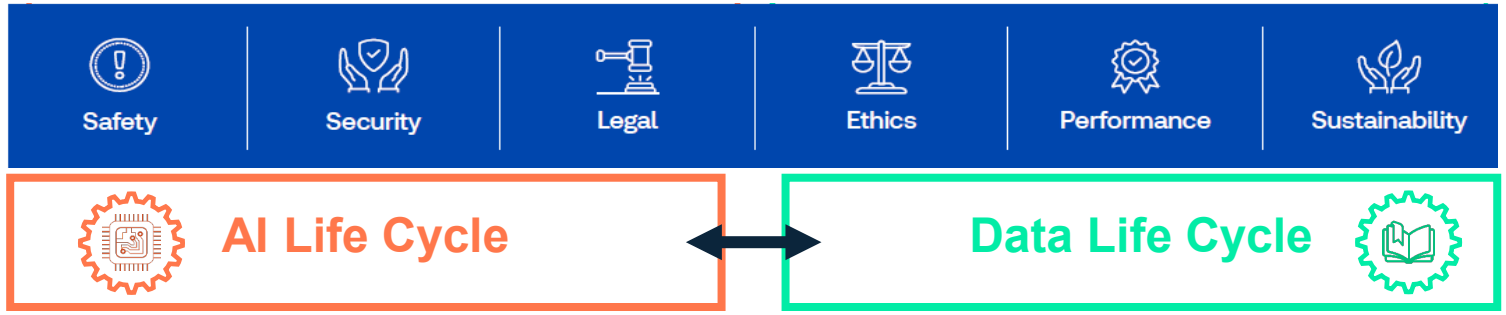


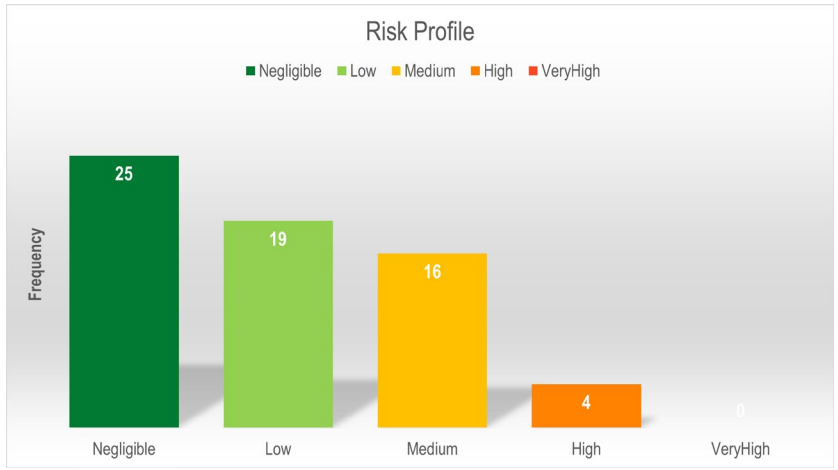






Risk profiling reveals risk characteristics of AI system

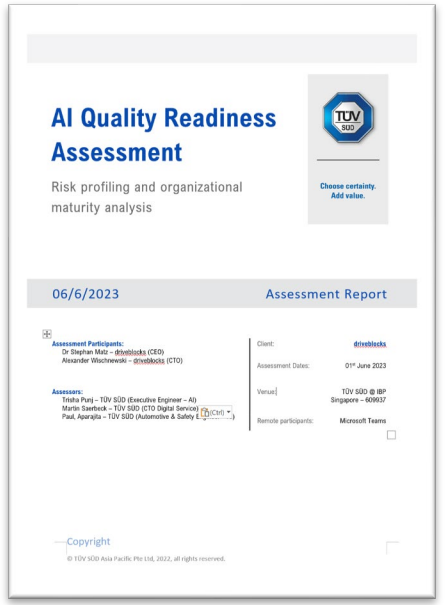




Identified residual risks

Key Findings

- Existing AI risk mitigation strategies
- Risk profiling
  - High risk: 4 priorities to focus on
  - Moderate risks: security, ethics, and performance



Thank You

Links:
- Infographic
- Question list
- Sharepoint
- Global Website