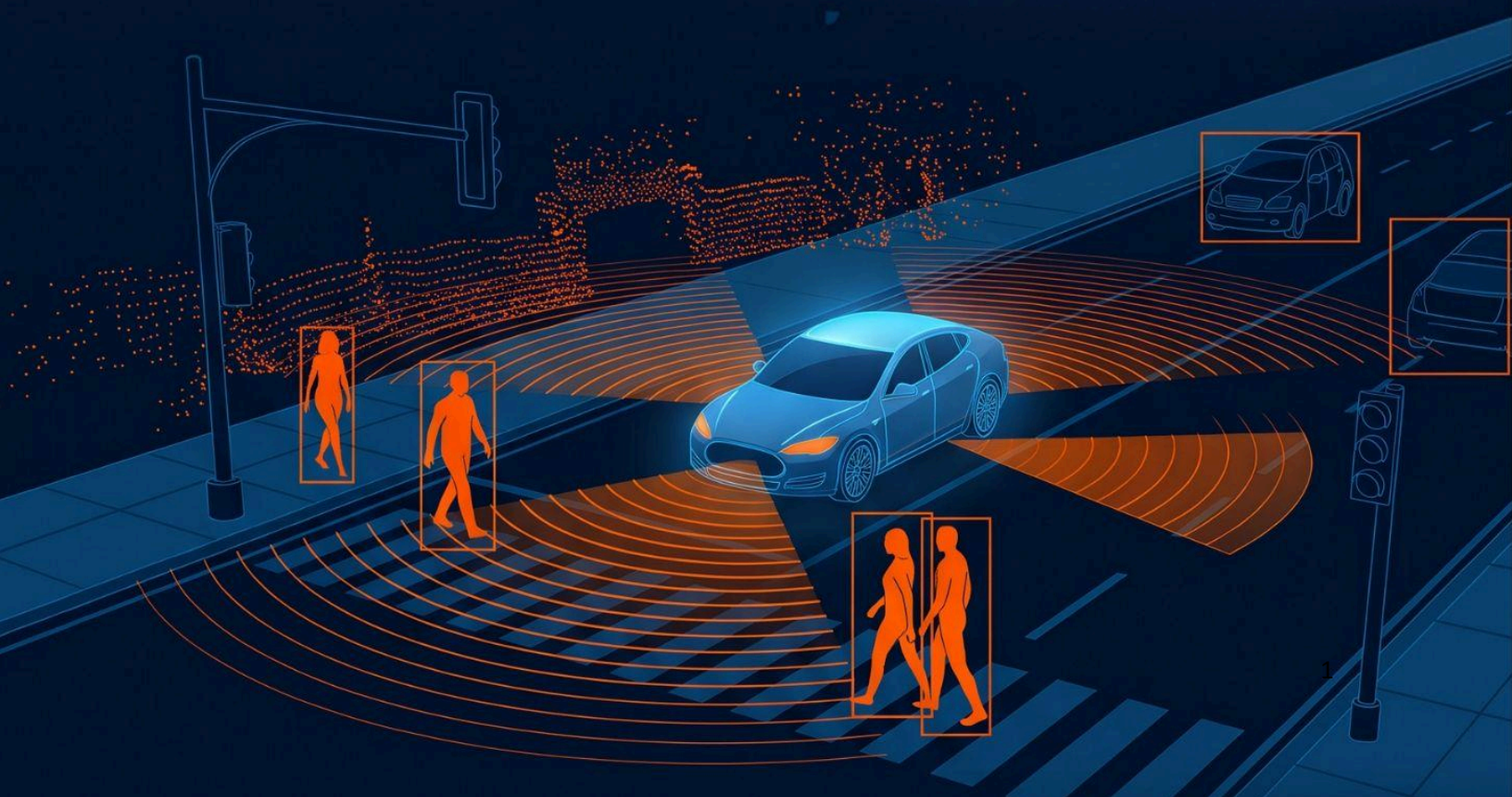
Synergy in AI Safety: Leveraging QuantPi and SGS-TÜV Saar for ISO/PAS 8800 Compliance



