

DIGITALISIERUNG FÜR KMU **MÖGLICH** MACHEN

DER DIGITAL INNOVATION HUB SÜD ALS
KOSTENLOSES SERVICE FÜR KMU



Data Driven Dialog, 26. November 2025, Lakeside Park Klagenfurt

Dokumentenverwaltung neu gedacht

ChatGPT & Co. für KMUs

Programm

- Begrüßung und Kurzeinführung Fraunhofer Austria und DIH Süd
Dr. Eva Eggeling
- Fachvortrag »Einführung in das Thema Sprachmodelle«
Ing. Dr. Marco A. Hudelist – Fraunhofer Austria
- Praxisvortrag »Einblicke in den Einsatz von Sprachmodellen in kleinen und mittleren Unternehmen«
Bernd Zwattendorfer – Praxis Querkopf
- Live Demonstratoren-Vorführung (Fraunhofer Austria)
Ing. Dr. Marco A. Hudelist, Dr. Christina Petschnigg
- Diskussion, Networking & Ausklang

Fraunhofer Austria Research GmbH

Innovative Lösungen für das Heute von morgen

ChatGPT 5 Thinking ▾

↑ Gemeinsam nutzen ...

Fraunhofer Austria Research GmbH ist... ?

Nachgedacht für 31s >

die rechtlich eigenständige Landesgesellschaft der Fraunhofer-Gesellschaft – Europas größter Organisation für angewandte Forschung. An vier Standorten (Wien, Graz, Wattens, Klagenfurt) arbeiten rund 136 Expert:innen in zwei Zentren an praxisnahen Lösungen **für Industrie und öffentliche Hand** – immer mit Blick auf Effizienz, Nachhaltigkeit und Wettbewerbsfähigkeit.

📄 👍 🗨️ ↗️ ↺️ ... 🌐 Quellen

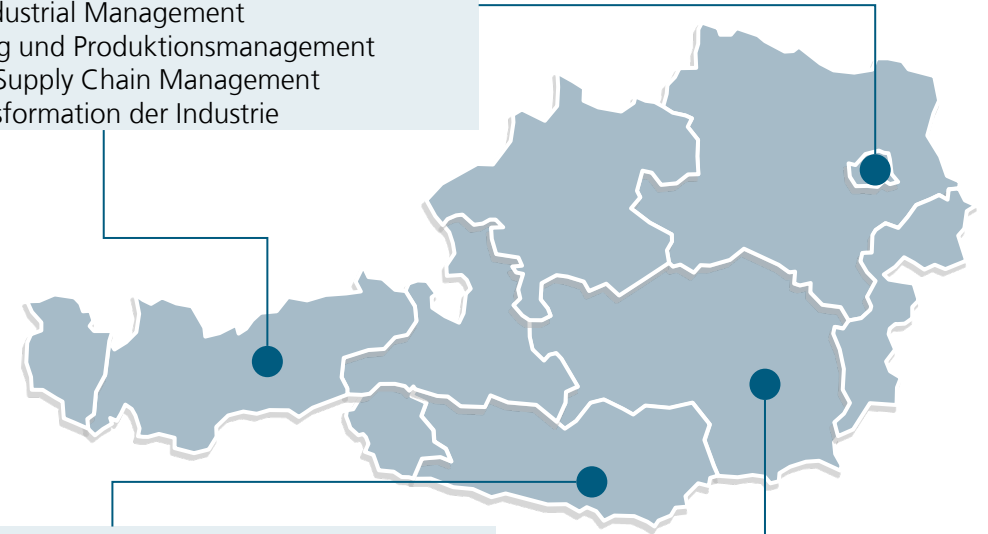
Stelle irgendeine Frage

+ 🔄 Umfassendes Nachdenk... ▾



Wien/Wattens Center für Nachhaltige Produktion und Logistik

- Advanced Industrial Management
- Fabrikplanung und Produktionsmanagement
- Logistik und Supply Chain Management
- Digitale Transformation der Industrie



Graz/Klagenfurt Center für Data Driven Design

- Digitalisierung und Künstliche Intelligenz KI4LIFE
- Visual Computing

Daten Driven Dialog: ChatGPT & Co. für KMUs

Kurzeinführung Sprachmodelle

Ing. Dr. Marco A. Hudelist

ChatGPT

■ Wie funktioniert ChatGPT?

VI Vervollständige einen Satz: Die Schule ist

Die Schule ist eine Bildungseinrichtung, in der Schülerinnen und Schüler Wissen und soziale Kompetenzen erwerben.

VI In einem Satz: Was ist Schule?

Schule ist eine Bildungseinrichtung, in der Schülerinnen und Schüler unterrichtet und auf ihre zukünftigen Lebens- und Berufsaufgaben vorbereitet werden.

Vom Modell erzeugte Wörter werden auch berücksichtigt

Die Schule ist

wahrscheinlich

unwahrscheinlich

eine
müde
ein
klein
doof
groß
toll
viereckig
lustig
kaputt

Einrichtung
Maßzahl
Karte
Einheit
Klasse
Katze
Institution
Bildungseinrichtung

Wie wird Text generiert?

Beispiel

Prompt

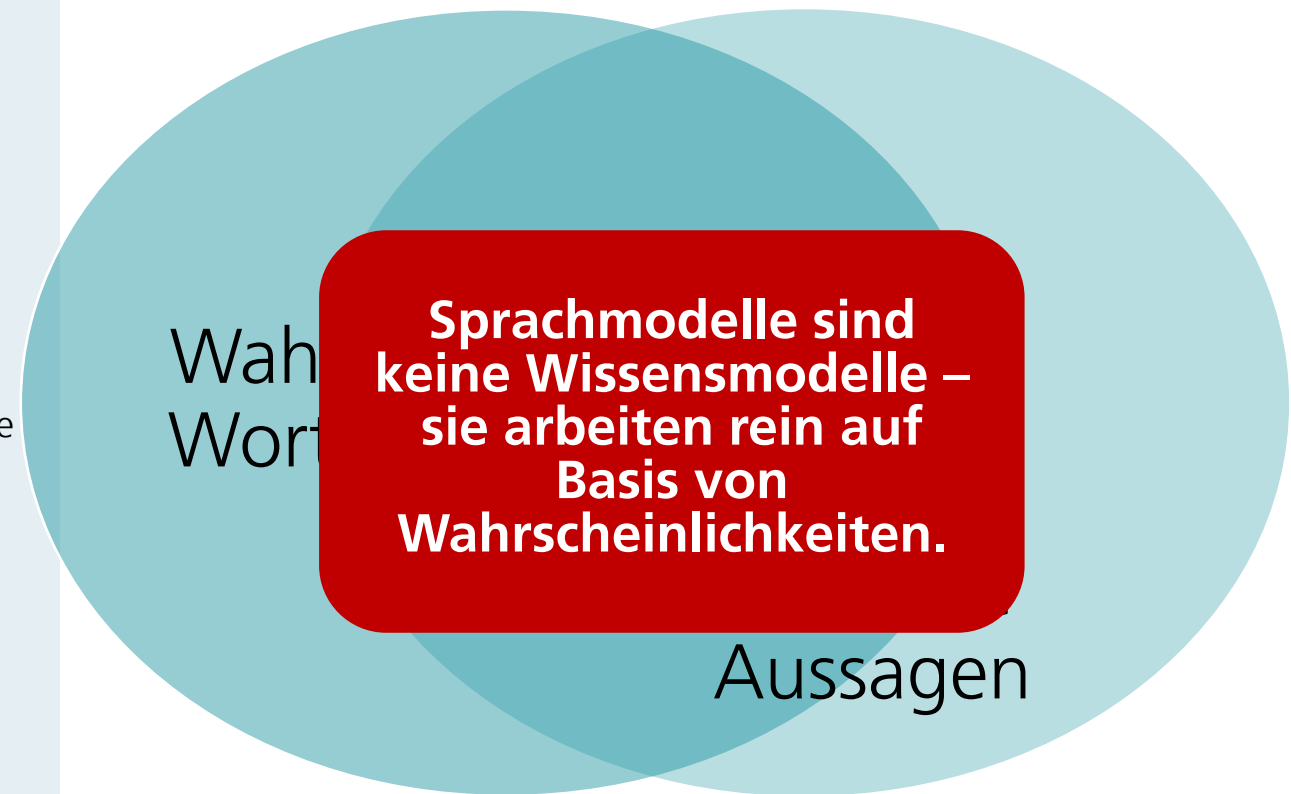
“Die Hauptstadt Frankreichs... ”

