



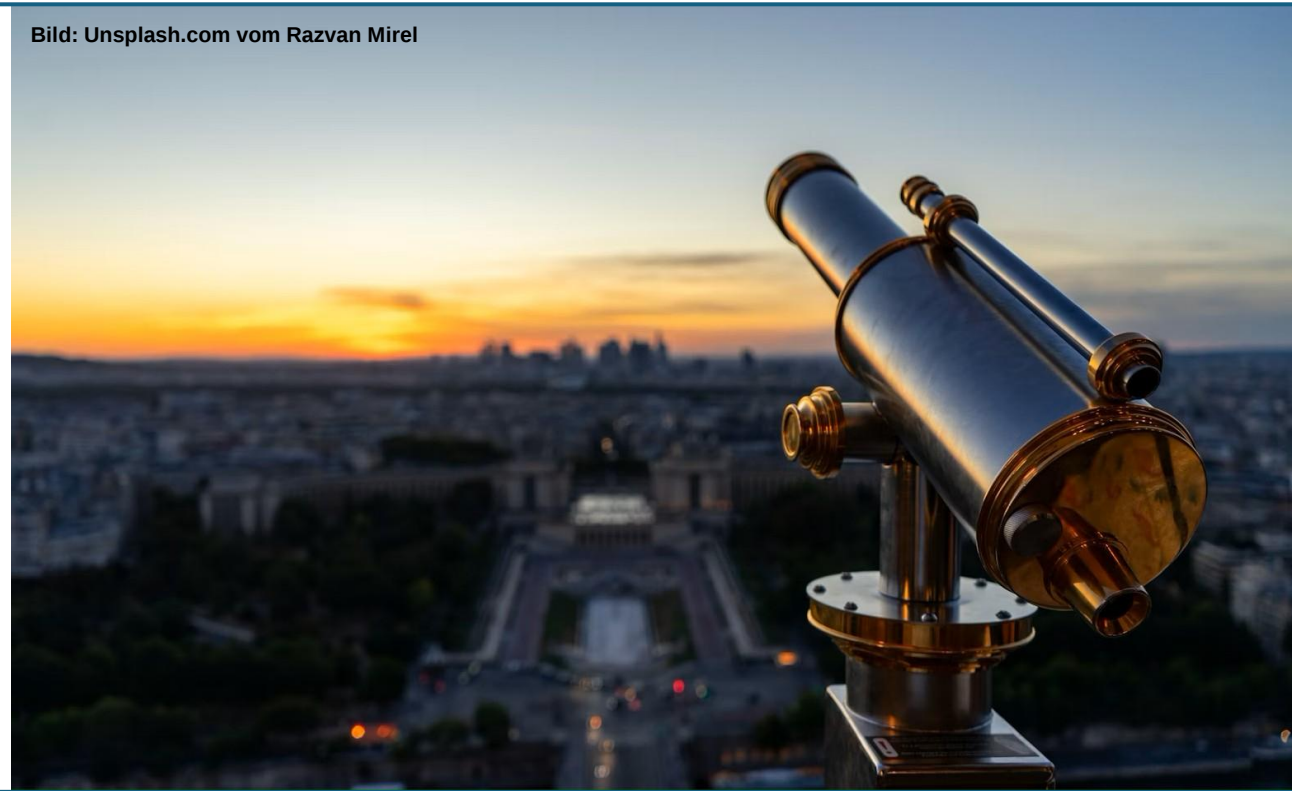
Bundesanstalt für Arbeitsschutz
und Arbeitsmedizin

- KI im Arbeitsschutz – Forschung zum unterstützenden Einsatz in der Gefährdungsbeurteilung
-

Inhaltsverzeichnis

- Übersicht: Künstliche Intelligenz
- Sensorassistierte Gefährdungsbeurteilung
- Große Sprachmodelle zur Unterstützung der Gefährdungsbeurteilung

Bild: Unsplash.com vom Razvan Mirel

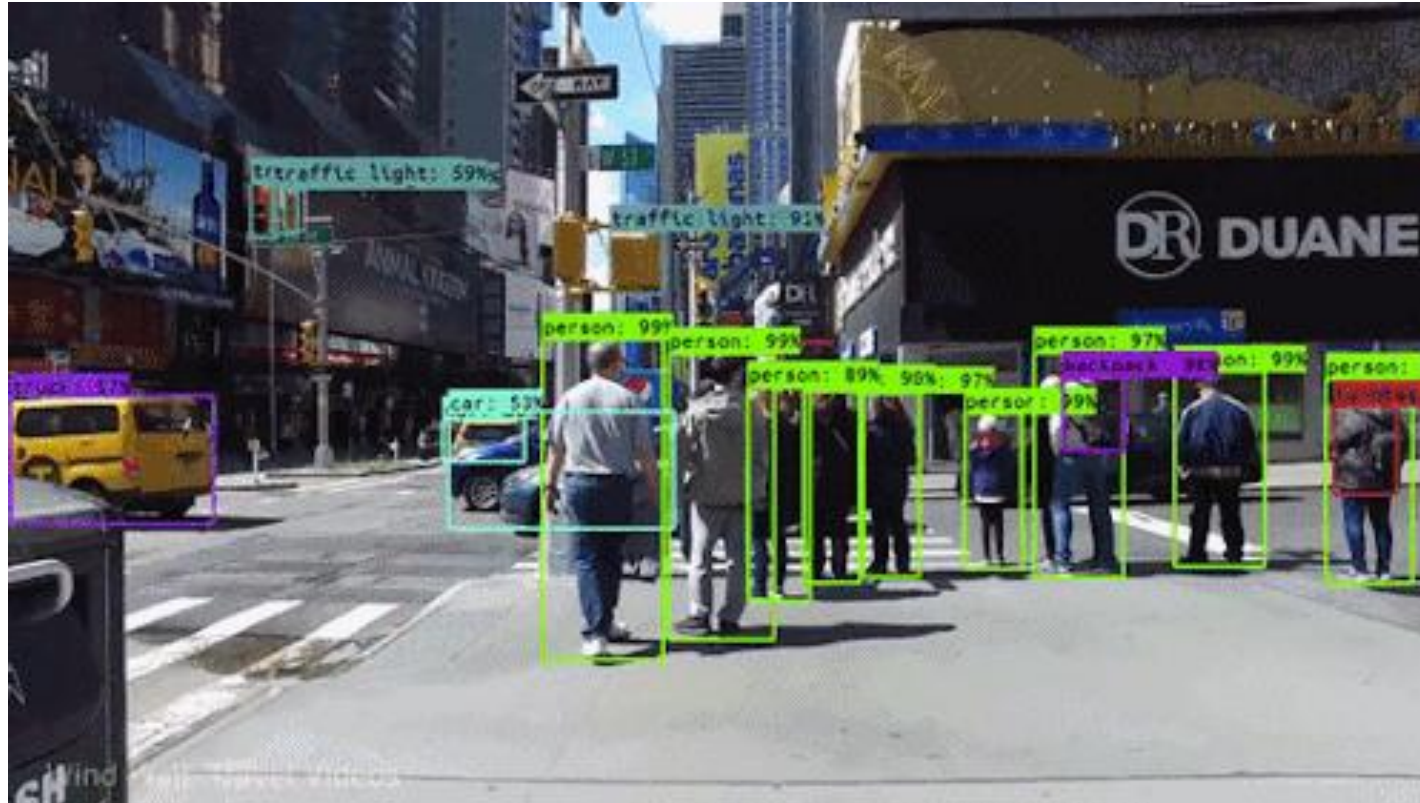


■ Künstliche Intelligenz – Was ist das?

■ Künstliche Intelligenz ist medienwirksam



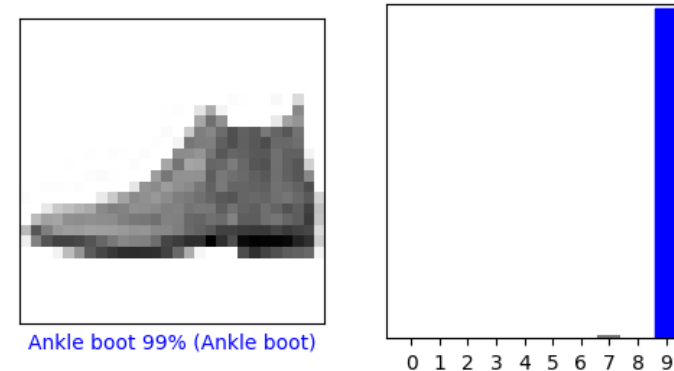
■ Künstliche Intelligenz ist medienwirksam