Abstract

Machine learning is increasingly used in safety-related functions of road vehicles, such as the detection of people, objects or lane markings. ISO/PAS 8800 sets out how the safety of such components can be argued. It calls for AI safety requirements to be expressed as probabilities of AI errors and supported by evidence with appropriate statistical confidence. Applying the standard in practice raises two challenges: interpreting requirements consistently for a given use case and producing reliable test evidence across the entire operational design domain. This whitepaper describes how SGS-TÜV Saar and QuantPi jointly offer a solution for these challenges. SGS-TÜV Saar provides normative interpretation, process and product assessment, and certification. QuantPi provides AI component testing that yields reliable technical evidence with quantified measurement uncertainty. Combined in one integrated quality gate, the two partners enable suppliers and OEMs to reach certification of their AI systems faster.

Authors

Jürgen Helmschmidt[1], Johannes Schneider[1], Ramya Vinod[1], Banupriya Valluvan[1], Sai Nikhil Menon[2], Lukas Bieringer[2]

[1]SGS-TÜV Saar, Munich, Germany

[2]QuantPi, Saarbrücken, Germany

Acknowledgements

The authors would like to thank Marcus Rau, Sergio Frauca Roca and Andrei Tolstykh for their input and valuable feedback on this work.

Date of Publication

The content of this document is based on the technological and legal framework conditions applicable as of November 2026.

Contact

SGS-TÜV Saar:

Jürgen Helmschmidt
Product Manager AI and Software
Mail: juergen.helmschmidt@sgs.com

QuantPi:

Lukas Bieringer
Head of Policy & Grants
Mail: lukas.bieringer@quantpi.com

1. The Normative Landscape for Safety-related AI

Machine learning is increasingly used in safety-related functions of road vehicles and mobile robots. Typical examples are perception functions that detect people, objects or lane markings from camera, radar or lidar data. If such a function fails, people can be harmed.

Established functional safety standards cover these functions only in part. ISO 26262 addresses systematic faults and random hardware failures in systems whose intended behaviour can be fully specified. ISO 21448 (SOTIF) addresses hazards caused by functional insufficiencies and triggering conditions. A machine-learned function, however, is not specified but trained on data. Its behaviour cannot be fully derived from its design, and it can fail on valid inputs that were poorly represented during training. Its correctness can therefore only be established empirically, as an error rate over the operational design domain (ODD), and with statistical confidence.

ISO/PAS 8800 addresses this gap for road vehicles. It builds on ISO 26262 and ISO 21448 and defines a safety lifecycle for AI elements. This lifecycle spans the derivation of AI safety requirements, data management, verification and validation, safety analysis and measures during operation. All of these activities are brought together in an assurance argument.

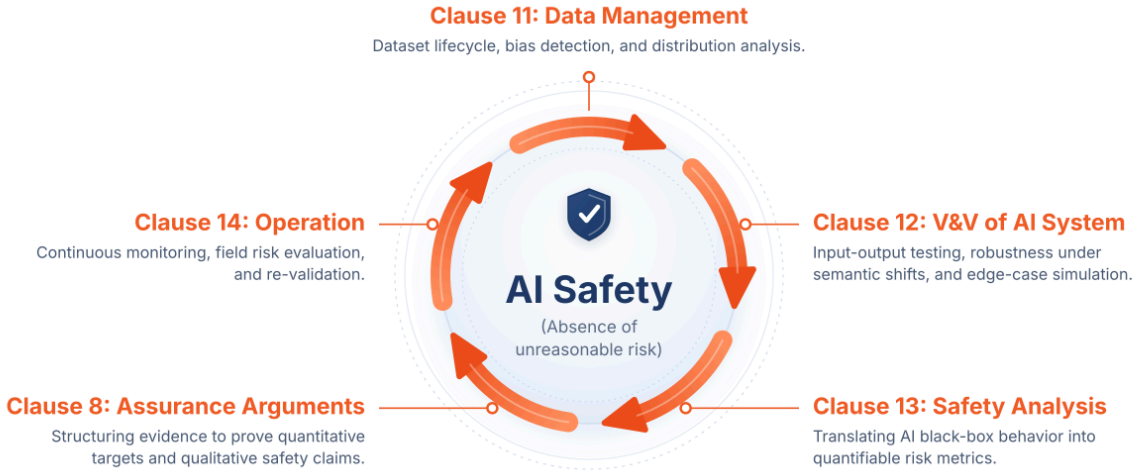


Figure 1: Selected clauses of ISO/PAS 8800 and their contribution to AI safety [1].

