
Wie können Qualitätsaspekte von KI-Systemen (und der zugrundeliegenden Daten und Annahmen) konkret nachgewiesen werden?

ANGEWANDTE KI – NUTZEN FÜR EINE BESSERE PATIENTENVERSORGUNG, Clinical Decision Support Symposium 2024,

Dr. Michael Rammensee

Dezember 2024

AIQ

AI QUALITY &
TESTING HUB



Die Agenda für heute

- 1 Was ist KI-Qualität und warum ist sie wichtig?**
- 2 Neue Herausforderungen bei der Entwicklung!**
- 3 Q&A**



digitales.hessen

Einzigartige Partnerschaft zwischen Staat und Berufsverband an der Schnittstelle von Forschung, Entwicklung, Industrie und politischen Rahmenbedingungen für den verantwortungsvollen Einsatz von Künstlicher Intelligenz - ein neutraler Partner

www.aiqualityhub.com

Unsere Angebote



Weiterbildung & Training

- Höhere Qualifikation der Arbeitskräfte
- Ausbildung von Führungskräften
- AI-Qualität in der Praxis



Tech & Tools

- Qualitätssicherung
- Anwendungsentwicklung
- Tests & Tools & Daten

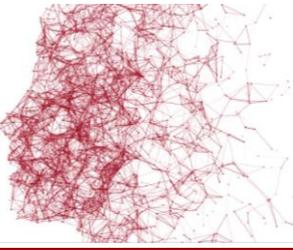


Beratung

- Standards & Guidelines
- Qualitätskriterien
- Audits & Red Teaming

Fokus auf Gesundheit, Mobilität, Finanzen und öffentlicher Sektor

www.aiqualityhub.com



01 Was ist KI-Qualität und warum ist sie wichtig?

Co-Piloten im medizinischen Kontext

The screenshot displays the Prof. Valmed AI chat interface. At the top left is the logo and name 'PROF. VALMED®'. In the top center, there is a 'Prompt' dropdown menu. At the top right, there is an 'Account' icon. On the left side, there is a 'Filter' section with a search bar and two expandable categories: 'Type' and 'Source'. The 'Type' category is expanded, showing a list of 10 filter options: 1. Guidelines, 2. Prescribing informatio..., 3. Quality Handbooks, Pa..., 4. Pubmed, 5. Overview Articles, 6. Original Articles, Journ..., 7. Cochrane library, 8. Books, Bookchapters, 9. Drug Databanks, and 10. Press releases. Below the filter list is a blue button labeled 'SELECT DOCUMENTS'. The main chat area on the right shows a conversation. A user message (labeled 'You') asks 'How to treat MS'. The AI response (labeled 'Prof. Valmed') provides a detailed answer: 'Multiple sclerosis (MS) is a lifelong disease with no known cure, but it can be treated with prescription drugs and rehabilitation. [1] The US FDA has approved several medications known as disease-modifying therapies (DMTs), thought to reduce the number and severity of relapses and slow the progression of disability. New medications are under development and review by the FDA. Medication, rehabilitation, and other interventions can also help relieve MS-related symptoms. [2] MS can be managed through medication and non-pharmacological management. [1] Medicines are available to treat fatigue and mobility problems associated with MS, but non-medical approaches can also help, such as supervised exercise programs and cognitive-behavioral techniques. [4] A combination of supervised aerobic and moderate progressive resistance activity and cognitive-behavioral techniques may be helpful for people with MS with'. At the bottom of the chat area, there is a text input field labeled 'Enter prompt ...' and a blue button with an upward arrow. Below the chat area, a small disclaimer reads: 'You are chatting with an AI: Consider checking important information.'

