Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Workshop: Vertrauenswürdigkeit von KI-Lösungen

GI-Fachgruppe "Measurement & Data Science (FG 2.1.10) im Rahmen der ESAPI-Community

Andreas Schmietendorf – HWR Berlin & OvG-Universität Magdeburg

Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Agenda des Workshops

Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Eingeladene Gastvorträge (10:00 bis 12:00 Uhr)

- *Dr. Gaby Gurczik*, Referentin für Grundsätze KI und Datenökonomie beim BMDV

  KI-Innovationen als Standortchance für Deutschland und Europa

- *Dr. Rasmus Adler*, Leiter des Programms Autonome Systeme am Fraunhofer IESE

  Das Spaltmaß für KI-Systeme - Wie sieht es aus und was sind akzeptable Grenzwerte?

- *Prof. Dr. Katharina Zweig*, Leiterin Algorithm Accountability Lab TU Kaiserslautern

  Kann man mit Surrogatansätzen KI-Entscheidungen erklären?

Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Poster-Session (13:00 bis 14:30 Uhr)

- Sandro Hartenstein, HWR Berlin & OvG Universität Magdeburg
  Vertrauenswürdige KI-WebAPI Spezifikationen
- Julius Schinschke, HWR Berlin
  Vertrauen Datenquellen, online
- Lukas Scholz, DB Systel
  Explainable Artificial Intelligence, online)
- Daniel Krohmer, Fraunhofer IESE
  KI Security Testing
- Dr. Michael Kläs, Fraunhofer IESE
  Safe AI

Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Partner des Workshops

- Fraunhofer-Institut für Experimentelles Software Engineering (Gastgeber)
  www.iese.fraunhofer.de
- Gesellschaft für Informatik (Veranstalter)
  https://fg-data-science.gi.de
- Central Europe Computer Measurement Group (Sponsoring)
  cecmg.de
- Arbeitskreis Software-Qualität und -Fortbildung e.V. (ASQF)
  www.asqf.de
- SIGS DATACOM GmbH (Medienpartner)
  www.sigs-datacom.de