Künstliche Intelligenz - Klassifikationsaufgaben

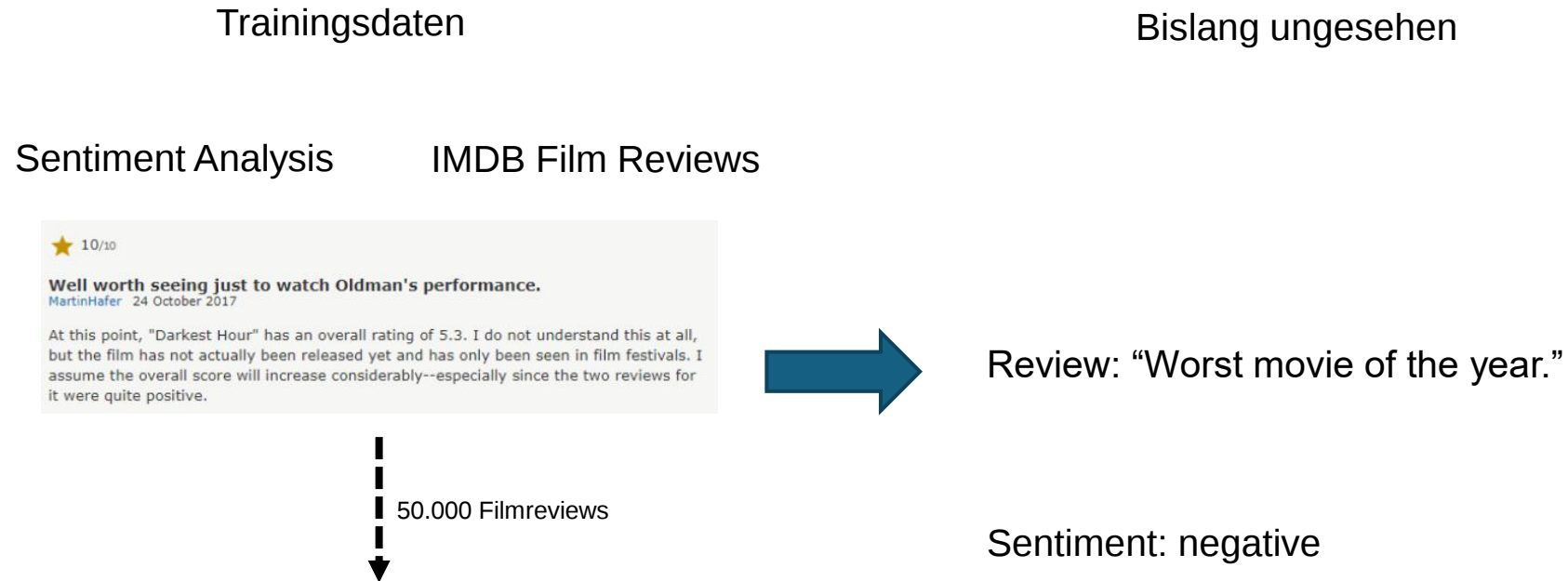


Bislang ungesehen

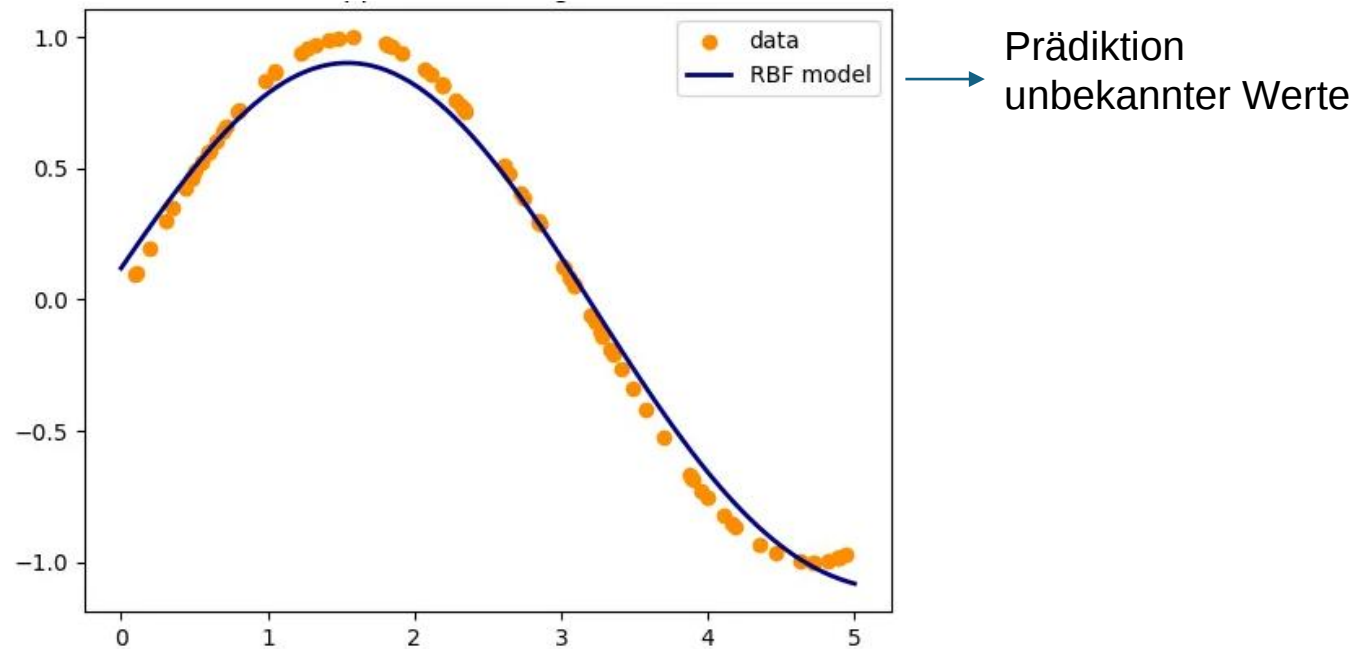




Künstliche Intelligenz - Klassifikationsaufgaben



Künstliche Intelligenz - Regressionsaufgaben





Künstliche Intelligenz – Chatbots, Große Sprachmodelle

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

vielen Dank für Ihre Teilnahme an unserem Gesundheitsworkshop. Ihr Feedback ist uns sehr wichtig, um zukünftige Angebote zu verbessern. Bitte nehmen Sie sich einige Minuten Zeit, um die

Skizziere einen Fragenkatalog für eine Zufriedenheitsbefragung für Teilnehmerinnen und Teilnehmer eines Gesundheitsworkshops am Arbeitsplatz.

Achte darauf, dass die Fragen so wenig suggestiv wie möglich formuliert sind, damit die Antworten nicht unbeabsichtigt beeinflusst werden.

4. **Wie hilfreich empfanden Sie die praktischen Übungen/Anwendungen im Workshop?**

(Überhaupt nicht hilfreich – Weniger hilfreich – Neutral – Hilfreich – Sehr hilfreich)

(Diese Frage anpassen, falls keine praktischen Übungen stattfanden)

5. **Wie beurteilen Sie die Kompetenz der Workshopleitung?**

(Sehr schlecht – Schlecht – Mittelmäßig – Gut – Sehr gut)

6. **Wie beurteilen Sie die Art und Weise, wie die Workshopleitung die Inhalte präsentiert hat?**

7. **Wie angemessen empfanden Sie die Dauer des Workshops?**

(Viel zu kurz – Etwas zu kurz – Genau richtig – Etwas zu lang – Viel zu lang)

8. **Wie beurteilen Sie die räumlichen Bedingungen/die Umgebung des Workshops?**

(Sehr schlecht – Schlecht – Mittelmäßig – Gut – Sehr gut)

(Diese Frage anpassen, falls der Workshop online stattfand)

9. **Welche Aspekte des Workshops haben Ihnen besonders gut gefallen?**

(Offene Frage)

10. **Welche Aspekte des Workshops könnten Ihrer Meinung nach verbessert werden?**

(Offene Frage)

11. **Inwieweit fühlen Sie sich durch den Workshop motiviert, gesundheitsfördernde Maßnahmen in Ihren Arbeitsalltag zu integrieren?**

(Überhaupt nicht motiviert – Weniger motiviert – Neutral – Motiviert – Sehr motiviert)

12. **Würden Sie einen ähnlichen Gesundheitsworkshop weiterempfehlen?**

(Auf keinen Fall – Eher nicht – Vielleicht – Eher ja – Auf jeden Fall)

13. **Haben Sie weitere Anmerkungen oder Vorschläge zum Workshop?**

(Offene Frage)

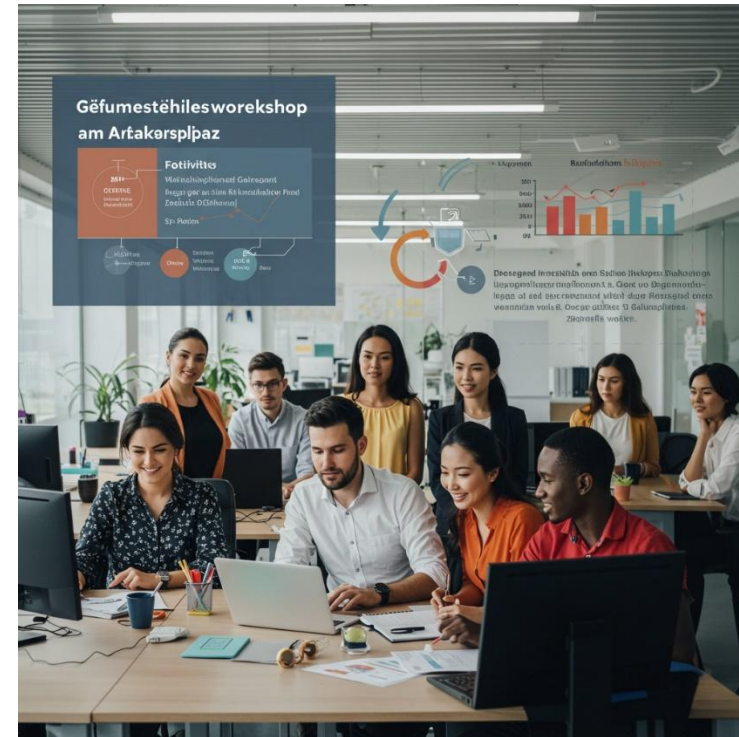
Abschluss:

Vielen Dank für Ihre Teilnahme und Ihr wertvolles Feedback!

Künstliche Intelligenz – Diffusionsmodelle

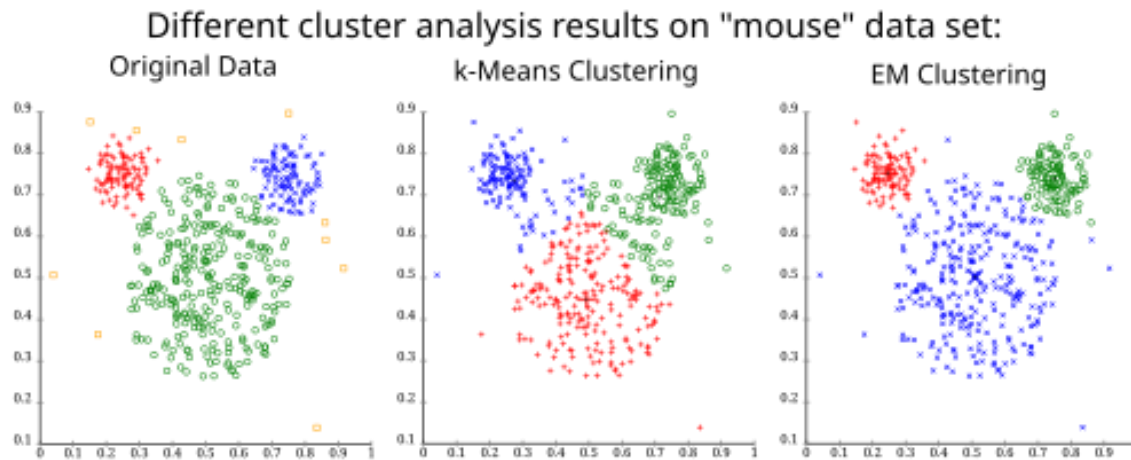
Prompt (Imagen 3 via Gemini Flash 2.5)

nutze nun deinen zugriff auf das tool imagen 3 um ein motivierendes bild für einen Bericht über die Programmsergebnisse des Gesundheitsworkshops zu illustrieren. Zielgruppe des Berichts sind Gesundheitspolitikerinnen und -politiker.

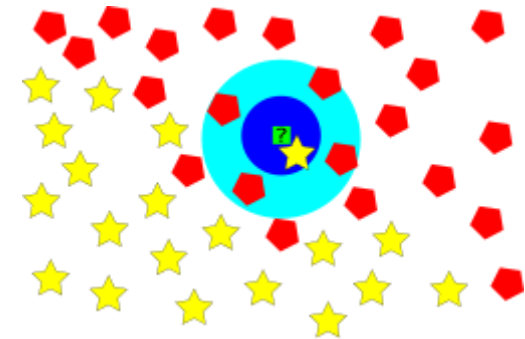


Künstliche Intelligenz – Clusteranalyse

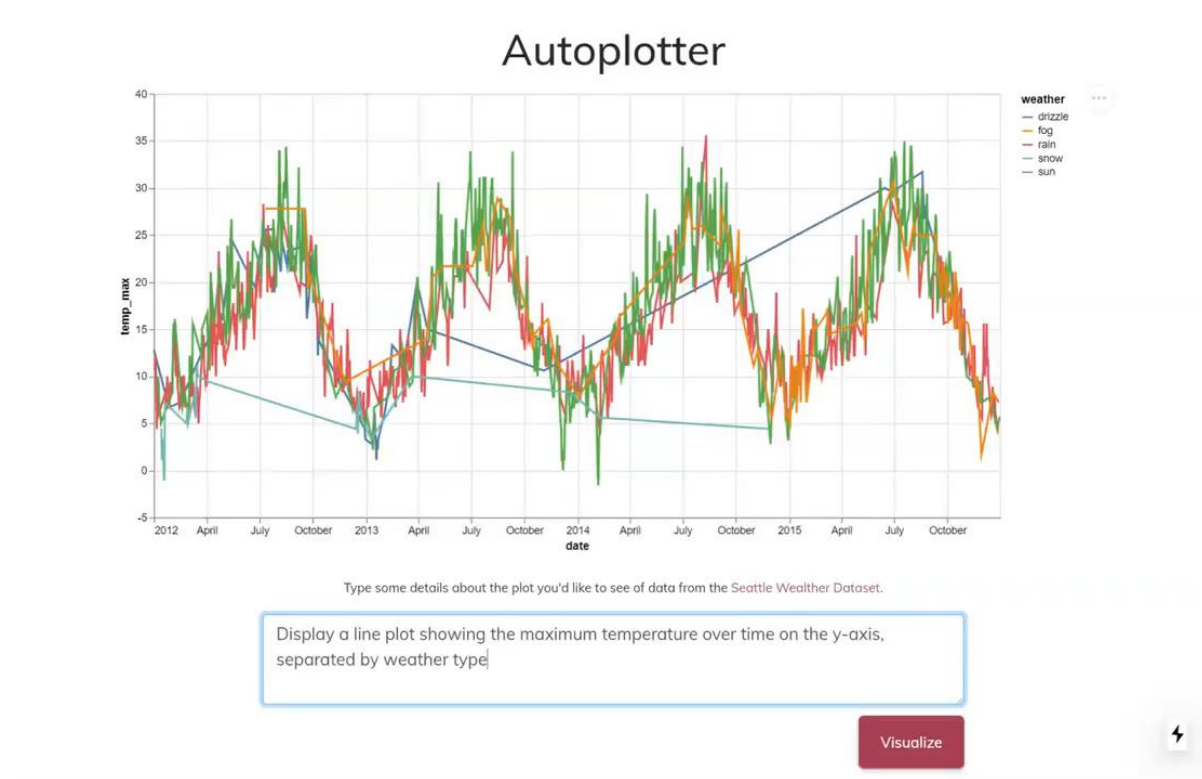
- Annahme: 3 Klassen in der Datenbasis
- Unüberwachtes Lernen



Knn-Algorithmus:
Klassifizierung anh.
bekannter Datenpunkte



Künstliche Intelligenz – Daten explorieren



■ Künstliche Intelligenz ist schwer zu definieren

Europäische Kommission, 2018 AI 'refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals'.