Prozess

1. Tokenisierung: “Die| Haupt| stadt| Frank| reichs |...|”
2. Das Modell weist möglichen nächsten Tokens Wahrscheinlichkeiten zu
3. Das wahrscheinlichste Token wird ausgewählt → „ist “
4. Der Vorgang wiederholt sich Token für Token, bis der vollständige Satz erzeugt ist.

Ergebnis

- Input: “Die Hauptstadt Frankreichs **ist**”
- Output: “Die Hauptstadt Frankreichs ist **Paris**”



Large Language Modelle (LLMs)

Grundlagen

Definition

- LLMs sind fortgeschrittene Deep-Learning-Modelle, die auf riesigen Mengen an Textdaten trainiert werden, um menschenähnliche Sprache zu verstehen und zu erzeugen.
- Oftmals wird »KI« mit LLMs heute gleichgesetzt (was nicht stimmt)

Kernpunkte

- Transformer Architektur
- Tokenisieren Text in kleinere Einheiten
- Können viele Aufgaben ohne aufgabenspezifisches Training ausführen
- Allgemein einsetzbar: Zusammenfassung, Übersetzung, Schlussfolgern/Reasoning, Code-Generierung, uvm.

Beispiele

- ChatGPT (OpenAI), Google Gemini, Anthropic Claude, Meta's LLaMa, ...
- Integriert in Microsoft Copilot, ...

LLM



Retrieval-Augmented Generation

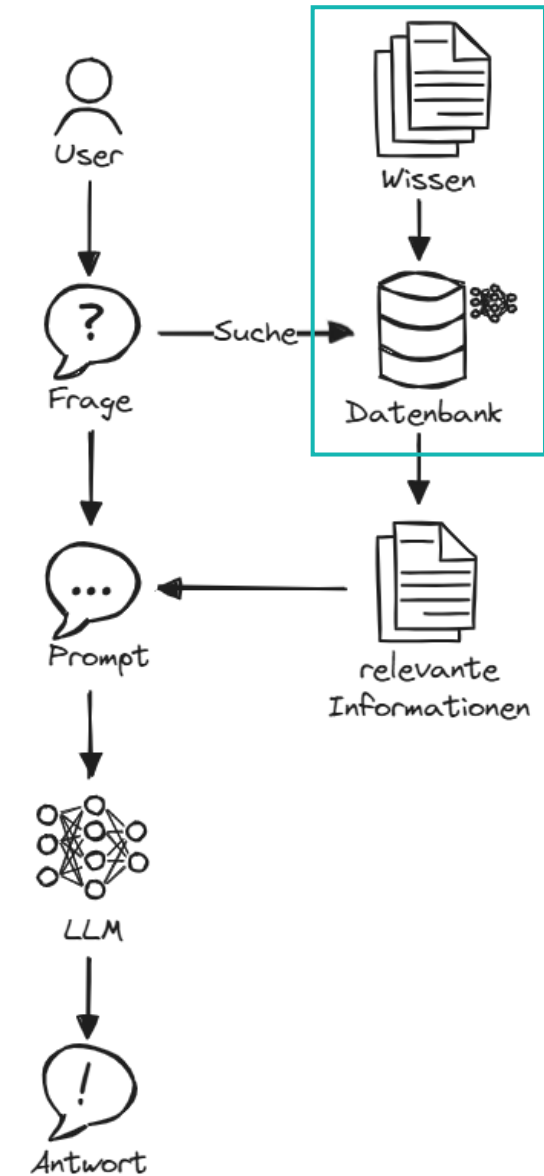
Wie funktioniert RAG?

RAG kombiniert LLMs und Dokumentensuche

- Antwort basiert nicht auf Trainingswissen, sondern auf externen Quellen
- Besonders geeignet für aktuelle oder unternehmensspezifische Fragen

Komponenten

1. Datenaufbereitung
 - Textaufbereitung
 - Aufbereitung einer Vektordatenbank

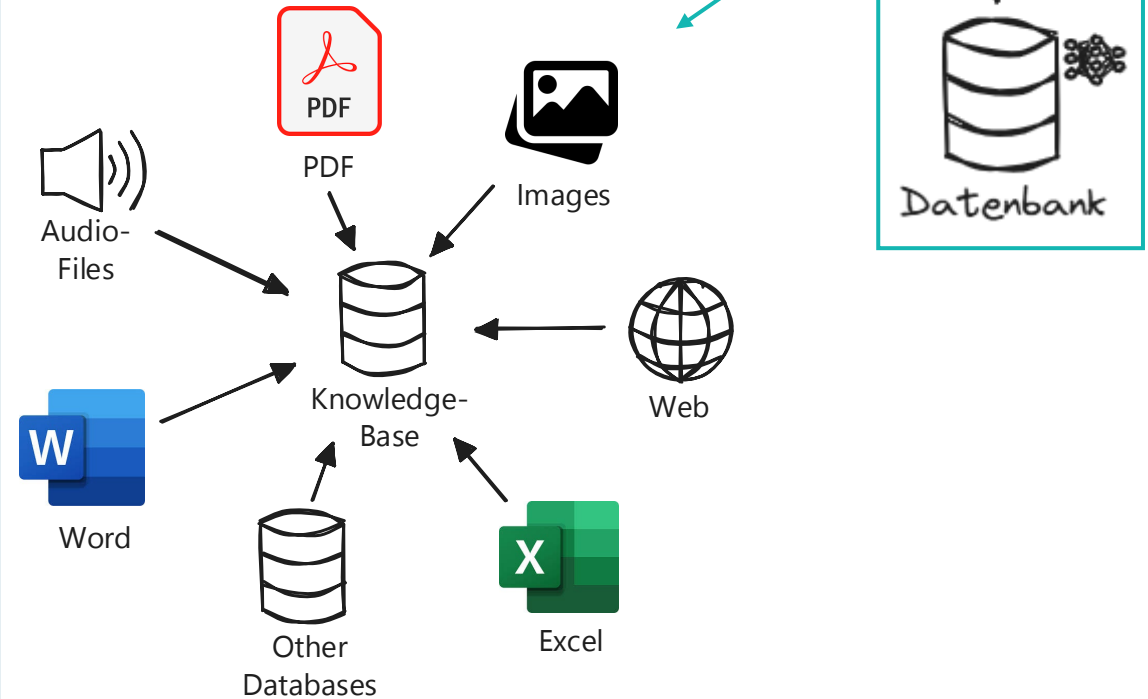


Retrieval-Augmented Generation

Vektordatenbank (Wissensdatenbank)