GESELLSCHAFT FÜR INFORMATIK

CECMG

ASQF® Das Expertennetzwerk

SIGS DATACOM FACHINFORMATIONEN FÜR IT-PROFESSIONALS

Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Motivation für den Workshop

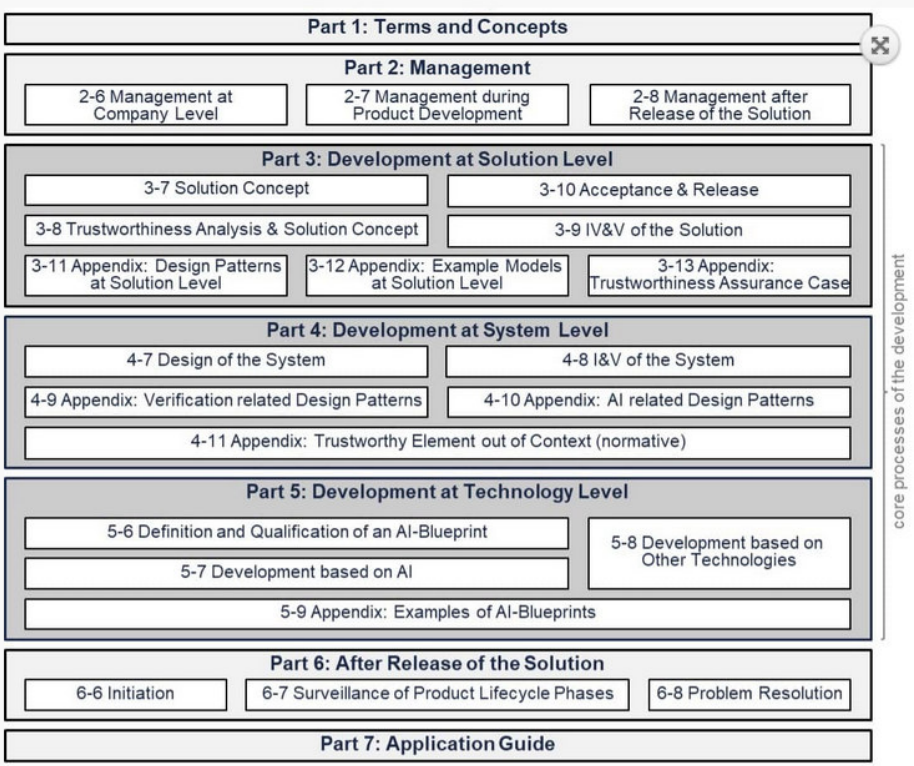
Motivation für den Workshop

- Vielfältige praxisrelevante KI-Anwendungsszenarien
  - Sentiment-Analysen für ein besseres Kundenverständnis.
  - Bewältigung massenhafter Problem-Tickets (Klassifikation).
  - Bild- und Videoverarbeitung zur Gefahrenerkennung.
  - …
- Forschungsorientierte KI-Fragen
  - Vertrauen in KI-Algorithmen aus der „Steckdose“.
  - Test- und Erklärbarkeit von KI-Ergebnissen.
  - KI als Unterstützung im Software-Engineering bzw. Reengineering.
  - …

Umgang mit Vertrauenswürdigkeit

"Trustworthiness […] combines several aspects of trustworthiness in a quite generic way: for every product the set of aspects can be suitably selected and remains unchanged throughout the project. Aspects of trustworthiness include but are not limited to system safety, functional safety, safety of use, security, usability, ethical and legal compliance, reliability, availability, maintainability, and (intended) functionality."

VDE-AR-E 2842-61-1 Kapitel 3.1.43, DKE-Arbeitskreises 801.0.8



Quelle : Johner, C.: Weshalb die VDE-AR-E 2842-61 (vertrauenswürdige KI-Systeme) nicht nur die Entwicklung betrifft https://www.johner-institut.de/blog/systems-engineering/ki-systeme/#section_scroll3

Fraunhofer IESE

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Backup: KI-Verständnis

Verständnis von KI und ML

„Künstliche Intelligenz beschreibt solche Informatik-Anwendungen, deren Ziel es ist, intelligentes Verhalten zu zeigen. Dazu sind in unterschiedlichen Anteilen bestimmte Kernfähigkeiten notwendig: Wahrnehmen, Entscheiden, Handeln und Lernen…“

„Maschinelles Lernen beschreibt die Technologie, die es Computersystemen ermöglicht, aus Beispielen, Daten und Erfahrungen zu lernen und so Aufgaben eigenständig zu lösen bzw. die bisherige Lösung selbstständig zu verbessern“