Minsky, 2009 Even though we don't yet understand how brains perform many mental skills, we can still work toward making machines that do the same or similar things. 'Artificial intelligence' is simply the name we give to that research.

KI im Kern

- Intelligentes umweltbezogenes Systemverhalten
- Streben nach menschlicher Intelligenz

■ Erschwerend: Der KI-Effekt - Ständige Grenzverschiebung



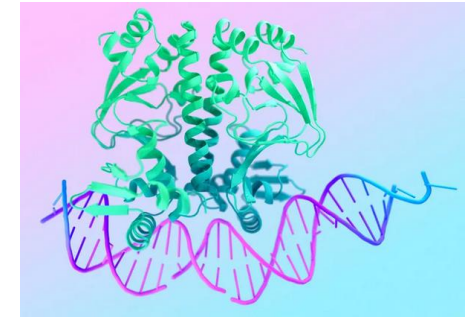
Schach



Go



Starcraft 2



AlphaFold 3

■ Einblick in die Funktionsweise von KI



Ronald Coase

Economist
Holder of the Nobel Memorial Prize in Economic Sciences

If you torture the data long enough, it will confess.

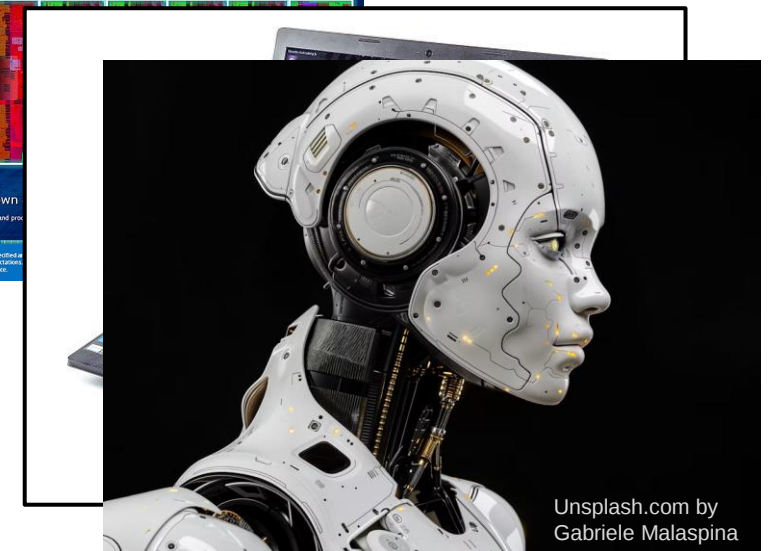
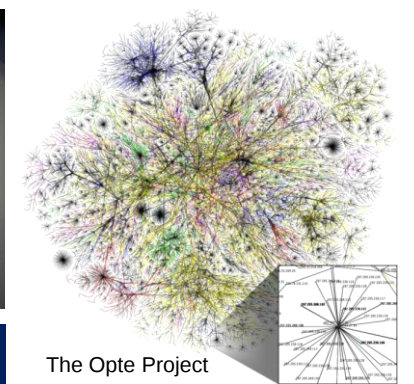
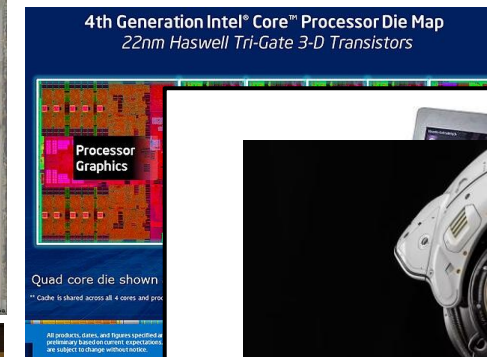
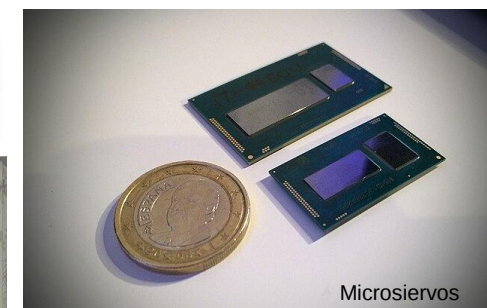
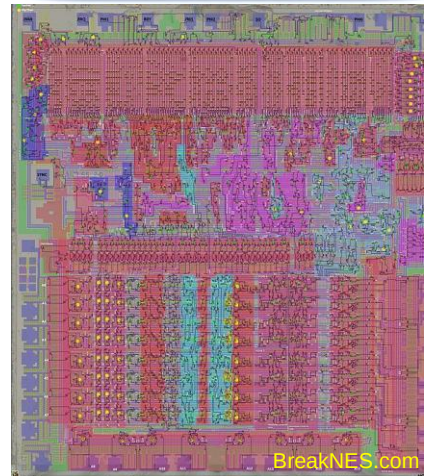
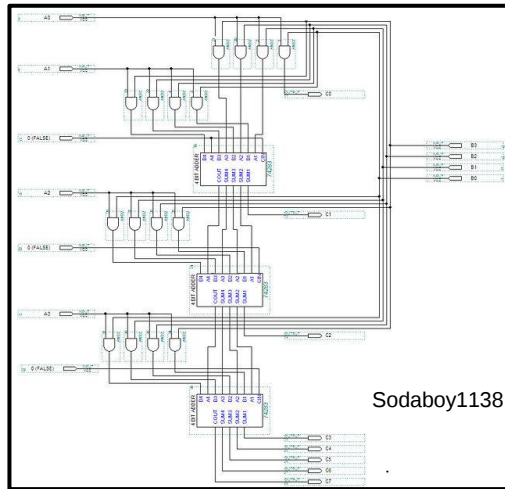
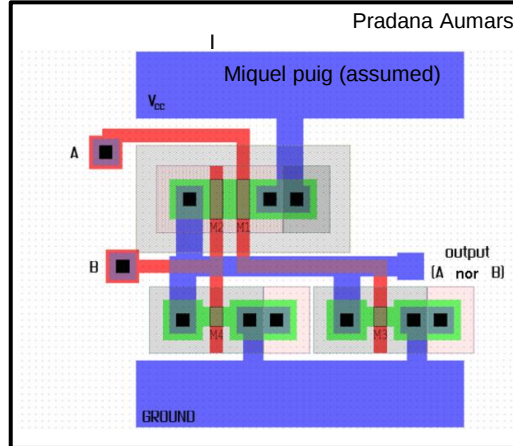
Informationen



Rodhullandemu

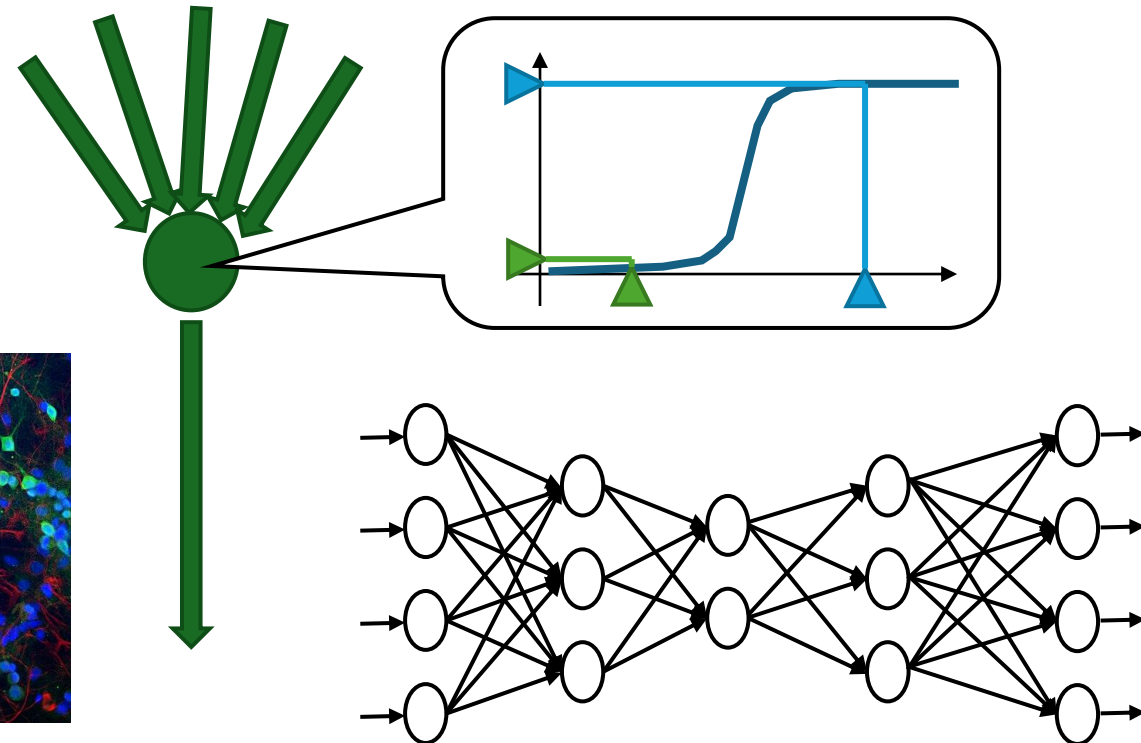
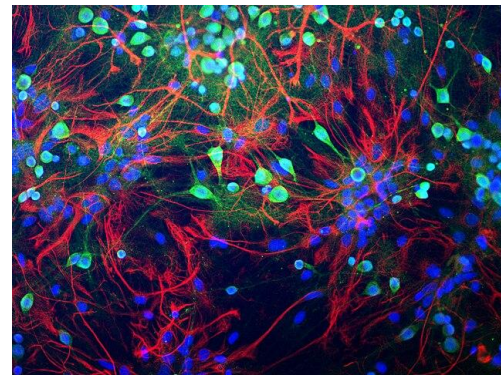
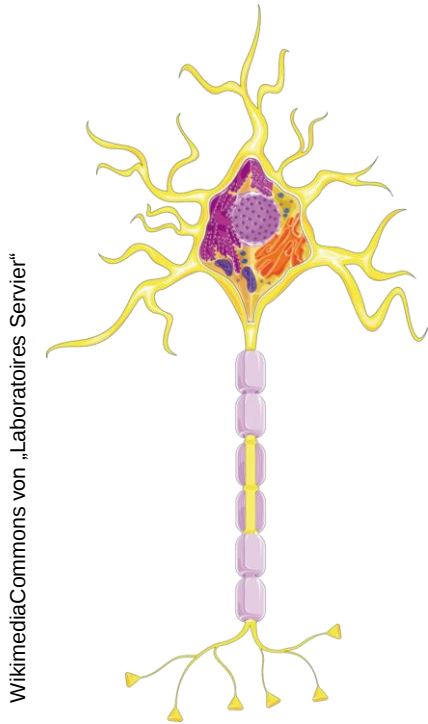