Welche Datentypen können eingebunden werden?

- Textdokumente (PDF, Word, PowerPoint, TXT)
- Tabellen (Excel, CSV)
- Inhalte von Webseiten
- Bilder
- Audioaufzeichnungen
- Datenbanken und APIs
- **Wichtig: Alle Inhalte müssen in Textform vorliegen**



Assistent für Produktinformationen

Ausgangssituation

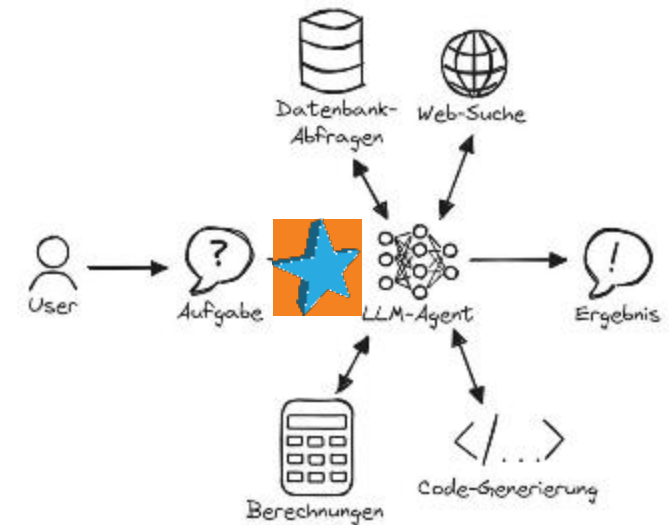
- Die Firma hat ein sehr **umfassendes Produktportfolio**. Jedes dieser Produkte hat bestimmte **Kennwerte und Eigenschaften**. Es ist daher für MitarbeiterInnen unmöglich, alle Details zu kennen und das Fehlerpotential durch Verwechslungen ist enorm. Ein Tool, das gewünschte Informationen effizient aus einer umfangreichen Wissensdatenbank extrahiert, ist erforderlich.

Umsetzung liefert

- Effizienzsteigerung** der Informationsverarbeitung
- Wissen wird **zentral verwaltet**
- Das Tool wurde als Chat-Agent implementiert, der Informationen effizient aus einer umfangreichen Wissensdatenbank extrahiert.
- Ein Tool für alle Bereiche des Unternehmens

Wie kommt das domänenspezifische Wissen in das Modell?
→ Domänen LLMs

Zentrale Wissensdatenbank für MitarbeiterInnen. Das umfangreiche Wissen zu Produkten wird allen MitarbeiterInnen effizient zugänglich gemacht



Unterschied »normales« Sprachmodell und »Reasoning Sprachmodelle«

Grundidee

- Normales Sprachmodell
 - Welche Wortfolge ist statistisch am wahrscheinlichsten als Nächstes?
- Reasoning Sprachmodell
 - Wie löse ich dieses Problem Schritt für Schritt logisch korrekt?
 - Zerlegt Aufgabe in Zwischenschritte
 - Insbesondere für:
 - Mathe- & Logikaufgaben, Programmierung, Analysen, Planungen, Strategien...
- Offensichtlichster Unterschied
 - Reasoning Sprachmodelle benötigen zum Antworten viel länger als normale Sprachmodelle.



Unterschied »normales« Sprachmodell und »Reasoning Sprachmodelle«

Beispiel

- Ein Zug fährt von Stadt A nach Stadt B. Die Fahrt dauert 2,5 Stunden. Nach 1 Stunde Fahrt hält er 10 Minuten an einem Bahnhof. Wie lange sind die Fahrgäste insgesamt unterwegs?



Unterschied »normales« Sprachmodell und »Reasoning Sprachmodelle«

»Normales Sprachmodell«



Erkennung der Zahlen im Prompt: »2,5 Stunden«, »1 Stunde«, »10 Minuten«



Auffälligste Zahl/Textteil: »2,5 Stunden dauert die Fahrt«



Antwort: »Die Fahrgäste sind 2,5 Stunden unterwegs.«

Unterschied »normales« Sprachmodell und »Reasoning Sprachmodelle«

»Reasoning Sprachmodell«



1. Was wird eigentlich gefragt?

- Nicht nur Fahrzeit, sondern „unterwegs“ = Fahrt + Halt.

2. Gegebene Infos sortieren

- Fahrzeit gesamt: 2,5 Stunden
- Halt: 10 Minuten = $\frac{1}{6}$ Stunde

3. Beziehung klären

- Die 2,5 Stunden sind nur die Zeit, in der der Zug fährt.
- Der Halt kommt zusätzlich dazu.

4. Rechnung aufbauen

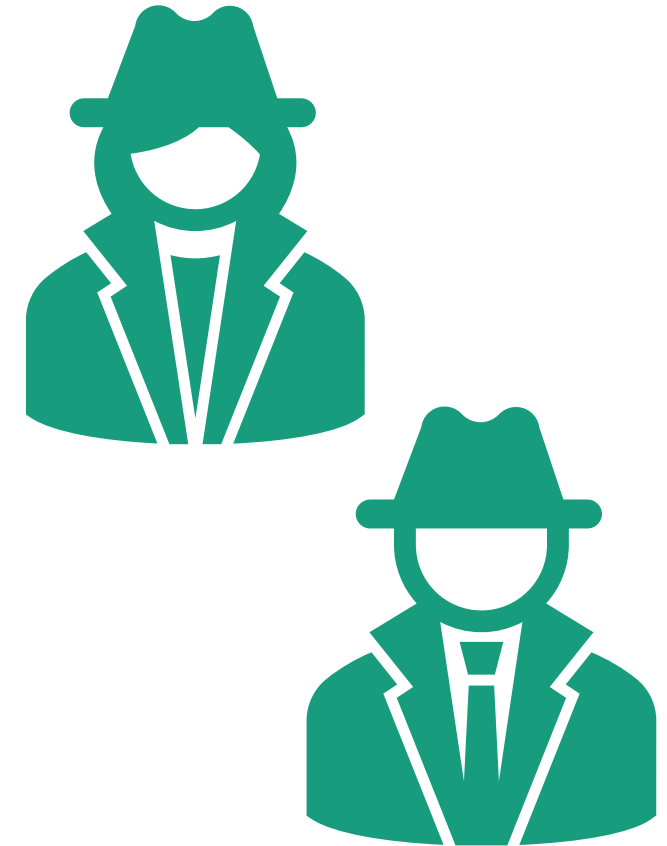
- Unterwegszeit = Fahrzeit + Haltezeit
- Also: 2,5 Stunden + 10 Minuten
- 10 Minuten sind $\frac{1}{6}$ Stunde $\approx 0,17$ Stunden
- Ergebnis in Stunden oder Minuten ausgeben.