Applying ISO/PAS 8800 in practice raises two kinds of challenges. The first is testing. AI safety requirements are expressed as probabilities of AI errors, and the standard asks for evidence that these are met with appropriate statistical confidence. Producing such evidence across an entire ODD, which can rarely be tested exhaustively, requires statistical methods and an explicit account of measurement uncertainty.

The second challenge is interpretation. As a Publicly Available Specification, ISO/PAS 8800 leaves considerable room for how its requirements apply to a given use case. Deployers must decide which requirements are relevant and which metrics are adequate proxies for their safety requirements. They must also justify their acceptance thresholds and their choice of confidence level. Good test results do not settle whether the assurance argument as a whole is complete.

Data management processes, the development lifecycle and the tools used to generate evidence must themselves meet the requirements outlined in the standard. And after deployment, model updates, sensor changes or shifts in the operating environment can invalidate parts of an argument that was once sound. Judging this consistently across products, and in line with how the standards themselves continue to evolve, requires independent expertise in functional safety and in standardisation.

Self-reported AI component performance is therefore not sufficient. Customers, users and market surveillance authorities must be able to rely on both the evidence and the safety claims built on it. This requires reliable technical evidence with quantified measurement uncertainty [2], combined with an independent assessment of whether the test method, the evidence and the argument are adequate for the safety case. Together, these turn internal confidence into external trust: a statistically qualified statement about the probability of hazardous AI errors, assessed against the requirements of ISO/PAS 8800.

2. QuantPi: The Technical Validation Layer

Safety cases for AI need more than benchmark scores. Generic leaderboards report averages. They hide rare but critical failures, and they say nothing about how a number itself can be trusted as a test result. QuantPi closes this gap. Its testing engine, rooted in research at the CISPA Helmholtz Center for Information Security, turns safety requirements into measurable evidence with quantified measurement uncertainty [3]. Testing AI systems with QuantPi is black-box and model-agnostic: any perception model, from any supplier, is assessed in the same uniform framework. When a model is retrained or replaced, test cases and acceptance thresholds can be reused.

Safety requirements as bounded probabilities

AI safety requirements are probabilistic by nature. A people detection system, for example, is assessed by how often it misses a person in a given time frame. QuantPi expresses each numerical result of a computed test scenario with an explicit confidence interval to help customers meet the requirements of Clause 8.4a) of ISO/PAS 8800. The interval methods it uses remain valid when failures are rare, which is exactly the regime that safety targets live in. Correlated data, such as consecutive video frames, is corrected for, so that confidence intervals are never narrower than evidence requirements allow. Each result is then compared against its target through a predefined pass/fail rule. AI system assessors receive numerical test results plus information on their inherent measurement uncertainty, the number of data points on which the test was executed, and the corresponding verdict derived by application of the decision rule.

ODD-stratified coverage and weak-spot discovery

An average performance figure can meet its target while the AI system under test fails in specific conditions. QuantPi therefore measures performance per operating condition of the ODD: lighting, distance, occlusion, pose, crowding, or environment type. Each condition receives its own test result, confidence interval and verdict. Conditions with too little data are flagged instead of reported. The worst-performing condition is made visible, not averaged away.

In line with Clause 12.5.2.2 of ISO/PAS 8800, this shows both how well the test data covers the ODD and where the AI system is performs weakest.