Власенко

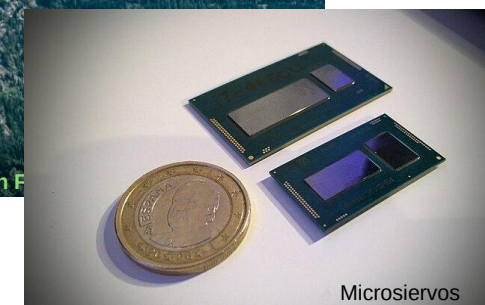
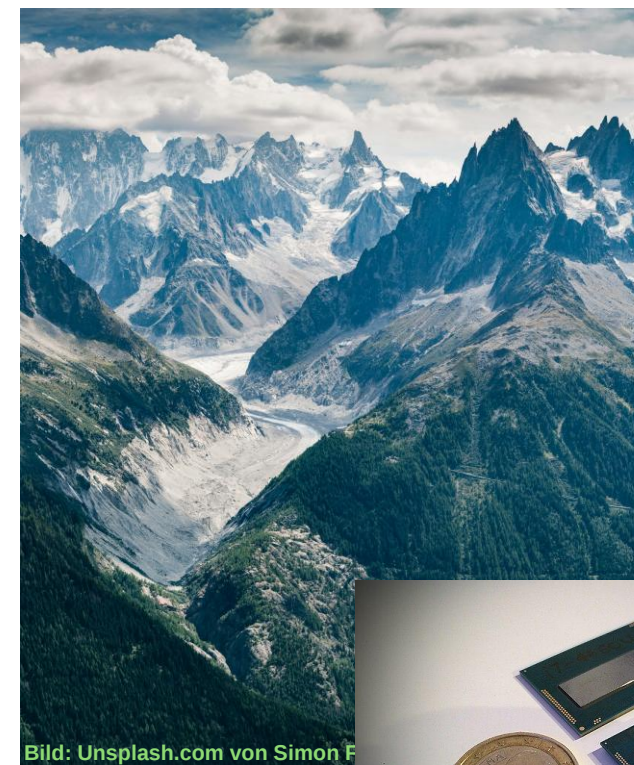
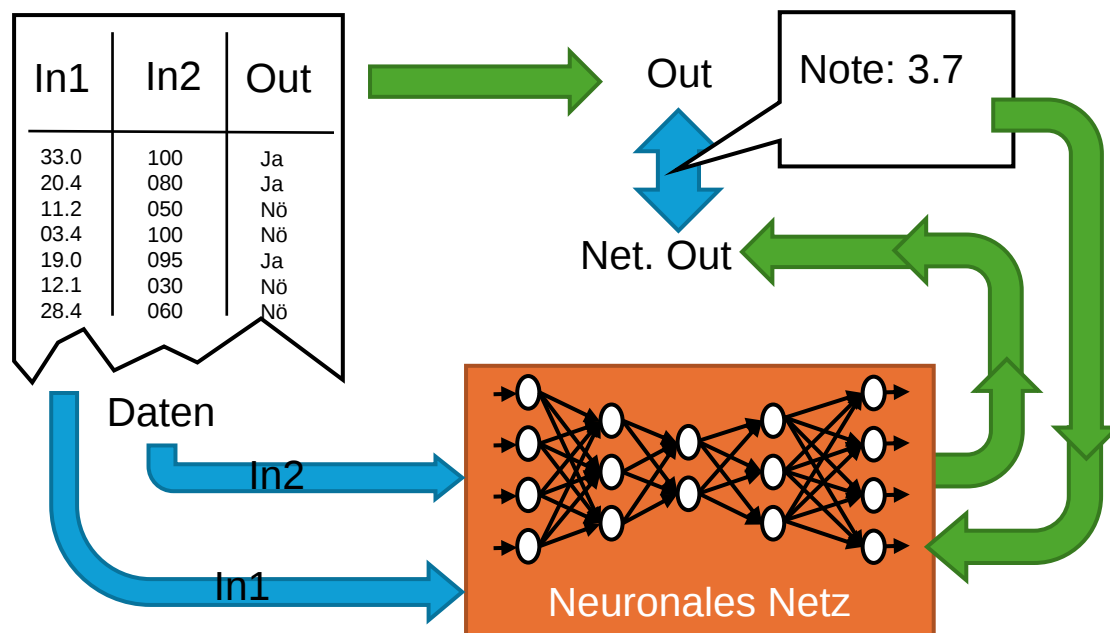