5. Antwort sauber formulieren

- **Die Fahrgäste sind ungefähr 2 Stunden und 40 Minuten unterwegs.**

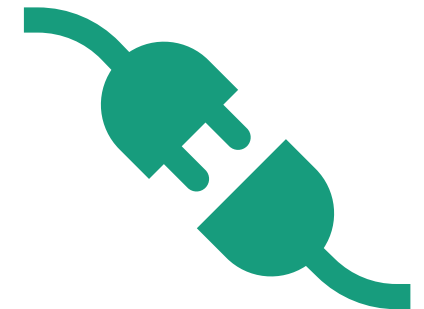
LLM-Agenten (Agents)

- LLM (Sprachmodell), das selbstständig andere Tools/Software nutzt
 - Chatbot: Antwortet in Text, nutzt kaum andere Software (außer genau von User angewiesen)
 - Agent: Handelt und führt mehrere Schritte und Workflows aus
- Beispiel
 - Erstelle einen Markt- und Risiko-Report auf Basis unserer Marktdaten + interner Kennzahlen
- Aktionen des Sprachmodells
 1. Schritte planen (Daten holen – Analyse – Tabellen/Grafiken erzeugen)
 2. Marktdaten über Internet und Finanzdatenanbieter holen
 3. Interne Software nutzen, um Unternehmens-KPIs abzuholen
 4. Durchführen der Analysen/Kalkulationen
 5. Internes Wiki ansprechen um Ergebnisse als Report abzulegen



Model Context Protocol (MCP)

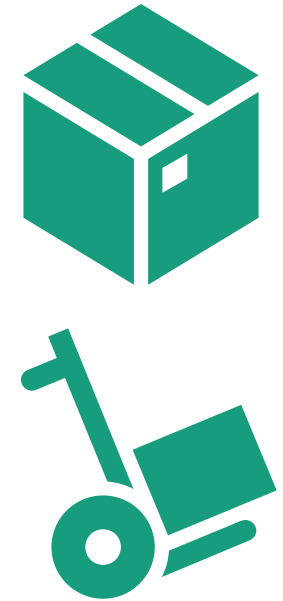
- Standard/Protokoll wie KI-Anwendungen (z.B. KI-Agenten) sich an Datenquellen bzw. andere Software andocken können
- Davor war jede Integration eine Sonder-/Einzellösung
- Mit MCP: Einheitliche Sprache zwischen KI-Agent und anderer Software
- Analogie: MCP ist wie USB für KI



Deployment von Sprachmodellen

Unterscheidungsmerkmal

- Gehostete Vs. On-premise Modelle
 - Gehostete LLMs laufen auf den Servern der Anbieter selber
 - On-premise LLMs laufen auf eigener Infrastruktur
 - Datenschutz/Betriebsgeheimnis-Bedenken
- GPT, CoPilot, Google Gemma, ...
 - Laufen bei den Anbietern selbst
 - Alle Daten, Anfragen, PDFs, Bilder, ... werden zu den Anbietern übertragen
- Meta Llama, Deepseek, Mistral, Qwen, Gemma...
 - Können auf eigenen Servern betrieben werden und so interner Software und Mitarbeitern zur Verfügung gestellt werden.



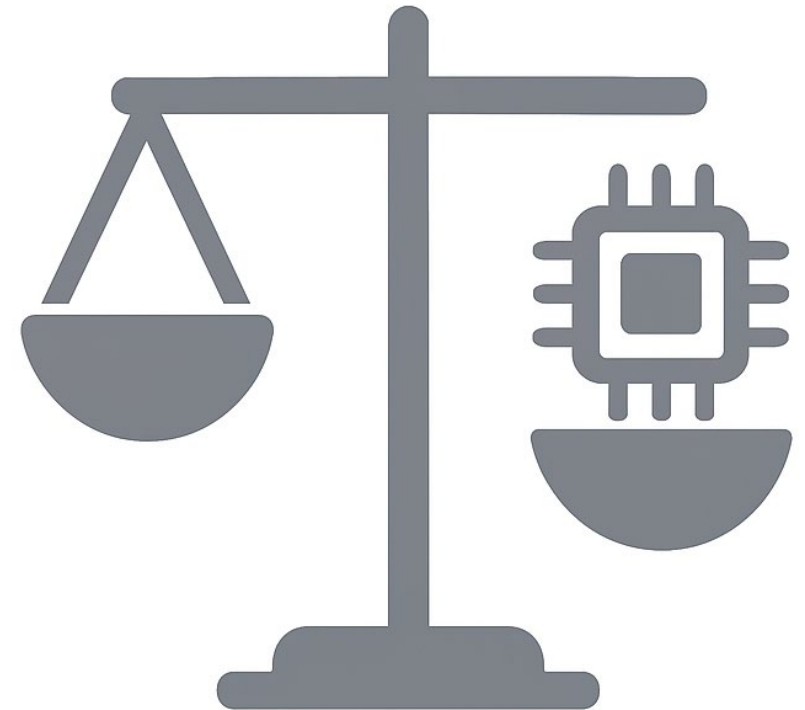
Worauf man beim Einsatz von LLMs achten sollte

Wichtige Entscheidungsbereiche im Umgang mit LLMs

- LLMs wirken einfach – doch ihr Einsatz ist komplex
- Wer LLMs nutzen will, muss Risiken und Rahmenbedingungen aktiv mitdenken

Zentrale Aspekte

- Datenschutz und Sicherheit
 - Welche Daten kommen zum Einsatz?
 - Was passiert mit den eingegebenen Daten?
- Recht und Lizenzierung
 - Wer haftet bei Fehlern?
 - Welche Nutzung ist erlaubt?
- Bias und Fairness
 - Welche Vorurteile reproduzieren Modelle unbemerkt?