Risiken generative KI

Chancen und Risiken generativer KI

Risiken von generativer KI



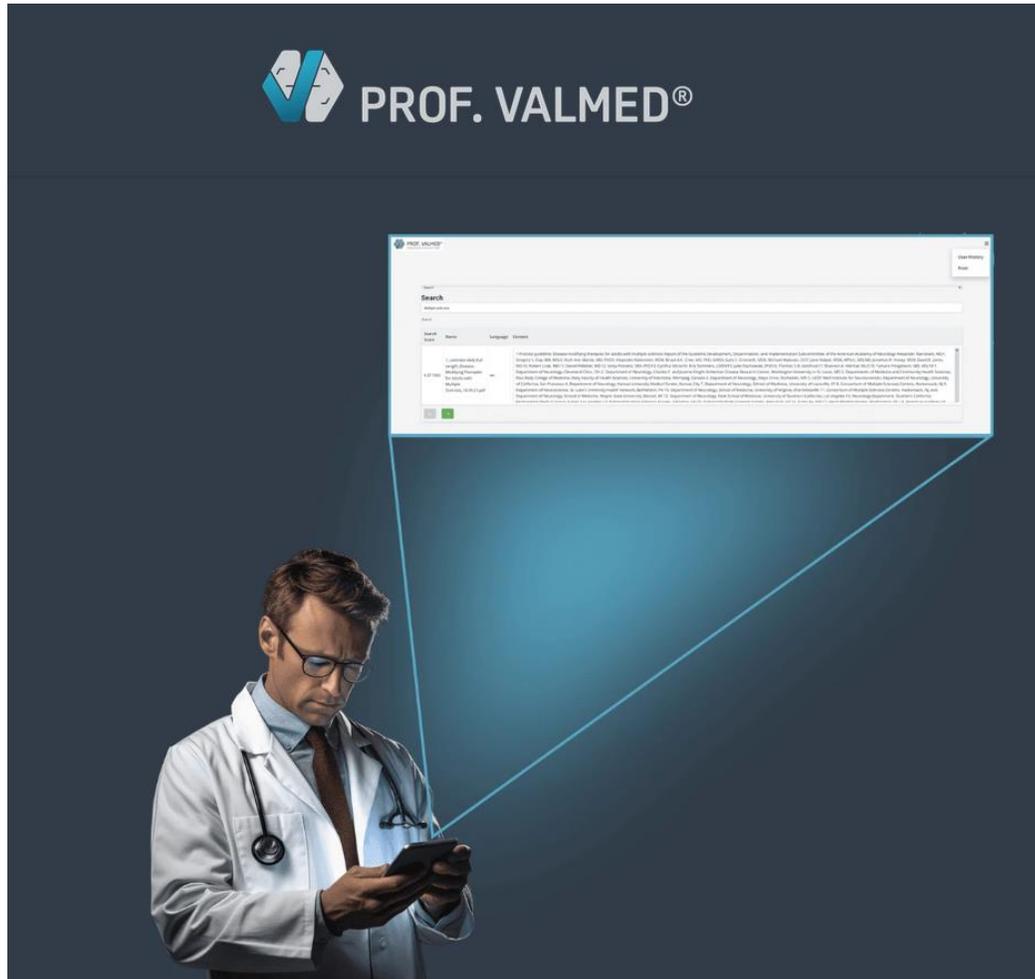
Gegenmassnahmen zu Risiken generative KI

Chancen und Risiken generativer KI

Gegenmaßnahmen im Kontext generativer KI



Definition von Umfang, Kontext und Kriterien der KI-Qualität



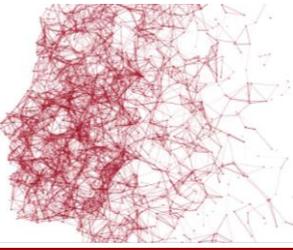
Large Language Model gestützte Anwendungen im Gesundheitswesen

Technische und wirtschaftliche Aspekte:

- Die Kundennachfrage verstehen.
- Abwicklung der Kundeninteraktion.
- Kann auf unbekannte Situationen in der menschlichen Interaktion reagieren.
- **Niedrige Fehlerraten erforderlich, einschließlich Failover-Mechanismen**
- **Schnelle Inferenz.**
- **Kosten wirtschaftlich vertretbar.**