Robustness under controlled variation, with validity checks

The verification and validation of AI system behaviour depends on suitable test datasets. However, some critical operating conditions might be rare in recorded data or dangerous to stage in reality. In line with Example 5 of Clause 8.7 in ISO/PAS 8800, QuantPi therefore simulates test scenarios with perturbations of the test data. Rule-based perturbers model physical effects such as glare, lens fouling, condensation, noise or blur at defined intensities. AI-based perturbers create semantic variations, such as transforming daytime scenes into night-time scenes or adding occlusions [4]. Performance is measured at each intensity level, which shows where and how fast it degrades. But since synthetic test data is only useful if it is valid, QuantPi can check perturbed data against real-world samples and reports where the simulated evidence holds.

Calibrated measuring instruments

Testing AI systems at scale needs automated annotation via AI-based testing components such as vision-language models (VLMs), and automated annotators make mistakes. In QuantPi's testing framework, AI-based annotators are treated as measuring instruments that introduce additional measurement uncertainty. Therefore, each AI-based annotator is calibrated against expert-labelled data. Its error is then mathematically propagated into every confidence interval it contributes to. This way, final test results remain valid even when test data labels are machine-generated. A small, smartly selected subset of expert-labelled data anchors the entire assessment. This keeps the cost of ground truth low without weakening the evidence.

QuantPi supplies the quantitative evidence for safety claims in context of ISO/PAS 8800. It does not certify safety. That judgement requires the independent interpretation of an accredited certification body.

3. SGS-TÜV Saar: Expert Assessment & Certification

In regulated industries, independent assessment is the final, non-negotiable gate for establishing trustworthy AI. As an accredited certification body, SGS-TÜV Saar provides the interdisciplinary expertise required to bridge the gap between technical output and normative compliance. This third-party validation is essential for mitigating liability risks and providing reliable assurance arguments to market surveillance authorities.

SGS-TÜV Saar utilizes an extensive portfolio of standards beyond the traditional safety standard ISO 26262 [5] to address the specific complexities of AI and robotics, including ISO/PAS 8800, ISO/IEC CD TS 22440 [6], ISO/IEC TR 5469 [7], and based on ISO 26262 and IEC 61508 [8]. Their services are designed to support the entire development lifecycle:

- Process Assessments: Process assessments ensure AI processes and general processes for functional safety match the relevant standards enabling a safe development.
- Gap analysis: Gap analysis gives the customer an early insight whether its processes and the way of the product development match the needs of functional safety standards for AI and in general.
- Product Assessments: SGS-TÜV Saar proves and certifies compliance with AI and functional safety standards within a product development.
- Training and Consulting: SGS-TÜV Saar offers training and consulting on functional safety and AI safety standards such as ISO 26262, IEC 61508, ISO 21448 (SOTIF) [9], ISO/SAE 21434 [10], ISO/IEC CD TS 22440, ISO/IEC TR 5469 and ISO/PAS 8800. This builds the knowledge development teams need to apply the standards correctly from the start and to prepare the evidence required for assessment and certification.

The SGS-TÜV Saar certification acts as a critical market catalyst. It transforms an engineering project into an assessed product, providing the legal and technical assurance necessary for market entry in highly regulated sectors. This path to certification is optimized when the technical evidence generation is pre-aligned with the assessor's rigorous methodology.

It is paramount SGS-TÜV Saar addresses functional safety and this is done typically on overall system or item level. In the automotive industry the item level is represented by the vehicle. In case of robotics the item is the humanoid or industrial robot. In case of automation it is the machinery. SGS-TÜV Saar always guarantees by the assessment safety in terms of absence of harm of any user imposed to an acceptable level. This is now extended from classical functional safety to AI functional safety. Hence it is the correct and failure free processing of the inputs coming from sensors and a failure free inference on the actuators, no matter whether they are brakes, the drive chain, a robot arm, etc. Ultimately, SGS-TÜV Saar enables Physical AI.

Additionally SGS-TÜV Saar offers services for cybersecurity of AI. A system must be protected against tampering through data or model poisoning. SGS-TÜV Saar also assesses the cybersecurity environment to minimize this risk. Together with QuantPi, SGS-TÜV Saar enables a statistical qualification of the AI system’s failure probability. QuantPi’s testing platform quantifies performance with confidence intervals across relevant conditions and user groups, while SGS-TÜV Saar assesses whether the test data is representative and whether the resulting probability meets the acceptable safety target.