Anwendungsfelder

Wissens- & Dokumentationsmanagement

- Intelligente Suchmaschinen für Produktionsunterlagen (Handbücher, Sicherheitsstandards, ...)
- Automatisierte Zusammenfassungen von Dokumenten (Berichte, Meeting-Transkripte, ...)
- FAQ-Bots für Mitarbeitende in der Produktion (Maschinenfehler, Qualitätsstandards, Sicherheitsregeln)
- Geführtes Troubleshooting bei Maschinenstillständen

Qualität & Compliance

- Unterstützung bei Fehler- und Beschwerdedokumentation (Strukturierung von Freitext)
- Automatische Checklistenprüfung gegen vordefinierte Anforderungen
- Erstellung von Auditberichten und Management-Zusammenfassungen



Anwendungsfelder

Wartung & Service

- Virtuelle Assistenz für Reparaturschritte (Schritt-für-Schritt-Anleitungen aus Handbüchern, geführtes Troubleshooting bei Maschinenstillständen)
- Analyse von Service-Tickets (Kategorisierung, Lösungsvorschläge)
- Erstellung von Ersatzteil- und Wartungsanfragen aus Maschinendaten

Produktion & Shopfloor

- Übersetzung von Arbeitsanweisungen in mehrere Sprachen (internationale Teams)
- Speech-to-Text-Dokumentation für Maschinenfehler und Schichtberichte
- Multimodale Eingaben: Kombination von Text und Bildern (z. B. automatische Fehlerdokumentation)



Homepage-Chatbot für Interessenten

Produktivsystem

Ziel

- Automatisierte Auskunft über Produkte und Dienstleistungen für Interessenten

Herausforderungen

- Einbindung von Firmen-Informationen aus verschiedenen Quellen
- Angabe verwendeter Informationsquellen
- Integration in die Firmen-Homepage

Projekthinhalte

- Aufbereitung von Daten verschiedenster Formate
- Aufbau einer Wissensdatenbank
- Antwort-Generierung über Retrieval-Augmented-Generation (RAG)
- Implementierung als Backend-Webserver



*Chatbot-Einbindung auf der Leeb-
Homepage
(<https://www.leeb.at>)*



Success Story | Symvaro GmbH

KI-basierter E-Mail-Agent für automatisierte Zukunftsgespräche



symvaro

Fact Box

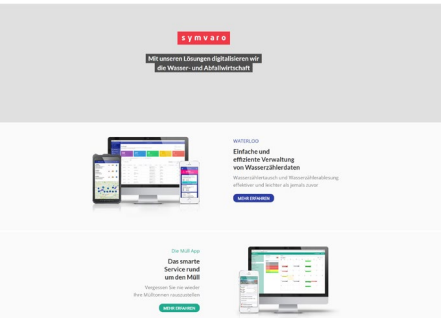
- **Kunde:** Symvaro GmbH
- **Projektdauer:** 3 Monate
- **Methoden:** • Large Action Models • KI-basiertes Agentensystem • Natural Language Processing • KI-basierte Textklassifikation
- **Ergebnisse:**
 - Funktionsfähiger E-Mail-Agenten-Prototyp für automatisiertes Führen von Zukunftsgesprächen via E-Mail

Projekthinhalte

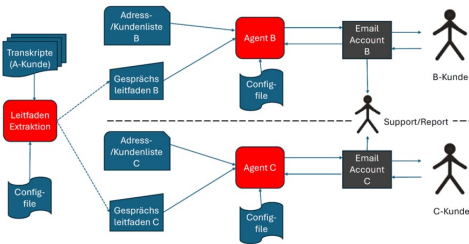
- Fraunhofer Austria KI4LIFE entwickelte ein KI-gestütztes Assistenzsystem, das Zukunftsgespräche erstmals automatisiert und skalierbar per E-Mail führt. Kern ist ein E-Mail-Agent, der aus realen Gesprächstranskripten strukturierte Leitfäden ableitet, personalisierte Startmails erstellt und automatisiert versendet. Antworten werden per Large Language Models analysiert und in drei Kategorien eingeteilt: beantworten, weiterleiten oder ignorieren. Über Konfigurationsdateien lässt sich die gesamte Kommunikation steuern; zusätzlich generiert das System wöchentliche Berichte mit relevanten Dialogen für interne Abteilungen. Der modulare Prototyp kann künftig um Kanäle wie Chat oder Voice erweitert werden und zeigt, wie KI den digitalen Kundendialog nicht nur automatisieren, sondern auch qualitativ verbessern kann.

Kundenstimme – Social Media Posting: »Wie lassen sich Kundenfeedback und Produktentwicklung noch enger verzweigen– effizient, skalierbar und datenbasiert? Fraunhofer Austria hat für die Symvaro GmbH ihr Konzept der „Zukunftsgespräche“ neu aufgesetzt und mit Künstlicher Intelligenz angereichert. Das Ergebnis: Ein KI-gestütztes Assistenzsystem, das Gespräche automatisiert per E-Mail führt, Antworten intelligent auswertet und relevante Informationen direkt an die richtigen Stellen im Unternehmen steuert«, **Symvaro, LinkedIn**

Details unter: [KI-basierter E-Mail-Agent für automatisierte Zukunftsgespräche für Symvaro](#)



© Symvaro GmbH





#allesaußergewöhnlich

BERND ZWATTENDORFER

USE CASE PRAXIS QUERKOPF

MIT SPRACHMODELLEN ZU MEHR EFFIZIENZ IM SOZIALBEREICH

26.11.2025

WER SIND WIR?

- 2014 als Einzelunternehmen gegründet
- Unternehmen im Sozialbereich
- ca. 80 Mitarbeiter:innen
- 4 Standorte
- Multikulturelles & multiprofessionelles Team

Praxis Querkopf



Stefanie Zwattendorfer

WAS TUN WIR?