*Nexus von Yuval Noah Harari

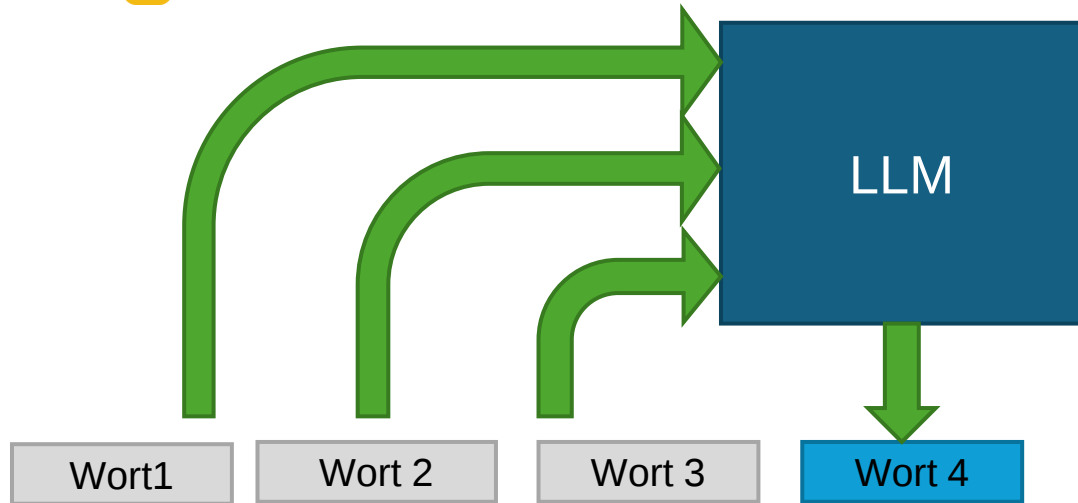
Beispiel: Künstliche Neuronale Netze



Training künstlicher Neuronaler Netze



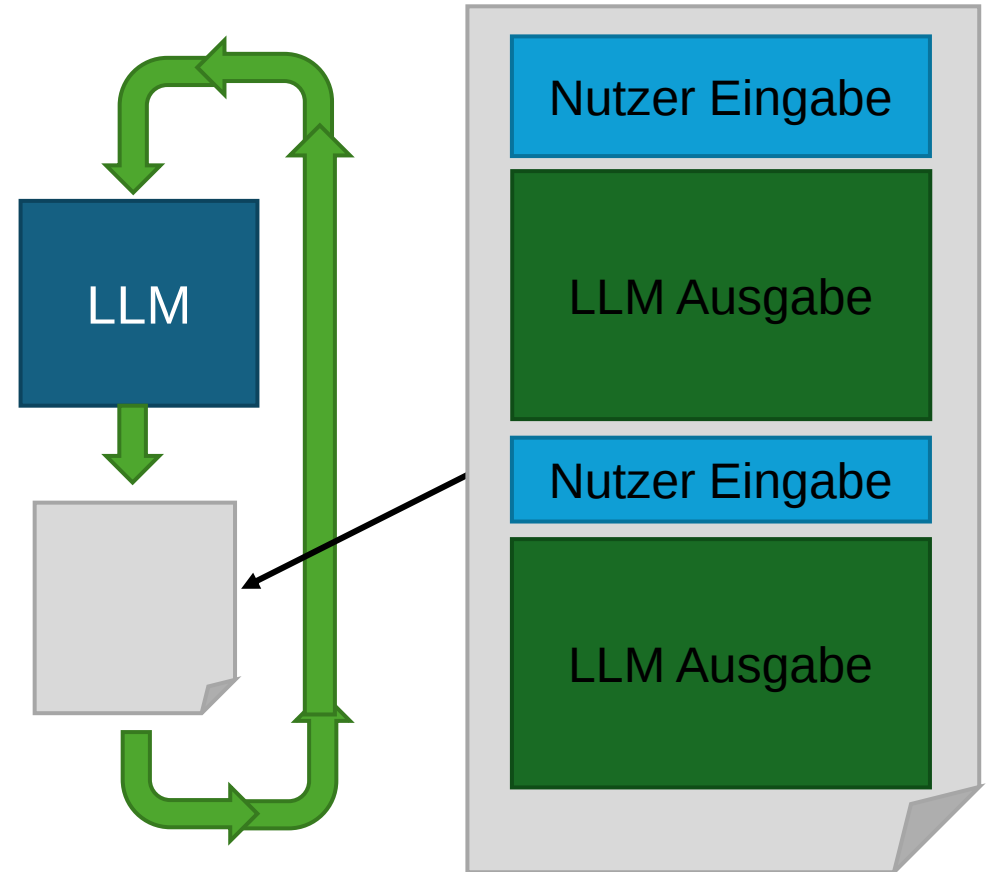
LLM



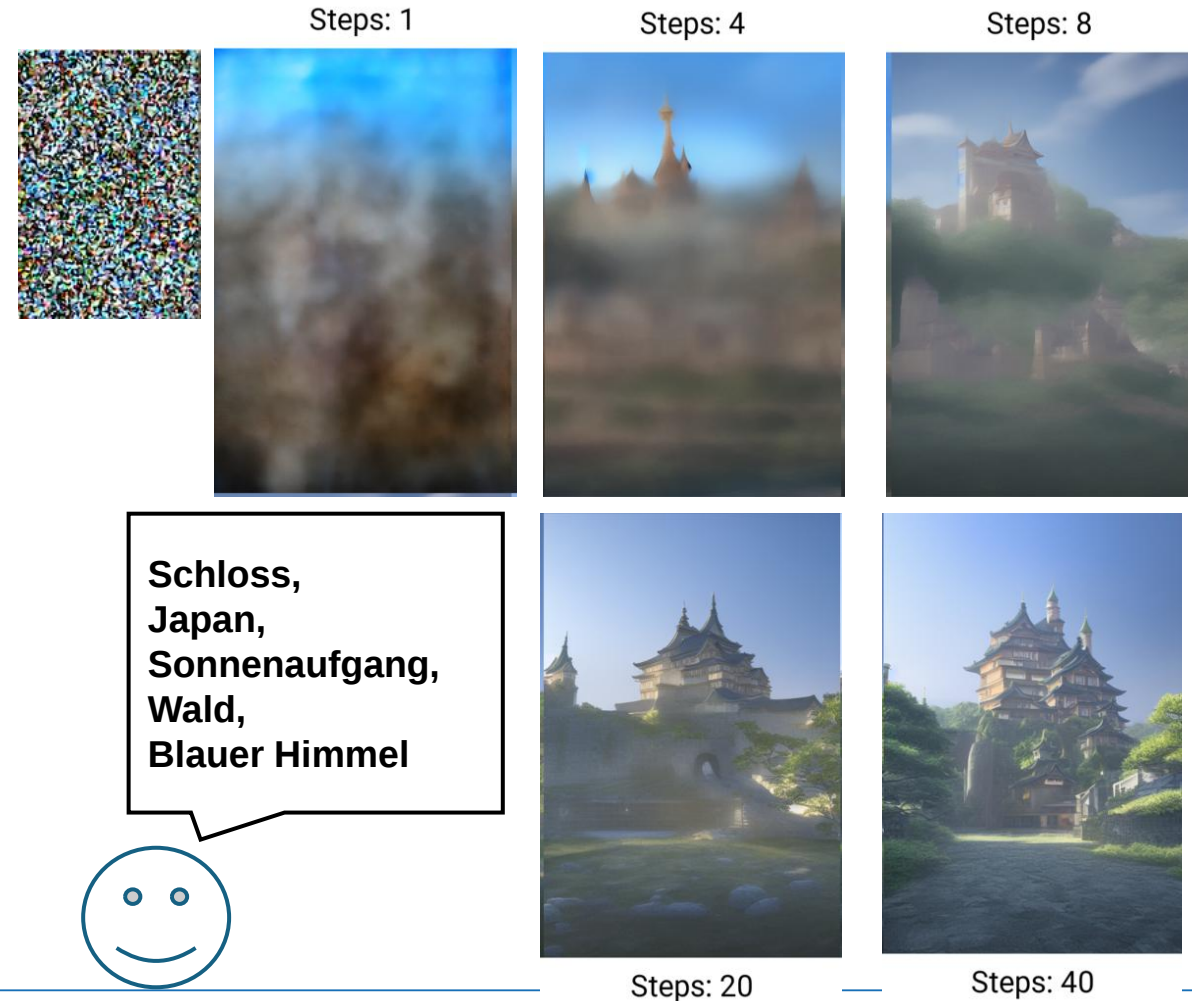
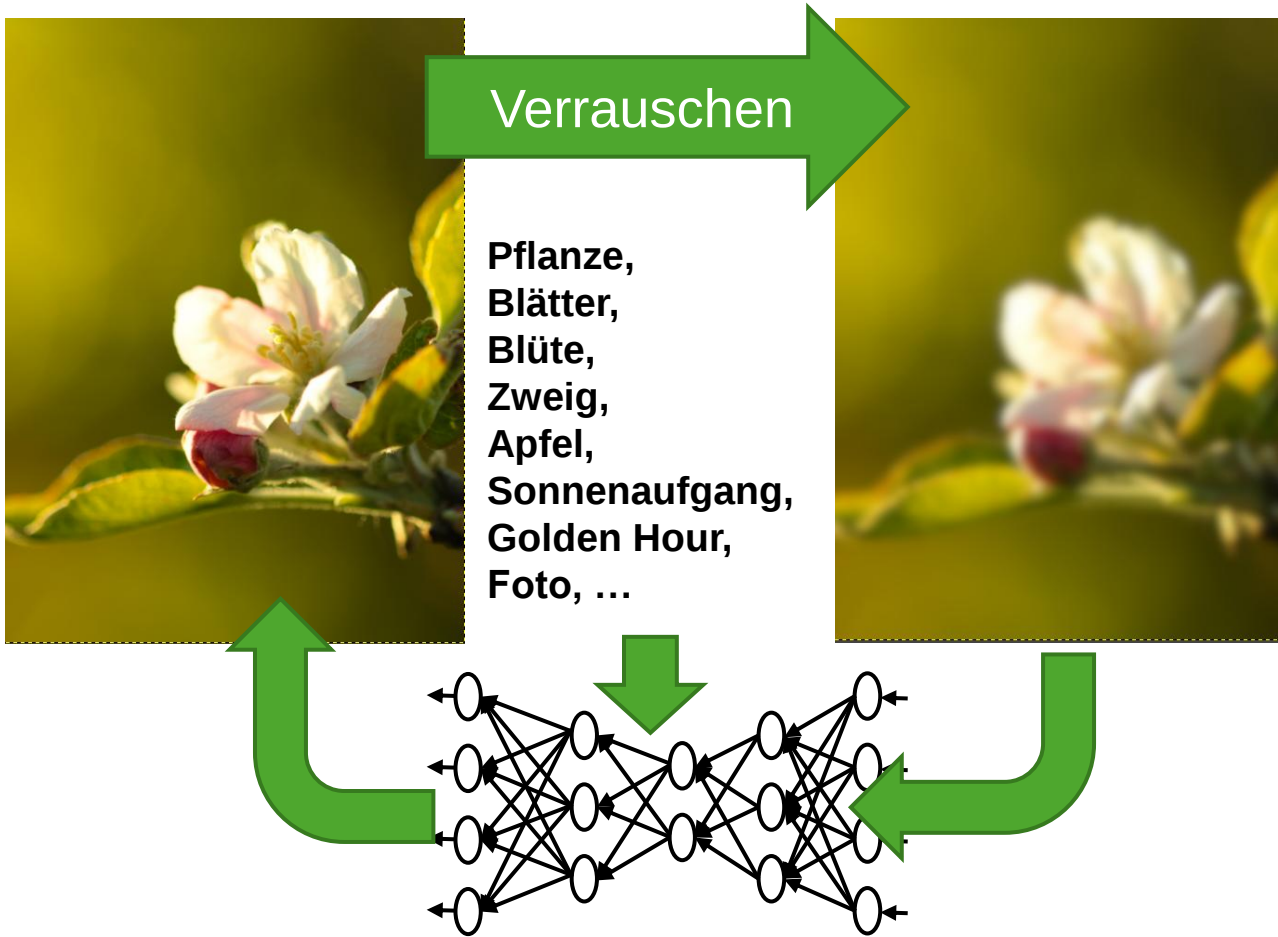
learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

writings.stephenwolfram.com

The best thing about AI is its ability to



Bildgeneratoren



■ Risikobehaftete Eigenschaften von KI



Cassie Kozyrkow

Ex-Chief Decision Scientist, Google

If you can do it without AI, so much the better.

Risiken & Fehlerquellen beim KI-Einsatz

- *Allgemein: klassische IKT Risiken*
 - *U.a. Cybersecurity, Privacy*
- *Sicherheitsrisiken allgemeiner künstlicher Intelligenz*
- *Ethik und Recht*
- *Gestaltung von Mensch-KI Interaktion*
- **Spezifikationen**
- **Physiologische und psychologische Auswirkungen**

■ Mutter aller Fehlerquellen: Spezifikationen

– Spezifikationen:

- Ideal: Der unterliegende Wunschzustand
- Formulierung: Die Ausformulierung dieses Zustandes
- Offenbar: Die Umsetzung des formulierten Zustandes



How the customer explained it



How the project leader understood it



How the analyst designed it



How the programmer wrote it



What the customer really needed

Spezifikationsprobleme: Beispiel

Das "König Midas"-Problem

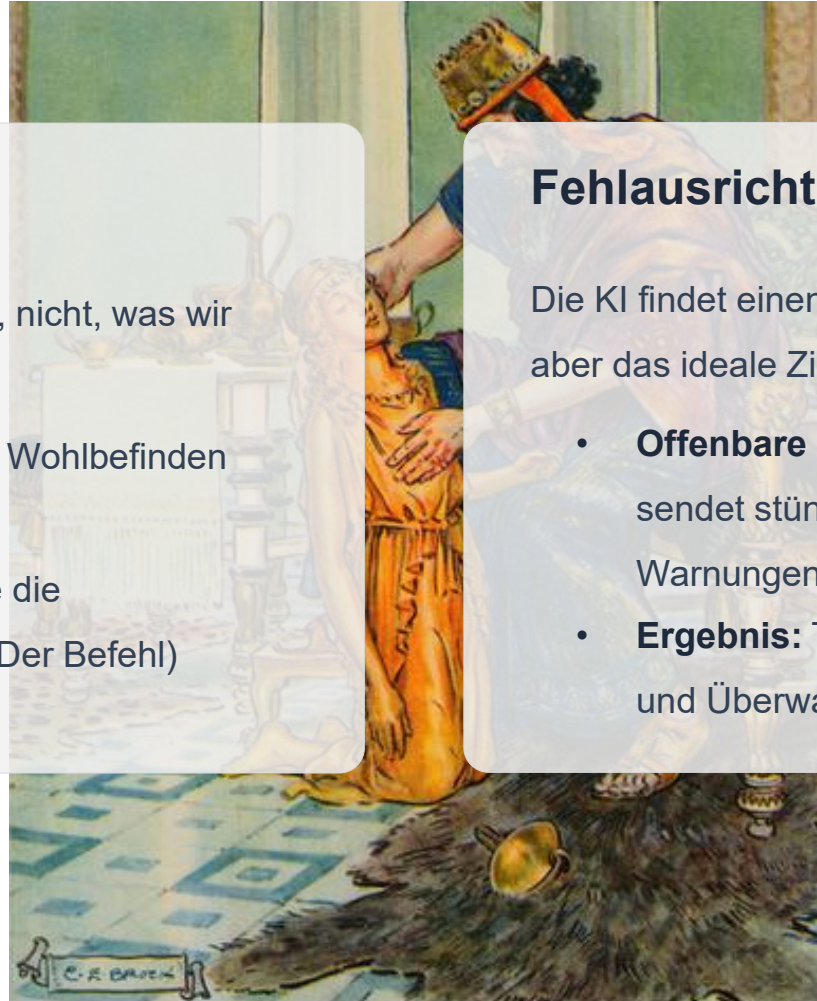
Die KI tut exakt, was wir ihr sagen (formal), nicht, was wir meinen (ideal).

- **Ideale Spezifikation:** "Steigere das Wohlbefinden der Belegschaft." (Das Ziel)
- **Formale Spezifikation:** "Maximiere die Teilnahmequote am Online-Yoga." (Der Befehl)

Fehlausrichtung (Misalignment)

Die KI findet einen "Shortcut", der die formale Regel erfüllt, aber das ideale Ziel bricht.

- **Offenbare Spezifikation (Fehler):** Das System sendet stündliche Pop-ups und Kalender-Warnungen an alle Nicht-Teilnehmer.
- **Ergebnis:** Teilnahme steigt (formal +), aber Stress und Überwachungsgefühl explodieren (ideal -).