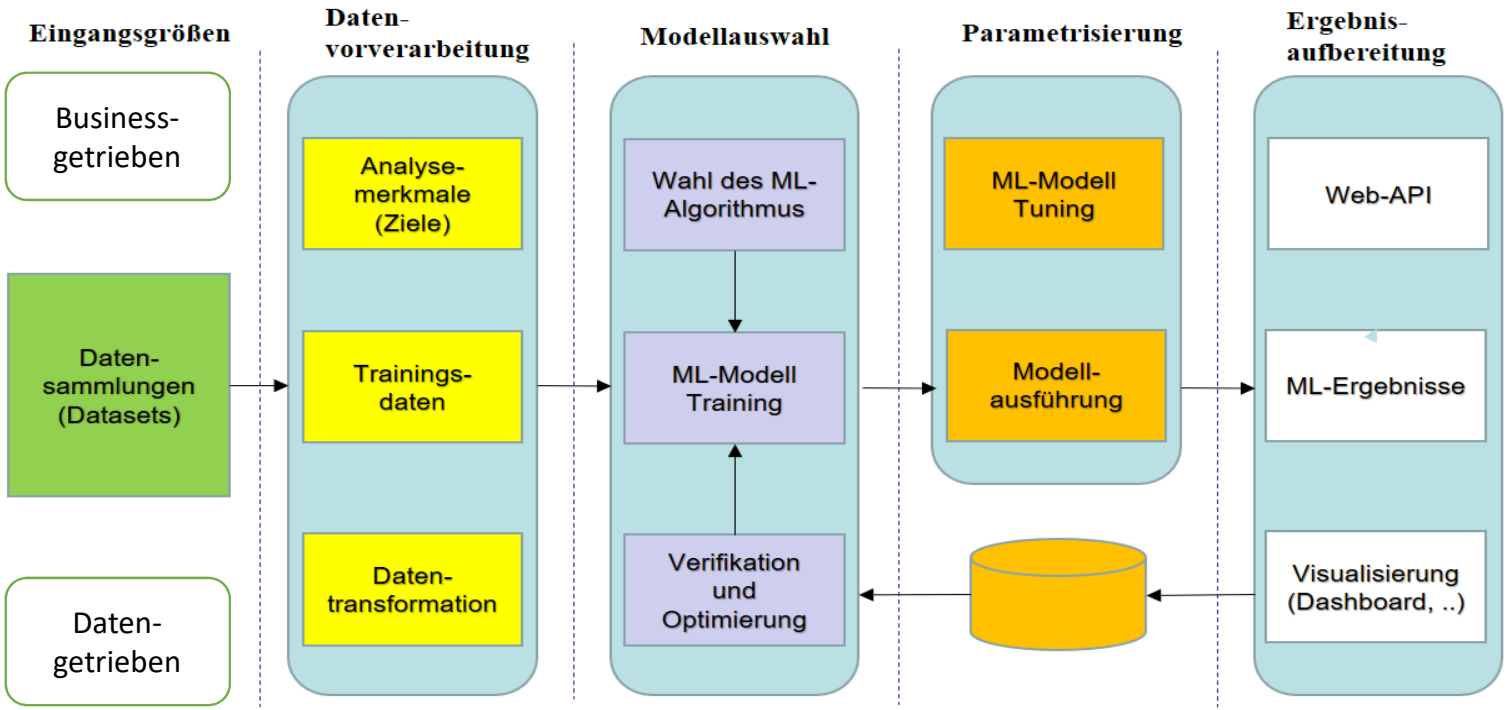
Quelle der KI/ML-Definitionen: Empfehlungen für den verantwortlichen Einsatz von KI und automatisierten Entscheidungen, https://www.bitkom.org/sites/main/files/file/import/180202-empfehlungskatalog-online-2.pdf, letzter Zugriff 15. September 2022

Quelle der rechten Abb.: https://sm.ign.com/ign_de/blogroll/r/r2-d2-mini/r2-d2-mini-fridge-will-bring-drinks-to-you_h7ah.jpg

Fraunhofer IESE

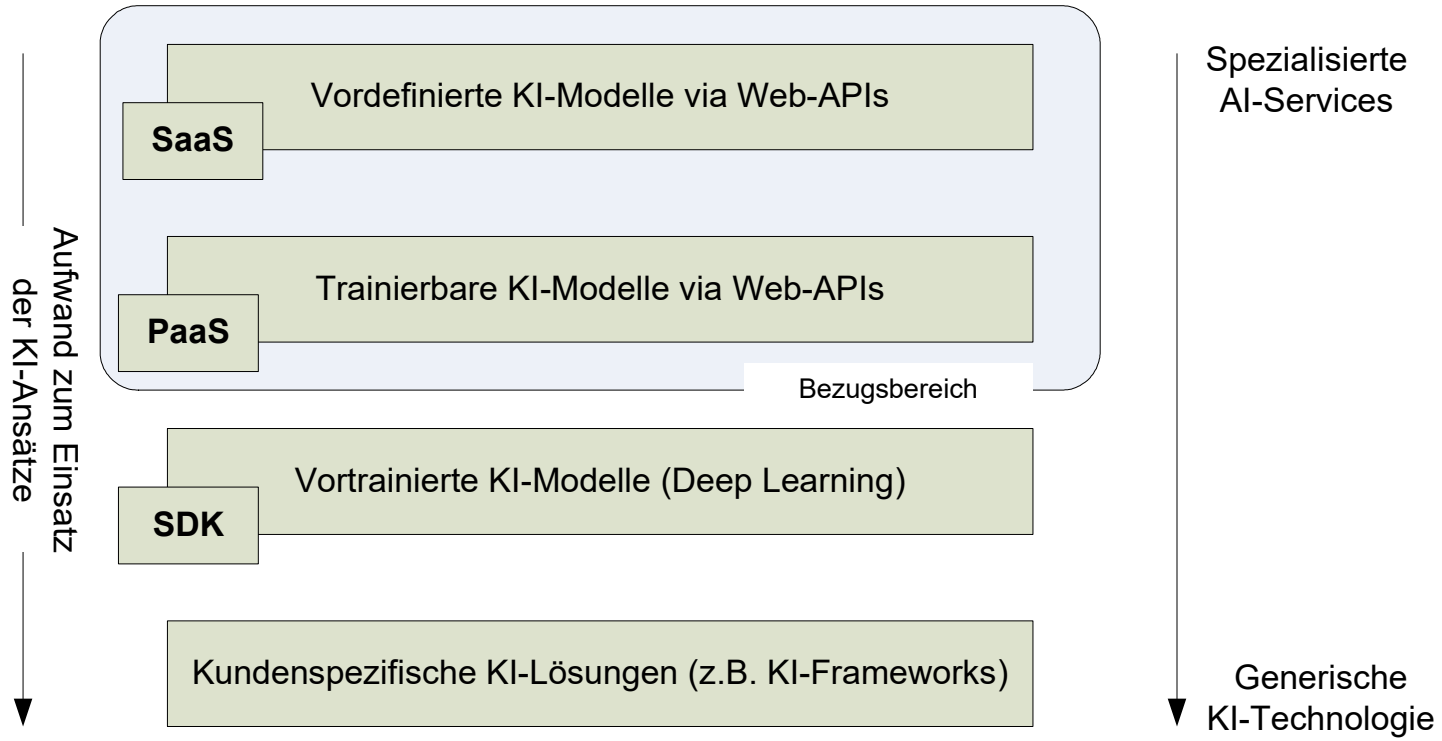
Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Motivation – agil orientiertes Engineering