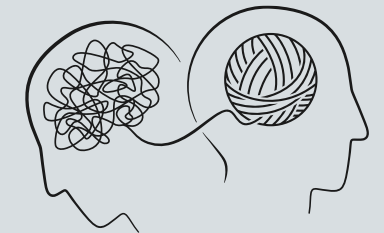
AMBULANTE FAMILIENINTENSIVBETREUUNG



PSYCHOLOGISCHE BERATUNG & DIAGNOSTIK



PSYCHOTHERAPIE



ERWACHSENENBILDUNG





ARBEITSALLTAG IN DER AMBULANTEN FAMILIENBETREUUNG

SELBSTORGANISATION

BETREUUNG UND BERATUNG

TELEFONIEREN

AUTO FAHREN

DOKUMENTATION



DOKUMENTATION ALS HERAUSFORDERUNG IM SOZIALBEREICH

WÖCHENTLICHE UPDATES

an fallführende Sozialarbeiter:innen

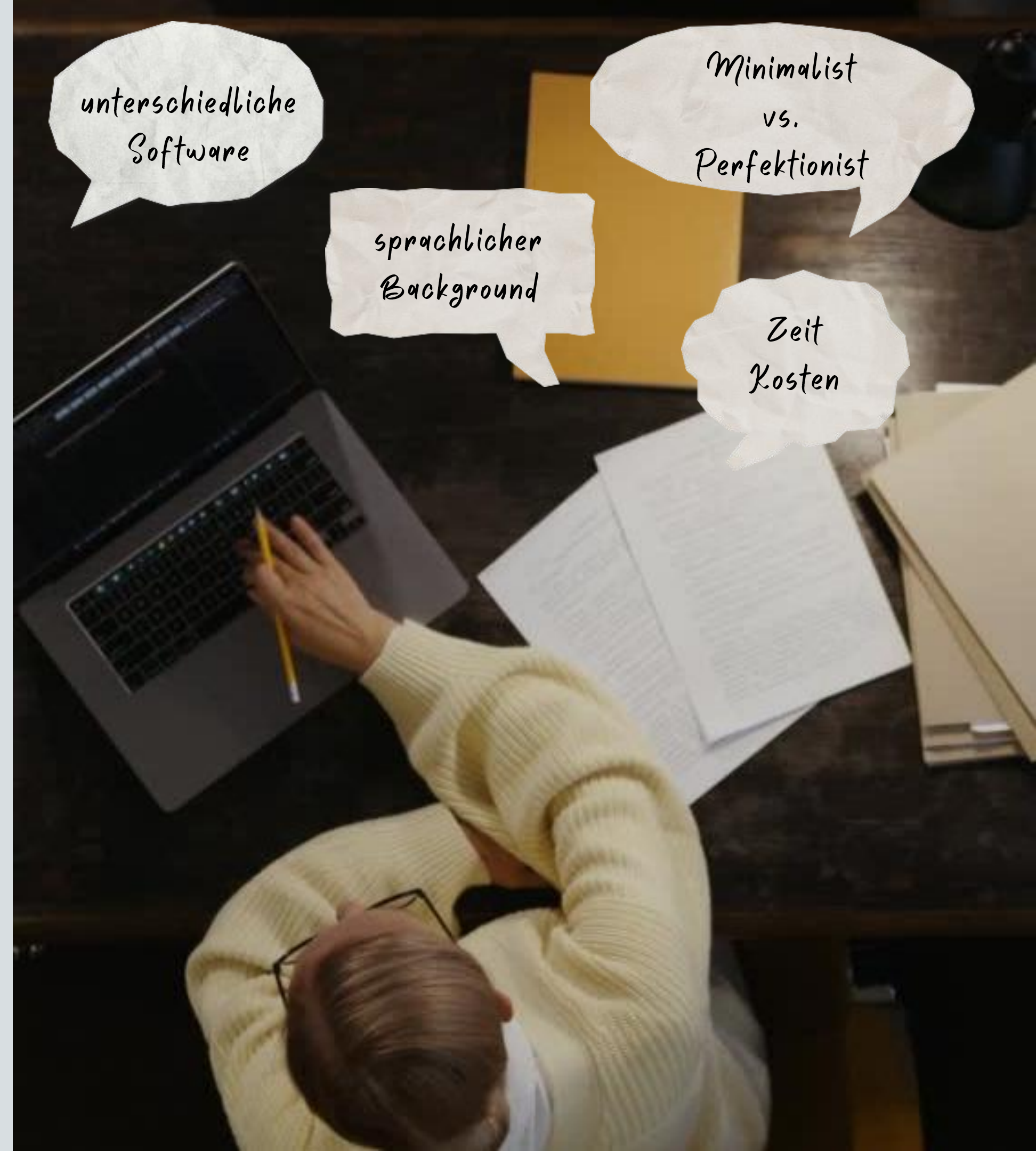
MONATLICHE BETREUUNGSDOKUMENTATION

mit Tätigkeitsbericht pro Termin

EINGANGSBERICHT SOZIALE DIAGNOSTIK

HALBJAHRESBERICHT

ABSCHLUSSBERICHT



IDEE

OPTIMIERUNG DER SOZIALEN DIAGNOSTIK MITTELS **KI - UNTERSTÜTZUNG**

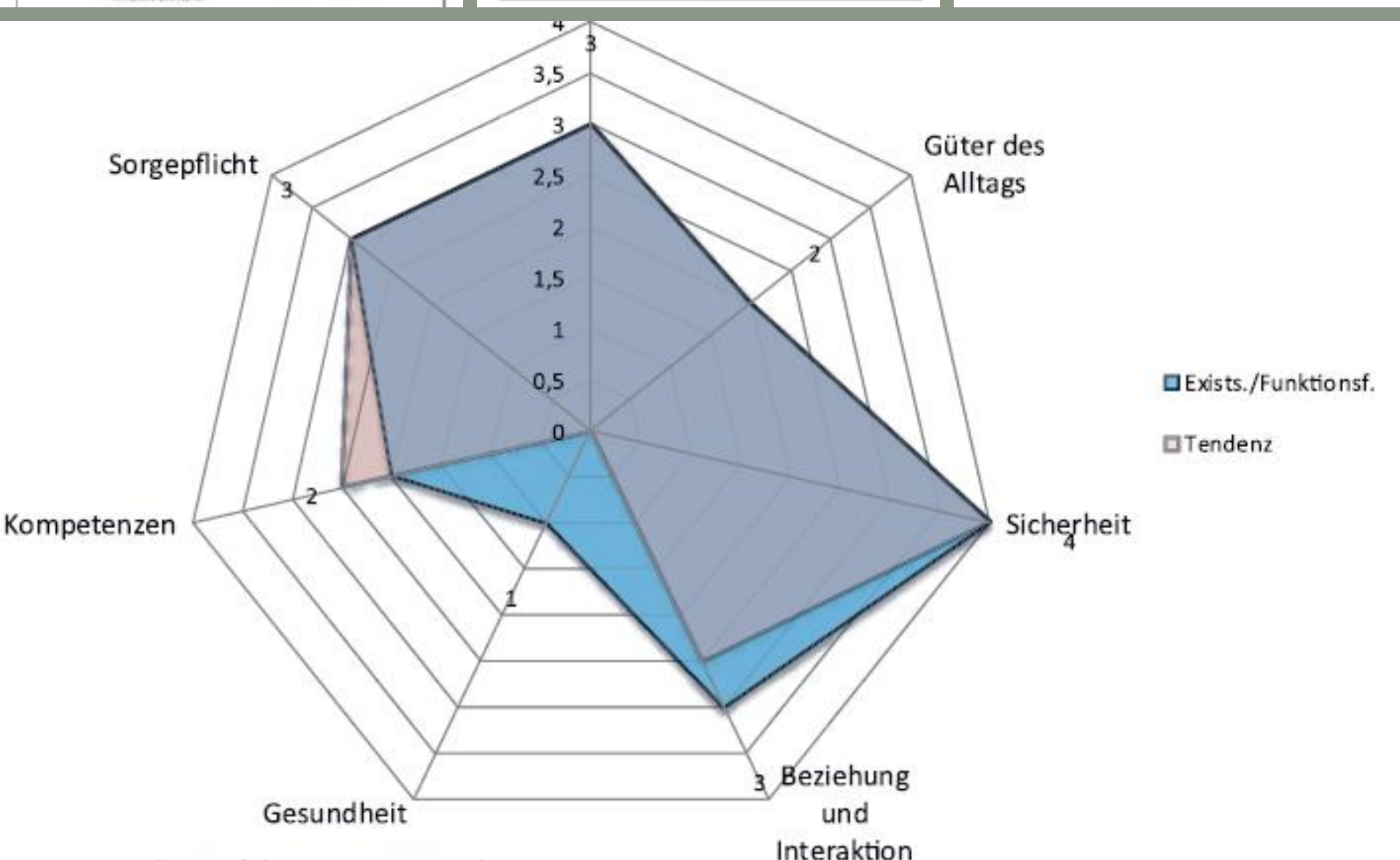
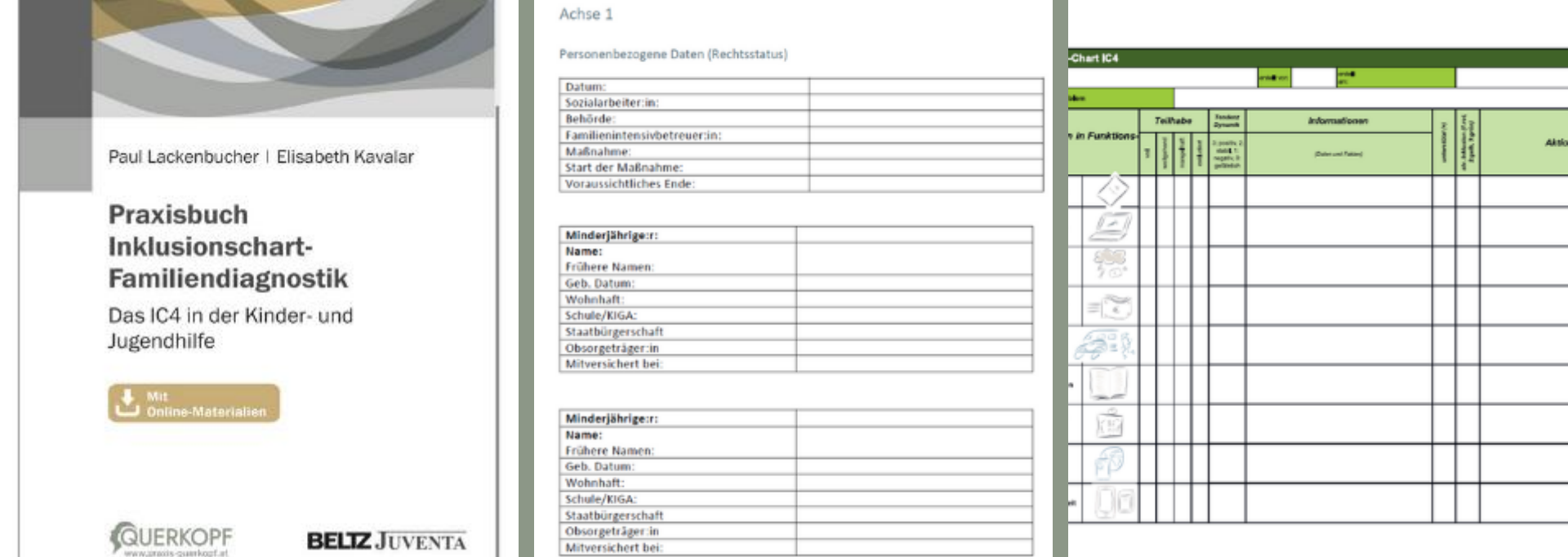
KI4LIFE KMU Digitalisierungs-Scheck