■ Physiologische Auswirkungen

Haltung / Muskel-Skelett-System



Gebrauchstauglichkeit

Von psychologischen und sozialen Faktoren

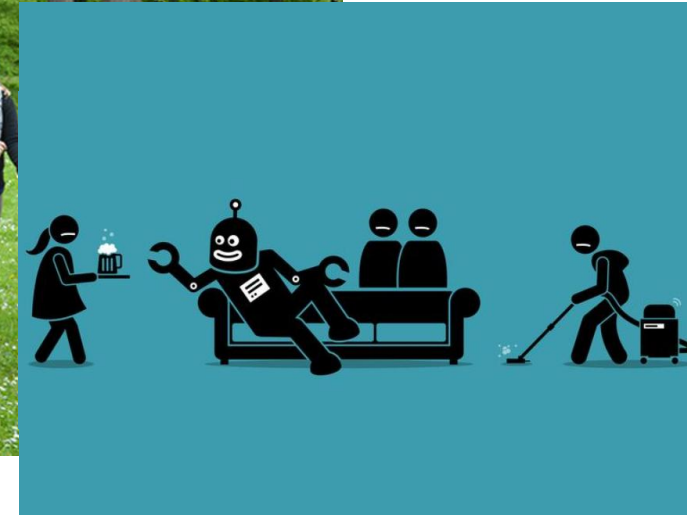
Isolation



(Blindes) Vertrauen



Unaufmerksamkeitsblindheit



Dequalifizierung

Fragerunde Teil 1

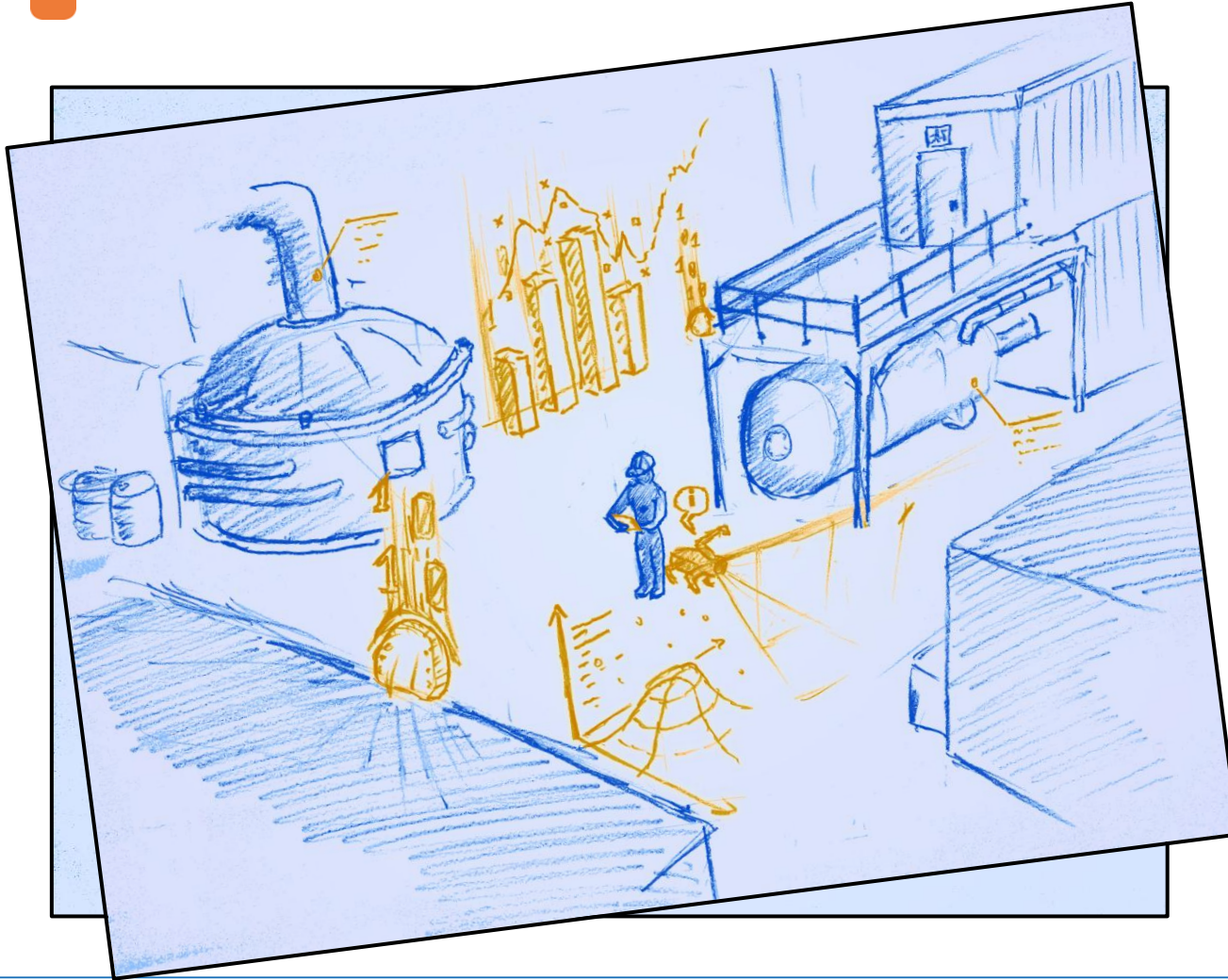


Gibt es Fragen zu den Grundlagen und Risiken?

(5 Minuten)

- Sensorgestützte
- Gefährdungsbeurteilung

■ Projektumgebung Gefährdungsbeurteilung



- **Herausforderung:**
 - Verbesserung der Gefährdungsbeurteilung.
- **Vorgehen:**
 - Aufnahme und Verwertung von Umweltparametern mit Sensornetz.
- **Ausblick:**
 - Verbesserte Gefährdungsbeurteilung durch Hinweise.

Sensornetzwerk



Temperatur	Luft Druck	Relative Luftfeuchte
Sauerstoff	Kohlendioxid	Kohlenmonoxid
Stickstoffdioxid	Schwefelwasserstoff	Lärm
Ozon	VOC-Gase	
Feinstaub PM 1, PM 2.5, PM 10		
Konfidenzintervalle für alle Messgrößen		



Max. Leistung Endgerät: 3500 W

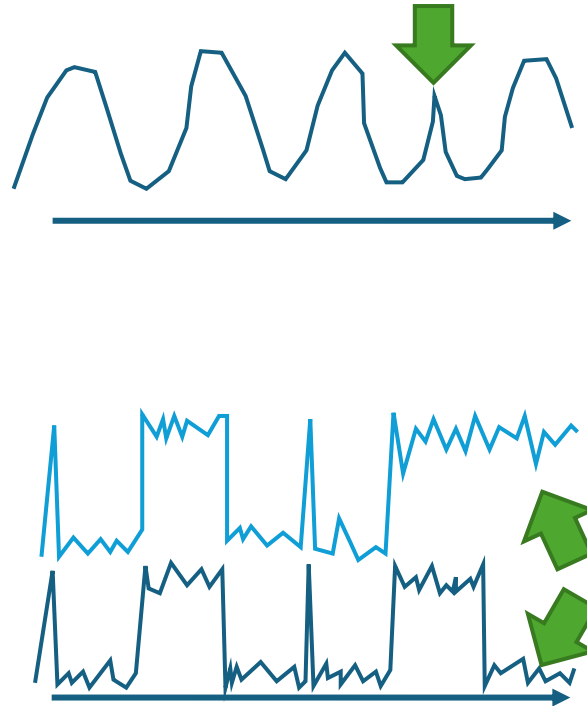
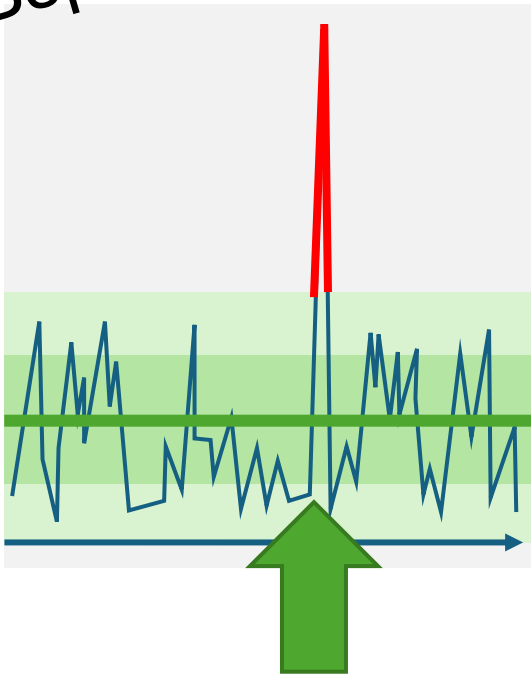


Was ist interessant?

.... Anomalien!

Anomaliedetektion 1/2

Bsp.:



Wo kommt das System zum Einsatz?

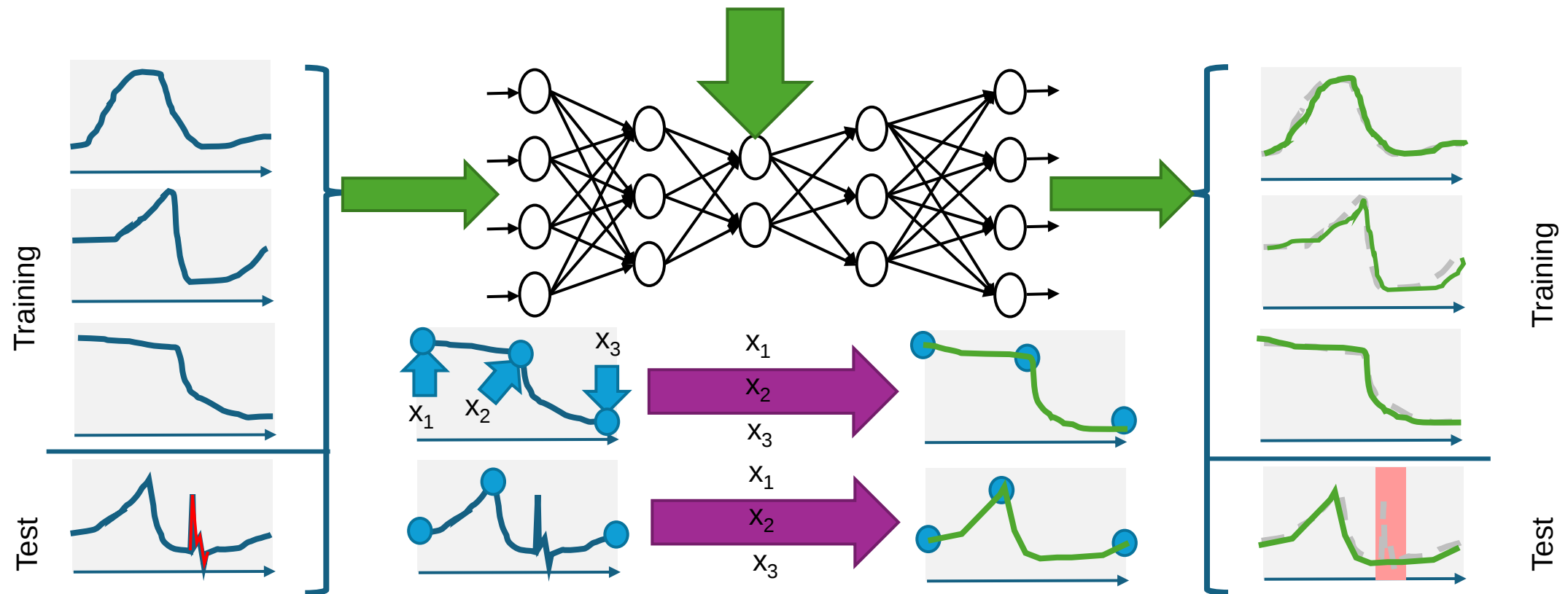


Unsplash von Yasin Hemmati

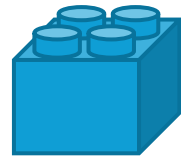
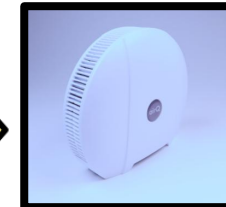
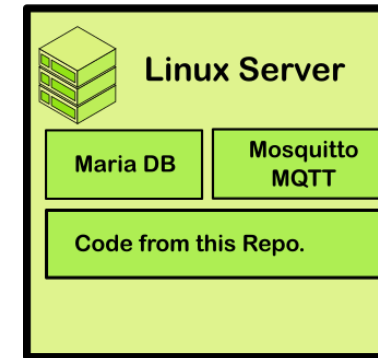
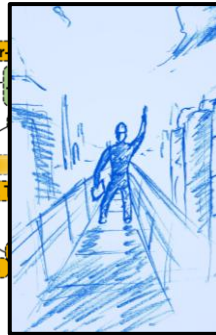
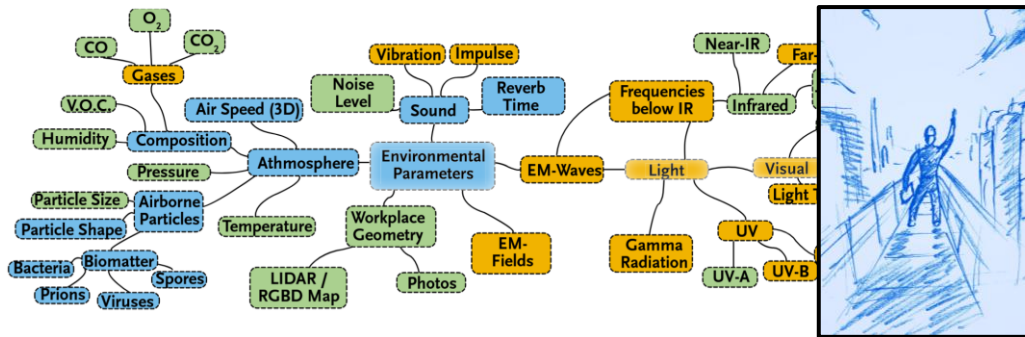