Normative und ethische Aspekte:

- **Safety:** Körperliche Schädigung bei falscher Diagnosehilfe.
- **Privacy:** Keine persönlichen/Unternehmensdaten, indem Sie dazu auffordern (z.B.) oder die Privatsphäre oder Würde von Personen verletzen, indem personenbezogene Daten ohne Zustimmung erhoben werden.
- **Das Urheberrecht muss beachtet werden.**
- Sicherstellung von Transparenz in der kritischen Kommunikation hin zur assistierten Diagnose.
- **Fairness: Alle erhalten die gleiche Antwortqualität.**



02 Neue Herausforderungen bei der Entwicklung!

Wichtige technische Prinzipien zur Erreichung von KI-Qualität

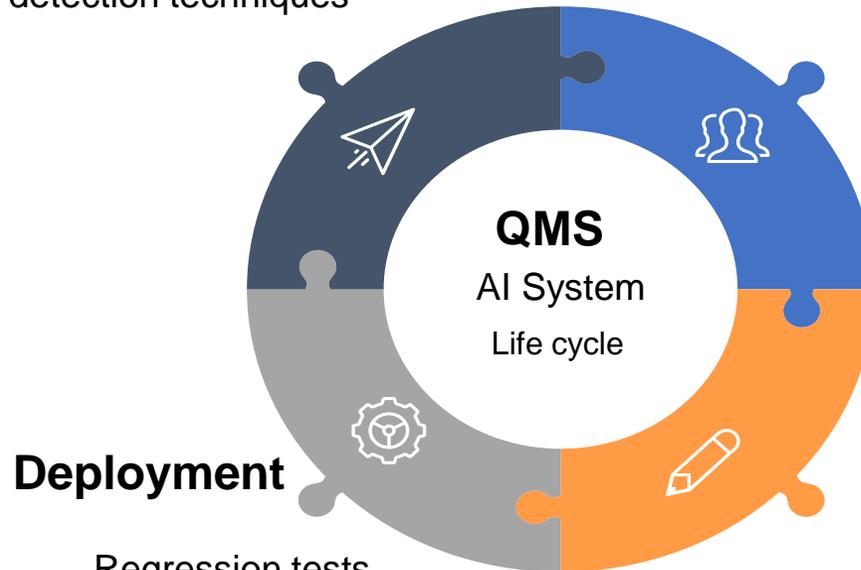
Monitoring

ODD Operation, analysis and monitoring
 “humans-in the loop”
Reference data sets
Data drift detection techniques

Development

Design, data and model planning
Data quality tools
Quality assured data sets (test/fine-tuning)
Annotation strategies (“human in the loop”)
Choice of foundational model/APIs
 Properties-by-design
 Meta-data descriptions (**Data sheets/Model cards**)
Fallback-mechanism (“human in the loop”)

AI Ops/MLOps



Deployment

Regression tests
Qualified data for adversarial testing
Intelligent regression tests
(e.g. other foundational models)
 Integration to larger systems