Fragestellung: Ist es möglich und wie gut funktioniert es?

Wunsch: Effizienz Steigerung

- weniger Zeitaufwand und Ressourcensteigerung (Speech-to-Text, Bulled Points, Ignorieren von Rechtschreibung- und Grammatik)
- geringere Kosten
- mehr Zeit für die Betreuung





Auch weitere für die Hilfeplanung relevante Daten können auf einen Blick abgelesen werden. So ist die derzeitige Situation im Bereich *Güter des Alltags* als mangelhaft eingestuft worden. Ebenso wird im Säulendiagramm (links unten) ersichtlich, dass hier derzeit noch keinerlei Unterstützungsleistung besteht

WAS IST SOZIALE DIAGNOSTIK UND WIE FUNKTIONIERT SIE?

Systematische Erhebung und Bewertung sozialer Daten (z.B. Finanzen, Gesundheit, Wohnen, etc.)

Ziel: Notwendigkeit, Dringlichkeit und Intensität sozialarbeiterischer Interventionen ableiten

ABLAUF

- Erhebung mittels mündlichen Fragebogens
- Übertragung in ein Excel-Sheet und Auswertung
- Berichts-Erstellung (Word)
- Weiterleitung und Archivierung (Online Tool)

UMSETZUNG DER SOZIALEN DIAGNOSTIK MITTELS KI-UNTERSTÜTZUNG

Auswahl des LLM-Models

Pre-Processing:

- Bearbeitung Excel-Sheet für automatisiertes Einlesen
- Datenschutz-Optimierung

KI-Input:

- Spezifisches Prompt
- Bericht-Template, existierende Berichte (anonymisiert), Fachbuch Soziale Diagnostik, Excel-Sheet

Output:

- KI-generierter Bericht





HERAUSFORDERUNGEN

- Uneinheitliche Excel-Strukturen
- Wo werden die Daten verarbeitet (lokales LLM)
- sensible Daten (Anonymisierung bzw. Pseudonymisierung)
- Prompting (Kontext, Rolle, Aufgabe etc.)



LESSONS LEARNED

Es funktioniert teilweise sehr gut, **ABER**

- Nicht zu unterschätzender Aufwand für das Pre-/Post-Processing.
- Sensible Daten und KI sind immer noch eine offene Baustelle.
- Unterschiedliche Ergebnis-Qualität bei gleichen Prompts.
- Evaluierung der Ergebnisse: Wurden Fakten verändert oder wichtige Schlussfolgerungen weggelassen? (Mensch vs. Maschine)
- Dokument-Freigabe: Qualitätskontrolle vs. Bequemlichkeit
- Tagge ich das Dokument als KI-unterstützt -> Was denkt der Empfänger?
- Wie binde ich KI-Unterstützung in ein bestehendes (Dokumentations-)System ein, das ich nicht selbst verändern kann?