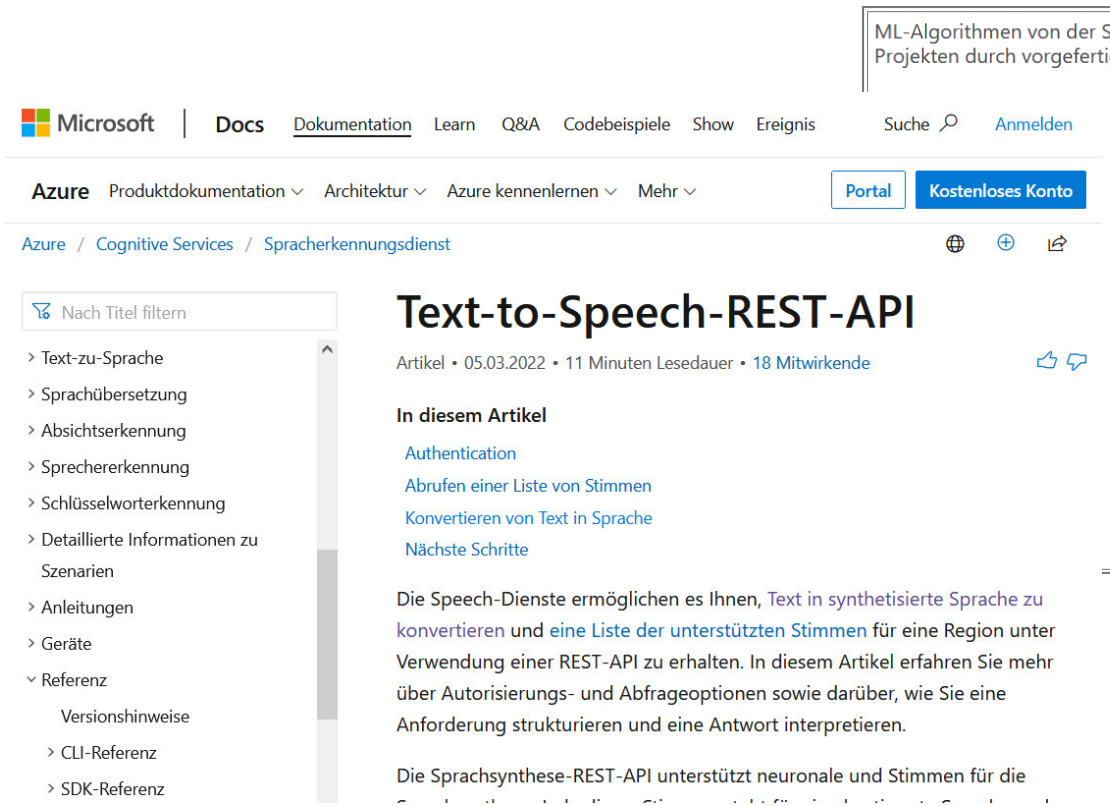
In Anlehnung an: El Shawi, R.; Mohamed Maher, M.; Sakr, S.: Automated machine learning: State-of-the-art and open challenges, University of Tartu, Estonia, Juni 2019.