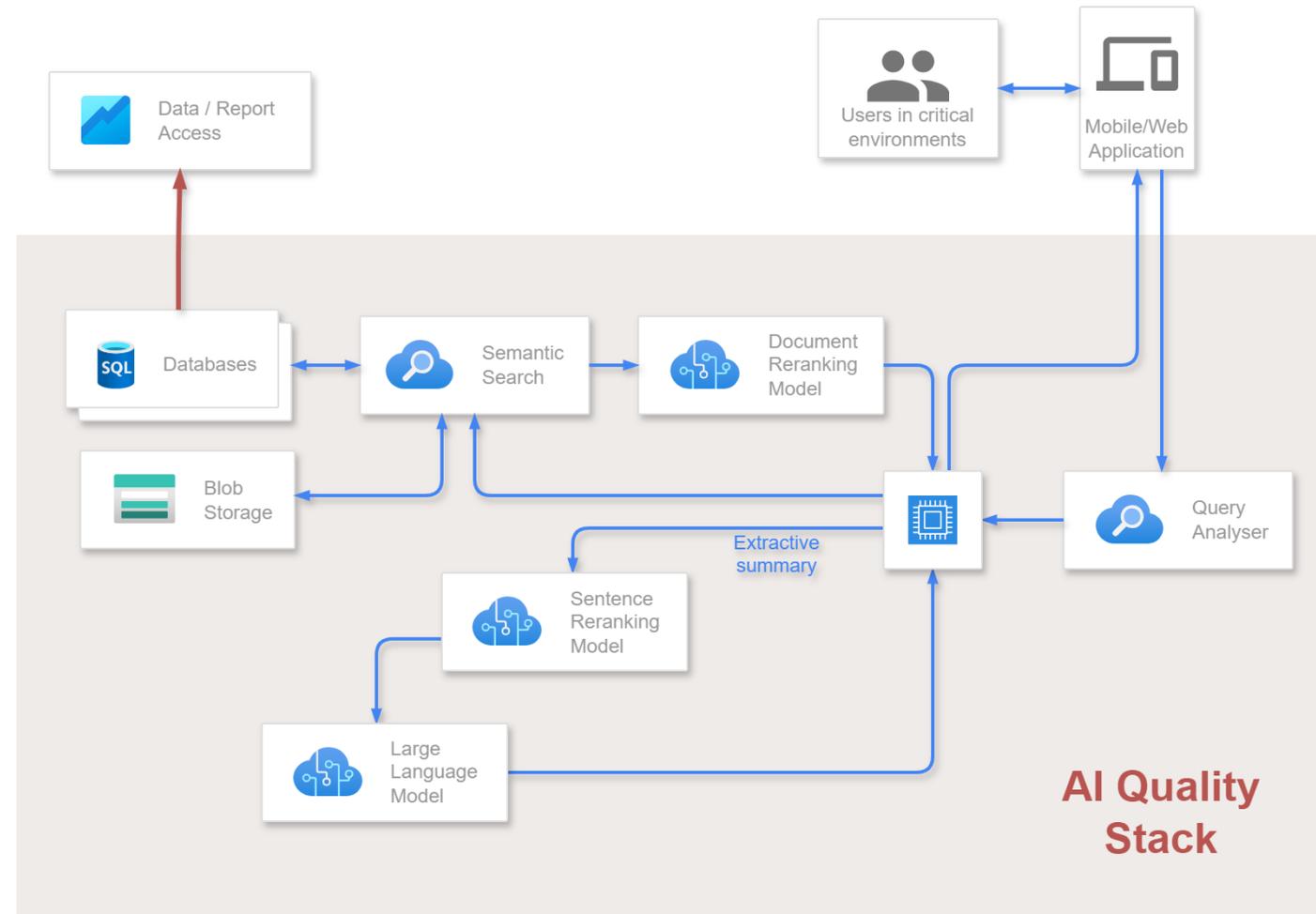
Verify

Verification and validation
Simulation Frameworks
 Metrics, e.g. **factuality**

Wichtige technische Prinzipien zur Erreichung von KI-Qualität

Key facts:

- Regulated environment
- Enterprise-ready with MS Azure
- Modern front-end
- Built-in AI quality
- Guaranteed performance, consistency and stability
- LLM interchangeable (**no lock-in**)
- Execution of trial runs for fine-tuning and acquisition of annotated data and measurement of performance
- Operational support available (Tier 1/2/3)



Retrieval Augmented Generation!