„KÜNSTLICHE INTELLIGENZ
KANN DATEN ORDNEN.
NUR EMPATHIE KANN SIE
VERSTEHEN.“

#allesaußergewöhnlich

www.querkopf-zentrum.at
www.praxis-querkopf.at

Data Driven Dialog, 26. November 2025, Lakeside Park Klagenfurt

Optische Qualitätskontrolle – Live Demo

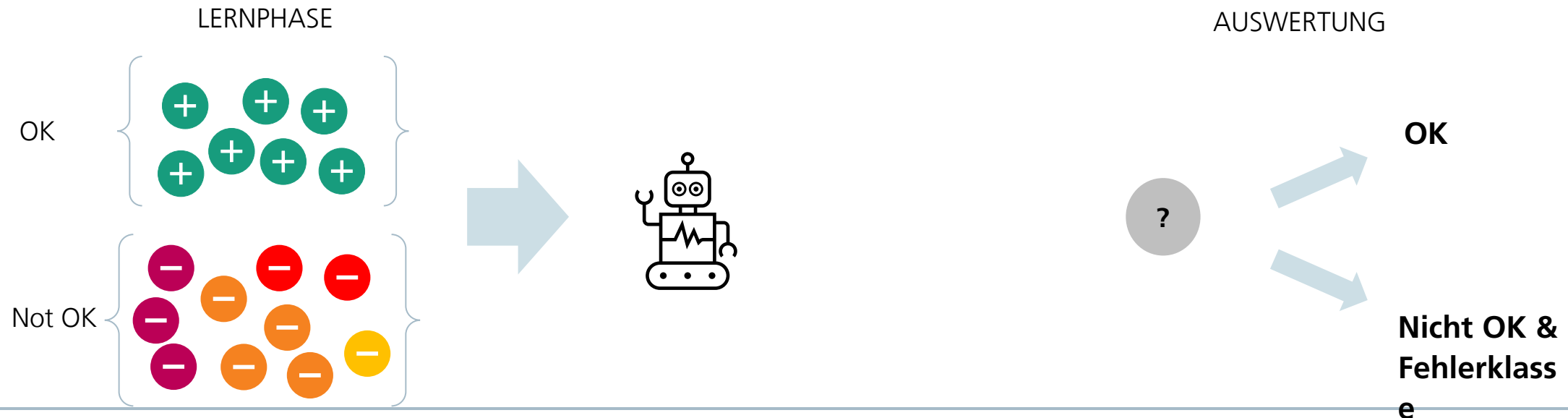
DI Dr. Christina Petschnigg
Gruppenleiterin Fraunhofer Austria Research GmbH

Bildbasierte Qualitätskontrolle

Einsatz von Künstlicher Intelligenz

Überwachtes Lernen

- »Gut«- und »Schlecht«-Beispiele sind notwendig
- Großer Aufwand, um ein Modell von Grund auf zu trainieren
- Erkennung vordefinierter Fehlermuster

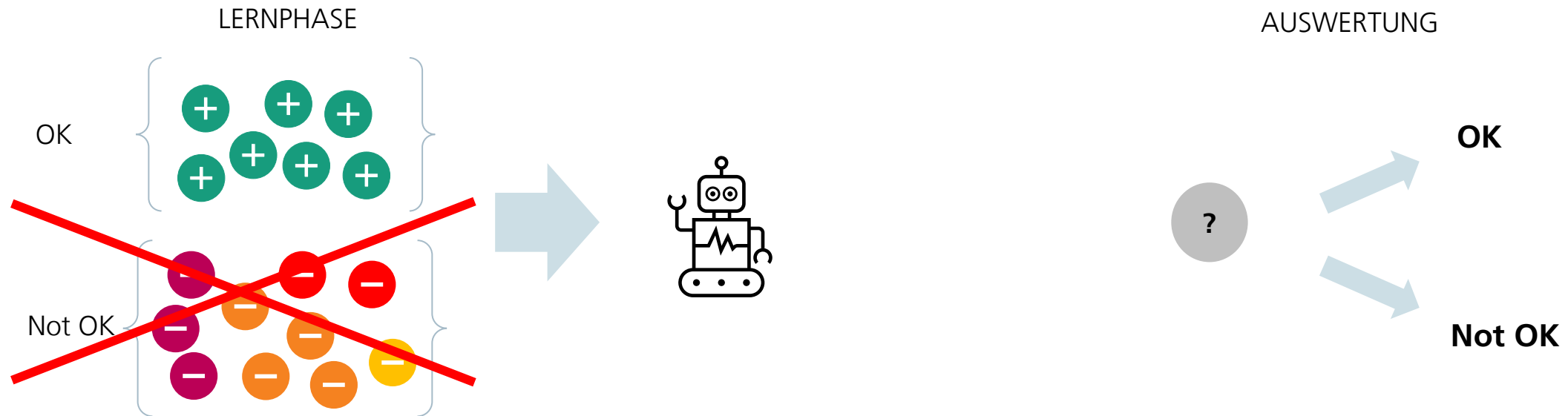


Bildbasierte Qualitätskontrolle

Einsatz von Künstlicher Intelligenz

Anomalieerkennung

- Nur wenige »Gut«-Beispiele sind notwendig
- Nutzen vortrainierter KI-Modelle
- Erkennung bekannter und unbekannter Fehlermuster



Bildbasierte Qualitätskontrolle in der Anwendung

Beispiel Lebensmittelindustrie

- KI-Modell mit lediglich 38 »Gut«-Beispielen trainiert
- Training: ca. 1 min
- Prüfung: ca. 33 ms
- Standardhardware zur Bildaufnahme (Webcam)
- Modellauswertung benötigt keine spezielle Hardware



Bildbasierte Qualitätskontrolle in der Anwendung

Beispiel Lebensmittelindustrie

→ **LIVE DEMO**

Kontakt Daten:



DI Dr. Christina Petschnigg

Projektleiterin

Fraunhofer Austria Research GmbH

Fraunhofer Innovationszentrum für Digitalisierung und
Künstliche Intelligenz KI4LIFE

Lakeside B13a | 9020 Klagenfurt | Austria

Mobil: +43 676 888 61 829

E-Mail: christina.petschnigg@fraunhofer.at

Bildbasierte Qualitätskontrolle in der Anwendung

Beispiel Lebensmittelindustrie

→ **LIVE DEMO**

Data Driven Dialog, 26. November 2025, Lakeside Park Klagenfurt

Willkommen bei » Do(KI)t «

Ing. Dr. Marco A. Hudelist
Gruppenleiter Fraunhofer Austria Research GmbH

→ LIVE DEMO

Willkommen!

Willkommen bei >>Do(KI)t<<!

Do(KI)t erzeugt eine individuelle Dokumentation für Ihren neuen benutzerdefinierten Computer. Dabei wird im Hintergrund Künstliche Intelligenz in Form von Large Language Models genutzt, um aus einer Komponentenliste, passenden Handbüchern, sowie Ihrem tatsächlichen Montagevorgang eine vollumfängliche Dokumentation zu generieren.

Komponentenliste



Montageaufzeichnung



Dokumentation erzeugen

Kontaktdaten:



Ing. Dr. Marco A. Hudelist

Projektleiter

Fraunhofer Austria Research GmbH

Fraunhofer Innovationszentrum für Digitalisierung und
Künstliche Intelligenz KI4LIFE

Lakeside B13a | 9020 Klagenfurt | Austria

Mobil: +43 676 888 61 810

E-Mail: marco.hudelist@fraunhofer.at



DIGITALISIERUNG IST

EASY

Folgen Sie
uns



für mehr Informationen zu Veranstaltungen,
Digitalisierung und Innovation.

www.dih-sued.at

Fraunhofer Austria Research GmbH
Center für Data Driven Design
Innovationszentrum für Digitalisierung
und Künstliche Intelligenz KI4LIFE

Lakeside B13a | 9020 Klagenfurt
Tel: +43 676 888 61 820
office.ki4life@fraunhofer.at
www.fraunhofer.at



Fraunhofer
AUSTRIA