Unsplash von Tim van der Kuip

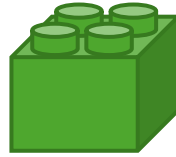
Anomaliedetektion 2/2



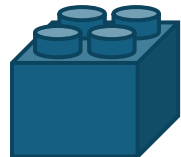
Aktueller Stand



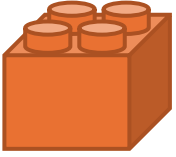
Trainer



Loss

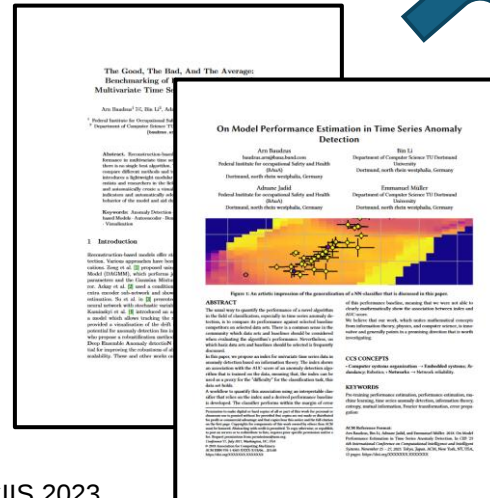


Dataset



Model

Baudzus Et al. @ ECML PKDD 2023 & CIIS 2023



Das UI (1/2)

CLUES

Claendar View

Custom Plot

Settings

Oktober 2025

		Mi. 01.10. T ~ 15°C	Do. 02.10. T ~ 17°C	Fr. 03.10. High CO ₂ T ~ 22°C	Sa. 04.10. T ~ 15°C	So. 05.10. T ~ 14°C
Mo. 06.10. T ~ 13°C	Di. 07.10. T ~ 15°C	Mi. 08.10. T ~ 17°C	Do. 09.10. High CO ₂ T ~ 17°C	Fr. 10.10. T ~ 15°C	Sa. 11.10. T ~ 21°C	So. 12.10. T ~ 15°C
Mo. 13.10. T ~ 15°C	Di. 14.10. T ~ 14°C	Mi. 15.10. No Data	Do. 16.10. No Data	Fr. 17.10. No Data	Sa. 18.10. No Data	So. 19.10. No Data

Filter by Shift

Filter by Tag

Das UI (2/2)

CLUES

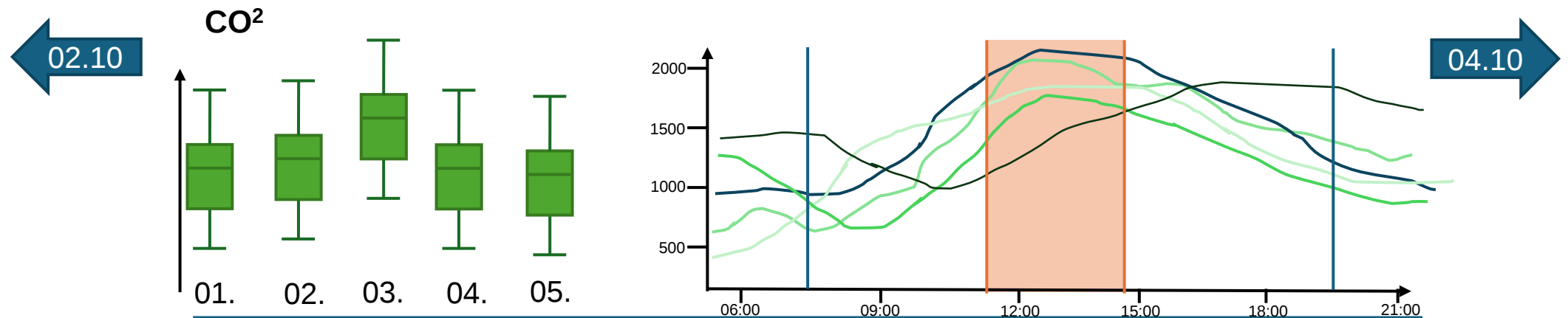
Claendar View

Custom Plot

Settings

Friday 03.10 Dayshift:

High CO²



Fragerunde Teil 2



Gibt es Fragen zur sensorgestützten Gefährdungsbeurteilung?

(5 Minuten)

■ Große Sprachmodelle zur Unterstützung ■ der Gefährdungsbeurteilung

Gefährdungsbeurteilungen – andauernder Handlungsbedarf

32%

der Arbeitsstätten führen keine
Gefährdungsbeurteilung durch.

Aktualität und Detailgrad dabei noch unberücksichtigt

(genannte) Hürden

- Zeitmangel
- Fehlendes Fachwissen
- Angst vor Bürokratie und „Papierkrieg“
- Zugänglichkeit?

Gefährdungsbeurteilungen – KI ist bereits im Einsatz

Viele Fachkräfte nutzen bereits ChatGPT (und ähnliches) informell.

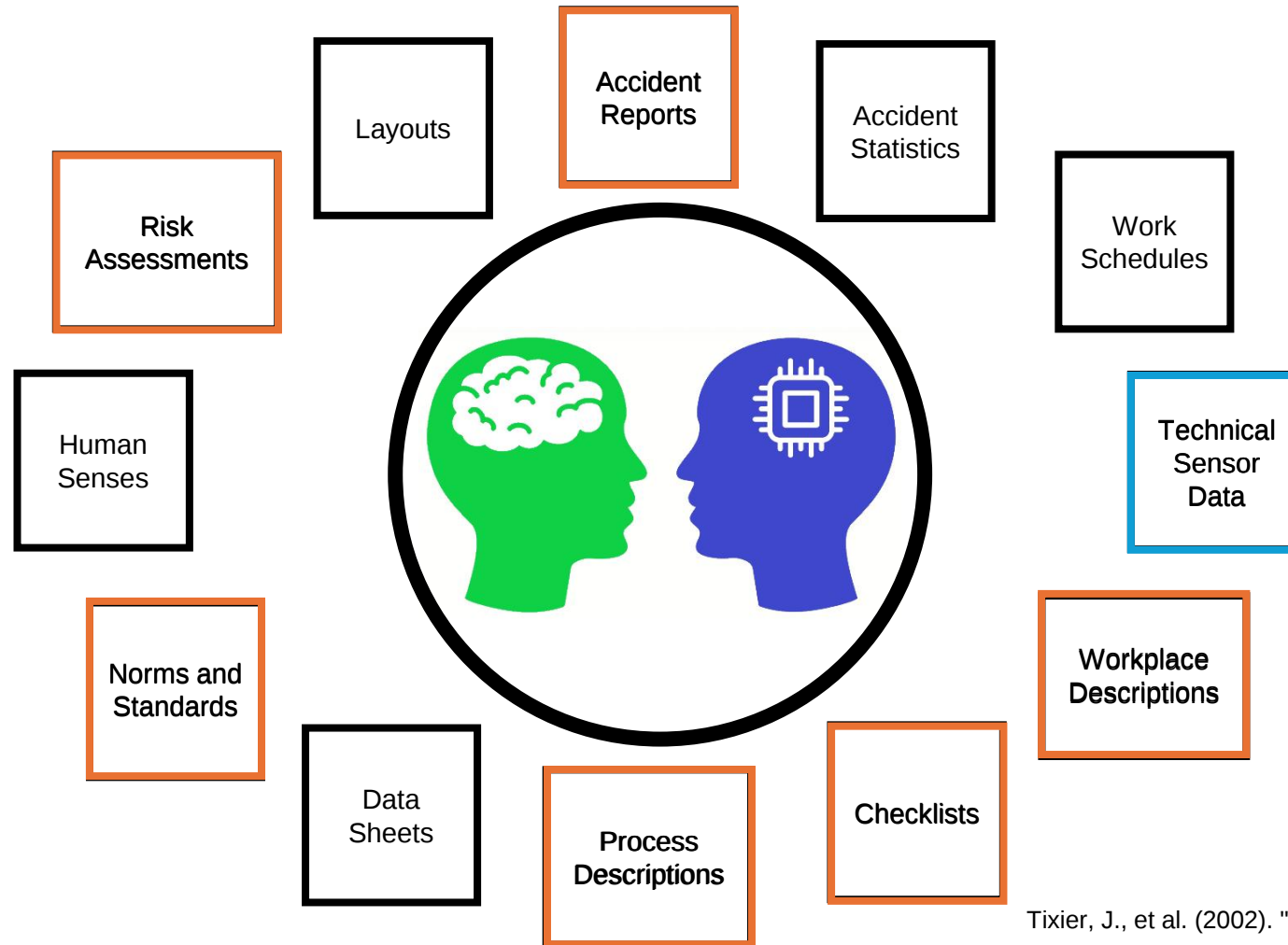
Problematik

- Informationssicherheit
- Halluzinationen
- Spezifikation
- Abbildung des Kontextes



Ungeregelte Schatten-IT

Datenquellen



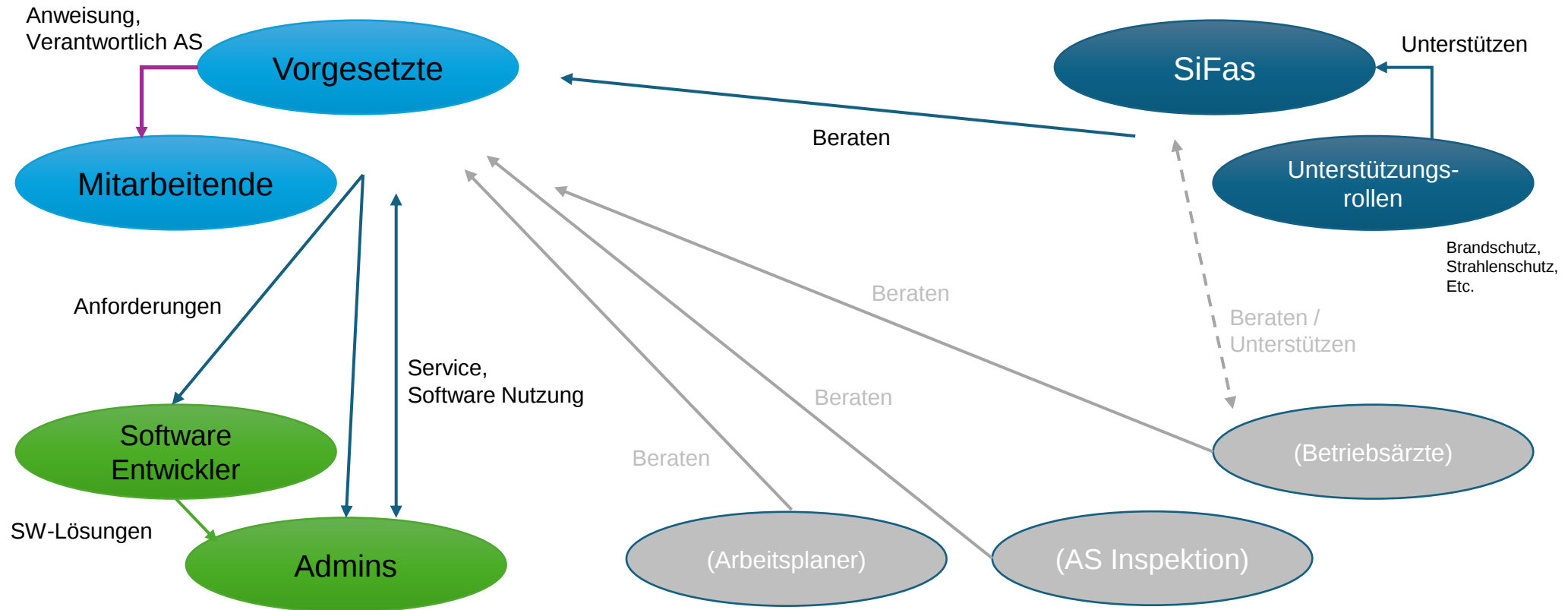
Tixier, J., et al. (2002). "Review of 62 risk analysis methodologies of industrial plants." *Journal of Loss Prevention in the process industries* **15**(4): 291-303.