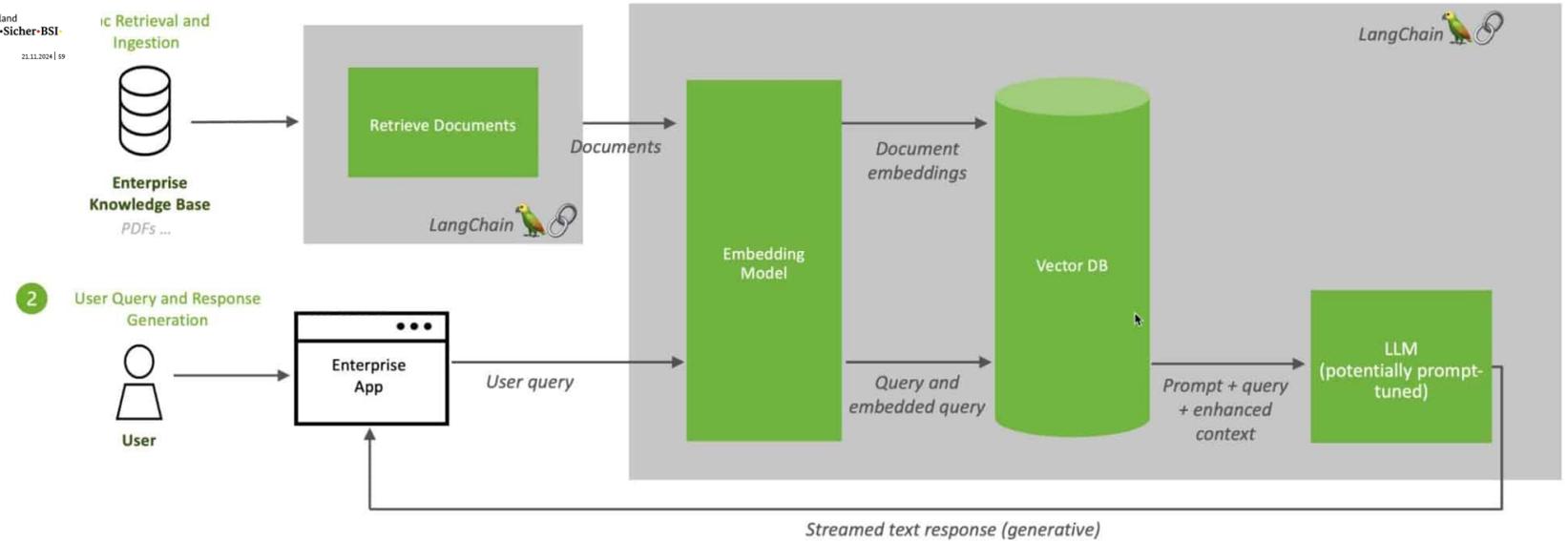
Chancen und Risiken generativer KI

Gegenmaßnahmen im Kontext generativer KI



Deutschland Digital-Sicher-BSI- 23.11.2024 | 39

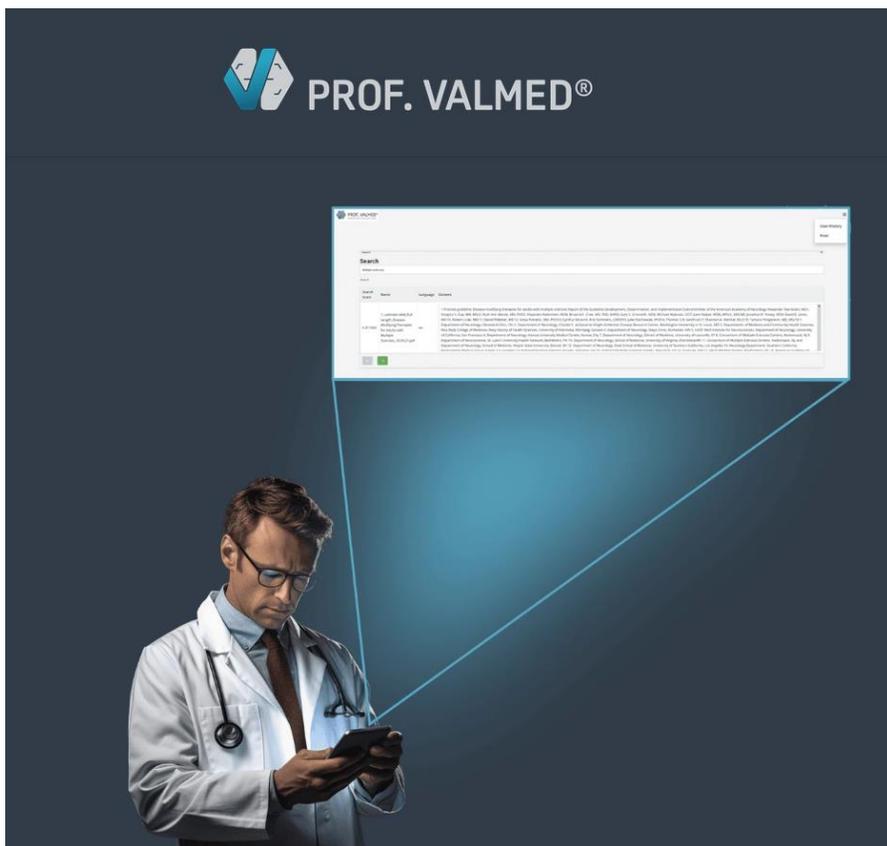
Retrieval Augmented Generation (RAG) Sequence Diagram