AI safety does not end with the first certification. AI systems are trained on data, and the real world they operate in keeps changing: new traffic situations, new environmental conditions, new user behavior (Clause 14, ISO/PAS 8800). Validation must therefore be a continuous process across the entire lifecycle. Not a one-time event and with respect to the data lifecycle, test and training data require continual updates. ISO/PAS 8800 explicitly requires measures during operation, including field data collection, continuous evaluation of field risks, and re-training, re-validation, re-approval and re-deployment of the AI system. SGS-TÜV Saar supports this by assessing whether manufacturers have effective processes in place to monitor their AI systems in operation and to re-evaluate the safety argument whenever new risks or changes arise.

Every update to an AI system, such as a retrained model, new training data or changed software, can change its behavior and must therefore be validated again before release. This includes checking new field data before it is used for training, and regression testing against established acceptance criteria to ensure that improvements in one area do not introduce new failures elsewhere. Equally important, updates must be delivered securely. A manipulated update could bring a poisoned model or harmful changes of test and training data into thousands of vehicles or robots at once. Updates must therefore be authenticated, protected against data and model tampering and traceable, in line with cybersecurity standards such as ISO/SAE 21434, prEN 18282 and, for vehicles, UNECE R155, UN R185 and R156. SGS-TÜV Saar assesses both sides: that every update has been properly validated for safety and security, and that the update process itself is secure from development to deployment in the field.

4. The Synergy: Integrating Test Tools with Independent Assessment

The partnership between QuantPi's technical test tool and SGS-TÜV Saar's assessment services solves the compliance bottleneck by synchronizing evidence generation with expert interpretation. This integrated approach allows enterprises to drastically reduce complexity of product verification, validation and certification, ensuring that the AI stack remains flexible and maintainable as technology shifts. Ultimately reducing time and cost to certification.

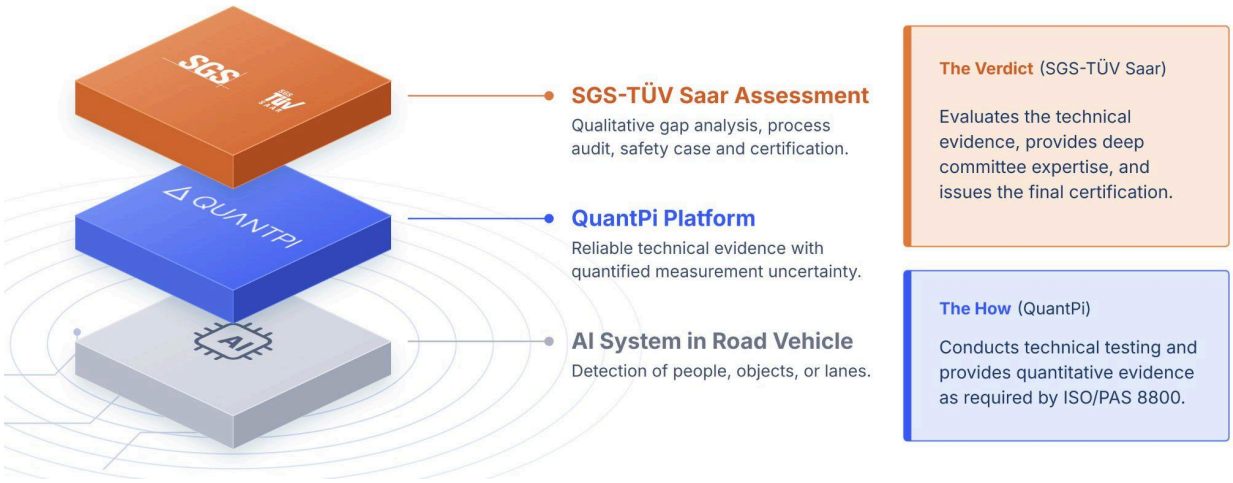


Figure 2: Holistic operating model for AI system deployment in compliance with ISO/PAS 8800. QuantPi software generates the evidence. Human experts from SGS-TÜV Saar build the safety case and certify.