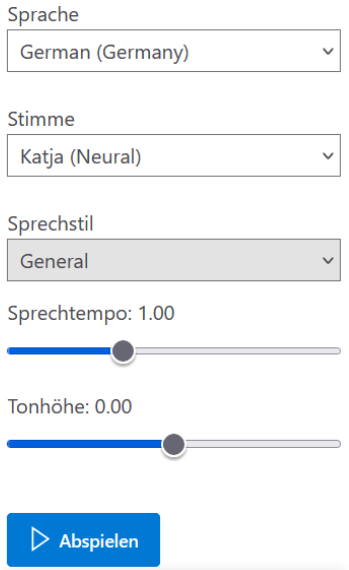
Fokus auf bereits implementierte KI-Algorithmen



Quelle: Fiegler, A.; Schmietendorf, A.: Entwicklung smarter Anwendungen mit Hilfe cloudbasiert angebotener KI-Algorithmen, in Proceedings ESAPI 2020, S. 01-08, Shaker-Verlag, Aachen, November 2020

Beispiel für PaaS/SaaS KI-Algorithmen





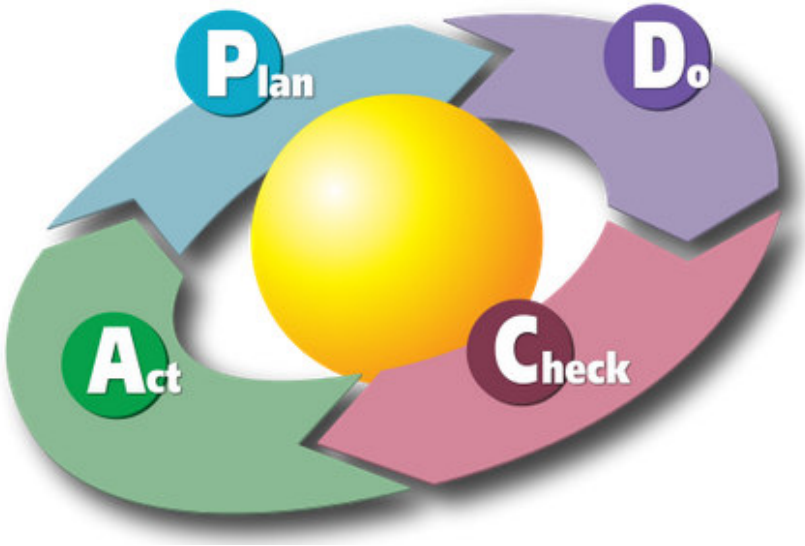
SaaS-Sicht – GUI

PaaS-Sicht – REST-APIs

Quelle: https://docs.microsoft.com/de-de/azure/cognitive-services/speech-service/rest-text-to-speech#convert-text-to-speech, letzter Abruf: 09.03.2022

Einhergehende Implikationen

- Agile durchgeführte KI-Experimente bzw. Tests.
- Einsatz problemadäquater (KI-) Algorithmen.
- Lernende Projektorganisationen akzeptieren.
- Qualitative Aspekte der KI-Lösung vgl. ISO 25.000.
- Gewährleistung von Sicherheit und Vertrauen.
- Berücksichtigung von Gesetzen und Compliance.
- Reproduzier- und Erklärbarkeit.
- Offene Ergebnisdiskussion mit Stakeholdern
- …



Quelle der Abbildung: https://en.wikipedia.org/wiki/PDCA, letzter Zugriff 15. September 2022