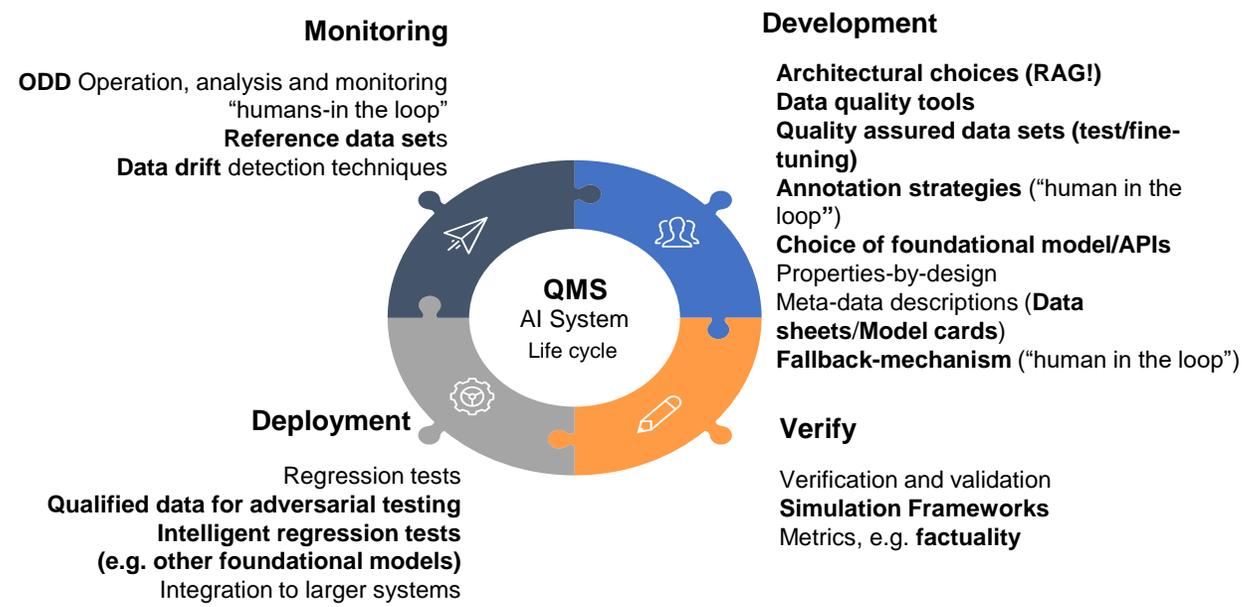
<https://blogs.nvidia.com/wp-content/uploads/2023/11/NVIDIA-RAG-diagram-scaled.jpg>