Figure 3 shows how the services of SGS-TÜV Saar and the AI component testing of QuantPi combine into one path to ISO/PAS 8800 certification for AI systems in road vehicles. In the inspection preparation phase, SGS-TÜV Saar identifies the norms and clauses relevant to the use case, provides consulting and training, and conducts a qualitative gap analysis. This defines the test scenarios, including acceptance thresholds, confidence levels, and decision rules before any testing begins. In the subsequent assessment phase, two work streams run in parallel. SGS-TÜV Saar assesses the product and the development process, audits the test

methodology and performs the actual conformity assessment. In the meantime, QuantPi tests the AI system and delivers statistically valid results based on auditable technology and processes. The result is quantified, reliable evidence that is structured from the outset as the assessor requires it. In the final phase, SGS-TÜV Saar validates the safety case and issues the certificate. Because the test scope is agreed early, quantitative testing and procedural assessment run side by side, and the evidence follows a methodology that has been pre-validated.

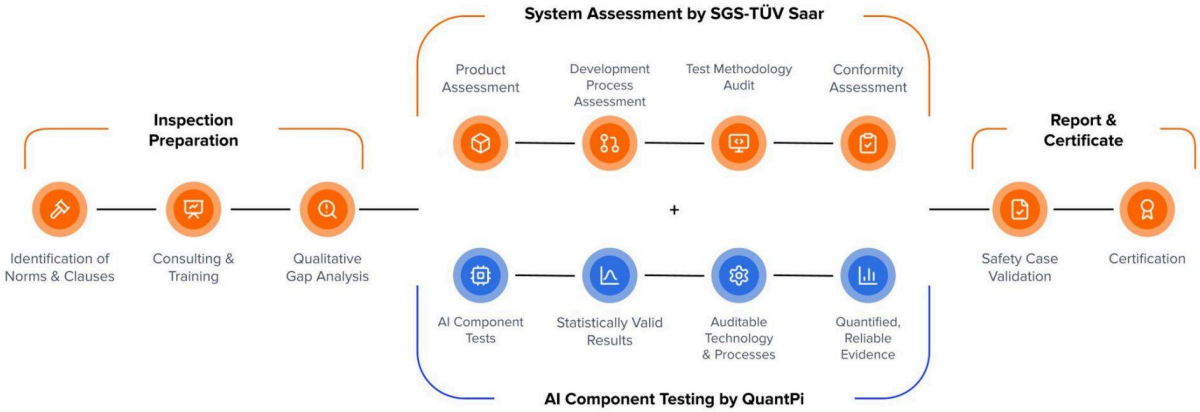


Figure 3: The integrated AI quality gate: SGS-TÜV Saar and QuantPi jointly offer a fast track to ISO/PAS 8800 certification based on reliable technical evidence. Orange: Services provided by SGS-TÜV Saar. Blue: AI component testing conducted by QuantPi.

This integrated model allows enterprises to maintain a consistent, certified safety posture even as they update models or iterate on workflows in the context of road vehicles. By providing both the evidence and the interpretation in a single, pre-engineered quality gate, suppliers and original equipment manufacturers (OEMs) can scale AI without compromising on governance.

5. Conclusion: Business Value through Verified Trust

Navigating the requirements of ISO/PAS 8800 requires a transition from experimental AI to industrial-grade, verified systems. The combination of QuantPi's technical validation and SGS-TÜV Saar's accredited expert assessment supports this transition.

Strategic Takeaways

1. **Efficiency:** Test scope, acceptance thresholds and decision rules are agreed with the assessor before testing begins. This reduces workload for iterations during third-party assessment. When a model is retrained or replaced, test cases can be reused. The result is lower time and cost to certification.

2. **Confidence:** Every test result is reported with a confidence interval, its data basis and a verdict per operating condition. Residual uncertainty is made explicit rather than averaged away.

3. **Market Access:** Independent assessment by an accredited certification body turns internal confidence into external trust. Customers, users and market surveillance authorities can rely on both the evidence and the safety claims built on it.
4. **Guided Development**: The same test scenarios can be run on every model iteration during development. Results per operating condition show exactly where performance falls short and which data closes the gap.