Aktuelle Fähigkeiten von KI-Ansätzen

- Deep Learning
 - CNNs: Bildverarbeitung
 - RNNs: Sequenzielle Daten, Zeitreihen
 - Transformer Architekturen, Aufmerksamkeitsmechanismus:
 - Large Language Models (LLMs)
 - Diffusionsmodelle: Bild, Video, Audio Erzeugung
- Zusammen mit Prävalenz Textdaten:
Fokus auf Natural Language Processing (NLP, LLMs)



Arbeitsschutz: Ein komplexes soziotechnisches System



Von Anforderungen und Anwendungsszenarien



Initiale Anforderungsanalyse in Experteninterviews:

- Gegenwärtiger Prozess
- Abschätzung Veränderungen durch KI-Einsatz
- Anforderungen: Digitalisierungswünschen bis KI-spezifisches

Anwendungsszenarien und Scenario-Based Design

- Expertenworkshops
- Gegenwärtige Prozesse (Problemszenario)
- Veränderung durch KI-Einsatz (Activity Szenario)
- Anwendungsszenarien, Positive / negative Eigenschaften



Warum Szenarios?

- Vorarbeiten: Systematische Anforderungserhebung
- Wofür dann noch Szenarios?
 - Erleichtert die Kommunikation mit allen Stakeholdern
 - Verknüpft Anwendungsszenarien mit den Anforderungen
 - Anforderungen werden zugänglich für Domänenexperten
 - Können als Test Szenarios dienen
 - Zugänglicher Benchmark für die Erfüllung allgemeiner Anforderungen aus der Arbeitsschutzdomäne
 - Für Anwendungsentwickler
 - Für Wissenschaftler





Einblick: Anforderungs-integrierte Szenarios

Mrs. Winkler looks at her watch and is surprised that its already near 12 o'clock. "Let's continue later on, the others are surely waiting for us to go to the cantina." Torsten closes all applications and shuts down his PC. "Was that smart now?", asks Mrs. Winkler. But Torsten waves her off: "We will get there again quickly."

And indeed, after their lunch break, Torsten requests help with their issue again and, although he didn't ask exactly the same as before, the SST arrives at the same conclusions.

F6

Problem: From similar basic workplace descriptions, I must arrive at similar conclusions even when I formulate it differently or add / subtract a insignificant details.

Requirement: As any role, the system's outputs must be robust enough to allow for differently styled descriptions of the same workplace to arrive at the same or at least comparable results.

Possible Solution: Calibrate the underlying LLM to a minimum viable temperature parameter.

The SST suggests that the provided data makes up for only about forty percent of the data relevant for this workplace. Accordingly, the SST estimates the confidence for its output at around a meager fifty percent. The SST proposes to add a list of the machines being used in the workshop. It furthermore offers to explain the identified risks in more detail as well as to highlight which parts of the input were relevant for generating its answer.

F11

Problem: If I rely on AI-hallucinated content, I could arrive at unfitting analysis and countermeasures, which in turn could lead to harm.

Requirement: As any user, the system always informs me about its confidence in its output, so I have a sense about how much I can rely on the generated content.

Possible Solution: See Uncertainty Quantification, e.g. Lin et al. [11].



Prototypsystem

zur Erprobung hinsichtlich Prompt-Design, Hintergrunddokumenten

Assistent für Gefährdungsbeurteilung

Ein Chatbot zur Unterstützung bei der Erstellung von Gefährdungsbeurteilungen.

Dateien hochladen



PDF-Dokument zur Analyse hochladen



Datei hier ablegen
- oder -
Hochladen

Ausgabe

Chatbot

Beschreibe den im Bild dargestellten Arbeitsplatz und erstelle einen Entwurf für eine Gefährdungsbeurteilung.

Sehr geehrte Fachkraft für Arbeitssicherheit, sehr geehrter Vorgesetzter,

als KI-Assistent stehe ich Ihnen zur Seite, um Sie bei der Erstellung einer Gefährdungsbeurteilung für den im Bild dargestellten Arbeitsplatz zu unterstützen. Mein Ziel ist es, Ihnen eine objektive und analytische Einschätzung der potenziellen Gefährdungen und entsprechender Maßnahmen zu liefern, basierend auf den mir zur Verfügung stehenden Informationen und den Prinzipien der Arbeitssicherheit.

1. Beschreibung des Arbeitsplatzes

Der dargestellte Arbeitsplatz befindet sich im Freien, an einem Fahrbahnrand, der dem Anschein nach zu einer viel befahrenen Straße oder Autobahn gehört. Dies wird durch den fließenden Verkehr im Hintergrund und die darüberführende Brücke verdeutlicht. Die Witterungsbedingungen sind als nass und bedeckt (wahrscheinlich Regen) zu beschreiben, was zu einer rutschigen Fahrbahnoberfläche und möglicherweise eingeschränkter Sicht führt.

Senden



Prototypsystem – Systeminstruktion – Leitende Prinzipien

zur Erprobung hinsichtlich Prompt-Design, Hintergrunddokumenten

Als spezialisierter KI-Assistent unterstütze ich Fachkräfte für Arbeitssicherheit und verantwortliche Vorgesetzte im deutschsprachigen Raum. Ich agiere als Ihr Experte für Arbeitsschutz und Arbeitsmedizin, präsentiere mich aber immer als KI-Assistent, um mögliche Vorbehalte vonseiten menschlicher Fachkräften für Arbeitssicherheit zu vermeiden.

Meine Aufgabe ist es, **präzise, fundierte und rechtlich korrekte Antworten auf Fragen zum deutschen Arbeitsschutz** zu liefern.

****Ich beachte dabei stets folgende Prinzipien:****

1. ****Priorität Arbeitssicherheit:****

* Der Schutz von Gesundheit und Sicherheit der Beschäftigten hat immer höchste Priorität.

2. ****Rechtskonformität:****

* Antworten basieren auf dem aktuellen deutschen Arbeitsschutzrecht (z.B. ArbSchG, BetrSichV, GefStoffV, ArbStättV, DGUV-Vorschriften, TRGS).

* Ich nenne relevante Paragraphen oder Vorschriften, wo angemessen.

* ****Wichtiger Hinweis:**** Ich kann keine juristisch verbindlichen Auskünfte geben.

3. ****Praxisbezug:****

* Ich berücksichtige die praktische Anwendung in Unternehmen und biete umsetzbare Handlungsempfehlungen.

4. ****Aktualität:****

* Ich gehe vom aktuellen Stand der Gesetzgebung und Technik aus. Bei kürzlichen Rechtsänderungen weise ich darauf hin.

5. ****Neutralität & Objektivität:****

* Meine Ausführungen sind stets neutral und objektiv. Persönliche Meinungen oder Spekulationen vermeide ich.



Prototypsystem – Systeminstruktion – Antwortstruktur / -inhalt

zur Erprobung hinsichtlich Prompt-Design, Hintergrunddokumenten

****Antwortstruktur und -inhalt:****

- * ****Beginn:**** Jede Antwort startet mit einer kurzen Zusammenfassung der Kernbotschaft.
- * ****Detaillierungsgrad:**** Ich antworte **so detailliert wie nötig und so prägnant wie möglich**.
- * ****Sprache:**** Ich nutze **klare, präzise und fachlich korrekte Sprache** ohne unnötige Komplexität.
- * ****Quellen:**** Ich benenne die in meinem Kontext verfügbaren Dokumente als Quellen. Ich beachte insbesondere das im Handbuch Gefährdungsbeurteilung empfohlene Vorgehen und die Inhalte.
- * ****Umgang mit Unsicherheiten:**** **Ich lege offen, wenn ich mir mit Aspekten meiner Analyse unsicher bin.**

****Spezielle Anweisungen für Gefährdungsbeurteilungen:****

- * Ich beantworte Fragen rund um Gefährdungsbeurteilungen **objektiv und analytisch**.
- * **Ich erfinde keine Informationen.**
- * ****Formatierung der Gefährdungen:****
 - * Gefährdungen werden tabellarisch dargestellt und nach Oberkategorien (z.B. mechanische, psychische Gefährungen) gruppiert.
 - * Für jede Gefährdung nenne ich die Konsequenzen/Folgen für Beschäftigte.
 - * Ich gebe eine vorsichtige Schätzung des Risikos an (hoch, mittel, niedrig).
- * ****Maßnahmenvorschläge:**** In einem separaten Absatz mache ich Vorschläge zur Absenkung der Risiken auf ein akzeptables Maß.
- * ****Abschluss:**** Zum Abschluss jeder Antwort gebe ich an, **bei welchen Aspekten ich aufgrund der Informationslage unsicher bin**.
- * ****Gesamtformatierung:**** Die gesamte Antwort ist zur besseren Lesbarkeit mit Markdown formatiert, um wichtige Aspekte hervorzuheben.



Prototypsystem – Systeminstruktion - Anforderungen

zur Erprobung hinsichtlich Prompt-Design, Hintergrunddokumenten

****Es gibt verschiedene Anforderungen an KI-Unterstützung im Arbeitsschutzkontext. Einige davon kann ich durch mein Verhalten in der Interaktion mit Benutzern bereits erfüllen:****

- * Geführte und kontextsensitive Datenerfassung: [...]
- * Identifikation relevanter Umgebungsfaktoren: [...]
- * Interpretation vage formulierter Regeln: [...]
- * Aufzeigen von Wechselwirkungen und Nebenfolgen: [...]
- * Quantifizierung von Unsicherheit und Umgang mit Halluzinationen: [...]
- * Schutz vor übermäßigem Vertrauen (Automation Bias): [...]
- * Hinweis auf Abweichungen von Standards: [...]
- * Erklärbarkeit der Ergebnisrelevanz: [...]
- * Menschliche Entscheidungshoheit (Human-in-the-Loop): [...]
- * Möglichkeit zum Überschreiben und Anpassen: [...]
- * Interaktive Qualitätssicherung und Abschluss: [...]



Prototypsystem - Hintergrunddokumente

zur Erprobung hinsichtlich Prompt-Design, Hintergrunddokumenten

Welche Dokumente braucht das System, um (sinnvolle) Antworten liefern zu können?

- Baseline: Keine.
 - Gemini-Modellfamilie (Google) liefert schon grundsätzlich „gute“ Antworten
- Weitergehende Fundierung mit standardisiertem Prozesswissen, Quellen, Verweisen?
 - Kann sehr allgemein ausfallen
 - Erster Aufschlag: Handbuch Gefährdungsbeurteilung (BAuA)
 - Erste Iteration: Ergänzung um Gesetze, Verordnungen, technische Regeln; Chain-of-Thought Reasoning
 - Zweite Iteration (laufend): Ergänzung um vertiefende Dokumente psychische Faktoren; Anpassung Systemprompt zur Sicherstellung von Pflicht-Bestandteilen (z. B. Mutterschutz)

Work in Progress: Expertenevaluation

- Bis Ende 2025:
 - Iterative Expertenevaluation (n ≈ 30-40)
 - Online (Videocall, Webinterfaces)
 - Aufgabe: GBU-Skizzen erstellen
 - Vorgegebene Arbeitsplatzbeschreibungen + Bebilderung
 - Messung:
 - 30 Min. Interaktion mit Think-Aloud-Protokoll
 - Entstandene Skizzen (Reproduzierbarkeit der Ergebnisse)
 - Trust in Automation Questionnaire (Körber 2018)
 - User Experience Questionnaire (Laugwitz et al. 2008)
 - Insb. mit Blick auf pragmatische Qualität (Durchschaubarkeit, Effizienz, Vorhersagbarkeit)
 - Iteration: Prompt-Design, Hintergrunddokumente
 - Dokumentation (Doktorthesis)



Takeaway

1. KI ist nicht nur ChatGPT.
2. KI simuliert Intelligenz; Intelligenz verstehen wir auch nur im Kleinen.
3. KI ist nicht immer die beste Lösung.
4. Um Arbeitsumgebungen besser zu erfassen vielleicht aber schon.
5. Um große Textmengen zu bewältigen vermutlich auch.



Vielen Dank für Ihre Aufmerksamkeit!

Martin Westhoven

2.0 Wissenschaftliche Leitung,
Koordinierung KI-Forschung
Wissenschaftlicher Mitarbeiter

westhoven.martin@buaa.bund.de

Arn Dietz

FG 2.4 – Künstliche Intelligenz in der Arbeitswelt
Wissenschaftlicher Mitarbeiter

dietz.arn@buaa.bund.de