Wichtige technische Prinzipien zur Erreichung von KI-Qualität

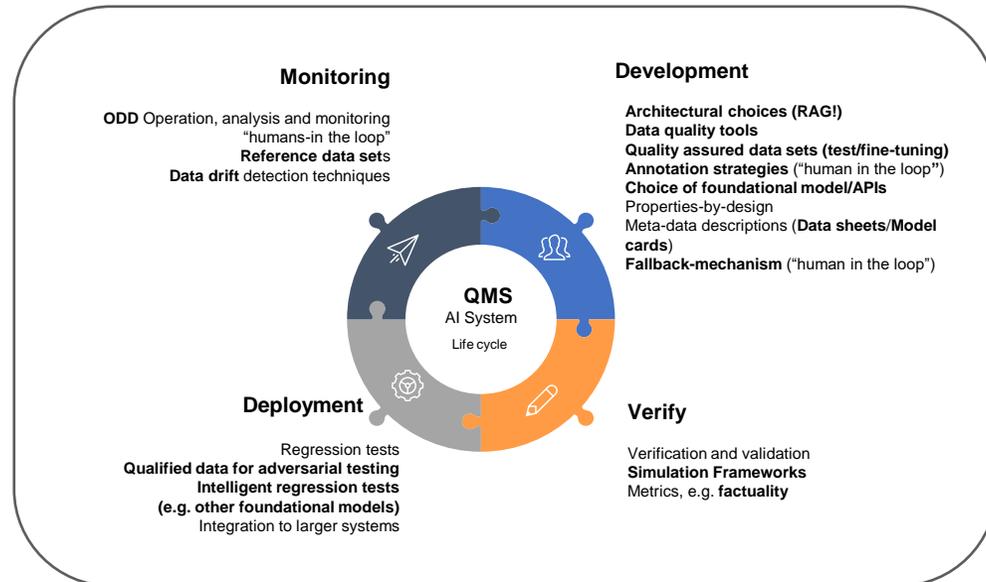
Gegenmaßnahmen gegen Halluzinationen sind kluge Entscheidungen der Systemarchitektur (Retrieval-Augmentation Generation) und intelligente Regressionstests zur Optimierung der Faktentreue



Konfabulation/Halluzinationen in Chatbots beziehen sich auf das Problem der scheinbar richtigen Antworten, die sachlich falsch sind.



Neue Herausforderungen im Qualitätsmanagement



Validierung im Sinne des Qualitätsmanagements

- Einbettung in das QM-System des Herstellers/Deployers
- Bewertung der Konformität mit MDR/In-Vitro-Verordnung, KI-Verordnung
- Benannte Stellen
- Klinische Studien, andere Validierungssysteme

Nachweis der Robustheit für Hoch-Risiko Systeme nach KI-Verordnung

Part of [Chapter III: High-Risk AI System](#) → [Section 2: Requirements for High-Risk AI Systems](#)

Article 15: Accuracy, Robustness and Cybersecurity

Date of entry into force: [2 August 2026](#)
According to: [Article 113](#)

See here for a [full implementation timeline](#).

2. To address the technical aspects of how to measure the appropriate levels of accuracy and robustness set out in paragraph 1 and any other relevant performance metrics, the Commission shall, in cooperation with relevant stakeholders and organisations such as metrology and benchmarking authorities, encourage, as appropriate, the development of benchmarks and measurement methodologies.

[Article 15: Accuracy, Robustness and Cybersecurity | EU Artificial Intelligence Act](#)

-> **Hersteller** von **Hoch-Risiko KI-Anwendungen** im medizinischen Bereich.