Suppliers and OEMs that develop AI for road vehicles, robots or machinery can engage the joint offering at any stage, from initial gap analysis to product assessment. This approach ensures your AI portfolio is de-risked, compliant, and ready for the complex demands of the physical world.

References

[1] **ISO, ISO/PAS 8800:2024** — Road vehicles — Safety and artificial intelligence. Geneva: International Organization for Standardization, 2024. [Online]. Available: https://www.iso.org/standard/83303.html

[2] **DIN, DIN SPEC 92006** — Artificial Intelligence — Requirements for AI testing tools. Berlin: DIN Media, 2024. [Online]. Available: https://www.dinmedia.de/en/technical-rule/din-spec-92006/398176545

[3] **A. Gautier, L. Bieringer, and S. Klößner**, "On challenges and approaches to test AI systems," Datenschutz und Datensicherheit – DuD, vol. 49, no. 4, pp. 215–219, 2025, doi: 10.1007/s11623-025-2074-7.

[4] **NVIDIA**, "Interactive World Simulation with Cosmos for Physical AI," NVIDIA Research, 2026. [Online]. Available: https://research.nvidia.com/labs/cosmos-lab/cosmos-simulation/technical-report.pdf

[5] **ISO, ISO 26262:2018** — Road vehicles — Functional safety. Geneva: International Organization for Standardization, 2018. [Online]. Available: https://www.iso.org/publication/PUB200262.html

[6] **ISO/IEC, ISO/IEC TS 22440** — Artificial intelligence — Functional safety and AI systems (under development). Geneva: International Organization for Standardization. [Online]. Available: https://www.iso.org/standard/89535.html

[7] **ISO/IEC, ISO/IEC TR 5469:2024** — Artificial intelligence — Functional safety and AI systems. Geneva: International Organization for Standardization, 2024. [Online]. Available: https://www.iso.org/standard/81283.html

[8] **IEC, IEC 61508 & Functional Safety.** Geneva: International Electrotechnical Commission, 2022. [Online]. Available: https://assets.iec.ch/public/acos/IEC%2061508%20&%20Functional%20Safety-2022.pdf

[9] **ISO, ISO 21448:2022** — Road vehicles — Safety of the intended functionality. Geneva: International Organization for Standardization, 2022. [Online]. Available: https://www.iso.org/standard/77490.html

[10] **ISO/SAE, ISO/SAE 21434:2021** — Road vehicles — Cybersecurity engineering. Geneva: International Organization for Standardization, 2021. [Online]. Available: https://www.iso.org/standard/70918.html

About SGS-TÜV Saar

SGS-TÜV Saar is a trusted partner for safety, precision and innovation. We support the development of safety-related systems. Our service portfolio includes training, standards consulting, assessments, audits, and certifications. We work across industries, including automotive, semiconductors, AI, automation, and the process industry, as well as in many other fields of application. We also cover complementary business areas such as cybersecurity, AI, and SOTIF. As an accredited partner, SGS-TÜV Saar provides independent assessments, standard-compliant evidence, and clear results for your decisions. With our support, it reliably guarantees the functional safety of all your products, systems, and processes and remains competitive. With our wide range of services, from inspections and certifications to testing and training, we help you to control risks, optimize performance, and comply with legal requirements and standards. SGS-TÜV Saar GmbH is the result of a joint venture between the SGS Group Germany and TÜV Saarland e.V. SGS-TÜV Saar is represented by employees at thirteen locations.

About QuantPi

QuantPi is a leading European technology provider for quality assurance of artificial intelligence. Its proprietary testing engine allows the uniform assessment of all types of AI (LLMs, computer vision, agentic systems, etc.). This end-to-end platform delivers actionable insights and operationalizes internal AI policies and regulatory frameworks, such as ISO/PAS 8800. QuantPi's testing technology is built side-by-side with leading auditing and certification organizations such as SGS-TÜV Saar. Customers can work towards audits or conformity assessments with streamlined processes and validated test criteria. Certification partners can easily model test scenarios and automate reliable testing with transparent reporting.