AIQ: Specialised in AI Quality by Design

LLM Testing Stack

Level 1) Consistency checks:

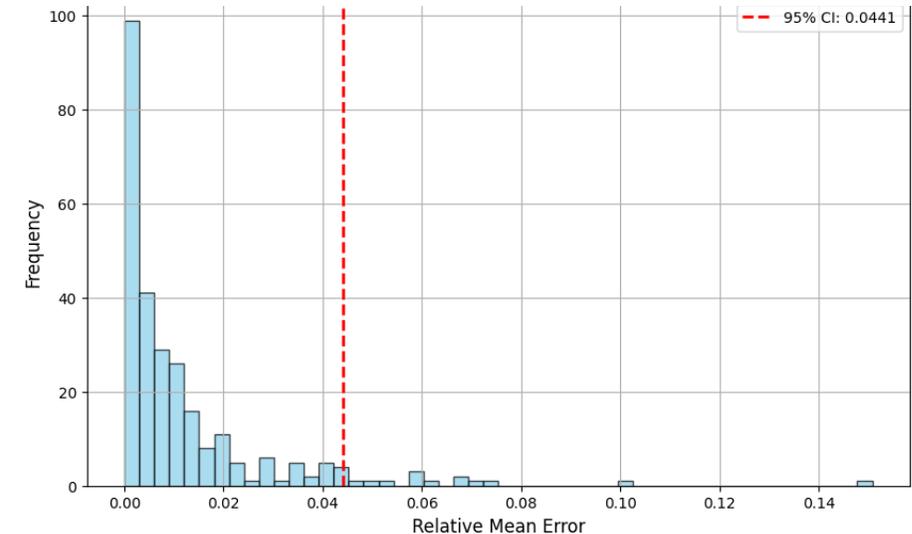
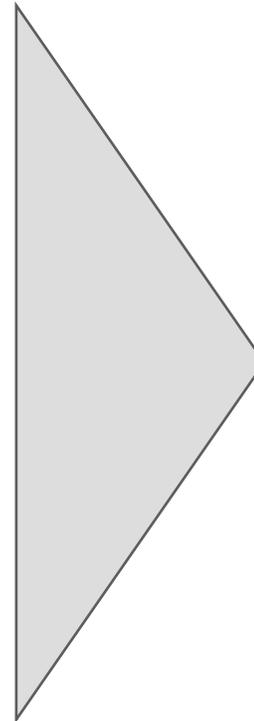
- No annotated data needed
- Hallucination countermeasure

Level 2) Answer Quality checks:

- No domain data needed
- Testing Language phase space

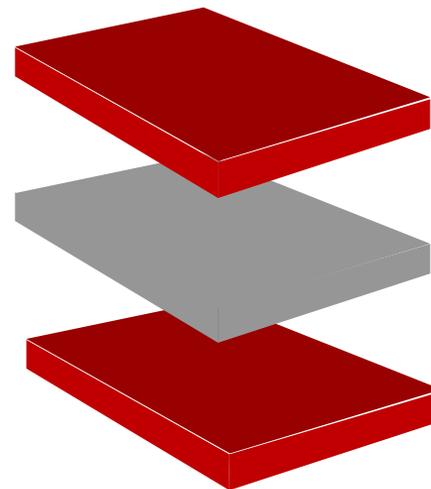
Level 3) Domain Quality checks:

- Domain data needed
- Testing domain phase space



Die Auswirkungen der KI-Verordnung auf die Produktentwicklung

- General Purpose AI-Modelle (GPAI) unterliegen wie großen Sprachmodellen je nach systemischem Risiko spezifischen Verpflichtungen, einschließlich technischer Dokumentation, Transparenz, Einhaltung des Urheberrechts und systemischer Risikobewertung.
- GPAI-Modelle mit hoher Wirkung unterliegen einer zusätzlichen Prüfung, einschließlich Adversarial Testing(!) und Cybersicherheitsmaßnahmen.
- Das KI-Gesetz soll formell Mitte 2026 in Kraft treten, wobei bestimmte Bestimmungen für verbotene und hochriskante KI-Systeme früher angewendet werden



EU AI ACT

GDPR

Industry-specific
regulation: **MDR, IN-
Vitro,...**



AI QUALITY &
TESTING HUB



Dr. Michael Rammensee

Managing Director

m.rammensee@aiqualityhub.com

+49 176 10553180